Check for updates

ESSAY

Voicing challenges: GDPR and AI research

[version 1; peer review: 2 approved with reservations]

Katherine Quezada-Tavarez, Lidia Dutkiewicz, Noémie Krack

Centre for IT & IP Law – imec, KU Leuven, Leuven, 3000, Belgium

v1 **First published:** 23 Nov 2022, **2**:126
https://doi.org/10.12688/openreseurope.15145.1
Latest published: 23 Nov 2022, **2**:126
https://doi.org/10.12688/openreseurope.15145.1

Abstract

EU data protection rules could be difficult for researchers to navigate, particularly when processing massive datasets containing personal data for Artificial Intelligence (AI) developments. This article examines how data protection intersects with AI research to elucidate the issues arising from the use of large-scale databases containing personal data to train, test and validate AI systems. The key objectives of this work are to (1) scrutinise the data protection requirements and limits for the processing of personal data in AI research, (2) reflect on possible complications regarding data quality requirements for trustworthy AI and General Data Protection Regulation (GDPR) compliance, and (3) present possible ways forward to reconcile GDPR requirements and AI research. While reviewing and mapping relevant provisions and guidance, we identify data protection challenges posed by the use of massive databases containing personal data for AI research. The findings suggest that, while the legal regime for research under the GDPR resolves some of the challenges identified, others, such as legal basis for processing and processing of special categories of data, remain unaddressed. We argue that the nature of these complications will make it difficult for EU researchers to advance in trustworthy AI efforts. The analysis concludes by suggesting possible ways to tackle the remaining issues.

Keywords

data protection, GDPR, scientific research purposes, AI research, AI Act

Open Peer Review

Approval Status ? ?

	1	2
version 1 23 Nov 2022	? view	? view

1. **Oya Deniz Beyan**, University of Cologne, Cologne, Germany

2. **Dimitra Georgiou**, University of Piraeus, Pireas, Greece
Hellenic Airforce Academy, Athens, Greece

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the Research on Research gateway.



This article is included in the Horizon 2020 gateway.



This article is included in the Artificial Intelligence and the Social Sciences and Humanities collection.



This article is included in the Research Ethics and Integrity collection.

Corresponding author: Lidia Dutkiewicz (lidia.dutkiewicz@kuleuven.be)

Author roles: Quezada-Tavarez K: Conceptualization, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Dutkiewicz L**: Conceptualization, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Krack N**: Conceptualization, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This research was financially supported by the European Union's Horizon 2020 research and innovation programme under the grant agreement No [883297](Deep AR Law Enforcement Ecosystem [DARLENE]) as well as under the grant agreement No [951911](A European Excellence Centre for Media, Society and Democracy[AI4Media]).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2022 Quezada-Tavarez K *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Quezada-Tavarez K, Dutkiewicz L and Krack N. **Voicing challenges: GDPR and AI research [version 1; peer review: 2 approved with reservations]** Open Research Europe 2022, **2**:126 https://doi.org/10.12688/openreseurope.15145.1

First published: 23 Nov 2022, **2**:126 https://doi.org/10.12688/openreseurope.15145.1

Plain language summary

EU data protection rules could be difficult for researchers to apply, particularly when processing (using) massive datasets containing personal data for Artificial Intelligence (AI) developments. This article examines how data protection intersects with AI research. The key objectives of this work are to (1) look at the data protection requirements and limits for the processing of personal data in AI research, (2) reflect on possible complications regarding data quality requirements for trustworthy AI, and (3) present possible ways forward.

Introduction

In recent years, aware of the importance of not lagging behind in the global competition on Artificial Intelligence (AI), EU leaders have made investing in AI a top priority,[1] with AI research and innovation initiatives playing a pivotal role in the EU policy efforts.[2] The European approach for AI focuses on supporting ethical and trustworthy AI,[3] making a significant portion of the available resources to foster the development and uptake of AI through funding programmes such as Horizon 2020 and Horizon Europe. The framework of these programmes make it that, for EU researchers, 'ethics is an integral part of [their] research from beginning to end',[4] with project beneficiaries also being strongly encouraged to contribute to the trustworthy AI efforts.

By nature, AI research and development efforts, particularly those involving Machine Learning (ML) techniques, require significant amounts of training, testing and validation data.[5] The quality of the data also matters for the downstream effects, since the characteristics of the data will translate into outputs of the system at deployments. The higher the quality of the data used, the better the system outputs (at least in theory). Hence, the data needed in AI research are also associated with the 'trustworthy' nature of EU-made AI systems, *i.e.* systems that respect human rights, adhere to ethical guidelines and promote human agency.[6] It is therefore no surprise that the 2021 proposal for an EU AI regulation[7] includes data quality criteria for certain systems.

However, the status of AI research in the General Data Protection Regulation (GDPR)[8] and existing data protection guidance is not clear. AI researchers mostly rely on large-scale databases containing massive volumes of (personal) data to 'instruct' their systems and ensure performance, and some of the personal data therein may even fall under special categories of data, particularly when engaging in debiasing activities.[9] Yet, the efforts to fulfil data quantity and quality requirements for AI research purposes may be at odds with the attempts to comply with certain data protection conditions, tensions that the current legal regime for research does not fully address, as shown in this article. Therefore, the shortcomings regarding the access to, (re)use and sharing of data by genuine researchers must be tackled.

Recent legal scholarship addresses questions about data protection in research[10] and the legal issues associated with AI,[11] while little attention has been paid to AI research specifically. This paper sheds light on the under-discussed issue of challenges and tensions between the processing of large-scale datasets for AI research and compliance with the GDPR. What data is needed to embed values and achieve 'trustworthiness' in EU-made AI systems? Is that processing allowed under EU data protection law? How to reconcile data protection requirements with legitimate data processing for genuine research

[1] European Commission, 'Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee of Regions - Fostering a European Approach to Artificial Intelligence (COM(2021) 205 Final)' (European Commission 2021) <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2021:205:FIN> accessed 21 April 2021.

[2] European Commission, 'The Pivotal Role of Research and Innovation in Artificial Intelligence Policy (Factsheet)' (European Commission 2020) <https://ec.europa.eu/info/sites/default/files/research_and_innovation/research_by_area/documents/ec_rtd_ai-pivotal-role.pdf> accessed 20 May 2020.

[3] European Commission, 'White Paper on Artificial Intelligence: A European Approach to Excellence and Trust (COM(2020) 65, Final)' (2020) 25 <https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificialintelligence-feb2020_en.pdf> accessed 3 May 2021.

[4] European Commission, 'Horizon 2020 Online Manual - Ethics Section' (n.d.) <https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/ethics_en.htm> accessed 21 November 2021.

[5] Bertin Martens, 'The Importance of Data Access Regimes for Artificial Intelligence and Machine Learning' [2018] SSRN Electronic Journal <https://www.ssrn.com/abstract=3357652> accessed 31 October 2022.

[6] High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (2019) <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> accessed 9 February 2021.

[7] 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)' (European Commission 2021) <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206> accessed 21 April 2021.

[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), OJ 2016 L 119/1.

[9] Indre Žliobaite and Bart Custers, 'Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models' (2016) 24 Artificial Intelligence and Law 183.

