Updating Purpose Limitation for AI

A normative approach from law and philosophy

Rainer Mühlhoff
rainer.muehlhoff@uni-osnabrueck.de
Universität Osnabrück
Osnabrück, Germany

Hannah Ruschemeier
hannah.ruschemeier@fernuni-hagen.de
FernUniversität in Hagen
Hagen, Germany

ABSTRACT

This paper addresses a critical regulatory gap in the EU's digital legislation, including the proposed AI Act and the GDPR: the risk of secondary use of trained models and anonymized training datasets. Anonymized training data, such as patients' medical data consented for clinical research, as well as AI models trained from this data, pose the threat of being freely reused in potentially harmful contexts such as insurance risk scoring and automated job applicant screening. To address this, we propose a novel approach to AI regulation, introducing what we term *purpose limitation for training and reusing AI models.* This approach mandates that those training AI models define the intended purpose (e.g., "medical care") and restrict the use of the model solely to this stated purpose. Additionally, it requires alignment between the intended purpose of the training data collection and the model's purpose.

The production of predictive and generative AI models signifies a new form of power asymmetry. Without public control of the purposes for which existing AI models can be reused in other contexts, this power asymmetry poses significant individual and societal risks in the form of discrimination, unfair treatment, and exploitation of vulnerabilities (e.g., risks of medical conditions being implicitly estimated in job applicant screening). Our proposed purpose limitation for AI models aims to establish accountability, effective oversight, and prevent collective harms related to the regulatory gap.

Originating from an interdisciplinary collaboration between ethics and legal studies, our paper proceeds in four steps, covering (1) the definition of purpose limitation for AI models, (2) examining the ethical reasons supporting purpose limitation for AI models, (3) critiquing the inadequacies of the GDPR, and (4) evaluating the proposed AI Act's shortcomings in addressing the regulatory gap. Through these interconnected stages, we advocate for amending current AI regulation with an updated purpose limitation principle to address one of the most severe regulatory loopholes.

Keywords: AI Act, AI Governance, AI regulation, collective privacy, data ethics, data protection, general purpose AI systems, GDPR, Ethics, LLMSs, EU regulation, secondary data use, power asymmetries, Open Source

1 INTRODUCTION

Powerful AI systems based on machine learning can have a far-reaching impact on society. In scholarly and public debate, many risks are documented, including novel privacy violations, infringements of fundamental rights, discrimination [35; 8] and unfair treatment [20], hate speech, influencing public debates and democratic election processes, monopolization of AI companies, lack of consumer protection [60], new forms of exploitation of human labor and socio-economic power differentials, particularly in countries of the Global South [76; 14; 13; 48; 24; 53]. Large Language Models (LLMs) synthesize eloquent and accurate sounding text or shiny images, up to the point of inventing new information and content [7; 85]. Debating AI risks is important as most AI systems scale very easily not only to large numbers of people they affect, but also across different sectors and areas of life, worldwide. This risk is particularly imminent when only a few models exist on the market that get reused and repurposed for ever wider purposes (for examples, see [64]).

Machine learning models are based on the principle of pattern recognition and therefore require significant amounts of data for training. This data is typically generated by thousands to millions of different individuals and extracted from multiple sources. These sources could include usage data from the web or smartphone apps, surveillance and tracking data [88], purchasing and transaction data, location data and communication metadata, or data explicitly produced in data labor (such as annotation, content moderation, customer support, digitalized services of all kinds) [56]. As most contemporary AI systems essentially rely on the exploitation of data stocks and streams, we shall refer to them as 'data-driven AI', simultaneously referring to both more classic forms of machine learning (e.g., scoring or ad targeting systems) and recent achievements such as generative AI systems.

The concept of purpose limitation has a long history in data protection.[1] Purpose limitation has also been vividly debated in relation to big data and the open-purpose "data mining" methodology in legal and empirical research [32; 51; 46; 87]. The majority of these contributions have found

[1]The principle of purpose limitation is laid down in various different data protection regulations beyond the GDPR: Chapt. 3 Cond. 3 Sec. 13 POPIA (South Africa); art. 4 FDPA (France); Sec. 202 f. ADPPA, Sec. 1798.100 (b) CCPA (US); art. 9 PIPL (China); art. 4 Ley Orgánica 3/2018 (Spain).

the purpose limitation principle incompatible with the promise of unexpected findings and innovative research that is associated with big data and data mining methodology. We have dealt with these arguments elsewhere [56], pointing out the differences between purpose limitation in data protection, which is linked to personal references in the training datasets, and purpose limitation for AI models, which is linked to the data that represents a trained model (which does not need to be personal data) to enable public oversight and risk control. Purpose limitation for models thus shifts the regulatory point of intervention from the input data (training data) to the trained model and its context of use – which might be different from the context of model training. As a consequence, argue in this paper, purpose limitation for models actually *promotes* data-driven research approaches as it provides a pivotal safeguard against potential abuse of the results (see sections 3.2 and 5.2).

The existence of data protection regulations such as the GDPR in the EU, however, does not seem to prevent the many risks and powerful ramifications of AI quoted above. In practice, data protection regulation is hardly effective in relation to data-driven AI because of both structural mismatches and enforcement deficits. For one, the ill-suited distinction between personal and non-personal data and the binding of defense rights to identified data subjects are conceptually at odds with big data applications such as machine learning. This pitfall is considerably amplified by gross enforcement deficits that are well documented [27; 70; 83].

Additionally, data protection regulation does not sufficiently address the highly aggregated and derived data that constitutes a trained model. Considering the internal weights and parameters of trained models as a specific kind of data – we refer to it as 'model data' [56] –, this data is not commonly considered personal data.[2] This poses an enormous risk of the uncontrolled secondary use of trained models in contexts and applications different from their original purpose. The secondary use of trained models is a considerable loophole in AI regulation that arises as regulatory regimes such as data protection only focus on the input stage that is concerned with training data [84].

At the same time, the many helpful debates about the risks of AI systems focus primarily on the *output* of an AI model when it is applied to concrete cases in a specific application context (by this we mean, for example, the concrete snippet of text that is produced by ChatGPT in reaction to a prompt, or the risk of developing a certain psychiatric disease calculated by a diagnostic AI model for a concrete person X). This critique then focuses on the consequences of actual applications (inferences) of a model and fails to address the general risks associated with the trained model due to its *potential* for circulating without control between different actors, application contexts and purposes. Amending these critical approaches, we thus seek to establish the trained model – which constitutes data and data processing *between* training and inference – as the object of regulatory interventions.

In this situation, we advocate for a zooming out and shifting the focus to the whole life-cycle of creation, use, and, above all, potential reuse of AI models. The narrow focus on individual contexts or data processing procedures often obscures the potential risks associated with the secondary use of data in trained AI models. This risk is a gateway to social inequality, unfair discrimination, and exploitation as unaccounted side effects of AI projects that often start with good intentions.

Our initial thesis is that the mere existence of a trained AI model inherently embodies a risk that is inversely proportional to the level of public governance over the model's potential applications or reapplications. We theorize this risk as a specific manifestation of informational power asymmetry that results from the possession of aggregate data and trained models (see section 3.3). This power is not sufficiently under public and democratic control given the ease with which a trained model can, in the current regulatory environment including the AIA, circulate without control and effective restrictions to other actors, application contexts and purposes. Controlling this power is one objective of our proposal for a regulation of trained models.

As we point out in comparing purpose limitation for models with purpose limitation from data protection (see section 4), the reuse of trained models poses an additional threat because it is risky to society and arbitrary third parties, and not only to the fundamental rights of the persons represented in the training data (the latter risk is already covered by the GDPR purpose limitation). Hence, it is the potential *collective* damages, manifest in potential threats to *anyone* (not only to the individuals in the training data) and in aggregate effects such as social inequalities, patterns of unfair discrimination and exploitation, that warrant an additional regulatory mechanism.

We will also justify our regulatory proposal of a purpose limitation for AI models pointing out the limitations of the Artificial Intelligence Act (AIA) of the European Union in this respect (section 5). In its risk-based and *ex post* oriented regulatory framework, the AIA's focus is predominantly on the deployment of models within specific application contexts, neglecting to fully address issues related to the whole life cycle of a model, including the training of models and their subsequent widespread distribution. Additionally, the Act provides extensive exemptions for open-source

[2]The regulatory gap addressed in this paper exists regardless of whether the model data is personal or anonymous data. In particular, this means that the severe societal risks arising from the secondary utilization of trained models is also present when the model data is anonymous data, in which case it would not fall in the scope of the GDPR. Consequently, trained models can be shared, traded, or reused in different contexts without the safeguards of the GDPR.

models and generative AI that further add to the problem of uncontrolled reuse (see section 5.2. In contrast to the deployment-based approach of the AIA, we advocate to shift the regulatory focus from the intended use to the *potential* uses and reuses, including different actors and their different positions, their potentially unforeseen paths of dissemination, the potentially affected legal interests, and, most importantly, the actual and potential collective effects [56].

As the example of an AI model for medical diagnosis shows 3, there are many potential applications of AI and machine learning that are (rightly) regarded as beneficial to society in public, political, and ethical debates. Often, funding decisions and political programs promote such applications. In these contexts, the perilous potential for the reuse of trained models in other contexts and other purposes is often overlooked and not included in the risk assessment of political actors and ethics committees. Crucially, the risk of unaccounted secondary use can materialize years later and involve different entities, such as various companies emerging from mergers or acquisitions of the original firm. Also, models that are created in public research and with public money might later be reused for doubtful or unwanted commercial purposes by private actors.[3] Often, these hazards go unrecognized or undiscussed in research ethics committees, funding decisions and public discourse surrounding the corresponding technological advancement [56]. We argue that to fully endorse AI for beneficial purposes, we need to ensure – both towards training data subjects and the public at large – that the models built from such projects remain with the original purposes.

In line with such a structural approach, in proposing purpose limitation for models, we follow three interrelated objectives (see section 4): (1) enabling accountability, (2) enabling public supervision, and (3) limiting collective and individual harms associated with the reuse of trained models.

