Article II

Salami, Emmanuel (2022) Balancing Competing Interests in the Reidentification of AI-Generated Data

Reprinted from European Data Protection Law Review, Vol 8, Issue 3, (2022) 362 – 376 with the permission of Lexxion publishers.

Balancing Competing Interests in the Reidentification of AI-Generated Data

*Emmanuel Salami**

AI systems generate valuable analytical information from (apparently) non-personal data with vast economic consequences. This information generated from non-personal data provides a competitive edge which serves as a key rationale for its appropriation to the exclusion of others. The proliferation of AI has made it possible for non-personal data (including anonymised data) to result in the reidentification of natural persons. There have been propositions from various quarters for the protection of non-personal data by Intellectual Property Rights (IPR). Should AI-generated data be protected by IPR, this can potentially result in data protection concerns in the event of data reidentification that singles out natural persons. This might particularly occur where reidentified data leads to the identification of natural persons in circumstances where the applicability of data protection law had neither been contemplated nor anticipated. This article highlights the concerns that might arise in the event of data reidentification and how this might raise interesting challenges for data protection compliance.

Keywords: AI | AI-Generated Data | Reidentification | Intellectual Property Rights

I. Introduction

Big data is very important to AI for reasons which include its role in training AI algorithms to perform specified tasks and its disposition as the form in which AI's output is generated. AI can be defined as systems that have been designed to carry out specific tasks in a manner that simulates human intelligence with varying levels of human intervention.[1] It has been predicted that 175 zettabytes (i.e. 175 trillion gigabytes) of newly created big data will be generated by 2025, an astronomical departure from the 35 zettabytes (i.e. 35 trillion gigabytes) predicted for 2020.[2] These figures indicate how indispensable big data is to AI and vice versa. In processing big data, AI systems create valuable information which provides further relevant insights. For instance, AI processes technical data from autonomous cars in a manner which provides insights about the efficiency of the car, air quality and traffic reports, etc.[3] Valuable information of this nature are referred to as AI-generated data in this article. This term (AI-generated data) will be used to depict those scenarios when AI systems have added value to data in such a manner that they generate some level of creativity worthy of consideration for (IP) protection. Throughout this article, AI-generated data refers to non-personal or anonymised data produced by AI systems.

The valuable nature of AI-generated data induces data producers to appropriate it to themselves to the exclusion of others. It is not unreasonable to expect that data producers who have invested their re-

DOI: 10.21552/edpl/2022/3/6

* The author is a PhD candidate and researcher at the University of Lapland, Rovaniemi, Finland. He is also an in-house data protection advisor for a global tech company. Any views expressed in this article are those of the author only. For correspondence: <Emmanuel.salami@outlook.com>

1 For further reading, see - Alan Turing, 'Computing Machinery and Intelligence' (1950) 433 - 460 <https://www.csee.umbc.edu/courses/471/papers/turing.pdf> accessed 12 March 2022. See also - Peter Norvig and Stuart Jonathan Russell, Artificial Intelligence: A Modern Approach (Pearson, 2010) 7.

2 Gil Press, '6 Predictions About Data In 2020 And The Coming Decade' (Forbes, January 6th 2020) <https://www.forbes.com/sites/gilpress/2020/01/06/6-predictions-about-data-in-2020-and-the-coming-decade/> accessed 12 July 2020.

3 Thomas Hoeren, 'Big Data and the Ownership in Data: Recent Developments in Europe' (2014) 36 EIPR 751.

sources towards AI development for the purpose of data generation will seek to assert some form of exclusionary right over it. However, such aspirations can potentially be antithetical to the freedom of expression and information and the free flow of data within the European Union (EU) digital single market.[4]

Data can generally be classified into personal and non-personal data. There is no doubt that IPR does not protect data, but as will be considered subsequently, various scholars have made propositions for some form of IPR protection for data.[5] The propositions for the ascription of IPR to data shows that the scope of such propositions are more commonly limited to non-personal data including anonymised data,[6] although similar arguments have also been made in respect of personal data.[7] For the sake of clarity, personal data is any information relating to an identified or identifiable natural person,[8] while non-personal data is any data other than personal data.[9] This means that any data which does not lead to any form of identification of natural persons would be considered as non-personal data. However, the distinction between personal and non-personal data has become very blurry because of various data reidentification techniques which have increased the reidentifiability of non-personal data. Ordinarily, non-personal data (including anonymised data) does not fall within the scope of data protection law and will normally have no impact on the right to personal data protection of data subjects. However, research has shown that non-personal data can be processed in a manner which results in the reidentification of data subjects thereby making data protection law potentially applicable. Therefore, the possible reidentification of natural persons might inadvertently create personal data protection concerns particularly because data producers will be under the impression that they are processing non-personal data. This article examines how data reidentification facilitates the transformation of non-personal data to personal data thereby blurring the distinction between both data types. This article further examines how the proposition for the ascription of IPR to non-personal data might result in violations of data protection law in the event of data reidentification. The reidentification of AI-generated data which results in the singling out of natural persons and the implications of such reidentification within the fields of data protection and intellectual property law are also considered. The interplay between the fields of data protection and intellectual property law when data reidentification of AI-generated data occurs will also be discussed herein. The large-scale data processing capability of AI and its technological consequences in the new age underlines the necessity of considering the theme of this article within the context of AI. The large volume of data that is being generated by AI means that if this issue is not adequately tackled, data reidentification can result in data protection risks of vast proportions. Data reidentification in this article principally refers to the singling out of natural persons from anonymised data but is also loosely used (for convenience) to refer to the identification of personal data through pieces of non-personal data. The reference to 'sui generis' in this article is used to indicate those circumstances where unique IPR regimes have already been created, or where such sui generis IPR is being proposed. This article also refers to data controllers which is distinguishable from its reference to data producers. Data controller means the natural or legal person which, alone or jointly with others, determines the purposes and means of the processing of personal data.[10] Data producers are used in this article to refer to AI inventors and other relevant persons who use technology (that is not limited to AI) to produce and generate data sets thereby giving them a vested interest in the data produced and generated.

4 European Commission, Proposal for a Regulation of the European Parliament and of the Council on a framework for the free flow of non-personal data in the European Union (COM) (2017) 495.

5 David Vaver, Intellectual Property Law: Copyright, Patents, Trademarks (2nd edn, Toronto: Irwin Law 2011) 59–60.

6 Anonymised data is data that does not relate to the identification of natural persons (Recital 26 GDPR).

7 See for instance: Václav Janeček, 'Ownership of personal data in the Internet of Things' (2018) 34 Computer Law & Security Review 5, 1039-1052.

8 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation, GDPR) (OJ L 119/2016), art 4(1).

9 Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, art 3(1).

10 Art 4(7) GDPR.

The existing IPR and related mechanisms for the protection of AI-generated data are considered in section II of this article, section III considers the existing propositions for the ascription of IPR protection to AI generated data. Section IV examines the implications of data reidentification in an AI-generated data context. The propositions aimed at the resolution of identified concerns in the reidentification of AI-generated data are addressed in Section V.

