AI and Regulation: Protecting the public without extinguishing the spark of innovation

By Gabriele Spina Ali and Ron Yu

Abstract

Recent instances of Artificial Intelligence (AI) systems causing harm through, *inter alia,* biased censorship of online speech, unfair denial of government benefits, unjust incarceration or attempted circumvention of laws has led governments to demand greater scrutiny of AI systems. This involves calls for more transparency, requiring service providers employing AI to disclose at least some features of their AI technology. On their side, corporations strongly oppose these demands for transparency. They point out that their technology constitutes valuable trade secrets, that disclosure implies losing their technology to competitors and that in turn this will stagnate innovation in the AI sector.

The present paper analyses the necessary trade-off between transparency and innovation and proposes some solutions, sometimes borrowed from other sectors, to strike a reasonable balance between public safety and innovation policies.

For inquiries please contact:

- Gabriele Spina Ali at Gabriele.spinaali@gmail.com
- Ronald Yu at ronaldyu@mac.com.

1. Introduction

There is considerable controversy surrounding AI systems. Beyond the fear that robots might one day rebel against humanity and subdue it, or at least displace workers in the labor market leading to higher inequality and unemployment rates, more imminent fears concern automated systems censoring speech, causing physical harm to humans, or encroaching citizens' rights. To provide some examples, the algorithms[1] of social platforms have been accused of interfering with the free speech of dissenting politicians and journalists,[2] doubts about the safety of AI powered vehicles have been raised,[3] AI systems have discriminated applicants on the basis of gender and race,[4] and even caused the unfair jailing of innocent people. [5] As that was not enough, AI has recently become a tool to be used to circumvent domestic laws and regulation, for instance in the field of public transportation.[6]

Thus, it should not be surprising that calls for greater regulation of AI systems have been ubiquitous, with reports issued by public bodies in Asia,[7] Europe[8] and the United States.[9]

However, regulating artificial intelligence is far from a straightforward process. There is indeed a danger that overly stringent oversight could throw water on the innovative spark by unintentionally undermining the intellectual property (IP) regime.

AI industries are reluctant to open their technologies to public scrutiny, fearing the loss of valuable trade secret, loss of competitiveness, or to the least, unwilling to bear the financial burdens associated with complying with the procedural, substantive and technological requirements prescribed by the law.

To understand why regulating AI is so difficult and the unintended consequences mandated algorithmic disclosure might bring about, we need to first understand how AI works.

---

[1] For purposes of this paper, 'algorithm'will be defined as a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer."

See: https://www.igi-global.com/dictionary/algorithms-aided-sustainable-urban-design/988

[2] Allum Bohkary, 'THE GOOD CENSOR': Leaked Google Briefing Admits Abandonment of Free Speech for 'Safety And Civility", *Breitbart*, 2 Oct. 2018, https://www.breitbart.com/tech/2018/10/09/the-good-censor-leaked-google-briefing-admits-abandonment-of-free-speech-for-safety-and-civility/ [Accessed 15 July, 2019] https://www.nytimes.com/2018/12/27/world/facebook-moderators.html [Accessed on 17 July 2019].

[3] https://www.wired.com/story/self-driving-cars-safety-metrics-miles-disengagements/ [Accessed on 17 July 2019].

[4] https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [Accessed on 17 July 2019]

[5] Adam Liptak,' Sent to Prison by a Software Program's Secret Algorithms', *New York Times*, 1 May 2017 https://www.nytimes.com/2017/05/01/us/politics/sent-to-prison-by-a-software-programs-secret-algorithms.html [Accessed 12, July 2019]

[6] For example, the controversy surrounding Uber's greyball scheme. See: Mike Isaac, ''How Uber Deceives the Authorities Worldwide', *New York Times*, 3 March 2017 https://www.nytimes.com/2017/03/03/technology/uber-greyball-program-evade-authorities.html [Accessed 11 July 2019]

[7] https://futureoflife.org/ai-policy-china/[Accessed on 17 July 2019].

[8] https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/artificial-intelligence-european-perspective [Accessed on 17 July 2019].

[9] https://futureoflife.org/ai-policy-united-states/[Accessed on 17 July 2019].

2. Artificial Intelligence: Components and Features

It is often said that AI systems are machines capable of displaying intelligence, i.e. the ability to combine knowledge base and deductive capabilities,[10] and that, however, they display an intelligence that is very different from human's in terms of generality and adaptability.[11] Overly broad definitions of this kind do not help regulators and policy makers, so that it is better to focus on the components and distinctive characteristics of AI systems. AI machines usually consist of "a sensor or input mechanism, controlling algorithm(s) and the capacity to give feedback to the outer world"[12] Leaving aside the input and output mechanisms, which are found also in traditional computers, the two main components of AI technology are:

- *The running algorithm(s):* This is the formal set of instruction given to the machine; the process or set of rules to be followed in calculations or other problem-solving operations by a computer. Unlike traditional computing, AI algorithms have the ability to self-modify based on past experience similarly to biological brains, so that they improve overtime.[13] This is done through specific computing techniques such as back-propagation, which allows the algorithm to propagate back from an undesired output to the origin of the mistake and to improve the process from that point onward.[14]

- *The training database:* AI algorithms can analyze, extract and identify significant patterns in large data sets. For instance, 'Google Translate' used a statistical machine engine, which identified significant linguistic patterns in millions of United Nations and EU parliament documents.[15] These databases are used to 'train' AI algorithms, which will start making predictions based on the patters identified in the training datasets. Also, unlike traditional computing, AI databases are constantly evolving, constantly changing – e.g. with social platforms algorithms being relentlessly fed new data from users and subscribers.

AI Systems produce different outputs based on their algorithms and the data they have been exposed to. While outputs are the end-result of a computing process rather than one of its components, two things are worth noticing.