(1) As regards accountability, both the institutions that develop AI models for what may be desirable purposes and the parties that seek to reuse such a secondary use should be accountable for ensuring that the models they develop or use do not constitute a case of abusive secondary use. (2) As regards supervision, we suggest that developers of AI models that allow for a *high-risk* secondary utilization (regardless of the primary purpose for which the model is developed) be registered with a supervisory authority, which could be one of the authorities installed by the DSA[4] or AI Act (see section 5.3). (3) In regard to the prevention of harm, we highlight the dogmatics of our proposition that purpose limitation for models serves the limitation of informational power asymmetries which arise from the potential use of trained models on *anyone* and for any purpose.

2 INTERDISCIPLINARY APPROACH AND METHODOLOGY

This work is the result of an in-depth interdisciplinary collaboration between philosophy and legal studies. In arguing for the protection of such collective and third-party interests that can be adversely affected by uncontrolled reuse of trained models, our approach starts from the assumption that regulating AI means regulating power (see section 3.3). Trained models are a specific manifestation of informational power [57]; regulating the distribution and use of trained AI models is an approach to control this power. From an ethical perspective, embracing this conception of power entails another form of interdisciplinarity that we strongly endorse, namely, the intersection of (classical) ethics and social philosophy [2; 82; 9; 52]. From a legal perspective, approaching AI regulation to limit the informational power accumulating in the hands of corporate or state actors that possess trained models means taking a *preventive* approach. We advocate for applying the theoretical foundations of the principle of proportionality within the realm of risk prevention law to regulate AI. This involves adapting the normative framework of proportionality testing, that is, assessing whether a means effectively achieves its purpose in light of the concurrent restrictions and negative side effects. This framework should encompass the safeguarding of individual, collective, and political interests in achieving the intended purpose. In this assessment, the presence of informational asymmetries should be acknowledged as they inherently limit the benefits and interests of various actors, thereby influencing the intensity of the regulation required. The greater the impact on people and critical legal interests, the more robust the justification for regulation becomes [56].

3 ETHICAL ARGUMENTS FOR PURPOSE LIMITATION

To make our normative discussion more vivid, we begin by outlining a plausible scenario to elucidate the regulatory gap addressed in this paper. Imagine a clinical research group at a public university hospital that aims to explore using Machine Learning models for predicting psychiatric

[3] "It's become standard practice for technology companies working with AI to commercially use datasets and models collected and trained by non-commercial research entities like universities or non-profits." [4]; examples: [3] for the Shutterstock data set used by Meta [36]. Furthermore, private companies actively fund research in public institutions to commercialize the results afterwards [see the Ommer-Lab at Ludwig Maximilian University of Munich and their contribution to the development of Stable Diffusion [78].

[4] Digital Services Act, Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC.

diagnoses based on voice data (recorded audio) of psychiatric patients. (Detecting psychiatric disorders from speech is a vivid research field that is also pursued in commercial applications, cf. [41; 79; 1].) The group assumes that this project could enhance psychiatric diagnostics and treatment – which also represents the purpose of the processing of the patients' data. Existing patients may volunteer to share anonymized medical records and audio recordings with the research project. We assume the research group successfully produces a model capable of predicting psychiatric dispositions, such as depression or anxiety disorders, based on audio data. We further assume that the training data was collected to ensure the anonymity of data subjects and that the model was trained in a way that effectively anonymizes its data (the matrix of weights and other internal parameters).

3.1 Defining purpose limitation for AI models

Under the current legislative framework of European data protection law, two subsequent forms of data processing are possible, which are highly critical and contestable on ethical grounds (see Figure 1 for an illustration):

Case 1: The clinical research group could distribute or sell the trained model to an external third party, such as an insurance company (see route ① in Figure 1). The insurance company could incorporate the model into their insurance risk assessment routines, for instance, by using audio probes that can be recorded on a telephone hotline. While the transfer of the trained model, which comprises anonymous data, doesn't encounter regulatory hurdles, the application of the model to calculate psychiatric diseases for specific insurance applicants falls within the scope of the GDPR [57]. However, in practice, insurance applicants are often effectively compelled to consent to the processing of their personal and sensitive data as part of insurance in risk assessment as otherwise their insurance application will fail. Moreover, it is plausible that the model for predicting psychiatric diseases could be employed by the insurance company as a component of a larger model or algorithmic routine for risk assessment (Model 2 in Figure 1). In such a scenario, it is conceivable that the predicted psychiatric condition may not be stored or explicitly output during the risk assessment procedure; however, this does not diminish the critical nature of this form of model reuse.

This first case presents a prototype of **secondary use of a trained model,** serving a purpose that exceeds or contradicts its original purpose and the purpose for which the training data was collected (clinical research and improvement of psychiatric treatment). We aim to prevent this scenario of the misuse of trained models by imposing a principle of purpose limitation on both the data processing that constitutes the training of model 1 from the training data and the transfer of model 1 to a third party.

Definition 1 (purpose limitation for models)**.** Purpose limitation here means that a machine learning model can only be trained, used and transferred for the purposes for which the training data was collected.

The copy of model 1 obtained by a third party would be restricted to the initial purpose for which the model was trained and the training data collected. Both direct and indirect reuse for insurance risk assessment would be prohibited as an effect of this provision.

Case 2: Rather than transferring the trained model to a third external party, the clinical research group could share the anonymized training dataset collected during their research. As detailed in section 4, processing anonymous data falls outside the scope of the GDPR. Consequently, obtaining the anonymized training dataset, for example, by an insurance company, faces no substantial hurdles. The insurance company could then use this dataset to train their own machine learning model (see route ② in Figure 1). Alternatively, the insurance company could also combine this specific training data with other data to train any other model from it. The recently revealed practice of the UK Biobank, which shared anonymized datasets with insurance providers, serves as a warning that this scenario is highly relevant [17].

If the insurance company utilizes the original training dataset that was collected by the research institution, or any model the company could train from this dataset, for insurance risk assessment, we regard this scenario as a **secondary use of the training data.** This secondary use surpasses and contradicts the original purpose for which the training data was collected by the research institution. This scenario of misuse of training data can be prevented by imposing a variation of the principle of purpose limitation that was already articulated for scenario 1:

Definition 2 (purpose limitation for training datasets)**.** Purpose limitation here means that an actor may only train a machine learning model from a data set X if they can prove the purposes for which the data set X was originally collected and if the training and use of the model follows these purposes.

While the processing (including transfer) of anonymous data is generally permitted, our proposal would introduce an accountability obligation on any processor of such data from the moment they start using the data to train a machine learning model. This accountability obligation would require the processor to trace the origin of the training data and the purpose for which it was originally collected. This "backward accountability" (cf. Figure 1) is crucial to prevent a diffusion of accountability that could occur if controlling the risks of trained models could no longer be

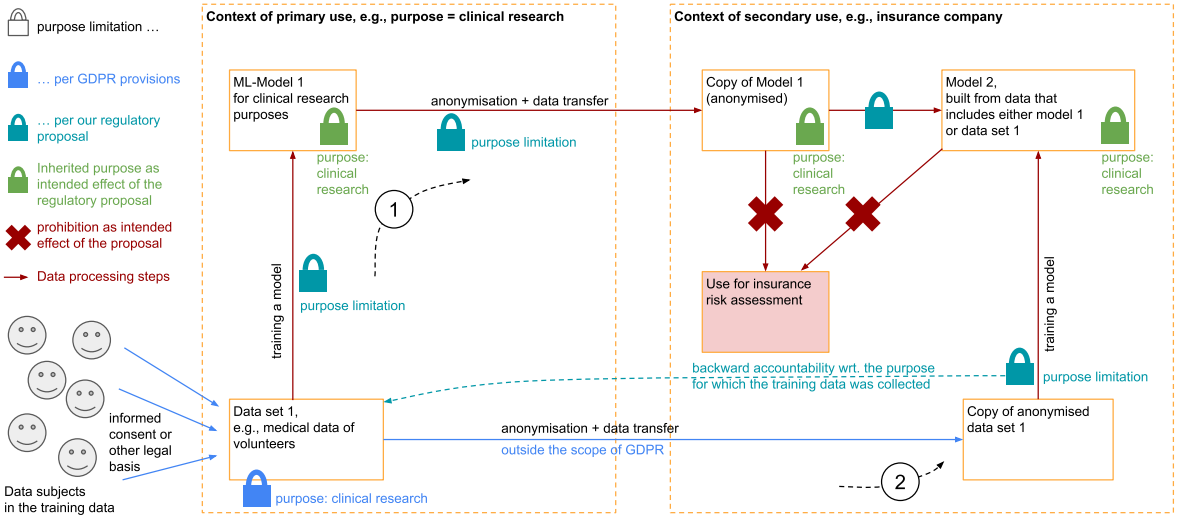

Figure 1: Flow chart illustrating different data processing steps relevant to our argument for purpose limitation for models and training data. © 2024 Rainer Mühlhoff & Hannah Ruschemeier.

assigned to specific actors after potentially many iterations of secondary use. Existing defense rights of the affected persons cannot counteract this, as both the individuals whose data was used to train the model and those who are subject to the model's application cannot oversee or control the chain of reuse of the training data.

In the remainder of this section, we will give ethical reasons for these provisions. In sections 4 and 5 we will then justify the necessity and proportionality of this proposal in legal terms.

3.2 Breach of trust towards the training data subjects

In many examples, machine learning models are trained from data that is collected from individuals such as medical patients (in medical research), clients, students, employees, users (e.g., of platform services, apps, devices), callers (e.g., to a telephone hotline), applicants (e.g., for jobs, insurances, educational programs), suspects (e.g., in relation to police or security services). As discussed in detail in section 4, the collection of personal data supposes a legal basis under the GDPR, which could be informed consent, but also specific (including national) legislation that enables the processing of personal data for tasks that are in the public interest (e.g., healthcare system, social services, public insurance companies). Regardless of which legal basis enables a specific data processing, in all these cases the architecture of the GDPR aims at ensuring that the processing is limited to a specific and legitimate purpose that is known ("explicit"), or at least knowable, to the data subjects.

The utilization of data or AI models for unintended purposes, even if not currently restricted by laws, jeopardizes the trust of data subjects in organizations and institutions. Even in cases where training datasets or models involve anonymous data falling outside the GDPR's scope, there exists a reasonable expectation among data subjects that deems the reuse of such data for powerful AI applications ethically questionable. This erosion of trust is particularly significant in public institutions and sectors like medical research, education, politics, law, welfare, and security. In the current legal landscape, anonymized datasets, including trained models, can be reused without limits, potentially leading to the creation of discriminatory AI models that exacerbate social inequalities. This not only harms corporate reputation, as seen in the big tech industry, but poses a more significant risk when it comes to public institutions (on the importance of public trust, c.f. [80]).

