# Licensing high-risk artificial intelligence: toward ex ante justification for a disruptive technology

Malgieri, G.; Pasquale, F.

Computer Law & Security Review 52 (2024) 105899

Contents lists available at ScienceDirect

Computer Law & Security Review: The International Journal of Technology Law and Practice

journal homepage: www.elsevier.com/locate/clsr

# Licensing high-risk artificial intelligence: Toward ex ante justification for a disruptive technology

Gianclaudio Malgieri [a,*], Frank Pasquale [b]

[a] eLaw Center for Law and Digital Technologies, Leiden University, the Netherlands
[b] Cornell Tech and Cornell Law School, United States

ARTICLE INFO

*Keywords:*
AI
Accountability
Justification
GDPR
AIA
Licensing
Regulation

ABSTRACT

The regulation of artificial intelligence (AI) has heavily relied on ex post, reactive tools. This approach has proven inadequate, as numerous foreseeable problems arising out of commercial development and applications of AI have harmed vulnerable persons and communities, with few (and sometimes no) opportunities for recourse. Worse problems are highly likely in the future. By requiring quality control measures *before* AI is deployed, an ex ante approach would often mitigate and sometimes entirely prevent injuries that AI causes or contributes to. Licensing is an important tool of ex ante regulation, and should be applied in many high-risk domains of AI. Indeed, policymakers and even some leading AI developers and vendors are calling for licensure in the area.

To substantiate licensing proposals, this article specifies optimal terms of licensure for AI necessary to justify its use. Given both documented and potential harms arising out of high-risk AI systems, licensing agencies should require firms to demonstrate that their AI meets clear requirements for security, non-discrimination, accuracy, appropriateness, and correctability before being deployed. Under this ex ante model of regulation, AI developers would bear the burden of proof to demonstrate that their technology is not discriminatory, not manipulative, not unfair, not inaccurate, and not illegitimate in its lawful bases and purposes. While the European Union's General Data Protection Regulation (GDPR) can provide key benchmarks here for ex post regulation, the proposed AI Act (AIA) offers a first regulatory attempt towards an ex ante licensure regime in high-risk areas, but it should be strengthened through an expansion of its scope and substantive content and through greater transparency of the ex ante justification process.

## 1. Introduction

Regulating AI is difficult. Complex technology, under-resourced regulators, substantial economic consequences, and high risks for fundamental rights all contribute to this difficulty. Thanks to the well-recognized "black box" problem, identifiable AI abuses are only the tip of an iceberg of problems.[1] AI systems can be opaque, nonlinear, and unpredictable, and they evolve rapidly. This makes it difficult to keep ex post, reactive regulations up to date with the latest technological advances. Years-long litigation will also often fail to set relevant

* Corresponding author.
*E-mail address:* g.malgieri@law.leidenuniv.nl (G. Malgieri).

[1] Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard Univ Press 2015).

https://doi.org/10.1016/j.clsr.2023.105899

Available online 6 November 2023

**Table 1**
A comparison between the governance safeguards for AI systems in the GDPR, in the draft AI Act and the Licensure model proposed in this paper.

| What the GDPR already provides for AI systems | What the EU AI Act (EC version) will require | AI licensure model based on ex ante justification |
|---|---|---|
| Every AI system processing personal data needs to be compliant with the GDPR principles (Article 5), including fairness, lawfulness, transparency.<br>If an AI system is used to make automated decisions, the processing (Article 22) must be based on consent, contractual necessity or EU or national law, and it must have specific safeguards for data subjects (at least the right to contest, the right to express one's view, and to obtain human involvement) and specific transparency measures (meaningful information about the logics, the significance and the envisaged effects) (Articles 13–15). | A conformity assessment is necessary prior to the commercialisation of high-risk AI systems (specifically listed in Annex III)<br>The conformity assessment (Article 19) must include the implementation of a risk management system (Article 9), accurate data governance principles (10), the duty of technical documentation (11) and record keeping (12), transparency measures (13), comprehensive human oversight duties (14), and accuracy, robustness and cybersecurity standards (15) | The Licensure model is based on a public authorisation for commercialisation of AI systems, based on an evaluation of an ex ante justification statement.<br>The ex ante justification is based on general principles similar to the GDPR principles (including fairness), but it would apply to any AI system (even if the GDPR does not apply).<br>The ex ante justification process (or a summary of its results) should be made available for public scrutiny as an accountability measure and be a basis for periodic reconsideration of risks. |

precedents and standards before major damage occurs. Meanwhile, many AI developers either lack legal expertise, or ignore potential legal problems, and they often have vastly more resources than the authorities supposedly monitoring and regulating them.

These asymmetries cause many problems, pressuring governments to prioritize innovation (however destructive its effects) at the cost of fundamental sacrifices of societal values.[2] Since jobs and growth are often far easier to quantify than, say, the negative effects of discrimination or disinformation (amongst the many harms unregulated AI can cause), inadequate regulations and enforcement are endemic to the field. In addition, AI regulatory frameworks cannot guarantee a good level of accountability of AI providers if they foresee small fines in case of AI misuse. A small dent in profits is not enough to deter bad behaviour; rather, it is treated as a cost of doing business. This can incentivize companies to take risks with their AI systems and prioritize profits over safety and ethical considerations. This would be understandable if AI were only a concern of a small number of scientists and laboratories. But it is now evident that the use of AI in business, policing, administration, and beyond, poses high risks to fundamental rights, such as privacy and equality, and can perpetuate and even amplify biases and discrimination, which can have a significant impact on individuals in a situation of vulnerability.[3]

This paper will criticise policymakers' over-reliance on ex post legal measures, including fines and penalties, and will advocate for AI licensure, taking inspiration jointly from the European Union's General Data Protection Regulation and the proposed AI Act, but going well beyond these approaches. Their approach might not prevent harm from occurring in the first place. A more proactive approach, ex ante rather than ex post, would require companies to meet certain safety and ethical standards before deploying AI systems, would be more effective in preventing harm and ensuring accountability. While the GDPR has essential principles for AI justification (including fairness and purpose limitation), it is generally more based on an ex post approach, since there is no requirement for prior administrative authorisation for high risk data processing. On the other hand, the proposed AI Act is based on an ex ante model (conformity assessment before commercialisation), but that model might prove limited in its scope (the rigid list of high-risk AI systems might be not adequate), substance (the proposed draft does not refer to, e.g., a fairness principle) and transparency (there is no duty to disclose the ex ante justification statement to the public).

A key regulatory tool for an ex ante regime is licensure. Under a licensing system, products, services, and activities are unlawful until the entity seeking to develop, sell, or use them has proven otherwise. High-risk AI's documented and potential harms indicate a strong case for a licensure regime here.[4] Under our proposal, to obtain a license, a high-risk AI provider must certify that its AI system meets clear requirements for security, non-discrimination, accuracy, appropriateness, and correctability before it is commercialized.[5] Such a standard may not seem administrable now, given the widespread and rapid use of AI at companies of all sizes. But such requirements could be applied, at first, to the largest companies' most troubling practices, and then gradually to other applications of AI. Under such a regime, AI providers may, for example, be required to demonstrate basic practices of fairness, accuracy, and validity once they have used an AI system in use by, or affecting, over 1 million people.[6] Since government often charges fees for licenses, this

[2] For a recent and suggestive list of such challenges, based on a catalog of harms caused by AI, see Electronic Privacy Information Center, 'Generating Harms: Generative AI's Impacts and Paths Forward,' at https://epic.org/documents/generating-harms-generative-ais-impact-paths-forward/ (2023) (accessed July 23, 2023). Numerous critics of AI have also documented problems of inaccurate, biased, or inappropriate data. See, e.g., Meredith Broussard, *More Than a Glitch* (M.I.T. Press 2023); Cathy O'Neil, *Weapons of Math Destruction* (Vintage 2016); Safiya Noble, *Algorithms of Oppression* (University of California Press 2018); Katharina Zweig, *Awkward Intelligence* (M.I.T. Press 2021); Eric Topol, *Deep Medicine* (2019); Ruha Benjamin, *Race After Technology* (Polity Press, 2019).

[3] Gianclaudio Malgieri, *Vulnerability and Data Protection Law* (Oxford University Press 2023).

[4] For purposes of this proposal, we follow the draft AIA's provisions for defining high-risk, to include both "AI systems that are used in products falling under the EU's product safety legislation [such as] toys, aviation, cars, medical devices and lifts…[and] AI systems falling into eight specific areas that will have to be registered in an EU database," a category comprised of "biometric identification and categorisation of natural persons, "management and operation of critical infrastructure," "education and vocational training," "employment, worker management and access to self-employment," "access to and enjoyment of essential private services and public services and benefits," "law enforcement," "migration, asylum and border control management," and "assistance in legal interpretation and application of the law." European Parliament, EU AI Act: first regulation on artificial intelligence, at https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#:~:text=AI%20systems%20that%20negatively%20affect,cars%2C%20medical%20devices%20and%20lifts., last visited Aug. 3, 2023.

[5] For earlier examples of this kind of move to supplement ex post regulation with ex ante licensure, see Saule Omarova, 'License to Deal: Mandatory Approval of Complex Financial Products' (2012) 90 Washington University Law Review 064; Andrew Tutt, 'An FDA for Algorithms' (2017) 69 Administrative Law Review 83; Pasquale, *The Black Box Society* (n 1) 181; The Federal Communications Commission's power to license spectrum and devices is also a useful precedent here as well. Data may usefully be considered as a public resource. Salomé Viljoen, 'A Relational Theory of Data Governance' [2021] The Yale Law Journal 82.

[6] For an overview of what such practices may entail, see Timnit Gebru and others, 'Datasheets for Datasets' (2021) 64 Communications of the ACM 86; Matthew Zook and others, 'Ten Simple Rules for Responsible Big Data Research' (2017) 13 PLOS Computational Biology e1005399.

system may also prove effective at providing much-needed resources to regulatory bodies now struggling to keep up with the AI revolution.

Our proposal builds on existing scholarship and regulatory proposals and practices. Scholars have argued that certain data practices should not be permitted; licensure would help ensure that they are indeed prohibited.[7] Rather than expecting underfunded, understaffed regulators to overcome monumental black box problems *after* harm has been done, responsibility could be built into the structure of data-driven industries via licensure schemes that require certain standards to be met before large-scale data practices expand even further.[8] Licensure should spur fundamental quality improvements in the realm of product-based and services-based AI, including automobiles, aircraft, logistics, smart infrastructures, financial and employment recommendations, and scoring. There is increasing concern about the validity of the data used in AI and the algorithms it is based on. Rather than addressing all these concerns in an ex post way via tort-based judicial actions or audits and litigation by regulators,[9] the *ex ante* approach of licensure must be part of the regulatory armamentarium. There are some wrongs that can arise out of AI that are too serious to be recompensed ex post.[10]

In addition, a solely ex post approach can create unnecessary risks for fundamental rights of consumers and end-users. Suppose that after a period of time of intensive use of an AI system (e.g., an App) by a massive number of consumers, regulators find that AI-driven app violates the law. A possible sanction might be to block the app and prevent those people to continue using that system. However, considering the period when the app was largely used, people might experience the need of that app, based on a psychological, economic or functional dependency from that AI system. Such harms occurred after the Italian ex post prohibition of Replika[11] and ChatGPT,[12] where many users experienced emotional distress and similarly significant adverse effects after that the Italian DPA prohibited those AI-driven systems. To be sure, the Italian moves here were warranted. Nevertheless, regulators' ex post approach created the paradox that both keeping an AI system in use and prohibiting it risked either harming or reducing the utility of individuals.[13] By contrast, conditioning the burden of proof on AI providers to provide a justification of fairness, safety, non-discrimination, and integrity ex ante would prevent such troubling double binds, and many other problems.

Beyond its value in preventing avoidable harms and double binds for regulators, a licensure regime for AI would also enable citizens to democratically shape technology's scope and proper use, rather than resigning themselves to forces beyond their control. To ground the case for more *ex ante* regulation, Part 2 catalogues the limitations of ex post approaches in the regulation of AI, while Part 3 examines the substantive foundation of licensure models by elaborating a jurisprudential conception of justification. Part 4 addresses the institutional dimensions of our licensure proposal and addresses objections. Part 5 concludes with reflections on the opportunities created by AI licensure frameworks and potential limitations upon them. This paper focuses mostly on the EU law. However, when formulating its proposal, it makes a necessary comparison with other legal systems, where the models of ex ante prohibition and licensures are already a reality or where the legal discussion can already offer some important food for thought.

## 2. Problems of the current frameworks for AI regulation

There are good reasons to be sceptical of artificial intelligence. Tesla crashes have delayed the dream of self-driving cars.[14] Even in areas where AI systems seem to be an unqualified good (as in machine learning to spot melanoma better), researchers worry that current data sets do not adequately represent all patients' racial backgrounds.[15] While machines are proving "better than humans" at some narrow tests, that superiority is fragile, given the dependence of many forms of AI on data sets that change over time.[16] Indeed, CHatGPT"'s performance at identifying prime numbers recently degraded significantly accoring to one study.[17] KataGo (once thought to be superior to any human player of

[7] Siddharth Venkataramakrishnan, 'Top Researchers Condemn "Racially Biased" Face-Based Crime Prediction' *Financial Times* (24 June 2020) <https://www.ft.com/content/aaa9e654-c962-46c7-8dd0-c2b4af932220> accessed 21 January 2022 ("More than 2,000 leading academics and researchers from institutions including Google, MIT, Microsoft and Yale have called on academic journals to halt the publication of studies claiming to have used algorithms to predict criminality. The nascent field of AI-powered 'criminal recognition' trains algorithms to recognise complex patterns in the facial features of people categorised by whether or not they have previously committed crimes."). ; For more on the problems of face-focused prediction of criminality by AI, see Frank Pasquale, 'When Machine Learning Is Facially Invalid' (2018) 61 Communications of the ACM 25, 25.

[8] See also, Data Ethics Commission of the Federal Government of Germany, 'Opinion of the Data Ethics Commission' <https://www.bmj.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN_lang.html> accessed 21 January 2022 (calling for "Preventive official licensing procedures for high-risk algorithmic systems"). The DEC observes that, "[I]n the case of algorithmic systems with regular or appreciable (Level 3) or even significant potential for harm (Level 4), in addition to existing regulations, it would make sense to establish licensing procedures or preliminary checks carried out by supervisory institutions in order to prevent harm to data subjects, certain sections of the population or society as a whole." Id. Such licensing could also be promulgated by national authorities to enforce the European Union's proposed AI Act. ; Frank Pasquale and Gianclaudio Malgieri, 'If You Don't Trust A.I. Yet, You're Not Wrong' *The New York Times* (30 July 2021) <https://www.nytimes.com/2021/07/30/opinion/artificial-intelligence-european-union.html> accessed 21 January 2022.

