Montinaro R, 'Responsible data sharing for AI: a test bench for EU Data Law' (2024) 1 EJPLT, 17-33.
DOI: https://doi.org/10.57230/EJPLT241RM

EJPLT
ISSN 2704-8012

Responsible data sharing for AI: a test bench for EU Data Law

Condivisione responsabile dei dati per l'IA: un banco di prova per la normativa UE sui dati

ROBERTA MONTINARO
Full Professor of Private Law
Università degli Studi di Napoli L'Orientale

Abstract
The paper addresses data sharing contracts for AI development and training, entered into by parties participating in AI value chains, in order to assess the incentives for responsible data sharing under European data law. As the governance of AI training data is mainly entrusted to party autonomy and the principles of the GDPR, the analysis first assesses whether contract theory, as it is traditionally conceptualized, is an appropriate tool to promote responsible data sharing. Then the paper delves into the impact of the mandatory data protection framework on data sharing contracts, to conclude that the principle of accountability in its organizational dimension requires the parties to 'contractualize' good data sharing practices. Lastly, the analysis comprises a brief overview of the relevant provisions of the AI Act that deal with data governance and AI value chains, to highlight their impact on the relationship between the entities participating in these value chains.

© The author(s) 2024, published by Suor Orsola Benincasa Università Editrice.
This contribution is licensed under a Creative Common Attribution 4.0 International Licence CC-BY-NC-ND, all the details on the license are available at:
https://creativecommons.org/licenses/by-nc-nd/4.0/

Abstract
Il documento affronta i contratti di condivisione dei dati per lo sviluppo e l'addestramento dell'IA, stipulati dalle parti che partecipano alle catene del valore dell'IA, al fine di valutare gli incentivi per una condivisione responsabile dei dati ai sensi della normativa europea sui dati. Poiché la governance dei dati relativi all'addestramento dell'IA è principalmente affidata all'autonomia delle parti e ai principi del GDPR, l'analisi valuta innanzitutto se la teoria dei contratti, così come è stata tradizionalmente concepita, sia uno strumento appropriato per promuovere una condivisione responsabile dei dati. In seguito, il contributo approfondisce l'impatto del quadro normativo obbligatorio in materia di protezione dei dati sui contratti di condivisione dei dati, per concludere che il principio di responsabilità nella sua dimensione organizzativa richiede alle parti di "contrattualizzare" le buone pratiche di condivisione dei dati. Infine, l'analisi comprende una breve panoramica delle disposizioni rilevanti della legge sull'IA che riguardano la governance dei dati e le catene di valore dell'IA, per evidenziare il loro impatto sulle relazioni tra le entità che partecipano a queste catene di valore.

Keywords: data sharing; contracts; AI value chains; data protection; European data law.

1. Introduction.

Responsible data governance with respect to Artificial Intelligence (hereinafter 'AI') is a field of research that tries to reconcile two perspectives which underpin European Union data law (hereinafter 'EU data law')[1]: on the one hand, fostering innovation by favoring the availability of data and, on the other hand, ensuring protection to personal data and, more broadly, to the fundamental rights of the individuals and groups affected by the deployment of AI applications. The Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act, hereinafter 'AI Act')[2] seeks as well to bridge these two perspectives. It states that one of the main purposes of the Regulation is to promote a high level of protection of fundamental rights, according to the Charter of Fundamental Rights of the European Union (hereinafter the 'Charter'), starting from the development phase of AI applications (Article 1 and Recital 1 AI Act). At the same time, the Ai Act acknowledges that access to

[1] N Riis, 'Shaping the field of EU Data Law' (2023) 14 JIPITEC, 54. See also T Streinz 'The Evolution of European Data Law', in P Craig, G de Búrca (eds), *The Evolution of EU Data Law* (Oxford University Press USA, 2021), 903.
[2] European Parliament and Council Regulation (EU) 1689/2024 of 13 June 2024 laying down harmonized rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 [2024] OJ L. (Artificial Intelligence Act).

data, namely to data of high quality, plays a vital role in preventing and mitigating any risk of harm to those rights (Recitals 66 and 67 AI Act).

When machine learning techniques are applied, data comes into consideration in view of its function as training data and any lack of data quality or any deficiency in the data generation and management processes can be a source of harm to the fundamental rights of data subjects, particularly when AI is used in decision-making processes that impact individuals or groups.[3]

Data sharing for AI mainly takes place by means of contracts that can be simple arrangements between two parties or multilateral complex agreements. These contracts are often entered into in the context of the so-called AI value chains in which multiple entities may participate, including entities contributing data.

The aim of this paper is to explore data sharing contracts[4] for the purpose of developing and training AI, entered into by parties participating in AI value chains, to assess the incentives to engage in responsible data sharing for AI under EU data law.

Responsible data sharing entails, at the very minimum, compliance with personal data regulations, prevention of bias and discrimination and respect for human rights.

As will be noted in paragraph 2, the normative instruments recently enacted by the EU fall short to promote the accountability of the parties to a data sharing contract. Regulation (EU) 2022/868 on European data governance (Data Governance Act, hereinafter 'DGA')[5] and Regulation (EU) 2023/2854 on harmonized rules on fair access to and use of data (hereinafter 'Data Act')[6] are both concerned with establishing a 'data economy' and entrust the protection of fundamental rights to Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data (General Data Protection Regulation, hereinafter 'GDPR')[7], as well as to other relevant EU and national laws.[8]

As a consequence, the governance of AI training data is mainly entrusted to party autonomy and to the principles of the GDPR.

Data sharing contracts are indeed subject to the applicable contract law regimes of Member States. Against this background the analysis will firstly consider the viewpoint of contract theory (as it is traditionally conceptualized),

---

[3] H Suresh, J Guttag, 'A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle' (2021) Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '21), 1 <https://doi.org/10.1145/3465416.3483305> accessed 1 February 2024. See also P Hacker, 'A Legal Framework for AI Training Data - From First Principles to the Artificial Intelligence Act' (2020) Law, Innovation and Technology, 12 <https://ssrn.com/abstract=3556598> or <http://dx.doi.org/10.2139/ssrn.3556598> accessed 1 February 2024.

