European Journal of Law and Technology, Vol 12 No.3 (2021)

EJLT European Journal of Law and Technology

Artificial Intelligence between Transparency and Secrecy: From the EC Whitepaper to the AIA and Beyond

Gabriele Spina Alì & Ronald Yu[*]

Abstract

As a response to the risks posed by Artificial Intelligence (AI), policymakers are establishing oversight mechanisms to ensure that AI technology complies with the applicable legal and ethical standards. These new regulatory frameworks affect a plethora of actors which store, record and disseminate AI proprietary information, possibly curtailing third-parties' intellectual property and trade secret in particular. This paper analyses the tension between transparency and secrecy in the context of the establishment of an EU conformity assessment for AI technology. After unveiling some of the shortcomings of the draft Artificial Intelligence Act, it proposes to borrow some of the solutions adopted in the pharmaceutical sector to AI. In doing so, it points to an on-demand access scheme that complies with the principle of proportionality and strikes a reasonable balance between public and commercial interests.

Keywords: Artificial Intelligence; Conformity Assessment; Trade Secret; Transparency; Data access; Artificial Intelligence Act.

1. Introduction

Legislators worldwide are striving to create a legal framework to best harness the benefits of artificial intelligence (AI) while mitigating the risks posed by the technology.[1] In particular, there have been warnings that AI can circumvent laws and regulations and even infringe upon fundamental rights, such as the principle of non-discrimination, freedom of

[*] Gabriele Spina Ali, Senior Research Fellow, Max Planck Institute for Innovation and Competition. Executive Editor, GRUR International: Journal of European and International IP Law; Ronald Yu, Research Associate, The Chinese University of Hong Kong.

[1] Future of Life Institute (2020) 'AI Policy China' [online], available at https://futureoflife.org/2018/07/12/ai-policy-china/ [Accessed 16 December 2021]; (2017) 'Draft AI R&D Guidelines for International Discussions' [online], available at https://www.soumu.go.jp/main_content/000507517.pdf [Accessed 27 April 2020]; Craglia, M., Annoni, A. et al. (2018) 'Artificial Intelligence: A European Perspective', *European Commission* [online], available at https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/artificial-intelligence-european-perspective [Accessed 17 July 2019].

expression and assembly, or the right to a fair trial.[2] These fears materialised in cases where algorithms have silenced the speech of dissenting journalists on social platforms,[3] have shown gender and racial biases,[4] and even caused the jailing of innocent people.[5] As the evidence suggests, such risks can only increase as governments gradually turn to AI to automate decision-making in public services.[6]

As a potential remedy, experts have intensively discussed the establishment of regulatory frameworks for the surveillance, monitoring, and ex-ante authorisation of AI technology. To reach informed decisions about product safety and reliability, these frameworks require the analysis of providers' confidential information, such as software and the training datasets of AI technology. For this reason, there is an inevitable tension between the commercial interest of providers in maintaining their technology secret on the on hand, and the public interest in a fair, transparent, and accountable regulatory environment on the other. It is clear that democratic control over public decisions can only be exercised through public access to the documentation held by public agencies.

This paper analyses the conflict between secrecy and transparency in the context of the establishment of an EU oversight system for AI technology. In particular, it will examine the recent European Commission (EC) proposal for an Artificial Intelligence Act (AIA), which describes an ex-ante conformity assessment and ex-post monitoring mechanism for high-risk AI products.[7] After carrying out an analysis of the relevant provisions on confidentiality and transparency, the paper submits that the AIA unreasonably privileges the former over the latter. As a solution, by drawing on the pharmaceutical sector, the research proposes an on-demand access scheme capable of balancing the public interest in a safe AI technology and the trade secrets of firms.

The remainder of the paper is structured as follows. Section two begins by describing the threats associated with artificial AI, the sources of machine misbehaviour and the case for

[2] EU Commission (2020) 'White Paper on Artificial Intelligence – A European Approach to Excellence and Trust', Brussels, 19.2.2020 COM(2020) 65 final, p. 9; European Commission (2021) 'Proposal for a Regulation of The European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, Brussels, 21.4.2021 COM(2021) 206 final 2021/0106 (COD), 2-3.

[3] Bohkary, A. (2018) ''The Good Censor': Leaked Google Briefing Admits Abandonment of Free Speech for 'Safety And Civility'', *Breitbart* [online], available at 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].

[4] Dastin, J. (2018) 'Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women', *Reuters* [online], available at https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [Accessed 17 July 2019].

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

[6] Ada Lovelace Institute (2021) 'AI Now Institute and Open Government Partnership. Algorithmic Accountability for the Public Sector' [online], available at: https:// www.opengovpartnership.org/documents/ algorithmic-accountability-public-sector/ [Accessed on 05 November 2021].

[7] See Chapter 4 and 5, EC (2021) AIA, *supra note* 2.

public oversight of AI technology. Section three explains companies' interests in secrecy. Section four delves into the topic of trade secret protection and how this constitutes a boundary to disclosure policies. Section five analyses the current European approach to AI transparency, tracing the evolution of the relevant rules from the 2020 Commission Whitepaper to the AIA. Section six proposes an on-demand access scheme modelled after pharmaceutical regulation, showing its applicability to AI. Section 7 emphasises both the advantages and compatibility with EU law of the proposed model. Section 8 recapitulates and concludes with some final thoughts on the future legislative action at the EU level.

2. Artificial Intelligence, its perils, and the call for transparency

AI systems based on machine learning identify recurring patterns in large sets of data and make predictions based on them. The human contribution to the machine self-learning process is normally limited to writing the initial algorithm and in some cases, supervising the machine during its learning process. AI systems will produce different outputs based on their algorithms and the data they have been exposed to. Thus, these two components are the natural targets of any attempt to regulate artificial intelligence.

2.1 Sources of machine misbehaviour

Presuming that the system was not deliberately designed for malicious purposes or was compromised by a cybersecurity attack,[8] flaws in the design or implementation of AI may cause machine misbehaviour. For instance, the limited ability of AI to generalise well from training data might result in failures when the machine is exposed to unexpected inputs.[9] Likewise, errors in the definition of the context in which a machine operates might lead to damages if an AI system is not properly programmed to avoid causing negative side effects to its working environment.[10] However, in most cases, failures depend either on flaws in the design of algorithms or the training datasets.[11]

Algorithms are the formal set of instructions given to the machine, allowing it to operate and to extract and analyse patterns in big data corpora. Unlike traditional computing, AI algorithms have the ability to self-modify based on past experience similarly to biological brains, improving over time.[12] 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 onwards.[13]

[8] Comiter, M. (2019) 'Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do about It', *Belfer Center for Science and International Affairs, Harvard Kennedy School* [online], available at https://www.belfercenter.org/publication/AttackingAI [Accessed 4 November 2021].

[9] European Commission (2021) 'AIA Impact Assessment' SWD(2021) 84 Final, p. 14.

[10] EC (2021) AIA Impact Assessment, p. 14.

[11] EC (2020) *supra note* 2, p. 11.

[12] 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].

[13] Moawad, A. (2018) 'Neural Networks and Back-propagation Explained in a Simple Way' [online], available at https://medium.com/datathings/neural-networks-and-backpropagation-explained-in-a-simple-way-f540a3611f5e [Accessed 27 December 2021].

Algorithms codify in a formal set of instructions the decisions to be taken by machines, and therefore, are not immune from the prejudices and biases of their programmers. Prominent public management cases well-exemplify the risk of poor algorithm design, with AI systems tripling the error rate in denying public benefits to eligible citizens,[14] or enforcing illegal debt notices to taxpayers.[15]

Databases are the second cornerstone of AI technology. AI algorithms analyse large data sets (Big Data) and identify statistical patterns from them. For instance, 'Google Translate' uses a statistical machine engine, which identifies linguistic patterns in millions of United Nations and EU Parliament documents.[16] These databases are used to 'train' AI algorithms, which will start making predictions based on the patterns identified in the training datasets. Unlike traditional computing, in some cases, AI databases are dynamic or 'open'. Social platform algorithms, for instance, are relentlessly fed new data from users and subscribers. Hidden biases in big data lead to misbehaviours anytime a machine replicates a biased pattern for future decisions. For instance, some AI have systematically discriminated against women based on past recruitment practices, or penalised minorities during financial and criminal justice profiling.[17]

2.2 The call for transparency

Not only is AI far from infallible, but its inherent characteristics make the call for oversight exceptionally compelling. This is the case due to the ability of AI systems to make decisions and implement them with minimal or without any human intervention (i.e., autonomy),[18] and their capacity to learn and self-modify without their programmers being aware of the changes (i.e., unpredictability).[19] The same goes for the difficulties in explaining how or why

[14] Eubanks, V. (2018) 'Automating Inequality', St Martins Press: US, p. 72; International Business Machines Corporation v. State of Indiana, acting on behalf of the Indiana Family & Social Services Administration, 49A02-1709-PL-2006.