[9] At the time we write this paper, the EU institutions are discussing the proposed AI Liability regulation. In the EU Parliament, the proposed position seems to be in line with our proposal: a rebuttable presumption of culpability of AI systems users, in case individuals had to allege/claim some damages. Link.

[10] Our model is meant to complement ex post approaches of tort and audit, with ex ante licensure. For more on the importance of audits (whose results could indeed feed into the information necessary for a valid licensing scheme, see Gregory Falco and others, 'Governing AI Safety through Independent Audits' [2021] 3 Nature Machine Intelligence <https://uwe-repository.worktribe.com/output/7562797/governing-ai-safety-through-independent-audits> accessed 21 January 2022.

[11] Garante per la Protezione dei Dati Personali, 'Provvedimento del 2 febbraio 2023 [9852214]' <https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9852214> accessed 1 May 2023.

[12] Garante per la Protezione dei Dati Personali, 'Provvedimento del 30 marzo 2023 [9870832]' <https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9870832> accessed 1 May 2023.

[13] Samantha Cole, '"It's Hurting Like Hell": AI Companion Users Are In Crisis, Reporting Sudden Sexual Rejection' (*Vice*, 15 February 2023) <https://www.vice.com/en/article/y3py9j/ai-companion-replika-erotic-roleplay-updates> accessed 1 May 2023; Natasha Lomas, 'Replika, a "virtual Friendship" AI Chatbot, Hit with Data Ban in Italy over Child Safety' (*TechCrunch*, 3 February 2023) <https://techcrunch.com/2023/02/03/replika-italy-data-processing-ban/> accessed 1 May 2023.

[14] Prescient commentators warned of this possibility. See Meredith Broussard, *Artificial Unintelligence: How Computers Misunderstand the World* (MIT Press 2018).

[15] Angela Lashbrook, 'AI-Driven Dermatology Could Leave Dark-Skinned Patients Behind' *The Atlantic* (16 August 2018) <https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/> accessed 3 May 2022.

[16] Eric Topol, *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again* (Illustrated edition, Basic Books 2019); Gary Marcus and Ernest Davis, *Rebooting AI: Building Artificial Intelligence We Can Trust* (Vintage 2019).

[17] Lingjiao Chen, Matei Zaharia, and James Zou, 'How Is ChatGPT's Behavior Changing over Time?', arXiv:2307.09009v2 [cs.CL] (2023).

Go) was recently defeated at Go by a relatively straightforward trick.[18] There are also many limits to the value of the "reinforcement learning by human feedback" (RLHF) pursued by many systems, particularly as some remote RLHF workers are starting to use AI itself to complete their own tasks (thereby potentially exacerbating, rather than fixing, the biases and errors they were hired to detect and correct).

Even when effective and accurate, AI can have many negative effects and externalities. As AI becomes more prevalent, massive companies are privy to exceptionally comprehensive and intimate details about individuals. Mysterious algorithms predict job applicants' performance based on little more than video interviews.[19] Similar technologies may soon be headed to the classroom, as administrators use "learning analytics platforms" to scrutinise students' written work and emotional states.[20] Financial technology companies use social media and other sensitive data to set interest rates and repayment terms.[21] In short, sectors ranging from transport, financial, retail, health, leisure, and entertainment are all being increasingly affected by AI. Once large enough stores of data are created, there are increasing opportunities to create AI-driven inferences about persons based on extrapolations from both humanly recognisable and ad hoc, machine learning-recognizable groups. Machines as well are increasingly directed by AI. This Part critically discusses some of the current approaches for the regulation of AI. In Section 2.1, self-help, disclosure, and notice and consent approaches are analysed. Section 2.2 drills down on the promise and limits of explanatory AI (XAI). Whatever the merits of extant approaches, they should be complemented by ex ante regulatory approaches based on licensing, at least with respect to some AI applications.

### 2.1. The limits of self-help, notice, and consent

Another simple way to regulate AI in reputational and evaluative contexts is to set a rule that persons must consent to its application before it may be applied. With respect to products, this would likely amount to a mere notification rule. Consumers would be notified if the product they were buying had significant use of AI processes in it and could then decide whether or not to purchase it. Similarly, employers might be required to disclose if they use AI tools in hiring. And litigants could be required to publicly acknowledge their utilisation of such tools, as Pasquale & Cashwell have recommended.[22]

How can a person with a job and family to take care of, try to figure out which of thousands of AI controllers has information about them, has correct information, and has used it in a fair and rigorous manner? In the U.S., even the diligent will all too often run into the brick walls of trade secrecy, proprietary business methods, and malign neglect if they do so much as ask about how their AI has been used, with whom it has been shared, and how it has been analysed.[23] Europeans may make Subject Access Requests (under Article 15 GDPR), but there are far too many AI-gathering and AI-processing companies for the average person to conduct a review of their results in a comprehensive way.[24]

The list of potential targets of disclosure is endless. However, the benefits of disclosure are not nearly as extensive. First, the growing prevalence of AI systems may make the "right" to avoid its use nugatory. Eventually, every automobile may include it, rendering the disclosure a mere notice without the opportunity to act upon it, much as such notices operate in the U.S. medical privacy context. In other words: if it is a near-inevitability that such technologies will be an increasingly important part of the products surrounding us, the question is less how to give individuals a chance to "opt-out," than how to ensure the inevitable accoutrements of their daily lives are functioning in a responsible and accountable manner. This leads to the infamous issue of consent fatigue and consent fallacy in a digitalised world.[25]

This notice-and-consentapproach has undoubtedly some merits, especially in the AI world. Indeed, it is of course vital that consumers are made aware that they are interacting with an AI system or that they are accessing a service (or reading content) that is generated by AI. This is especially true in the field of online newspapers and with any chatbot and virtual assistants. Awareness and consent are vital.to inform expectations of consumers and protect their autonomous choices. However, this approach has also multiple infirmities.[26] Much AI arises out of observation unrestricted by even theoretical contracts. To give an example: a person may be put in situations where it is impractical to "consent" to AI use —for example, when entering another person's car,

[18] Tony T. Wang et al., 'Adversarial Policies Beat Superhuman Go AIs,' arXiv: 2211.00241v4 [cs.LG] (2022).

[19] Drew Harwell, 'A Face-Scanning Algorithm Increasingly Decides Whether You Deserve the Job' *Washington Post* <https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/> accessed 3 May 2022; See also Zoë Corbyn, '"Bossware Is Coming for Almost Every Worker": The Software You Might Not Realize Is Watching You' *The Guardian* (27 April 2022) <https://www.theguardian.com/technology/2022/apr/27/remote-work-software-home-surveillance-computer-monitoring-pandemic> accessed 3 May 2022.

[20] Deborah Lupton and Ben Williamson, 'The Datafied Child: The Dataveillance of Children and Implications for Their Rights' (2017) 19 New Media & Society 780.

[21] Kristin Johnson, Frank Pasquale and Jennifer Chapman, 'Artificial Intelligence, Machine Learning, and Bias in Finance: Toward Responsible Innovation' (2019) 88 Fordham Law Review 31.

[22] Frank Pasquale and Glyn Cashwell, 'Prediction, Persuasion, and the Jurisprudence of Behaviorism' [2018] Faculty Scholarship <https://digitalcommons.law.umaryland.edu/fac_pubs/1604>.

[23] Even in the health care system, where access to such information is supposed to be guaranteed by federal health privacy laws, patients find considerable barriers to the exercise of their rights.

[24] See Jef Ausloos and Pierre Dewitte, 'Shattering One-Way Mirrors – Data Subject Access Rights in Practice' (2018) 8 International Data Privacy Law 4; See also, in general, René Mahieu, 'Right of Access to Personal Data: A Genealogy' [2021] Technology and Regulation 62.

[25] Solon Barocas and Helen Nissenbaum, 'On Notice: The Trouble with Notice and Consent' [2009] Proceedings of the Engaging Data Forum: The First International Forum on the Application and Management of Personal Electronic Information 7; Bart W Schermer, Bart Custers and Simone van der Hof, 'The Crisis of Consent: How Stronger Legal Protection May Lead to Weaker Consent in Data Protection' (2014) 16 Ethics and Information Technology 171; Rossana Ducato and Enguerrand Marique, 'Come to the Dark Side: We Have Patterns. Choice Architecture and Design for (Un)Informed Consent' (Social Science Research Network 2018) SSRN Scholarly Paper ID 3365952 <https://papers.ssrn.com/abstract=3365952> accessed 31 May 2020; Benjamin Bergemann, 'The Consent Paradox: Accounting for the Prominent Role of Consent in Data Protection' in Marit Hansen and others (eds), *Privacy and Identity Management. The Smart Revolution: 12th IFIP WG 9.2, 9.5, 9.6/11.7, 11.6/SIG 9.2.2 International Summer School, Ispra, Italy, September 4-8, 2017, Revised Selected Papers* (Springer International Publishing 2018) <https://doi.org/10.1007/978-3-319-92925-5_8> accessed 4 April 2020.

[26] Julie E Cohen, 'Turning Privacy Inside Out' (2019) 20 Theoretical Inquiries in Law <http://www7.tau.ac.il/ojs/index.php/til/article/view/1607> accessed 23 January 2019; Gabriela Fortuna-Zanfir, 'Forgetting about Consent. Why the Focus Should Be on "Suitable Safeguards" in Data Protection Law' in Serge Gutwirth, Ronald Leenes, Paul De Hert (ed), *Reloading Data Protection* (Springer 2014); Schermer, Custers and van der Hof (n 26).

home, or office. There are also practices that it may be unwise to permit persons to consent to. For example, a driver may freely choose an autonomous vehicle programmed to save the driver in cases of unavoidable tragedy, even if that means taking the lives of many others (imagine, for instance, a car facing an oncoming truck which can only avoid a head-on collision by colliding with a crowd on a sidewalk).

### 2.2. The limits of AI "explanation"

A deeper version of a disclosure approach involves AI explanation. Such a rule would require that vendors not only disclose the presence of AI in a product or service but also explain how it works. Legal scholars and computer scientists have discussed widely *how* to reach a good level of AI explainability and a good level of algorithmic accountability and fairness.

In general, *explaining* decision-making is a complex task.[27] Many commentators have interrogated the notion of explanation in AI in particular.[28] In general terms, explaining means making (an idea or a situation) clear to someone by describing it in more detail or revealing relevant facts.[29] In other terms, the explanation is an act of spotting the main reasons or *factors* that led to a particular consequence, situation or decision.[30] In the field of Computer Science, explanation (of AI) has been referred to as making it possible for a human being (designer, user, affected person, etc.) to understand a result or the whole system.[31] Miller, analysing the structure and expectations of explanations, identified four characteristics of explanations.[32] They are a) *contrastive*, i.e. mostly in response to some counterfactuals ;[33] b) *selected*, i.e. not comprehensive, but based only on the few main factors that influenced the final decision; c) *causal* rather than correlational/statistical; d) *social and contextual*, i.e. depending on the specific social relations and contexts at stake.[34] As affirmed in legal theory, an explanation attempts to render a situation or a process understandable under a *causal, intentional, or narrative* perspective.[35] The causal nature of the explanation is based on the link between cause and effect ("what are the causes behind this decision?"), while its intentional nature is based on the motives of the actor and their beliefs regarding reality ("what are the purposes or intentions behind this decision?"). Considering these two sides of the coin, the explanation is the "answer to the question of *why* something happened or why someone acted as he did." Said in other terms, an explanation is a framework for understanding the action that has happened.[36]

To be meaningful, explanations should go beyond mere description, addressing context, function, intention, and alternative courses of action where possible. Indeed, AI explanations should enable individuals to contest the automated decision[37] and to exercise their rights.[38] This appeared evident also in the CJEU case law. In particular, in the case "Ligue Droits Humains",[39] the Courts affirmed that the opacity of AI technology might make impossible to understand a decision and this might even deprive the data subjects of their individual rights, including the right to an effective judicial remedy. This is why, in our proposal about AI justification below, we believe that explanation and transparency are vital components. Actually, justification and transparency are inextricably related. Not only AI explanation should be part of the AI justification (justifiying why certain AI systems are not black boxes and why they respect the transparency principle), but AI justification statements should be disclosed to the end-users as an act of transparency and so to enable end-users to better exercise their rights.

The GDPR (and in particular the provisions in Article 22 and Recital 71) is often interpreted as referring to only "one" kind of explanation. Actually, there is no unique explanation in practice;[40] each form of explanation highly depends on the context at issue.[41] More importantly, the capability to give a fair and satisfactory explanation depends also on the possibility of showing *causal* links between the input data (and, in particular, some crucial factors within the input information) and the final decision. However, this is not always possible: while for traditionally data-based decision-making, it might be easier to give adequate explanations, addressing the causes, the determining factors and the counterfactuals, in more complex AI-based decisions, it might be hard to reach this high level of explainability. Indeed, looking at the quick development of deep learning in different forms of automated decisions (even COVID-19 automated diagnosis based on, e.g., lung images), explaining the specific reasons and factors of an individual decision might be nearly impossible.[42] An explanation which is neither causal nor

[27] Charles Tilly, *Why?* (2008) <https://press.princeton.edu/books/paperback/9780691136486/why> accessed 21 January 2022.

[28] Tim Miller, 'Explanation in Artificial Intelligence: Insights from the Social Sciences' (2019) 267 Artificial Intelligence 1.

[29] 'EXPLAIN', Meaning & Definition for UK English | Lexico.Com' (*Lexico Dictionaries | English*) <https://www.lexico.com/definition/explain> accessed 21 January 2022.

[30] Andrew D Selbst and Julia Powles, 'Meaningful Information and the Right to Explanation' (2017) 7 International Data Privacy Law 233.

[31] Clement Henin and Daniel Le Métayer, 'A multi-layered approach for interactive black-box explanations' (2021), in Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15. Proceedings, Part, III. Berlin, Heidelberg: Springer-Verlag. p. 5–19. https://doi.org/10.1007/978-3-030-68796-0_138.

[32] Miller (n 28), 3.

[33] Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' [2018] Harvard Journal of Law & Technology <http://arxiv.org/abs/1711.00399> accessed 16 September 2019.

[34] Miller (n 28), 3.

[35] Aulis Aarnio, *The Rational as Reasonable: A Treatise on Legal Justification* (Springer Science & Business Media 1986).

[36] ibid.

[37] Antoni Roig, 'Safeguards for the Right Not to Be Subject to a Decision Based Solely on Automated Processing (Article 22 GDPR)' (2018) 8 European Journal of Law and Technology <http://ejlt.org/article/view/570> accessed 15 January 2019, 6.