II. Existing IPR and Related Mechanisms for the Protection of AI-Generated Data

Data producers would typically seek to protect the AI-generated data produced by their AI systems with the aim of recouping their investments therein. Practically, trade secrets, copyright, contracts, etc. are being used to protect data by industry practitioners. Patents have been excluded from these considerations because even though they might be remotely capable of protecting inventive elements of non-personal data,[11] the remoteness of this possibility has caused it to receive limited attention from both existing literature and this very article.[12]

Trade secrets protect confidential information of a secret and commercial nature, which may only be licensed at the behest of the trade secret holder.[13] To qualify as a trade secret, information ought to be commercially valuable because of its secrecy, be known only to a limited number of persons, and subject to reasonable steps aimed at keeping it secret.[14] It is expected that AI-generated data will qualify for trade secrets protection if it can be proven to be 'commercially valuable because of its secrecy'.[15] Generally, information which gives an entity competitive advantage may be protected as a trade secret. In the event that AI-generated data is easily obtainable by other data producers using their own algorithms, then such AI-generated data ought not be protected because the independent discovery of a trade secret by a third party will be lawful under trade secrets law.[16] It would appear that trade secrets are used more often in practice for the protection of eligible data.[17] However, the shortcoming of trade secrets in this regard is that it may potentially prevent public access to data that may otherwise have been publicly available to members of the public if protected by a less opaque system. Therefore, in eligible circumstances described above, trade secrets can be said to be a veritable tool for protecting AI-generated data today.

Copyright is another tool worth considering for the protection of AI-generated data superficially because copyright protects texts, which is the very form of AI-generated data. The copyright eligibility requirements are the authorship[18] and the originality requirements.[19] The originality requirement is more relevant in determining the applicability of copyright protection to AI-generated data. In the many cases handed down by the Court of Justice of the European Union (CJEU), the originality criteria has been interpreted as requiring an author to leave room for "creative freedom for the purposes of copyright"[20] and also "stamp his personal touch" or "reflect his person-

11 Josef Drexl, Reto Hilty, Luc Desaunettes-Barbero, Franziska Greiner, Daria Kim, Heiko Richter, Surblyte Gintare, and Klaus Wiedemann, 'Data Ownership and Access to Data - Position Statement of the Max Planck Institute for Innovation and Competition of 16 August 2016 on the Current European Debate' (August 16, 2016) Max Planck Institute for Innovation & Competition Research Paper No. 16-10 <https://ssrn.com/abstract=2833165 > accessed 21 December 2020.

12 Taina E. Pihlajarinne and Rosa Maria Ballardini, 'Owning Data via Intellectual Property Rights: Reality or Chiemera?' in Rosa Maria Ballardini, Olli Pitkänen and Petri Kuoppamäki (eds), Regulating Industrial Internet through IPR, Data Protection and Competition Law (Alphen aan den Rijn: Kluwer Law International 2019) 115-133.

13 Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure (15 June 2016), OJ L 157/1 (Trade Secrets Directive), art 2(2). In the EU, trade secrets are not intended to 'override the exercise of the right to freedom of expression, information and the pluralism of the media'. Therefore, attempts to use trade secrets to hinder the free flow of information within the EU will fail. See Recital 19 and Art 1(2) Trade Secrets Directive.

14 Art 2 Trade Secrets Directive.

15 Art 2 Trade Secrets Directive.

16 Art 3(a) Trade Secrets Directive.

17 Taina E. Pihlajarinne and Rosa Maria Ballardini (n 12).

18 In respect of the authorship requirement, some EU copyright instruments can be interpreted as requiring that the author of a work could be a natural and/or a legal person. See Rosa Maria Ballardini, Kan He, and Teemu Roos, 'AI Generated Content: Authorship and Inventorship in the age of Artificial Intelligence' <https://www.cs.helsinki.fi/u/ttonteri/pub/aicontent2018.pdf> accessed 21 December 2020.

19 L Bently and B Sherman, Intellectual Property Law (4th edn, OUP 2014) 93-108.

20 Joined Case C-403/08 and C-429/08 Football Association Premier League Ltd et al v. QC Leisure et al [2011] ECLI:EU:C:2011:631, Paras. 95, 96, 98, 99, 155, 156, 159.

ality in the sense that he expresses his creative abilities in an original manner by making free and creative choices".[21] The CJEU held inter alia, "that any production in the literary, scientific and artistic domain, whatever the mode or form of its expression would amount to a copyrightable work so long as they are more than ideas, procedures, methods of operation or mathematical concepts as such and they are expressions of the authors' intellectual creation".[22] Therefore, AI-generated data might be protectible under EU copyright law if AI can be said to make free and creative choices in the course of producing AI-generated data. However, this is largely impossible as of today because of the pervasiveness of weak AI systems which lack the consciousness to perform tasks independently of human intervention. In other words, most if not all AI systems are unable to meet the originality requirement of copyright.[23] To ascertain authorship where both the human and the AI author have contributed to the generation of data, it is necessary to determine the extent to which the free and creative choices of both parties have resulted to the data generated.[24] For these reasons, it is safe to say that copyright protection of AI-generated data is largely an impossibility in this day.

Arguably due to the limited options available for the protection of AI-generated data, contractual agreements are largely used to protect it. These contractual agreements would typically outline the conditions for data licensing and usage, data access, etc. However, the use of contracts to protect AI-generated data is not without its own shortcomings which includes the unharmonized nature of European contract law, which causes unpredictability and complexity in the free flow of data.[25] Such complexity is visible, for instance, in the principle of the privity of contract which can be interpreted as preventing third parties from enforcing agreements that they are not directly party to.[26] Despite the identified challenges, contracts are one of the more realistic and widely used methods for the protection of AI-generated data today.

III. Propositions for the Ascription of IPR to (AI-Generated) Data

In order to bypass the limitations of IPR protections to (AI-generated) data highlighted above, various scholars and stakeholders have made propositions for the IPR protection of data. Before substantively considering these propositions, it is important to note that the concept of creativity and what constitutes same is key to the ascription of IPR to AI-generated data. Traditionally, natural, and (in some cases), legal persons generally create works which are eligible works for IPR protection.[27] Creativity is that element that justifies the incentivisation of creatives through IPR. The determination of what is protectible within the context of AI-generated data is a particularly knotty issue because of the relative ease with which it might amount to protecting facts which are typically not protected under IP law.[28] Traditionally, creative value could only be added to non-personal data by natural/legal persons, but this has changed because AI systems are now capable of doing the same. One factor that might put a strain on the ascription of IPR to AI generated non-personal data are the divergent roles played by multi-parties thereby creating a challenge as to who the proper holder of the IPR ought to be. From a traditional IPR perspective, there is usually an eligibility requirement that necessitates the author to express his creativity in the work. The originality requirement performs this function within the scope of copyright law where the work is required to both be an expression of the author's creative abilities in the work as well as his/her own intellectual creation.[29] Similar objectives can be gleaned from the inventiveness and distinctiveness requirements of patents and trademarks

21 Case C-145/10 Painer v. Standard VerlagsGmbH ECLI:EU:C:2013:138.

22 See Case C-310/17 Levola Hengelo BV v. Smilde Foods ECLI:EU:C:2018:899, para. 39.

23 It has been proposed that some AI systems are close to having consciousness. An example of such AI is the Aican which designs art on its own but still requires human intervention to name it. See Ahmed Elgammal, 'Meet AICAN, a machine that operates as an autonomous artist' (The Conversation, October 17 2018) <https://theconversation.com/meet-aican-a-machine-that-operates-as-an-autonomous-artist-104381> accessed 29 July 2022.

24 Anette Alen-Savikko, Rosa Maria Ballardini and Taina Pihlajarinne, 'Tekoälyn Tuotokset ja Omaperäisyysvaatimus – Kohti Koneorientoitunutta Tekijänoikeutta?' (2018) 116 Lakimies/7-8.