The issue extends to foundational research, heavily reliant on volunteers providing data for the common good. Trust is crucial in ensuring that voluntarily provided data serves its agreed-upon purpose [59]. If such research serves common interests, it should also be in the common interest to create and maintain trust that the voluntarily provided data stays with the purpose that was agreed by the data subjects [68]. Without this trust, willingness to participate in such studies may decline.

Hence, it is part of our shared political responsibility to bring the reuse of trained models and of anonymised training data under legal control. Such a provision is a cornerstone of consistently *enabling* data sharing and AI

for the common good as many AI projects that serve common interests rely on collective data troves to be available. This holds both for data that is voluntarily provided for a specific project, e.g., when patients participate in a clinical study, and for data troves that are made available by law for research purposes, such as when state health insurance services are mandated to provide anonymised patients records for research purposes (see, for example, the Health Data Hub in France and the obligations for public insurances in Germany §§ 303a et seq. Social Code 5; [16]). As explained further in section 4, article 89 GDPR exempts research institutions from the principles of purpose and storage limitation for research data. This special status of research data should be amended by the provision that derived data, such as trained models or anonymised training data sets, must stay within the purpose of research and not be reused for other purposes. Last but not least, when a project involving AI is publicly funded, such as in foundational research, both the models obtained in such projects as well as the data collected for its training should be safeguarded against secondary use that serves commercial interests at the expense of vulnerable groups.

3.3 Controlling informational power asymmetry

Another reason to advocate for protection against the uncontrolled reuse of training data and trained models is the potential amplification of significant forms of informational power asymmetries. These asymmetries arise between those who possess the models and data on one hand, and individuals and society on the other hand (see on power accumulation in relation to AI the diverse debates in [2; 13; 62; 64; 75; 82]). Certain AI models that are trained on personal data have the ability to predict personal information about whatever target individual or case they get applied to ("predictive modeling", for an ethical discussion [55]). In the example above, a model trained on audio recordings and medical records of psychiatric patients could be used to predict psychiatric diseases *for any other individual* for whom audio data is available.

Hence, the mere existence of such a trained model poses a potential threat that doesn't specifically target the data subjects in the training data but applies to *anyone* out there. Of course, the moment the model gets applied to a concrete person to derive an estimation of their disposition towards depression, this data processing (inference by means of the model) falls in the scope of the GDPR. But we argue that already the possession and potential circulation of the model must be regulated because this model comes with the *potential* to be used on anybody and in any context, and already this potential, before it actually manifests in the calculation of personal data about a known individual, is necessary to control.

This is because the potential to derive certain personal data about nearly anybody constitutes a form of informational power. Controlling this power, rather than only its singular manifestations, should be the objective of a preventative regulation. Leaving the problem to defense rights of the target subjects does not prevent the actors from actually obtaining this power in an uncontrolled fashion, if not to speak of the many enforcement deficits with respect to individual defense rights whose violation is hard to prove, often of minor damage (if only the single case is considered) and rarely brought to court [58]. Moreover, there are many realistic situations where the informational power asymmetry effectively coerces individuals to waive their rights, for instance, when job or housing applications are only processed on the condition that the applicants consent to the use of predictive models for the assessment of their applications. Preventing that trained models and training data are even available for reuse in these contexts is the aim of our proposal.

3.4 Privacy violations towards future target individuals

Purpose limitation for models is not only essential to address the betrayal of trust by the subjects of training data and the risk of accumulating power asymmetry. It is also bolstered by ethical considerations related to potential target individuals of reused AI models. One of the fundamental ethical values that is at stake in such reuse scenarios is privacy. For example, the model discussed earlier can be utilized to evaluate any individual and determine potential psychiatric disorders based on an audio recording. Such an estimation of personal information (in this case even sensitive medical information) about a target individual may result in a new form of privacy infringement (see the concepts of "inferential privacy" [40] and "predictive privacy" [55], and on predictions and privacy: [34]). The novelty about this infringement lies in the fact that private information can be *estimated* by means of profiling and pattern matching AI models; thus, the violation of privacy happens by means of derived information and not through, for instance, a re-identification in anonymised data, or a data leak (cf. [84]).

Privacy infringements through predicted information pose a systemic risk to individual rights and democratic societies that is particularly induced by uncontrolled reuse of AI models in other contexts and for altered purposes. The systemic nature of this risk means that it could affect *anyone* if only the input data to the model (an audio recording in our example) is available. This is due to the fact that a privacy infringement on a target individual T is enabled by the model that was trained on the data about individuals $X_1 - X_n$. Crucially, T does not need to be among the n training data subjects. This means that the privacy

decisions of individuals $X_1 - X_n$, in this case, to participate in the medical research project to improve psychiatric diagnosis, has implications for the privacy level that is guaranteed to *any other* individual on whom the model could potentially be used (see for this collective aspect of privacy [67; 43; 54; 45]) as long as we do not implement appropriate safeguards such as purpose limitation for models. The collective nature of privacy that is apparent in this scenario has long been debated in scholarship on privacy and anti-discrimination, such as in relation to profiling [33; 34; 42], in the debate on "group privacy" [77; 50; 22; 30; 40], in the debates on predictive privacy [55; 54; 57], with respect to a "right to reasonable inferences" [84], as a limit of individualism in data protection [81; 6], and in relation to privacy as contextual integrity [73; 61].

Our proposition of purpose limitation for models and training data seems to be an apt solution to this collective privacy problem that has for so long been overlooked in the individualistic framing of privacy as self-control [6]. Potential privacy violations by trained models are a systemic risk as they could affect any target individual. Handling this risk must therefore be separated from the individual privacy rights of the data subjects in the training data. The risk pertains to the trained model itself, as creating a model from training data is equivalent to creating a capability to derive about *any* other individual the personal information that is only known about *some* individuals whose data is used for the training.

4 PURPOSE LIMITATION IN DATA PROTECTION LAW

4.1 Goal and History

On paper, the purpose limitation principle is a core element of data protection law [21; 28], cf. art. 8 (2) ECFR (see [25, 109 et seq.] for the constitutional background), art. 5 (1 b) GDPR; a similar principle can already be found among the Fair Information Practice Principles (FIPPS) from 1973 (no. 3: "There must be a way for a person to prevent information about the person that was obtained for one purpose from being used or made available for other purposes without the person's consent"). Purpose limitation means that data controllers are obligated to define the purpose for data collection no later than at the point of collecting personal data, and they are restricted from processing the data in any way that deviates from the initially stated purpose. These purposes shall be specific, explicit and legitimate to define the aim and goal of the data processing. In essence, the main goal of purpose limitation is to protect the data subject and to enable the controllability of further data processing and its compliance with data protection law [65]. Data subjects shall be enabled to make informed choices about which actors process their data and for what purposes.

Hence, the purpose limitation principle legitimizes data processing and serves as the reference point for assessing its necessity, appropriateness, completeness, and duration. It has a dichotomous structure with a temporal element: a purpose is first established, to which subsequent data processing activities are then bound. Notably, the purpose limitation principle comes with specific requirements of accountability since it does not only obligate the processor who collected the data to consider the specific purposes of their own data processing, but also to transmit these purposes to any secondary processor, cf. art. 19 GDPR.

In the reasoning behind the GDPR, the purpose limitation principle reacts to the fact that once data has been collected and stored, it could in theory be used for *any* purpose, thus repeatedly infringing the right to data protection and the right to informational self-determination of the data subjects (critical in the context of big data [32]). To limit these potential infringements of the data subject's rights, it is not sufficient to merely regulate the admissibility of certain types of data processing for certain types of controllers through provisions of permission; rather, it is the determination of the processing purpose that is specific to the affected individual and the particular matter at hand which limits the processing possibilities to a scope that is legitimate, comprehensible for the affected individual, and verifiable for the supervisory authorities. It ensures transparency and fairness in the handling of personal data, and also provides a clear expectation to data subjects about how their data will be used.

4.2 Why the purpose limitation Principle of the GDPR is insufficient in the context of Big Data and AI

As it turns out in practice, the GDPR's purpose limitation principle is rather toothless ("forgotten" [37], "enormous disconnect between law and reality" [39, p. 256]), for multiple reasons. First, alongside the requirement of lawful data processing, purpose limitation often lacks independent significance. This is primarily because after a certain time delay, the risks associated with secondary usage by other processors or subsequent users tend to be neglected during assessments. Second, the purpose limitation requirement is watered down if data is repeatedly re-processed over several stages since, per the GDPR, the principle is not strictly binding to the original purpose, but the secondary data use has to be *compatible* with the original purpose; cf. art. 5(b). Although art. 6(4) GDPR concretizes the requirement of "compatibility", the criteria for this are assessed by the responsible processors themselves, and these criteria tend to be somewhat insubstantial and arbitrary. Coupled with the fact that an assessment of compatibility is not required if the secondary purpose can be based on consent, art. 6 (4) GDPR opens up numerous possibilities for

further processing. This exception is particularly problematic since consent is an inadequate normative category in times of digitalization and big data [18; 31; 44, p. 222; 69; 86]. Third, the principle of purpose limitation becomes unenforceable, or at the very least, untraceable, in scenarios involving a multitude of actors and vast quantities of data such as in the case of big data and training data for machine learning models [83]. In such cases, where data subjects are no longer identifiable, normative categories like "personal data" effectively become obsolete (cf. ECJ C-252/21). This shows that individual rights are not able to break the power asymmetries between powerful players, such as global digital companies, and affected data subjects. Neither is establishing only individual rights therefore a promising approach to AI regulation, because the same powerful players are involved: all successful AI companies have considerable data power, as all popular and successful AI applications have so far relied on extremely large databases. Systemic solutions are therefore required to ensure the accountability of the actors profiting from the massive data extraction that is currently evident in the training of LLMs.