[15] Karp, P. & Knaus, C. (2018) 'Centrelink Robo-debt Program Accused of Enforcing Illegal Debts', *The Guardian* [online], available at https://www.theguardian.com/australia-news/2018/apr/04/centrelink-robo-debt-program-accused-of-enforcing-illegal-debts [Accessed 14 July 2020].

[16] Adams, T. (2010) 'Can Google Break the Computer Language Barrier', *The Guardian* [online], available at https://www.theguardian.com/technology/2010/dec/19/google-translate-computers-languages [Accessed 27 February 2018].

[17] Gutierrez, D. (2019) 'AI Black Box Horror Stories', *ODSC* [online], available at https://opendatascience.com/ai-black-box-horror-stories-when-transparency-was-needed-more-than-ever/ [Accessed 11 March 2020].

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

[19] European Group on Ethics in Science and New Technologies (2018) 'Statement on Artificial Intelligence, Robotics and Autonomous Systems', *European Commission*, p. 6; Pappas, S. (2019) 'AI Created a 3D Replica of Our Universe. We Have No Idea How It Works' [online], available at https://www.livescience.com/65832-ai-creates-model-universe-mysteriously.html [Accessed 21 April 2020].

a machine has taken a given decision (opacity), a problem that will become even more challenging as AI becomes more complex.[20-21]

It is true that human decision-making too is far from infallible and that AI often replicates human biases. However, several considerations make the flaws in automated decisions far more dangerous than human mistakes. First, the higher performance of machines entails a much larger effect on society.[22] For example, while a bank employee might unconsciously assign a higher mortgage rate to an applicant from a minority group, a software processing thousands of files per day might generalise this bias to any applicant with an Afro-American sounding name.[23] Secondly, AI have a self-replicating nature, and a tendency to reinforce small-scale biases. This can happen when an AI system self-feeds on its outputs leading to the self-replication of flawed outputs[24] or when a system treats data as immutable information and traps people into social stereotypes.[25] Finally, human conduct is controlled by social and legal mechanisms that, although far from perfect, are meant to correct misbehaviours in the short and long term. Wrong human decisions can be appealed, whereas monitoring procedures, audits and periodical reviews can remedy flawed decision making. All these measures are difficult to apply to automated processes, considering both the opacity of AI technology and the unwillingness of providers to open up to public scrutiny.[26]

Against this background, the call for transparency rests on the need to look inside AI technology, in order to try to fully understand its logic and regulate its behaviour. Under this vantage, regulating AI can be considered a two-step process. The first step consists of establishing new legal standards to make AI more equitable, trustworthy, and fair. Respect for human rights must be embedded into the technology, while rules regarding the traceability and impartiality of training data as well as guaranteeing a minimum of human oversight over AI decisions must be observed.[27] The second step relates to the enforcement

[20] Knight, W. (2017) 'The Dark Secret at the Heart of AI', *MIT Technology Review* [online], available at https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ [Accessed 30 March 2019].

[21] For instance, an AI called Deep Patient self-trained to diagnose psychiatric disorders such as schizophrenia with great accuracy. This happened notwithstanding the struggle of the scientific community to understand the causes of schizophrenia, so that it remains unexplained how Deep Patient reached its conclusions. See Gutierrez (2019) *supra note* 17*.*

[22] EC (2020) *supra note* 2, p. 11.

[23] Bartlett, R., Morse, A. et al. (2019) 'Consumer Lending Discrimination in the Fintech Era' [online], available at https://faculty.haas.berkeley.edu/morse/research/papers/discrim.pdf [Accessed 21 April 2020].

[24] Garcia, M. (2016) 'Racist in the Machine: The Disturbing Implications of Algorithmic Bias', *World Policy Journal* 33(4), p. 113; Caliskan, A., Bryson, J. et al. (2017) 'Semantics Derived Automatically from Language Corpora Contain Human-like Biases', *Science* 356, pp. 183-186.

[25] Chowdhury, R. (2019) 'How to Stop AI from Reinforcing Biases', *Accenture* [online], available at https://www.accenture.com/us-en/insights/artificial-intelligence/stop-ai-reinforcing-biases [Accessed 21 April 2020].

[26] EC (2020) *supra note* 2, p. 11.

[27] See Article 15, EC (2021) AIA, *supra note* 2, prescribing standards of accuracy, robustness and cybersecurity.

of the prescribed standards, which includes transparency and accessibility mechanisms to ensure the compliance of AI providers.[28] At stake is the accountability of automated decision-making processes, the possibility to prevent potentially harmful conducts and to correct any source of unequal, illegal or undesirable behaviour.[29]

3. Corporate secrecy

It is common knowledge that corporations look at their confidential information as one of their most valuable strategic assets. To give some figures, some reports have estimated the cost of corporate espionage to be as high as $1.7 trillion worldwide in 2018 alone,[30] while the impact of cybercrime alone to be $600M in the same year.[31] Hence, companies go to great lengths to secure the secrecy of their AI systems. These include security and access control mechanisms, confidentiality agreements, clauses in employment contracts, and even frequent changes in the algorithms powering their AI systems to thwart unscrupulous parties. For example, in 2018 Google made 3,234 updates to its search algorithms[32] and over 4,500 changes in 2020.[33] It should therefore be no surprise that corporations openly oppose transparency and try to resist the forced disclosure of their AI components.[34] Two main reasons explain the opposition by companies towards transparency.

The first leverages on the rationale underpinning IP protection: opening up to public scrutiny would allow competitors to free-ride on innovator-based technology and reduce the latter's competitive edge. In the end, this could have a chilling effect on innovation, since firms would not commit large sums to R&D without being able to reap the benefits of their investments.[35] This is especially true considering the absence of clear and suitable

[28] EC (2020) *supra note* 2, p. 9-19.

[29] Maggiolino, M. (2019) 'EU Trade Secrets Law and Algorithmic Transparency', *AIDA*, pp. 4-5.

[30] Wimmer, B. (2018) 'Why There Are Bigger Threats to Your Business than Cyber Attacks, G4S.com [online], available at https://www.g4s.com/news-and-insights/insights/2018/09/03/why-there-are-bigger-threats-to-your-business-beyond-cyber-attacks [Accessed 12 March 2020].

[31] Lewis, J. (2018) 'The Economic Impact of Cybercrime,' *McAfee – Center for Strategic and International Studies* [online], available at https://www.csis.org/analysis/economic-impact-cybercrime [Accessed 12 March 2020]; Grobman, S. (2018) 'Impact of Cybercrime: Why Cyber Espionage isn't Just the Military's Problem'. *McAfee* [online]., available at https://www.mcafee.com/blogs/enterprise/economic-impact-cybercrime-cyber-espionage-isnt-just-militarys-problem/ [Accessed 16 December 2021].

[32] (2019) 'Google Algorithm Update History', *Moz.com* [online], available at https://moz.com/google-algorithm-change [Accessed 12 March 2020].

[33] Schwartz, B. (2021), 'Google made 4,500 changes to search in 2020', *Search Engine Land*, available at https://searchengineland.com/google-made-4500-changes-to-search-in-2020-351445 [Accessed 5 November 2021].

[34] For instance, the 2019 'Corporate Accountability Index' authored by the Ranking Digital Right project, outlined several transparency issues across among the 24 most important powerful internet, mobile, and telecommunications companies. See Ranking Digital Rights (2019) 'Corporate Accountability Index', *Rankingdigitalrights.org* [online], available at https://rankingdigitalrights.org/index2019/ [Accessed 11 May 2020].

[35] Maggiolino (2019) *supra note* 29, p. 1.

IP entitlements over the core components of AI technology, which impedes provider efforts to effectively protect their R&D and makes them reluctant to open up to public scrutiny.[36]

Secondly, there is a fear that disclosure would compromise company operations, allowing other parties to manipulate or otherwise exploit vulnerabilities in their systems. To provide an anecdote, a basic understanding of the functioning of Google Maps allowed a German artist to pull a prank on Silicon Valley's tech giant. He simply had to carry around Berlin a cart filled with 99 smartphones connected to the map service, tricking the navigation system to believe that a traffic jam was occurring in the area.[37] Companies could also point at the multi-billion-dollar Search Engine Optimisation (SEO) industry[38] whose main focus is deciphering how search engines work to produce better rankings for their clients. From this perspective, Google's outcry for secrecy is understandable. Disclosure could result in numerous attempts to manipulate search results and to the appearance of numerous copycat search websites that would start competing with the company.[39]

More recently, big corporations have made some timid initiatives towards transparency. For instance, after months of resisting government oversight,[40] in April 2019, Mark Zuckerberg finally called for government regulation.[41] However, this move was quickly

[36] As for the limit of copyright to protect AI technology, see Otero-Gonzales, B. (2021) 'Machine Learning Models Under the Copyright Microscope: Is EU Copyright Fit for Purpose?', *GRUR International* 70(11), 1043-55; Craglia (2018) *supra note* 1, p. 64; Gervais, D. (2019) 'Exploring the Interfaces between Big Data and Intellectual Property Law', *JPITEC* 10(3), p. 6-9; Regarding the sui generis database protection: Hugenholtz, P. (2018) 'Data Property: Unwelcome Guest in the House of IP', *Kritika - Essays on Intellectual Property* 3, p. 27; 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; Malgieri, G. (2016) 'Quasi-Property on Customer Information: Trade Secrets and Consumer Rights in the Age of Big Personal Data', *Journal of Internet Law* 6(2), p. 102. On Patents: European Patent Office (2018) 'Guidelines for Examination', sec. 3.3.1.