25 Taina E. Pihlajarinne and Rosa Maria Ballardini (n 12).

26 Taina E. Pihlajarinne and Rosa Maria Ballardini (n 12).

27 Directive 2009/24/EC of the European Parliament and of the council of 23 April 2009 on the legal protection of computer programs (2009) OJ L111, Art 2 (1).

28 David Vaver (n 5) 59–60.

29 Case C-145/10 Eva-Maria Painer v Standard Verlages GmbH et al. [2013] ECLI:EU:C:2013:138, 42.

respectively.[30] From the perspective of AI-generated data, AI systems will add value to data which can be gleaned as the 'creativity' of such AI systems.[31] This creative value added by AI will therefore merit some form of IPR protection in accordance with the existing propositions for the protection of non-personal data. It is necessary to consider these propositions for the ascription of IPR protection to (AI-generated) data as a necessary condition precedent to appreciating the intersection between IPR and data protection law which will be highlighted in subsequent parts of this article.

The concept of computational creativity has been coined to depict the capability of AI to produce works which are based on its ability to exercise judgement and some level of randomness/unpredictability and not merely based on replications of existing works.[32] Hristov argues that such randomness/unpredictability, just like autonomously learned behavior, is distinguishable from the input of the human programmer and cannot be attributed to the human programmer of AI as a result.[33] The ability of AI to reason at different levels of abstraction or work in more than one domain without reprogramming has also been said to potentially result in creativity.[34] Based on this position, AI systems that display such levels of creativity ought to be provided with IPR protection for the datasets that they generate. This ability is more synonymous with strong AI which can operate without human intervention and, at the time of writing this article, is yet to become prevalent.[35] The European Commission (EC) acknowledged the proposition of Zech which makes a case for the protection of the data producer's right in the generated data.[36] In outlining the scope of the proposed data producer's right, the EC (advancing Zech's proposition) noted that this could include statistical analyses.[37] Therefore, it can be inferred that the EC views statistical analyses of data (and not just the data itself) as a creative improvement of data that merits some form of protection postulated as the data producer's right. Hugenholtz reckons that this proposed data producer's right will most likely qualify as an IP right.[38] Since this data producer's right does not materially correlate to the existing IPR framework, it is safe to refer to it as some form of sui generis protection. It would appear that this approach has also been adopted in the US where the court has recognised statistical analyses of data as an IPR (specifically, a copyright).[39] Within the context of AI, AI systems have been developed to carry out analysis on data in a manner that produces random and unpredictable results which might not have been foreseen at the time when the AI system was fed with the data.[40] Therefore, an analysis which forms the output of an AI system would itself be the subject matter of some form of IPR (the data producer's right).

Data scientists have also weighed in on the concept of creativity in relation to data. Vlastelica argues that during data analytics, creativity can be deployed in various ways including hypothesis generation, feature engineering, workflow management, simplicity, and finally visualization.[41] Howev-

30 An invention is required to be novel and must not be anticipated by prior art which could be interpreted as requiring a patentable invention to be the creative work of the inventor. Trademarks are required to be distinctive in the sense of distinctly identifying the product or service with which they pertain. This might also require that the trademark is an intellectual creation of its creator. See World Intellectual Property Organisation (WIPO), Understanding Copyright and Related Rights (2nd edn, Geneva 2016) 17, 82.

31 There is no doubt that the use of the term creativity in reference to AI is very disputable because most AI systems lack the consciousness to be creative. See Jonathan Zittrain, 'The Hidden Costs of Automated Thinking' (The New Yorker, July 23 2019) <https://www.newyorker.com/tech/annals-of-technology/the-hidden-costs-of-automated-thinking> accessed 14 March 2022.

32 Margaret Boden, 'Computer Models of Creativity' (2009) 30 AI Magazine 3, 23-24 <https://ojs.aaai.org/index.php/aimagazine/article/view/2254/2100> accessed 14 March 2022. See also Roger Schank and Christopher Owens, 'The Age of Intelligent Machines: the Mechanics of Creativity' in Raymond Kurzweil (ed), The Age of Intelligent Machines 394 (1991) 149-151. See also David Cope, Computer Models of Musical Creativity 12 (MIT 2005).

33 Kalin Hristov, 'Artificial Intelligence and the Copyright Dilemma' (September 1, 2016) 57 IDEA: The IP Law Review, 3 <https://ssrn.com/abstract=2976428> accessed 14 March 2022.

34 Francisco Câmara, Creativity and Artificial Intelligence: A Conceptual Blending Approach (De Gruyter Mouton 2008) 10.

35 Ismail Bello, 'Beginners guide to Artificial Intelligence' (Becominghuman July 17 2017) <https://becominghuman.ai/beginners-guide-to-artificial-intelligence-ai-ec8a409b6424> accessed 14 May 2022.

36 Herbert Zech, 'A Legal Framework for a Data Economy in the European Digital Single Market: Rights to Use Data' (November 20, 2016) 11 JIPLP 460-470.

37 European Commission (n 4) 33.

38 Bernt Hugenholtz, 'Data Property: Unwelcome Guest in the House of IP' (2017) <https://pure.uva.nl/ws/files/16856245/Data_property_Muenster> accessed 9 March 2021.

39 Lowry v Legg Mason, 271 F Supp 2d 737 (D Md 2003).

40 Nikolaus Forgó, Stefanie Hänold, and Benjamin Schütze, 'The principle of purpose limitation and big data' in Marcelo Corrales et al. (eds), New Technology, big data and the law (Springer Publishers, 2017) 20.

41 Damjan Vlastelica, 'On the importance of creativity in Data Analytics' (DataScience, 4 Nov 2019) <https://towardsdatascience.com/on-the-importance-of-creativity-in-data-analytics-469efc9c3ac5?source=social.tw> accessed 9 March 2021.

er, it is yet to be seen if these activities will be protectible either as IPR or some other form of sui generis right.[42]

Another means through which IPR protection can be ascribed to AI-generated data is through codification. To put it simply, IPR protection for AI-generated data will apply if the law-making apparatus (of the EU) so decides. The possible use of codification is another means through which a sui generis right might be created to bypass the limitations that might be imposed by traditional IP rights. Therefore, whether the lawmaker decides to ascribe traditional IPR to data or create some sort of sui generis protection, it can be expected that such protection will be intended to provide incentivization to data producers. The Database Directive (DB directive) is significant in this regard because it outlines sui generis rights aimed at incentivising database producers, the DB directive itself being an example of a legislation that does something similar. In creating a sui generis set of rights for databases, Article 4 of the DB Directive designates the natural person(s) who created the base or, where the legislation of the Member States so permits, the legal person designated as the rightholder as an author. The EU Database protection regime covers copyright protection granted to eligible databases and also a sui generis right granted to databases that are produced as a result of the substantial investment of the database producer.[43] It is argued that this adaptation of EU copyright law to incentivize the investment(s) of database producers in databases is further proof that if deemed necessary, the EU commission is able to create a sui generis set of rights for the protection of the rights of the data producer.