[4] Within the present paper, the terms 'supply' and 'sharing' of data will be used interchangeably, to describe a one-way provision of data, as well as a two-way provision of data, where the parties provide data to one another.

[5] European Parliament and Council Regulation (EU) 868/2022 of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 [2022] OJ L152/1 (Data Governance Act).

[6] European Parliament and Council Regulation (EU) 2854/2023 of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 [2023] OJ L. (Data Act)

[7] European Parliament and Council Regulation (EU) 679/2016 of 27April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC [2016] OJ L119/1. (General Data Protection Regulation).

[8] See in this regard Recitals 7 and 8 Data Act.

referencing some relevant soft law instruments which have been devised with the aim to set out model rules on data sharing contracts.

Then paragraph 3 will analyze how the GDPR interacts with contract law, to conclude that the data protection framework constrains the autonomy of the parties to a data sharing agreement by shaping the contractual relationship.

Lastly, the analysis will conclude with a brief overview of the relevant provisions of the AI Act which deal with data governance and AI value chains, to highlight their impact on the relationship between the parties to data sharing contracts (paragraph 4).

The AI Act sets out requirements on the quality of training data and mandates the use of adequate data governance practices. This results in due diligence obligations, but only for providers of AI applications that fall under the definition of high-risk systems. The AI Act Proposal was only intended for providers and users of AI systems, not including those who provide data for the development and training of such systems,[9] whereas the AI Act recognizes the complexity of AI value chains and consequently requires third parties involved in the development of AI systems to cooperate with providers of high-risk AI systems to enable them to comply with the requirements therein set out (Recital 88).[10] In this way, the Regulation entrusts the accountability of all participants in AI value chains to the agreements between the parties and to 'model clauses' to be drafted by the AI Office.

2. Data sharing contracts seen from the viewpoint of traditional contract theory.

2.1. Recent Evolutions in EU Data Law.

As above mentioned, the EU legal framework currently comprises the DGA and the Data Act, which have been added to the GDPR in order to implement the 'European strategy for data'[11]. As clarified by their respective preambles, both the DGA and the Data Act are mainly concerned with fostering the establishment of a European 'data economy'. As a consequence, they do not directly address the role played by data in the context of AI development and training.

The DGA focuses on ensuring the neutrality and accountability of data intermediaries falling within its scope of applications with regard to non-personal and personal data in general. It entrusts the protection of the

[9] Commission, 'Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act)' COM (2021) 206 final.

[10] It is worth noting that the AI Act Proposal with the amendments adopted by the European Parliament on 14 June 2023 (hereinafter 'AI Act Proposal of June 2023') contained explicit reference to those providing data for the development of AI systems: "all relevant third parties, in particular those that are involved in the development, sale and the commercial supply of […] data incorporated into the AI system" shall cooperate with providers of AI systems to enable them to comply with the AI Act (Recital 60 of the AI Act of June 2023).

[11] See Commission 'Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A European strategy for data' COM (2020) 66 final.

fundamental rights of data subjects to the relevant EU legislations, most notably, to the GDPR.

There is much evidence to support this. Under the DGA, the data intermediaries only bear a fiduciary duty towards the data subjects - to act in their best interest - when their business model is centered on facilitating the exercise of data subjects' rights (Article 12 (l)). Duties of care of data intermediaries with regard to third parties' rights are only provided for with respect to non-personal data, such as: *i)* taking the necessary legal, organizational and/or technological measures (see Articles 10 and 12 DGA) to avoid transmission and access to non-personal data which is unlawful under Union or national law; *ii)* adopting measures necessary to ensure an adequate level of security for the storage and transmission of non-personal data (Article 10, (1) DGA).

Moreover, the scope of application of the DGA only covers data sharing services aiming at intermediating between an indefinite number of data holders and data users, excluding data sharing services that are meant to be used by a closed group of data holders and users (see Recital 28 DGA). This requirement may leave outside the reach of the DGA several data governance models that are especially suitable for the development and training of AI, for instance, data clearinghouses (which may come in various forms; e.g. data transfer, consent management etc.) and data pools, [12] whenever the governance of the pooled data is entrusted to an intermediary that facilitates data sharing between predefined data holders and users.[13]

The Data Act sets out obligations for data holders to make both personal and non-personal data available to users of connected products and related services; it establishes rules for business-to-business and business-to-consumer data sharing, as well as rules on unfair terms related to data access and use between enterprises (see Article 1 Data Act).[14]

Examining this in detail, its scope of application covers data agreements on: *i)* access and use of data concerning the performance, use and environment of connected products and related services [Chapter II]; *ii)* private sector data that is subject to statutory data sharing obligations under Article 5 of the Data Act or under national legislation adopted in accordance with Union law (Chapter III); and *iii)* private sector data accessed and used on the basis of a contract between enterprises (Chapter IV).

Like the GDA, the Data Act is not directly concerned with protecting the rights of third parties that may be affected by the data sharing agreement, such as a right arising from the GDPR. Rather, the Data Act merely clarifies that its provisions do not affect the protection of personal data and that the sharing of

[12] See A Wernick, C Olk, M Grafenstein, 'Defining Data Intermediaries. A Clearer View through the Lens of Intellectual Property Governance' (2020) 2 Technology and Regulation, 70 <https://doi.org/10.26116/techreg.2020.007> accessed 28 January 2024.

[13] See Recital 28 DGA: "Examples of data intermediation services include […] data pools established jointly by several legal or natural persons with the intention to license the use of such data pools to all interested parties in a manner that all participants that contribute to the data pools would receive a reward for their contribution".

[14] M Hennemann, G Ebner, B Karsten, 'The Data Act Proposal – Literature Review and Critical Analysis - Part I (Art. 1-13, 35)' (2023) University of Passau Institute for Law of the Digital Society Research Paper 1/2023, 6 <www.jura.uni-passau.de/irdg/publikationen/research-paper-series> accessed 25 January 2024.

personal data can only lawfully take place if there is a legal basis in accordance with the GDPR (see Recital 7[15] and Article 4 (12) Data Act)[16].

