My data to develop AI: How data protection and intellectual property laws shape AI research

Amanda Costa Novaes*

Abstract: In the realm of Artificial Intelligence (AI) research and experimental development, the implications of data protection and intellectual property (IP) laws take center stage. These legal fields dictate if and how researchers can use data and databases to train AI systems. The study explores the research exception of the General Data Protection Regulation (GDPR) and IP-relevant directives, emphasizing the aimed balance between privacy, creator's rights, and innovation. While the first offers a broad approach, IP laws create a distinction based on the profit goals of the research institution. The complex interplay of these regulations raises questions about their collective impact on fostering AI innovation, particularly concerning biases and public interest.

Keywords: AI, research, intellectual property, TDM, data protection, GDPR

1. Introduction: AI development as scientific research

The Universal Declaration of Human Rights states that everyone has the right to 'share in scientific advancement and its benefits.' Such importance of individual participation in scientific endeavours is also integral to the European Union (EU). Art. 3(3) of the Treaty on the European Union highlights the EU's focus on promoting scientific and technological advances.[1] The European research and innovation policy serves as a driving force for the continent's move towards sustainability and technological progress.[2]

The Organisation for Economic Co-operation and Development (OECD) divides research into three categories: basic research, applied research, and experimental development. What unites them is not an emphasis on the non-commercial nature of the endeavour, but rather a shared commitment to acquire and advance knowledge. This commitment shall entail some characteristics such as novelty, addressing uncertainty, fostering creativity,

* Amanda Costa Novaes, Master's Student in Law, Researcher at NOVA Knowledge Centre on Intellectual Property and Sustainable Innovation (IPSI), Collaborator at Centro de Investigação sobre Direito e Sociedade (CEDIS).

[1] Consolidated version of the Treaty on European Union [2012] OJ C326/13.

[2] European Commission, 'Strategic Plan 2020-2024 DG Research and Innovation' [2020] Ref.Ares(2020)5352987 <https://commission.europa.eu/documents_en?f%5B0%5D=document_title%3AStrategic%20plan%202020-2024%20%E2%80%93%20Research%20and%20Innovation > accessed 1 September 2023.

employing a systematic approach, and providing transferable knowledge. Those are the attributes that distinguish research from purely commercial product development.[3]

Within this context, artificial intelligence (AI) emerges as a valuable instrument for pushing forward scientific realms. Research done by software engineers on these systems plays a pivotal role that extends beyond just shaping them – it also involves ensuring their potential societal advantages. To illustrate, the European Commission's strategic plan spanning from 2020 to 2024 underscores how AI is anticipated to significantly influence EU crime and cybersecurity policies.[4]

To provide a concrete example, by employing data-driven deep learning techniques, facial recognition systems have the capacity to determine whether two images depict the same individual, relying on the match of their facial characteristics. This AI technology serves diverse purposes, spanning from locating missing children to identifying people using counterfeit documents.

However, increased limitation on the data available to train these systems directly corresponds to a decline in the technology's accuracy.[5] For instance, due to a lack of diversity in the training data, researchers found that facial recognition systems tend to be less accurate on people of colour, especially on women,[6] perpetuating bias towards these minorities. Alarmingly, these inaccuracies have already resulted in significant harm in the US, for example. [7]

The EU, on the other hand, focuses on fostering trustworthy AI systems.[8] Situations where law enforcement authorities use facial recognition systems are specifically affected by the guarantees of Art. 22 of the General Data Protection Regulation (GDPR),[9] which

---

[3] OECD, *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities,* (OECD publishing 2015), 50-52

[4] European Commission, 'Strategic Plan 2020-2024 DG Research and Innovation', 27.

[5] Michele Merler, 'Diversity in Faces' [2019] 1901.10436 arXiv, 2 <http://arxiv.org/abs/1901.10436> accessed 1 September 2023.