Flowing from the above propositions for the ascription of some form of IPR to non-personal data, it is arguable that AI-generated data might, by extension, also be eligible for some form of IPR protection. This is because AI systems are imbued with algorithms which have been developed to analyse data and provide insights which would otherwise have been unavailable and unpredictable at the beginning of the processing activity.[44] For instance, data analysis conducted on vehicular (brake usage) data, has been used to elicit behavioral and driving patterns from distinct vehicle drivers even though this finding would not ordinarily have been anticipated in the traditional processing of vehicle data.[45] There is no doubt that any attempt to ascribe IPR to AI-generated data will face other challenges including the eligibility of AI systems to own or hold IPR, the necessity of such protection, etc.[46] At the time of writing this article, the (proposition(s) for the) IPR protection of AI-generated data is at best an aspiration which is yet to find acceptance in EU law and policy.[47]

For the sake of completeness, it is necessary to also consider the possibility of ascribing IPR to personal data especially because this article addresses both data types. As will be discussed subsequently,[48] AI has blurred the distinction between personal and non-personal data. This is because AI can be deployed to reidentify non-personal data thereby bringing it within the scope of data protection law.[49] It is interesting to consider the possibility of IPR protection within the scope of data protection law as this enhances the appreciation of the intersections between IPR and data protection law, as will be subsequently highlighted. Under the current EU legal framework, it is doubtful that any person other than the data subject can exercise control over their own data. The rationale behind this position is traceable to Recitals 7 and 68 of the General Data Protection Regulation (GDPR) which gives 'natural persons control over their own personal data'. It is doubtful that in the light of said control, third parties can exercise 'control' over the personal data of others outside of statutorily permitted circumstances. Even the `control' granted to data subjects does not translate to

42 Annette Kur and Thomas Dreier, European Intellectual Property Law: Text, Cases and Materials (Edward Elgar Publishing, 2013).

43 Art 3 and 7 Database Directive

44 Forgó et. al. argue that during a data processing activity, big data (including personal data) may be further processed in way which might not have been envisaged at the inception of a processing activity. Nikolaus Forgó et. al. (n 40) 20.

45 Miro Enev, Alex Takakuwa, Karl Koscher and Tadayoshi Kohno 'Automobile Driver Fingerprinting' (2016) 1 Proceedings on Privacy Enhancing Technologies 34-50 <https://doi.org/10.1515/popets-2015-0029> accessed 24 August 2020.

46 Some of these concerns have been addressed in: Emmanuel Salami, 'AI-generated works and copyright law: towards a union of strange bedfellows' (2021) 16 JIPLP 2, 124-135 <https://doi.org/10.1093/jiplp/jpaa189> accessed 10 March 2021.

47 Some of these aspirations are reflected in scholarly works such as: Wolfgang Kerber, 'A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis' (2016) GRUR 989. See also Taina E. Pihlajarinne and Rosa Maria Ballardini (n 12) 115-133; Herbert Zech (n 36) 463-464.

48 Section 4 of this article.

49 Such non-personal data could either be data which never related to a natural person or personal data which no longer relates to a natural person as a result of manipulation.

property rights over their personal data.[50] An example of the control being exercised by data subjects is reflected in the right to data portability under which data subjects can request data controllers to transfer their personal data to other data controllers to make entry into new digital markets less strenuous.[51] Another difficulty that can be possibly encountered in the ascription of IPR to personal data is the ethical argument against the commercialisation of human rights[52] with the right to personal data protection being guaranteed within the EU legal framework.[53] The European Data Protection Supervisor (EDPS) in support of this position notes that there should be no market for personal data just as a market for human organs should not exist.[54] Despite this position, various propositions have been made towards the ascription of IPR to personal data.[55] Some businesses within and outside the EU are involved in the sale of personal data and have justified their business model on grounds which include the non-prohibition of such transactions within the legal framework of EU law.[56] Despite the EDPS' counter position, Article 3(1) of the Digital Content Directive regulates transactions 'where traders supply or undertake to supply digital content or digital services to consumers, who provide or undertake to provide personal data in exchange'.[57] By virtue of this provision of the Directive, individuals are allowed to rely on contractual remedies and data protection rights even after they have given their personal data in exchange for free access to social media platforms. This provision gives some form of legitimacy to the barter of personal data for digital content and services although this practice will still be subject to the provisions of the GDPR. By implication, data subjects will still be able to exercise control over their personal data.

IV. Data Reidentification and the Implications for AI-Generated Data

Having considered the propositions for the ascription of IPR to AI-generated data, it is necessary to examine the implications of data reidentification on such AI-generated data because this is the point where an intersection might occur between IPR and data protection law. An analysis of the concept of (non) personal data under the GDPR is necessary to properly understand the concept of data reidentification that is addressed in this article. It is trite law that data protection law only protects personal data that relates to an identified or identifiable natural person.[58] In expatiating on the definition of personal data, the Article 29 Working Party (A29WP) sheds some light on the elements which constitute the definition of personal data. These quadripartite elements are - 'any information', 'relating to', 'an identified or identifiable', 'natural person'.[59] Though the opinion of the A29WP was particularly tailored to the Data Protection Directive, we can still make some comparisms between both definitions especially to the extent that they remain unchanged under the GDPR. [60] The use of the term 'any information' is sug-

50 Herbert Zech (n 36) 463-464.

51 Recital 68 and Art 20 GDPR.

52 Lisa Cosgrove and Allen F. Shaughnessy, 'Mental Health as a Basic Human Right and the Interference of Commercialised Science' (2020) Health and Human Rights Journal <https://www.hhrjournal.org/2020/06/mental-health-as-a-basic-human-right-and-the-interference-of-commercialised-science/> accessed 10 July 2020.

53 Art 8 Charter of Fundamental Rights of the European Union [2012] OJ C326/391.

54 European Data Protection Supervisor, 'Opinion 4/2017 on the Proposal for a Directive on certain aspects concerning contracts for the supply of digital content' (2017) 7 <https://edps.europa.eu/sites/edp/files/publication/17-03-14_opinion_digital_content_en.pdf> accessed 25 September 2021.

55 World Economic Forum, 'Personal Data: The Emergence of a New Asset Class' (2011) 10 <https://perma.cc/T7JL-BZXK> accessed 11 July 2020.

56 Sanna Toropainen, 'Buying and selling personal data directly from consumers' (2020) Privacy Laws & Business International Report 19-21.

57 Directive (EU) 2019/770 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning contracts for the supply of digital content and digital services (22 May 2019), OJ L 136/1.

58 Recital 26 GDPR. See also Art 4(1) GDPR which defines personal data as any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

59 Article 29 Working Party, 'Opinion 04/2007 on the concept of personal data' (WP 136) 01248/07/EN, 6.

60 One of the key differences in the definition of personal data between the Data Protection Directive and the GDPR lies in the fact that the latter legislation places more emphasis on the term 'natural persons' than the former legislation. See Article 2(1) Directive 95/46/EC of the European Parliament and of the Council on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data, 24 October 1995 (Data Protection Directive), Official Journal L 281 , 23/11/1995 P. 0031 - 0050.

gestive of the intention of the draftsman of the GDPR to adopt a wide interpretation of personal data which could include both objective and subjective information.[61] This view has also received judicial support from the CJEU.[62] In respect of the second element, data can be said to relate to a natural person if it is about the relevant person.[63] For the third element of the definition of personal data, a natural person will be deemed 'identified or identifiable' if they can be distinguished or singled out from other members of a group. These 'identified or identifiable' natural persons can be singled out 'directly or indirectly' either through their names or the combination of various datasets.[64] To ascertain whether a natural person can be identified within a dataset, factors such as the costs of, and the amount of time required for identification, the available technology at the time of the processing and technological developments, ought to be taken into consideration.[65] The fourth element which is its applicability to 'natural persons' embodies the fact that data protection law ought to apply only to human beings.[66]