It has been discussed that AI technologies provide unprecedented opportunities for the secondary use of data, including sensitive data such as health data [47; 11; 19; 12]. The example of the secondary use of anonymized data from the UK Biobank [17]) shows that the purpose limitation principle of the GDPR does not provide sufficient protection. This is due to the fundamental fact that anonymization of (training) data breaks the purpose limitation of that data. Furthermore, it is widely accepted that the purpose limitation principle does not prevent health data to be used for big data analytics, as the collection of this data is often based on the "broad consent" of the data subjects, i.e., consent for multiple potential processing purposes [5; 23; 29; 49], see also recital 33 GDPR. Arguments in favor of the permissibility of broad consent are based on the fact that the societal benefits of certain medical research outweigh the data protection rights of the data subjects [29]. This balancing of interests is not transferable to the secondary use of training datasets or trained models, especially when this reuse serves private interests rather than the public good. In these cases, the benefits for the public good do not prima facie outweigh the limitations of data protection for the data subjects concerned or the risks of discrimination against people subject to the secondary use of models. Rather, a few actors benefit from originally useful models for supra-individual purposes.

In addition to the systematic deficiency of the GDPR purpose limitation principle concerning anonymous data, the purpose limitation principle is also not sufficiently enforced even in cases where personal data are processed. The data processing chain in the life cycle of models consists of the three steps of training, storage, and application of the model (in this analysis we follow [56]).

(1) In the first steps, data for the training of the model is collected. In extremely large databases, it is impossible to distinguish between the legal categories of personal and non-personal data, as illustrated by the example of ChatGPT [70]. (2) The second step is the storage of the trained model. The model data (calibrated weights and internal parameters) differ from the training data. If state-of-the-art anonymization techniques such as differential privacy and federated machine learning are used during training, the model data is anonymous even if the original training data is not. (3) Following this, the third step is the application of the model in which it generates output. This output may again be personal data, but it is no longer subject to the purpose limitation of the *collection of the training data*, as that data had been anonymized in the second step. In addition, there are different data subjects involved in the first step (training) and the third step (application). Personal information about any individual X might be inferred from applying the model, although X is not in the training data. This means that the objective of purpose limitation, which is to give the individual data subject control over the processing of *their* data, can no longer be achieved, as the purpose for which the training data was collected is not linked to the third party on which the model might later be applied.

Another reason why the purpose limitation principle of the GDPR is insufficient to regulate the risks arising from trained models is the fact that the rightfully standardized privileges for certain purposes of data processing are undermined by unregulated secondary data use. Art. 89 (1) GDPR names privileged purposes of data processing such as archiving purposes in the public interest, scientific or historical research, or (public) statistical purposes. The obvious aim of this norm is to prioritize the named purposes, which benefit the public, over other data processings. For this reason, e.g. exceptions to the rights of the data subjects under art. 15, 16, 18, and 21 GDPR apply, cf. art. 89 (no 2, 3). These exemptions also include purpose limitation: for privileged purposes (art. 89 GDPR), it is assumed that they are compatible with the original purpose, art. 5 (1 b) GDPR, art. 6 (4) GDPR. Despite these exceptions, when it comes to research and policy making, some recent contributions challenge the validity of purpose limitation [47].

4.3 Transferring purpose limitation to purpose limitation for models

In this section, we shall compare purpose limitation in data protection with the proposed purpose limitation for models and training data (see section 3.1) in terms of the

risks to which they respond and in terms of the objectives that motivate the respective provision.

For the risks to which they respond, both principles are similar: Purpose limitation in data protection law is meant to prevent data, once it is collected, to be stored and further processed for an unlimited number of purposes. Similarly, purpose limitation for AI models seeks to control the purposes for which trained models and training datasets can be reused.

Regarding the objectives they intend to achieve, we will analyze (1) accountability, (2) supervision and (3) limiting harm. (1) First, both concepts aim at establishing accountability. In data protection law, accountability means that the entity which collects and processes the data is responsible for transmitting the purposes to subsequent processors. In the case of trained models, purpose limitation shall establish accountability of those who subsequently reuse, modify, or transfer the model or the training data: As was discussed in section 3.1 and in reference to Figure 1, accountability in the case of the reuse of a trained model means that those who reuse a trained model are limited to purposes that match with the purpose for which the model was originally trained (and, consequently, for which the training data was collected). Accountability in the case of the reuse of training data means that those who reuse (anonymized) data for the training of an AI model are obliged to determine the purpose for which that data was originally collected (before anonymization) and are bound to use it as training data only for compatible purposes.

(2) As a second objective, both purpose limitation principles aim at enabling control of supervisory authorities over the data processing or the model. Under the GDPR, the purpose limitation principle facilitates oversight over the appropriateness and necessity of the data processing. For models and training datasets, the requirement to define purposes would allow the supervisory authorities established by the DSA, the AI Act or the GDPR to supervise the training and reuse of models and training datasets, and the processor's compliance with the relevant legal frameworks. This supervision would not be limited in its scope to the processing of personal data and could include, for example, the assessment of whether the secondary use of training data or trained models could be a systemic risk under the DSA or the AIA. Moreover, the role of the supervisory authority in the case of purpose limitation for the training and reuse of models could be designed in such a way that, for the first time, a comprehensive overview of models would be gained (see section 5.3). This could be relevant in cases such as GPT-3, where individual data processing steps are no longer traceable, but the potential impact of the system is enormous. Consequently, in contrast to data protection law, a potential legal implementation of purpose limitation for models should also encompass a documented procedural obligation

(3) There is a third objective connected to our proposal, and in this point purpose limitation for AI models diverges from the aims of purpose limitation in data protection law. Purpose limitation in data protection intends to safeguard the individuals' right to informal self-determination, which means that data subjects should have control over who can process information about them (and for what purposes) by setting limits on how data processors may use and reuse their personal data. This empowerment of each individual *with respect to their own data* serves to prevent data uses that the data subject considers as unexpected, inappropriate, potentially harmful, or otherwise objectionable [66].

With purpose limitation for models, in contrast, our motivation goes beyond the promotion of *individual* rights such as informational self-determination which is too much focused on the data subject in the training data and the potential consequences of data processing on this same subject. Our objective regarding purpose limitation for models is controlling potential consequences *on others* – individuals other than the data subjects, groups or the society at large. We do assume that it is also in line with the data subjects' interests and expectations that personal data they provide won't be misused to the harm of others or society at large. This is in line with recital 50 of the GDPR, which explicitly refers to "the reasonable expectations of the data subjects based in their relationship with the controller as to their further use" and also with the doctrine of "reasonable expectations of privacy" in US law [15; 72; 74]. However, we argue that such a doctrine of "reasonable expectation" of a person X concerning their data should be extended to the reuse of AI models that were trained from X's data, although they no longer contain personal data about X. To pick up our example from section 3, it cannot generally, be assumed that data subjects donating their personal data to a research project also expect their data to be used to build a model for risk assessment in the insurance industry. Likewise, students whose personal data is processed by their school or university, or clients of healthcare services whose data is processed as part of the healthcare system's operations, cannot be assumed to reasonably expect that their data, or a model trained on that data, will be reused, for instance, for the assessment of job applicants or placement of personalized advertisement.

5 WHY DOES THE AI ACT NOT SUFFICIENTLY REGULATE THE RISK OF SECONDARY DATA USE?

5.1 Context oriented risk classifications and purposes in the AI Act

The AIA follows a risk-based approach that categorizes AI models into four risk classes: unacceptable risk, high

risk, limited risk, and low risk (for a critical analysis see [63]).[5] In Annex III, the AIA lists various contexts of use, the application of which should lead to categorization as a high-risk system. When it comes to secondary data use, the AIA states that it cannot be considered as providing legal grounds for processing personal data, especially special categories of data, cf. recital 41 (eventually there are some exceptions for regulatory sandboxes [47]). Even not specifically written in the text, the criteria of whether a system falls under this category depends on the intended purpose, cf. art. 6 (2) AIA, since there is no other yardstick to evaluate how an AI system is used before being placed on the market. Intended purpose means "the use for which an AI system is intended by the provider, including the specific context and conditions of use, as specified in the information supplied by the provider in the instructions for use, promotion or sales materials and statements, as well as in the technical documentation", art. 3 (12).

We are critical of this for the following reasons: the objectives of the AI system may be different from the intended purpose of the AI system in a specific context (cf. recital 6) and not the provider alone, but a supervisory authority should decide based on the potential use cases and context of development and use about the purpose of the system. This is particularly important because providers can also decide for themselves that their system is not a high-risk system, despite the relevant application contexts in Annex III, art. 6 (2a) AIA. In these cases, the providers are obliged to notify the authorities who then shall review and reply within three months, but the notion can "take the form of a one-page summary of the relevant information on the AI system in question, including its intended purposes", recital 32a. Additionally, there is a risk that providers could pretend to have a specific purpose to avoid falling into the high-risk category. It is therefore all the more important to document purposes explicitly and to be able to check them against the actual context of use outside the risk classifications.

Yet, according to the AI Act, an AI system has to undergo a new conformity assessment when the intended purpose of the system changes, recital 66. We argue that it is not sufficient to document purposes of models as one subpoint of the risk assessment, but rather they shall be publicly registered and documented, especially when the system is reused by another deployer. Only if the system is classified as high risk do the obligations of art. 10(2)(aa) AIA apply, which in its current version requires transparency measures about the original purposes of the data collection. This mere transparency requirement, however, does not mean that the intended purpose, which is considered high-risk under the AIA, must be compatible with the purposes of the original data collection.

Furthermore, the risk classification requirement of Art. 6 (1) does not apply at the time of training, but when the system is placed on the market or put into use. With regard to the applications that constitute a high-risk classification according to Annex III, the time is not further specified by the wording of Art. 6 (2) AIA. The training of a model or the re-use can also take place before the market placement or use according to art. 3 (9, 11) AIA. Although art. 9 AIA refers to the whole lifecycle of a model regarding the requirements of an ongoing risk assessment, these testing requirements are only applicable when deemed suitable for achieving the intended purpose of the AI system, art. 9 (6) AIA.

The critical areas of application of Annex III do not provide for the training or secondary use of models as a specific risk. Medical applications are missing in order not to impede research in this area, but it is precisely here that a particular risk of abusive secondary use arises. Therefore, the obligations of the AIA should not be extended to research projects, but the purposes of the trained models should be registered and limited according to our proposal after the model has been trained. Overall, the AIA regulates high-risk use cases but not the transfer of a (high-risk) model to another provider or use case; meaning the selling or transfer of the model itself is not considered a risk under the AIA; as argued before, this creates a considerable gap of accountability.