[10] See *e.g.*, Rossana Ducato, 'Data Protection, Scientific Research, and the Role of Information' (2020) 37 Computer Law & Security Review 105412; Catherine Jasserand, 'Massive Facial Databases and the GDPR: The New Data Protection Rules Applicable to Research', *Data Protection and Privacy: the Internet of Bodies* (Hart Publishing/Bloomsbury Publishing Plc 2018); Katrin Weller and Katharina E Kinder-Kurlanda, 'A Manifesto for Data Sharing in Social Media Research' (2016).

[11] See *e.g.*, Woodrow Barfield and Ugo Pagallo, *Research Handbook on the Law of Artificial Intelligence* (Edward Elgar Publishing 2018); S Barocas and A Selbst, 'Big Data's Disparate Impact' (2016) 104 California Law Review; Jan De Bruyne and Cedric Vanleenhove (eds), *Artificial Intelligence and the Law* (Intersentia 2021).

(collection, storage, sharing and (re)use of massive volumes of data, including sensitive data)? Does the GDPR threaten to hinder the progress in AI research? These are ongoing challenges explored in this paper.

The analysis is structured as follows. The section 'AI training, testing and validation data' refers to the importance of data quantity and quality for AI development purposes, and discusses the data quality criteria in the proposed EU AI Act. Following this, 'Legal and ethical challenges and complications of personal data use in AI research' examines how the GDPR applies to AI research, referring to the allowances, limits and requirements for data processing in research, and explores key issues regarding the incorporation of massive amounts of (sometimes sensitive) personal data in AI developments. Considering the findings of this analysis, 'Ways forward' suggests some measures to tackle unaddressed challenges and to find the proper balance between privacy and innovation. The final section contains the concluding remarks.

AI training, testing and validation data

Quantity and quality of data are crucial factors for the research and development of AI. In this section, we refer to the importance of quality of data for AI research given the impact on system outcomes and downstream effects at deployments. The significance of this matter is illustrated by the inclusion of quality criteria in the proposal for an EU AI Act presented by the European Commission in April 2021, also referred to in the analysis.

The importance of data quality for AI

AI/ML training, testing and validation require large volumes of data, which also needs to meet certain quality criteria to ensure model performance. This is evidenced in current literature on the quality and representation issues with training, testing and validation data, especially those in publicly available datasets and databases.[12] The problem is that AI systems built on insufficient or flawed data are prone to have significant downstream effects. In particular, ML systems that are trained on limited datasets, which only represent some part of the society, will reproduce and even exacerbate biased and partial perspectives.[13] This refers to the 'garbage in, garbage out' principle, a phenomenon that can result in flawed outcomes and decisions based in incorrect conclusions, and eventually even lead to human rights violations if left unaddressed.[14] There is also growing interest in ways to detect and mitigate data quality risks in AI research. For instance, a study on facial recognition technology showed that access to more personal data about people from different origins, races, and ethnicities may increase accuracy of the system, decrease systemic bias, and improve the ability for developers to prove this to regulators.[15]

Data quality in AI and ML is not only a recurrent subject in research but also forms part of AI policy discussions and documents. For instance, the European Council has acknowledged that 'high-quality data are essential for the development of [AI]'.[16] Also, the Ethics Guidelines for Trustworthy AI[17] specifically refer to data governance as one of the requirements for AI systems, highlighting the impact of biases in training datasets. Similarly, the White Paper on AI by the European Commission essentially considers the availability of data as a 'key driver' for AI initiatives, enabling the development and improvement of products and services in support of better decision-making.[18] It can therefore be said that data quantity and quality play a crucial role in the attempts to develop trustworthy and unbiased AI systems.[19]

Understanding data quality issues can help address potential problems, including the risks of bias and discrimination. Against this backdrop, the EU AI Act proposal establishes specific quality criteria for training, validation and testing data, further explored in the following.

Quality criteria in the EU AI Act proposal

On 21 April 2021, the European Commission presented a proposal for an AI Act (also known as the AI Regulation), laying down harmonised rules on AI.[20] This heavily anticipated

[12] Inioluwa Deborah Raji and others, 'Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing' (2020).

[13] Osonde A Osoba and William Welser IV, 'An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence' (Rand Corporation 2017) <https://www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1744/RAND_RR1744.pdf> accessed 25 July 2021.

[14] Rashida Richardson, Jason Schultz and Kate Crawford, 'Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice' (2019) 94 New York University Law Review 42.

[15] John Roach, 'Microsoft Improves Facial Recognition Technology to Perform Well across All Skin Tones, Genders' (*The AI Blog*, 26 June 2018) <https://blogs.microsoft.com/ai/gender-skin-tone-facial-recognition-improvement/.]>.

[16] 'European Council Meeting of 28 June 2018 – Conclusions (EUCO 9/18)' (2018) <https://www.consilium.europa.eu/media/35936/28-euco-final-conclusions-en.pdf> accessed 15 July 2021.

[17] High-Level Expert Group on Artificial Intelligence (n 6).

[18] 'White Paper on Artificial Intelligence: A European Approach to Excellence and Trust (COM(2020) 65, Final)' (n 3) 2–3.

[19] Still, we cannot overlook the fact that, contrary to popular belief, more data does not per se guarantee better AI models. A recent research trend suggests that blindly adding a vast amount of data may lead to diminishing returns in model performance. See *e.g.* Joel Hestness and others, 'Deep Learning Scaling Is Predictable, Empirically' [2017] arXiv preprint <http://arxiv.org/abs/1712.00409> accessed 10 February 2022; Chen Sun and others, 'Revisiting Unreasonable Effectiveness of Data in Deep Learning Era' (2017) <https://openaccess.thecvf.com/content_iccv_2017/html/Sun_Revisiting_Unreasonable_Effectiveness_ICCV_2017_paper.html> accessed 10 February 2022; Divya Shanmugam and others, 'Learning to Limit Data Collection via Scaling Laws: Data Minimization Compliance in Practice' [2021] arXiv preprint <http://arxiv.org/abs/2107.08096> accessed 10 February 2022; In addition, some algorithmic techniques now permit to reduce the amount of data processed while improving the accuracy such as outlier detection, feature selection, learning framework with data collection stopping points (see Michele Finck and Asia J Biega, 'Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems' [2021] 2021 Technology and Regulation 44). Therefore, the traditional assumption that more data improves the outputs is nuanced. What one needs is not necessarily more data, but better data. Clearly, a massive amount of inaccurate data will produce inaccurate results.

[20] 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)' (n 7).

document represents a major regulatory development of our time since it is the first ever attempt worldwide to regulate in that area. The AI Act aims to foster a coherent and harmonised digital single market, in a way that AI systems or products powered by AI technology can circulate in the EU market according to common rules, ensuring the safeguard of European fundamental values.

The proposal establishes requirements and obligations for the development, placement on the market and use of AI systems. The initial proposal by the European Commission is framed following a risk-based approach, laying down specific obligations that apply to different categories of AI systems based on their characteristics and risks. The draft regulation classifies AI systems into four categories based on whether they pose: (i) unacceptable risks, entailing the outright or qualified prohibition of such systems; (ii) high risks, which are subject to ex-ante and ex-post requirements, as well as a supervision and enforcement regime; (iii) limited risks, subjects to specific transparency obligations; or (iv) minimal risks, which are free to circulate in the EU market.[21] As becomes apparent, the vast majority of requirements concern high risk systems, including data quality requirements.