[38] Laurens Naudts, Pierre Dewitte and Jef Ausloos, 'Meaningful Transparency through Data Rights: A Multidimensional Analysis' in Eleni Kosta, Irene Kamara and Ronald Leenes (eds), *Research Handbook on EU Data Protection Law* (Edward Elgar Publishing 2022), 530-571, 537.

[39] *Ligue des droits humains ASBL v Conseil des ministres* [2022] ECJ Case C-817/19. ("given the opacity which characterises the way in which artificial intelligence technology works, it might be impossible to understand the reason why a given program arrived at a positive match. In those circumstances, use of such technology may deprive the data subjects also of their right to an effective judicial remedy enshrined in Article 47 of the Charter, for which the PNR Directive, according to recital 28 thereof, seeks to ensure a high level of protection, in particular in order to challenge the non-discriminatory nature of the results obtained.").

[40] Miller (n 28), 3; Margot E Kaminski, 'Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability' (2019) 92 Southern California Law Review, 1529, 1546-1547.

[41] Clement Henin and Daniel Le Métayer, 'Beyond explainability: justifiability and contestability of algorithmic decision systems' (2022), AI and Society 37 (4):1397–1410. https://doi.org/10.1007/s00146-021-01251-8, 1402.

[42] Ronan Hamon and others, 'Impossible Explanations? Beyond Explainable AI in the GDPR from a COVID-19 Use Case Scenario', *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2021) <https://doi.org/10.1145/3442188.3445917> accessed 27 May 2021. Of course, in present medical practice, such clinical decision support AI is better described as an automated *description* (giving some likelihood that a given lung is infected with COVID based on algorithmic analysis of an image of the lung), rather than an automated *decision* (which would presumably be the course of treatment recommended by a physician on the basis of the AI description and other factors). So it is not clear whether the mere imaging of the lung without more would trigger Article 22 duties to provide meaningful information.

contextual is perhaps inadequate to show to the data subject eventual grounds for challenging the decision and then unsuitable under Article 22(3) of the GDPR.

## 3. Toward justification of high-risk AI

Considering the meaningful limitations and paradoxes of the explanation-consent model based entirely on ex post sanctions, this paper advocates for an ex ante justification model to complement ex post regimes. Complementing ex post with ex ante regulation will better protect fundamental rights and freedoms of individuals impacted by AI systems. This model would also overcome the current difficulties of enforcement bodies, induce a stricter accountability of AI providers and prevent serious harms for individuals and society. As we will see below, the proposed EU AI Act already goes in this direction. However, its limitations induced us to look at alternative complementary models, like the GDPR, which articulate a strong set of broad principles, and delineate comprehensive risk-based accountability duties.

In particular, one potential solution for addressing the limitations of disclosure, notice and consent, and explanation-driven approaches to regulating AI is to draw inspiration from certain elements of the GDPR that focus on the legitimacy and value of data use. While Article 22(3) and Recital 71 of the GDPR mention possible measures for making automated decisions more accountable, such as the right to an individual explanation, they also include other complementary tools, such as the right of contestation, rights to human involvement, and algorithmic auditing. In the context of algorithmic decision-making, various principles and concepts could shape the interpretation of accountability duties, including the fairness principle, lawfulness principle, accuracy principle, the risk-based approach, and the data protection impact assessment model. These provisions suggest that justifying automated decisions is not only more feasible but also more effective and desirable than alternative approaches that have been considered thus far. Justifying AI means not merely explaining the logic and the reasoning behind it but also explaining why it operates in a legally acceptable (correct, lawful and fair) way[43] (e.g., why decisions made by the AI comply with the core of the GDPR and are based on proportional and necessary data processing, using pertinent categories of data and relevant profiling mechanisms).[44]

This justification process will be addressed in the next section. However, at this moment, we can already affirm that justification and explanation complement each other: when explanations are not satisfactory or feasible, the data controller should still implement some alternative accountability tools.[45] Kaminski and Malgieri propose to disclose meaningful information about a Data Protection Impact Assessment (DPIA) on the algorithmic decision-making system. The DPIA, as mentioned in Article 35 of the GDPR, is a process to assess and mitigate the impact of data processing operations on the fundamental rights and freedoms of data subjects.[46] This paper, in addition to that proposal, introduces a practical description of a possible *justification test* on the AI, where `the data controller explains why the algorithm (analysed on the aggregated final effects on different data subjects, but also analysed in its purposes, intentions, etc.) is not unfair, unlawful, inaccurate, and beyond the purpose limitation of relevant data. This Part proposes a shift from disclosure/explanation to justification of AI.

### 3.1. The nature of justification

Before delving into the specifics of a potential justification model and its benefits, it's important to understand the meaning of justification both in general and in the legal and data protection contexts. Generally speaking, justification involves taking action to prove or demonstrate that something - be it a person, action, opinion, etc. - is either just, or right, or desirable, or reasonable depending on the context.[47] Indeed, the meaning of justification varies across different fields and contexts. .[48] For example, in theology, justification refers to the act of declaring or making someone "righteous" in the eyes of God,[49] while in philosophy, it involves proving the validity of a theory, opinion, or approach to a problem based on meta-ethical criteria such as utilitarianism or deontology.[50] In the scientific context, justification means proving that a theory or statement is correct and verified through the scientific method.[51]

Unlike an explanation, which seeks to enhance understanding of why a decision was made, a justification aims to persuade an observer that the decision is "just" or "right" based on different standards of correctness or validity in different fields.[52] Whereas explanations are intrinsic and descriptive because they rely solely on the system itself, justifications are extrinsic and normative because they draw on external references - namely, a "norm" against which the validity of the decision can be evaluated.[53] Therefore, a justification requires two elements: a reference norm and proof that the decision aligns with that norm.

While the proof can adhere to logical reasoning standards, the "norm" depends on the specific context. As demonstrated earlier, the norm can be derived from theological, philosophical, scientific, or legal sources. In legal terms, justification means proving that a particular action or act complies with the current law.[54] Loi et al. argue that the two-dimensional justification - consisting of the norm and proof - should be hybrid in nature.[55] This means that the norms can originate from various sources, such as utilitarian and legal norms. Decision-makers may justify their decisions based on their primary goals, which align with utilitarian norms, such as business objectives. However, they are also required to justify their decisions based on "constraining goals" imposed by the law or other ethical values, such as privacy and fairness.[56] Justifying a decision based on primary goals aims to demonstrate

[43] Kaminski (n 40), 1546.

[44] Gianclaudio Malgieri, '"Just" Algorithms: Justification (beyond) Explanation Of Automated Decisions under the GDPR' (2021) 1 Law and Business.

[45] Lilian Edwards and Michael Veale, 'Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?' (2018) 16 IEEE Security & Privacy 46.

[46] Margot Kaminski and Gianclaudio Malgieri, 'Multi-Layered Explanation from Algorithmic Impact Assessments in the GDPR', *FAT 2020 Proceedings* (ACM publishing 2020).

[47] 'JUSTIFICATION', Meaning & Definition for UK English, Lexico.Com' (*Lexico Dictionaries | English*) <https://www.lexico.com/definition/justification> accessed 21 January 2022.

[48] Luc Boltanski and Laurent Thévenot, *On Justification* (2006), Princeton University Press.

[49] 'JUSTIFICATION', Meaning & Definition for UK English, Lexico.Com' (n 47).

[50] See, in general, Larry Alexander and Michael Moore, 'Deontological Ethics' in Edward N Zalta (ed), *The Stanford Encyclopedia of Philosophy* (Winter 2020, Metaphysics Research Lab, Stanford University 2020) <https://plato.stanford.edu/archives/win2020/entries/ethics-deontological/> accessed 1 December 2020.

[51] Paul K Moser, 'Justification in the Natural Sciences' (1991) 42 The British Journal for the Philosophy of Science 557; Mario Bunge, *Philosophy of Science: From Problem to Theory* (Transaction Publishers 1998).

[52] Or Biran and Courtenay V Cotton, 'Explanation and Justification in Machine Learning : A Survey', (2017) IJCAI-17 workshop on explainable AI (XAI), </paper/Explanation-and-Justification-in-Machine-Learning-%3A-Biran-Cotton/02e2e79a77d8aabc1af1900ac80ceebac20abde4> accessed 26 November 2020.

[53] Henin and Métayer (n 41), 1404.

[54] Aarnio (n 35), 256; Mireille Hildebrandt, *Law for Computer Scientists and Other Folk* (Oxford University Press 2020), 257 and 267.

[55] Michele Loi, Andrea Ferrario and Eleonora Viganò, 'Transparency as Design Publicity: Explaining and Justifying Inscrutable Algorithms' [2020] Ethics and Information Technology <https://doi.org/10.1007/s10676-020-09564-w> accessed 30 November 2020.

[56] ibid.

that the decision is not morally arbitrary, while justifying it based on constraining goals aims to prove its legality.[57]

### 3.2. Legal justification

Coming back to the notion of *legal* justification, scholars have proposed different approaches to it,[58]Judgements and the reasoning behind judicial actsdisclose normative rationales for action and simultaneously function to clarify issues on appeal.[59] In regulatory contexts, an agency must justify the rules it promulgates. As large companies deploying AI increasingly govern aspects of common life,[60] they should expect to see more societal demands that their products provide similar justifications.[61] Before explaining how such reason-giving may be institutionalised, it is helpful to review the special case of legal justification.

In general terms, there are two forms of legal justifications: a strict legal positivist approach (i.e., a valid law in itself is a sufficient justification) and a more balanced approach that concedes the dependence of some degree of legal validity on normative legitimacy (i.e., a justification lies on balance between the letter of the law and other grounds having significance in the decision-making).[62] The second approach might better solve different issues related to the law's open nature and the defeasible nature of legal justification (if additional information is taken into account, the status of a conclusion can change).[63] These considerations are also evident in criminal law, where the "justification" is an exception to the prohibition of committing certain offences that renders a nominal violation of the criminal law lawful and, therefore, exempt from criminal sanctions. In doing so, such a justification balances a general legal norm with other contextual interests at issue.[64]

A desirable justification should not merely show compliance with the "law" but with the *core* of the legal principles, i.e., with the *legality* principle.[65] As we will argue below, the core of data protection in the GDPR is summarised in the data protection principles in Article 5. Accordingly, justifying automated decision-making under the data protection *goals* and *norms* means – at least – showing respect for the principles of data protection in Article 5.

### 3.3. The proposed EU AI Act as a timid example of ex ante justification of AI systems

Interestingly, the proposed EU Artificial Intelligence Act (AIA) is adopting a (timid) form of ex ante approach to the regulation of AI. Firstly, it recognises the potential risks associated with AI systems, including the risks to fundamental rights and safety. In particular, the AIA ex ante prohibits the deployment and commercialization of some forms of AI applications for which the risks to the fundamental rights and freedoms of individuals are considered unbearable (see Article 5). In addition, for high-risk AI systems (listed in Annex III), the AI Act requires a conformity assessment of AI systems "prior to their placing on the market or putting into service" (Article 19(1)). Such a conformity assessment is a process where specific assessment bodies (under the oversight of public regulators, "notifying authorities") verify whether the requirements set out in the AI Act relating to a high-risk AI system have been fulfilled (Article 3(20). These requirements include the implementation of a risk management system (Article 9), of accurate data governance principles (Article 10), the duty of technical documentation (Article 11) and record-keeping (Article 12), transparency measures for companies that will use those AI systems (Article 13), comprehensive human oversight duties (Article 14) and important accuracy, robustness and cybersecurity standards (Article 15). The AI providers need to comply with these principles and "justify" their compliance through a conformity assessment. In case the conformity assessment is not carried out, or is carried out incorrectly or irregularly, the market surveillance authority can "take all appropriate measures to restrict or prohibit the high-risk AI system being made available on the market or ensure that it is recalled or withdrawn from the market" (Article 68(2)) in addition to monetary sanctions (Article 71).

In sum, the main core of the draft AI Act is based on an ex ante "licensure" approach, where the AI providers need to "justify", through some technical documentations, that their system is adequate according to specific principles (transparency, accountability, human oversight, accuracy, security).[66] In sum, the authorisation for the commercialisation of AI systems that might produce high risks for fundamental rights and freedoms of individuals, is conditional to the ex ante justification of those AI systems and a form of administrative pre-authorisation for the commercialisation.

However, this approach can be considered a merely "timid ex ante approach" to regulating AI for three main reasons: the material scope of the conformity assessment, the limited content of the conformity assessment and the non-transparency about the conformity justification (See Table 1 for a comparison between the governance safeguards for AI systems in the GDPR, in the proposed AI Act and in our proposal). Firstly, the conformity assessment mechanism only applies to high-risk AI systems, leaving lower-risk systems largely unregulated. This means that potentially harmful AI systems may still be deployed without undergoing the necessary risk assessments and conformity checks. A clear example is AI systems that can interact with humans, like chatbots (especially when these chatbots can affect the emotions of end-users), or AI systems that can detect emotions or perform biometric categorisation of individuals, or even AI systems producing "deep fakes" (highly realistic images, audios or videos who are generated by the AI). These systems are considered "limited risk" by Article 52 and no ex ante conformity assessment is needed for them. But at least one person's actually consummated suicide has been encouraged by a chatbot, and there are doubtless enormous dangers in the realm of commercial and political manipulation by such AI.

Secondly, the principles for the conformity assessment (Articles 10–15 of the proposed AI Act) are appropriate and reasonable, but they are incomplete, since there is no reference to fairness, privacy, and data protection, and little reference to power imbalances, non-discrimination and protection of vulnerable individuals. As we will see below, the concept of fairness is key when analysing interactions between individuals and powerful AI providers. These AI providers can either strongly influence or affect the daily life of consumers or take significant

[57] Federico Cabitza and others, 'Quod Erat Demonstrandum? - Towards a Typology of the Concept of Explanation for the Design of Explainable AI' (2023) 213 Expert Systems with Applications 118888.

[58] Aarnio (n 35), 256; Arno R Lodder, *Dialaw: On Legal Justification and Dialogical Models of Argumentation* (1999 ed, Kluwer Academic Pub 1999).

[59] Aarnio (n 35), 187.

[60] Elizabeth Anderson, *Private Government: How Employers Rule Our Lives* (Princeton University Press 2017); Frank Pasquale, *New Laws of Robotics: Defending Human Expertise in the Age of Ai* (Belknap Pr 2020).

[61] Anderson (n 72); Frank Pasquale, 'Licensure as Data Governance' [2021] Knight First Amendment Institute at Columbia University <https://knightcolumbia.org/content/licensure-as-data-governance> accessed 21 January 2022.

[62] Aarnio (n 35), 187 and 256.

[63] Lodder (n 58), 8-31.

[64] JC Smith, *Justification and Excuse in the Criminal Law* (Stevens 1989); Donald L Horowitz, 'Justification and Excuse in the Program of the Criminal Law' (1986) 49 Law and Contemporary Problems 109.

[65] Hildebrandt (n 54), 267.