The Data Act does take into account the quality of the data covered by the data sharing agreement in some of its provisions. In particular, Articles 4 (1) and 5 (1) Data Act stipulate that the data holder must make accessible to the counterparty data of the same quality as is available to the data holder. The rationale of these provisions, however, seem to lie in the need to regulate the interests of the parties. At the same time, according to Article 13 (4) (g) Data Act, contract terms relating to data quality cannot be modified by a unilaterally imposed term, which would be unfair and thus not binding on the other party. Since this is a type of unfair clause, here too it is reasonable to argue that such a provision is intended to address contractual imbalances that may hinder fair access to and use of data by the parties to such agreements.

Thus, despite these recent regulations that have been added to European data law, data governance for AI remains predominantly in the realm of party autonomy, as will be demonstrated in the following analysis. However, it will also be argued that the GDPR - and, albeit to a much lesser extent, the AI Act - shape the autonomy of the parties to data-sharing contracts, in ways that will be discussed in paragraphs 3 and 4.

2.2. Shortcomings of Traditional Contract Theory.

Traditional conceptualizations of private autonomy - whereby the parties to a contract are only bound by what is agreed between them with a view to protecting their own interests - are put 'under strain' when data is the subject matter of the contract. Personal data, in particular, is at the heart of several regulations, such as the GDPR. The latter, on the one hand, takes into account the private dimension inherent in the circulation of personal data and its use for economic purposes; on the other hand, it also aims to protect the rights and interests of the natural persons concerned by the processing.

The parties to a data sharing agreement regard data as a commodity, but, at the same time, they are aware that the data sharing may involve a number of legally protected interests, including rights protected by mandatory legislations. As a consequence, they tend to be risk-averse and inclined to design the contract terms in such a way to satisfy their interests and to manage the risk of being held accountable for non-compliance with public law regimes or incurring liability for any harm to third parties' rights under tort law regimes enacted at a national level or at the level of EU law.

Indeed, one of the functions of such regimes is precisely to deter unlawful forms of data sharing. This, however, provided that they have an adequate level of effectiveness.

If we consider tort law, and namely, the EU product liability regime, as it

[15] According to Recital 7 Data Act the legal basis for sustainable and responsible data processing, even where data sets include a mix of personal and non-personal data, is provided for solely by the GDPR.

[16] Article 4 (12) of the Data Act :"Where the user is not the data subject whose personal data is requested, any personal data generated by the use of a connected product or related service shall be made available by the data holder to the user only where there is a valid legal basis for processing under Article 6 of Regulation (EU) 2016/679 and, where relevant, the conditions of Article 9 of that Regulation and of Article 5(3) of Directive 2002/58/EC are fulfilled." In this vein, see also Article 5(7) of the Data Act.

currently stands, we can find that it only insufficiently contributes to promoting responsible data sharing for AI. It suffices to mention two aspects that are relevant for the purpose of this analysis.

Under the Council Directive 85/374/EEC of 25 July 1985 concerning liability for defective products (Product Liability Directive, hereinafter "PLD")[17], it is debated whether AI systems and input data, due to their nature as intangible goods, can be regarded as a product (or a component thereof); at the same time, the complexity and opacity of AI systems entail that it is difficult for the injured party to meet the burden of proof.[18]

These obstacles are likely to be overcome if the PLD will be updated by the Proposal for a Directive on Liability for Defective Products (proposal for a New Liability Directive, hereinafter "NLD")[19]. Therein, the definition of 'component' appears to be broad enough to include the data supplied to the value chains of an AI system (see Article 4 (2) NLD)[20]. A mechanism of disclosure is provided for in Article 8 NLD to alleviate the burden of proof resting on the claimant.[21]

Entrusting the governance of AI training data to party autonomy entails further shortcomings.

First, contract terms providing for limitation or exclusion of liability for breach of contract agreed upon by the parties to a data sharing contract may shift the risk of being held accountable onto one party. Under Article 13 of the Data Act, such contract terms are not binding when they are included in data contracts between enterprises,[22] but only if the breach of contracts results from intentional acts or gross negligence of the party who imposed the

[17] Council Directive 1985/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products (Product Liability Directive) OJ L 210/29.

[18] C de Meeus 'The Product Liability Directive at the Age of the Digital Industrial Revolution: Fit for Innovation?' (2019) 8(2) EuCML, 149. See also BEUC The European Consumer Bureau, 'How to make product liability fit for consumers in the digital age' (07 May 2020, BEUC-X-2020-024) <www.beuc.eu/sites/default/files/publications/beuc-x-2020-024_product_liability_position_paper.pdf> accessed 27 January 2024.

[19] Commission, 'Proposal for a Directive of the European Parliament and of the Council on Liability for Defective Products' COM (2022) 495 final/2.

[20] Article 4 (2) NLD: "'component' means any item, whether tangible or intangible, or any related service, that is integrated into, or inter-connected with, a product by the manufacturer of that product or within that manufacturer's control". Whether this definition covers, as well, data is a matter of interpretation.

[21] However, the mechanism of disclosure provided for in Article 8 NLD can only be used in the course of the litigation and on condition that the plaintiff has provided sufficient elements to substantiate the plausibility of its claim. In the same vein, under Article 3 of the Proposal for a Directive of the European Parliament and the Council of 28 September 2022 on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) COM (2022) 496 final), the presumption of causation also applies to damage caused by high-risk AI systems whenever the court deems it excessively complex for the injured party to prove the causal link. See R Gellert, A Janssen, 'The Impact of Artificial Intelligence on European Contract Law: Talking Stock to an Ongoing Process', in A Janssen, M Lehmann, R Schulze (eds.), *The future of European Private Law,* (Nomos 2023) 169-194.