Before diving into the data quality criteria in the EU AI Act proposal, it is first necessary to assess to what extent the proposed regulation would apply to research. The initial proposal by the European Commission appears to indicate that scientific research would not fall under the scope of that law given the non-commercial nature of these activities, thus not belonging to the market context. The so-called 'research exemption' was however only tackled in a recital[22] and not in the main text, thus arguably only addressing the corresponding provisions[23] leaving the open question of whether, *a contrario*, research generally does fall within the scope of the proposed regulation.[24]

In the presidency compromise text,[25] the Council in turn introduced a broad exception for 'AI systems, including their output, specifically developed and put into service for the sole purpose of scientific research and development'.[26] The '*sole purpose*' requirement needs to be distinguished from a situation where a research activity, by *e.g.* a company's research and development department, leads to or entails placing an AI system on the market or putting it into service.[27] In such a case the AI Act would apply. If adopted, this research exemption would mean that AI researchers will not be subject to the requirements in the proposed legislation. It does not mean, however, that they do not have to foresee the impact or risks for fundamental rights and freedoms of the AI systems they develop as researchers still need to comply with relevant rules, including the conditions and requirements to engage in research. For instance, in the context of EU-funded initiatives, projects involving the development, deployment and/or use of AI are required, from the outset, to assess and provide minimum information on the potential ethical risks regarding AI developments and the foreseen risk mitigation measures to be taken throughout the action implementation.[28]

Still, one may wonder whether the Council's choice could lead to undesirable effects given the other factors that come into play in a scientific research context, particularly the ethical obligation to make the results open and publicly accessible. Given their dissemination duties, AI researchers have to ensure that project outcomes are widely available, which could make research prone to be used by other actors and for other purposes than initially envisaged, *e.g.*, exploited by private entities to pursue commercial interests. This could result in the research exemption leading to a heightened risk of misuse of research findings, and thus calls for a need to examine measures to avoid possible side effects. Something to consider is to strengthen the rules on potential misuse of research findings, or go for an alternative approach consisting of facilitating access to sandboxes in which the rules could be suspended for research reasons, as commentators have suggested.[29]

When it comes to AI datasets, the EU AI Act proposal sets out quality criteria for the data used to train, validate and test high-risk AI systems. Article 10 specifies that the data used shall be relevant, representative, free of errors and complete, and with application-area specific properties.[30] To ensure representativeness of the data and to prevent risks of bias, the proposal foresees the processing of special categories of data, provided that adequate safeguards for fundamental rights are in place.

[21] The presidency compromise text introduces a new category of AI systems, namely general purpose AI systems, understood as AI systems that are able to perform generally applicable functions such as image/speech recognition, audio/video generation, pattern detection, question answering, translation etc. (EU AI Act proposal – Council compromise text, Recital 70(a)).

[22] EU AI Act proposal, Recital 16.

[23] In this case, arts 5(1)(a) and (b) of the AI Act proposal.

[24] Nathalie A Smuha and others, 'How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act' [2021] SSRN Electronic Journal <https://www.ssrn.com/abstract=3899991> accessed 20 October 2021.

[25] Council of the EU, '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 - Presidency Compromise Text (14278/21)' (Council of the EU 2021) <https://data.consilium.europa.eu/doc/document/ST-14278-2021-INIT/en/pdf> accessed 3 October 2022.

[26] EU AI Act proposal – Council compromise text, Art. 2(6).

[27] EU AI Act proposal – Council compromise text, Recital 12.

[28] See European Commission, 'EU Grants - How to Complete Your Ethics Self-Assessment' (2021) s 8 <https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/how-to-complete-your-ethics-self-assessment_en.pdf> accessed 10 February 2022.

[29] Gianclaudio Malgieri and Vincenzo Tiani, 'How the EU Council Is Rewriting the AI Act' (*Brussels Privacy Hub*, 6 December 2021) <https://brusselsprivacyhub.eu/publications/how-the-eu-council-is-rewriting-the-ai-act> accessed 14 March 2022.

[30] EU AI Act proposal, Article 10.

Despite establishing a commendable requirement, the draft regulation does not, however, provide any guidance as to how the data quality criteria should be accomplished considering the GDPR requirements for the processing of personal data. Also, the proposed Act does not specify what is representative data,[31] neither does it mention data integrity, precision or recall, which are considered a more appropriate measures to data classification.[32] At the same time, the draft regulation requires training, validation, and testing datasets to be 'free of errors and complete', which is considered to be technically impossible.[33] While it specifies that these requirements are to be met 'sufficiently' and 'in view of the intended purpose of the system',[34] it is however difficult to determine what constitutes 'sufficient' statistical properties[35] or 'adequate' requirements in terms of accuracy or robustness.[36] Such a prior assessment appears particularly difficult in the case of third-party or publicly available datasets. Moreover, the proposed Act does not specify how the data quality of such datasets should be checked, assessed, and mitigated.[37]

Legal and ethical challenges and complications of personal data use in AI research

Scientific research involving the processing of personal data of EU persons is subject to a specific regime under EU data protection law allowing for certain flexibility of the data protection requirements. Specifically, the GDPR puts forward the research regime for processing personal data, composed of specific derogations from certain obligations, as well as the requirement of appropriate safeguards. This regime applies insofar as the research takes place within an ethical framework and is carried out with the broader goal of contributing to the development of knowledge and wellbeing in society. While it does help avoid some issues when processing personal data for research purposes, important challenges remain as it does not fully accommodate the demands of AI research. This section examines how the GDPR intersects with trustworthy AI research practices and explores challenges in this context.

Compliance with data protection principles: a largely resolved issue

It would be fair to start this analysis by acknowledging the positive effects of the special GDPR regime for research. While it may initially seem challenging to comply with data protection provisions in a research context, the favourable data processing conditions for research in the GDPR seem to address relevant issues (or at least a great part of them). A clear example of this is compliance with data protection principles.

At first, it may seem hard to comply with data protection principles when processing large scale datasets for AI research. However, scientific research might be less affected by possible issues since those are largely addressed by the special data protection regime for research. Specifically, the GDPR exceptions to some data protection principles, such as purpose limitation and data minimisation, insofar as appropriate technical and organisational safeguards for the rights and freedoms of data subjects are provided,[38] help address this challenge.

To illustrate how the special GDPR regime helps address some issues, one can consider the example of purpose limitation. Under the purpose limitation principle, personal data shall not be further processed for purposes incompatible with the original purposes of collection.[39] In AI research, the data used are typically obtained from publicly available databases, which contain personal information likely to have been collected for different purposes.[40] Given the so-called presumption of compatibility between (secondary) processing for research purposes and the original purpose of collection, further processing for scientific research purposes does not raise purpose limitation issues. It should be noted, however, that this presumption is not a general authorisation to further process data in all cases for research.[41]

Lack of a definition of what is research

Despite containing a special regime for research, the GDPR does not specify what is to be understood by 'research' or 'scientific research', which are concepts with no generally agreed-upon definition. Some clarification may be found in recital 159 of the GDPR, which hints that the processing of personal data for scientific research purposes should be interpreted broadly, including for example technological development and demonstration, fundamental research, applied research and privately funded research as well as studies conducted in the public's interest in the area of public health.