First, AI applications may feed on their own outputs, i.e. utilize their own end-results as training data. This leads to the problem of the self-reinforcing nature of AI technology.

Second, in some fields it may be necessary to regulate AI outputs independently from considerations related to their databases and algorithms. For instance, significant

[10] Butler, T. (1981) 'Can a Computer be an Author? Copyright Aspects of Artificial Intelligence', *Entertainment Law Society* 4, p 710.
[11] Bostrom. N. & Yudkowski. E. (2019) 'The Ethics of Artificial Intelligence', Machine Learning Institute [online]. Available at https://intelligence.org/files/EthicsofAI.pdf [Accessed 23 April 2019].
[12] Eidenmüller, H. (2017), 'The Rise of Robots and the Law of Humans', Oxford Legal Studies Research Paper No. 27/2017 *[online]*. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2941001 [Accessed 18 July 2019].
[13] University of Toronto (2018) 'Artificial Neural Networks' [online]. Available at http://www.psych.utoronto.ca/users/reingold/courses/ai/nn.html [Accessed 4 May 2018].
[14] https://medium.com/datathings/neural-networks-and-backpropagation-explained-in-a-simple-way-f540a3611f5e
[15] Adams, T. (2010) 'Can Google Break the Computer Language Barrier', *TheGuardian.com* [online]. Available at https://www.theguardian.com/technology/2010/dec/19/google-translate-computers-languages [Accessed 27 February 2018].

literature has been dedicated to the problem whether automated art deserves copyright protection.

Other aspects of artificial intelligence lead to two challenging features:

–*Autonomy*: Which can be defined as the ability to take decisions and implement them with minimal or without any human intervention.[16] For instance, in composing music a user might need to specify the track length, style or instrumentation, but modern AI composers can create music with no further human influence.[17]

– *Unpredictability:* Which derives from the ability of AI systems to learn autonomously from the surrounding environment and to self-modify their internal algorithms to achieve better results. Programmers may not know exactly to what stimuli and data the system has been exposed to, nor how or when it has reprogrammed itself.[18]

Moreover, as with all computer systems, AI systems can be hacked and could behave unpredictably as a result.

As AI systems become more complex, experts will be increasingly challenged to explain the way algorithms make decisions based on deep learning and neural networks,[19]not only because of the inherent complexity of the technologies and algorithms but also because of the unpredictability of the machine learning process. Perhaps the most mind-blowing example of AI unpredictability was the ability of an AI system to create a 3D replica of our universe, not only without being programmed to do so, but also performing a task that, according to the researchers, exceeded the ability of the machine.[20]

3. Sources of Bad Behaviour

Presuming that the system was not deliberately designed to 'misbehave' and that there are neither (software) bugs nor malware issues, flaws in the design of algorithms, problems with the training data, as well as faults in the implementation of AI systems can lead to skewed automated decisions, which may violate existing laws or commonly accepted principles of morality.

- *Algorithms:* may codify in a formal set of instructions which kind of decision should be taken by a machine and thus are not immune from the prejudices and biases of their programmers.[39] They can even reinforce these prejudices, since they are written and maintained by people and because machine learning algorithms adjust what they do

[16] Palmerini, E., Bertolini, A. et al. (2016) 'RoboLaw: Towards a European Framework for Robotics Regulation', *Robotics and Autonomous Systems,* 86, p 79; Holder, Khurana et al., *ibid* (2016), p 385; Hristov, K. (2017) 'Artificial Intelligence and the Copyright Dilemma', *IDEA* 57(3), p 434; European Parliament resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).

[17] Farrell, J. (2015) 'Artificial Composers: Tools of Modern Composers or Affront to Human Creativity?', *Inquiries Journal,* 7(3), p 1. Available at http://www. inquiriesjournal.com/articles/1017/2/artificial-composers-tools-of-the-modern-musician-or-affront-to-human-creativity [Accessed 26 February 2018].

[18] European Group on Ethics in Science and New Technologies (2018) 'Statement on Artificial Intelligence, Robotics and Autonomous Systems', European Commission, p. 6.

[19] Will Knight, 'The Dark Secret at the Heart of AI', *MIT Technology Review*, Apr. 11, 2017, https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ (visited Mar. 30, 2019)

[20] https://www.livescience.com/65832-ai-creates-model-universe-mysteriously.html.

based on people's behaviour.[40] Prominent cases of poor algorithm design concerned some image recognition applications. For instance, it was revealed that Hewlett-Packard's implementation of a feature-based face localization algorithm did not detect black people as having a face.[41] The algorithm measured the difference in intensity of contrast between the eyes and the upper cheek and nose of a human face and, because of the choice of these parameters, it did not work properly on darker faces in certain light conditions.[42] Similarly, Google Photo's image recognition algorithm started tagging black people as gorillas. In the short term, Google was unable to fix the algorithm and solved the problem by removing words relating to monkeys from Google Photo's search engine.[43]

Data and implementation issues are even more relevant in the field of deep learning, since these algorithms tend to take decisions based on the patterns found in the training dataset and there may be problems of data mismatch. These issues featured in two prominent public management cases. *IBM v Indiana* concerned IBM's problematic modernization of Indiana's welfare case management. The related AI software was full of inherent biases, which tripled Indiana's error rate in denying benefits to eligible citizens, causing harm to applicants and recipients, including some of the most vulnerable families in the state.[21-22] Another case concerned Australian government's automated Online Compliance Intervention (OCI) system, which earned the nickname RoboDebt, that was accused of enforcing 'illegal' debts and forced the Australian government to either wipe or change one of every six debts the OCI raised against welfare recipients.[23]

4. Obsessive secrecy v Transparency

Unsurprisingly, such controversies have led to calls for greater oversight over AI systems, going as far as advocating for the disclosure of algorithms and datasets. The main challenge of public oversight is the impossibility to anticipate the exact behavior of AI machine; because, as seen earlier on, once the AI runs autonomously, it is impossible to foresee the machine behavior in any accurate or meaningful way.

