

# Machine Learning & EU Data Sharing Practices

By Mauritz Kop<sup>1</sup>

**Stanford - Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 1/2020**

## Abstract

*This article connects the dots between intellectual property (IP) on data, data ownership and data protection (GDPR and FFD), in an easy to understand manner. It also provides AI & Data policy and regulatory recommendations to the EU legislature.*

*Data sharing is a prerequisite for a successful Transatlantic AI ecosystem. Hand-labelled, annotated training datasets (corpora) are a sine qua non for supervised machine learning. But what about intellectual property (IP) and data protection?*

*Data that represent IP subject matter are protected by IP rights. Augmented machine learning training datasets are awarded with either a database right or a sui generis database right in Europe. Unlicensed (or uncleared) use of machine learning input data potentially results in an avalanche of copyright (reproduction right) and database right (extraction right) infringements.*

*The article offers three solutions that address the input (training) data copyright clearance problem and create breathing room for AI developers: the implementation of a broadly scoped, mandatory TDM exception covering all types of data (including news media) in Europe, the Fair Learning principle in the United States and the establishment of an online clearinghouse for machine learning training datasets. A right to machine legibility that drastically improves access to data, will greatly benefit the growth of an AI ecosystem.*

*Introducing an absolute data property right or a (neighboring) data producer right for annotated machine learning training datasets or other classes of data is not opportune. Legislative gaps concerning ownership of data can be remedied by contracts. Implementing a sui generis system of protection for AI-generated Creations & Inventions is -in most industrial sectors- not necessary since machines do not need incentives to create or invent. Where incentives are needed, IP alternatives exist.*

*Autonomously generated non-personal data should fall into the public domain. It should be open data, excluded from protection by the Database Directive (DD), the Copyright Directive (CDSM) and the Trade Secrets Directive (TSD).*

*As legal uncertainty about the patentability of AI systems is causing a shift towards trade secrets, legal uncertainty about the protection and exclusive use of machine generated databases is causing a similar shift towards trade secrets. This general shift towards trade secrets to keep competitive advantages results in a disincentive to disclose information and impedes on data sharing. In an era of*

---

<sup>1</sup> Mauritz Kop, Stanford Law School TTLF Fellow, Stanford University; Managing Partner at AIRecht, Amsterdam, The Netherlands. Correspondence: [advies@airecht.nl](mailto:advies@airecht.nl). The author would like to thank Mark Lemley, Begoña Gonzalez Otero, Teresa Quintel, Suzan Slijpen and Nathalie Smuha for valuable remarks on an earlier draft of this article, and Christophe Geiger for his lecture on Big Data, Artificial Intelligence, Freedom of Information and the TDM exception, organized by IViR, 10 March 2020.

*exponential innovation, it is urgent and opportune that both the TSD, the CDSM and the DD shall be reformed by the EU Commission with the data-driven economy in mind.*

*Informed IP policy seeks to compose a regime that balances underprotection and overprotection of IP rights per economic sector. Freedom of expression and information are core democratic values that should be internalized in our IP framework. The article argues that strengthening and articulation of competition law is more opportune than extending IP rights.*

*More and more datasets consist of both personal and non-personal machine generated data. Both the General Data Protection Regulation (GDPR) and the Regulation on the free flow of non-personal data (FFD) apply to these 'mixed datasets'. Based on these two Regulations, data can move freely within the European Union. The article contends that in some cases, GDPR legislation causes market barriers for early-stage AI-startups (SME's). The GDPR also has some important advantages for European SME's since it is now the international data protection standard.*

*Besides the legal dimensions, the article describes the technical dimensions of data in machine learning. Most AI models need centralized data. Federated learning, in contrast, trains algorithms by bringing the code to the data, instead of bringing the data to the code. Data sharing is not required.*

*Both data sharing practices and AI-Regulation are high on the EU Commission's agenda. The article discusses -inter alia- the EC's 'White Paper On Artificial Intelligence - A European approach to excellence and trust' and the 'EU Data Strategy'.*

*Important European initiatives in the field of open data and data sharing are: the Support Centre for Data Sharing (focused on data sharing practices), the European Data Portal (EDP, data pooling per industry i.e. sharing open datasets from the public sector), the Open Data Europe Portal (ODP, sharing data from European institutions) and the EU Blockchain Observatory and Forum.*

*Transformative technology is not a zero sum game, but a win-win strategy that creates new value. When developing inclusive transformative tech related policies, the goal should be a Pareto optimum and if possible a Pareto improvement by increasing overall prosperity.*

*Society should actively shape technology for good. The alternative is that other societies, with perhaps different social norms and democratic standards, impose their values on us through the design of their technology. With built-in public values, including Privacy by Design that safeguards data protection, data security and data access rights, the federated learning model is consistent with Human-Centered AI and the European Trustworthy AI paradigm.*

## Introduction

Data sharing or rather the ability to analyse and process high quality training datasets (*corpora*) to teach an Artificial Intelligence (AI) model to learn, is a prerequisite for a successful Transatlantic AI ecosystem. But what about intellectual property (IP) and data protection?