[37] Patel, B. (2020) 'Artists Creates a Traffic Jam on Google Maps by Dragging a Cart Full of 99 Smarthphones', *Dailymail* [online], available at https://www.dailymail.co.uk/news/article-7962413/Artist-creates-traffic-jam-Google-Maps-dragging-cart-99-smartphones.html [Accessed 11 March 2020].

[38] Wood, L. (2021) 'Worldwide Search Engine Optimization Services Industry to 2030 - Featuring Google, Bing and Baidu among Others', *Globenewswire* [online], available at https://www.globenewswire.com/news-release/2021/04/02/2203806/28124/en/Worldwide-Search-Engine-Optimization-Services-Industry-to-2030-Featuring-Google-Bing-and-Baidu-Among-Others.html [Accessed 7 September 2021].

[39] See Tower, N. (2014) 'Why Google Shouldn't Reveal Its Search Algorithm', *Perrill blog* [online], available at https://www.perrill.com/google-shouldnt-reveal-search-algorithm/ [Accessed 3 May 2020].

[40] Volz, D. & Ingram, D. (2018) 'Zuckerberg Resists Effort by U.S. Senators to Commit Him to Regulation', *Reuters* [online], available at https://www.reuters.com/article/us-facebook-privacy-zuckerberg/zuckerberg-resists-effort-by-u-s-senators-to-commit-him-to-regulation-idUSKBN1HH1CU [Accessed 20 April 2019].

[41] Albeit in the area of social media content but as AI systems are involved, there are related ethical issues. See Jackie Wattles, J. & O'Sullivan, D. (2019) 'Facebook's Mark Zuckerberg Calls for more Regulation of The Internet' [online], available at

questioned by critics, who hinted that Big Tech were merely attempting to shape future regulations in their favour or even outsourcing their oversight responsibilities.[42] Others pointed out that such regulation may benefit current players such as Google and Facebook (now Meta[43]) by making it harder for other businesses to compete with them.[44]

4. Trade secrets as a limit to AI transparency

From a legal standpoint, corporations are correct in pointing out that their AI components qualify as trade secrets and deserve some form of protection against disclosure. Trade secret protection (also known as "protection of undisclosed information") is a tort-like action that sanctions unfair commercial conducts such as espionage, theft, and breach of contract. However, it does not protect against competitors that have acquired information through honest practices such as reverse engineering or independent discovery.[45] In this sense, trade secret exclusively concerns the factual control over the tangible means where information is stored, but not the information per se, and has a more limited scope of protection than other IPRs.[46] Two main reasons explain the advantages of trade secret over other IPRs. These relate to the ease to meet the legal requirements for protection and its cost-effectiveness.

4.1 Trade Secrets and its applicability to Artificial Intelligence

The Treaty on the Trade-related Aspects of Intellectual Property Protection (TRIPs) was the first international instrument to explicitly regulate and frame trade secrets under the umbrella of IPRs.[47] The provisions enshrined in Article 39 later served as a role model for the 2016 EU Trade Secret Directive (TSD). The Directive reiterates that, to qualify for trade secret protection, commercial information must satisfy three requirements[48]:

- *It must be secret*. More precisely, the information must not "as a body or in the precise configuration and assembly of its components, be generally known among or readily accessible to persons within the circles that normally deal with the kind

https://edition.cnn.com/2019/03/30/tech/facebook-mark-zuckerberg-regulation/index.html [Accessed 11 April 2019].

[42] Harper, N. (2019) 'Zuckerberg Call for Tech Rules Gets Cold Reception' *The Hill* [online], available at https://thehill.com/policy/technology/437055-zuckerberg-call-for-tech-rules-gets-cold-reception [Accessed 12 March 2020].

[43] Lyons, K. (2021), 'Facebook just Revealed its New Name: Meta'*, The Verge* [online], available at https://www.theverge.com/2021/10/28/22745234/facebook-new-name-meta-metaverse-zuckerberg-rebrand [Accessed 04 November 2021].

[44] Hawkins, J. (2018) 'The Conservative Case for Breaking Up Monopolies Such as Google and Facebook', *National Review* [online], available at https://www.nationalreview.com/2018/05/breaking-up-tech-giants-conservative-case/ [Accessed 12 March 2020].

[45] Article 3, Directive (EU) 2016/943.

[46] Mezzanotte (2017) *supra note* 36*,* p. 169-70; Gervais (2019) *supra note* 36, p. 5; Malgieri, G. (2016) 'Ownership of Customer (Big) Data in the European Union: Quasi-Property As Comparative Solution?', *Journal of Internet Law* 20(3).

[47] See Article 1(2) and Article 39(2) TRIPs Agreement (1994).

[48] Recital 3 and Article 2, Directive 2016/943.

of information in question".[49] As such, 'secrecy' is an autonomous legal requirement, distinct from similar concepts in other branches of the law. In this sense, the concept does not coincide with the novelty requirement in the patent system.[50]

- *The information must have commercial value because of its secrecy*. In contrast to other IPRs, trade secrets protect information of any kind, irrespective of its artistic or technical value. In other words, trade secret protection does not depend upon the inherent characteristics of the information (e.g., its 'originality' or 'non-obviousness) but mostly on external factors and thus can protect even mere ideas and purely abstract concepts.[51] However, it is key that the information offers some sort of commercial advantage when kept secret. In some cases, secrecy concerns some product features, preventing competitors from imitating successful products, as in well-known examples concerning drink recipes and seasoning powders. In other instances, the secret information concerns know-how*,* business methods or other commercial information, i.e., internal protocols and procedures that give an edge by maximising efficiency in the manufacturing, logistics, governance and distribution channels.[52] Delving further, some have identified four categories of information that normally constitute trade secrets: a) highly specific products that businesses decide to exploit secretly rather than seeking patent protection, b) know-how, i.e., information that allows harnessing the full potential of technological innovations, c) strategic business information that essentially improves decision-making, and d) collections of publicly available information, which, especially in the modern computational era confers an edge in developing and implementing IT technology.[53]

- *It must be subjected to reasonable measures to keep it secret:* These measures must both protect the information from external threats as well as impede internal leakages. As such, they include both binding contracts – e.g., non-disclosure agreements and restrictive covenants for employees - as well as technological and physical precautions against industrial theft and espionage. As for the reasonability requirement, it is meant to relieve companies from implementing measures that might prove excessively burdensome in proportion to their financial means. In this sense, the provision gives entrepreneurs sufficient room to implement cost-effective measures and to strike an adequate balance between security and financial expenditure.[54]

[49] Article 39(2) TRIPs Agreement (1994).

[50] Carvalho, N. (2008) 'The TRIPs Regime of Antitrust and Undisclosed Information', Kluwer Law International: The Hague, pp. 224-36.

[51] See Article 4 Directive (EU) 2016/943.

[52] Carvalho (2008) *supra note* 50.

[53] See European Commission (2013) 'Staff Working Document, Trade Secret Directive Impact Assessment', SWD/2013/0471 final, pp. 109-110.

[54] Grimes, S. & Murphy, S. (2019) 'EU Trade Secret Directive: What Are Reasonable Steps?', *Winston.com* [online], available at https://www.winston.com/en/thought-leadership/eu-trade-secrets-directive-what-are-reasonable-steps.html [Accessed 2 April 2020]; Hoeren, T. (2018) 'The EU

The TSD establishes a general right to redress against the unlawful acquisition, use and disclosure of confidential information and requires member states to implement measures to allow rightsholders to enforce this right in a fair, timely and effective manner.[55] Monetary damages are only one of the foreseen redress measures.[56] Secret holders can invoke injunctions against the utilisation of unlawfully acquired information, the destruction or delivery of any document or item in which the information is stored and of any infringing material created through the utilisation of the secret information.[57] The TSD also goes well-beyond the discipline of the TRIPs when clarifying that liability extends to any person that "knew or ought, under the circumstances, to have known that the trade secret had been obtained directly or indirectly from another person who was using or disclosing the trade secret unlawfully".[58]

There is little doubt that both constituents of AI technology, i.e., algorithms and databases, can fall within the scope of trade secret protection.[59] Moreover, regardless of normative considerations, providers tend to keep AI innovations for themselves and refuse to share them with competitors and the public.[60] Secrecy allows data holders to engage in monopolistic behaviours as *de facto* proprietors of AI assets, by enforcing those defensive measures that safeguard possessions against third parties intrusions. Under this perspective, the normative framework against trespassing and cybercrime complements trade secret protection.[61] Moreover, while secrecy does not protect from endeavours such as parallel invention 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 it extremely challenging to reverse engineer such systems.[62]

4.2 The advantages of secrecy

Furthermore, two other reasons explain the success of trade secret protection as a cost-effective means to protect AI systems.