[22] See Article 13 (1) Data Act: "1. A contractual term concerning access to and the use of data or liability and remedies for the breach or the termination of data related obligations, which has been unilaterally imposed by an enterprise on another enterprise, shall not be binding on the latter enterprise if it is unfair". According to Article 13 (2) Data Act: "In particular, a contractual term shall be unfair for the purposes of paragraph 3, if its object or effect is to: (a)exclude or limit the liability of the party that unilaterally imposed the term for intentional acts or gross negligence; (b) exclude the remedies available to the party upon whom the term has been unilaterally imposed in the case of non- performance of contractual obligations, or the liability of the party that unilaterally imposed the term in the case of a breach of those obligations; (c) give the party that unilaterally imposed the term the exclusive right to determine whether the data supplied are in conformity with the contract or to interpret any contractual term".

contract term. Therefore, this rule aims to rebalance unequal relationships between economic operators active in data markets. It differs from provisions which prohibit exclusion or limitation clauses contravening public order[23] and, in particular, clauses on the exclusion or limitation of contractual liability for physical harm or harm to personality rights.[24]

Second, liability under contract law does not extend beyond the parties to the contract and, hence, it fails to promote the accountability of all those who share data for AI development (for instance, in the case of a chain of data sharing contracts).

Moreover, the third parties impacted by the deployment of AI applications are not entitled to claim damages under contract law, unless they qualify as end-users that can seek compensation from the seller under Directive (EU) 2019/770 on certain aspects concerning contracts for the supply of digital content and digital services (hereinafter 'DCDS Directive')[25] and Directive (EU) 2019/771 on certain aspects concerning contracts for the sale of goods (hereinafter 'CSD')[26]. In this case, incentives for developers of AI systems to comply with data protection principles can only arise from the right of redress that the seller can exercise against a person in the previous links of the transaction chain, whose action or omission resulted in the lack of compliance (see Article 18 CSD and, in the same vein, Article 20 DCDS Directive).[27]

2.3. Model Rules on Contracts for the Provision of Data: Fit for AI Training Data?

The party autonomy logic above described is reflected in some soft law documents aiming to set out uniform model rules and implied terms for the cross-border flows of data.

According to the American Law Institute and the European Law Institute's Principles for a Data Economy: Data Transactions and Data Rights (hereinafter 'ALI/ELI Principles for a Data Economy')[28], in contracts for the 'supply' or 'sharing' of data, the data recipient must be enabled "rightfully to exercise control over the data" (Principle 7(2)(c)(ii))[29]. At the same time, "the data

[23] The Italian civil code qualifies as null and void any terms that exclude or limit liability for breach of contract for intentional acts and gross negligence, irrespective of whether these terms are unilaterally imposed (Article 1229 Italian civil code).

[24] See H Claes, M Herbosch 'Artificial Intelligence and Contractual Liability Limitations: A Natural Combination?' (2023) ERPL 31 (2-3), 475.

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

[26] European Parliament and Council Directive (EU) 2019/771 of 20 May 2019 on certain aspects concerning contracts for the sale of goods, amending Regulation (EU) 2017/2394 and Directive 2009/22/EC, and repealing Directive 1999/44/EC [2019] OJ L 136/28.

[27] P Hacker, 'A Legal Framework for AI Training Data - From First Principles to the Artificial Intelligence Act' (2020) 18 Law, Innovation and Technology, 11 <https://ssrn.com/abstract=3556598> or <http://dx.doi.org/10.2139/ssrn.3556598> accessed 1 February 2024.

[28] The American Law Institute, The European Law Institute, 'Principles for a Data Economy: Data Transactions and Data Rights' (2023) <www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ALI-ELI_Principles_for_a_Data_Economy.pdf> accessed 20 January 2024.

[29] The rationale behind Principle 7(2)(c)(ii) of ALI/ELI Principles for a Data Economy is explained as follows: "the usefulness of data to the recipient would be undermined if the recipient did not obtain rightful control over the data at the time it is supplied". "The supplier must therefore ensure that, for example, there are no

recipient may utilize the data […] for any lawful purpose and in any way that does not infringe the rights of the supplier or third parties […]" (Principle 7(2)(c)(iv)).[30]

In the same vein, under the Default rules for data provision contracts (hereinafter 'UN Default rules for data provision contracts') issued by the United Nations Commission on International Trade Law[31], the data must be provided lawfully (Article 7 (2)(d)). Moreover, the data recipient is entitled to use the data for any lawful purpose (so "that the data recipient can exercise its rights to use the data under the contract")[32] and in a manner that does not infringe the rights of third parties (Article 8 (1) (a) and (2) (a)).

In a nutshell, the lawfulness of the data provided or made accessible under these contracts pertains to the conformity of the data[33] and covers compliance with any applicable legal requirements;[34] at the same time, lawfully using the data is a contractual obligation towards the data supplier that the data recipient must comply with, to ensure that the data provider is not held liable under the law.

Some conclusions can be drawn from these soft law documents. The contract terms therein devised are shaped by the private interests of the parties. They do not reflect the legal status of data which differs from tradable commodities.[35]

The ELI/ALI Principles for a Data Economy do recognize the existence of third parties' rights, including rights with *erga omnes* effect, which may be affected by data sharing or 'data activities' (e.g. how data are handled, their quality, security, etc.). However, these Principles: *i)* on the one hand, adopt a 'tort law logic' in that they merely reiterate that the data processing activity may be unlawful towards third parties holding *erga omnes* rights; ii) on the other hand, entrust the determination of legal remedies available to the affected third parties to the applicable law (see Principles 4, 28, 29). Similarly, with regard to the matters of the effects of the transmission of data on the protection of others and direct actions by a data supplier towards a downstream recipient these Principles refer the matter to the applicable law(see Part IV, Chapter B,

---

legal barriers that would prevent the recipient from rightfully gaining control. Legal barriers could be barriers stemming, e.g., from data privacy/data protection law, from intellectual property law, or from trade secrets law".