Non-personal data could either be - inherently incapable of identifying natural persons because they never related to identified or identifiable natural persons (e.g. statistical data about the number of traffic signs on a particular road) or manipulated in a manner that it becomes anonymised and incapable of leading to the identification of natural persons. Recital 26 GDPR makes it abundantly clear that pseudonymized data which could be attributed to a natural person by the use of additional information should be considered as personal data.[67] The recital goes further to exclude anonymised data, (which is data that does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not identifiable), from the scope of data protection law. Therefore, any data (including pseudonymous data), which relates to an identified or identifiable natural person constitutes personal data and falls within the scope of data protection law while those that do not, are excluded. The implication of Recital 26 GDPR is that it adopts a risk-based approach to the definition of personal data owing to the fact information would constitute personal data when identification of a natural person is 'reasonably likely'. When this is not the case, such information will be deemed non-personal data. Finck and Pallas point out a long line of conflicting interpretations of the risk-based approach adopted by the GDPR.[68] For instance, the UK Information Commissioner's Office (ICO) in its interpretation of Recital 26 GDPR adopts the position that the determination of whether information will constitute personal or non-personal data is the 'the identification or likely identification' of a data subject. The ICO further acknowledges the risk of re-identification through data linkage which could prove unpredictable because of the uncertainty surrounding which data is already available or which data might be made available in the future.[69] The A29WP while acknowledging the risk-based approach of the (then draft to the) GDPR, took the position that anonymisation ought to irreversibly prevent identification.[70] In other words, the A29WP viewed anonymisation as a technique which ought to make data reidentification impossible. The French supervisory authority, Commission Nationale de l'informatique et des Libertés (CNIL) while acknowledging an inherent difficulty, posits that anonymisation ought to make identification practically impossible thereby making anonymisation irreversible and not permissive of any further processing of personal data.[71] The Irish supervisory authority takes a position

61 Such objective information could include the fact that a person is terminally ill. Using the same example, subjective information about this could include an opinion that a personal should not be offered an employment opportunity because of their terminal illness. In both cases, the objective and subjective views on a data subject will constitute personal data. For further readings, see Article 29 Working Party (n 59) 6-9.

62 Case C-434/16 Peter Nowak [2017] EU:C:2017:582, para 34.

63 Article 29 Working Party (n 59) 9-12.

64 Article 29 Working Party (n 59) 12-14. For further reading, see Nadezhda Purtova 'The law of everything. Broad concept of personal data and future of EU data protection law' (2018) 10 Law, Innovation and Technology 1 40-81.

65 Recital 26 GDPR.

66 Art 6 Universal Declaration of Human Rights (UDHR), 1948.

67 Art 4(5) GDPR.

68 Michèle Finck and Frank Pallas, 'They Who Must Not Be Identified - Distinguishing Personal from Non-Personal Data Under the GDPR' (2019) Max Planck Institute for Innovation & Competition Research Paper No. 19-14, 7-10 <https://ssrn.com/abstract=3462948> accessed 23 July 2022.

69 Information Commissioner's Office, 'Anonymisation: Managing Data Protection Risk Code of Practice' (November 2012) <https://ico.org.uk/media/1061/anonymisation-code.pdf> accessed 13 July 2022.

70 Article 29 Working Party, Opinion 05/2014 on Anonymisation Techniques (WP 216) 0829/14/EN.

71 Commission Nationale de l'Informatique et des Libertés, 'Comment prévenir les risques et organiser la sécurité de vos données?' (16 April 2019) <https://www.cnil.fr/fr/comment-prevenir-les-risques-et-organiser-la-securite-de-vos-donnees> accessed 22 July 2022.

similar to that of the A29WP though it differs to some extent by stating that it suffices to prove that the identification of data subjects is unlikely given the circumstances of the individual case and the state of technology. Therefore, while acknowledging irreversibly anonymised data as non-personal data in accordance with the A29WP, the Irish supervisory authority still leans towards Recital 26 GDPR by listing the circumstances of such anonymisation and the state of technology as factors to be considered in determining the irreversibility and effectiveness of the anonymisation.[72] These diverging interpretations have prevented legal certainty as to what test ought to be applied in practice.

Pursuant to Recital 26 GDPR, the relevant criterion for assessing whether data is personal or non-personal remains 'identifiability' of the data subject. To determine such 'identifiability', account should be taken of all the means reasonably likely to be used to identify the natural person, such as singling out.[73] To ascertain whether there is reasonable likelihood that these means for identification of the data subject will be used, objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments, ought to be used. The A29WP lists three conditions to be considered for determining if de-identification has occurred. These factors are (i) is it still possible to single out an individual; (ii) is it still possible to link records relating to an individual, and (iii) whether information concerning an individual can still be inferred.[74] It is noteworthy that only criterion (i) can be traced to the GDPR.[75] As will be shown subsequently, in spite of these criteria, the identification of data subjects from apparently anonymised data has become more prevalent and any attempt to present data as irreversibly anonymised will likely not be infallible.

Personal data such as (user)names, home addresses, IP addresses, etc. are clearly identifiers which undoubtedly fall within the framework of (EU) data protection law. However, other data types (such as anonymised customer feedback, movie reviews, brake pedal usage in cars etc.), though conceivable as not directly capable of identifying natural persons have in fact led to the identification of natural persons. While the possibility of identifying natural persons is the primary distinction between personal and non-personal data, the endless possibilities that abound in singling out natural persons from non-personal and/or anonymised data has obscured this distinction. The concept of reidentification science has been proposed as the reidentification of natural persons through non-personal data in circumstances where the reidentification of natural persons would otherwise have been impossible.[76] In practice, the tendency is to view non-personal and/or anonymised data as incapable of singling out natural persons. However, the possibility of data reidentification particularly through AI, has changed this view.

Several notable data reidentification experiments aimed at proving that anonymisation may not necessarily deidentify data have been undertaken by scholars. In the AOL case, anonymised data, made up of search queries, were publicly released to support open research. Researchers were able to single out natural persons by combining the search queries with other data categories.[77] In another case, 'anonymised' medical records including those of the then Governor of the American state of Massachusetts were singled out and other natural persons reidentified through the combination of ZIP codes, gender, and birthdays.[78] Scholars have also been able to identify Netflix customers through their movie rating by comparing the Netflix rating data to similar data from the Internet Movie Database.[79] AI is also increasing the possibility of data reidentification thereby casting fundamental doubts on the effectiveness of data anonymisation. Two students developed an AI system capable of combing through large vol-

72 Data Protection Commission, 'Guidance on Anonymisation and Pseudonymisation' (June 2019) 2 accessed 31 July 2019. < https://www.dataprotection.ie/sites/default/files/uploads/2019-06/190614%20Anonymisation%20and%20Pseudonymisation.pdf> accessed 23 July 2022.

73 'Singling out' refers to 'the possibility to isolate some or all records which identify an individual in the dataset'. See Article 29 Working Party (n 70) 3.

74 ibid.

75 Recital 26 GDPR.

76 Paul Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2010) 57 UCL Law Review 1723-1726.

77 Ibid 1717-1718.

78 Latanya Sweeney, 'Matching Known Patients to Health Records in Washington State Data' (June 5, 2013) <https://ssrn.com/abstract=2289850> accessed 26 August 2020.