[6] Patrick J Grother, Mei L Ngan and Kayee K Hanaoka, 'Face Recognition Vendor Test Part 3: Demographic Effects' [2019] 8280 Interagency/Internal Report NISTIR <https://www.nist.gov/publications/face-recognition-vendor-test-part-3-demographic-effects> accessed 1 September 2023.

[7] Khari Johnson, 'How Wrongful Arrests Based on AI Derailed 3 Men's Lives' (Wired, 7 March 2022) <https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/> accessed 1 September 2023.

[8] Commission, 'Building Trust in Human Centric Artificial Intelligence' COM (2019), 168.

[9] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the

prohibits automated decision-making and profiling. Also, the current Proposal for the AI Act has specific rules in Art. 5(1)(d), (2), (3), and (4),[10] imposing serious limitations on this use to avoid damaging innocent citizens.

Nonetheless, beyond simply addressing the potential adverse outcomes of AI, does the EU's legislative framework promote trust in AI by adequately fostering research to develop the technology? Recognizing the vital role of training datasets and the influence of GDPR and IP laws in safeguarding these content and/or databases, this paper seeks to explore the influence of such research-related legislation on the experimental development of AI systems.

To achieve this objective, the forthcoming chapter will examine the implications of both data protection and intellectual property regulations on the research of AI systems. Subsequently, a more detailed assessment will be devoted to the distinct framework established for researchers, starting with the GDPR's provisions, then delving into the research and Text and Data Mining (TDM) exceptions within IP law. The analysis aims to evaluate whether these legal provisions effectively facilitate scientific innovation in the field of AI, as originally intended, or if they hinder it instead.

2. How IP and data protection rules affect AI research

While a well-balanced and comprehensive dataset is indispensable for optimizing AI performance, it is vital to understand that AI researchers cannot freely use any available data or datasets for training such technologies. Take facial recognition systems, for instance. The datasets consist of a multitude of facial measurements extracted from images of individuals gathered by the system. Yet, due to their specific format and content, these images and databases will raise issues both on data protection and intellectual property dimensions.

a. AI data vs. Users' privacy

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 [2016] OJ L 119.

[10] European Commission, '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', COM (2021) 206 Final.

Initially, the GDPR has rules to ensure the privacy of the users' data used to train AI systems. Art. 4(1) of the GDPR defines personal data as 'any information relating to an identified or identifiable natural person', including 'factors specific to the physical, […] cultural or social identity of that natural person'. Given that facial measurements have the capability to uniquely identify an individual, they naturally fall under the scope of personal data. These biometric measurements even qualify as sensitive data, a special category highlighted in Art. 9 of the GDPR.

As a result, to process such personal data in AI systems, a legal basis is needed. One possibility would be to acquire the specific consent of the data subjects to use their images and input their facial measures in the training system.[11] However, to collect valid consent from all the data subjects of the multitude of data necessary to develop an AI system would be unfeasible.[12]

Nevertheless, the GDPR encompasses alternative legal bases for processing personal data that can, in some scenarios, be used. For instance, the public interest present in Art. 6(1)(e) can hold relevance, particularly when mitigating bias in facial recognition systems, for example. Moreover, according to Art. 9(2)(e) of the GDPR, sensitive data can be processed when the data subject has manifestly made it public, an occurrence that could be deemed to happen when such content is publicly shared online.[13]

Additionally, as will be further explained, the scope of the research exception can play a significant role in allowing AI system developments specifically for the public good. Art.9(2)(j) of the GDPR introduces a legal basis grounded in the public interest for processing sensitive data within the context of research. This provision, however, is contingent on the implementation of appropriate safeguards, as outlined in Art. 89 of the GDPR.

b. AI data/databases vs. Creators' rights

[11] Legal basis for processing explicitly on Art.6(1)(a) of the GDPR and Art. 9(1)(a) of the GDPR for sensitive data.
[12] For more, see Marvin van Bekkum, and Frederik Zuiderveen Borgesius 'Using sensitive data to prevent discrimination by artificial intelligence: Does the GDPR need a new exception?' [2023] 48 Computer law & Security Review 1, 6.
[13] European Union Agency for Fundamental Rights, European Court of Human Rights, European Data Protection Supervisor, *Handbook on European Data Protection Law* (Publications Office of the European Union 2018), 162.