- *Compatibility with other forms of protection:* Trade secret protection is highly compatible with other IPRs, meaning that often companies do not have to renounce to secrecy to enjoy other forms of protection. For instance, providers might both keep their algorithms secret and enforce copyright over them. As for patents, while the publication of the invention eventually entails the loss of secrecy, hybrid patent-secrecy strategies are a common solution. For instance, patents can protect the visible features of a technology while the hidden ones are kept secret, such as in the case of the know-how necessary for the correct maintenance, functioning and optimal performance of a machine. Trade secrets can also

Directive on the Protection of Trade Secret and its Relation to Current Provisions in Germany' 9, *JIPITEC*, p. 140.

[55] Article 6, Directive (EU) 2016/943.

[56] Article 14, Directive (EU) 2016/943

[57] Article 12, Directive (EU) 2016/943.

[58] Article 4, Directive (EU) 2016/943.

[59] Maggiolino (2019) *supra note* 29*,* pp. 4-7; Malgieri (2016) *supra note* 36, p. 102.

[60] Mezzanotte (2017) *supra note* 36*,* p. 159.

[61] Mezzanotte (2017) *ibid,* p. 167.

[62] Maggiolino (2019) *supra note* 29, p. 2.

complement non-IP assets. For instance, in the AI sector secrecy normally works in tandem with the *de facto* ownership of data, i.e., the factual control over servers and mainframe computers where data are stored and processed.[63]

- *Formalities, requirements, and length:* Unlike some other IPRs, trade secrets do not require registration or disclosure. This saves companies the risks and the financial commitments associated with the prosecution of other forms of protection. For instance, patent applicants might expect to spend at least 15.000 euros for comprehensive patent protection in Europe between patent and attorney fees.[64] This adds up to an examination procedure that normally takes between three to five years,[65] with a relatively high rejection rate.[66] This is particularly relevant in the AI field considering the challenges of patent protection for this new technology.[67] Also, trade secrets last indefinitely, as long as the requirements for protection are met. Thus, trade secret protection potentially lasts longer than any other IPR, outlasting not only the 20-year patent term but also copyright protection.[68]

4.3 Limitations to trade secret protection in Europe

As with other rights, trade secret protection finds some boundaries in the necessity to safeguard countervailing rights and in particular fundamental rights. In this sense, one of the most innovative aspects of the TSD lies in its attempt to settle potential conflicts between secrecy and disclosure in the public interest. The Directive explicitly stipulates that its application does not affect the right to freedom of expression,[69] and in particular, any violation of secrecy made with the intent of revealing misconducts, wrongdoings, or illegal activities (the so-called whistle-blower exception).[70] Likewise, it guarantees the right of workers to join labour unions and to relay relevant information to them in order to exercise a right foreseen by the law.[71]

Most importantly for our discussion, Art. 1(2) clarifies that the Directive does not affect: "..b) the application of EU or domestic rules requiring trade secret holders to disclose, for reasons of public interest, information to the public or public authorities for the performance of their duties; c) the application of rules requiring or allowing public authorities to disclose information submitted by businesses which those authorities hold

[63] Craglia (2018) *supra note* 1, p. 65.

[64] Patent Pilot (2020) 'The Cost of Obtaining a Patent' [online], available at https://www.patent-pilot.com/en/obtaining-a-patent/costs-of-obtaining-a-patent/ [Accessed 2 April 2020].

[65] European Patent Office (2020) 'FAQ - Procedure & law' [online], available at https://www.epo.org/service-support/faq/procedure-law.html [Accessed 2 April 2020].

[66] European Patent Office (2019) 'Patent Index 2019' [online], available at http://documents.epo.org/projects/babylon/eponet.nsf/0/BC45C92E5C077B10C1258527004E95C0/$File/Patent_Index_2019_statistics_at_a_glance_en.pdf [Accessed 2 April 2020].

[67] See Prange, D. & Lawson, A. (2018) 'Re-evaluating Companies' Al Protection Strategies', *Managing IP* 2, 35-38; Bader, M. & Stummeyer, C. (2019) 'The Role of Innovation and IP in AI-Based Business Models, in Baierl, R. et al. (eds.), 'Digital Entrepreneurship', Springer: Switzerland, pp. 36-40.

[68] See Council Directive 93/98/EEC.

[69] See Articles 1(2) and 5(2), Directive (EU) 2016/943.

[70] See Article 5(2), Directive (EU) 2016/943.

[71] See Articles 1(2) and 5(2), Directive (EU) 2016/943.

in compliance with the obligations and prerogatives set out in the law".[72] Put simply, these provisions allow public authorities to request and further disclose confidential information. Therefore, they are extremely relevant for the oversight of AI, and in particular, in the assessment of the conformity between new technology and the applicable legal requirements. Indeed, this involves the transmission of information on AI to public agencies and its possible communication to the third parties involved in the assessment. Other provisions in the TSD are meant to safeguard secrecy in the course of legal proceedings, confirming that judicial authorities can order the disclosure of trade secrets.[73] This allows courts to scrutinise AI components and assess their faulty or unlawful behaviour, as already done by the EU Commission in leading competition cases.[74]

Whereas these provisions serve the public interest well, they are not without flaws. The main perplexity, which will be dealt in greater detail further on, relates to the Directive's unclear position in respect to proportionality. This is one of the cornerstones of the international human rights system, EU law and the constitutional tradition of European countries.[75] While minor differences exist from a jurisdiction to another, in its essence proportionality requires state measures interfering with the rights of citizens to be 'proportionate', in the sense of not constraining them beyond what necessary to achieve the public goal set by the law.[76] In these respects, the Directive does not explicitly oblige courts or public authorities to take into consideration the countervailing interests of trade secret holders, which seems confirmed by the black letter of the law stating that the Directive "shall not affect" the ability of public authorities to request and disseminate trade secret information.[77] Only recital 34 stipulates the right to access files must be exercised while respecting secrecy,[78] a clarification that it would have been wiser to insert in the main text of the Directive.[79]

5. The EU Strategy: From the White Paper to the AIA

The monitoring and certification of AI technology will be one of the cornerstones of the upcoming EU legal framework on artificial intelligence. In 2017, the European Parliament

[72] Article 5, Directive (EU) 2016/943.

[73] Article 9, Directive (EU) 2016/943.

[74] Maggiolino (2019) *supra note* 29, p. 13.

[75] Article 52(3) Charter of Fundamental Rights of the European Union (2012/C 326/02): "In so far as this Charter contains rights which correspond to rights guaranteed by the Convention for the Protection of Human Rights and Fundamental Freedoms, the meaning and scope of those rights shall be the same as those laid down by the said Convention. This provision shall not prevent Union law providing more extensive protection".

[76] Christoffersen, J. (2015) 'Human rights and Balancing: The Principle of Proportionality', in Geiger, C. (ed.) 'Research handbook on human rights and intellectual property', Edward Elgar Publishing: Cheltenham, pp 19–37.

[77] See Articles 1(2)(b) and 1(2)(c) of the Directive (EU) 2016/943. Article 5(1), employs the even more categoric wording the application of the measures provided by the Directive must be "dismissed" when the exceptions foreseen by the provision apply.

[78] Recital 34, Directive (EU) 2016/943.

[79] For instance, national authorities might oversee Recital 34 during the implementation of the Directive.

had already proposed the establishment of an agency with certification competencies,[80] while certification systems have already been launched in member states such as Denmark and Malta.[81] More recently, the EC first described a certification system in its 2020 Whitepaper on AI and one year later in the AIA draft.