79 Arvind Narayanan and Vitaly Shmatikov, 'Robust De-Anonymisation of Large Sparse Datasets' (2008) In Proceedings of the 2008 IEEE Symposium on Security and Privacy 111, 121 <https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf> accessed 26 April 2022.

umes of consumer datasets with the aim of identifying personal data from non-personal/anonymised data.[80] Personal data categories such as emails, usernames etc. which are mostly obtained from data leaks are inserted into the tool which then attempts to match it with non-personal data. Having used the said AI to analyze thousands of datasets, it was found that it was not so difficult to identify natural persons particularly with the mass of publicly available information from data leaks. The students concluded that pieces of data from various leaks could be put together to create large volumes of personal data. Researchers have also developed a tool capable of correctly identifying 99.98 percent of Americans in any anonymised dataset using just 15 characteristics.[81] Another study at the Massachusetts Institute of Technology revealed that anonymised credit card data could result in the identification of natural persons 90 percent of the time through the use of four relatively vague points of information.[82] One research focused on anonymised user vehicle data found that 15 minutes worth of data from brake pedal use could let researchers identify the right driver with the aid of machine learning techniques.[83] This was despite the collection of very small amounts of sensor data.

In the proposed ascription of IPR protection for AI-generated data, it is expected that data producers will seek to profit off AI-generated data by licensing them to third-party licensees. If not properly regulated, licensees might be able to reidentify anonymised/non-personal data either using AI or the combination of different databases. Such an occurrence can give rise to various data protection issues such as compliance with the principles of data protection law[84] including the requirement for a justifiable legal basis for processing such data,[85] the provision of requisite information about the nature of the processing activity to data subjects,[86] the existence of a clear and defined purpose of processing at the time of obtaining the data,[87] etc. The primary and foundational data protection law issue that will arise upon data reidentification will be the lack of a justifiable legal basis. This is because upon reidentification, a personal data processing activity will be in motion without a justifiable legal basis.[88] The result of this will be an unprecedented infraction of the right to personal data protection. Furthermore, the determination of the appropriate party to be designated as the data producer may not always be devoid of complication. For instance, this can occur in complex processing operations involving multi parties supplying various Internet of Things (IOT) used in the collection and processing of data.[89] An example of this occurrence can be found in autonomous vehicles where various parties supply the IOT that supports its operation.[90]

This possibility might even become further aggravated should data subjects attempt to enforce their data subject rights in cases where (re)identified AI-generated data is protected by IPR or some sui generis right that might be created for this purpose.[91] For instance, this raises the question of how the data sub-

80 Karl Bode, 'Researchers find 'Anonymised' data is even Less Anonymous than We Thought' (Motherboard, 3 February 2020). <https://www.vice.com/en_us/article/dygy8k/researchers-find-anonymised-data-is-even-less-anonymous-than-we-thought> accessed 26 August 2020.

81 Luc Rocher, Julien M. Hendrickx and Yves-Alexandre de Montjoye, 'Estimating the success of re-identifications in incomplete datasets using generative models' (2019) 10 *Nat Commun* 3069 <https://doi.org/10.1038/s41467-019-10933-3> accessed 21 August 2020.

82 Rob Matheson, 'The Privacy Risks of Compiling Mobility data, Merging different types of location-stamped data can make it easier to discern users' identities, even when the data is anonymised' (MIT News Office, 7 December 2018). <https://news.mit.edu/2018/privacy-risks-mobility-data-1207> accessed 24 August 2020.

83 Miro Enev, Alex Takakuwa, Karl Koscher, and Tadayoshi Kohno (n 45).

84 The principles of data protection law are encapsulated in Art 5 GDPR. See also Lee A. Bygrave, Data Protection Law: Approaching Its Rationale, Logic and Limits (Kluwer, The Hague 2002) 57-89.

85 Ibid. See also Art 6 GDPR.

86 Art 13 GDPR.

87 This is known as the purpose limitation principle. Art 5 (1) (b) GDPR.

88 Article 29 Working Party, 'Guidelines on Consent under Regulation 2016/679' (2017) 17/EN WP 259, 17-22.

89 IoT is any physical device capable of connectivity that directly interfaces the physical world, such as embedded devices, sensors etc. Singh Jatinder, Thomas Pasquier, Jean Bacon, Hajoon Ko, and David Eyers, 'Twenty Security Considerations for Cloud-Supported Internet of Things' (2016) 3 IEEE Internet of Things Journal 3 269-284.

90 Andreas Herrmann, Walter Brenner, Rupert Stadler, Autonomous Driving: How the Driverless Revolution will Change the World (Emerald Group Publishing 2018) 9.

91 Art 12-23 GDPR. See also Christopher Kuner, Lee Bygrave, Christopher Docksey, Laura Drechsler (eds), The EU General Data Protection Regulation: A Commentary/Update of Selected Articles (OUP, 2020) <https://ssrn.com/abstract=3839645> accessed 12 February 2021.

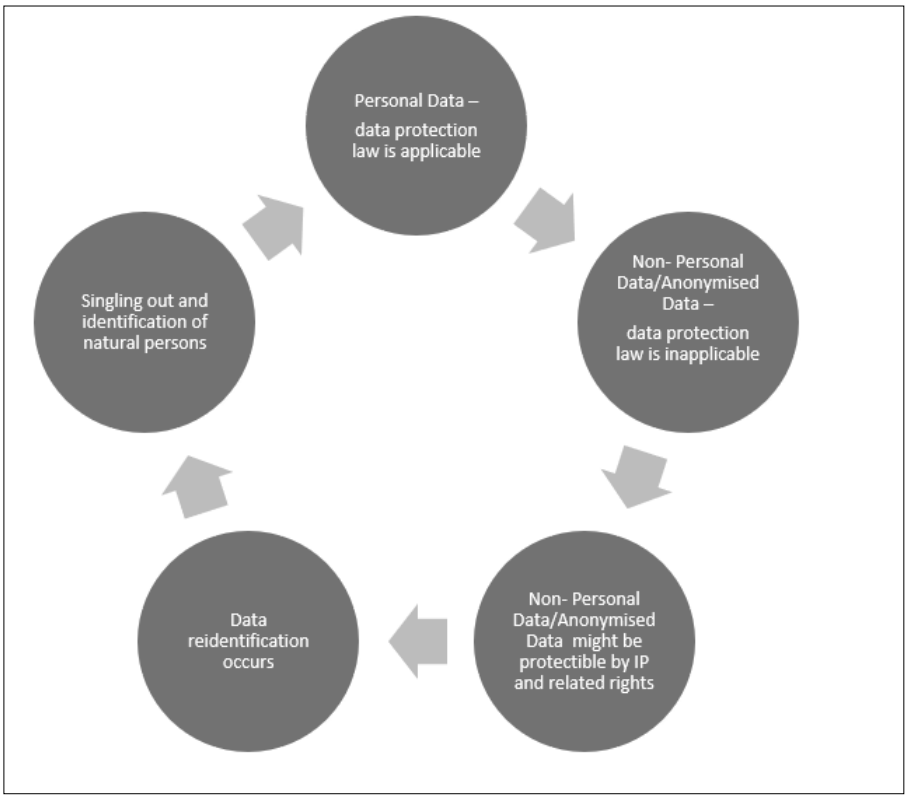

Figure 1: This figure depicts the transformation of personal data to non-personal data, the possible protection of the non-personal data by IPR, and the possibility of data reidentification which can result in the singling out of data subjects. This indicates the interaction between data protection and IP law in the transformation of data from personal data to non-personal data and possibly back to personal data.

ject rights will be fulfilled where components of (re)identified personal data have been processed as part of protected data. Such an occurrence raises questions pertaining to the balancing of competing interests between IPR and data protection law.