As that would not be enough, AI vendors and developers try to resist the forced disclosure of their AI algorithms. They normally go to great lengths to secure the secrecy of their AI systems, implement a variety of measures to protect secrecy, including security and access control mechanisms, confidentiality agreements, clauses in employment contracts, and even frequent changes in the underlying algorithms powering their AI systems to thwart unscrupulous parties. Google, for example, reportedly made 3,234 updates to its search algorithms[24] - more than eight times the number of updates in 2009[25]. Similarly, Amazon makes frequent changes to its search algorithms that not only

[21] Virginia Eubanks, *Automating Inequality*, St Martins Press, 2018 p. 72

[22] *International Business Machines Corporation v. State of Indiana, acting on behalf of the Indiana Family & Social Services Administration*, 49A02-1709-PL-2006

[23] Paul Karp, Christopher Knaus, '' Centrelink robo-debt program accused of enforcing 'illegal' debts', *The Guardian*, 4 April 2018 [Accessed 14 July, 2019]

[24] RankBrain is an artificial intelligence Google uses in order to serve better search results, particularly for the 15% of daily search queries that Google has never seen before.

[25] https://moz.com/google-algorithm-change

determine how and what products get shown when people are searching on Amazon but are also used to search and remove fake or fraudulent sellers on Amazon's site.[26]

Corporations regard AI as their proprietary trade secret and argue that disclosure would compromise their operations, allowing other parties to manipulate or otherwise exploit vulnerabilities in their systems, undermining their IP assets and consequently have a chilling effect on innovation.

Companies could, for instance, point to the multi-billion dollar cottage industry – Search Engine Optimization (SEO) – that was worth US$65 billion in 2016 and is estimated to grow to US$79 billion in 2020[27] and that is focused, *inter alia,* on attempting to guess how search engine algorithms work in order to attract customers to clients' sites and/or produce better search engine rankings for their clients. While much attention in AI innovation seems to focus on, for example, new forms of machine learning, breakthroughs in artificial vision or natural language processing, this constant creation and implementation of new AI algorithms itself is a form of innovation that keeps large numbers of mathematicians and data scientists employed.

In April 2019, Facebook's founder, Mark Zuckerberg even called for government regulation[28] after months of resisting government oversight[29] though Zuckerberg's move was quickly questioned by critics seeing such a move by Facebook and other companies as attempting to shape future regulations in their favor or counter more drastic proposals[30] or merely outsourcing their oversight responsibilities. Others also pointed out that such regulation may actually benefit current players such as Google and Facebook by making it even tougher for other businesses to compete with them[31].

5. Protecting AI through Intellectual Property

Furthermore, while AI vendors can rely on various forms of intellectual property to keeping competitors from appropriating their technology, there are important reasons why trade secrecy remains an indispensable asset for the protection of AI technology.

---

[26] https://www.trickc.com/blog/amazon-a10-algorithm-updates-its-effect-over-sale-page/

[27] Jayson DeMers, "The SEO industry is worth $65 billion, will it ever stop growing?", 9 May 2016, https://searchengineland.com/seo-industry-worth-65-billion-will-ever-stop-growing-248559 [Accessed 10 July 2019]

[28] Albeit in the area of social mdia content but as AI systems are involved, there are related ethical issues. See Jackie Wattles, Donie O'Sullivan, Facebook's Mark Zuckerberg calls for more regulation of the internet, *CNN*, Mar. 30, 2019, https://edition.cnn.com/2019/03/30/tech/facebook-mark-zuckerberg-regulation/index.html (visited Apr 11, 2019).

[29] Dustin Volz, David Ingram, 'Zuckerberg resists effort by U.S. senators to commit him to regulation' *Reuters*, Apr. 10, 2018, https://www.reuters.com/article/us-facebook-privacy-zuckerberg/zuckerberg-resists-effort-by-u-s-senators-to-commit-him-to-regulation-idUSKBN1HH1CU (visited Apr. 20, 2019).

[30] Harper Neidig, Zuckerberg call for tech rules gets cold reception' *The Hill*, Apr. 3, 2019 https://thehill.com/policy/technology/437055-zuckerberg-call-for-tech-rules-gets-cold-reception (visited Apr. 11, 2019).

[31] John Hawkins, 'The Conservative Case for Breaking Up Monopolies Such as Google and Facebook', *National Review*, May 16 2018, https://www.nationalreview.com/2018/05/breaking-up-tech-giants-conservative-case/ (visited Apr. 20, 2019).

5.1 Copyright and database protection

Leaving aside the problem of copyright protection on AI outputs, the main limitation of copyright, is that while it protects the software behind AI systems, its scope does not extend to the underlying ideas in the software – i.e. the AI algorithms and other processes.[32] Furthermore, traditional copyright does not protect training datasets, since these are not work of art.[33]

Copyright protection for AI was further eroded in the monkey selfie case (*Naruto v Slater*, 2016 WL 363321 (N.D California)) whereby the court reaffirmed that copyright protection was available only to human authors[34]. The U.S. Copyright Office also affirmed this declaring that it would not register works created by non-human authors[35]; meaning that works, including new algorithms, created by an AI would not receive copyright protection

The question is more complex in relation to the sui generis database right under EU law. Some advocate that, in principle, AI datasets might be protected under the EU database directive.[36] Others argue that AI databases seem to fall short of meeting the requirements of being arranged in a systematic or methodical way, of being individually accessible and of being obtained through substantial investments directed at obtaining or presenting the information. Also, the *sui generis right* does not seem to grant control on the semantic level of the content of the data.[37-38] This is also the position of the European Commission, according to which the sui generis right does not apply broadly to the data economy (in particular machine-generated data, big data, AI), covering only databases that contain data obtained from external sources.[39] Other have pointed out that among the difficulties in applying sui generis database protection to big data is to establish what a "substantial part" of big data is.[40]