[30] From the perspective of the ALI/ELI Principles for a Data Economy, third party rights, including fundamental rights, help to define the data rights conferred on the recipient of the data under the contract. See ALI/ELI Principles for a Data Economy, p. 10: "Among the policy choices recommended by these Principles in the context of supply or sharing of data is the default position that, when the data is fully transferred, the data may be used by the recipient for any lawful purpose that does not infringe the rights of third parties".

[31] Working Group IV on Electronic Commerce, 'Default rules for data provision contracts (first revision): Note by the Secretariat' (United Nations Commission on International Trade Law, Vienna 16-20 October 2023) <https://documents-dds-ny.un.org/doc/UNDOC/LTD/V23/064/75/PDF/V2306475.pdf?OpenElement> accessed 22 January 2024.

[32] UN Default rules for data provision contracts (2023) 13.

[33] In this respect, these soft law instruments mirror Article 10 DCDS Directive: Third-party rights: "Where a restriction resulting from a violation of any right of a third party, in particular intellectual property rights, prevents or limits the use of the digital content or digital service in accordance with Articles 7 and 8, Member States shall ensure that the consumer is entitled to the remedies for lack of conformity […]".

[34] See Article 7 (2) (d) UN Default rules for data provision contracts.

[35] See V Janeček, G Malgieri, 'Data extra commercium', in S Lohsse, R Schultze and D Staudenmayer (eds), *Data as Counter-Performance - Contract Law 2.0?* (Hart Publishing/Nomos 2020), 2 <https://ssrn.com/abstract=3400620> accessed 1 February 2024.

Principles 32 and 33 respectively), only suggesting that a 'model rule' is one that provides for due diligence obligations for the party supplying the data with respect to wrongful activities of the other parties in the chain of contracts for the provision of data.

Above all, the soft law instruments already mentioned do not take into account the peculiarities of the data used in the context of AI development/training. As explained by experts in the field of machine learning, the data used for this purpose is not a static artefact; on the contrary, it is the result of choices and practices to be implemented during the phases of data collection and generation and model development. Possible sources of harm to fundamental rights may result from poor choices and practices at these stages.

In order to prevent such harms, the terms of data-sharing agreements should include good practices, such as: *i)* requiring 'data traceability', *ii)* setting forth transparency requirements (for instance, documentation on how input data are generated and managed), *iii)* mandating to identify adequate technical standards/practices.

Soft law alone appears ill-suited to ensure that good practices and technical standards become binding and thus subject to enforcement. It is therefore relevant to identify the incentives offered by European data law to engage in responsible data governance, in particular, with regard to personal data used for the development of AI applications.

3. The Impact of the GDPR on Data Sharing Contracts for AI.

The interaction of the GDPR with contract law challenges the corollaries of traditional contract theory above mentioned.

Owing to the peculiar nature of data, when the subject matter of a contract is the provision of data, several legally protected interests may come into play, including the right to the protection of personal data and other fundamental rights, such as the right not to be discriminated against.

Indeed, the provision of personal data to the recipient of the data pursuant to a data sharing agreement implies a form of processing within the meaning of Article 4 (2) GDPR. Accordingly, the data protection legal framework applies, which aims, inter alia, to protect the fundamental rights and freedoms of natural persons (see Article 1 (2) GDPR).

Under EU data law, where non-personal and anonymized data are processed, and technological developments make it possible to turn this data into personal data, the data protection framework comes into play (see Recital 9 of the Regulation (EU) 2018/1807 on a framework for the free flow of non-personal data in the European Union (hereinafter 'FFDR')[36]). In light of this, the key concept for defining the character of data is 'identifiability' (see Article 4(1)

[36] European Parliament and Council Regulation (EU) 1807/2018 of 14 November 2018 on a framework for the free flow of non-personal data in the European Union [2018] OJ L303. According to Recital 9 FFDR "Recital 9 of the FFDR "If technological developments make it possible to turn anonymized data into personal data, such data are to be treated as personal data, and Regulation (EU) 2016/679 is to apply accordingly".

GDPR). As highlighted by legal scholars and case law,[37] identifiability is a dynamic and context-dependent concept. There is a strong argument that the qualification of data as personal must take into account inter alia the 'life cycle' of the data. This can be deduced from Recital 26 GDPR, which refers to anonymous data, i.e. data rendered anonymous in such a manner that the data subject is not or is *no longer* identifiable.

Against this background, the relevant international technical standards (ISO/IEC 27701) rely on the concept of *Personally Identifiable Information* (hereinafter 'PII')[38]. These standards suggest that any data that *i)* can be used to establish a link between the information and the natural person to whom it relates or *ii)* can be directly or indirectly linked to a natural person should be regarded as PII and managed by taking into account any risk of harm to third parties. Therefore, the substantive scope of application of the data protection framework is not narrow, as some scholars point out.[39]

The public law regime set out by the GDPR entails constraints on party autonomy, since personal data can only be processed in conformity with the principles of lawfulness and fairness enshrined in article 5 GDPR. The requirements arising from the data protection framework translate into limits on entitlements and rights of the parties to data sharing contracts, as well as into duties for the same parties.[40] These constraints and duties shape the contractual relationship between the parties to such contracts.

For instance, in the Italian legal system, as interpreted by the Supreme Court (no. 15824/2014), compliance with duties of care arising from public law regulations, such as the EU Regulation on food safety, constitutes a contractual obligation resting with both the parties to a sales contract: they are required to proactively comply with the regulation in order to ensure a high level of protection of the rights of consumers; failing to fulfil such obligations may result in liability for breach of contract and/or contract termination. Some legal scholars point out that duties of this kind lie in the grey area between tort and contractual liability, having their origin in the fact that the liable party is entrusted with the protection of the person or property of the other party and/or a third party.[41]

A similar finding holds with regard to Article 16 DCDS Directive: "In respect of personal data of the consumer, the trader shall comply with the obligations

---

[37] Judgment of 19 October 2016, *Patrick Breyer vs Bundesrepublik Deutschland*, C-582/14, EU:C:2016:77 <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:62014CJ0582>. See, as well, I Graef, R Gellert, M Husovec, 'Towards a holistic regulatory approach for the European data economy: why the illusive notion of non-personal data is counterproductive to data innovation' (2018) 29 TILEC Discussion Paper, 7 <https://ssrn.com/abstract=3256189> or <http://dx.doi.org/10.2139/ssrn.3256189>.