In the realm of IP law, both the content of these datasets and the databases themselves are safeguarded. With regards to the data within the training dataset, copyright law grants authors exclusive rights over such content if, overall, it results from the author's free and creative choices, bearing his/her personal mark. Notably, this protection extends even to short literary extracts and the so-called 'point-and-click photos'.[14]

In this scenario, the creation needs a specific license from the right holder to be extracted and copied into the database, which normally requires the payment of a fee.[15] Considering the extensive scope of AI systems datasets – which can encompass a diverse range of data types such as literary excerpts, digital artwork, images, and audio – copyright protection will inherently extend to a substantial portion of these elements. For instance, an authorization of the photographer is needed to use his/her pictures of individuals to train facial recognition systems.

To subsist, copyright needs original content. However, databases composed of raw data (content with no originality), can still be copyright-protected. A protection that, in the EU, lasts 70 years after the author's death. In those cases, the arrangement of the database needs to bear the author's creativity, being his/her own intellectual creation, as stated in Art. 3 of the Database Directive.[16]

Even when that is not the case, according to Art. 7 of the Database Directive, a database composed of independent data, individually accessible and systematically arranged, can also be protected by a shorter-term *sui generis* right. In this case, those who made a considerable investment can prevent the extraction and/or re-utilization of the whole or a substantial part of the database for 15 years after the date of completion, as stated in Art. 10 of that directive.

This protection and the consequent need for licensing agreements to use the data and datasets necessary to train AI systems generates an adverse net effect on innovation.

---

[14] Justine Pila and Paul Torremans, *European Intellectual Property Law* (2nd edn, OUP 2019), 252.
[15] Directorate-General for Internal Policies of the Union, 'The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market: Technical Aspects' [2018] (Publications Office) <https://data.europa.eu/doi/10.2861/480649> accessed 1 September 2023.
[16] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77 (Database Directive).

Empirical studies demonstrates that stricter copyright rules hinder 'the wide adoption of novel ways to build on copyright works and generate derivative works'.[17]

In sum, both the restrictions provided by the GDPR and IP law can generate the need for safeguards and legal authorizations that are expensive and time-consuming for the volume of data necessary to train a more accurate system. As a result, AI researchers are prompted to prioritize the use of 'easily available, legally low-risk works as training data, even when those data are demonstrably biased'.[18] For instance, openly licensed photos or images publicly available online are often favoured due to their reduced risk of triggering infringement actions, although they tend to accentuate problems with the efficiency of the AI outcome.

The fear of being held accountable for the unauthorized use of data/databases in the training dataset can also discourage its disclosure.[19] Consequently, the non-accessibility of such datasets causes *de facto* a restriction to bias mitigation techniques. First, it diminishes the transparency of the AI system. Moreover, it helps prevent reverse engineering processes that could provide for algorithmic accountability.[20] Both transparency and accountability are important principles of the EU legislative approach towards AI.[21]

While the legislative process for the AI Act has been underway, it is worth noting that a provision under discussion about copyright law has been leaked. Art. 28b–5a would potentially require companies involved in AI experimental development to disclose any copyrighted content used during the model's training.[22] However, it is important to emphasize that since this provision has not officially become part of the discussed legislation, this paper will not delve into its implications.