5.2 Patents

AI developers consistently seek patent protection for their AI technology and even though patents do not protect algorithms *per se*, they can protect systems implementing

---

[32] Craglia, M., Annoni, A. et al. (2018) 'Artificial Intelligence: A European Perspective', European Commission Joint Research Centre, p. 64.
[33] Craglia et al. (2018) *ibid*, p. 64.
[34] Edward Klaris, Lisa Oratz, Protecting Artificial Intelligence Under Copyright Law: Protectability, Authorship, Registration" 17 April, 2018 http://media.straffordpub.com/products/protecting-artificial-intelligence-under-copyright-law-protectability-authorship-registration-2018-04-17/presentation.pdf [Accessed 26, July 2019]
[35] Chapter 306 of the compendium of U.S. Copyright Office Practices states that: "the Office will refuse to register a claim if it determines that a human being did not create the work."

https://www.copyright.gov/comp3/chap300/ch300-copyrightable-authorship.pdf [Accessed 23, July 2019]

[36] Craglia et al. (2018) *ibid*, p. 65; Zeno-Zencovich, V. & Giannone-Codiglione, G. (2016) 'Ten Legal Perspectives on the "Big Data Revolution', *Concorrenza e Mercato,* p. 32.
[37] See Article 1 and 3 Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of database.
[38] Mezzanotte (2017) 'Access to Data: The Role of Consent and the Licensing Scheme' in Lhosse, S., Schulze, R. & Staudenmayer (eds) 'Trading Data in the Digital Economy: Legal Concepts and Tools', NOMOS, p. 165.
[39] (2018) 'The Database Directive, AI and the Data Economy', LexUniversal.com [online]. Available at https://lexuniversal.com/en/news/20629 [Accessed on 26 April 2019].
[40] Zeno-Zencovich, V. & Giannone-Codiglione, G. (2016) *ibid*, p. 32.

those algorithms. However, the patent system is far from an efficient way to protect the interests of innovators.

A first problem is that filing a patent requires the disclosure of important technical details regarding the operation and/or implementation of the invention, with companies potentially giving up more benefits that they receive from patent monopolies, especially in the case of inventions having multiple technical applications. Patents also require time to prosecute, making them less than ideal for protecting new inventions in a very fast moving field.

Enforcing AI patents might also prove a difficult and expensive task. The very nature of AI technology makes monitoring competing products and identifying potential infringers more difficult than for tangible, physical products. For instance, a producer of electric fans could buy one of its competitors' fans, reverse-engineer it, and figure out how it works and whether it was infringing its patent. As an alternative, if its rival's fan is patented, find ways to design around its competitor's patent. Conversely, AI systems may run on the cloud giving no clue about the instructions and processes the system is using. This greatly complicates assessing infringement, because patent holders can only observe the outcomes of the technology used by competitors but not the underlying processes. Machine learning further complicates matters because of difficulties in determining either the relationships that were learned and used to make an inference or how an underlying neural network makes those determinations – or both. That SEO consultants have been able to glean important bits of the operation of algorithms employed by major technology players such as Google, but not reveal the algorithms themselves, demonstrates the difficulties involved. A present, a party interested in determining whether another party were infringing its AI patent would likely have to commence some form of legal action to access the necessary information and even after doing so it may be unable to determine if there was infringement making this a potentially risky and expensive undertaking.

5.3 Trade secret and its advantages

At the international level, the TRIPS agreement protects undisclosed information insofar as the relevant information is kept secret (i.e. not generally known or accessible to competitors), has commercial value because it is secret and has been subject to reasonable steps to keep it secret.[41] The Treaty protects information holders against unfair commercial conducts such as data espionage, misappropriation and breach of contract. Trade secret becomes a form of *de facto* property, i.e. the factual exercise of proprietary privileges by keeping information hidden from competitors. As seen earlier on, trade secrecy is well-known to AI developers, with show a tendency to keep the data to themselves, refusing to share them with competitors and third parties.[42] This practice allows data holders to engage in monopolistic behaviours as *de facto* proprietors of data assets, by enforcing those defensive measures meant to safeguard possessions against

[41] See (1994) Article 39(2) TRIPS Agreement.
[42] Mezzanotte (2017) *ibid,* p. 159.

third parties' intrusions.[43] The inherent limitation of de facto property/trade secret protection is that it protects exclusively to the factual control over the tangible means where the data have been stored, but not the information per se.[44]

There are several reasons that explain the success of trade secret protection:

- *Compatibility with other forms of protection:* Trade secret normally works in tandem with the *de facto* ownership of the data, i.e. the availability of mainframe computers where the big data are stored and processed and of the algorithms capable of extracting economic value from the data.[45] Contractual arrangements might reinforce the de facto exclusive control over the data.[46] It is also very compatible with other form of IP protection, e.g. copyright on the AI algorithm. Mixed strategies are also a solution, with patents protecting the visible features of their technology while trade secret protects the hidden features.

- *Scope of protection:* While trade secrecy does not prevent data holders from competitors' endeavors such as parallel inventing and data self-collection (e.g. linguistic patterns in public documents), trade secret protection works well for the core features of AI systems, since competitors find extremely challenging to reverse engineering such systems.

- *Formalities and length:* Moreover, trade secret does not require registration or disclosure, saving companies from the uncertainties and financial commitments associated with the prosecution of other forms of protection, e.g. patents. Also, it lasts indefinitely, as long as the requirement for protection are met, potentially lasting longer than other forms of protection, and especially patents.