In its Preliminary Opinion on data protection and scientific research, the European Data Protection Supervisor (EDPS)[42] considers that academic researchers, as well as non-profit organisations, government institutions and profit-seeking commercial companies can engage in scientific research. According to the EDPS opinion, scientific research 'applies the "scientific

[31] Smuha and others (n 24).

[32] Brendan Juba and Hai S Le, 'Precision-Recall versus Accuracy and the Role of Large Data Sets' (2019) 33 Proceedings of the AAAI Conference on Artificial Intelligence 4039.

[33] Martin Ebers and others, 'The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)' (2021) 4 J 589.

[34] EU AI Act proposal, Recital 44.

[35] EU AI Act proposal, art 10(3).

[36] EU AI Act proposal, recital 38.

[37] Also, the proposal does not define the criteria for measuring the quality of datasets but leaves it to the discretion of providers, which calls into question the effectiveness of the proposed regulation.

[38] GDPR, art 89(1).

[39] GDPR, art 5(1)(b).

[40] Jasserand (n 10).

[41] European Data Protection Supervisor, 'A Preliminary Opinion on Data Protection and Scientific Research' (2020) <https://edps.europa.eu/sites/edp/files/publication/20-01-06_opinion_research_en.pdf>.

[42] (n 40).

method" of observing phenomena, formulating and testing a hypothesis for those phenomena, and concluding as to the validity of the hypothesis'.[43]

After a case law analysis on this issue, a legal scholar concluded that there are three potential criteria to define what constitutes scientific research, namely: (1) the role of the legal entity; (2) the role of the person carrying out the activity; and (3) quality standards, including scientific methodology and scientific publication.[44] EU-funded AI research and innovation initiatives under EU-funded programmes such as Horizon 2020 and Horizon Europe are likely to meet those criteria. When it comes to commercial research on AI some argue that it should not benefit from the GDPR research exemption except when there is a public interest and similar safeguards to those that apply in a scientific research context.[45] In practice, as demonstrated by the Cambridge Analytica scandal, the relations between private companies and academic researchers can become entangled.[46] This case showed how when too much connected, scientific research and commercial activity can lead to unethical results. It also showed that the line between scientific research and commercial activity is not always easy to trace.[47]

Challenge to comply with legal basis

AI researchers typically rely on existing datasets as training data.[48] While not all the data used is personal data (and thus not covered by data protection law), often these datasets may contain personal data, which may at times fall under special categories of data, particularly during efforts for bias detection and mitigation in AI. Research shows that in order to make non-discriminatory decision models, for instance, with respect to race, the sensitive information about race needs to be used in the model building process.[49]

An example of this is attempting to have a more diverse dataset which accurately represents the relevant population, including sufficient data of historically under sampled groups of individuals, an activity that may require sensitive and biometric information.[50] Despite the presumption of compatibility that applies in a research context (section c*ompliance with data protection principles: a largely resolved issue*), purpose specification and lawfulness are two separate and cumulative requirements. As a consequence, the initial collection of data, re-used for scientific research purposes, even with the presumption of compatibility, would still require a legal basis.

When personal data are further processed for compatible purposes, such as scientific research, 'no legal basis *separate from* that which allowed the collection of the personal data is required'.[51] The question arises as to what extent is it the responsibility of the re-user (in this case, the AI researcher) to investigate the legal basis for the initial data processing. When it comes to publicly available datasets, it might prove difficult if not impossible to find out whether the initial collection of data was based on a valid legal basis meeting the conditions under the GDPR. Should that be the case, the researcher would nonetheless have to find and rely on an applicable legal basis.

Further, regarding special categories of data specifically, which are subject to stricter conditions under the GDPR, those data require an extra layer of protection and their processing is in principle prohibited, unless one of the ten specific exceptions contained in the GDPR is met.[52] In the context of scientific research, there are three legal bases that are most likely to apply, namely: the consent of the data subject,[53] the performance of a task carried out in the public interest[54] or a legitimate interest of the controller or a third party[55]. Yet, there are some issues with those legal grounds.

Regarding the first option (consent), it is not likely to provide a sufficient legal basis for various reasons. In case of large-scale datasets with millions of personal data points, some of which may be sensitive data, consent is almost impossible to realise. It is also problematic to contact data subjects to ask for consent in the first place. Relying on a derived consent from the consent that data subjects gave, for example, in the terms and conditions of a social media platform on which they initially uploaded their personal data, would hardly meet a threshold of 'real' genuine informed consent.

The second legal basis could be the performance of a task carried out in the public interest.[56] There is however lack of clear guidance how to interpret the 'public interest' in a research context. What constitutes a 'public interest' is a matter of national legislation and only a few Member States (*e.g.*, Finland) have considered scientific research as such.[57] In the case

[43] (n 40).

[44] Heidi Beate Bentzen, 'In the Name of Scientific Advancement: How to Assess What Constitutes "Scientific Research" in the GDPR to Protect Data Subjects and Democracy', *9781780689753* (Intersentia 2020) <https://www.duo.uio.no/handle/10852/82314> accessed 1 March 2022.

[45] Janos Meszaros and Chih-hsing Ho, 'AI Research and Data Protection: Can the Same Rules Apply for Commercial and Academic Research under the GDPR?' (2021) 41 Computer Law & Security Review 105532.

[46] Information Commissioner's Office, 'Investigation into Data Analytics for Political Purposes - A Report to Parliament' (2018) <https://ico.org.uk/media/action-weve-taken/2260271/investigation-into-the-use-of-data-analytics-in-political-campaigns-final-20181105.pdf> accessed 3 March 2022.

[47] Meszaros and Ho (n 44).

[48] Some examples include CommonCrawl, used to train large language models, ImageNet for object recognition, or MS COCO for computer vision tasks.

[49] (n 9)

[50] Raji and others (n 12) 4.

[51] GDPR, recital 50.

[52] GDPR, art 9(2).

[53] GDPR, art 6(1)(a).

[54] GDPR, art 6(1)(e).

[55] GDPR, art 6(1)(f).

[56] GDPR, art 6(1)(e).

[57] European Commission, 'Report on the implementation of specific provisions of Regulation (EU) 2016/679', < https://www.dataguidance.com/sites/default/files/1609930170392.pdf>.

of special categories of data, data processing must moreover be 'necessary for reasons of *substantial* public interest'.[58] Again, it is a matter for Member State law to substantiate this provision.

A third legal basis that would arguably be applicable for training AI models with personal data is legitimate interest.[59] Although this may hold true for the *training operation* itself, depending on the degree of anonymisation, the legitimate interest must be assessed also at the very first stage of data processing, namely the collection of data to be fed into the ML pipeline. As provided by the Article 29 Working Party,[60] the 'interest of carrying out scientific research', subject to appropriate safeguards, is one of the 'interests [that] may be compelling and beneficial to society at large'.[61] This legal basis requires a case-by-case assessment and a balancing test between the interests of researchers (and society as a whole) and data subjects' rights. The GDPR seems to weigh decisively in favour of third parties' interest when it states that 'for scientific or historical research purposes or statistical purposes, the legitimate expectations of society for an increase of knowledge should be taken into consideration'.[62] However, an important hurdle concerns that the basis of legitimate interests is only available to private entities, not public authorities (*e.g.*, public universities) 'in the performance of their tasks'. This is not to say that public authorities, or those having a hybrid status, depending on the task they perform are *a priori* excluded from the scope of the provision. What matters it the national laws' interpretation of what constitutes 'the performance of their tasks'.