[66] See, in general, Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act — Analysing the Good, the Bad, and the Unclear Elements of the Proposed Approach' (2021) 22 Computer Law Review International 97.

decisions that can have profound impact on their fundamental rights and freedoms. Respecting individuals' interests and expectations and preventing adverse effects on them is vital in this context. Regarding data protection, although the GDPR applies to all AI systems using personal data, the success of the conformity assessment under the proposed AI Act is not conditional to the proof of GDPR compliance. While there is a whole article about data and data governance (Article 10), it does not refer directly to the EU data protection principles (as stated in the GDPR and in related legislation).

Thirdly, as argued below, justification should have not a merely pre-authorisation role, but also an accountability role with an instrumental function. End-users (as well as the information intermediaries they rely on) should receive clear indications on why a system is not unfair, not discriminatory, not inaccurate, and not manipulative or harmful. This will be necessary not only for better transparency of the system, but also to enable individuals to exercise rights, to contest algorithmic decisions, to lodge complaints to competent supervisory authorities or to start judicial proceedings (e.g., for civil liability claims). In the proposed AI Act, all technical documentation referring to the ex ante conformity assessment (what we can call here a "justification" statement) is under strict confidentiality (see Article 70) and cannot be disclosed to the public. This means that the "justification statement" will be deprived of its functional role and cannot be used to enhance transparency and contestability and to strengthen the agency of affected users.

These problems can be solved through a GDPR-orientated reform of AI regulatory efforts. The limited scope of the justification process can be extended to all AI systems (as Article 5 GDPR applies to all data processing activities), the principles for compliance can be complemented with the data protection principles, and the transparency of the justificatory process can be already a reality in light of the GDPR provisions. Thus, our proposal applies GDPR principles of ex ante compliance assurances (articulated in Articles 5 and 44 of the GDPR, inter alia) as an ex ante requirement for all AI systems before their commercialisation. Before diving into this proposal, we should first clarify what "justification" means from the perspective of the GDPR principles.

### 3.4. Justification of data processing in the GDPR

In the GDPR, we observe several references to the justification of data processing in general, and of automated decision-making in particular. In different parts of the GDPR, when there is a prohibition (e.g., the prohibition to repurpose the data processing as stated in Article 5(1)(b); the prohibition to process sensitive data as stated in Article 9(1); the prohibition against conducting automated decision-making as stated in Article 22(1); the prohibition of transferring data outside the EU as mentioned in Article 44, etc.), there is always a list of exceptions, often accompanied by some safeguards to protect fundamental rights and freedoms of the data subject. This combination of exceptions and safeguards is the basis of what we can consider a *justification*. In addition, in these cases, the GDPR often refers to the "principles of data processing" as the overarching norm or goal that the data controller needs to comply with in order to *justify* the legality of some nominally illegal acts (see, e.g., Recital 72 about profiling or recital 108 about data transfer).[67]

We might observe another strong example of justification in the GDPR: it is the case of high-risk data processing (Article 35). Under the Data Protection Impact Assessment (DPIA) model, data controllers must prove the legal proportionality and necessity of the data processing and, thus, the legal necessity and proportionality of eventual automated decisions taken (Art. 35(7)(d)). This may constitute a form of *justification of data processing* on the basis of the "core" of data protection.[68]

In addition, the Article 29 Working Party Guidelines on profiling recommend that data controllers (in order to comply with Articles 13–15) explain the pertinence of categories of data used and the relevance of the profiling mechanism.[69] Assessing whether the data used are pertinent and the profile is relevant for a decision, as well as assessing the necessity and proportionality of the data processing in an automated decision-making system seems to constitute a call for justification. The purpose of such assessment is not just transparency about the technology and its processes but a justification about the lawfulness, fairness, necessity, accuracy and legitimacy of certain automated decisions.[70]

Interestingly, empirical research revealed that the justification of algorithms (defined as showing the fairness of goals and rationales behind each step in the decision) is the most effective type of explanation in changing users' attitudes towards the system.[71]

It is interesting to notice that, even before the GDPR, the EU Data Protection Directive already provided for a system of ex ante authorisation of certain data processing activities at high risks. Recital 54 gave the Member States the possibility to foresee that Data Protection Authorities could give ex ante authorisations. As an example, Article 36 of the Italian Data Protection Law implementing that directive stated that sensitive data could be processed only after the prior approval of the national DPA.[72] The French data protection law had a whole Section dedicated to "authorisation" of the CNIL (the French DPA) for special cases of high-risk data processing activities (e.g., automatic processing of genetic or biometric data, automatic processing leading to decisions about social and economic conditions of people, etc.).[73]

### 3.5. Specific grounds for AI justifications in the GDPR

While some scholars have already addressed the need for justification of automated decision-making (rather than a mere need for explanation), very few authors tried to clarify *what* this AI justification should be and *how* it should be conducted under the GDPR rules. This article argues that, considering the meaning of "legal justification" as mentioned in the previous sections, justifying an algorithmic decision should lead to *proving the legality of that decision*. For "legality", we mean not just lawfulness but also accountability, fairness, transparency, accuracy, integrity, and necessity.

[67] Malgieri, '"Just" Algorithms: Justification (beyond) Explanation of Automated Decisions under the GDPR' (n 44), 20.

[68] Dariusz Kloza and others, 'Data Protection Impact Assessment in the European Union: Developing a Template for a Report from the Assessment Process' (LawArXiv 2020) DPiaLab Policy Brief <https://osf.io/7qrfp> accessed 1 December 2020.

[69] Article 29 Working Party, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679' (2017), 31.

[70] Kaminski and Malgieri (n 46), 77.

[71] Biran and Cotton (n 52), 1; Kaminski (n 40), 1549; Tom R Tyler, 'Procedural Justice, Legitimacy, and the Effective Rule of Law' (2003) 30 Crime and Justice 283, 283.

[72] Decreto Legislativo 30 giugno 2003 , n. 196, Art. 26 "(Garanzie per i dati sensibili) 1. I dati sensibili possono essere oggetto di trattamento solo con il consenso scritto dell'interessato e previa autorizzazione del Garante, nell'osservanza dei presupposti e dei limiti stabiliti dal presente codice, nonché dalla legge e dai regolamenti".

[73] See Article 25 of Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, Modifié par Loi n°2004-801 du 6 août 2004 - art. 4 JORF 7 août 2004.

In the last years, scholars have called for fair algorithms,[74] or accountable algorithms[75] or for transparent algorithmic decisions[76] or, again, for lawful, accurate and secure automated decisions. Justifying AI means calling for algorithmic decision processes that prove to have *all* the aforementioned characteristics and respect the *essence* or the core of data protection.[77] The authors argue that the core of data protection in the GDPR consists of the data protection *principles* in Article 5. Accordingly, *justifying* automated decisions means proving that they comply (or adjusting them in order to comply) with the data protection principles in Article 5.

Interestingly, the principles of data protection seem to lead to the desirable characteristics of automated decision-making, as mentioned above. We will now analyse them one by one, contextualising them to the case of algorithmic decision-making.

Article 5(1)(a) refers to lawfulness, transparency and fairness. As regards *lawfulness*, automated decision-making should be lawful, i.e. having a lawful ground and respecting fundamental rights and freedom. Such a lawful basis should be found not only in Article 6(1) (or in Article 9(2) in case of special categories of personal data) but also in Article 22. Since Article 22(1) is interpreted as a *prohibition* of automated decision-making,[78] in order to make it lawful, it is necessary to prove that one of the exceptions in Article 22(2) (consent, contract, Union or national law) applies, with the related requirements in Article 22(3) (suitable measures to safeguard the data subject's rights, including at least the right to human intervention, to express his or her point of view and to contest the decision). This part of "justification" is the most formal one: the controller needs to *justify* why an activity which is apparently unlawful (profiling individuals or taking significant decisions on automated bases) is instead lawful. In this sense, this part of justification is reminiscent of the legal justification in criminal law, as mentioned above.[79]

As regards *fairness* justification, the data controller should prove that the decision-making processing is fair. The concept of fairness is not well defined in the text of the GDPR nor in authoritative guidelines or in caselaw. Several scholars have suggested the link between fairness and two other principles, namely lawfulness and transparency.[80] Different data protection authorities have referred to fairness as consisting in, at least in part, the respect of expectations of data subjects and as a tool for preventing adverse impacts on them.[81] Compliance with GDPR rules entails respect for data subjects' expectations and preventing adverse effects in practice, beyond mere formalistic compliance. And whatever data subject expectations may be, there is also an enduring role in fairness determinations for principles of non-discrimination, non-manipulation and the prevention of biases in AI-driven data processing.

Thus a fairness justification should require proof that the AI system respects the expectations and the interests of end users, including expectations of non-discriminatory, unbiased, and non-manipulative interactions. Fairness also entails that the AI not exploit a significant imbalance between the controller and the subject in particular contexts (e.g., its treatment of vulnerable individuals).[82] In general, algorithmic processing of information should not violate the expectations of data subjects,[83] and its effects should not impair human dignity, autonomy, safety and other fundamental rights set out in the EU Charter of fundamental rights.[84]

As regards *transparency* justification, the data controller should prove that the algorithmic processing is legible[85] in the sense that, at least, meaningful information about the logic, the significance, and envisaged consequences of the decision-making are communicated to the subject at the beginning of the data processing (Articles 13(2)(f) and 14(2)(g)) and, upon request, after the processing has started (Article 15(1)(h)). Adding the transparency requirement in our justificatory models is not a contradiction of our shift from transparency to justification: explanations and justifications are not alternative elements, but they should be read in conjunction. In other words, transparency is *part* of our proposed justification process, but not the main focus.

There are at least two levels of possible transparency: general (or "global") information, or individual (or "local") explanation

[74] Future of Privacy Forum, 'Unfairness By Algorithm: Distilling the Harms of Automated Decision-Making' (2017) <https://fpf.org/2017/12/11/unfairness-by-algorithm-distilling-the-harms-of-automated-decision-making/> accessed 8 February 2020; Sainyam Galhotra, Yuriy Brun and Alexandra Meliou, 'Fairness Testing: Testing Software for Discrimination', *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering - ESEC/FSE 2017* (ACM Press 2017) <http://dl.acm.org/citation.cfm?doid=3106237.3106277> accessed 31 May 2019; Andrew D Selbst, 'Disparate Impact in Big Data Policing' (2018) 52 Georgia Law Review 109.

[75] Joshua Kroll and others, 'Accountable Algorithms' (2017) 165 University of Pennsylvania Law Review 633.

[76] Bruno Lepri and others, 'Fair, Transparent, and Accountable Algorithmic Decision-Making Processes' (2018) 31 Philosophy & Technology 611; Bilyana Petkova and Philipp Hacker, 'Reining in the Big Promise of Big Data: Transparency, Inequality, and New Regulatory Frontiers' [2016] Lecturer and Other Affiliate Scholarship Series <https://digitalcommons.law.yale.edu/ylas/13>; Mireille Hildebrandt, 'Profile Transparency by Design? Re-Enabling Double Contingency' (2013), in Mireille Hildebrandt and Katia De Vries. 'Privacy, Due Process and the Computational Turn' (Routledge, 2013), 221-247.

[77] Maja Brkan, 'The Essence of the Fundamental Rights to Privacy and Data Protection: Finding the Way Through the Maze of the CJEU's Constitutional Reasoning' (2019) 20 German Law Journal 864.

[78] Article 29 Working Party (n 69), 19; Michael Veale and Lilian Edwards, 'Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on Automated Decision-Making and Profiling' (2018) 34 Computer Law & Security Review 398.

[79] Smith (n 64), 109.

[80] Damian Clifford and Jef Ausloos, 'Data Protection and the Role of Fairness' (2018) 37 Yearbook of European Law 130; Michael Butterworth, 'The ICO and Artificial Intelligence: The Role of Fairness in the GDPR Framework' (2018) 34 Computer Law & Security Review 257; Gianclaudio Malgieri, 'The Concept of Fairness in the GDPR: A Linguistic and Contextual Interpretation', *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery 2020) <https://doi.org/10.1145/3351095.3372868> accessed 29 January 2020.

[81] Datatilsynet, 'Advance Notification of Order to Rectify Unfairly Processed and Incorrect Personal Data - International Baccalaureate Organization' (2020) <https://www.datatilsynet.no/contentassets/04df776f85f64562945f1d261b4add1b/advance-notification-of-order-to-rectify-unfairly-processed-and-incorrect-personal-data.pdf>; Commission National Informatique and Libertés, 'How Can Humans Keep the Upper Hand? The Ethical Matters Raised by Algorithms and Artificial Intelligence, Report on the Public Debate Led by the French Data Protection Authority (CNIL) as Part of the Ethical Discussion Assignment Set by the Digital Republic Bill' (2017); Information Commissioner's Office, 'Big Data, Artificial Intelligence, Machine Learning and Data Protection' (2017) <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>.

[82] Clifford and Ausloos (n 80); Malgieri, 'The Concept of Fairness in the GDPR' (n 80).

[83] Butterworth (n 80), 263.

[84] European Parliament Resolution, 'Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies - Tuesday, 20 October 2020' <https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html> accessed 21 January 2022.

[85] Gianclaudio Malgieri and Giovanni Comandé, 'Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 243.

(implementing recital 71).[86] Each level of transparency should depend on the level of risk of that algorithmic decision-making process.[87] Recently the Advocate General of the CJEU has specified that transparency of automated decision-making should be meaningful and should imply sufficiently detailed explanations on the method used for the decisions and the reasons that led to a certain result.[88] This would mean to give at least aggregate information, in particular on the factors taken into account for the decision-making process and their respective importance at an aggregate level, which is also useful for challenging any decision.