- *Uncertainty:* Lastly, trade secret is also a natural response to the legal uncertainty surrounding artificial intelligence and big data. It is still unclear whether big data are an object of property, especially considering that exclusive rights on non-material entities are normally considered a closed number and that the common take of scholars is that no branch of the IP system currently confers exclusive rights over datasets *per se*.[47] Even admitting that the data can be an object of property, questions remain about the allocation of associated property rights. For instance, it is up to debate whether ownership of personal data falls upon the person from which the data originate or its collector (e.g. data on sleep/wake patterns, or on preferred routes and time of daily commuting).[48] Finally, there are also questions raised on whether data should be considered as a form of consideration,[49] and whether this implies the application of consumer protection rules anytime the data are exchanged as a counter-performance between a consumer and a

[43] Mezzanotte (2017) *ibid,* p. 167.
[44] Mezzanotte (2017) *ibid,* p. 169-70.
[45] Zeno-Zencovich, V. & Giannone-Codiglione, G. (2016) *ibid*, p. 32.
[46] Craglia et al. (2018) *ibid*, p. 65.
[47] Mezzanotte (2017) *ibid*, p. 165; Zeno-Zencovich & Giannone-Codiglione (2016) *ibid*, p. 31.
[48] See Zeno-Zencovich & Giannone-Codiglione (2016) *ibid*, p. 33-6.
[49] See Zeno-Zencovich & Giannone-Codiglione (2016) *ibid*, p. 33-9.

professional.[50] This complex scenario perplexes the relationships between data holders and competitors since the establishment of clear property rules is a necessary pre-requisite to the smooth functioning of a free market. In spite of this, experts seem to convene that it is too early to come up with new rules on data ownership, considering that AI technology is still in its infancy and that the characteristics and functioning or bid data markets are yet to be understood.[51]

6. The challenges for regulation

While calls for greater AI transparency and oversight have increased, the actual process of formulating, monitoring and enforcing relevant guidelines is complicated. Oversight for ethical purposes, for instance, will be difficult, if not impossible to achieve in practice and will be complicated by cultural and political considerations. Google, for example, tried and failed to form an AI ethics board. The rapid dissolution of its Advanced Technology External Advisory Council was apparently due to some members of the group founding the right wing views of others in the group objectionable, highlighting the difficulty of even forming a body.[52]

Another problem concerns the interdisciplinary nature of AI studies, with different academic disciplines are studying AI ethics issues from a variety of perspectives. This often translates into a lack of a shared language and common methods making discourse, synthesis, and coordination a challenge, which in turn complexes the job of policy makers. Indeed, the latter find it nearly impossible to process and understand this avalanche of research and thinking, and in particular to determine which risks are already being tackled through technical measures or better practices, or what risks are relatively underserved.[53]

Finally, different jurisdictions endorse different views of what the optimal trade-off between transparency and innovation/corporate interests should be. In Europe, there seems to be a movement towards mandating greater transparency of AI, which has been embedded in the General Data Protection Regulation (GDPR).[54] More in detail, some of the provisions of the GDPR[55] prohibit made solely on the basis of automated processes,[56] and specify the need for safeguards to protect citizens[57] as stepping stones to new regulations and mandated algorithmic disclosure. Conversely, the US appear to be putting trade secrets ahead of transparency.[58] For instance, the anti-circumvention provisions of

[50] See Zeno-Zencovich & Giannone-Codiglione (2016) *ibid*, p. 40-3.
[51] Craglia et al. (2018) *ibid*, p. 65.
[52] Mariella Moon, Google dissolves newly formed AI ethics board', *Engaget*, Apr. 4, 2019 https://www.engadget.com/2019/04/04/google-dissolves-ai-ethics-board/ (visited Apr. 13, 2019)
[53] Valerie Frissen, Gerhard Lakemeyer, Georgios Petropoulos, Ethics and artificial intelligence, *bruegel*, http://bruegel.org/2018/12/ethics-and-artificial-intelligence/#_ftnref1 (visited Apr. 13, 2019).
[54] See for instance Recital 71 as well as Articles 13.2(f), 14.2(g), 15.1(h) GDPR.
[55] Notably in Arts. 13.2(f), 14.2(g),15.1(h), Recital 71 GDPR
[56] Art. 22.1 GDPR
[57] Art. 22.4 GDPR
[58] Rebecca Wexler, "Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System", *Stanford Law Review*, May 2018, Volume 70, 70 STAN. L. REV. 1343 (2018)

the Digital Millennium Copyright Act (DMCA) [59] can be used to block research on AI algorithms and even accessing AI systems to study their algorithms might contravene provisions the Computer Fraud and Abuse Act of 1986 (CFAA).[60] However, it should also be noted that in June 2019, US Senator Josh Hawley introduced the Ending Support for Internet Censorship Act to amend s. 230 of the Communications Decency Act and force large technology companies[61]to submit to external audits to ensure the political neutrality of their systems.[62]

As for China, it remains to be seen in which direction China will take, since the country is in the verge of introducing new laws on online privacy[63] and cyber-security.[64]

7. Potential Solutions

Two implications of mandated algorithmic disclosure are the potential weakening of trade secret protection and, because it makes discovery of algorithms easier, the lowering of the cost and risk for AI patent holders to initiate patent infringement actions thus raising the specter of aggressive patent trolling making it harder for the AI community to undertake innovative activities.

If regulation is inevitable, we therefore need to explore potential solutions to strike a reasonable balance between transparency and innovation.

7.1 Explainable Artificial Intelligence

Some have suggested the use of Explainable Artificial Intelligence' (Explainable AI or XAI) as a potential solution to the dimness of AI decision-making process. The expression refers to a suite of machine learning techniques that enable human users to understand,

[59] 17 U.S.C. §§ 1201-1202 (2012).