In terms of establishing the second legal basis required when processing sensitive data, a way forward would be for researchers to rely on the possibility to process such data when 'manifestly made public by the data subject'[63]. As the EDPS has said, 'this provision has to be interpreted to imply that the data subject was aware that the respective data will be publicly available which means to everyone'[64] including, in this case, researchers.[65] In this line, the EDPB confirmed that 'the word "manifestly" implies that there must be a high threshold for relying on this exemption'.[66] A combination of the elements such as whether the data subject took a specific action to change the private settings into public ones, the nature of the social media platform, the fact whether the data subject has published the sensitive data himself/herself, or whether the data has been published by a third party, need to be considered for controllers to demonstrate that the data subject has clearly manifested the intention to make the data public.[67] A case-by-case assessment is always needed. It should be noted, however, that the practice of 'web scraping' is generally not permitted as a method to access data and is often explicitly prohibited by platforms' terms of service.

Difficulties to comply with data subject rights

Another challenge for re-using personal data in AI research relates to data subject rights as it may prove hard for AI researchers to comply with data subject requests. The major challenge is how to comply with the right to information.[68] The GDPR provides that, where personal data have not been obtained from the data subject, as it is the case for processing data in publicly available datasets, the controller must provide the data subject with information 'within a reasonable period after obtaining the personal data, but at the latest within one month'.[69] But there is an exception to this rule, which stipulates that controllers do not have to provide information to the data subject where appropriate safeguards are in place and when the provision of information proves impossible, requires a disproportionate effort or risks seriously compromising the achievement of the research.[70]

This was one of the matters considered in a Belgian case regarding the further processing of personal data extracted from social media for a research study on the fight against disinformation.[71] The Belgian Data Protection Authority (DPA) found that the non-profit organisation that carried out the study was dispensed from the obligation to provide prior information to the persons concerned in the context of journalistic activities, to the extent that informing them could have compromised the performance of the study and its subsequent publication. In its decision, the Belgian DPA highlighted the fact that the scientific exception requires more transparency, in particular on the methodology for collecting data and their degree of anonymisation and pseudonymisation. Moreover, to benefit from the research exemption, appropriate safeguards for the rights and the persons concerned should be adopted, as required by the special regime in the GDPR. In this case it would have included a registrar, external

[58] GDPR, art 9(2)(j).

[59] GDPR, art 6(1)(f). See Philipp Hacker, 'A Legal Framework for AI Training Data—from First Principles to the Artificial Intelligence Act' [2021] Law, Innovation and Technology 1.

[60] Currently, European Data Protection Board.

[61] Article 29 Working Party, 'Opinion 06/2014 on the Notion of Legitimate Interests of the Data Controller under Article 7 of Directive 95/46/EC' (2014) WP217 <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf> accessed 11 June 2021.

[62] GDPR, recital 113.

[63] GDPR, art 9(2)(a).

[64] (n 37).

[65] The Court of Justice of the European Union might clarify this question in the near future. The request for a preliminary ruling by the Higher Regional Court of Düsseldorf (C-252/21 - Facebook and Others), concerns, among others, whether visiting social media website or app and/or entering information and/or clicking or tapping on the buttons integrated into them by a provider such as Facebook ('Like', 'Share') constitute manifestly making the data public within the meaning of Article 9(2)(e) of the GDPR.

[66] 'Guidelines 8/2020 on the Targeting of Social Media Users | European Data Protection Board' <https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-82020-targeting-social-media-users_en> accessed 14 March 2022.

[67] (n 61).

[68] In arts 13 and 14 of the GDPR.

[69] GDPR, art 14(3)(a).

[70] GDPR, art 14(5)(b).

[71] Belgian DPA, *Case 13/2022* [2022].

privacy policy on the methodology and confidentiality measures. Therefore, notwithstanding the moral legitimacy of the purposes pursued by the researchers, the DPA concluded that releasing raw and un-pseudonymised personal data containing sensitive personal data is a disproportionate way to demonstrate the research methodology.[72]

Another potential challenge for fulfilling individuals' rights is the difficulty involved in identifying the individuals concerned by each dataset. Indeed, singling out data belonging to a specific individual is not always easy. The fact that the data subject can exercise his or her rights at different stages of the lifecycle of an AI system processing personal data, namely when gathering and preparing data, when training or testing the model, or during deployments, makes it even more challenging.

It can be argued that given the large scale of training data, identifying the individuals whose data belongs to, would indeed 'prove impossible'. Should this be the case, the controller should, however, 'take appropriate measures to protect the data subject's rights and freedoms and legitimate interests, including making the information publicly available'.[73] The GDPR does not provide further guidance on how to realise this obligation in practice. One solution would be for AI researchers to be transparent about the datasets they use. However, this raises other questions, in particular on whether and how that publicly available information could help data subject. Provided data subjects would indeed receive information about their data being used for AI research, how could they exercise their other rights?

As far as the rights to restriction and objection are concerned, in certain circumstances, a large number of individuals objecting to all or part of a scientific research may have a negative effect on the representativeness and reliability of the research data, and thus on the integrity of research.[74] For instance, personal data could be retained when the AI system would need re-training, quality search or evaluation processes. To this end, the GDPR[75] contemplates the possibility for researchers to ignore an erasure request where it would render impossible or seriously impair the processing of personal data for scientific research purposes. However, a balance must be struck between the research objectives and the rights of data subjects.

[72] The Belgian DPA found that releasing the personal data for justifying the integrity of the study is a legitimate aim but could be achieved in a way that is less detrimental to the interests, rights and freedoms of the persons concerned. The decision highlights some measures that could have been taken to mitigate risks, namely: (i) some data were irrelevant and could have been deleted (notably containing biographies); (ii) prior access control and access limitation for recipients should have been set up; (iii) the methodology should have been communicated to scientific peers in a limited way and scope; (iv) confidentiality agreement should have been signed to prevent the dissemination of the data (Belgian DPA, *Case 13/2022* [2022], §189).

[73] GDPR, art 14(5).

[74] (n 37).

[75] Art 17(3)(d).

Impact on the quality of the resulting AI system

Ethical and legal issues are integral to EU-funded research, with researchers having to comply with ethical principles and relevant legislation being part of the funding schemes.[76] Accordingly, EU-funded research strives to achieve high ethical and legal commitments from inception to end, with AI projects allocating part of their budgets to addressing ethical, legal and societal issues.[77] One of the main ethical issues in AI research is algorithmic bias.[78] While not everything is valid in fighting bias risks, large amounts of data can be of help here. Yet, data protection rules may have an impact on the quality of the developed systems. In particular, the GDPR may interfere with, and could potentially impair, the possibilities for AI researchers to collect or be able to use the data necessary for detecting and correcting bias in AI systems, or even to develop a system that complies with AI training data criteria more broadly. If access to relevant data and features is restricted, it is much more difficult to detect bias in the model, and even harder (in some cases, impossible) to fix it without access to the relevant data and features.