Transparency will also help assure compliance (or at least detection of noncompliance) with general data protection principles. For example, Article 5(1)(b) refers to *purpose limitation,* an increasingly important value as more persons realize that their data is being used to train AI. According to this principle, the justification should also prove that the ADM system is based just on data collected for the specific (licit and declared) purpose of obtaining an automated decision affecting the data subject. Under a broader perspective, the purpose limitation justification should also clarify that the algorithm was not originally developed for other purposes (military, commercial, etc.) and then eventually repurposed for the processing at stake.[89] This would help to prevent algorithmic biases caused by a decontextualisation of algorithms.[90]

Article 5(1)(c) mentions the principle of *data minimisation*. Under this principle, the justification of the data controller should prove that the ADM is based on the processing of only data that are adequate, relevant and limited to what is necessary for the purpose of taking that automated decision. For example, if the controller is an employer that needs to hire a new employee and she declares that the automated decision-making processing has the purpose of selecting the worthiest candidate, any information about, e.g., the sexual orientation, the ethnic origin, the religion or the possibility to take maternity leave (fertility, marital status, etc.), are unnecessary and should not be collected. This might also be a way to prevent intentional discrimination[91] hidden through "masking"[92]: when the data controller tries to cover intentional discrimination behind the shield of data analytics. In such cases, the data minimisation justification could be helpful at revealing opportunistic repurposing of data for illicit ends.[93]

Article 5(1)(d) refers to data *accuracy*. When justifying AI, accuracy is also central. The data controller should prove that the algorithmic decision is based on correct and accurate data. Recital 71 (addressing ADM) requires data controllers to ensure "that factors which result in *inaccuracies* in personal data are corrected, and the risk of errors is minimised" (italics added). Indeed, accuracy has generally been considered one of the main elements to justify the use of certain algorithms.[94] WP29 has referred to inaccuracy as one of the main problems posed by automated decision-making, since these errors in data or in the AI process might result in "incorrect classifications" and "assessments based on imprecise projections that impact negatively on individuals."[95] The European Bank Authority, in its report on advanced analytics, has given great importance to data accuracy for justifying algorithms in the bank sector and has developed that concept through different subconcepts: accuracy and integrity, timeliness, consistency and completeness of data.[96] In our view, a correct and comprehensive application of the accuracy justification should result not only in proving the accuracy of input data but also in proving that the chosen algorithm is fit-for-purpose, i.e. produces accurate results. Indeed, often discriminatory decisions are also inaccurate and incorrect.[97] Empirical studies also confirm that the "usefulness" of an algorithmic decision is a key component in their social acceptance.[98] We are aware that accurate decision-making process can bring to unfair results (e.g., perpetuating social injustice, exacerbating individual or group vulnerabilities), but this principle is just one component in the bigger picture of justification and should be, thus, read in conjunction with the other principles described in this section.

Article 5(1)(e) mentions the principle of *storage limitation*. Although in the field of AI, this principle seems not so pertinent, its function is also important. This principle requires that data should be stored for no longer than necessary for the purpose of the processing. This time limitation should also apply to algorithmic decision-making. In other words, ADM should not be based on data that are no longer necessary (e.g., outdated or inappropriate data) for the purpose and the context of the decision. At the same time, controllers should not use algorithms that are no longer necessary for the declared purposes.

Article 5(1)(f) mentions the principle of *integrity and confidentiality*. In the context of AI, it is central that algorithmic decisions do not lead to cybersecurity risks that could adversely affect the safety (or any other fundamental right or freedom) of the data subject. Recital 71 also indirectly refers to these "risks" when mentioning automated decisions. Cybersecurity, safety and integrity are central elements to consider when justifying algorithms. A "just" algorithm is based on and produces integrous data, is based on clear and safe steps, and does not endanger the (digital or physical) safety of the data subject.[99]

The last principle in Article 5 is *accountability* (Article 5(2)). Accountability of AI is an overarching goal that is considered the final objective of legally desirable AI, in particular in the data protection framework.[100] This is a "meta-principle", i.e., a methodology to apply and implement all the other data protection principles in Article 5. We

[86] Margot E Kaminski and Gianclaudio Malgieri, 'Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations' (2021), International Data Privacy Law, Volume 11, Issue 2, April 2021, Pages 125–144, https://doi.org/10.1093/idpl/ipaa020, 143.

[87] Kaminski and Malgieri (n 86), 140.

[88] *Advocate General Opinion, M P Pikamäe, C-634/21* (ECJ) Par. 58.

[89] Malgieri and Comandé (n 85), 259.

[90] Jonida Milaj, 'Privacy, Surveillance, and the Proportionality Principle: The Need for a Method of Assessing Privacy Implications of Technologies Used for Surveillance' (2016) 30 International Review of Law, Computers & Technology 115.

[91] Pauline T Kim, 'Data-Driven Discrimination at Work', (2017) 58 William & Mary Law Review 3, 857.

[92] Solon Barocas and Andrew D Selbst, 'Big Data's Disparate Impact' (2016) 104 California Law Review 671; Cynthia Dwork and Deirdre K Mulligan, 'It's Not Privacy, and It's Not Fair' (2013) 66 Stanford Law Review 6; Kroll and others (n 75), 682.

[93] Sandra Wachter, 'Affinity Profiling and Discrimination by Association in Online Behavioural Advertising' (2019) 35 Berkeley Technology Law Journal <https://papers.ssrn.com/abstract=3388639> accessed 2 June 2019.

[94] Kroll and others (n 75); Cynthia Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead' (2019) 1 Nature Machine Intelligence 206; Zachary C Lipton, 'The Mythos of Model Interpretability' (2018) 61 Communications of the ACM 36.

[95] Article 29 Working Party (n 69), 27.

[96] Benjamin T Hazen and others, 'Data Quality for Data Science, Predictive Analytics, and Big Data in Supply Chain Management: An Introduction to the Problem and Suggestions for Research and Applications' (2014) 154 International Journal of Production Economics 72.

[97] Julia Dressel and Hany Farid, 'The Accuracy, Fairness, and Limits of Predicting Recidivism' (2018) 4 Science Advances eaao5580.

[98] Theo Araujo and others, 'In AI We Trust? Perceptions about Automated Decision-Making by Artificial Intelligence' (2020) 35 AI & SOCIETY 611.

[99] European Parliament Resolution (n 84).

[100] Sonia K Katyal, 'Private Accountability in the Age of Artificial Intelligence' (2019) 66 UCLA Law Review 88; Kroll and others (n 75); Kaminski (n 40); Lepri and others (n 76).

can identify two perspectives of accountability justification in the GDPR: a practical perspective and a methodological one. The practical accountability justification should lead to a demonstration that the data controller has proactively implemented some suitable ADM measures under Article 22(3) and recital 71,[101] that she is ready to enable data subjects exercise their ADM-related rights (within and beyond Article 22), and that those rights are effective (for example, the right to contest the algorithm, should be made effective through clear information about the system[102] and the decision and there should be concrete technical or organisational steps to take into account the eventual data subjects' contestation, by either complying with it or explaining why such a request is unreasonable).[103]

On the other hand, the methodological perspective of accountability indicates *how* the justification should be conducted, i.e. how the justificatory auditing should be carried out (see the section below) and what the *legal approach* to justification should be. In particular, the accountability principle – as Article 5(2) indicates – puts the burden of proving data processing compliance on the data controller.[104] This means that there is a rebuttable presumption (*praesumptio iuris tantum*) that the data processing activity at stake – and, thus, any ADM processing too – is not compliant with the data protection principles. The burden of proof of legality is on the data controller.[105] In other terms, we should consider that algorithmic decisions are illegal by default unless the data controller *justifies* them through a valid justification, meant both as a process of justificatory auditing and an eventual final justification statement.

## 4. Institutionalising justification via licensure

To summarise, our proposal here is to impose a licensure model on the providers of AI systems. The conformity assessment model for high-risk AI systems in the proposed AI Act is a good starting point. However, considering the limited scope of the conformity assessment (limited to the high-risk systems), the limited transparency of the justification documents produced by AI providers, and the limited principles to which the AI providers should prove compliance in the conformity assessment (limited reference to fairness, data protection, vulnerable users' protection) we propose that the AI Act and any AI regulation across the world might be based on a more comprehensive licensure model based on AI justification. A very good model for justification is offered by the GDPR, which requires the respect of broad principles (fairness, lawfulness, transparency, purpose limitation, accuracy, data storage limitation, integrity and accountability) applying to all data controllers. Auditing for such values and standards should be a critical first step in the licensure process. As Mökander and Floridi have argued, "ethics-based auditing of AI holds the potential to complement and enhance other tools and methods like human oversight, certification, and regulation" like the licensing we propose.

Of course, all these values and goals, as expressed in law, are mere dead letters if they are not realised in an institutional framework for their effective realisation (or progressive realisation, to borrow terminology from the discourse of cultural and social rights).[106] One way to ensure proper justification of AI along the lines developed above is to create mechanisms that promote proper scrutiny occurs before the collection, analysis, and use of the data and algorithms fuelling AI, to be followed by ongoing monitoring of AI's effects and results. If enacted via a licensure regime, this scrutiny would enable a true industrial policy for AI, deterring misuses and thereby helping to channel AI development in more socially useful directions. As AI becomes more invasive and contested, there will be increasing calls for licensure regimes. To be legislatively viable, proposals for licensure need theoretical rigour and practical specificity. Note, too, that licensure is not a substitute for ex post regulation, but a complement. An entity may obtain a license by promising to abide by certain standards, and then abandon them over time. This possibility requires ongoing, ex post actions by regulators when they become aware of scofflaws, and by courts when litigation with a proper basis is pursued before them. A licensure scheme can also adopt some safeguards against ex post abuse. For example, it may require ongoing audits, or license renewals. Several agencies in the U.S. routinely inspect regulated entities to determine how well they are complying with relevant regulations. License renewals are also common for drivers, particularly as they age, so that authorities can ensure they still have the necessary visual and auditory acuity to safely operate a motor vehicle. Both audits and license renewal may be included in a more general licensing scheme in order to stop and/or punish illegal activities.

Cognisant of these queries, some legislators and regulators have begun to develop an explicitly justification-driven approach to AI.[107] While not embracing licensure, U.S. Sen. Sherrod Brown has demonstrated how substantive limits may be enforced with respect to the large-scale data collection, analysis, and use at the heart of so much AI. His proposed Data Accountability and Transparency Act would amount to a Copernican shift in U.S. governance of data, putting civil rights protection at the core of public concern.[108] This reflects a deep concern about the dangers of discrimination against minoritised or disadvantaged groups, as well as against the "invisible minorities" previously

[101] See, e.g., Roig (n 37).

[102] Kaminski and Malgieri (n 46), 77.

[103] Gianclaudio Malgieri, 'Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations' (2019) 35 Computer Law & Security Review 105327.

[104] Information Commissioner's Officer, 'Accountability and Governance' (1 October 2020) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/> accessed 29 November 2020.

[105] Raluca Oprișiu, 'Reversal of "the Burden of Proof" in Data Protection | Lexology' <https://www.lexology.com/library/detail.aspx?g=e9e8c734-23d9-41bb-a723-5d664b3c86cc> accessed 29 November 2020.

[106] Eric Lander, 'Americans Need a Bill of Rights for an AI-Powered World' *Wired* <https://www.wired.com/story/opinion-bill-of-rights-artificial-intelligence/> accessed 21 January 2022.

[107] European Data Protection Board, 'Guidelines 1/2018 on Certification and Identifying Certification Criteria in Accordance with Articles 42 and 43 of the Regulation - Version Adopted after Public Consultation | European Data Protection Board' (2018) 20 <https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-12018-certification-and-identifying_en> accessed 21 January 2022.

[108] Press release, 'Brown Releases New Proposal That Would Protect Consumers' Privacy from Bad Actors | U.S. Senator Sherrod Brown of Ohio' <https://www.brown.senate.gov/newsroom/press/release/brown-proposal-protect-consumers-privacy> accessed 21 January 2022.

described in *The Black Box Society*.[109]

### 4.1. Case studies in health predictions and facial recognition

Consider a concrete example of an AI diagnostic technology that could have dual uses, some to be licensed and some not to be (and thus forbidden). Researchers have analysed certain activities of people who extensively searched for information about Parkinson's disease on Bing, including their mouse movements, six months before they entered those search terms.[110] Most internet users are probably unaware that not just what they click on but how fast and smoothly they move their mouse to do so can be recorded and traced by the sites they are using. The group of Bing users who searched for Parkinson's—which it is probably safe to assume is more likely to have Parkinson's than the population as a whole—tended to have certain tremors in their mouse movements distinct from other searchers. These tremor patterns were undetectable by humans—only machine learning could distinguish the group identified to have a higher propensity to have Parkinson's, based in part on microsecond-by-microsecond differences in speed and motion of hand movement.

A licensure regime would likely forbid the calculation of the inference itself by entities that intend to discriminate based on it (or, more broadly, entities that have not demonstrated a personal or public health rationale for creating, disseminating, or using the inference).[111] But licenses could be granted to health care providers to use these inferences to give early diagnosis and support to the person whose data was analysed in this way. General inferences that enable other diagnostic programs may be permissible as a way of conducting "public or peer-reviewed scientific, historical, or statistical research in the public interest."[112] Thus, the generalisable finding may be made public, but its harmful use against an individual would be precluded by preventing a firm with no reasonable method of improving the person's health from making the inference. This avoids the "runaway AI" problem described in Pasquale's *Black Box Society*, where predictive analytics, initially deemed promising and helpful, becomes a bane for individuals stigmatised by them.

Sensitive to misuses of AI, ethicists have called for restrictions on certain types of AI, with a presumption that it be banned. For example, facial recognition is widely regarded as particularly dangerous and deserving of a ban.[113] The proposed EU AI Act already provides a blacklist of AI practices that should be banned (Article 5), but for the large majority of risky AI (the so-called high-risk AI), there is neither a ban nor a justificatory requirement, but only some specific design and organisational duties (Articles 6–15). But licensure allows for society to permit some of the highest value cases of facial recognition while preventing all others. For example, it may be reasonable to develop highly specialised databases of the faces of terrorists. But to deploy such powerful technology to ticket speeders or ferret out benefits fraud is inappropriate, like using a sledgehammer to kill a fly.[114] A rational government would not license the technology for such purposes, even if it would be entirely reasonable to do so for other purposes (for example, to prevent pandemics via early detection of infection clusters). Nor would it enable many of the forms of discrimination and mischaracterisation now enabled by light-to-non-existent regulation of large-scale AI. A licensure regime would help ensure that inaccurate, irresponsible, and damaging AI is limited. Rather than assuming that AI use is, in general, permitted and that regulators must struggle to catch up and outlaw particular bad acts, a licensure regime flips the presumption. Under it, companies would need to apply for permission for their AI to be deployed in mission-critical and sensitive contexts (at the very least for new AI applications if older ones are "grandfathered" and thus assumed to be licensed).

### 4.2. The finance precedent

The shift to thinking of AI use as a privilege, instead of as a right, may seem jarring to American ears, given the expansion of First Amendment coverage over the past century. However, even in the U.S. it is roundly conceded that there are certain particularly sensitive pieces of "information" that cannot simply be collected and disseminated. A die-hard cyberlibertarian or anarchist may want to copy and paste bank account numbers or government identification numbers onto anonymous websites, but that is illegal because complex sociotechnical systems like banks and the Social Security Administration can only function on a

[109] Pasquale, *The Black Box Society* (n 1) 2; Sandra Wachter and Brent Mittelstadt, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (2019) 2 Columbia Business Law Review <https://papers.ssrn.com/abstract=3248829> accessed 18 December 2018; Gianclaudio Malgieri and Jedrzej Niklas, 'The Vulnerable Data Subject' (2020) 37 Computer Law & Security Review.