V. Resolving Identified Concerns in the Reidentification of AI-Generated Data

As indicated above, the proliferation of AI in culmination with various data reidentification techniques means that the lines between personal and non-personal data have become less clear.[92] Since the propositions pertaining to the ascription of (sui generis) IPR to data are particularly focused on non-personal data, it is necessary to ensure that no elements of personal data are included therein. This requirement is actually quite tricky because of the possibility of AI to single out personal data from apparently non-personal data. A framework of data categories capable of singling out natural persons from IPR protected non-personal data may be excluded from protection to prevent any overlap that could violate data protection law. To achieve this, it is necessary to reduce the scope of IPR protection afforded to AI-generated data by not ascribing such protection to data contain-

92 Karl Bode (n 80).

ing personal identifiers (such as names, emails, address, gender, etc.) and/or data that can potentially result in the identification/singling out of natural persons. This is particularly because previous attempts at data reidentification have proven that personal identifiers were important in the reidentification of natural persons, an occurrence which can be reduced by their exclusion.

It might also be necessary to treat anonymisation merely as a technique aimed at securing or restricting access to data (just like encryption and pseudonymisation)[93] rather than a technique deemed as preventing the identification of natural persons and a justification for the non-applicability of data protection law. The possibilities for data reidentification (highlighted in section 4 of this article), creates some difficulty in distinguishing between data anonymisation and data pseudonymisation since both approaches can, (in reality), result in the singling out of natural persons. Article 4(5) GDPR defines 'pseudonymisation' as 'the processing of personal data in a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person'. Based on the provision of Recital 26 GDPR, the primary difference between 'pseudonymisation' and 'anonymisation' pertains to the possibility of 'identifiability', which is determined by taking all the means that are reasonably likely to be used to identify the natural person, such as singling out, into consideration. Therefore, while pseudonymisation is ordinarily expected to result in the identification of natural persons, anonymisation is not. It is argued that based on the seemingly boundless possibilities of data reidentification, pseudonymisation and anonymisation, though defined differently, may well have the same implication from a personal data protection perspective. Flowing from these reidentification possibilities, it is also arguable that anonymisation might not be very different from pseudonymisation particularly because of the impact of AI which will potentially result in the identification of natural persons. It is therefore necessary to reconsider the status ascribed to data anonymisation as a technique that excludes the applicability of data protection law in order to prevent data protection risks that might occur as a result of data reidentification. This approach is not only practical but also realistic because as research has shown, there is no absolute anonymisation. In acknowledging the limitations posed by the heavy reliance of data protection law on anonymisation, Ohm proposes the scrapping of the term altogether and suggests replacing it with a term that suggests an 'attempt to achieve anonymity'.[94] Finck and Pallas recommend that the risk based approach stipulated in Recital 26 GDPR is a more meaningful approach at distinguishing between personal and non-personal data rather than assuming that data is automatically incapable of identifying natural persons once they have been anonymised.[95] Whether by an express abolition of the term as currently used or by revisiting the meaning/expectation ascribed to the term, it is clear that the singling out of data subjects remains highly likely despite anonymisation. Therefore, it is necessary to reevaluate the practice of the non-applicability of data protection law once anonymisation techniques have been applied.

Despite the concerns expressed about data anonymisation, its advantage of being a preventive and a pre-emptive measure that facilitates the protection of personal data rather than a remedial approach to the protection of personal data cannot be denied. This is because techniques such as data anonymisation, (as well as pseudonymisation, encryption),[96] etc. are very important to data protection law because of their role in the prevention of data protection risks at the inception of the processing activity, rather than as remediation actions implemented after the occurrence of a breach. This is an effective way of enforcing said right. Where possible, (technological) measures which might prove helpful in the improvement of data anonymisation should be adopted. Techniques such as interactive techniques, aggregation, access controls, audit trails, etc. which have been introduced to improve the anonymisation of data can be adopted for enhancing the data anonymisation process with the overarch-

93 For further readings on pseudonymisation and encryption techniques, see Article 29 Working Party (n 70) 20-23.

94 Ohm suggested replacing the term 'anonymisation' with 'scrub' because the latter reflects an attempt to anonymise which takes away the expectation that identification is impossible. For further reading, see Paul Ohm (n 76) 1744-1755.

95 Michèle Finck and Frank Pallas, (n 68) 18.

96 Orin Kerr, and Bruce Schneier, 'Encryption Workarounds' (March 20, 2017) 106 Georgetown Law Journal 989.

ing objective of protecting personal data.[97] These techniques can form part of the efforts to be channeled towards data anonymisation to further strengthen it and reduce the possibilities of the reidentification of natural persons. To prevent data (re)identification, it might be effective to shift the focus of relevant legal instruments to the prevention of data reidentification rather than focusing on data anonymisation. This approach will completely prohibit data reidentification in circumstances which unlawfully result in the singling out of natural persons. This will also be a clear acknowledgement of the highlighted shortcomings of data anonymisation within the context of data reidentification. It is expected that the heavy reliance on data anonymisation as a technique for data deidentification will be reduced by taking its shortcomings into more consideration.

Furthermore, to prevent the reidentification of 'anonymised' data,[98] data subjects will benefit from exercising their data subject rights to prevent weak anonymisation of their data. Data subjects (and even regulators) can hold data controllers/processors accountable in a manner that mandates them to comply with relevant provisions of data protection law, particularly when it comes to ensuring that anonymised data is truly anonymised. Complementarily, data controllers/processors (as the case may be) ought to be tasked with informing data subjects, when data which though seems non-personal and/or anonymised, might result in their (re)identification. Using autonomous vehicles as an example, data subjects ought to be informed in advance that they can possibly be identified (no matter how remotely) through their automobile sensors and brake usage.[99] This will at least create some level of awareness, transparency and accountability in respect of the nature of relevant processing activities between the data subject and the data controllers/processors respectively. Data controllers ought also to be tasked with a responsibility to inform data subjects of the controls and safeguards aimed at preventing reidentification and protecting the right to personal data protection. This can, for instance, be added to the transparency requirements of the GDPR.[100] This will also potentially make the quality/effectiveness of data anonymisation standards one of the factors to be considered by data subjects in the selection of service providers. The desire to be preferred by data subjects as compliant service providers will in turn, motivate data controllers/processors to improve the standards of data anonymisation deployed by them as practically as possible. This will not only improve the standard of data anonymisation, but also the consequent possibility of reduction in data reidentification. It is also suggested that compliance measures which may help restrict or define the permissible circumstances for the reidentification of anonymised data should be codified. Should AI-generated data be protected under IP law, standardised contractual terms which prohibit and/or clearly define the conditions for data reidentification must be established. To ensure compliance with the codified measures and standardised contractual terms, the existence/imposition of steep fines aimed at compelling compliance might further motivate data controllers and processors to act appropriately.

As previously highlighted, data reidentification will raise possible implications for AI-generated data protected by IPR, thus necessitating the balancing of competing interests. One of such implications can occur in the event of data (re)identification which results in data subjects requesting for the enforcement of their access rights, thereby making the creation of a balance between the right to data protection and IPR necessary. This is particularly possible where protected AI-generated data results in the reidentification of data subjects who then exercise their access rights thereby possibly resulting in the sharing of (parts of) IPR protected AI-generated data. The CJEU provides some insight into balancing competing interests in several cases including Bonnier Audio AB and Others v. Perfect Communication Sweden AB.[101] In that case, five publishing companies holding the copyright on 27 audiobooks approached the Swedish court requesting it to compel an Internet Service Provider (ISP) to disclose the contact details of the infringers of their copyright. The ISP challenged the application on the ground that it violated the (now

97 Paul Ohm (n 76) 1751.

98 In an ideal scenario, when personal data is properly anonymised, data controllers ought not to know the identity of the data subjects. Therefore, the anonymisation referred to above is for those circumstances when data anonymisation has been incorrectly executed.