[38] E Podda, M Palmirani 'Inferring the Meaning of Non-personal, Anonymized, and Anonymous Data', in V Rodríguez-Doncel, M Palmirani, M Araszkiewicz, P Casanovas, U Pagallo, G Sartor (eds), *AI Approaches to the Complexity of Legal Systems XI-XII: AICOL 2020, AICOL 2018, XAILA 2020. Lecture Notes in Computer Science*, vol 13048 (Springer, 2021) 4.

[39] N Purtova 'The law of everything. Broad concept of personal data and future of EU data protection law' (2018) 10(1) Law Innov. Technol., 40-81<https://doi.org/10.1080/17579961.2018.1452176> accessed 29 January 2023.

[40] S Orlando 'Il coordinamento tra la direttiva 770/2019 e il GDPR. L'interessato-consumatore' (2023) 2 Persona e mercato, 232.

[41] C Castronovo 'L'obbligazione senza prestazione ai confini tra contratto e torto', in G Alpa and others, *Le ragioni del diritto. Scritti in onore di Luigi Mengoni* (Giuffré, 1995) vol I, 147. See L Nivarra 'Alcune precisazioni in tema di responsabilità contrattuale' (2014) 1 Europa e diritto privato, 73.

applicable under Regulation (EU) 2016/679". This provision can be read as it entails that non-compliance with the GDPR is also relevant from the point of view of contract law.[42]

Moreover, the data protection framework covers automated decision-making that often involves the deployment of AI systems which may have significant adverse impacts on individuals (Article 22 (1) GDPR).[43] Automated decision-making of this sort is per se prohibited, but Article 22 (2) GDPR provides for several exceptions. When one of these exceptions applies, transparency requirements are set out and the adoption of adequate measures is mandated (Article 22 (3)) according to the principles of accountability and privacy by design (see Articles 5, 22, 24 and 25; Recital 71 GDPR), in order to prevent such impacts.[44]

These principles interact with party autonomy by establishing duties of care on the parties that share and manage data for the purpose of training/developing AI applications.

As highlighted by some authors, the principle of accountability has an organizational dimension, since it places on controllers the duty to also adopt the 'legal' measures that enable them to comply with the GDPR. Therefore, the parties to these contracts are required to arrange adequate data sharing architectures, starting with the choice of the contract type and contract terms,[45] which should be made taking into account the complexity and severity of the risks entailed in the processing (which are a variable of the purpose of the AI system and type of data needed). As for the choice of contract type, for instance, where there is a large number of data contributors, the creation of data pools instead of a chain of contracts can be considered in line with these principles. In fact, this type of agreement allows each party to monitor whether the other parties comply with data protection requirements. In this way, all the parties to the contract are required to responsibly manage the data.

Moreover, when an AI application is meant to be used for decision making purposes, the parties that confer the training data are required to ensure data quality and use adequate processes of data generation and management. This aspect has been underlined by the EDPB-EDPS Joint Opinion on the interplay between the AI Act and the GDPR:[46] AI systems should be conceived and developed having in mind data protection principles and rules, such as the ones which require to adopt adequate safeguards to prevent any harmful impact arising from automated decision making having legal or similarly significant effects on a data subject (Article 22 and Recital 71 GDPR).

---

[42] G Resta, *Autonomia privata e diritti della personalità* (Jovene, 2005) 281.

[43] Judgment of 7 December 2023, *OQ vs Land Hessen*, C-634/21, EU:C:2023:957 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62021CJ0634>.

[44] ME Kaminski 'The right to explanation, explained' (2019) 34(7) Berkeley Technology Law Journal <https://ssrn.com/abstract=3196985> or <http://dx.doi.org/10.2139/ssrn.3196985> accessed 31 January 2024.

[45] S Stalla-Bourdillon and others, 'Data protection by design: Building the foundations of trustworthy data sharing' (2020) 2 Data & Policy, 5.

[46] EDPB-EDPS, 'Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)' (21 June 2021) point 58 <https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_en> accessed 31 January 2024.

One way to comply with the principle of accountability and privacy by design is to choose adequate contract terms. This means that the duties of care arising from the data protection framework can be 'contractualized' and turned into due diligence contract terms included in the data sharing contracts.

Contract terms of this sort may compel the parties to such contracts to adopt: i) technical and procedural standards and practices on matters such as the ones mentioned in Paragraph 2.2 ii) mechanisms to oversee compliance with these standards/practices by all the parties (for instance, mandating to carry out internal audits or reporting) and iii) enforcement mechanisms (for instance, penalty clauses, contract termination clauses etc.) in the event of non fulfilment of these duties. Along the same lines are contract terms requiring the parties to monitor whether sub-suppliers comply with data protection requirements.

As a consequence, the adoption of good practices and standards is not left to market drivers.[47] On the contrary, it is triggered by the public law provisions at play and notably by the principle of accountability which is a regulatory model for the protection of personal data based on a hybrid form of enforceable self-regulation.[48]

The GDPR already provides for some form of private enforcement. According to Article 82 GDPR, data controllers (and processors) are liable for damage resulting from a processing that does not comply with the rules and principles laid down in the GDPR. However, a right to compensation is afforded under certain conditions. i.e. an infringement of the Regulation, the existence of a material or non-material damage and a causal link between these two elements (Article 82(1) GDPR). Furthermore, when a processing is complex and involves multiple parties and processing operations, these parties may qualify as *i)* processors or separate and independent controllers (or a combination of the two) or *ii)* as joint controllers. The claimant must therefore prove the existence of all such conditions. It is correct that an easing of the burden of proof is provided for in Article 82 (4) GDPR, according to which the multiple parties involved in the same processing may be held jointly and severally liable. In this case, the full amount of compensation may be recovered from one of these multiple parties, but only provided that the requirements of Article 82 (2) and (3) GDPR are met, in addition to their '*involvement*' in the processing.