[110] Ryen W White, P Murali Doraiswamy and Eric Horvitz, 'Detecting Neurodegenerative Disorders from Web Search Signals' (2018) 1 npj Digital Medicine 1; In this case, the source of the information was clear: Microsoft itself, which operates Bing, permitted the researchers to study anonymized databases. In the U.S., such data is now well beyond the scope of the privacy and security protections guaranteed pursuant to the Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health (HITECH) Act], see Bill Stead, NCVHS Chair and Linda Kloss, 'Health Information Privacy Beyond HIPAA: A Framework for Use and Protection' 21.

[111] Data Accountability and Transparency Act (DATA Act), S. 20719, 116th Cong. § 102(b)(4) (as proposed to the Senate, 2020) [hereinafter AI Act]. The proposed act states that data aggregators "shall not collect, use, or share, or cause to be collected, used, or shared, any personal data unless the aggregator can demonstrate that such personal data is strictly necessary to carry out a permissible purpose under section 102." *Id. at* § 101.

[112] European Data Protection Supervisor, 'Preliminary Opinion on Data Protection and Scientific Research | European Data Protection Supervisor' (2020) <https://edps.europa.eu/data-protection/our-work/publications/opinions/preliminary-opinion-data-protection-and-scientific_en> accessed 21 January 2022.

[113] Woodrow Hartzog and Evan Selinger, 'Why You Can No Longer Get Lost in the Crowd' *The New York Times* (17 April 2019) <https://www.nytimes.com/2019/04/17/opinion/data-privacy.html> accessed 21 January 2022.

[114] For an example of other such potential excessive uses, see Robert Pear, 'On Disability and on Facebook? Uncle Sam Wants to Watch What You Post' *The New York Times* (10 March 2019) <https://www.nytimes.com/2019/03/10/us/politics/social-security-disability-trump-facebook.html> accessed 21 January 2022.

predicate of privacy and informational control.[115] AI that enables, say, the automation of constant attempts to break into websites, or massive misuse and wasting of computational powers, should be similarly suspect and restricted.

Just as there is regulation of many forms of human subjects research, similar patterns of review and limitation must apply to the new forms of human classification and manipulation now enabled by AI.[116] A licensure regime for AI also puts some controls on the speed and ubiquity of the correlations such systems can make. Just as policymakers may want to prevent automated bots from dominating forums like Twitter (while permitting their development in other settings), we can and should develop a societal consensus toward limiting the degree to which automated correlations of often biased, partial, and secret AI influence our reputations and opportunities.[117] This commitment is already a robust part of finance regulation. For example, when credit scores are calculated, the Fair Credit Reporting Act imposes restrictions on the AI that can affect them.[118] Far from being a forbidden content-based restriction on the "speech" of scoring, such restrictions are vital to a fair credit system.[119] The Equal Credit Opportunity Act takes the restrictions further regarding a creditor's scoring system.[120] Such scoring systems may not use certain characteristics—such as race, sex, gender, marital status, national origin, religion, or receipt of public assistance—as a factor regarding a customer's credit worthiness. .[121] Far from being a relic of the activist 1970s, restrictions like this are part of contemporary efforts to ensure a fairer credit system.[122]

European examples abound as well. In Germany, the United Kingdom, and France, agencies cannot use ethnic origin, political opinion, trade union membership, or religious beliefs when calculating credit scores.[123] Germany and the United Kingdom also prohibit the use of health status in credit score calculations.[124] Such restrictions might be implemented as part of a licensure regime for use of AI-driven propensity scoring in many fields. For example, authorities may license systems that credibly demonstrate to authorized testing and certification bodies that they do not process AI on forbidden grounds, while denying a license to those that do.

Moreover, credit scores themselves feature as forbidden AI in some other determinations. For example, many U.S. states prevent them from being used by employers.[125] California, Hawaii, and Massachusetts ban the use of credit scoring for automobile insurance.[126] A broad coalition of civil rights and workers' rights groups reject these algorithmic assessments of personal worth and trustworthiness.[127] The logical next step for such activism is to develop systems of evaluation that better respect human dignity and social values in the construction of actionable reputations—those with direct and immediate impact on how we are classified, treated, and evaluated. For example, many have called for the nationalization of at least some credit scores.[128] Compared with that proposal, a licensure regime for such algorithmic assessments of propensity to repay is moderate.

To be sure, there will be some difficult judgement calls to be made, as in the case with any licensure regime. But size-based triggers can blunt the impact of licensure regimes on innovation by small and medium sized entities, focusing restrictions on companies with the most potential

[115] For a broader argument on the limits of First Amendment protection for operational code, see David Golumbia, 'Code Is Not Speech' (Social Science Research Network 2016) SSRN Scholarly Paper ID 2764214 <https://papers.ssrn.com/abstract=2764214> accessed 21 January 2022.

[116] For an analysis of the potential and limits of this analogy, see James Grimmelmann, 'Law and Ethics of Experiments on Social Media Users' [2015] Cornell Law Faculty Publications <https://scholarship.law.cornell.edu/facpub/1487>.

[117] On policy rationales for limiting automated bot speech, see Frank Pasquale, 'Preventing a Posthuman Law of Freedom of Expression' in David E Pozen (ed), *The Perilous Public Square: Structural Threats to Free Expression Today* (Columbia University Press 2020).

[118] U.S. Fair Credit Reporting Act (FCRA) § 609, 15 U.S.C. § 1681(g) (2011)

[119] The FCRA provides further language limiting what information may by contained in a consumer report. 15 U.S.C. 1681(c) (2011). Consumer reports cannot contain: Title 11 cases over ten years old; civil suits, judgments, or arrest records over seven years old; paid tax liens over seven years old; accounts placed for collection or charged to profit and loss over seven years old; or any other adverse information, other than criminal convictions, over seven years old. These restrictions have not been successfully challenged as content-based restrictions under the First Amendment.

[120] A creditor is defined by the Equal Credit Opportunity Act as those who "extend, renew, or continue credit." 15 U.S.C. § 1691(a)(e) (2010).

[121] 15 U.S.C. § 1691(a).

[122] Keshia Clukey, 'Social Networks Can't Go Into Credit Decisions Under N.Y. Ban (1)' (News Bloomberg Law) <https://news.bloomberglaw.com/banking-law/social-networks-cant-go-into-credit-decisions-under-n-y-ban> accessed 21 January 2022.

[123] Nicola Jentzsch, *Financial Privacy: An International Comparison of Credit Reporting Systems* (Springer Science & Business Media 2007).

[124] *Id.* The same restriction applies in the U.S. "A consumer reporting agency shall not furnish … a consumer report that contains medical information (other than medical contact information treated in the manner required under section 1681(c)(a)(6) of this title) about a consumer, unless—the consumer affirmatively consents, … if furnished for employment purposes, … the information is relevant to the process or effect the employment or credit transaction, … the information to be furnished pertains solely to transactions, accounts, or balances relating to debts arising from the receipt of medical services, products, or devises, … a creditor shall not obtain or use medical information … in connection with any determination of the consumer's eligibility, or continued eligibility, for credit." Fair Credit Reporting Act, 15 U.S.C. § 1681(b)(g) (2020).

[125] Microbilt, 'State Laws Limiting Use of Credit Information For Employment' <https://www.microbilt.com/Cms_Data/Contents/Microbilt/Media/Docs/MicroBilt-State-Laws-Limiting-Use-of-Credit-Information-For-Employment-Version-1-1-03-01-17-.pdf>.

[126] ibid.

[127] NYC Commission on Human Rights Legal Enforcement Guidance on the Stop Credit Discrimination in Employment Act, N.Y.C. Admin. Code §§ 8-102 (29), 8-107(9)(d), (24); Local Law No. 37 (2015), 'Stop Credit Discrimination in Employment Act: Legal Enforcement Guidance' <https://www1.nyc.gov/site/cchr/law/stop-credit-discrimination-employment-act.page> accessed 21 January 2022.

[128] McKenna Moore, 'Biden Wants to Change How Credit Scores Work in America' *Fortune* <https://fortune.com/2020/12/18/biden-public-credit-agency-economic-justice-personal-finance-racism-credit-scores-equifax-transuion-experian-cfpb/> accessed 21 January 2022; Amy Traub, 'Establish a Public Credit Registry' *Demos* <https://www.demos.org/policy-briefs/establish-public-credit-registry> accessed 21 January 2022; 'The Biden Plan for Investing in Our Communities through Housing' (*Joe Biden for President: Official Campaign Website*) <https://joebiden.com/housing/> accessed 21 January 2022.

to cause harm. Many of these companies are so large and powerful that they are almost governmental in their own right.[129] The EU's Digital Services Act, for example, includes obligations that would only apply to platforms that reach 10 percent of the EU population (at least 45 million people).[130] The Digital Markets Act includes obligations that would only apply to companies that provide "at least 45 million monthly active end users established or located in the Union and at least 10 000 yearly active business users established in the Union."[131] In the U.S., the California Consumer Privacy Act applies to companies that have AI on 50,000 California residents.[132] Many U.S. laws requiring security breach notifications generally trigger at around 500–1000 records breached.[133] In short, a nuanced licensing regime can be developed that is primarily aimed at the riskiest collections of AI and only imposes such obligations (or less rigorous ones) on smaller entities as the value and administrability of requirements for larger companies is demonstrated.

Nevertheless, close attention to the risks of AI may dictate wider applicability of such laws immediately. The size of the AI provider might not be directly proportional to the level of risks that their AI system poses for the fundamental rights and freedoms of individuals. In other words, even if an AI provider is an SME, potential impact of the application of the AI system that they provide might be highly risky (e.g., a start-up producing students' scoring algorithms or biometric categorisation on sensitive data, etc.). That is why where we refer to "nuanced" licensing regimes, rather than exemptions or derogations.

### 4.3. Anticipating objections

There will, of course, be many objections to our proposal. The division of responsibilities amongst the European Commission and member states can become dizzyingly complicated, as evidenced by recent concerns about the EU AI Act's apparent delegation of important functions to standardisation bodies. Veale and Borgesius have complained that the standardisation bodies that are slated to play an important role in EU AI regulation are not, at present, constituted to fully grasp (let alone regulate) the full panoply of civil rights, safety, and other normative issues raised by AI.[134] We agree that it would take some investment and empowerment of such institutions to address the full array of concerns raised. However, until more apt regulatory bodies are proposed, it may well be necessary to house licensure and justification regimes in institutions that will need to adapt to the role.

Given their regulation of information and information flows, licensure regimes will face challenges in some jurisdictions based on free expression rights.[135] For some commentators, AI and robots are tantamount to persons and thus deserve free speech rights.[136] While understandable as a futuristic possibility, the problems of such "rights for machines" become clear upon further reflection. As Birhane and van Dijk argue, so-called "intelligent machines" are "increasingly used in sustaining forms of oppression."[137] Consider the case of facial recognition. It is one thing to go to a protest when security personnel watch from afar. It is quite another when the police can immediately access your name, address, and job from a quick face scan purchased from an unaccountable private firm using machine vision.

This may be one reason why the American Civil Liberties Union decisively supported the regulation of Clearview AI (a firm providing facial recognition services) under the Illinois Biometric Information Privacy Act (BIPA), despite Clearview's insistence (to courts and the public at large) that it has a First Amendment right to gather and analyse AI unimpeded by BIPA. If unregulated, the firm's activities seem far more likely to undermine a robust public sphere than to promote it. Moreover, even if its AI applications were granted free expression protections, such protections may be limited by "time, place, and manner" restrictions. In that way, the licensure regime proposed here is much like permit requirements for parades, which recognise the need to balance the parade organisers' and marchers' free expression rights against the public need for safe and orderly streets. Given the privacy, security, and safety concerns raised by many forms of AI, a tailored licensing regime may be subject to only intermediate scrutiny in the U.S (ACLU v. Clearview AI, Case 20 CH 4353, Aug. 27, 2021: "BIPA's speaker-based exemptions do not appear to favour any particular viewpoint. As

[129] Frank Pasquale, 'From Territorial to Functional Sovereignty: The Case of Amazon' [2017] LPE Project <https://lpeproject.org/blog/from-territorial-to-functional-sovereignty-the-case-of-amazon/> accessed 21 January 2022.

[130] Article 33(1) of the 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). Such thresholds reflect a risk-focused model of regulation commended by the German AI Ethics Commission. AI Ethics Comm'n Fed. Gov't Ger., *Opinion of the AI Ethics Commission* (2019), 177.

[131] Art 3(2)(b) of the 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).

[132] Cal. Civ. Code § 1798.140(c)(1)(B) (West 2020) (covering businesses that "[a]lone or in combination, annually buys, receives for the business's commercial purposes, sells, or shares for commercial purposes, alone or in combination, the personal information of 50,000 or more consumers, households, or devices").

[133] *See, e.g.,* 16 C.F.R. § 318.5(b)–(c) ("A vendor of personal health records or PHR related entity shall provide notice to prominent media outlets serving a State or jurisdiction, following the discovery of a breach of security, if the unsecured PHR identifiable health information of 500 or more residents of such State or jurisdiction is, or is reasonably believed to have been, acquired during such breach."); Security Breach Notification Laws, https://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx [https://perma.cc/BS39-J2RE] (last visited May 13, 2021) (36 states set notification thresholds at 500 or 1,000).

[134] Veale and Borgesius (n 78).

[135] Jane Bambauer, 'Is Data Speech?' (2014) 66 Stanford Law Review 57; These rights claims will be particularly salient in the U.S., whose courts have expanded the scope of the First Amendment to cover many types of activity that would not merit free expression elsewhere, or would merit much less intense free expression protection, given the importance of competing rights to privacy, security, and AI protection. On the general issue of information processing being categorized as speech, see Jack M Balkin, 'Information Fiduciaries and the First Amendment' (2016) 49 UC Davis Law Review 52; Paul M Schwartz, 'Free Speech vs. Information Privacy: Eugene Volokh's First Amendment Jurisprudence' (2000) 52 Stanford Law Review 1559; James Hilmert, 'The Supreme Court Takes on the First Amendment Privacy Conflict and Stumbles: Bartnicki v. Vopper, the Wiretapping Act, and the Notion of Unlawfully Obtained Information' (2002) 77 77 Indiana Law Journal 639 (2002) <https://www.repository.law.indiana.edu/ilj/vol77/iss3/5>; Eric Easton, 'Ten Years After: Bartnicki v. Vopper as a Laboratory for First Amendment Advocacy and Analysis' [2011] SSRN Electronic Journal <http://www.ssrn.com/abstract=1986895> accessed 21 January 2022.

[136] John Frank Weaver, 'Why Robots Deserve Free Speech Rights' [2018] *Slate* <https://slate.com/technology/2018/01/robots-deserve-a-first-amendment-right-to-free-speech.html> accessed 21 January 2022.