5.1 The centralised system in the 2020 Whitepaper

The 2020 Whitepaper envisaged the creation of a double-track system for the certification and oversight of AI technology. 'High-risk' systems were to be subjected to prior conformity assessment before being publicly used or distributed commercially. Conversely, 'low-risk' technology did not have to obtain prior approval, but was subjected to ex-post monitoring and reviewed in cases of reported malfunctioning or misbehaviour. The assessment of the risk of a given technology mainly depended on its purpose and industrial sector.[82] This scheme was a simplified version of a proposal originally advanced by the German Data Ethics Commission. The proposal envisaged five different levels of regulation, starting from no regulation for harmless systems to a complete ban on research, production, and distribution for the most dangerous applications.[83]

The conformity assessment had to be carried out by a centralised European agency, to which providers had to submit all the documentation necessary for the assessment. For harmless technology, the required information was limited to the name of the software, its proprietor and field of application. For high-risk AI, traders were required to submit the software source, the object code and the AI training datasets, as well as any other relevant information. This was to allow potentially problematic behaviours to be traced back and verified.[84] Providers also had to include information on the variable used with their values and deviations and the amount and type of training data used.[85] Considering the dynamic nature of AI, the agency was also meant to establish a post-release monitoring system, to ensure the proper functioning of AI technology overtime.[86]

In this context, a centralised agency was seen as a necessary instrument to avoid the fragmentation of competencies within the EU, with the view of creating a unified framework for the testing and certification of AI-enabled products and services.[87] The main idea was for the agency to carry out the conformity assessments between AI technology and the mandatory requirements set by the law or at least coordinate the work of domestic agencies entrusted with this task.[88] Besides regulatory oversight, the agency could perform a variety of other tasks, including being a forum for the exchange of information and best

[80] See European Parliament (2017) 'Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics' (2015/2103(INL)); Craglia (2018) *supra note* 1, p. 69.

[81] EC (2020) *supra note* 2*,* p. 10.

[82] EC (2020), pp. 17-23.

[83] EC (2020), pp. 10.

[84] EC (2020), p. 19-20.

[85] Craglia (2018) *supra note* 1, p. 59.

[86] EC (2020) *supra note* 2, pp. 23-24.

[87] EC (2020), p. 24.

[88] EC (2020), p. 25.

practice, the identification of emerging trends and providing advice on standardisation activity, including certification at the local level.[89]

5.2 The decentralised conformity assessment in the AIA

In April 2021, the EC published the draft AIA regulation, with the aim of fostering the development of a single market for AI applications and ensuring that AI technology is safe, trustworthy and respects existing laws and fundamental rights.[90] After spelling out a number of prohibited AI practices,[91] the central section of the proposal deals with the authorisation and monitoring of AI technology through the so-called conformity assessment mechanism of potentially dangerous AI technology.[92]

Expanding on the classification of the 2020 Whitepaper,[93] the proposal reserves the conformity assessment to “high-risk” AI systems. The AIA does not define this concept, but it identifies the systems falling within this category in its annexes.[94] These include, inter alia, systems for the management and operation of critical infrastructure (e.g. railways and lifts) or AI to be used for the recruitment of personnel or to evaluate the creditworthiness of natural persons.[95] High-risk systems must comply with a set of complex and detailed legal requirements.[96] Some of them are ancillary duties relating to the establishment of appropriate risk management systems,[97] the drafting of technical documentation,[98] and instructions for users,[99] as well as obligations relating to human oversight.[100] Other requirements relate to the design of the technology itself, such as the proper criteria for the training, validation and testing of data,[101] standards relating to accuracy, robustness, cybersecurity and consistency throughout a product lifecycle,[102] and the ability of AI systems to automatically create logs of its activity.[103]

Departing from the Whitepaper, the proposal establishes a *decentralised* system where conformity assessments are carried out at the domestic level.[104] ‘Conformity assessment

[89] EC (2020), p. 25.

[90] EC (2021) AIA *supra note* 2, p. 3.

[91] EC (2021) AIA, Title II; on the topic see Veale, M. & Zuiderveen Borgesius, F. (2021) ‘Demystifying the Draft EU Artificial Intelligence Act’, *Computer Law Review International* 4, p. 98-102.

[92] EC (2021) AIA, Title III.

[93] EC (2021) AIA, Titles II, III, IV, IX. See also Smuha, N. Ahmed-Renger, E.et al. (2021) ‘How The EU Can Achieve Legally Trustworthy AI: A Response to the European Commission’s Proposal for an Artificial Intelligence Act’, p. 2 [online], available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3899991 [Accessed 19 October 2021].

[94] EC (2021) AIA, Art. 6; Smuha (2021) *ibid*, p. 29.

[95] EC (2021) AIA, Annexes II and III.

[96] EC (2021) AIA, Art. 8 and 16.

[97] EC (2021) AIA, Art. 9.

[98] EC (2021) AIA, Art. 11.

[99] EC (2021) AIA, Art. 13.

[100] EC (2021) AIA, Art. 14.

[101] EC (2021) AIA, Art. 10.

[102] EC (2021) AIA, Art. 15.

[103] EC (2021) AIA, Art. 12.

[104] EC (2021) AIA, Art. 43.

bodies' (also known as 'notified bodies') will carry out the actual tests to verify the conformity of high-risk systems with the prescribed requirements.[105] These bodies are independent technical organisations, typically private sector certification firms.[106] They will have to be accredited by 'national notifying authorities[107] that, as suggested by the name itself, must notify the EC of their decisions.[108]

Despite the emphasis on third-party review, self-assessment plays an even bigger role in the AIA. In fact, Art. 40 stipulates that conformity is presumed when the technology complies with the harmonised standards published in the Official Journal of the European Union.[109] Under the supervision of the EC, these standards are normally developed by European Standardisation Organisations (ESOs) such as the CEN (European Committee for Standardisation) and CENELEC (European Committee for Electrotechnical Standardisation).[110] Considering that the AIA contains an explicit requirement to consult harmonised standards during the implementation of risk management procedures,[111] and that the EC expects the first standards to be published by 2025,[112] some have pointed out that standard compliance will become the preferred route for AI certification.[113] This choice is not without criticism, with experts pointing out to the unreliable, cloudy and discretionary nature of self-assessment and strongly advocating for the strengthening of ex-ante audit obligations in the AIA draft.[114]

Against this background, the preference towards self-assessment explains the choice to abandon the 2020 Whitepaper's centralised certification system. The AIA avoids disrupting the frameworks applicable to high-risk systems that are safety components of complex products and are already subject to third party conformity by the sectorial regulations.[115] Moreover, charging a single agency with the review of all AI technology regardless of its technical sector might have proven challenging.[116] This is because assessing and certifying AI technology requires broad and diversified sectorial competencies (e.g. finance or healthcare), in addition to transversal expertise in areas such as privacy, consumer protection and information technology. The multidisciplinary nature of AI translates into a lack of a shared language and common methods among experts and policymakers making discourse, synthesis, and coordination a challenge.[117] These points were emphasised by UK

[105] EC (2021) AIA, Art. 33.

[106] Veale (2021) *supra note* 92*,* p. 106.

[107] EC (2021) AIA *supra note* 2, Art. 31.

[108] EC (2021) AIA, Art. 32.

[109] EC (2021) AIA, Art. 40.

[110] Veale (2021) *supra note* 92*,* p. 105.

[111] EC (2021) AIA *supra note* 2, Art. 9(3).

[112] EC (2021) AIA Impact Assessment *supra note* 9*,* p. 58.

[113] Veale (2021) *supra note* 92*,* p. 106.

[114] See Smuha (2021) *supra note* 94, p. 39.

[115] EC (2021) AIA Impact Assessment *supra note* 9, p. 58.

[116] EC (2021) AIA Impact Assessment*,* p. 23.

[117] Frissen, V. Lakemeyer, G. et al. (2018) 'Ethics and Artificial Intelligence', *Bruegel* [online], available at http://bruegel.org/2018/12/ethics-and-artificial-intelligence/#_ftnref1 [Accessed 12 March 2020]; See Mercurio, B. & Yu, R. (2021), 'An AI policy for the (near) future' in Borchert, I. & Winters, L. (eds), 'Addressing Impediments to Digital Trade', CEPR Press: London, pp. 73-104.

Under Secretary of State at the Department for Business, Energy and Industrial Strategy, Amanda Solloway, during the November 2020 World Intellectual Property Organisation (WIPO) Conversation on AI.[118]

5.3 Transparency and confidentiality in the AIA

The AIA endorses strict confidentiality standards. Conformity assessment bodies, as well as their subcontractors and associates, must put in place documented procedures to ensure the strictest confidentiality of the files submitted during the assessment procedure.[119] The obligation of confidentiality extends to any other domestic authority involved in the application of the AIA.[120] The violation of confidentiality can be sanctioned with monetary penalties.[121]

As a counterweight, the AIA spells out a broadly defined exception to confidentiality for disclosures required by the law.[122] Furthermore, conformity assessment bodies are not bound to secrecy with the respect to the communications to notifying authorities to the Member state in which the conformity assessment is carried out.[123] Market surveillance authorities[124] can be granted access to the training, validation and testing datasets used by the provider,[125] and even to the source code of the AI system under certain conditions.[126] National public authorities which supervise or enforce the respect of obligations protecting fundamental rights in relation to the use of high-risk AI systems in predetermined sectors can also obtain access to conformity documentation, when access is necessary for the fulfilment of their institutional tasks and within the limits of their jurisdiction.[127] Finally, as a general rule, public authorities can exchange information in their possession as long as confidentiality is maintained.[128]

These exceptions are overall consistent with the ones foreseen in the TSD, which the AIA explicitly recalls.[129] However, the coordination between the AIA and Art. 5 TSD is probably insufficient to adequately serve the public interest in the early disclosure of AI threats. This is particularly evident in relation to the whistle-blower exception: on one side the letter of the law does not seem to grant to third parties a right to access to the documentation in order to unveil suspected AI flaws, while on the other the provision presupposes a

[118] WIPO (2021) 'Conversation on Intellectual Property and Artificial Intelligence', Third Session, WIPO/IP/AI/3/GE/20/INF/5, 8 January [online]. Available at: https://www.wipo.int/meetings/en/details.jsp?meeting_id=59168 [Accessed 16 December 2021].