In our turbulent technological era, tangible information carriers such as paper and storage media are declining in importance. Information is no longer tied to a continent, state or place. Information technology such as AI is developing at such a rapid, exponential pace that the legal problems that arise from it are to a large extent unpredictable.

## 1. Legal dimensions of data

Data, or information, has a large number of legal dimensions.<sup>2</sup> Data sharing is associated with IP law (right to prohibit and reimburse), fundamental rights (privacy, data protection, freedom of expression and other constitutional rights)<sup>3</sup>, fiscal law (taxation), contract law and international commercial law (e-commerce, trade treaties, anti-trust law, consumer protection).<sup>4</sup> In addition, the handling of personal data has ethical, social and techno-philosophical facets.

### *Legal ownership of data does not exist*

In most European countries, the law of property is a closed system.<sup>5</sup> This means that the number of proprietary rights *in rem*, which are rights enforceable against everyone, are limited by law. Legal ownership of data therefore does not yet exist. From a property law point of view, data cannot be classified as “*res*”, as an intangible good or as a thing in which property rights can be vested. Data does have proprietary rights aspects and represents value.

### *Data that represent IP subject matter*

Data that represent IP subject matter are protected by IP rights.<sup>6</sup> Data that embody original literary or artistic works are protected by copyright. New, non-obvious and useful inventions represented by data are protected by patents. Data that epitomize independently created new and original industrial designs are safeguarded by design rights.<sup>7</sup> Confidential data that have business or technological value are protected by trade secret rights.<sup>8</sup>

---

<sup>2</sup> Data and information are not always interchangeable terms. From a European trade secrets perspective, it is not clear whether data or datasets fulfill the requirements of Article 2(1) of the EU [Trade Secrets Directive](#) (TSD). When data is mentioned in the TSD, the terms seems to be not understood as “datasets” but rather in the context of customer/supplier lists – “commercial data” in recital 2 or “personal data” in Article 9(4). The TSD was not developed with the data-driven economy in mind, but rather on the information society (recitals 1 and 4).

<sup>3</sup> Privacy and data protection are not always interchangeable terms. Privacy is a human right as enshrined in Article 12 of the [Universal Declaration of Human Rights](#).

<sup>4</sup> See for international commercial law aspects: Kristina Irion & Josephine Williams (2019). ‘Prospective Policy Study on Artificial Intelligence and EU Trade Policy’. Amsterdam: The Institute for information Law (IViR) 2019. See for consumer protection: Gabriele Accardo and Maria Rosaria Miserendino, ‘Big Data: Italian Authorities Published Guidelines and Policy Recommendation on Competition, Consumer Protection, and Data Privacy Issues’, TTLF Newsletter on Transatlantic Antitrust and IPR Developments Stanford-Vienna Transatlantic Technology Law Forum, Stanford University, 2019 Volume 3-4.

<https://tlfnews.wordpress.com/2019/11/29/big-data-italian-authorities-published-guidelines-and-policy-recommendation-on-competition-consumer-protection-and-data-privacy-issues/>. See for unfair competition law, data sharing and social media platforms: Catalina Goanta, ‘Facebook’s Data Sharing Practices under Unfair Competition Law’, TTLF Newsletter on Transatlantic Antitrust and IPR Developments Stanford-Vienna Transatlantic Technology Law Forum, Stanford University, 2018 Volume 2.

<https://tlfnews.wordpress.com/2018/06/08/facebooks-data-sharing-practices-under-unfair-competition-law/> See for competition law as a driver for digital innovation and its relationship with IP law: Josef Drexl, ‘Politics, digital innovation, intellectual property and the future of competition law’, *Concurrences Review* 4 (2019), 2-5. <https://www.concurrences.com/en/review/issues/no-4-2019/foreword/politics-digital-innovation-intellectual-property-and-the-future-of-competition>

<sup>5</sup> All European Member States have civil law systems. Great Britain, as the USA, has a common law system.

<sup>6</sup> WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI), Second Session, Draft Issues Paper on Intellectual Property Policy and Artificial Intelligence, prepared by the WIPO Secretariat, December 13, 2019 [https://www.wipo.int/about-ip/en/artificial\\_intelligence/policy.html](https://www.wipo.int/about-ip/en/artificial_intelligence/policy.html)

<sup>7</sup> *Ibid*. See also: [https://www.wipo.int/meetings/en/doc\\_details.jsp?doc\\_id=470053](https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=470053)

<sup>8</sup> WIPO is planning to launch a digital time stamping service that will help innovators and creators prove that a certain digital file was in their possession or under their control at a specific date and time. See: ‘Intellectual