[137] Abeba Birhane and Jelle van Dijk, 'Robot Rights? Let's Talk about Human Welfare Instead' [2020] Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society 207.

BIPA's restrictions are content neutral, the Court finds that intermediate scrutiny is the proper standard."). Far less free expression protection would be due in the EU, Canada, and Australia.[138] And the Chinese government, a leader in this space, has even more freedom of manoeuvre.[139]

## 5. Conclusion

Without proper assurances that the abuse of AI has been foreclosed, citizens should not accede to the large-scale application of AI now underway. Not only *ex post* enforcement but also ex ante licensure procedures are necessary to ensure that AI is only used for permissible purposes and is "justified", i.e. is not merely "explainable" but also lawful, fair, non-biased, non-manipulative, non-discriminatory, secure, and purpose-limited, respecting both data minimisation and storage limitation requirements. The ex ante approach is preferable to ex post approaches for several reasons, including market predictability (for AI providers), the limited deterrent impact of ex post sanctions, and the risk that an ex post regime might even be dangerous for some fundamental rights as it avers to preserve others (e.g., prohibiting an AI-driven app that has been massively used could create adverse effects on its previous users that might be in a situation of psychological or economic dependency on that app).

Building on present regulatory models, including the GDPR and AIA, this article has proposed a presumption of unlawfulness for high-risk AI models. Developers, vendors, and users of such models should bear the burden of proof to justify why their AI system is not illegitimate (and thus not unfair, not discriminatory, and not inaccurate). Such a standard may not seem administrable now, given the widespread and rapid use of AI at companies of all sizes. But such requirements could be applied, at first, to the most troubling practices and only gradually (if at all) to smaller companies and less menacing practices. This article has sketched the first steps toward translating the general normative construct of a "social license" for justifying AI use into a specific licensure framework. Our starting point is the proposed AI Act model, which seems a reasonable model for ex ante authorisation of AI systems (through the ex ante conformity assessment procedure that providers of high-risk AI systems need to perform in order to show the compliance with several principles and design safeguards, including data governance, integrity and human oversight). However, the proposed AI Act represents only a "timid" ex ante approach due to the limited scope of its licensure model, the lack of transparency of justification documents, and the limited content of the AIA justification (fairness and non-discrimination were not referred to in the original European Commission proposal of the AIA). That is why this article advocates for complementing the AIA's approach with a data-focused GDPR approach, especially inspired by GDPR Article 5 principles (fairness, lawfulness, transparency, accuracy, purpose limitation, data storage limitation, data minimisation, integrity, and accountability).

Of course, more conceptual work remains to be done, both substantively (elaborating grounds for denying a license) and practically (to estimate the resources needed to develop the first iteration of the licensing proposal).[140] The notice and consent model has enjoyed the benefits of such conceptual work for decades; now, it is time to devote similar intellectual energy to a licensing model. *Ex ante* licensure of large-scale AI use should become common in jurisdictions committed to enabling democratic governance of AI. Defining permissible purposes for the licensure of AI will take up an increasing amount of time for regulators, and law enforcers will need new tools to ensure that regulations are actually being followed. The articulation and enforcement of these specifications will prove an essential foundation of an emancipatory industrial policy for AI.

## Declaration of Competing Interest

No conflict of interest to declare.

## Data availability

No data was used for the research described in the article.

## Further reading


[1] Aarnio A. The rational as reasonable: a treatise on legal justification. Springer Science & Business Media; 1986.
[2] Alexander L, Moore M. Deontological Ethics. In: Zalta Edward N, editor. The stanford encyclopedia of philosophy. Winter 2020, Metaphysics Research Lab, Stanford University; 2020. <, https://plato.stanford.edu/archives/win2020/entries/ethics-deontological/. > accessed 1 December 2020.
[3] Anderson E. Private government: how employers rule our lives. Princeton University Press; 2017.
[4] Araujo T, others. In AI we trust? Perceptions about automated decision-making by artificial intelligence. AI Soc 2020;35:611.
[5] Article 29 Working Party, 'Guidelines on automated individual decision-making and profiling for the purposes of regulation 2016/679' (2017).
[6] Ausloos J, Dewitte P. Shattering one-way mirrors – data subject access rights in practice. Int Data Privacy Law 2018;8:4.
[7] Balkin JM. Information fiduciaries and the first amendment. UC Davis Law Rev 2016;49:52.
[8] Bambauer J. Is data speech? Stanford Law Rev 2014;66:57.
[9] Barocas S, Nissenbaum H. On notice: the trouble with notice and consent. ]. In: Proceedings of the engaging data forum: the first international forum on the application and management of personal electronic information; 2009. p. 7.
[10] Barocas S, Selbst AD. Big data's disparate impact. Calif Law Rev 2016;104:671.
[11] Benjamin R. Race after technology. Polity Press; 2019.
[12] Bergemann B., 'The consent paradox: accounting for the prominent role of consent in data protection' in Marit Hansen and others (eds), Privacy and identity management. the smart revolution: 12th IFIP WG 9.2, 9.5, 9.6/11.7, 11.6/SIG 9.2.2 International Summer School, Ispra, Italy, September 4-8, 2017, Revised Selected Papers (Springer International Publishing 2018) <https://doi.org/10.1007/978-3-319-92925-5_8> accessed 4 April 2020.
[13] Biran O. and Cotton C.V., 'Explanation and justification in machine learning : a survey', (2017) IJCAI-17 workshop on explainable AI (XAI), </paper/explanation-and-justification-in-machine-learning-%3A-Biran-Cotton/02e2e79a77d8aabc1af1900ac80ceebac20abde4> accessed 26 November 2020.
[14] Birhane A, van Dijk J. Robot rights? Let's talk about human welfare instead. In: Proceedings of the AAAI/ACM conference on AI, ethics, and society; 2020. p. 207.
[15] Boltanski L, Thévenot L. On Justification. Economies of Worth. Princeton University Press; 2006.
[16] Brkan M. The essence of the fundamental rights to privacy and data protection: finding the way through the maze of the CJEU's constitutional reasoning. German Law J 2019;20:864.
[17] Broussard M. Artificial unintelligence: how computers misunderstand the world. MIT Press; 2018.


[138] Office of the Privacy Commissioner of Canada, 'Joint Investigation of Clearview AI, Inc. by the Office of the Privacy Commissioner of Canada, the Commission d'accès à l'information Du Québec, the Information and Privacy Commissioner for British Columbia, and the Information Privacy Commissioner of Alberta' (3 February 2021) <https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2021/pipeda-2021-001/> accessed 21 January 2022.

[139] 'Translation: Internet Information Service Algorithmic Recommendation Management Provisions – Effective March 1, 2022' (*DigiChina*) <https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/> accessed 21 January 2022.

[140] To provide the proper level of resources, the "self-funding agency" model is useful. Certain financial and medical regulators are funded in part via fees paid by regulated entities which must apply to engage in certain activities. For example, fees paid pursuant to the Prescription Drug User Fee Act (PDUFA) fund the Food and Drug Administration (which essentially licenses drugs for sale in the U.S.). For background on this Act and its Amendments, see U.S. Food and Drug Administration, Prescription Drug User Fee Amendments, at https://www.fda.gov/industry/fda-user-fee-programs/prescription-drug-user-fee-amendments (last updated Aug. 25, 2021).

[18] Broussard M. More than a glitch. MIT Press; 2023.
[19] Bunge M. Philosophy of science: from problem to theory. Transaction Publishers; 1998.
[20] Butterworth M. The ICO and Artificial Intelligence: the Role of Fairness in the GDPR Framework. Computer Law Secur Rev 2018;34:257.
[21] Cabitza F, others. Quod erat demonstrandum? - Towards a typology of the concept of explanation for the design of explainable AI. Expert Syst Appl 2023; 213:118888.
[22] Canada Office of the Privacy Commissioner, 'Joint Investigation of Clearview AI, Inc. by the office of the privacy commissioner of Canada, the Commission d'accès à l'information Du Québec, the information and privacy commissioner for British Columbia, and the information privacy commissioner of alberta' (3 February 2021) <https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-businesses/2021/pipeda-2021-001/>Accessed 21 January 2022.
[23] Chen, L. et al., 'How Is ChatGPT's Behavior Changing over Time?', arXiv:2307.09009v2 [cs.CL] (2023).
[24] Clifford D, Ausloos J. Data Protection and the Role of Fairness. Yearbook Eur Law 2018;37:130.
[25] Clukey K. Social networks can't go into credit decisions under N.Y. ban (1). News Bloomberg Law; 2022. <, https://news.bloomberglaw.com/banking-law/social-networks-cant-go-into-credit-decisions-under-n-y-ban. > accessed 21 January.
[26] Cohen JE. Turning privacy inside out. Theoretical inquiries in law, 20; 2019. <, http://www7.tau.ac.il/ojs/index.php/til/article/view/1607. > accessed 23 January 2019.
[27] Cole S. "It's hurting like hell": ai companion users are in crisis, reporting sudden sexual rejection. Vice; 2023. 15 February<, https://www.vice.com/en/article/y3py9j/ai-companion-replika-erotic-roleplay-updates. > accessed 1 May 2023.
[28] Commission National Informatique and Libertés, 'How Can Humans Keep the Upper Hand? The Ethical Matters Raised by Algorithms and Artificial Intelligence, Report on the Public Debate Led by the French Data Protection Authority (CNIL) as Part of the Ethical Discussion Assignment Set by the Digital Republic Bill' (2017).
[29] Corbyn Z., '"Bossware Is Coming for Almost Every Worker": the Software You Might Not Realize Is Watching You' The Guardian (27 April 2022) <https://www.theguardian.com/technology/2022/apr/27/remote-work-software-home-surveillance-computer-monitoring-pandemic> accessed 3 May 2022.
[30] Data Ethics Commission of the Federal Government of Germany, 'Opinion of the Data Ethics Commission' <https://www.bmj.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN_lang.html> accessed 21 January 2022.
[31] Datatilsynet, 'Advance notification of order to rectify unfairly processed and incorrect personal data - international baccalaureate organization' (2020) <https://www.datatilsynet.no/contentassets/04df776f85f64562945f1d261b4add1b/advance-notification-of-order-to-rectify-unfairly-processed-and-incorrect-personal-data.pdf>.
[32] Dressel J, Farid H. The accuracy, fairness, and limits of predicting recidivism. Sci Adv 2018;4:eaao5580.
[33] Ducato R. and Marique E., 'Come to the dark side: we have patterns. choice architecture and design for (Un)informed consent' (social science research network 2018) SSRN Scholarly Paper ID 3365952 <https://papers.ssrn.com/abstract=3365952> accessed 31 May 2020.
[34] Dwork C, Mulligan DK. It's Not Privacy, and It's Not Fair. Stanford Law Rev 2013; 66:6.
[35] Easton E. Ten Years After: Bartnicki v. Vopper as a laboratory for first amendment advocacy and analysis. ] SSRN Electron J 2011. <, http://www.ssrn.com/abstract=1986895. > accessed 21 January 2022.
[36] Edwards L, Veale M. Enslaving the algorithm: from a "right to an explanation" to a "right to better decisions"? IEEE Secur Priv 2018;16:46.
[37] Electronic Privacy Information Center, Generating harms: generative AI's impacts and paths forward, at https://epic.org/documents/generating-harms-generative-ais-impact-paths-forward/(2023), accessed July 20, 2023.
[38] European Data Protection Board, 'Guidelines 1/2018 on certification and identifying certification criteria in accordance with articles 42 and 43 of the regulation - version adopted after public consultation | European data protection board' (2018) <https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-12018-certification-and-identifying_en> accessed 21 January 2022.
[39] European Data Protection Supervisor, 'Preliminary opinion on data protection and scientific research | european data protection supervisor' (2020) <https://edps.europa.eu/data-protection/our-work/publications/opinions/preliminary-opinion-data-protection-and-scientific_en> accessed 21 January 2022.
[40] European Parliament Resolution, 'Framework of ethical aspects of artificial intelligence, robotics and related technologies - Tuesday, 20 October 2020' <https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html> accessed 21 January 2022.
[41] European Parliament, EU AI Act: first regulation on artificial intelligence, at https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#:~:text=AI%20systems%20that%20negatively%20affect,cars%2C%20medical%20devices%20and%20lifts, accessed Aug. 3, 2023.
[42] 'EXPLAIN', Meaning & definition for UK English, Lexico.Com' (Lexico Dictionaries | English) <https://www.lexico.com/definition/explain> accessed 21 January 2022.
[43] Falco G. and others, 'Governing AI safety through independent audits' (2021) 3 nature machine intelligence <https://uwe-repository.worktribe.com/output/7562797/governing-ai-safety-through-independent-audits> accessed 21 January 2022.
[44] Fortuna-Zanfir G. Forgetting about consent. Why the focus should be on "suitable safeguards" in data protection law. In: Gutwirth Serge, Leenes Ronald, Hert Paul De, editors. Reloading data protection. Springer; 2014.
[45] Future of Privacy Forum, 'Unfairness by algorithm: distilling the harms of automated decision-making' (2017) <https://fpf.org/2017/12/11/unfairness-by-algorithm-distilling-the-harms-of-automated-decision-making/> accessed 8 February 2020.
[46] Galhotra S, Brun Y, Meliou A. Fairness testing: testing software for discrimination. In: Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering - ESEC/FSE 2017. ACM Press; 2017. <, http://dl.acm.org/citation.cfm?doid=3106237.3106277. > accessed 31 May 2019.
[47] Garante per la Protezione dei Dati Personali, 'Provvedimento del 2 febbraio 2023 [9852214]' <https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9852214> accessed 1 May 2023.
[48] Garante per la Protezione dei Dati Personali, 'Provvedimento del 30 marzo 2023 [9870832]' <https://www.garanteprivacy.it:443/home/docweb/-/docweb-display/docweb/9870832> accessed 1 May 2023.
[49] Gebru T, others. Datasheets for datasets. Commun ACM 2021;64:86.
[50] Golumbia D., 'Code is not speech' (social science research network 2016) SSRN Scholarly Paper ID 2764214 <https://papers.ssrn.com/abstract=2764214> accessed 21 January 2022.
[51] Grimmelmann J., 'Law and ethics of experiments on social media users' [2015] Cornell Law Faculty Publications <https://scholarship.law.cornell.edu/facpub/1487>.
[52] Hamon R, others. Impossible explanations? Beyond explainable AI in the GDPR from a COVID-19 use case scenario. In: Proceedings of the 2021 acm conference on fairness, accountability, and transparency. Association for Computing Machinery; 2021. https://doi.org/10.1145/3442188.3445917. <> accessed 27 May 2021.
[53] Hartzog W. and Selinger E., 'Why you can no longer get lost in the crowd' The New York Times (17 April 2019) <https://www.nytimes.com/2019/04/17/opinion/data-privacy.html> accessed 21 January 2022.
[54] Harwell D., 'A face-scanning algorithm increasingly decides whether you deserve the job' washington post <https://www.washingtonpost.com/technology/2019/10/22/ai-hiring-face-scanning-algorithm-increasingly-decides-whether-you-deserve-job/> accessed 3 May 2022.
[55] Hazen BT, others. Data quality for data science, predictive analytics, and big data in supply chain management: an introduction to the problem and suggestions for research and applications. Int J Prod Econ 2014;154:72.
[56] Henin C, Métayer DL. Beyond explainability: justifiability and contestability of algorithmic decision systems. AI Society 2022;37(4):1397–410. https://doi.org/10.1007/s00146-021-01251-8.
[57] Henin C, Métayer DL. 'A multi-layered approach for interactive black-box explanations'. In: Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15. Proceedings, Part III. Berlin, Heidelberg: Springer-Verlag; 2021. p. 5–19. https://doi.org/10.1007/978-3-030-68796-0_138.
[58] Hildebrandt M. Profile transparency by design? *Re*-Enabling double contingency. Mireille Hildebrandt and Katia De Vries. "Privacy, Due Process and the Computational Turn" (Routledge, 2013). 2013. p. 221–47.
[59] Hildebrandt M., Law for computer scientists and other folk (Oxford University Press 2020).
[60] Hilmert J. The supreme court takes on the first amendment privacy conflict and stumbles: Bartnicki v. Vopper, the Wiretapping Act, and the Notion of Unlawfully Obtained Information. Indiana Law J 2002;77(77):639 (2002) <, https://www.repository.law.indiana.edu/ilj/vol77/iss3/5. >.
[61] Horowitz DL. Justification and Excuse in the Program of the Criminal Law. Law Contemp Probl 1986;49:109.
[62] Information Commissioner's Office, 'Big data, artificial intelligence, machine learning and data protection' (2017) <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>.
[63] Information Commissioner's Officer, 'Accountability and governance' (1 October 2020) <https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/> accessed 29 November 2020.
[64] Jentzsch N. Financial privacy: an international comparison of credit reporting systems. Springer Science & Business Media; 2007.
[65] Johnson K, Pasquale F, Chapman J. Artificial intelligence, machine learning, and bias in finance: toward responsible innovation. Fordham Law Rev 2019;88:31.
[66] 'JUSTIFICATION', Meaning & Definition for UK English, Lexico.Com' (Lexico Dictionaries | English) <https://www.lexico.com/definition/justification> accessed 21 January 2022.
[67] Kaminski M, Malgieri G. Multi-layered explanation from algorithmic impact assessments in the GDPR. In: FAT 2020 Proceedings. ACM publishing; 2020.
[68] Kaminski ME. Binary governance: lessons from the GDPR's approach to algorithmic accountability'. South Calif Law Rev 2019;92. <, https://papers.ssrn.com/abstract=3351404. > accessed 23 April 2019.
[69] Kaminski ME, Malgieri G. Algorithmic impact assessments under the GDPR: producing multi-layered explanations. Int Data Priv Law 2021;11(2):125–44.
[70] Katyal SK. Private accountability in the age of artificial intelligence. UCLA Law Review 2019;66:88.
[71] Kim P.T., 'Data-driven discrimination at work' 58 81.