[119] EC (2021) AIA *supra note* 2, Art. 33(6).

[120] EC (2021) AIA.

[121] EC (2021) AIA, Art. 70(2).

[122] EC (2021) AIA, Art. 33(6).

[123] EC (2021) AIA.

[124] These are defined as "the national authority carrying out the activities and taking the measures pursuant to Regulation (EU) 2019/1020". EC (2021) AIA, Article 3(26).

[125] EC (2021) AIA, Art. 64(1).

[126] EC (2021) AIA, Art. 64(2).

[127] EC (2021) AIA, 64(3).

[128] EC (2021) AIA, Art. 70(2).

[129] EC (2021) AIA*,* Art. 70(1)(a).

"misconduct, wrongdoing or illegal activity" and therefore might not be triggered by genuine mistakes in the design or implementation of AI systems.

Furthermore, the choice to limit the ability to access conformity assessment documentation to public authorities is lamentable for overlooking the role that the civil society and researchers play in unveiling the flaws of AI technology and the associated risks for the collectivity.[130] For instance, independent studies uncovered the bias of self-driving cars in detecting pedestrians with darker skin,[131] revealed discriminations in health care management programs,[132] and even discovered that an algorithm used to predict the likelihood of recidivism for on-parole probation discriminated against black convicts.[133] In most of the cases, independent review cannot take place without access to the training data, since this might be the most fast and effective, if not the only, way to spot bias in the utilised datasets.[134] In this sense, it is not a coincidence that medical researchers have been particularly effective in identifying AI biases, considering the wider availability of anonymised clinical documentation.[135]

Even the establishment of a publicly accessible database for standalone high-risk systems,[136] despite its intended purpose, does not really play out in favour of independent researchers.[137] This is because the information to be registered is limited to the data needed to identify AI providers and their technology such as names and addresses.[138] Clearly enough, the register will be of no use in assessing the risk of the registered

[130] Mehrabi, N., Morstatter, F. et al. (2021) 'A Survey on Bias and Fairness in Machine Learning', *ACM Computing Surveys* 54(6), 1-35.

[131] Wilson, B., Hoffman, J. et al (2019) 'Predictive Inequity in Object Detection', *Arxiv* [online], available at https://arxiv.org/pdf/1902.11097.pdf [Accessed 10 April 2020].

[132] Obermeyer, Z. & Mullainthan, S. (2019) 'Dissecting Racial Bias in an Algorithm that Guides Health Decisions for 70M People', Proceedings of the Conference on Fairness, Accountability, and Transparency [online], available at https://dl.acm.org/doi/10.1145/3287560.3287593 [Accessed 10 April 2020].

[133] Larson, J., Mattu, S. et al. (2016) 'How We Analyzed the COMPAS Recidivism Algorithm', *Pro Publica* [online], available at https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm [Accessed 16 December 2021].

[134] On a classification of data biases, Mehrabi (2021) *supra note* 131*,* pp. 4-7.

[135] For some example of biased AI datasets, Manrai, A., Funke, B. et al (2016) 'Genetic Misdiagnoses and the Potential for Health Disparities, *New England Journal of Medicine* 375(7), 655–665; Larrazabala, A., Nieto, N. et al. (2020) 'Gender Imbalance in Medical Imaging Datasets Produces Biased Classifiers for Computer-aided Diagnosis, *PNAS* 117(23), 12593-12594.

[136] EC (2021) AIA *supra note* 2, Art. 60.

[137] The European Commission discarded the registration of high-risk AI systems that are safety components of products or devices on the observation that these might already be subject to registration according to the existing product safety legislation and duplication of databases should be avoided. Furthermore, in the scenario where sectoral safety legislation does not establish a registration obligation for the products, the registration was considered of limited value for the public and the market surveillance authorities considering that the product as a whole would not be subject to central registration obligations. See EC (2021) AIA Impact Assessment, *supra note* 9, p. 58; See also Veale (2021) *supra note* 92, p. 112.

[138] EC (2021) AIA *supra note* 2, Art. 60 and Annex VIII.

AI-systems, leading some scholars to advocate for an extension of the information to be recorded therein.[139]

Overall, these choices are particularly unfortunate, because the AIA takes a step back from the Whitepaper which suggested the establishment of a more liberal access system meant to ensure that the (then) centralised agency could perform its assessment duties while allowing for a high degree of transparency. To this end, the Whitepaper delineated an on-demand scheme under which citizens could access AI documentation as long as confidentiality was preserved.[140] Interestingly, the Whitepaper hinted that the expertise and procedures developed in other technical fields could serve as a model for AI.[141] Following this lead, the next section will discuss how the rules adopted in the pharmaceutical sector, and in particular, the ones foreseen in the latest European Medicine Agency (EMA) policy on the disclosure of clinical dossiers,[142] are particularly fit to set a reasonable trade-off between transparency and secrecy.

Indeed, even if the AIA does not foresee a centralised marketing agency as the gatekeeper of AI documentation, access rules still have a role to play in at least two contexts. The first one is the establishment of a central database for standalone AI systems pursuant to Art. 60, especially under the auspices of the expansion of the documentation to be submitted. The second is the access granted by the organisations involved in the certification and monitoring of AI technology at the domestic level, such as conformity assessment bodies and market surveillance authorities.[143]

6. Borrowing from the pharmaceutical sector

The transparency rules in the AIA appear overly strict when compared to other fields. In particular, the pharmaceutical sector had its own struggles in balancing transparency and trade secret protection, which led to the fine-tuning of the rules and procedures on data access. Pharmaceutical agencies evaluate the safety, quality, and efficacy of a drug on the basis of a dossier submitted by a sponsor, summarising the drug testing on thousands of patients.[144] Given the high cost of clinical trials, which according to some appraisals can overcome the 1 billion dollars,[145] it is not surprising that pharmaceutical companies have a strong interest in the secrecy of their clinical dossiers. Their stance is that the dossier contains proprietary information, and that disclosure might harm their commercial interests and strategies. There are also legitimate concerns that disclosure may allow competitors to disguise the disclosed data as independently developed and submit them to regulatory agencies in foreign jurisdictions. This would allow competitors to enter

[139] See Smuha (2021) *supra note* 94, pp. 52-53.

[140] EC (2020) *supra note* 2, pp. 19-20.

[141] EC (2020), pp. 24-25.

[142] EMA (2019) 'European Medicines Agency policy on publication of clinical data for medicinal products for human use', EMA/144064/2019.

[143] EC (2021) AIA *supra note* 2, Art. 64.

[144] See Spina Alì, G. (2017) 'TRIPS and Disclosure of Clinical Information: An Intellectual Property Perspective on Data Sharing', *Journal of World Intellect Property*, 20(1-2), 24–56.

[145] Mestre-Ferrandiz, J., Sussex, et al. (2012). 'The R&D Cost of a New Medicine', *Office of Health Economics*, United Kingdom.

foreign markets before data developers, gaining important first-mover advantages. Among these, most jurisdictions grant to data developers a period of exclusivity over the utilisation of the data, which impedes competitors to rely on the first authorisation to market their bio-equivalent drugs.[146]

Unfortunately, full data secrecy comes with severe backlashes.[147] Some of them are a prerogative of the pharmaceutical sector and are mostly irrelevant for AI. This is the case of foreclosing a better understanding of drugs pharmacodynamics or potentially leading to unethical duplicative trials of drugs whose negative effects are already known. Nevertheless, transparency is a compelling argument both in the pharmaceutical and AI sector, since it allows public oversight over potentially harmful technology. Indeed, exactly as in the AI sector, on more than one occasion independent reviews have unveiled the unsafety or inefficacy of authorised drugs, leading to the removal of dangerous products from the market or their relabelling. Regardless of whether the error in the authorisation of a drug is the result of a genuine mistake from the regulatory agency[148] or the malicious conduct of the firm sponsoring the product,[149] public scrutiny ensures the democratic control over collective choices and, more pragmatically, allows potential hidden threats to be identified and removed.