5.2 General purpose AI and open source exceptions

After the spectacular market launch of ChatGPT and an increasing relevance of open-source and foundational models, negotiations on the AI Act over the year 2023 sought to include more specific provisions for generative AI and open-source models. First, minimum standards for generative models were introduced, although they have already been criticized as extremely weak and falling short of the industry's voluntary commitments as they provide for mere transparency and limited copyright requirements [26]. Second, as it was announced in the initial press release, "providers of free and open source AI models" will be exempted from the scope of the AI Act unless they are general purpose AI models that pose a "systemic risk" or are trained with computing power of more than 10^{25} FLOPS (floating-point operations).[6] Here, general-purpose

[5] At the time of submitting this paper, the final text of the agreement on the AIA was not yet publicly known. In our analysis, we refer to the information available here: https://ec.europa.eu/commission/presscorner/api/files/document/print/en/qanda_21_1683/QANDA_21_1683_EN.pdf; https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6473; https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai.

[6] https://ec.europa.eu/commission/presscorner/api/files/document/print/en/qanda_21_1683/QANDA_21_1683_EN.pdf

AI means an AI system that can be used in and adapted to a wide range of applications for which it was not intentionally and specifically designed, Art. 3 (1d). The term "open-source model", in contrast, is so far not explicitly defined in the AI Act. We assume that the term refers to the free publication of a trained model's internal parameters (weights etc.), that is, to the public release of the trained model for reuse by anyone and for any purpose.

In these provisions, the concept of a "systemic risk" plays a central role. On the one hand, regulation of general-purpose models hinges on their positing a systemic risk, on the other hand, a broad exemption for open-sourced models is made as long as they are *not* general purpose and do *not* pose a systemic risk. A systemic risk is acknowledged in the AIA if the training of a model exceeds the computing power threshold of 10^{25} FLOPS: providers are then mandated to assess and mitigate risks, report serious incidents, conduct state-of-the-art tests and model evaluations, ensure cybersecurity and provide information on the energy consumption of their models. Additionally "they are asked to engage with the European AI Office to draw up Codes of Conduct as the central tool to detail out the rules in cooperation with other experts. A scientific panel will play a central role in overseeing general-purpose AI models." [10]

The clear figures of the computing power threshold seem to be easy to verify and ensure legal certainty; this provision thus standardizes a rigid, numerical criterion. The term "systemic risk", in contrast, suggests an assessment involving the evaluation and consideration of different societal factors and interests, including normative, ethical, social and societal implications of AI, which the AIA indeed claims to address. It is not plausible how such a complex assessment can be reduced to the numerical criterion of computing power (see for a similar criticism of the number of users as a regulatory threshold: [71]). This approach comes with the additional problem that most models currently on the market do not cross this threshold (Bard, GPT 3.5, maybe Gemini), although they can still pose significant risks that are arguably even systemic [26].

From an abstract perspective, open-sourcing a trained model significantly amplifies the risks of uncontrolled secondary utilization, as anyone can adopt the open-sourced model for various purposes. Many scholarly and political debates over the past years have highlighted systemic risks such as unfair discrimination, biases, social inequality, and privacy infringements in AI systems operating below the computing power threshold of 10^{25} FLOPS. The broad exemption of open-source models below this threshold in the AIA implies that highly problematic AI applications, which have been the focus of these debates, could escape regulation if their models were open-sourced. It is by no means plausible how the risk-based treatment of AI should validly be interrupted depending on whether a model is open or closed source. With this provision, the current version of the AI Act not only fails to address concerns related to the open-ended reuse of open-source models for other purposes, but it might contribute to additional risks. The extensive exemptions for open-source models from the AI Act's provisions create an incentive for developers to make their models publicly available, avoiding the high compliance costs. This was evident in the case of Mixtral 8x7B, an extremely powerful AI model developed by a French startup, which was promptly released as open source following the public announcement of the exemption rule in December 2023. The risks stemming from the uncontrolled expansion of the purposes for which an AI model is trained are more pronounced in the case of open-source models due to their wider distribution. This distribution is challenging to control, making it difficult, for instance, to retract or revise biased or otherwise flawed models.

In light of these shortcomings in how the AI Act handles open-source models concerning their risk of purpose creep, our proposed purpose limitation for models (see section 3.1; [38]) would fill the gap as it also applies to open-source models. It is important to point out that purpose limitation does *not* prevent models from being freely published, but it regulates how they may be utilized. According to our purpose limitation principle, the creators of a model (irrespective of whether this model will be open-source or not) would have to state the purpose of the model *ex ante*, in line with the purposes for which the training data was collected. Anyone using, modifying or re-publishing such a model would then be bound to the stated purpose. Hence, our proposal of a purpose limitation for models actually contributes to *enabling* open-source for trained models in an ethically and politically viable fashion as it introduces relevant provisions to prevent the risks that come with uncontrolled reuse.

5.3 Governance: Registration and supervisory authority

To address the shortcomings of purpose limitation in data protection law and mitigate the risk of enforcement deficits, we propose to implement purpose limitation for models in combination with a tiered system of procedural obligations. At the initial stage, anyone who trains or reuses an AI model is subject to two types of obligations: one retrospective (backward-looking) and the other prospective (forward-looking):

(1.1) According to what we termed "backward accountability" above (see section 3 and figure 1), the entity is obliged to ensure that the training data is compatible with the purpose for which the model is being trained. If (anonymized) training data was obtained from elsewhere, the purpose for which it was originally collected is to be determined.

(1.2) An *ex ante* risk assessment is to be made concerning the potential secondary use cases of the model that is to be trained. This risk assessment specifically includes use cases that are *not* intended by the entity in question.

(2) If the *ex ante* risk assessment identifies high-risk secondary use cases among the potential uses of a model, the entity training the model must then proceed to the next stage in the tiered system of obligations: registering the model with a central authority. This is regularly to be expected for models that are either exceptionally large or carry particularly high risks in the assessment of the AI Act. Specifically, this applies to models that permit secondary use cases listed in Annex III of the AIA, that are capable of causing systemic risks under the DSA, or that could impact a significant number of people, for example, through integration in office applications, as well as powerful open-source models. The models that allow for such high-risk secondary uses would then be documented in a publicly accessible database.

These obligations also hold for developers or organizations intending to release their model as open source (see section 5.2). If in this case, the *ex ante* risk assessment reveals a potential high-risk secondary use case and, in consequence, a registration of the model with the supervisory authority is mandatory, a decision of this supervisory authority must be awaited as to whether the model is permitted for open-source publication. As an alternative, the authority could mandate the creator to share the model on a "hosted access" scheme. Hosted access has already been mentioned in the debates of open-source models, and is a scheme where a model would not be published, but API access would be made available, cf. [26]. If an open-source model is published (either as it falls below the registration threshold or after the authority's permission), the *ex ante* risk assessment must be published together with the model. We argue that this enables better enforcement of purpose limitation, which is particularly relevant in high-risk cases. With hosted access, the creator of the model could be held responsible for providing access only to certain actors and certain application contexts that are compatible with the model's purpose. Hosted access would prevent the trained model from circulating in an uncontrolled way, while the model could still be opened to independent scrutiny for systemic risks by independent researchers.

In terms of competences and procedures, integrating purpose limitation for models with the governance structures of the AIA could be an available approach. Given the regulatory scope of the AIA, it is highly probable that the scope of purpose limitation for models will intersect with some high-risk systems. The future governance structures of the AIA could therefore be used accordingly for the implementation of purpose limitation for models: during the registration of high-risk systems in the EU-wide database (art. 51, 60 AIA), the purposes of the trained models could also be documented. Additionally, the AI Office at the Commission, as the designated supervisory authority, would have the responsibility to oversee adherence to the registration requirements, ensuring that secondary uses align with and are compatible with the registered purposes.

6 CONCLUSION AND OUTLOOK

Our paper addressed a critical regulatory gap in the EU's digital legislation, including the proposed AI Act and the GDPR: the risk of secondary use of trained models and anonymized training datasets. As a solution, we introduced what we term purpose limitation for training and reusing AI models. In brief, this approach mandates that those training AI models define the intended purpose and restrict the use of the model to this stated purpose.

As such, purpose limitation is a well-known concept in data protection law which, despite its important objectives, has so far played a subordinate, almost overlooked role, and largely failed to prevent dangerous secondary uses of trained AI models or training data. Our proposal for an update of the purpose limitation principle seeks to overcome these shortcomings, which are, as we discussed, both conceptual and enforcement deficits. The main ingredient in our recipe is the shift of regulatory focus from training data to trained models represented through "model data", which is a dataset in its own right, completely distinct from the training data. This shift in perspective involves zooming out beyond the regulatory aspects of data processing or the introduction of an AI system to the market, considering the lifecycle of an AI system instead.

We developed this discussion by looking at ethical obligations towards the data subjects in the training data, towards data subjects on which an AI model could potentially be applied, towards society at large which faces ever-increasing power asymmetries from tech companies. A subsequent comparison of purpose limitation for models with the regulatory regimes of the GDPR and the AIA revealed both systematic and enforcement-related reasons why the risk of unaccounted secondary use of trained models is not sufficiently addressed in current and future EU legislation. Concerning the AI Act, we contend that its regulatory framework fails to acknowledge the risk of abusive secondary use because the self-conducted risk assessment by providers may result in an ambiguous specification of the models' purposes. The real risk associated with an AI model is not diminished by the model being open-sourced; the AI Act's exemption for a large class of open-source models is therefore counter-productive in addressing the societal risks that stem from open-ended reuse of open-source models. Based on this, we argue that a purpose limitation for models could help promote open source in an ethical and legally compliant way.