Furthermore, some research may even require access to data considered more sensitive given the need for data that reflect the problem the envisaged system aims to address and the real-world conditions where it is supposed to operate. For instance, security research requires data imitating the operational environment in question.[79] It is unfeasible to reproduce the high-stakes situations, such as cybercrime events, without access to relevant and representative data. Although currently available datasets capture some of the real-world conditions of the field, the ability to reproduce realistic environments and the genuine effects concerned by security situations and forensic practices might require access to criminal offense data, which are subject to stricter conditions under the GDPR.[80]

[76] See art 19 of the Horizon 2020 Regulation (Regulation (EU) No 1291/2013 of the European Parliament and of the Council of 11 December 2013 establishing Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020) and repealing Decision No 1982/2006/EC, OJ L 347) and art 19 of Horizon Europe Regulation (Regulation (EU) 2021/695 of the European Parliament and of the Council of 28 April 2021 establishing Horizon Europe – the Framework Programme for Research and Innovation, laying down its rules for participation and dissemination, and repealing Regulations (EU) No 1290/2013 and (EU) No 1291/2013, OJ L 170).

[77] See *e.g.* Katherine Quezada-Tavárez and Eva Houtave, 'Augmented Reality Technology to Counter Crime and Terrorism: Introduction to DARLENE' (*CiTiP Blog*, 16 April 2021) <https://www.law.kuleuven.be/citip/blog/augmented-reality-technology-to-counter-crime-and-terrorism-introduction-to-darlene/> accessed 16 April 2021; Stergios Aidinlis and Agata Gurzawska, 'Responsible Innovation in Multidisciplinary Research and Innovation Projects: Moving from Principle to Practice' (The International Society for Professional Innovation Management (ISPIM) 2021). See *e.g.* Quezada-Tavárez and Houtave; Aidinlis and Gurzawska.

[78] As the European Commission seems to acknowledge in its guidance for researchers. See European Commission, 'EU Grants - How to Complete Your Ethics Self-Assessment' (n 25) 40.

[79] European Commission, 'Horizon Europe: Work Programme 2021-2022 — 6. Civil Security for Society' (2021) 27 <https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2021-2022/wp-6-civil-security-for-society_horizon-2021-2022_en.pdf> accessed 27 July 2021.

[80] Art 10.

In this context, a number of complex legal, ethical and technical pre-requisites may make it difficult for the relevant community (including researchers, practitioners, industry, and policy makers) to have access to real, diverse and high-quality datasets, or alternatively to create representative datasets to engage in scientific research. Continuing with the same example, in security research the challenges include confidentiality constraints, data protection restrictions, and legal uncertainties regarding the legal basis to share operational data for research purposes.[81] Hence, it has become very complex for some research projects, particularly in more sensitive domains, such as security, health or defence, to develop reliable AI tools due to the absence of legislation addressing these issues.[82]

The AI Act proposal seems to recognise these types of hindrances. The draft regulation provides that, in order to protect the right of others from the discrimination that might result from the bias in AI systems, developers should be able to process special categories of personal data, as a matter of substantial public interest, in order to ensure the bias monitoring, detection and correction in relation to high-risk AI.[83] The processing of special categories of data is governed by the GDPR, which does not contain such an exemption for bias detection. The draft AI regulation seems to attempt to fill this gap by allowing the processing of special categories of personal data for bias mitigation, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons.[84] However, the exemption can only be used in relation to high-risk systems, leaving, for no evident reason, non-high-risk developers unable to rely on it.[85] Another question is whether the AI Act proposal provides a new legal ground for processing special categories of sensitive personal data. Should that be the case, the proposal could interfere with the application of data protection rules, including the GDPR.[86] Overall, there are various arguments in favour of and against adopting an exception in the AI Act that would enable the use of special categories of data to prevent AI-driven discrimination[87] and there is a need to find a proper balance.

Confronted with possible adverse consequences

Due to the uncertainties associated with the use of large-scale datasets as training, testing and validation data, and lack of guidance on certain aspects of the legal regime for research, AI projects and researchers themselves face certain risks in their balancing of the different interests at stake during their activities. The risks include possible GDPR infringement and subsequent breach of contractual obligations, potential reputational damage, difficulties to publish their research results in prestigious journals and conferences, and eventually even threaten to hamper societal trust in scientific research.

First, if the activities undertaken during the AI developments were to be considered unlawful processing of personal data, that could trigger the GDPR sanction regime,[88] leading to a significant liability (*e.g.*, fines of up to €20 million or 4% of annual revenue). In the context of EU-funded research, beyond the potential penalties, a GDPR infringement will also entail a breach of the contractual obligations with the funding agency, meaning that project participants may face other sanctions contemplated in the contract.

Second, the use of datasets eventually flagged as 'problematic' as training data raises ethical questions about the legitimisation of such research and about possible strategies to mitigate such risks, with likely reputational consequences for the project and parties involved. This is illustrated by the controversies associated with publicly available corpora, such as the MegaFace database and the Casia WebFace.[89] While initially widely used as benchmark datasets in AI research, these and other datasets were eventually flagged by researchers, activists and journalists due to the many concerns raised by the data and its use, particularly in terms of privacy and human rights.[90] The controversies ultimately also affected the reputation of projects where such data were used, as well as that of the centres and researchers involved in such projects.

Third, the difficulties to find and be able to use relevant data can also affect the possibilities for researchers to produce meaningful findings to disseminate in relevant fora, particularly high-ranking journals. In this context, both publishers and academics are confronted with difficulties. Scientific journals aim for the timely publication of correct results, at the right time and to the right audience, with the review of ethical aspects (including data ethics for AI research) being one of the success

[81] Community of Users for Safe, Secure and Resilient Societies (currently the Community for European Research and Innovation for Security (CERIS)), 'Summary of Workshop "Digital Tools for Law Enforcement – Issue of Training and Testing Data", Held within the Scope of the 13th Conference of the Community of Users on Safe, Secure and Resilient Societies

(Brussels, 28th- 29 March 2020)' (2019) <https://www.ffg.at/sites/default/files/downloads/Digital%20tools%20for%20LEAs%20-%20DATA%20issue.pdf> accessed 4 November 2022.

[82] While existing guidelines and laws aim to create the conditions (*e.g.*, the special GDPR regime for research), much of the existing guidance and policy initiatives aim to account for a broad array of data types, and do not accommodate the special characteristics of more sensitive sectors.

[83] EU AI Act proposal, recital 44.

[84] EU AI Act proposal, art 10(5).

[85] Michael Veale and Frederik Zuiderveen Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act' (SocArXiv 2021) preprint <https://osf.io/38p5f> accessed 20 July 2021.

[86] European Data Protection Board and European Data Protection Supervisor, 'Joint Opinion 5/2021 on the Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)' (2021) <https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf> accessed 20 June 2021.

[87] Marvin van Bekkum and Frederik Zuiderveen Borgesius, 'Using sensitive data to prevent discrimination by AI: Does the GDPR need a new exception?.' arXiv preprint arXiv:2206.03262 (2022).

[88] Art. 83.

[89] Set up by the University of Washington and the Chinese Academy of Science, respectively.