This is the reason why a wisely crafted access scheme should be an important tool to unveil bad faith attempts to circumvent regulation, expose unforeseen risks or detect flaws in public surveillance and monitoring. These claims are even stronger considering the defining features of AI. The opacity and interdisciplinary nature of AI require the intervention of several experts to sufficiently grasp the logic of the technology under examination, while its ability to self-program and change over time calls for the continuous monitoring of such systems. The following sub-section will elucidate how the rules envisaged for the pharmaceutical sector can be easily transposed to AI.

6.1 Conditions to access the information submitted to the AI agency

In a nutshell, establishing an access scheme means answering the questions of who, when, for what purposes and under which conditions access to the documentation held by public authorities should be granted. These questions often translate into regulations that limit the amount of accessible information or procedures that adequately protect the commercial interest of providers. As the result of a multi-staged process involving civil litigation, public consultations and legislative reform, the EU legislator has refined these criteria in the pharmaceutical sector. The relevant regulations now provide a gold standard access scheme to pharmaceutical documentation, which, with the needed tweaks, can serve as a role model for other sectors. This is achieved by providing cumulative measures that complement each other to create an effective and balanced disclosure policy.

[146] IFPMA (2011) 'Data Exclusivity' [online], available at https://www.ifpma.org/wp-content/uploads/2016/01/IFPMA_2011_Data_Exclusivity__En_Web.pdf [Accessed 11 April 2020].

[147] See the list of organizations mentioned in Gøtzsche, P. (2011). 'Why We Need Easy Access to all Data from All Clinical Trials and How to Accomplish it', *Trials Journal*, 12(249), 1–14.

[148] Gøtzsche (2011) *ibid,* p. 3.

[149] Gøtzsche (2011) *ibid*, pp. 3-8.

6.1.1 Subjects entitled to access the information

The latest EMA policy on the publication of clinical data for medicinal products conforms to the overarching principle that EU citizens and residents are entitled to access the documents held by EU institutions.[150] The subjects interested in accessing clinical information can do this by undergoing a registration process. A simplified procedure consisting in the creation of a username and a password allows users to access the information only 'on-screen' and for general information purposes. Conversely, users willing to disclose identifying data (e.g., name, surname, a valid ID) can download and store the documentation on their terminals. This form of registration is mandatory for research organisations wanting to utilise the data for academic purposes.[151]

It is relatively straightforward to see how a similar double access scheme can be adopted in the AI sector, even though some important adjustment will be required. The first one is granting access to AI information in 'machine-readable form', rather than merely as documentary records.[152] From a legal standpoint, given the different nature of clinical data and AI components, it is important to ensure some degree of coordination with the principles set in other EU instruments. To provide an example, since both the AI algorithm and training dataset might contain copyrighted material, it might be necessary to comply with the limitations set by the EU copyright directives for text and data mining (TDM) and scientific research.[153] This might be achieved in two ways. The first one is to limit the possibility to download and store the information to 'not-for-profit' research organisations, so to harmonise the disclosure policy with the provisions of the DSM copyright directive.[154] An alternative solution is imposing contractual limitations over providers, i.e., by requiring them to forfeit their copyright claims for conducts concerning exclusively the review of the technology submitted to the agency. The second solution carries the advantage to allow for the disclosure of the relevant documentation to users that might want to engage in independent review but do not qualify as non-profit organisations. The backlashes are that it is overall less consistent with the current copyright framework and overly restricts the rights of AI providers.

6.1.2 Confidential information and partial disclosure

It is straightforward that an on-demand scheme should allow reviewers to access all the documentation necessary to carry out an accurate and meaningful review. By contrast, purely commercial information conferring an edge to competitors can be withheld by the agency. The partial disclosure of clinical documentation is a well-known measure in the pharmaceutical sector, taking the form either of the publication of excerpts and summaries or the redaction of clinical dossiers. The idea underlying the release of excerpts is to provide sufficient information for the purpose of independent review while withholding

[150] Article 2(1), Regulation 1049/2001; EMA (2019) *supra note* 143.

[151] EMA (2019) *supra note* 143.

[152] See for instance, Campos, M., Silva, E. (2021) 'Towards Machine-Readable (Meta) Data and the FAIR Value for Artificial Intelligence Exploration of COVID-19 and Cancer Research Data', *Front. Big Data* [online], available at doi.org/10.3389/fdata.2021.656553 [Accessed 18 October 2021].

[153] See Article 3, Directive (EU) 2019/790 and Article 5(3)(a) Directive 2001/29/EC.

[154] Articles 3 and 4, Directive 2019/790.

some of the core information necessary to obtain marketing authorisation in foreign jurisdictions or replicate the technology. For instance, this can be achieved by publishing the data on drug safety, whilst at the same time keeping the data on drug pharmacokinetics and pharmacodynamics confidential. In fact, safety trials alone are insufficient to support marketing authorisation without evidence of bioavailability, thus preventing competitors from obtaining licenses in most foreign jurisdictions.[155] The principles underlying partial disclosure can be easily adapted to AI. Reviewers need not to have access to complete datasets anytime running simulations on randomised samples is sufficient for review purposes. This avoids full disclosures and possible leakages resulting in competitive harm to providers.

The redaction of the clinical dossier works in a slightly different manner. The agency does not limit *ex officio* the information to be released, resting on the data owner to object to the publication of information that constitutes commercial confidential information (CCI).[156] The EMA policy arguably adopts a broader definition of CCI than the TSD, defining it as any information not publicly available whose disclosure may undermine the commercial interest of the drug applicant.[157] Among the factors to evaluate the confidential nature of the information are the nature of the product, the competitiveness of the relevant market, the approval status in other jurisdictions, the novelty of the drug, and the possibility to develop follow-on drugs.[158] However, clinical trials are not, *ipso facto*, considered confidential information.[159]

This is in sharp contrast to the AIA whose default rule prescribes strict confidentiality for all the documentation submitted for conformity assessment. This is all the more unfortunate when the submitted technology, or part thereof, does not qualify for trade secret protection. Common examples include AI trained on public databases, or providers who have no interest in secrecy and have already made their technology public (e.g. through a creative commons license). By contrast, it is key to tailor the definition of CCI to the needs of the AI sector. This will require a significant deal of public consultation, with traders playing an important role in the definition of the concept and the establishment of acceptable transparency practices.

6.1.3 Delaying publication

Some jurisdictions allow the disclosure of clinical dossiers only after a fixed period of time has elapsed from the date of authorisation. This is to allow developers to comply with different regulatory requirements and obtain authorisation in all the territories of commercial interest, as well as allowing sponsors to seek approval for second medical uses of the compound.[160] For instance, in the US, the Hatch-Waxman Act stipulates that trials

[155] Kesselheim, A. & Mello, M. (2007) 'Confidentiality Laws and Secrecy in Medical Research: Improving Public Access to Data on Drug Safety', *Health Affairs* 26(2), pp. 489–491.

[156] EMA (2019) *supra note* 143, pp. 6-7.

[157] EMA (2019), p. 3.

[158] EMA (2019), pp. 6-7.

[159] EMA (2019), p. 4.

[160] Institute of Medicine of the National Academies (2015) 'Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk', *The National Academies of Sciences, Engineering, Medicine* [online],

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.[161] Similarly, Regulation 536/2014 stipulates that, in general, data included in clinical reports should not be treated as confidential information after marketing authorisation has been granted, the granting procedure is completed or the application has been withdrawn.[162]

The opportunity to implement delayed disclosures in the AI sector mainly depends on the legal effect attached to the act of registration. If registration confers upon the registrant proof, or at least a legal presumption of ownership, it might be reasonable to give sufficient time to AI owners to register their products in all the jurisdictions of interest. Thus, the option to allow disclosure only after a reasonable period of time (e.g., one year) from the moment of submission or authorisation should not be discarded, also with a view of giving some time to work on derivative applications of the authorised technology.

6.1.4 Confidential agreements

Some agencies allow access to clinical documents to private parties under the express acceptance of confidentiality terms. Terms of use normally include, inter alia, clauses on the obligation of recipients not to utilise the data in support to marketing applications. For instance, the EMA 2019 policy allows disclosure only to those recipients who pledge not to seek marketing authorisation outside the EU and use it only for non-commercial purposes.[163] It is easy to see how these measures could be extended to AI. Agencies could also evaluate the guarantees offered by information recipients to take all necessary measures to avoid data leakage and prevent free-riding conducts. Moreover, information recipients might be asked to further reinforce confidential obligations through monetary deposits, performance bonds, fines or penalties. Access may be refused to applicants not offering adequate warranties.