Due to the significant costs associated with legally obtaining the substantial volume of data required to thoroughly train an AI system, the outcome is the centralization of

---

[17] Christian Handke, Lucie Guibault and Joan-Josep Vallbé, 'Copyright's Impact on Data Mining in Academic Research' (2021) 42 Managerial and Decision Economics 1999, 2012.
[18] Amanda Levendowski, 'How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem' (2018) 93 Washington Law Review 579, 597.
[19] Ibid, 599.
[20] Ibid, 602.
[21] Commission, 'Building Trust in Human Centric Artificial Intelligence', 5-6.
[22] João Pedro Quintais, 'Generative AI, Copyright and the AI Act' (Kluwer Copyright Blog, 9 May 2023) <https://copyrightblog.kluweriplaw.com/2023/05/09/generative-ai-copyright-and-the-ai-act/> accessed 1 September 2023.

innovation within a handful of dominant global technology corporations. These companies essentially monopolize the authority to negotiate and amass the extensive datasets needed for effective AI training, testing, and experimental development. On the other hand, research affiliated with public institutions often lacks the financial resources to pursue innovations similar to those accessible to private entities.[23]

In response to this challenge, both IP and data protection rules establish a distinct framework for researchers. However, while in the IP realm there is a distinction between researchers with and without a profit goal, data protection rules focus on the greater public good rather than being reliant on specific economic gains.[24]

Consequently, certain questions remain: How do these provisions align in a practical scenario of AI research? Furthermore, which approach more effectively tackles and remedies the current discrepancy when it comes to the research and experimental development of these systems?

3. Research exception for AI innovation

a. The GDPR research exception

In addition to the aforementioned general regulations within the GDPR, a distinctive framework exists for researchers. This framework is designed to encompass studies that have a scientific method and, according to Recital 175 of the GDPR, 'improve the quality of life for a number of people and improve the efficiency of social services'.[25] The European Data Protection Supervisor (EDPS) issued a preliminary opinion on research affirming that the 'respect for personal data is wholly compatible with responsible research'.[26]

---

[23] Levendowski, 'How Copyright Law Can Fix Artificial Intelligence's Implicit Bias Problem', 609.
[24] European Data Protection Supervisor, 'Preliminary Opinion on Data Protection and Scientific Research' (30 March 2023) <https://edps.europa.eu/data-protection/our- work/publications/opinions/preliminary-opinion-data-protection-and-scientific_en> accessed 1 September 2023.
[25] See also the Opinion of AG Mancini in C-234/83 *Gesamthochschule Duisburg v Hauptzollamt München-Mitte* [1985] when interpreting scientific activities as 'including activities carried on by a public or private establishment engaged in education or research for the purpose of further the acquisition, development, exposition or dissemination of scientific knowledge (...)'.
[26] European Data Protection Supervisor, 'Preliminary Opinion on Data Protection and Scientific Research'

This preliminary opinion from the EDPS also clarified how academia and the commercial sector can intertwine, due to the funding large technology companies provide for vast amounts of academic research and their partnerships with research institutions and public bodies.[27] Notably, Recital 159 clarifies that technological advancements, including those privately funded, can be considered as falling within this research concept. As a result, research endeavours aimed at developing AI systems with reduced bias could align with this description.

Art. 89 of the GDPR details the processing of data under a research regime. The primary requirement is that the study incorporates proper safeguards for data protection, ensuring the principle of minimisation of data, meaning that only the necessary data for the research purpose shall be collected.[28] An exemplified measure to achieve it is the pseudonymisation of data.

In this regard, the GDPR establishes a pathway that facilitates the pursuit of research endeavours aligned with the broader public interest, as long as required safeguards, including data pseudonymization, are diligently implemented.[29] The GDPR's overarching scope, devoid of dichotomies based on for-profit or not-for-profit distinctions, underscores a unified approach to fostering research.

In this context, researchers must remain vigilant in adhering to stringent data protection protocols, ensuring that data utilization remains minimal and safeguarded through pseudonymization techniques. However, most importantly, this framework does not preclude researchers from advancing their investigative pursuits.