We identified the training of a model as the pivotal stage where the risks and potential harms of AI originate. This includes scenarios where a dataset initially used in one context is repurposed for training a different model, as well as situations where a trained model is subsequently utilized secondarily by another entity. Our proposal of a purpose limitation for trained models and training data is motivated by the three main objectives of (1) enabling accountability of all the processors in a potential chain of reuses of the same model or training data; (2) enabling supervision by a public authority such as the supervisory authorities established by the DSA or the future AI Act; and (3) limiting both collective and individual harms. The latter point particularly emphasizes the need to control potential implications of AI models on individuals that are not in the training data. Trained models can be applied on *anybody*, potentially causing the discrimination, infringements of fundamental rights, or unfair treatment of groups and harmful effects on society at large.

In the spirit of a progressive interdisciplinary discussion, this paper introduced the conceptual foundation of a novel regulatory approach to govern trained models. To translate the proposal of a purpose limitation for models into effective regulation, the following questions need to be clarified in future research:

A coherent integration of the concept into the existing EU digital legislation and the clarification of the relationships to existing legal acts is required – a task that the AI Act has not taken on so far. Particularly, it needs to be analyzed whether and to what extent purpose limitation for models could be implemented in the governance structures of the AIA. This includes whether the AI Office at the Commission would be a suitable oversight body and how this would relate to the national competencies of the member states. Further questions are how purposes can be exactly documented, at which time purposes must be indicated (for example, at training of a model, or placement on the market? – the AIA refers to the latter). In addition, it should be examined whether the regulatory sandboxes provided for in the AIA (Art. 53 ff. AIA) can be used to determine purposes or risk assessments.

Regarding the definition of purposes, it needs to be examined whether a definitive list of desirable and prohibited purposes should be legally implemented like in the proposal for a regulation about a European Health Data Space (EHDS – Com2022/197-final). The "positive list" of art. 34 includes purposes which are activities in the public interest like public health surveillance and protection against cross-border threats (a), supporting public sector bodies (b), producing statistics (d), education or teaching (e). In contrast, art. 35 excludes purposes like taking decisions concerning natural persons or groups of natural persons to exclude them from the benefit of an insurance contract. It must be determined how exactly purposes are to be defined and whether secondary purposes must coincide with the original purposes or only be compatible. An argument against adopting the criteria for determining compatibility from Art. 6 (4) of the GDPR is that they are formulated much too vaguely and allow for far-reaching deviations. A reference to the level of specification of the EHDS seems more promising.

REFERENCES

[1] Carmen Molina Acosta and Lisa Weiner. "Artificial Intelligence Could Soon Diagnose Illness Based on the Sound of Your Voice". In: *NPR. Health* (Oct. 10, 2022). URL: https://www.npr.org/2022/10/10/1127181418/ai-app-voice-diagnose-disease (visited on 10/14/2023).

[2] Louise Amoore. *Cloud Ethics: Algorithms and the Attributes of Ourselves and Others*. Duke University Press, May 1, 2020. 143 pp. ISBN: 978-1-4780-0927-6. Google Books: tVbZDwAAQBAJ.

[3] Max Bain et al. *Frozen in Time: A Joint Video and Image Encoder for End-to-End Retrieval*. May 13, 2022. DOI: 10.48550/arXiv.2104.00650. arXiv: 2104.00650 [cs]. URL: http://arxiv.org/abs/2104.00650 (visited on 01/08/2024). preprint.

[4] Andy Baio. *AI Data Laundering: How Academic and Nonprofit Researchers Shield Tech Companies from Accountability*. Waxy.org. Sept. 30, 2022. URL: https://waxy.org/2022/09/ai-data-laundering-how-academic-and-nonprofit-researchers-shield-tech-companies-from-accountability/ (visited on 01/08/2024).

[5] Gaia Barazzetti et al. "Broad Consent in Practice: Lessons Learned from a Hospital-Based Biobank for Prospective Research on Genomic and Medical Data". In: *European Journal of Human Genetics* 28.7 (7 July 2020), pp. 915–924. ISSN: 1476-5438. DOI: 10.1038/s41431-020-0585-0. URL: https://www.nature.com/articles/s41431-020-0585-0 (visited on 01/08/2024).

[6] Lemi Baruh and Mihaela Popescu. "Big Data Analytics and the Limits of Privacy Self-Management". In: *New Media & Society* 19.4 (Apr. 1, 2017), pp. 579–596. ISSN: 1461-4448. DOI: 10.1177/1461444815614001. URL: https://doi.org/10.1177/1461444815614001 (visited on 04/02/2022).

[7] Emily M. Bender et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? " In: *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '21. New York, NY, USA: Association for Computing Machinery, Mar. 1, 2021, pp. 610–623. ISBN: 978-1-4503-8309-7. DOI: 10.1145/3442188.3445922. URL:

https://dl.acm.org/doi/10.1145/3442188.3445922 (visited on 01/12/2024).

[8] Joy Buolamwini and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification". In: *Conference on Fairness, Accountability and Transparency*. Conference on Fairness, Accountability and Transparency. PMLR, Jan. 21, 2018, pp. 77–91. URL: http://proceedings.mlr.press/v81/buolamwini18a.html (visited on 09/21/2020).

[9] Mark Coeckelbergh. *AI Ethics*. Cambridge, MA: The MIT Press, 2020. ISBN: 978-0-262-53819-0.

[10] European Commission. *Artificial Intelligence – Questions and Answers*. European Commission, 12.12.2023, updated 14.12.2023.

[11] Marcelo Corrales Compagnucci et al. "Technology-Driven Disruption of Healthcare and 'UI Layer' Privacy-by-Design". In: *AI in eHealth: Human Autonomy, Data Governance and Privacy in Healthcare*. Ed. by Till Bärnighausen et al. Cambridge Bioethics and Law. Cambridge: Cambridge University Press, 2022, pp. 19–67. ISBN: 978-1-108-83096-6. DOI: 10.1017/9781108921923.005. URL: https://www.cambridge.org/core/product/43DED5E535EFD1B70EFCFBE543F5902F.

[12] P. Coorevits et al. "Electronic Health Records: New Opportunities for Clinical Research". In: *Journal of Internal Medicine* 274.6 (2013), pp. 547–560. ISSN: 1365-2796. DOI: 10.1111/joim.12119. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/joim.12119 (visited on 01/08/2024).

[13] Nick Couldry and Ulises Ali Mejias. *The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism*. Culture and Economic Life. Stanford, California: Stanford University Press, 2019. 323 pp. ISBN: 978-1-5036-0366-0 978-1-5036-0974-7.

[14] Kate Crawford. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press, 2021. 327 pp. ISBN: 978-0-300-20957-0.

[15] Brandon T. Crowther. "(Un)Reasonable Expectation of Digital Privacy Comment". In: *Brigham Young University Law Review* 2012.1 (2012), pp. 343–370. URL: https://heinonline.org/HOL/P?h=hein.journals/byulr2012&i=352 (visited on 01/16/2024).

[16] Marc Cuggia and Stéphanie Combes. "The French Health Data Hub and the German Medical Informatics Initiatives: Two National Projects to Promote Data Sharing in Healthcare". In: *Yearbook of Medical Informatics* 28.1 (Aug. 2019), pp. 195–202. ISSN: 0943-4747, 2364-0502. DOI: 10.1055/s-0039-1677917. URL: http://www.thieme-connect.de/DOI/DOI?10.1055/s-0039-1677917 (visited on 01/22/2024).

[17] Shanti Das. "Private UK Health Data Donated for Medical Research Shared with Insurance Companies". In: *The Observer. Technology* (Nov. 12, 2023). ISSN: 0029-7712. URL: https://www.theguardian.com/technology/2023/nov/12/private-uk-health-data-donated-medical-research-shared-insurance-companies (visited on 01/06/2024).

[18] Sourya Joyee De and Abdessamad Imine. "Consent for Targeted Advertising: The Case of Facebook". In: *AI & SOCIETY* 35.4 (Dec. 1, 2020), pp. 1055–1064. ISSN: 1435-5655. DOI: 10.1007/s00146-020-00981-5. URL: https://doi.org/10.1007/s00146-020-00981-5 (visited on 01/09/2024).

[19] Francisco de Arriba-Pérez, Manuel Caeiro-Rodríguez, and Juan M. Santos-Gago. "Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios". In: *Sensors (Basel, Switzerland)* 16.9 (Sept. 21, 2016), p. 1538. ISSN: 1424-8220. DOI: 10.3390/s16091538. pmid: 27657081.

[20] Theodoros Evgeniou, David R. Hardoon, and Anton Ovchinnikov. "What Happens When AI Is Used to Set Grades?" In: *Harvard Business Review* (Aug. 13, 2020). ISSN: 0017-8012. URL: https://hbr.org/2020/08/what-happens-when-ai-is-used-to-set-grades (visited on 01/09/2024).

[21] Michele Finck and Asia J. Biega. "Reviving Purpose Limitation and Data Minimisation in Data-Driven Systems". In: *Technology and Regulation* 2021 (Dec. 7, 2021), pp. 44–61. ISSN: 2666-139X. DOI: 10.26116/techreg.2021.004. URL: https://techreg.org/article/view/10986 (visited on 07/10/2023).

[22] Luciano Floridi. "Open Data, Data Protection, and Group Privacy". In: *Philosophy & Technology* 27.1 (Mar. 1, 2014), pp. 1–3. ISSN: 2210-5441. DOI: 10.1007/s13347-014-0157-8. URL: https://doi.org/10.1007/s13347-014-0157-8 (visited on 06/27/2021).

[23] Nikolaus Forgó, Stefanie Hänold, and Benjamin Schütze. "The Principle of Purpose Limitation and Big Data". In: *New Technology, Big Data and the Law*. Ed. by Marcelo Corrales, Mark Fenwick, and Nikolaus Forgó. Perspectives in Law, Business and Innovation. Singapore: Springer, 2017, pp. 17–42. ISBN: 978-981-10-5038-1. DOI: 10.1007/978-981-10-5038-1_2. URL: https://doi.org/10.1007/978-981-10-5038-1_2 (visited on 01/08/2024).

[24] Tarleton Gillespie. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press, June 26, 2018. 297 pp. ISBN: 978-0-300-23502-9. Google Books: cOJgDwAAQBAJ.

[25] Maximilian von Grafenstein. *The Principle of Purpose Limitation in Data Protection Laws*. Nomos Verlagsgesellschaft mbH & Co. KG, Apr. 9, 2018. ISBN:

978-3-8487-4897-6 978-3-8452-9084-3. DOI: 10.5771/9783845290843. URL: https://www.nomos-elibrary.de/10.5771/9783845290843/the-principle-of-purpose-limitation-in-data-protection-laws?hitid=00&search-click&page=1 (visited on 01/07/2024).