6.1.5 Legal effects of registration

The AIA does not attach any proprietary effects to registration. In other words, the rights over the registered technology arise exclusively from the rules that govern intellectual property rights, especially copyright and patents. This choice is laudable since it avoids potential conflicts between registration and the IP regime. Furthermore, in a decentralised registration model, it also avoids potential conflicting proprietary claims arising from multiple registrations by different providers.

By contrast, even though the matter might vary among Member States, domestic courts might attach some evidentiary value to registration. In particular, registration could serve as a rebuttable presumption that: a) the registrant owns the copyright over the registered algorithm and b) that the technology originated from the registrant, i.e., the provider had

available at http://www.nap.edu/catalog/18998/sharing-clinical-trial-data-maximizing-benefits-minimizing-risk [Accessed 27 December 2021].

[161] See Hatch Waxman Act, 21 U.S.C. § 355(l)(l).

[162] Recital 68, 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.

[163] EMA (2019) *supra note* 143, pp. 10-11.

lawful access to the registered technology at the date of submission. This solution seems consistent with the very wording of the AIA, which on one side defines the technology provider as the "natural or legal person... that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark"[164] and on the other mandates the registration of the information necessary to identify the AI provider.[165] In case of conflict between the rule on attribution of IPRs and registration, the former would prevail over the latter, or, more precisely, the presumption conferred by registration would capitulate. The presumption could increase legal certainty over the entitlement of AI technology, especially for non-registered IPRs. Furthermore, it also offers non-monetary compensation to AI developers for the administrative burden associated with registration and provides some incentives to spontaneously comply with the system.

7. The proportionality of an on-demand scheme

An on-demand access scheme as the one in place in the pharmaceutical sector is fully compatible with the principles set by the TSD, insofar as Articles 1(2)(b) and 1(2)(c) already allow public bodies to acquire and disclose confidential information for reasons of public interest. A further advantage is that the scheme conforms to the principle of proportionality, overcoming one of the deficiencies of the TSD. Even with some variations from a jurisdiction to another, proportionality is a principle requiring limitations to individual rights to be prescribed by the law, pursue a public interest and be 'proportionate' in a strict sense, i.e., to not go beyond what necessary to achieve the public goal at stake.

There is little doubt that the relevant exceptions of the TSD meet the requirements of being prescribed by the law and pursuing a public interest. By contrast, it is the strict sense proportionality assessment that requires further reflection. This assessment normally encompasses two different considerations. First, the measure must be suitable to achieve the public interest pursued by the law. Secondly, the measure must not go beyond what is necessary to achieve said goal, considering both the availability of less-restrictive measures and the precautions to be taken to safeguard the countervailing interest. Furthermore, when the interest pursued by the law and the one limited by the state intervention both have constitutional ranking, proportionality becomes a balancing exercise meant to ensure that both rights reciprocally limit each other but none of them succumbs.[166]

The legal interests involved in setting up an access scheme are numerous. Depending on the technological field, disclosure might protect interests such as health, property, equality, privacy and non-discrimination. More transversally, the right of EU citizens to receive relevant information on the safety of AI products in commerce is always at stake. This is one of the many facets of the right to receive and impart information, i.e., free speech.[167]

[164] EC (2021) AIA Proposal *supra note* 2, Art. 3.

[165] EC (2021) AIA Proposal *ibid*, Art. 60 and Annex VIII.

[166] Christoffersen (2015) *supra note* 77, pp. 19–37.

[167] Hugelier, S. (2011) 'Freedom of Expression and Transparency: Two Sides of One Coin', *Jura Falconis* 47, 61-91.

All these interests are to be balanced against the interest of undertakings in not having their secrets disclosed to competitors. In spite of trade secret's characterisation as an unfair competition tort-based action,[168] both statute and prominent case law confirm its intellectual property connotation.[169] This elevates trade secrets to the standing of a fundamental right, pursuant to Article 17(2) of the European Charter.[170-171] Furthermore, trade secrets fall within the scope of 'possession' pursuant to Article 1, Protocol 1 of the European Convention of Human Rights (ECHR), since this includes not only intangible and intellectual assets but even comprises a vast array of ''concrete proprietary interests having economic value'', including licenses, leases, contractual rights and even business goodwill and clientele.[172] Even a legitimate expectation to acquire property falls within the scope of the provision when backed up by a proper legal basis under domestic law, such as a mere trademark application.[173] Therefore, it is beyond doubt that trade secret qualifies as a 'possession' deserving fundamental right protection, insofar as it constitutes one of the most precious assets of a business and has a clear and often measurable economic value.[174]

This in turn entails that measures impinging on providers confidential information must obey the principle of proportionality. In better terms, setting up a balanced disclosure policy requires mechanisms that safeguard both the commercial assets of traders and the public interest in a transparent and accountable technology. Whereas disclosure policies are undoubtedly appropriate and suitable to ensure higher AI accountability, proportionality also entails that disclosures must not go beyond what is necessary to achieve the goals underpinning AI transparency. This implies taking into consideration the availability of less restrictive measures than full disclosure, including those precautions meant to safeguard the interest of providers. Against this background, the safeguards foreseen in favour of trade secret holders in the previous section aim at creating a legal framework that, in ensuring a reasonable trade-off between the countervailing interests in question, is fully compliant with the proportionality principle.

Besides normative considerations, there is a more pragmatic side to proportionality. It concerns the idea that legislators should strive to create a regulatory environment that is both safe for consumers but also appealing to entrepreneurs. Exceedingly burdensome regulatory requirements, disregard for corporations' financial interests and intellectual assets might stifle innovation, reduce foreign investments and lead to a delay in the

[168] Hoeren (2018) *supra note* 54, p. 139.

[169] See Article 1(2) TRIPs Agreement (1994); Recital 1 and 2, Directive (EU) 2016/943; Judgement of the Court of First Instance of 17.9.2007, case T-201/04, Microsoft.

[170] Article 17(2), European Charter of Fundamental Rights.

[171] Smuha (2021) *supra note* 94, p. 48 arguing that trade secret might also fall under the scope of the right to conduct business pursuant Art 16 of the European Charter.

[172] Kopecky´ v. Slovakia, App. No. 44912/98, ECtHR, at 144 (2004); 'Iatridis v. Greece', App. No 31107/96, ECtHR (1999).

[173] Anheuser-Busch Inc. v. Portugal, App. No. 73049/01, ECtHR, at 13–24 (2007).

[174] EC (2013) Trade Secret Directive Impact Assessment, *supra note* 53, pp. 145-6.

adoption and marketing of new technologies. This is a mistake that should be avoided in a period when Europe is struggling to impose itself as a worldwide leader in the AI sector.[175]

8. Conclusions

The AIA has already been criticised for its potential added costs in compliance, some say as high as €31 billion over the next five years,[176] which could deter investments in European start-ups and steer a brain drain of entrepreneurs, scientists and developers to locations with fewer bureaucratic hurdles.[177] On top of these criticisms, the AIA deviates from the guidelines set in the 2020 Whitepaper, which sketched a balanced access scheme to be modelled after the procedures in place in other regulated sectors. In doing so, the above exposition showed how the AIA prescribes an overly strict transparency regime for the information submitted during the conformity assessment of AI technology. In particular, the choices to treat all the submitted information as confidential information and to allow disclosure only between public authorities are particularly lamentable. Not only do they exceed the TSD requirements, but it also neglects the important role that researchers and citizens can play in unveiling AI threats.

But not all hope is lost. Not only the AIA is only at the proposal stage, but more liberal access rules can be implemented through soft-law instruments meant to clarify the conditions for accessing AI documentation.[178] As for the specific rules to be implemented, the paper sketched an on-demand access scheme that, in full respect of the proportionality principle, tries to reap the major benefits of disclosure without compromising providers commercial interests. To achieve this, the paper suggested the fine-tuning of some of the mechanisms elaborated for clinical dossiers to AI documentation, such as the rules governing qualified access, confidential agreements, delayed disclosures and the publication of summaries and excerpts.

[175] EC (2020) *supra note* 2, p. 18-23.

[176] Mueller, B. (2021) 'How Much Will the Artificial Intelligence Act Cost Europe?', *Center for Data Innovation* [online], available at https://datainnovation.org/2021/07/how-much-will-the-artificial-intelligence-act-cost-europe/ [Accessed 23 December 2021].

[177] Mueller, B. (2021) 'The Artificial Intelligence Act is a Threat to Europe's Digital Economy and Will Hamstring the EU's Technology Sector in the Global Marketplace', *Center for Data Innovation* [online], available at https://datainnovation.org/2021/04/the-artificial-intelligence-act-is-a-threat-to-europes-digital-economy-and-will-hamstring-the-eus-technology-sector-in-the-global-marketplace/ [Accessed 16 September 2021].

[178] See for example in the pharmaceutical sector EMA (2019) *supra note* 143.