[90] Jasserand (n 10).

factors, given the implications of the data quality.[91] At the same time, researchers have to cope with the difficulties associated with the research data, including the varied terms of use, diverse legal frameworks and unclear provisions applicable to their activities.[92] Therefore, the challenges regarding data for AI research could hamper the possibilities for researchers to publish their work in scientific fora to gain recognition for their efforts, earn academic advancement and maximise the impact of their endeavours.

Finally, a broader aspect to consider is the societal trust in researchers and scientists to produce knowledge and make relevant discoveries that help solve problems. Trust is essential in scientific research. Yet, concerns and controversies associated with scientific research, such as scandals associated with the use of the data available for AI research, can hinder public trust in science. Mistrust and rejection of scientific findings could in turn significantly reduce the return on research and ultimately entail a major waste of public money.

Ways forward

Navigating between the need for large amounts of data, data quality requirements and GDPR compliance is a challenge in AI research. Researchers often face questions about the quality, legality, and ethics of the openly accessible datasets that they use because those are the ones available. There are possible ways to address these issues and find a balance between the different interests at stake, particularly privacy and innovation.

More availability of high-quality datasets

The EU AI Act proposal provides that for the development of high-risk AI systems, certain relevant entities, such as digital innovation hubs, testing experimentation facilities and researchers, 'should be able to access and use high-quality datasets within their respective fields of activities' related to the regulation.[93] In this context, it is no coincidence that the EU AI policy is accompanied by the European data strategy,[94] which amongst others aims to address the challenge of lack of available data necessary to achieve EU's digital transformation. In particular, European common data spaces established by the Commission and the facilitation of Business to Business and Business to Government data sharing in the public interest are expected 'to provide trustful, accountable and non-discriminatory access to high-quality data for the training, validation and testing of AI systems'.[95] As an example, a European health data space will facilitate access to health data and the training of AI algorithms on those datasets.[96]

The EU legislator seems to assume that data-related initiatives will provide a sufficient framework for researchers' access to high-quality datasets, an aspect that the AI Act proposal does not address. In particular, a horizontal governance framework for data access and use should 'facilitate decisions on which data can be used how and by whom for scientific research purposes in a manner compliant with the GDPR'.[97]

Data altruism, as contemplated in the recently adopted Data Governance Act (DGA)[98] aims to make sharing of personal data on the basis of the consent of data subjects possible 'for objectives of general interest'.[99] However, the DGA does not specify what is considered 'an objective of general interest'. It hints in the recital that the support to scientific research should also considered to be an objective of general interest.[100] Yet, it is left to national laws to determine whether scientific research will fall under this definition. A wide binding data access framework for AI research would be more than welcome to enable researchers to access data in a harmonised and legally-compliant way.

Setting up data spaces that can be used in research

According to the Data Strategy, next to the horizontal data governance framework, the Commission intends to fund the establishment of EU-wide common, interoperable data spaces in strategic sectors.[101] For instance, the difficulties regarding data availability in security research is one of the lessons learned from previous research projects and intended to be addressed in its the Horizon Europe framework by fostering the creation of a European common training and testing data repository for this field of research.[102]

Although such data spaces can be used for multiple purposes, the EDPS points out that it should be clearly defined what

[91] As an important step in this direction, some journals and academic conferences on AI have announced extra ethics checks on the research to be published. A case in point is the Conference and Workshop on Neural Information Processing Systems (abbreviated as NeurIPS and formerly NIPS), a prestigious AI meeting where papers could be rejected due to ethical and legal doubts associated with the data used. See Alina Beygelzimer and others, 'Introducing the NeurIPS 2021 Paper Checklist' (*Medium*, 26 March 2021) <https://neuripsconf.medium.com/introducing-the-neurips-2021-paper-checklist-3220d6df500b> accessed 20 June 2021.

[92] Anna Rogers, Tim Baldwin and Kobi Leins, 'Just What Do You Think You're Doing, Dave?'A Checklist for Responsible Data Use in NLP' [2021] arXiv preprint arXiv:2109.06598.

[93] Recital 45.

[94] European Commission, 'White Paper on Artificial Intelligence: A European Approach to Excellence and Trust (COM(2020) 65, Final)' (n 3); European Commission, Communication on a European strategy for data 2020 [COM(2020) 66].

[95] EU AI Act proposal, recital 45.

[96] 'European Health Data Space' <https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en> accessed 31 October 2022.

[97] European Commission Communication on a European strategy for data (n 94).

[98] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act), OJ L 152, 3.6.2022, p. 1–44.

[99] DGA, art 2(16), recital 45.

[100] Recital 45.

[101] European Commission Communication on a European strategy for data (n 94).

[102] European Commission, 'Horizon Europe: Work Programme 2021-2022 — 6. Civil Security for Society' (n 79).

are the permitted purposes of its creation (*i.e.*, research and non-research). The limitations on the cross-context use of data, including prohibitions on the use of sensitive personal data for other purposes remain applicable to personal data in a data space.[103] The data gleaned in these data spaces could, however, benefit the research community.

Avoiding problematic datasets from proliferating and being used in research

The efforts to foster more availability of data should not be done in a vacuum, but as part of larger initiatives to tackle different issues, including preventing and restricting the use of problematic datasets. This is important because, despite the scandals associated with widely used datasets (*e.g.*, MegaFace, which was retracted in 2019 due to problematic practices outlined above), AI researchers have continued using those. The original datasets, and derivatives thereof, have been cited in several publications released months after retraction of the data.[104]

This calls for measures to avoid troublesome datasets from continuing to be relied upon in AI research. In this context, it is important that the efforts towards improving access to relevant and sufficient data for AI research endeavours are accompanied by initiatives to prevent problematic data from continuing to proliferate and being used. To address this, a recent study[105] calls for, amongst others, making ethically salient information about datasets clear and accessible, active stewardship of the data and its uses, employing ethics review procedures to promote responsible data uses, and the advance review of datasets and publications.

Developing standards or codes of conduct

Another path that could be taken in providing more clarity on how to access and use datasets in a GDPR-compliant way would be standards or codes of conduct. The GDPR[106] specifically encourages the drawing up of codes of conduct intended to contribute to a proper GDPR application. According to the EDPB, codes of conduct should specify how to apply the GDPR in a 'specific, practical and precise manner', provide 'unambiguous, concrete, attainable and enforceable (testable)' standards and rules.[107]

In this context, codes of conduct could be developed to resolve the tensions around accessing data for AI research in the GDPR-compliant way. This approach is being explored by the European Digital Media Observatory (EDMO), which in May 2022 released its report on researchers access to platform data together with a draft Code of Conduct on how platforms can share data with independent researchers while protecting users' rights.[108] This work provides the groundwork for a Code of Conduct on Platform-to-Researcher Data Access under Article 40 of the GDPR.

Conclusions

This article has examined the interactions between AI research and the GDPR, particularly when it comes to the use of large-scale databases containing personal data to train, test and validate AI systems. This analysis identifies data protection challenges posed by the use of the data required for AI research, particularly when engaging in trustworthy AI efforts. The findings suggest that the GDPR regime for research provides for (at least partial) solutions and interpretation keys to resolve relevant complications. However, this is unfortunately not the case for all aspects of AI research. In particular, AI researchers face hurdles to meet the quantity and quality criteria for developing trustworthy AI systems, while at the same time attempting to comply with data protection requirements. This paper elucidates these issues and proposes some ways to address remaining challenges.