99 Miro Enev, Alex Takakuwa, Karl Koscher, and Tadayoshi Kohno (n 45).

100 Art 5 (1) (a), 13, 14 GDPR.

101 C-461/10, Bonnier Audio AB, Earbooks AB, Norstedts Förlagsgrupp AB, Piratförlaget AB, Storyside AB v. Perfect Communication Sweden AB, 19 April 2012.

invalidated) data retention directive.[102] The case was referred to the CJEU to determine whether the then data retention directive precluded the enforcement of Article 8 of the Intellectual Property Rights Enforcement Directive under which an injunction requiring ISPs to transmit the contact details of infringing subscribers to copyright holders can be issued.[103] In balancing the competing interests, the CJEU held that transmitting the personal data to the copyright holders was necessary in civil proceedings to ensure copyright protection.[104] The CJEU further held that both directives (pertaining to IP right and the right to personal data protection) must be interpreted in a way that does not preclude the enforcement of the IP right. In the case of Promusicae v. Telefónica de España,[105] Telefónica (a Spanish ISP) had refused to furnish Promusicae with a list of its customers who had infringed upon the copyright of some Promusicae members. Promusicae sought the list of infringers to enable it initiate copyright enforcement proceedings against them. The Spanish court referred the matter to the CJEU asking if such personal data (the name and contact details of the customers) could be released to Promusicae for the purpose of copyright enforcement. The CJEU held *inter alia* that the right to privacy does not preclude Member States from laying down an obligation to disclose personal data in the context of civil proceedings to ensure effective copyright protection. The court further held that member states must strike a fair balance between fundamental human rights protected by EU law and IP rights.[106] It is clear from the above that the CJEU's approach favours the creation of a balance between fundamental human rights (for the purpose of this article, the right to data protection), and IP rights. In both cases, the court found that the transmission of personal data for the purpose of IPR enforcement was justified. It is important to note that these cases are not necessarily on all fours with the facts considered in this article. This is because the two cases cited above addressed the balancing of competing interests where data access was contested for between two entities in different scenarios. This is quite distinguishable from the considerations herein because in this article, the potential competing interest can arise between the data subject and the controller/processor in respect of the data subject's data. However, the CJEU cases cited above provide some insights into the reasoning of the court on this topic.

In this article, the scenario under consideration remains how competing interests will be balanced should there be data (re)identification which results in data subjects requesting for the enforcement of their access rights. The above cited CJEU decisions can provide some sort of guidance on how the balancing of competing interests between fundamental human rights and IPR ought to be resolved. Therefore, data subject rights can still be complied with without releasing the analytical information generated from any processed personal data. Generally, AI generated data cannot be the personal data of natural persons but rather, information derived from analytics attached to such personal data. For instance, AI-generated data could be analytical data (such as most/least preferred travel habits, destinations, stops, routes, musical preferences, restaurants, etc.) from anonymised autonomous vehicle data. Should such AI-generated data be protected by some form of IPR, it is highly doubtful that the identity of the natural persons who are processed in a manner that generates such analytical information remain relevant to the generated analytical data. The test in these cases would be whether such AI-generated data are capable of reidentifying natural persons particularly taking the vast possibilities of data reidentification into consideration. Furthermore, even though the personal data of natural persons can be the subject matter of an IPR (e.g. photographs), the IPR holder must still comply with data protection law in processing the relevant personal data. Hypothetically, a photographer who has copyright in a photograph will still be expected to comply with the principles of data protection law before posting the pictures on his website. Therefore, data protection compliance cannot be discarded in the event that IPR protected AI-generated data contains or includes personal data in

102 Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications' services or of public communications' networks and amending Directive 2002/58/EC, OJ 2006 L 105.

103 Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights, [2004] OJ L 195/16 (Enforcement Directive).

104 Bonnier Audio AB case, (n 101), paras 52-54.

105 C-275/06, Productores de Música de España (Promusicae) v. Telefónica de España SAU [GC], 29 January 2008, paras. 62 – 68.

106 Ibid, paras. 65 and 68.

any way. Flowing from the decisions of the CJEU, it can be inferred that a fair balance which ensures the enforcement of the right to data protection and IP rights ought to be pursued in this scenario. Data subjects should be able to obtain their personal data that was processed during data analytics without being necessarily entitled to the anonymised/deidentified results of such analytics which reveals their identity in no way.[108] In designing internal processes for complying with data subject access requests, data controllers must ensure that privacy by design principles are adopted to prevent the inadvertent transmission of data protected by IP and related rights to competitors.[107]

The data producer is critical to the protection of AI generated data because the rights in it as well as the amelioration of any (data protection) concerns that might arise therefrom rests with the data producer. Once data re(identification) occurs, the party that originally intended to generate the data will be the most appropriate party to be accorded as the data producer. However, identifying the data producer may not always be straightforward in complex data processing activities where there are multiple parties involved in supplying the IOT used therein. In such cases, the party that convened and commissioned the suppliers of all IOT used in the processing activity would most likely be the proper data producer. Such party (the data producer) would more likely have been responsible for the investment in the data generation including the payment of all suppliers of all IOT involved therein. For instance, in the case of telerobotic surgeries, the medical institution on whose request and purpose-definition the medical devices have been supplied will be the appropriate data producer, and by extension, the data controller. Where an anonymised analysis generated from such telerobotic surgery is reidentified, the party designated as the data producer will be responsible for assuming the role of an ad-hoc data controller that will take precautionary steps to protect the rights of data subjects. The use of 'intention to process data' as the standard for designating the adhoc data controller is even more significant because such party will be designated as the data controller being the party that determines the purpose and means of processing.[109] For instance, where non-personal/anonymised data which is the subject matter of own IPR are reidentified, the data producer in his capacity as 'ad-hoc data controller' will bear certain responsibilities - mitigating potential risks, notifying the supervisory authorities and/or data subjects (if necessary), etc.

VI. Conclusion

This article has considered the possibility of data producers deploying AI systems to create AI-generated data. This article concludes that if not properly addressed, the IP protection of AI-generated data might result in infractions of the right to personal data protection thereby creating an intersection between the fields of data protection and intellectual property law. As indicated in this article, data reidentification blurs the distinction between personal and non-personal data thereby enabling the concerns that arise from the intersection between the fields of data protection and intellectual property law, which this article fully addresses. Data producers have been identified as a necessary stop-gap that can help prevent any data protection risks that might arise in the event of the data reidentification of AI-generated data protected by IPR. Furthermore, the use of AI and other data reidentification techniques have made it necessary to reconsider the effectiveness of data anonymisation as a technique for the deidentification of data, and the consequent exclusion of the applicability of data protection law. Finally, despite the conflicts identified in this article, both fields of IPR and data protection can co-exist and function side by side without hindering the free flow of business or hinder relevant rights. To achieve this, the attempts of the EU at data regulation within the union must take the fields of data protection and IP law into joint consideration to prevent the scenarios highlighted in this article particularly with the continuous proliferation of AI. This approach will also be contemporaneous to the general approach of the CJEU when it comes to balancing competing interests between the fields of data protection and IPR.

108 Even though it would appear that the Proposal for the Data Act seeks to provide transparency obligations for non-personal. See Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) COM/2022/68.

107 For further readings on privacy by design, see Lee A. Bygrave, 'Data protection by design and by default: deciphering the EU's legislative requirements' (June 20, 2017) 4 Oslo law review 2 105-120.

109 Art 4(7) GDPR