[26] Philipp Hacker. "What's Missing from the EU AI Act: Addressing the Four Key Challenges of Large Language Models". In: *Verfassungsblog* (Dec. 13, 2023). DOI: 10.17176/20231214-111133-0. URL: https://verfassungsblog.de/whats-missing-from-the-eu-ai-act/ (visited on 01/06/2024).

[27] Philipp Hacker, Andreas Engel, and Marco Mauer. "Regulating ChatGPT and Other Large Generative AI Models". In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '23. New York, NY, USA: Association for Computing Machinery, June 12, 2023, pp. 1112–1123. ISBN: 9798400701924. DOI: 10.1145/3593013.3594067. URL: https://dl.acm.org/doi/10.1145/3593013.3594067 (visited on 07/13/2023).

[28] Isabel Hahn. "Purpose Limitation in the Time of Data Power: Is There a Way Forward?" In: *European Data Protection Law Review* 7.1 (2021), pp. 31–44. ISSN: 2364284X. DOI: 10.21552/edpl/2021/1/7. URL: https://edpl.lexxion.eu/article/EDPL/2021/1/7 (visited on 07/04/2023).

[29] Dara Hallinan. "Broad Consent under the GDPR: An Optimistic Perspective on a Bright Future". In: *Life Sciences, Society and Policy* 16.1 (Jan. 6, 2020), p. 1. ISSN: 2195-7819. DOI: 10.1186/s40504-019-0096-3. URL: https://doi.org/10.1186/s40504-019-0096-3 (visited on 01/08/2024).

[30] Paula Helm. "Group Privacy in Times of Big Data. A Literature Review". In: *Digital Culture & Society* 2.2 (Dec. 1, 2016), pp. 137–152. ISSN: 2364-2122, 2364-2114. DOI: 10.14361/dcs-2016-0209. URL: https://www.degruyter.com/document/doi/10.14361/dcs-2016-0209/html (visited on 08/09/2022).

[31] Mireille Hildebrandt. "Profile Transparency by Design? Re-enabling Double Contingency". In: *Privacy, Due Process and the Computational Turn*. Routledge, 2013. ISBN: 978-0-203-42764-4.

[32] Mireille Hildebrandt. "Slaves to Big Data. Or Are We?" In: *IDP. REVISTA DE INTERNET, DERECHO Y POLÍTICA* 17 (2013), pp. 7–44. DOI: 10.7238/idp.v0i17.1977. URL: https://works.bepress.com/mireille_hildebrandt/52/ (visited on 07/04/2023).

[33] Mireille Hildebrandt. "Who Is Profiling Who? Invisible Visibility". In: *Reinventing Data Protection?* Ed. by Serge Gutwirth et al. Dordrecht: Springer Netherlands, 2009, pp. 239–252. ISBN: 978-1-4020-9497-2 978-1-4020-9498-9. DOI: 10.1007/978-1-4020-9498-9_14. URL: http://link.springer.com/10.1007/978-1-4020-9498-9_14 (visited on 08/18/2022).

[34] Mireille Hildebrandt and Serge Gutwirth, eds. *Profiling the European Citizen: Cross-Disciplinary Perspectives*. New York: Springer, 2008. 373 pp. ISBN: 978-1-4020-6913-0.

[35] Kashmir Hill. "Wrongfully Accused by an Algorithm". In: *The New York Times. Technology* (June 24, 2020). ISSN: 0362-4331. URL: https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html (visited on 01/12/2024).

[36] Shutterstock Inc. *Shutterstock Expands Long-standing Relationship with Meta*. URL: https://www.prnewswire.com/news-releases/shutterstock-expands-long-standing-relationship-with-meta-301719769.html (visited on 01/08/2024).

[37] Catherine Jasserand. "Subsequent Use of GDPR Data for a Law Enforcement Purpose:" in: *European Data Protection Law Review* 4.2 (2018), pp. 152–167. ISSN: 2364284X. DOI: 10.21552/edpl/2018/2/6. URL: https://edpl.lexxion.eu/article/EDPL/2018/2/6 (visited on 07/03/2023).

[38] Bert-Jaap Koops. "The Concept of Function Creep". In: *Law, Innovation and Technology* 13.1 (Jan. 2, 2021), pp. 29–56. ISSN: 1757-9961. DOI: 10.1080/17579961.2021.1898299. URL: https://doi.org/10.1080/17579961.2021.1898299 (visited on 01/22/2024).

[39] Bert-Jaap Koops. "The Trouble with European Data Protection Law". In: *International Data Privacy Law* 4.4 (Nov. 1, 2014), pp. 250–261. ISSN: 2044-3994. DOI: 10.1093/idpl/ipu023. URL: https://doi.org/10.1093/idpl/ipu023 (visited on 07/14/2023).

[40] Michele Loi and Markus Christen. "Two Concepts of Group Privacy". In: *Philosophy & Technology* 33 (2020), pp. 207–224. ISSN: 2210-5433, 2210-5441. DOI: 10.1007/s13347-019-00351-0. URL: http://link.springer.com/10.1007/s13347-019-00351-0 (visited on 12/20/2019).

[41] Daniel M. Low, Kate H. Bentley, and Satrajit S. Ghosh. "Automated Assessment of Psychiatric Disorders Using Speech: A Systematic Review". In: *Laryngoscope Investigative Otolaryngology* 5.1 (Jan. 31, 2020), pp. 96–116. ISSN: 2378-8038. DOI: 10.1002/lio2.354. pmid: 32128436. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7042657/ (visited on 01/19/2024).

[42] Monique Mann and Tobias Matzner. "Challenging Algorithmic Profiling: The Limits of Data Protection and Anti-Discrimination in Responding to Emergent Discrimination". In: *Big Data & Society* 6.2 (July 1, 2019), p. 2053951719895805. ISSN: 2053-9517. DOI: 10.1177/2053951719895805. URL: https://doi.org/10.1177/2053951719895805 (visited on 09/13/2020).

[43] Alessandro Mantelero. "Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection". In: *Computer Law & Security Review* 32.2 (Apr. 2016), pp. 238–255. ISSN: 02673649. DOI: 10.1016/j.clsr.2016.01.014. URL: https://linkinghub.elsevier.com/retrieve/pii/S0267364916300280 (visited on 04/16/2020).

[44] Alessandro Mantelero. "The Future of Consumer Data Protection in the E.U. Re-thinking the "Notice and Consent" Paradigm in the New Era of Predictive Analytics". In: *Computer Law & Security Review* 30.6 (Dec. 1, 2014), pp. 643–660. ISSN: 0267-3649. DOI: 10.1016/j.clsr.2014.09.004. URL: https://www.sciencedirect.com/science/article/pii/S026736491400154X (visited on 01/09/2024).

[45] Tobias Matzner. "Why Privacy Is Not Enough Privacy in the Context of "Ubiquitous Computing" and "Big Data"". In: *Journal of Information, Communication and Ethics in Society* 12.2 (May 6, 2014), pp. 93–106. ISSN: 1477-996X. DOI: 10.1108/JICES-08-2013-0030. URL: https://www.emerald.com/insight/content/doi/10.1108/JICES-08-2013-0030/full/html (visited on 05/24/2021).

[46] Viktor Mayer-Schönberger and Yann Padova. "Regime Change? Enabling Big Data through Europe's New Data Protection Regulation". In: *Science and Technology Law Review* 17.2 (2 May 24, 2016), pp. 315–335. ISSN: 1938-0976. DOI: 10.7916/stlr.v17i2.4007. URL: https://journals.library.columbia.edu/index.php/stlr/article/view/4007 (visited on 07/14/2023).

[47] Janos Meszaros, Jusaku Minari, and Isabelle Huys. "The Future Regulation of Artificial Intelligence Systems in Healthcare Services and Medical Research in the European Union". In: *Frontiers in Genetics* 13 (2022). ISSN: 1664-8021. URL: https://www.frontiersin.org/articles/10.3389/fgene.2022.927721 (visited on 01/08/2024).

[48] Milagros Miceli, Martin Schuessler, and Tianling Yang. *Between Subjectivity and Imposition: Power Dynamics in Data Annotation for Computer Vision*. July 30, 2020. DOI: 10.48550/arXiv.2007.14886. arXiv: 2007.14886 [cs]. URL: http://arxiv.org/abs/2007.14886 (visited on 11/09/2022). preprint.

[49] Rasmus Bjerregaard Mikkelsen et al. "Broad Consent for Biobanks Is Best – Provided It Is Also Deep". In: *BMC Medical Ethics* 20.1 (Oct. 15, 2019), p. 71. ISSN: 1472-6939. DOI: 10.1186/s12910-019-0414-6. URL: https://doi.org/10.1186/s12910-019-0414-6 (visited on 01/08/2024).

[50] Brent Mittelstadt. "From Individual to Group Privacy in Big Data Analytics". In: *Philosophy & Technology* 30.4 (2017), pp. 475–494. ISSN: 2210-5433, 2210-5441. DOI: 10.1007/s13347-017-0253-7. URL: http://link.springer.com/10.1007/s13347-017-0253-7 (visited on 12/20/2019).

[51] Lokke Moerel and Corien Prins. *Privacy for the Homo Digitalis: Proposal for a New Regulatory Framework for Data Protection in the Light of Big Data and the Internet of Things*. May 25, 2016. DOI: 10.2139/ssrn.2784123. URL: https://papers.ssrn.com/abstract=2784123 (visited on 07/14/2023). preprint.

[52] Rainer Mühlhoff. "Automatisierte Ungleichheit: Ethik der Künstlichen Intelligenz in der biopolitischen Wende des Digitalen Kapitalismus". In: *Deutsche Zeitschrift für Philosophie* 68.6 (Dec. 16, 2020), pp. 867–890. ISSN: 0012-1045, 2192-1482. DOI: 10.1515/dzph-2020-0059. URL: https://www.degruyter.com/view/journals/dzph/68/6/article-p867.xml (visited on 12/19/2020).