This approach engenders a delicate balance between safeguarding the privacy of data subjects and catalysing advancements in scientific research across diverse domains. When specifically considering the realm of AI, where accurate results demand substantial volumes of data, such possibility provided to researchers through regulatory provisions assumes pivotal significance. This becomes a cornerstone in realizing the ambitious

(30 March 2023) <https://edps.europa.eu/data-protection/our- work/publications/opinions/preliminary-opinion-data-protection-and-scientific_en> accessed 1 September 2023.

[27] Ibid, 7.

[28] GDPR, Art. 5(1)(c) GDPR.

[29] For more, see Ciara Staunton, Santa Slokenberga, and Deborah Mascalzoni, 'The GDPR and the research exemption: considerations on the necessary safeguards for research biobanks' [2019] 27 European Journal of Human Genetics 1159, 1165.

innovation objectives set forth by the EU within the research landscape. In essence, this harmonized approach can align the protection of users' privacy with scientific progress and innovation across multifaceted horizons.

b. The IP Law research exception regime

IP law provisions, on the other hand, will vary depending on the for-profit or not-for-profit goal of the research activity. When developing AI technology, the main legislations on the use of IP-protected works for research are: (i) the research exception in Art. 5(3)(a) InfoSoc Directive[30] and Art. 6(2)(b) and 9(b) Database Directive; and (ii) the Text and Data Mining (TDM) exception in Art. 3 and 4 of the Directive on Copyright in the Digital Single Market (CDSM Directive).[31]

These provisions collectively furnish a framework that aims to enable researchers to integrate copyrighted content into their investigative undertakings. The overarching intent behind these provisions is to strike a balance between the imperative of upholding IP rights and expanding the frontiers of research within the realm of AI technology.

Regarding the research exception of the InfoSoc and Database Directives, they apply for a wide-ranging definition of research,[32] yet there is a non-uniform approach because member states are free to adopt it or not and to determine its scope.[33] This creates confusion, hampering cross-border and collaborative research. Hence, without legal certainty to use research data, researchers cannot maximize their innovative potential.[34]

Moreover, even though these exceptions do not require any specific organizational structure and means of funding of the establishment concerned, there is the need for the research to be carried out on a purely non-commercial nature. Notably, even if the

---

[30] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society [2001] OJ L 167 (InfoSoc).
[31] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC [2019] OJ L 130 (CDSM Directive).
[32] InfoSoc Directive, Recital 42 and Database Directive, Recitals 50 and 51.
[33] Madeleine de Cock Buning, Lucky Belder, and Roeland de Bruin, 'Research Exceptions in EU Copyright Law' (2012) 20 (4) *European Review of Private Law* (Kluwer Law and Business) 933, 957.
[34] Linda Kuschel and Jasmin Dolling, 'Access to Research Data and EU Copyright Law' (2022) 13 JIPITEC 247, 247.

research is conducted within public universities or institutions, if an economic objective underpins the investigation, the exception will not apply.[35]

This implies that individuals or entities engaged in commercial research and experimental development activities, such as enhancing the efficacy of facial recognition systems for subsequent sale to law enforcement agencies, cannot rely on such a provision. This limitation significantly curtails the scope of application of the InfoSoc research exception within the domain of AI development methodologies, since, as well captured in the GDPR, the academia and commercial sector tend to be intertwined.

By its turn, the TDM exception, delineated in the CDSM Directive, offers a distinct regime in support of this endeavour. TDM is a specialized technique designed to uncover significant patterns within extensive datasets. This technique holds paramount significance in the realm of AI systems, as their architecture and functionality are inherently reliant on these datasets to yield accurate outcomes.

The underlying principle of the TDM exception is rooted in the notion that this technique neither infringes upon the expressive elements of a work nor reconstructs a substantial portion of a safeguarded database. [36] This mechanism is crafted to facilitate the extraction of valuable insights without compromising the foundational integrity of copyrighted content or comprehensive databases.

In this scenario, the CDSM Directive offers a fairly uniform approach by setting a minimum mandatory standard for all member states. However, it establishes two distinct regimes essentially based on the profit motives of research initiatives.