[60] Concerns were raised that accessing a computer to study the AI algorithms it is running would exceed authorized access thus violating site User Agreements and thus contravening 18 U.S. Code § 1030 which states that "Fraud and related activity in connection with computers ….(2) intentionally accesses a computer without authorization or exceeds authorized access, and thereby obtains—….(C) information from any protected computer"

[61] Defined as companies with over 30 million active monthly users in the U.S., over 300 million active monthly users worldwide or more than US$500 million in global annual revenue

[62] Senator Hawley Introduces Legislation to Amend Section 230 Immunity for Big Tech Companies', 19 June 2019 , https://www.hawley.senate.gov/senator-hawley-introduces-legislation-amend-section-230-immunity-big-tech-companies

[63] On June 13, 2019, the Cyberspace Administration of China (the "CAC") released Draft Measures on Security Assessment of Cross-Border Transfer of Personal Information ("Draft Measures") for public comment https://www.huntonprivacyblog.com/2019/06/19/china-issues-draft-regulation-on-cross-border-transfer-of-personal-information/ [Accessed 10 July 2019]

[64] On May 28, 2019, the Cyberspace Administration of China ("CAC") released draft Data Security Administrative Measures (the "Measures") for public comment. The Measures, which, when finalized, will be legally binding, supplement the Cybersecurity Law of China (the "Cybersecurity Law") that took force on June 1, 2017, with detailed and practical requirements for network operators who collect, store, transmit, process and use data within Chinese territory. https://www.huntonprivacyblog.com/2019/06/10/china-issues-draft-of-data-security-administrative-measures/ [Accessed 10 July 2019].

appropriately trust, and effectively manage artificial intelligence outputs,[65] and that improve the transparency of AI systems[66] by clearly laying out their objectives.

Several actors, including the US Department of Defence are working in this direction.[67] One of the most interesting projects is run by a team at the University of Berkeley and involves lashing two neural networks together, tasking one to describe the inner procedures running inside the other.[68] Other research projects are trying to teach AI to explain itself by means such as providing examples, describing the evidence relied upon or how it weighted different variables.[69]

Scholars and experts have already emphasized how we cannot blindly outsource moral decisions to machines.[70] Understanding AI internal logic is a necessary step to build trust and confidence in AI decisions, with experts emphasizing the need to support research on AI explainability.[71] XAI would help AI meeting the requirements set by the EU GDPR, which, as anticipated, entitles data subjects with a right to explanation, i.e., the right to obtain meaningful information about the logic involved in automated decision making.[72] XAI would be especially useful considering that the GDPR prescribes that the relevant information has to be provided in a concise, transparent, intelligible and accessible manner.[73] [74]

However, XAI is far from being universally accepted. Some technologists have warned against potential performance degradations, and additional overheads associated with XAI[75] such as documentation to explain how algorithms are supposed to work. Moreover, it remains unclear whether non-technical reviewers could understand such documentation.

---

[65] Gunning, D. (2018) 'Explainable Artificial Intelligence (XAI)', *DARPA* [online]. Available at https://www.darpa.mil/program/explainable-artificial-intelligence [Accessed 22 May 2018].

[66] Finale Doshi-Velez*, Mason Kortz, Accountability of AI Under the Law: The Role of Explanation, Nov. 3, 2017, arXiv:1711.01134v2 (visited Apr 20, 2019).

[67] Gunning, D. (2018) 'Explainable Artificial Intelligence (XAI)', *DARPA* [online]. Available at https://www.darpa.mil/program/explainable-artificial-intelligence [Accessed 22 May 2018].

[68] Kuang, C. (2017) 'Can AI Be Taught to Explain Itself?', *The New York Times* [online]. Available at https://www.nytimes.com/2017/11/21/magazine/can-ai-be-taught-to-explain-itself.html [Accessed 22 May 2018].

[69] Kuang (2017) *ibid.*

[70] Zeynep Tufekci (2016) 'Machine Intelligence Makes Human Morals More Important' [online]. Available at https://www.ted.com/talks/zeynep_tufekci_machine_intelligence_makes_human_morals_more_important [Accessed 23 October 2018].

[71] Villani et al. (2018) *ibid*, p. 115-6.

[72] Article 14 'Regulation 201/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data'.

[73] Article 12 'Regulation 201/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data'.

[74] See Andrew D Selbst Julia Powles (2017) 'Meaningful Information and the Right to Explanation' *International Data Privacy Law*, 7(4), pp. 233–242.

[75] David Weinberger, 'Optimization over Explanation', *Medium*, Jan. 28, 2018, https://medium.com/berkman-klein-center/optimization-over-explanation-41ecb135763d (visied Apr. 24, 2019).
See also: David Weinberger, 'DON'T MAKE AI ARTIFICIALLY STUPID IN THE NAME OF TRANSPARENCY', *Wired*, Jan. 28, 2018, https://www.wired.com/story/dont-make-ai-artificially-stupid-in-the-name-of-transparency/ (visited Apr. 24 2019)

7.2 Applying drug disclosure to Artificial Intelligence

Another possibility is to entrust a centralized agency with keeping the confidential information on artificial intelligence and managing an on-demand access system whereby access to said information is granted to third parties only when particular conditions are met. This solution is well-known in the pharmaceutical sector, which had its own struggles to balance the need for greater disclosure of information against protection of trade secrets.