[53] Rainer Mühlhoff. "Human-Aided Artificial Intelligence: Or, How to Run Large Computations in Human Brains? Toward a Media Sociology of Machine Learning". In: *New Media & Society* 22.10 (Oct. 2020), pp. 1868–1884. ISSN: 1461-4448, 1461-7315. DOI: 10.1177/1461444819885334. URL: http://journals.sagepub.com/doi/10.1177/1461444819885334 (visited on 02/21/2021).

[54] Rainer Mühlhoff. "Predictive Privacy: Collective Data Protection in Times of AI and Big Data". In: *Big Data & Society* (2023), pp. 1–14. DOI: 10.1177/20539517231166886.

[55] Rainer Mühlhoff. "Predictive Privacy: Towards an Applied Ethics of Data Analytics". In: *Ethics and Information Technology* 23 (2021), pp. 675–690. ISSN: 1388-1957, 1572-8439. DOI: 10.1007/s10676-021-09606-x. URL: https://link.springer.com/10.1007/s10676-021-09606-x (visited on 09/12/2021).

[56] Rainer Mühlhoff and Hannah Ruschemeier. "Democratising AI via Purpose Limitation for Models". In: (July 19, 2023). URL: https://dx.doi.org/10.2139/ssrn.4599869. preprint.

[57] Rainer Mühlhoff and Hannah Ruschemeier. "Predictive Analytics and the Collective Dimension of Data Protection". In: *Law, Innovation and Technology forthcoming* (2024 forthcoming). URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4614268.

[58] Stephen Mulders. "Collective Damages for GDPR Breaches: A Feasible Solution for the GDPR Enforcement Deficit?" In: *European Data Protection Law Review* 8.4 (2022), pp. 493–506. ISSN: 2364284X. DOI: 10.21552/edpl/2022/4/8. URL: https://edpl.lexxion.eu/article/EDPL/2022/4/8 (visited on 01/22/2024).

[59] Shadreck Mwale. "'Becoming-with' a Repeat Healthy Volunteer: Managing and Negotiating Trust among Repeat Healthy Volunteers in Commercial Clinical Drug Trials". In: *Social Science & Medicine* 245 (Jan. 1, 2020), p. 112670. ISSN: 0277-9536. DOI: 10.

1016/j.socscimed.2019.112670. URL: https://www.sciencedirect.com/science/article/pii/S0277953619306653 (visited on 01/21/2024).

[60] Nathan Newman. "The Costs of Lost Privacy: Consumer Harm and Rising Economic Inequality in the Age of Google". In: *William Mitchell Law Review* 40 (2013–2014), p. 849. URL: https://heinonline.org/HOL/Page?handle=hein.journals/wmitch40&id=889&div=&collection=.

[61] Helen Nissenbaum. "A Contextual Approach to Privacy Online". In: *Daedalus* 140.4 (2011), pp. 32–48. DOI: 10.1162/DAED_a_00113.

[62] Safiya Umoja Noble. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press, 2018. 229 pp. ISBN: 978-1-4798-4994-9 978-1-4798-3724-3.

[63] Claudio Novelli et al. "Taking AI Risks Seriously: A New Assessment Model for the AI Act". In: *AI & SOCIETY* (July 12, 2023). ISSN: 1435-5655. DOI: 10.1007/s00146-023-01723-z. URL: https://doi.org/10.1007/s00146-023-01723-z (visited on 01/14/2024).

[64] Cathy O'Neil. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. First edition. New York: Crown, 2016. 259 pp. ISBN: 978-0-553-41881-1 978-0-553-41883-5.

[65] Art. 29 Data Protection Working Party. *Opinion 03/2013 on Purpose Limitation. WP 203, 00569/13/EN*. 2013.

[66] Judith Rauhofer. "Of Men and Mice: Should the EU Data Protection Authorities' Reaction to Google's New Privacy Policy Raise Concern for the Future of the Purpose Limitation Principle?" In: *European Data Protection Law Review (EDPL)* 1.1 (2015), pp. 5–15. URL: https://heinonline.org/HOL/P?h=hein.journals/edpl1&i=10 (visited on 01/09/2024).

[67] Priscilla M. Regan. "Privacy as a Comon Good in the Digital World". In: *Information, Communication & Society* 5.3 (2002), pp. 382–405. DOI: 10.1080/13691180210159328.

[68] David B. Resnik. "Scientific Research and the Public Trust". In: *Science and Engineering Ethics* 17.3 (Sept. 1, 2011), pp. 399–409. ISSN: 1471-5546. DOI: 10.1007/s11948-010-9210-x. URL: https://doi.org/10.1007/s11948-010-9210-x (visited on 01/21/2024).

[69] Hannah Ruschemeier. *Competition Law as a Powerful Tool for Effective Enforcement of the GDPR*. Verfassungsblog. July 7, 2023. URL: https://verfassungsblog.de/competition-law-as-a-powerful-tool-for-effective-enforcement-of-the-gdpr/.

[70] Hannah Ruschemeier. *Squaring the Circle*. Verfassungsblog. July 4, 2023. URL: https://verfassungsblog.de/squaring-the-circle/.

[71] Hannah Ruschemeier. "Wettbewerb Der Aufsicht Statt Aufsicht Über Den Wettbewerb?" In: *GRUR Junge Wissenschaft, Tagungsband 2023*. Ed. by Johannes Buchheim et al. Baden-Baden: Nomos, 2024 - forthcoming.

[72] Alicia Shelton. "A Reasonable Expectation of Privacy Online: Do Not Track Legislation". In: *University of Baltimore Law Forum* 45.1 (2014), pp. 35–56. URL: https://heinonline.org/HOL/P?h=hein.journals/ublfo45&i=39 (visited on 01/16/2024).

[73] Patrick Skeba and Eric PS Baumer. "Informational Friction as a Lens for Studying Algorithmic Aspects of Privacy". In: *Proceedings of the ACM on Human-Computer Interaction* 4 (CSCW2 2020), pp. 1–22.

[74] Shaun B. Spencer. "Reasonable Expectatoins and the Erosion of Privacy". In: *San Diego Law Review* 39.3 (2002), pp. 843–916. URL: https://heinonline.org/HOL/P?h=hein.journals/sanlr39&i=853 (visited on 01/16/2024).

[75] Nick Srnicek. *Platform Capitalism*. Theory Redux. Cambridge, UK ; Malden, MA: Polity, 2017. 171 pp. ISBN: 978-1-5095-0486-2 978-1-5095-0487-9.

[76] Jasmina Tacheva and Srividya Ramasubramanian. "AI Empire: Unraveling the Interlocking Systems of Oppression in Generative AI's Global Order". In: *Big Data & Society* 10.2 (July 1, 2023), p. 20539517231219241. ISSN: 2053-9517. DOI: 10.1177/20539517231219241. URL: https://doi.org/10.1177/20539517231219241 (visited on 01/09/2024).

[77] Linnet Taylor, Luciano Floridi, and Bart van der Sloot. *Group Privacy: New Challenges of Data Technologies*. New York: Springer, 2016. ISBN: 978-3-319-46606-4.

[78] Teytaud. *Genetic Stable Diffusion*. June 17, 2023. URL: https://github.com/teytaud/genetic-stable-diffusion (visited on 01/08/2024).

[79] Han Tian, Zhang Zhu, and Xu Jing. "Deep Learning for Depression Recognition from Speech". In: *Mobile Networks and Applications* (Jan. 26, 2023). ISSN: 1572-8153. DOI: 10.1007/s11036-022-02086-3. URL: https://doi.org/10.1007/s11036-022-02086-3 (visited on 10/14/2023).

[80] Cynthia Townley and Jay L. Garfield. "Public Trust". In: *Trust: Analytic and Applied Perspectives*. Ed. by Pekka Mäkelä and Cynthia Townley. Brill, Jan. 1, 2013, pp. 95–107. ISBN: 978-94-012-0941-0. DOI: 10.1163/9789401209410_007. URL: https://brill.com/display/book/9789401209410/B9789401209410-s007.xml (visited on 01/21/2024).

[81] Anton Vedder. "KDD: The Challenge to Individualism". In: *Ethics and Information Technology* 1.4 (1999), pp. 275–281.

[82] Pieter Verdegem, ed. *AI for Everyone? Critical Perspectives*. University of Westminster Press, Sept. 20, 2021. ISBN: 978-1-914386-16-9. DOI: 10.16997/book55.

URL: https://www.uwestminsterpress.co.uk/site/books/e/10.16997/book55/ (visited on 04/20/2023).

[83] Sandra Wachter. "Data Protection in the Age of Big Data". In: *Nature Electronics* 2.1 (Jan. 2019), pp. 6–7. ISSN: 2520-1131. DOI: 10.1038/s41928-018-0193-y. URL: http://www.nature.com/articles/s41928-018-0193-y (visited on 01/04/2020).

[84] Sandra Wachter and Brent Mittelstadt. "A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI". In: *Columbia Business Law Review* 2019.1 (2019), pp. 1–130. DOI: 10.31228/osf.io/mu2kf. URL: https://osf.io/mu2kf (visited on 12/20/2019).

[85] Laura Weidinger et al. *Ethical and Social Risks of Harm from Language Models*. Dec. 8, 2021. DOI: 10.48550/arXiv.2112.04359. arXiv: 2112.04359 [cs]. URL: http://arxiv.org/abs/2112.04359 (visited on 11/08/2023). preprint.

[86] Gabriela Zanfir. "Forgetting About Consent. Why The Focus Should Be On "Suitable Safeguards" in Data Protection Law". In: *Reloading Data Protection: Multidisciplinary Insights and Contemporary Challenges*. Ed. by Serge Gutwirth, Ronald Leenes, and Paul De Hert. Dordrecht: Springer Netherlands, 2014, pp. 237–257. ISBN: 978-94-007-7540-4. DOI: 10.1007/978-94-007-7540-4_12. URL: https://doi.org/10.1007/978-94-007-7540-4_12 (visited on 01/09/2024).

[87] Tal Zarsky. "Incompatible: The GDPR in the Age of Big Data". In: *Seton Hall Law Review* 47.4 (Dec. 1, 2017), pp. 995–1020. URL: https://scholarship.shu.edu/shlr/vol47/iss4/2.

[88] Shoshana Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Paperback edition. London: Profile Books, 2019. 691 pp. ISBN: 978-1-78125-685-5.