Art. 3 of this Directive is applicable when scientific activities are conducted by research organizations and cultural heritage institutions. This applies specifically when these activities are carried out only on a non-profit basis or align with a public interest mission recognized by a Member State. In such cases, there cannot be an entity with preferential access to results or exercising substantial influence over the process.[37]

---

[35] InfoSoc Directive, Recital 42 and Database Directive, Recital 50.
[36] Rossana Ducato and Alain Strowel, 'Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out' [2021] *European Intellectual Property Review* 43(5) 322, 323.
[37] CDSM Directive, Art. 2(1).

Art. 4, conversely, applies to the other entities. Essentially, this implies that any institution engaging in research with profit motives or having an entity with preferential access to research outcomes falls under this regime. In the present research landscape, where financial compensation plays a pivotal role in sustaining such initiatives, it is reasonable to anticipate that this second regime will be more commonly employed when considering AI experimental development.

In any case, whoever is carrying out the TDM must possess previous lawful access to the content. This pertains to cases like using collections of images or texts, where prior payment is essential for accessing or using the content. Additionally, as outlined in the Recital 17 CDSM Directive, member states shall not provide compensation to the rightholders for TDM activities. Consequently, when an AI system is being developed, the researcher entity must cover any subscription fees required for legitimate content access.

Conversely, if the rightholders of copyrighted content want to impede its use for TDM in research, there is nothing prohibiting them from abruptly increasing subscription fees for the research entities developing AI systems, for example.[38] For instance, this scenario could materialize even to the detriment of not-for-profit research institutions that are dedicated to advancing technology for the collective welfare, such as mitigating bias techniques on facial recognition systems.

These two regimes of the CDSM Directive, regarding for and not-for-profit research, will also carry key distinct consequences. The first, and one that is notably clear, lies in the fact that only not-for-profit research can maintain copies of the works. The objective is to facilitate peer review and thorough validation of the results, yet security measures must be in place for that.

Another critical distinction pertains to the ability of copyright holders to prevent TDM through contractual or technical means, outlined in Art. 3(3) and 4(3) CDSM Directive. In the context of non-profit research, rightholders possess the option to employ safeguards to ensure the security and integrity of their works. However, these measures cannot be used to hinder TDM activities, either contractually or technologically.

[38] Ducato, 'Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out', 19.

Conversely, in the context of for-profit research, the rightholder can impede TDM using contractual or technical methods, including those that are machine-readable, even if the content is freely accessible online. This distinction introduces a significant shift in the landscape of AI technology experimental development. It clarifies uncertainties surrounding the necessity of authorization for using seemingly 'free' online content, while also empowering copyright holders to prevent its use.

However, the automated nature of these interactions can lead to complexities in practical application. Technological measures cannot inherently discern whether a research institution pursuing profit goals or not is attempting to engage in TDM. As a result, not-for-profit researchers can be, *de facto*, inadvertently restricted from using the content for AI experimental development.[39]

In summary, all these provisions are relevant when assessing a practical scenario involving research aimed at addressing hidden biases within AI systems. In such a scenario, the following elements come into play.

First, if the research institution is profit-oriented, even if the objective is to construct an AI system with a comprehensive dataset to counteract biases in facial recognition, only the regulatory framework of Art. 4 in the CDSM Directive is applicable. Under this regime, the right holder effectively possesses the means to hinder TDM activities through contractual or technological methods.

On the other hand, if the research is driven by non-profit motives, and devoid of any entity enjoying preferential access to research outcomes, it can benefit from the exceptions granted by the InfoSoc, Database, and Art. 3 of the CDSM Directives.

In this case, with proper security measures, the researcher could retain the results for peer review. However, two aspects can lead to a *de facto* restriction for research activities carried out by these entities. First, they must possess lawful prior access to the content and nothing impedes rightholders from substantially raising the fees specifically for them. Also, they can be inadvertently affected by the automated nature of technological measures in place to prohibit TDM activities of for-profit research.