Beyond the GDPR, this paper discusses the EU AI Act proposal. It is not yet clear whether the proposed regulation will come into force, and if its requirements (such as the data quality criteria) will apply to AI research, as has become apparent in this study. In any case, it is important to ensure a balanced approach in the final version of the text. For instance, the research exemption that appears in the latest proposal draft at the time of writing (*i.e.* the presidency compromise text) may ultimately become a sort of mixed blessing. On the one hand, AI research would not be subject to stringent data quality requirements applicable to commercial, high-risk applications – which would prove difficult for researchers in any case. At the same time, researchers will not be able to benefit from the AI Act affordances such as the new possibility to process special categories of personal data for bias mitigation that actors that do fall under the scope may benefit from. As a result, the EU AI Act may ultimately not change the status quo of data (un)availability for scientific research, an outcome that the AI research community would hope for and welcome.

Hopefully, these and other issues identified in this research will be addressed soon, in a way that EU law is well equipped

[103] European Data Protection Supervisor, 'Opinion 3/2020 on the European Strategy for Data' (2020) <https://edps.europa.eu/sites/edp/files/publication/20-06-16_opinion_data_strategy_en.pdf>.

[104] Karen Hao, 'Deleting Unethical Data Sets Isn't Good Enough | MIT Technology Review' <https://www.technologyreview.com/2021/08/13/1031836/ai-ethics-responsible-data-stewardship/> accessed 31 October 2022.

[105] Kenny Peng, Arunesh Mathur and Arvind Narayanan, 'Mitigating Dataset Harms Requires Stewardship: Lessons from 1000 Papers' [2021] arXiv preprint arXiv:2108.02922.

[106] Art 40.

[107] EDPB, 'Guidelines 1/2019 on Codes of Conduct and Monitoring Bodies' (2019) <https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-12019-codes-conduct-and-monitoring-bodies-0_en>.

[108] EDMO, Report of the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access, 31 May 2022, https://edmo.eu/wp-content/uploads/2022/02/Report-of-the-European-Digital-Media-Observatorys-Working-Group-on-Platform-to-Researcher-Data-Access-2022.pdf.

to capture these situations, finding the right balance between enabling to reap the benefits of AI research while mitigating privacy and other fundamental rights concerns. Research is a core part of the EU digital strategy, thus more clarity on the requirements and greater support for researchers would be welcome.

Ethics and consent

Ethical approval and consent were not required

Data availability

No data are associated with this article.

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 02 May 2023

https://doi.org/10.21956/openreseurope.16383.r30995

© **2023 Georgiou D.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Dimitra Georgiou
[1] Department of Digital Systems, University of Piraeus, Pireas, Greece
[2] Department of Informatics, Hellenic Airforce Academy, Athens, Greece

The paper provides insight into the interesting topic of GDPR and AI research. It presents light on the under-discussed issue of challenges and tensions between the processing of large-scale datasets for AI research and compliance with the GDPR. The main objectives of the paper are three: 1) that it analyzes the data protection requirements and limits for the processing of personal data in AI research, 2) it reflects on possible complications regarding data quality requirements for trustworthy AI and (GDPR) compliance, and 3) it presents possible ways forward to reconcile GDPR requirements and AI research.

Although the topic is quite innovative and of interest to the research community, the paper lacks proper research maturity for being a winning paper. I recommend that this paper be indexed after revisions:

Firstly, I found the paper difficult to read, mainly because I think the paper tries to fit too much in the space it has. It is not possible to discuss all related aspects in a paper.

Another criticism of the paper is that the paper presents (In the section AI training, testing and validation data') many important research questions *(e.g., What data is needed to embed values and achieve 'trustworthiness' in EU-made AI systems? Is that processing allowed under EU data protection law?...*), that are not analyzed and discussed well in this article. The questions should be more targeted at the research, including only the information relevant to the immediate study subject and the reasons the writers are doing the research. The research questions could help the audience- readers to easily understand the contribution of this research. In addition to this, in this section, the writer should motivate the reader to continue reading.

In the section "Legal and ethical challenges and complications of personal data use in AI research" the challenges are presented well. But it could be useful for the reader if the writers could categorize the requirements or present them in an understandable way (e.g Figures or tables).

Furthermore, It could improve the readability of the paper, if the writers could present the important results at a glance.

In the section “Ways forward”, a stronger comparison with existing literature would have helped the paper’s argument about the innovation and novelty of the work.

The final section of this paper contains the concluding remarks. As the purpose of the conclusion section is to explain the results and show how they help to answer the research questions posed in the Introduction, the writer should present a more contextualized conclusion targeting the innovation and contribution of this paper.

In conclusion, much more justification is needed to make the discussion of the paper meaningful, and the ideas have to be better structured and better explained/motivated.

Is the topic of the essay discussed accurately in the context of the current literature?
Partly

Is the work clearly and cogently presented?
Partly

Is the argument persuasive and supported by appropriate evidence?
Yes

Does the essay contribute to the cultural, historical, social understanding of the field?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Computer Security and Data Protection, Information Security and Privacy Policies,Privacy Enhancing Technologies, Data Protection and impact assesment,Compliance of information systems with GDPR

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 14 December 2022

https://doi.org/10.21956/openreseurope.16383.r30416

© 2022 Beyan O. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Oya Deniz Beyan

Institute for Medical Informatics, Faculty of Medicine, University Hospital Cologne, University of Cologne, Cologne, Germany

The topic is very wide and it is not possible to discuss all related aspects in a single article. For example, special cases such as patient data and consent might have other relevant aspects. Nevertheless, the article draws a good scope and discusses the selected issues in the context of the literature. It provides evidence, although each EU country and their regions might have different restrictions and implementation of GDPR.

Data quality for the AI research is very worth to highlight. In our research (as the director of a medical data integration center and biomedical institute at the medical faculty), we experience many trade offs between anonymization of data (shuffling, reducing, summarizing, etc.) and the required quality of data for research. Specifically, if you model medical conditions, less data items are likely to represent a minority group, such as rare conditions (not an outlier). And at the end they are not represented in the model.

Also removing sensitive fields such as race, might be a source of bias too (rather then reducing bias). Treatments might be working differently in different race groups. If they are not successful for the majority of the minority, ignoring this minority (because you can not associate false negative cases to specific minority group), might end up creating disadvantage for this group.

Another point is that the “purpose limitation and data minimization” does not always work for real life cases. If an AI developer needs to integrate multi source data, from primary sources, it is very difficult to minimize the data (since after the integration, mostly data size reduces).

As article mentions legal uncertainties are very important aspect, and that might become blockers for approval processes.
In ways forward section, authors mention data spaces as a solution. Although they are very necessary, close cooperation with legal stakeholders, and without establishing common understanding for reusing data for AI this way forward might be blocked by local legal uncertainties.

Is the topic of the essay discussed accurately in the context of the current literature?
Partly

Is the work clearly and cogently presented?
Yes

Is the argument persuasive and supported by appropriate evidence?
Yes

Does the essay contribute to the cultural, historical, social understanding of the field?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Medical Data Science

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.