Centralized pharmaceutical agencies are tasked with the evaluation of the safety, quality and efficacy of a drug, features that are proved by the applicant through the submission of a dossier containing the result of the drug testing in thousands of patients.[76] Given the high cost of drug testing, which according to some appraisals can overcome the 1 billion dollars,[77] it should not surprise that pharmaceutical companies have a strong interest in keeping the dossier secret, especially from competitors. Their stance is that the dossier contains proprietary information, and that disclosure might harass their commercial interests and strategies. There are also legitimate concerns that disclosure of clinical dossier may allow competitors to disguise the disclosed data as independently developed, and then submit the doctored data to regulatory agencies in foreign jurisdictions. This would allow competitors to enter foreign markets even before data developers, gaining important first-mover advantages. These are also guaranteed by the legal system as a period of exclusivity over the utilization of the data, which impedes competitors to rely on the first authorization to market their generic drugs.[78]

Unfortunately, keeping the data secret presents also severe backlashes and scholars, physicians and non-governmental organizations have all advocated for greater clinical trials disclosure.[79] There are several benefits to disclosure:

- *Independent review:* Disclosure of trials reports might allow for independent review of clinical trials results aimed at detecting flaws in the regulatory assessment potentially leading to revelations on drug characteristics in turn leading to the removal of dangerous products from the market, as well as paving the way towards important decisions on drugs relabeling or change of dosage. Moreover, independent review could be a useful counterweight to pharmaceutical companies push to ensure pharmaceutical authorization independent from the real clinical value of their drugs.

- *Innovation:* Science relies on the accumulation of knowledge and making clinical data freely available to the scientific community research would undoubtedly fuel the progress of the medical science and innovation. More in detail, publication of information could

[76] Gabriele Spina Alì (2017) 'TRIPS and disclosure of clinical information: An intellectual property perspective on data sharing', *Journal of World Intellect Prop.* 2017; 20:24–56. https://doi.org/10.1111/jwip.12071.
[77] Mestre-Ferrandiz, J., Sussex, J., & Towse, A. (2012). The R&D cost of a new medicine. United Kingdom: Office of Health Economics.
[78] https://www.ifpma.org/wp-content/uploads/2016/01/IFPMA_2011_Data_Exclusivity__En_Web.pdf
[79] See the list of organizations mentioned in Gøtzsche, P. C. (2011). Why we need easy access to all data from all clinical trials and how to accomplish it. Trials Journal, 12(249), 1–14.

foster follow-on innovation and second medical use discoveries, as well as provide valuable information for understanding human physiology and drug absorption.

- *Personalized medicine:* More accurate knowledge of drugs pharmacodynamics and adverse effects could provide physicians with a better understanding of drugs functioning, that would foster more accurate drug prescriptions and could allow patients to take informed therapeutic decisions. Greater disclosure may even enable development of predictive models for patient selection to appropriate treatments through machine learning systems and could obviate the need for wasteful and expensive redundant trials.

Pharmaceutical agencies, and especially the European Medicines Agency (EMA), have employed three main strategies to strike a fair balance between the corporate interest in secrecy and the public interest in disclosure. Notably, these strategies can be combined in a single policy to best serve the interests of both innovators and the public.

- *Confidential agreements:* Some agencies allow access to clinical documents only under the express acceptance of confidentiality terms. These should include, inter alia, the obligation of recipients not to utilize the data in support to marketing applications. Data access may be refused to applicants not offering adequate warranties. For instance, the EMA 2015 policy allows disclosure only to those recipients who pledge not to seek marketing authorization outside the EU and use it only for non-commercial purposes.[65] Similarly, the US Secretary's Advisory Committee on Human Research Protection has called for terms of use, with penalties for violation "under which the recipient would pledge to use the data only for the purposes specified; not to disclose the data to others.. and not to try at any point to re-identify subjects".[80]

- *Delaying publication:* Another measure is to allow disclosure after a fixed period of time has elapsed from the moment of national authorization. This is to allow developers to comply with different regulatory requirements and obtain authorization in all the territories of commercial interest, as well as allowing sponsors to seek approval for second medical uses of the compound.[81] For instance, the Hatch-Waxman Act stipulates that trials information shall be made available to the public at the end of the data exclusivity period, or earlier in case the drug application is abandoned or refused.[82] Similarly, EU Regulation 536/2014 stipulates that in general, data included in a clinical trials report should not be considered commercially confidential once a marketing authorization has been granted, the granting procedure is completed or the application has been withdrawn.[83] Also Brazilian law allows disclosure after the expiry of data exclusivity.[84]

- *Partial disclosure:* A last measure is to limit publications to trials excerpts or summaries. Ideally, they should be sufficiently detailed to allow independent review, but they should lack some of the core information necessary for marketing authorization.

[80] Institute of Medicine of the National Academies (IOM). (2015). Sharing clinical trial data: Maximizing benefits, minimizing risk. http://www.nap.edu/catalog/18998/sharing-clinical-trial-data-maximizing-benefits-minimizing-risk. (Accessed on March 20, 2016), p. 115.
[81] Institute of Medicine, 2015, p. 115.
[82] See Hatch Waxman Act, 21 U.S.C. § 355(l)(l).
[83] (2014) Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC, Premises, (68).
[84] See Article 3(1) of Acts of the executive department, Presidential Decree No. 69 of September 26, 2002, Brazil.

Some scholars have proposed to publish only the data on a product safety (and perhaps efficacy) while retaining the pharmacokinetic and pharmacodynamics data. Safety trials alone indeed are not enough to support marketing authorization without evidence of bioavailability, thus preventing competitors from obtaining licenses in most jurisdictions.[85] However, the problem with partial disclosures would be that part of the information could still be used for marketing purposes.[73] In the EU, clinical data are to be published in a publicly accessible database, after any commercial confidential information is removed from the dossier.[86] As a general rule, clinical trials are not considered confidential information, but applications can object to the publication of the trials and propose a redaction of the information to be disclosed.[87] Among the relevant factor to evaluate are the nature of the product, the competitiveness of the relevant market, the approval status in other jurisdictions and the novelty of the drug, the clinical tests as well as the opportunity to develop follow-on drugs.[88]