Above all, the one just described is a form of tort liability having a compensatory function.[49] A further policy objective is to prevent infringements of data protection regulations and the resulting harm to data subjects. This objective underpins Article 26 GDPR which mandates joint controllers to clarify their roles and responsibilities in an agreement (Article 26 GDPR). The agreement to be adopted by the joint controllers under this provision is

[47] K Peterkova Mitkidis, 'Sustainability Clauses in International Supply Chain Contracts: Regulation, Enforceability and Effects of Ethical Requirements' (2014) 1 Nordic Journal of Commercial Law, 7 <www.researchgate.net/publication/299269785_Sustainability_Clauses_in_International_Supply_Chain_Contracts_Regulation_Enforceability_and_Effects_of_Ethical_Requirements> accessed 22 August 2024.

[48] A Spina, 'Alla ricerca di un modello di regolazione per l'economia dei dati. Commento al regolamento (Ue) 2016/679' (2016) 1 Rivista della regolazione dei mercati, 148. See also G Comandé 'Intelligenza artificiale e responsabilità tra liability e accountability. Il carattere trasformativo dell'IA e il problema della responsabilità' (2019) 1 Analisi giuridica dell'economia, 184.

[49] See Judgment of 4 May 2023, *UI vs Österreichische Post AG*, C-300/21, EU:C:2023:370 <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=CELEX%3A62021CJ0300>.

intended to ensure the lawfulness of joint data processing activities and, therefore, to perform an ex ante protection function, whereas liability under Article 82 GDPR only comes into play once damage has occurred.

However, providing evidence of the existence of joint control in a concrete case can be arduous and generate uncertainty. In this respect, the EDPB Guidelines on the concepts of controller and processor in the GDPR are emblematic when they state that "The fact that several actors are involved in the same processing does not mean that they are necessarily acting as joint controllers of such processing. Not all kind of partnerships, cooperation or collaboration imply qualification of joint controllers as such qualification requires a case-by-case analysis of each processing at stake and the precise role of each entity with respect to each processing".[50] The decisive factor is the existence of purposes and means of processing determined jointly by the controllers. In this respect, it is added that "the exchange of the same data or set of data between two entities without jointly determined purposes or jointly determined means of processing should be considered as a transmission of data between separate controllers".[51] Being well aware of these challenges, the French CNIL felt the need to provide guidance that specifically addresses where joint controllers can be involved in the use of personal data for the development and training of AI applications.[52]

On the contrary, good practices/standards on data governance, once turned into contract terms, become binding on all parties to a data sharing contract, regardless of how these parties qualify under the data protection legal framework. Furthermore, they can be enforced through the contractual remedies available under the applicable contract law regime, such as contract termination and/penalty clauses.

Contract law remedies stand alongside the forms of public and private enforcement envisaged by the GDPR by contributing to promoting the accountability of all those who share data for the training and development of an AI application, even in the context of complex value chains.

4. The AI Act: Data Governance and AI Value Chains.

The AI Act aims to establish a legal framework that is mainly addressed to providers of high-risk AI systems (which may not coincide with data controllers) and deals with all the stages (input data collection, design and development of models etc.) leading to the deployment of AI systems. Inter alia, it mandates providers of high-risk systems to adopt adequate data governance practices and to ensure the quality of training data (Article 10 AI Act). It also sets out transparency and documentation duties for providers of systems of this kind

[50] EDPB, 'Guidelines 07/2020 on the concepts of controller and processor in the GDPR' (07 July 2021) point 69 <www.edpb.europa.eu/system/files/2023-10/EDPB_guidelines_202007_controllerprocessor_final_en.pdf>.

[51] Ibidem, point 70.

[52] See CNIL, 'Déterminer la qualification juridique des acteurs' (8 April 2024) <www.cnil.fr/fr/determiner-la-qualification-juridique-des-fournisseurs-de-systemes-dia>: "Plusieurs acteurs peuvent intervenir dans le développement d'un système d'IA, avec divers degrés d'implication sur les traitements de données personnelles".

with reference, as well, to data governance and data quality (see Articles 11 and 13 (3) (vi) AI Act). These requirements, however, only apply to high-risk AI systems.

Moreover, the AI Act Proposal did not take into account the fact that the provider of an AI system rarely coincides with its developer and almost never carries out in-house all the steps leading up to the AI system being placed on the market or put into service. As a result, it was left to AI providers to promote the accountability of the other participants in the value chain, also by adopting contract terms under which the latter are required to guarantee compliance with the AI provisions on data quality, documentation etc. This policy option relied on the assumption that the AI provider is the economic operator in a dominant position vis-à-vis these participants. This assumption, however, was criticized for being too simplistic and failing to take into account that possible imbalances of economic power between the parties do not always see AI providers assuming the role of 'strong' parties in the relationship with participants in the AI value chain.

In the light of such criticism the approved text of the AI Act acknowledges the complexity of AI value chains, which may include multiple third parties involved in the development and in the supply of components. Hence, it prescribes that these third parties must cooperate, as appropriate, with providers of AI systems to enable them to monitor compliance with the requirements set out by the AI Act. To this end, third parties are required to disclose in writing to AI providers all information necessary for this purpose (Article 25 (4) AI Act).[53]

This obligation to cooperate sets an incentive for AI providers and value chain participants to 'contractualize' and document how they share responsibility for complying with requirements on high-risk AI systems, including data governance. This brings a number of benefits but also possible risks.

AI providers and value chain participants can theoretically agree on a fair distribution of the responsibilities between them. On the other hand, risks arise from possible imbalances of bargaining power among the contracting parties, for instance, where AI providers are in a position of economic dominance vis-à-vis the other participants in the value chain, such that the former can impose contractual terms under which the costs of compliance with the requirements of the IA Act or the fines provided for therein are passed on to the latter.