[72] Kloza D. and others, 'Data protection impact assessment in the european union: developing a template for a report from the assessment process' (LawArXiv 2020) DPiaLab Policy Brief <https://osf.io/7qrfp> accessed 1 December 2020.
[73] Kroll J, others. Accountable algorithms. Univ PA Law Rev 2017;165:633.
[74] Lander E. Americans need a bill of rights for an AI-powered world. Wired 2022. <, https://www.wired.com/story/opinion-bill-of-rights-artificial-intelligence/. > accessed 21 January.
[75] Lashbrook A., 'AI-driven dermatology could leave dark-skinned patients behind' the atlantic (16 August 2018) <https://www.theatlantic.com/health/archive/2018/08/machine-learning-dermatology-skin-color/567619/> accessed 3 May 2022.
[76] Lepri B, others. Fair, transparent, and accountable algorithmic decision-making processes. Philos Technol 2018;31:611.
[77] Lipton ZC. The mythos of model interpretability. Commun ACM 2018;61:36.
[78] Lodder A.R., Dialaw: On legal justification and dialogical models of argumentation (1999 ed, Kluwer Academic Pub 1999).
[79] Loi M, Ferrario A, Viganò E. Transparency as design publicity: explaining and justifying inscrutable algorithms. ] Ethics Inf Technol 2020. https://doi.org/10.1007/s10676-020-09564-w. <> accessed 30 November 2020.
[80] Lomas N., 'Replika, a "Virtual friendship" AI Chatbot, hit with data ban in Italy over child safety' (TechCrunch, 3 February 2023) <https://techcrunch.com/2023/02/03/replika-italy-data-processing-ban/> accessed 1 May 2023.
[81] Lupton D, Williamson B. The datafied child: the dataveillance of children and implications for their rights, 19. New Media & Society; 2017. p. 780.
[82] Mahieu R. Right of access to personal data: a genealogy. ] Technol Regulation 2021:62.
[83] Malgieri G. Automated decision-making in the EU member states: the right to explanation and other "suitable safeguards" in the national legislations. Comput Law Secur Rev 2019;35:105327.
[84] Malgieri G. The concept of fairness in the GDPR: a linguistic and contextual interpretation. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery; 2020. https://doi.org/10.1145/3351095.3372868. <> accessed 29 January 2020.
[85] Malgieri G. "Just" algorithms: justification (beyond) explanation of automated decisions under the GDPR. Law Bus 2021;1.
[86] Malgieri G., Vulnerability and data protection law (Oxford University Press 2023).
[87] Malgieri G, Comandé G. Why a right to legibility of automated decision-making exists in the general data protection regulation. Int Data Privacy Law 2017;7:243.
[88] Malgieri G, Niklas J. The vulnerable data subject. Computer Law Secur Rev 2020;37.
[89] Marcus G. and Davis E., Rebooting A.I.: Building artificial intelligence we can trust (Vintage 2019).
[90] Microbilt, 'State laws limiting use of credit information for employment' <https://www.microbilt.com/Cms_Data/Contents/Microbilt/Media/Docs/MicroBilt-State-Laws-Limiting-Use-of-Credit-Information-For-Employment-Version-1-1-03-01-17-.pdf>.
[91] Milaj J. Privacy, surveillance, and the proportionality principle: the need for a method of assessing privacy implications of technologies used for surveillance. Int Rev Law, Comput Technol 2016;30:115.
[92] Miller T. Explanation in artificial intelligence: insights from the social sciences. Artif Intell 2019;267:1.
[93] Moore M. Biden wants to change how credit scores work in America. FortuneFortune 2022. <, https://fortune.com/2020/12/18/biden-public-credit-agency-economic-justice-personal-finance-racism-credit-scores-equifax-transuion-experian-cfpb/. > accessed 21 January.
[94] Moser PK. Justification in the natural sciences. Br J Philos Sci 1991;42:557.
[95] Naudts L, Dewitte P, Ausloos J. Meaningful transparency through data rights: a multidimensional analysis. In: Kosta Eleni, Kamara Irene, Leenes Ronald, editors. Research handbook on EU data protection law. Edward Elgar Publishing; 2022, 530-571..
[96] Noble S. Algorithms of oppression. University of California Press; 2018.
[97] NYC. Commission on human rights legal enforcement guidance on the stop credit discrimination in employment act, N.Y.C. Admin. Code §§. Stop credit discrimination in employment act: legal enforcement guidance, 8-102; 2015. p. 8–107. <, https://www1.nyc.gov/site/cchr/law/stop-credit-discrimination-employment-act.page. > accessed 21 January 2022.
[98] Omarova S. License to deal: mandatory approval of complex financial products. Washington University Law Rev 2012;90:64.
[99] Cathy O'Neil, Weapons of math destruction (Vintage 2016).
[100] Oprișiu R., 'Reversal of "the burden of proof" in data protection | Lexology' <https://www.lexology.com/library/detail.aspx?g=e9e8c734-23d9-41bb-a723-5d664b3c86cc> accessed 29 November 2020.
[101] Pasquale F. The black box society: the secret algorithms that control money and information. Harvard Univ Pr; 2015.
[102] Pasquale F. From territorial to functional sovereignty: the case of amazon. ] LPE Project 2017. <, https://lpeproject.org/blog/from-territorial-to-functional-sovereignty-the-case-of-amazon/. > accessed 21 January 2022.
[103] Pasquale F. When machine learning is facially invalid. Commun ACM 2018;61:25.
[104] Pasquale F., New laws of robotics: defending human expertise in the age of Ai (Belknap Pr 2020).
[105] Pasquale F. Preventing a posthuman law of freedom of expression. In: David E Pozen, editor. The perilous public square: structural threats to free expression today. Columbia University Press; 2020.
[106] Pasquale F. Licensure as data governance. ] Knight First Amend Institute at Columbia University 2021. <, https://knightcolumbia.org/content/licensure-as-data-governance. > accessed 21 January 2022.
[107] Pasquale F, Cashwell G. Prediction, persuasion, and the jurisprudence of behaviorism. ] Faculty Scholarship 2018. <, https://digitalcommons.law.umaryland.edu/fac_pubs/1604. >.
[108] Pasquale F. and Malgieri G., 'Opinion | if you don't trust A.I. yet, you're not wrong' The New York Times (30 July 2021) <https://www.nytimes.com/2021/07/30/opinion/artificial-intelligence-european-union.html> accessed 21 January 2022.
[109] Pear R., 'On disability and on Facebook? Uncle Sam wants to watch what you post' The New York Times (10 March 2019) <https://www.nytimes.com/2019/03/10/us/politics/social-security-disability-trump-facebook.html> accessed 21 January 2022.
[110] Petkova B, Hacker P. Reining in the big promise of big data: transparency, inequality, and new regulatory frontiers. ] Lect Other Affiliate Scholar Series 2016. <, https://digitalcommons.law.yale.edu/ylas/13. >.
[111] Press release, 'Brown releases new proposal that would protect consumers' privacy from bad actors | U.S. senator Sherrod Brown of Ohio' <https://www.brown.senate.gov/newsroom/press/release/brown-proposal-protect-consumers-privacy> accessed 21 January 2022.
[112] Price WN, Gerke S, Cohen IG. Potential Liability for Physicians Using Artificial Intelligence. JAMAJAMA 2019;322:1765.
[113] Roig A. Safeguards for the right not to be subject to a decision based solely on automated processing (Article 22 GDPR). Eur J Law Technol 2018;8. <, http://ejlt.org/article/view/570. > accessed 15 January 2019.
[114] Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intell 2019;1:206.
[115] Schermer BW, Custers B, van der Hof S. The crisis of consent: how stronger legal protection may lead to weaker consent in data protection. Ethics Inf Technol 2014;16:171.
[116] Schwartz PM. Free speech vs. information privacy: Eugene Volokh's first amendment jurisprudence. Stanford Law Rev 2000;52:1559.
[117] Selbst AD. Disparate impact in big data policing. Georgia Law Review 2018;52:109.
[118] Selbst AD, Powles J. Meaningful information and the right to explanation. Int Data Privacy Law 2017;7:233.
[119] Smith J.C., Justification and excuse in the criminal law (Stevens 1989).
[120] Stead B., Chair N. and Kloss L., 'Health information privacy beyond HIPAA: a framework for use and protection' 21.
[121] 'The biden plan for investing in our communities through housing' (joe biden for president: official campaign website) <https://joebiden.com/housing/> accessed 21 January 2022.
[122] Tilly C., Why? (2008) <https://press.princeton.edu/books/paperback/9780691136486/why> accessed 21 January 2022.
[123] Topol E, Medicine Deep. How artificial intelligence can make healthcare human again. Illustrated ed. Basic Books; 2019.
[124] 'Translation: Internet information service algorithmic recommendation management provisions – Effective March 1, 2022' (DigiChina) <https://digichina.stanford.edu/work/translation-internet-information-service-algorithmic-recommendation-management-provisions-effective-march-1-2022/> accessed 21 January 2022.
[125] Traub A. Establish a public credit registry. Demos 2022. <, https://www.demos.org/policy-briefs/establish-public-credit-registry. > accessed 21 January.
[126] Tutt A. An FDA for Algorithms. Adm Law Rev 2017;69:83.
[127] Tyler TR. Procedural justice, legitimacy, and the effective rule of law. Crime Justice 2003;30:283.
[128] Veale M, Borgesius FZ. Demystifying the draft eu artificial intelligence act — analysing the good, the bad, and the unclear elements of the proposed approach. Comput Law Review Int 2021;22:97.
[129] Veale M, Edwards L. Clarity, surprises, and further questions in the article 29 working party draft guidance on automated decision-making and profiling. Comput Law Secur Rev 2018;34:398.
[130] Venkataramakrishnan S., 'Top researchers condemn "racially biased" face-based crime prediction' Financial Times (24 June 2020) <https://www.ft.com/content/aaa9e654-c962-46c7-8dd0-c2b4af932220> accessed 21 January 2022.
[131] Viljoen S. A relational theory of data governance. ] Yale Law J 2021;82.
[132] Wachter S. Affinity profiling and discrimination by association in online behavioural advertising. Berkeley Technol Law J 2019;35. <, https://papers.ssrn.com/abstract=3388639. > accessed 2 June 2019.
[133] Wachter S, Mittelstadt B. A right to reasonable inferences: *re*-thinking data protection law in the age of big data and AI'. Columbia Bus Law Rev 2019;2. <, https://papers.ssrn.com/abstract=3248829. > accessed 18 December 2018.
[134] Wachter S, Mittelstadt B, Russell C. Counterfactual explanations without opening the black box: automated decisions and the GDPR. ] Harv J Law Technol 2018. <, http://arxiv.org/abs/1711.00399. > accessed 16 September 2019.
[135] Wang T. et al., 'Adversarial policies beat superhuman Go AIs,' arXiv: 2211.00241v4 [cs.LG] (2022).

[136] Weaver JF. Why robots deserve free speech rights. ] Slate 2018. <, https://slate.com/technology/2018/01/robots-deserve-a-first-amendment-right-to-free-speech.htm. l> accessed 21 January 2022.
[137] White RW, Doraiswamy PM, Horvitz E. Detecting neurodegenerative disorders from web search signals. NPJ Digital Med 2018;1:1.
[138] Zook M, others. Ten simple rules for responsible big data research. PLoS Comput Biol 2017;13:e1005399.
[139] Zweig K., Awkward intelligence (M.I.T. Press 2021).
[140] Advocate General Opinion, M P Pikamäe, C-634/21 (ECJ).
[141] Ligue des droits humains ASBL v Conseil des ministres [2022]ECJ Case C-817/19.