7.2.3 Registration of AI technology: Current Proposals

Even though the utilization of third party data to seek authorization in foreign jurisdictions remains a prerogative of the pharmaceutical sector, some of the solutions adopted in this field can be transplanted to AI with the necessary adjustments. Firstly, it is worth recalling that while committees with a merely advisory function to governments and legislative bodies are already in place,[89] some countries are considering the creation of a specialized AI agency.[90] However, it remains unclear what would be the precise tasks of said agency, with scholars arguing that these agency should perform advisory and enforcement functions, including helping citizens to find redress for violation of AI rules, but also raising public awareness on AI issues. [91] Most importantly, the agency would be responsible for the protection of public welfare through the scientific evaluation and supervision of AI systems. The agency would also establish a post-release monitoring system, to ensure that AI technology functions properly overtime.[92]

To help the agency in its supervisory function, some have advocated the creation of an AI database, where information on AI systems should be compulsory registered.[93] To a lesser level, the European Commission has already proposed the development of a common open-access European Library of algorithms to help the private and the public sectors to identify and acquire whichever solution works best for their needs.[94]

However, more comprehensive proposals on AI repositories are yet to be formulated, and

[85] Kesselheim & Mello, 2007, pp. 489–490; Perehudoff, 2013, p. 31.
[86] (2014) EMA Policy/0070, European Medicines Agency policy on publication of clinical data for medicinal products for human use, EMA/240810/2013, 10-5,
[87] (2014) EMA, *ibid* 16-20.
[88] (2014) EMA, *ibid* 16-20.
[89] See for instance the EU CommissionHigh-Level Expert Group on AI (AI-HLEG). See Craglia et al. (2018) *ibid*, p. 38.
[90] See European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)); Craglia et al. (2018) *ibid*, p. 69.
[91] Floridi, Cowls et al. (2018) *ibid*, p. 703; Craglia et al. (2018) *ibid*, p. 69.
[92] Floridi, Cowls et al. (2018) *ibid*, p. 703.
[93] Craglia et al. (2018) *ibid*, p. 69.
[94] Craglia et al. (2018) *ibid*, p. 38.

legislators need to come up with clear policies regarding:

- *What information to register:* Should said information be limited to the name of the software, its proprietor, and information on its technical field and abilities or should also comprise also source code, the algorithm on which the AI is running and the databases on which it has been trained?
- *Legal effects:* Rules will have to be established on the legal effects of registration, e.g. whether registration confers ownership or absolves a mere evidentiary function.
- *The rules on the accessibility of the database:* This includes rules on who, when and for what purposes should be allowed to obtain information from the database. This is a particularly delicate topic since the information might be covered by intellectual property and trade secret laws.

7.2.4 Applying disclosure to AI system

The present paper advocates for the creation of a network of AI repositories to allow parties with legitimate needs (e.g. to assess the safety of an algorithm) to get access to the relevant technology. These repositories should be kept offline, to preclude hacking or other cyber attacks. They are also managed at the national level, to account for national cultural differences that may impact*, inter alia*, ethical standards applied to AI-related reviews or even training data. Nevertheless, information could be shared among agencies in a manner similar to the international sharing of safety data for aircraft among national air safety regulators. As for the rules on registration, these should as follows:

- *What to register*: The AI repository should hold information on AI algorithms and the AI training data, since access to both of them is necessary to assess and review AI technology. Also, since both AI data and algorithms tend to evolve overtime, follow-up procedures whereby AI owners update the information on periodic basis (e.g. six months) should be established. Besides that, the repository should also contain information such as reports made by recognized, qualified reviewers. They should also include information on the variable used with their values and deviations and the amount and type of training data used.[95]

- *Rules on data access:* Information in each repository will be strictly guarded to preclude competitive abuse. Access would be allowed only under the express acceptance of strict confidentiality terms including, *inter alia*, restrictions on recipients' ability to utilize the information, for instance in support of short term product development, circumvention of legal rights or even marketing applications. Agencies tasked with maintaining such repositories should evaluate the guarantees offered by information recipients to take all necessary measures to avoid data leakage and prevent free rider conducts. Agencies may further reinforce confidential obligations through monetary deposits, performance bonds, fines, or penalties. Information access may be refused to applicants not offering adequate warranties.

[95] Craglia et al. (2018) *ibid*, p. 59.

- *Legal effects:* Finally, since conditioning the ownership of AI systems to registration might create friction with the rules stipulated for other IP rights, and copyright in particular, registration should serve as a legal presumption of ownership (e.g. copyright in the algorithm). This solution appears consistent with copyright rules, might increase the certainty of the law and lessen the workload of courts in the case of registration. It also confers on AI developers a legal benefit to contrast their diffidence towards a registration system.

Finally, one must note that the other two remedies utilised in the pharmaceutical sectors, i.e. *delayed* and *partial disclosure* have little role to play in AI disclosure policies. As for the former, there is little necessity to give AI holders a period of time to seek registration in other jurisdictions once AI technology is not divulgated to competitors. As for partial disclosure, this might make sense in relation to pharmaceutical dossiers, which contain separate sections about safety, efficacy and quality. By contrast, having access to only a sample of the training data might affect the ability of reviewers to ascertain how the AI behaves and the same goes for disclosing the algorithms separately from the training data or vice versa.

8. Conclusions

In conclusion, while there are legitimate safety and other concerns surrounding AI, there is also a need to protect innovation in this area. Unchecked disclosure could expose hitherto secret algorithms potentially compromising existing IP regimes and placing a chill on innovation.

Establishing such system of restricted disclosure would be a challenge and strict access criteria would need to be established, something that could be a potential subject for additional research. Nevertheless, such as system could be a workable compromise providing the opportunity to access AI algorithms for legitimate review or other purposes while protecting vendors' trade secrets.

To balance the legitimate needs AI vendors and developers have to protect their works and parties against equally valid concerns about AI safety and fairness requires compromise. While the proposed national repositories scheme would not address some issues, such as whether reviewers could be adequately vetted, and whether companies based outside the geographic boundaries of some of the repositories would subject their AI to such treatment, it does offer one solution to the problems of modern day AI.