[39] For more, see João Pedro Quintais, 'The New Copyright in the Digital Single Market Directive: A Critical Look ' (2020) 1 European Intellectual Property Review 28.

In all scenarios, IP law provisions protecting the content essential for training AI systems can potentially hinder the progress of this technology for the collective benefit, as seen in the context of addressing biases in facial recognition systems. On IP provisions, there is no uniform and encompassing approach, such as the one existing in the GDPR, allowing for research for the public good as long as appropriate safeguards are in place.

Hence, in the realm of researching and developing AI systems, it becomes imperative to meticulously weigh the implications of both data protection and IP legislation. This comprehensive approach is essential to ensure that the research journey effectively upholds the principles of data privacy and respects the IP rights of content creators. Nonetheless, the divergent approaches taken by these legislations can present significant challenges to the advancement of AI and innovation, potentially obstructing progress even when the goal is to benefit the broader public.

4. Conclusion: is the legislative framework promoting or hindering innovation?

The right to research materialized in the EU legal system is a pillar present in many strategies for the forthcoming years. However, when considering AI research and innovation, the intersection of data protection and IP rules significantly influences this trajectory. Striking the delicate equilibrium between safeguarding users' privacy and creators' rights becomes crucial in this intricate landscape, potentially influencing the broader societal benefits derived from the research.

When examining the GDPR's impact on AI research, it becomes evident that the regulation's provisions create a nuanced environment. The GDPR's recognition of the public interest in research aimed at enhancing societal well-being presents a valuable opportunity for AI experimental development that serves the greater good. By stipulating data protection safeguards such as pseudonymisation and embracing a broad concept of research, the GDPR can uniformly accommodate a range of research activities, including those directed at addressing AI bias, without the possibility of prohibiting them. This balancing between safeguarding data subjects and propelling AI advancement is pivotal for the evolution of responsible and beneficial AI technologies.

Shifting the focus to the domain of IP law, the landscape for AI research takes on a distinctive complexion. While the GDPR primarily focuses on privacy concerns, IP law operates on a different axis, seeking to balance the rights of content creators and promote creativity. Unlike the GDPR's unified approach regarding public good goals, IP law draws a significant differentiation between for-profit and not-for-profit research endeavours.

The research exceptions within the InfoSoc and Database Directives offer a broad scope, yet they apply solely to non-commercial research and provide for a fragmented regulatory approach. The TDM exception, on the other hand, carries mandatory elements but also segregates research into profit and non-profit categories. However, these provisions fail to comprehensively address the imbalance regarding access to IP-protected content.

Notably, the IP landscape provides for a possibility that the GDPR does not encompass: AI research, even when done for the public good, can, *de facto,* be prevented by the rightholders. This can happen by abruptly raising subscription fees for the content or employing technological protection measures that do not distinguish the profit goal or not of the institution carrying out TDM activities.

Consequently, the IP legal regime promotes an environment where it still is very expensive and time-consuming to acquire all the necessary licenses to develop an AI system, even for the public good. Meanwhile, researchers are likely to resort to openly available content and databases that tend to be biased or do not disclose the content used in the AI system. Consequently, it can disincentivize the trustworthy and transparent AI development aimed at the EU.

As demonstrated, the intersection of data protection and intellectual property within the domain of AI research presents a significant challenge, yet also an opportunity for innovation. The GDPR provides a crucial framework for safeguarding personal data and upholding the public good, permitting research to thrive while maintaining privacy.

Thus, a research exception within IP law that encompasses both copyright and the databases *sui generis* right – providing that AI researchers have access to data independently of their profit goals, as long as they are developing research for the public good – is essential to safeguard the right to research and foster such fundamental activity for our society. With that, an intricate balance between IP rightholders protection to foster

creativity and the subject's privacy could be achieved without undermining AI research and innovation.