In order to address possible inequalities in bargaining power within AI value chains, the AI Act Proposal of June 2023 explicitly qualified contract terms unilaterally imposed by AI providers on a SME or start-up aimed at excluding or limiting the liability of the party imposing the term for intentional act and gross negligence (or excluding or limiting access to remedies available to the other party) as unfair and non-binding.[54] In addition, the AI Act Proposal of June 2023

[53] Article 25 (4) AI Act: "The provider of a high-risk AI system and the third party that supplies an AI system, tools, services, components, or processes that are used or integrated in a high-risk AI system shall, by written agreement, specify the necessary information, capabilities, technical access and other assistance based on the generally acknowledged state of the art, in order to enable the provider of the high-risk AI system to fully comply with the obligations set out in this Regulation".

[54] See Article 28a AI Act Proposal of June 2023. See, as well, Recital 62 AI Act Proposal of June 2023: "In order to ensure a high level of trustworthiness of high-risk AI systems, those systems should be subject to a

stated that "*A contractual term is also unfair if it has the effect of shifting penalties referred to in Article 71 or associated litigation costs across parties to the contract, as referred to in Article 71(8)*" (Article 28a).

These provisions have not been included in the final text of the IA Act. Instead Article 28 (4) AI Act empowers the AI Office to develop and recommend voluntary model terms for contracts between providers of high-risk AI systems and those involved in the value chain. The fact that such model terms can be adopted on a voluntary basis is further evidence that the governance of AI value chains is left to national or EU regulations on unfair terms in business contracts. Likewise, the validity of clauses on penalties or claims arising from non-compliance with the IA Act included in contracts between suppliers and participants in IA value chains must be assessed under national contract law.[55]

5. Conclusions.

The paper addressed data sharing contracts for AI development and training, entered into by parties participating in AI value chains, in order to assess the incentives for responsible data sharing under European data law.

The normative instruments recently enacted by the EU, such as the DGA and the Data Act, are concerned with establishing a 'data economy' and fall short to promote the accountability of the parties to a data sharing contract. As a consequence, the governance of AI training data is mainly entrusted to party autonomy and to the principles of the GDPR, where the latter is applicable.

As demonstrated, the terms included in data sharing contracts tend to be exclusively shaped by the private interests of the parties. These terms do not reflect the legal status of data which differs from tradable commodities, nor do they take into account the peculiarities of the data used in the context of AI development/training. The data used for this purpose is not a static artefact; on the contrary, it is the result of choices and practices to be implemented during the phases of data collection and generation and model development, in order to tackle possible sources of harm to fundamental rights.

The analysis showed that leaving the governance of AI training data to party autonomy entails several shortcomings. The parties are incentivized to comply with mandatory regulations, such as those which protect personal data and the right to privacy, only if the public law and tort law regimes at play show an adequate level of effectiveness. Moreover, liability under contract law does not extend beyond the parties to the contract and, therefore, it fails to promote the accountability of all those who share data for AI development.

---

conformity assessment prior to their placing on the market or putting into service. *To increase the trust in the value chain and to give certainty to businesses about the performance of their systems, third-parties that supply AI components may voluntarily apply for a third-party conformity assessment.*"

[55] In reality, the AI Act Proposal of June 2023 contained a provision (Article 71 (8a)) under which: "*The penalties referred to in this article as well as the associated litigation costs and indemnification claims may not be the subject of contractual clauses or other form of burden-sharing agreements between providers and distributors, importers, deployers, or any other third parties*". Although clarifying, this provision merely reasserted a principle common to many legal regimes, i.e. that private autonomy is not allowed to alter the allocation of liability between the parties, which is established by public law remedies.

However, it was demonstrated that the interaction of the data protection framework with contract law challenges the corollaries of traditional contract theory. The public law regime set out by the GDPR entails constraints on party autonomy, since data can only be processed in conformity with the principles of lawfulness and fairness enshrined in article 5 GDPR. Moreover, the data protection framework also covers automated decision-making that often involves the deployment of AI systems which may have significant adverse impacts on individuals. In these cases, transparency requirements are set out and the adoption of adequate measures is mandated according to the principles of accountability and privacy by design (see Articles 5, 22, 24 and 25; Recital 71 GDPR), in order to prevent such impacts. These principles interact with party autonomy by establishing duties of care on the parties that share and manage data for the purpose of training/developing AI applications.

As highlighted in the paper, the principle of accountability has an organizational dimension and places on controllers the duty to take the 'legal' measures that enable them to comply with the GDPR. Therefore, the parties to these contracts are required to arrange adequate data sharing architectures, starting with the choice of the contract type and contract terms.

This means that the duties of care arising from the data protection framework can be 'contractualized' and turned into due diligence contract terms included in the data sharing contracts. Contract terms of this sort may compel the parties to such contracts to adopt: *i)* technical and procedural standards and practices *ii)* mechanisms to oversee compliance with these standards/practices by all the parties (for instance, mandating to carry out internal audits or reporting) and *iii)* enforcement mechanisms (for instance, penalty clauses, contract termination clauses, etc.) in the event of non fulfilment of these duties. Along the same lines are contract terms requiring the contracting parties to monitor whether sub-suppliers comply with data protection requirements.

As a consequence, the adoption of good practices and standards is not left to market drivers. On the contrary, it is triggered by the public law provisions at play and notably by the principle of accountability which is a regulatory model for the protection of personal data based on a hybrid form of enforceable self-regulation. Contract law remedies stand alongside the forms of public and private enforcement envisaged by the GDPR and contribute to promoting the accountability of all those who share data for the training and development of an AI application, even in the context of complex value chains.

Lastly, the analysis comprised a brief overview of the relevant provisions of the AI Act that deal with data governance and AI value chains. This new body of legislation acknowledges the complexity of AI value chains, which may comprise multiple entities, such as parties involved in the supply of data. It is therein prescribed that these third parties must cooperate, as appropriate, with providers of AI systems to enable them to monitor compliance with the requirements set out by the AI Act. However, the AI Act does not address inequalities of bargaining power between the various economic actors involved. This gap risks undermining the policy objective of making all those participating in value chains for high-risk AI systems accountable.