## *Sui generis database rights*

Hand-labelled, annotated machine learning training datasets are awarded with either a database right or a *sui generis* database right in Europe.<sup>9</sup> Although the 1996 Database Directive was not developed with the data-driven economy in mind, there has been a general tendency of extensive interpretation in favor of database protection.<sup>10</sup> A database right can be qualified as either a neighboring (ancillary or related) right (however shorter in duration i.e. 15 years), or a true *sui generis* IP right, but not as a full copyright. A *sui generis* database right is an IP right with characteristics of a property right, and is awarded after a substantial investment in creating and structuring the database, be it money or time, has been made. Businesses usually consider hand-labelled, tagged training *corpora* to be an asset that they can license or sell to another company. This applies to the AI system's output data as well. As all IP rights, (*sui generis*) database rights are subject to exhaustion.<sup>11</sup> In the USA, no *sui generis* database right exists on augmented input or output data.<sup>12</sup> What Europe and the USA do have in common, is that any existing IP rights on input data need to be cleared before processing.

Feeding training data to the machine qualifies as a reproduction of works, and requires a license.<sup>13</sup> The training *corpus* usually consists of copyrighted images, videos, audio, or text. If the training *corpus* contains non-public domain (copyrighted) works or information protected by database rights - and no text and datamining (TDM)<sup>14</sup> exception applies- *ex ante* permission to use and process must

---

property in a data-driven world', WIPO Magazine October 2019

[https://www.wipo.int/wipo\\_magazine/en/2019/05/article\\_0001.html](https://www.wipo.int/wipo_magazine/en/2019/05/article_0001.html) The time stamping initiative is a digital notary service that resembles the BOIP i-Depot, see <https://www.boip.int/en/entrepreneurs/ideas>

<sup>9</sup> Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (Database Directive):

<https://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML> For an analysis of the rules on authorship and joint authorship of both databases and database makers' *sui generis* rights, and how to overcome potential problems contractually see: Michal Koščík & Matěj Myška (2017), 'Database authorship and ownership of *sui generis* database rights in data-driven research', *International Review of Law, Computers & Technology*, 31:1, 43-67, DOI: [10.1080/13600869.2017.1275119](https://doi.org/10.1080/13600869.2017.1275119)

<sup>10</sup> See also CJEU, Case [C-490/14 Verlag Esterbauer](https://eur-lex.europa.eu/legal-content/EN/TJ/?uri=CELEX:62014CJ0490), The CJEU notes that the term "database" is to be given a wide interpretation. In the case of hand-labelled data for supervised machine learning, application of the Database Directive is not really straight forward. The Database Directive does not distinguish between hand and machine coding in what it protects, only between digital and analogue databases. It has been evaluated for the second time in 2018, see <https://ec.europa.eu/digital-single-market/en/protection-databases>

<sup>11</sup> Mezei, Péter, Digital First Sale Doctrine Ante Portas -- Exhaustion in the Online Environment (June 7, 2015). *JIPITEC – Journal of Intellectual Property, Information Technology and E-Commerce Law*, Vol. 6., Issue 1., p. 23-71, 2015. Available at SSRN: <https://ssrn.com/abstract=2615552>. This rule has two exceptions: online transmission of the database and lending or rental of databases do not result in exhaustion.

<sup>12</sup> Bernt Hugenholtz, 'Something Completely Different: Europe's *Sui Generis* Database Right', in: Susy Frankel & Daniel Gervais (eds.), *The Internet and the Emerging Importance of New Forms of Intellectual Property* (2016), 205-222. See also SCOTUS landmark decision *Feist: Feist Publications, Inc., v. Rural Telephone Service Company, Inc.*, 499 U.S. 340 (111 S.Ct. 1282, 113 L.Ed.2d 358), No. 89-1909. <https://www.law.cornell.edu/supremecourt/text/499/340>

<sup>13</sup> See also James Grimmelmann, 'Copyright for Literate Robots' (101 *Iowa Law Review* 657 (2016)), U of Maryland Legal Studies Research Paper No. 2015-16) 678, <https://scholarship.law.cornell.edu/facpub/1481/>. Access to out-of-commerce works held by cultural heritage institutions also requires clearance. In Europe, this license can be obtained from collective rights organisations (Article 8 CDSM Directive).

<sup>14</sup> The non-technologically neutral definition of 'text and data mining' in the CDSM Directive is '*any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations*'.

be obtained from the rightsholders (for both scientific, commercial and non-commercial training purposes).

### *Clearance of machine learning training datasets*

Unlicensed (or uncleared) use of machine learning input data potentially results in an avalanche of copyright (reproduction right) and database right (extraction right) infringements.<sup>15</sup> Some content owners will have an incentive to prohibit or monetize data mining.<sup>16</sup> Three solutions that address the input (training) data copyright clearance problem and create breathing room for AI developers, are the implementation of a broadly scoped, mandatory TDM exception (or even a right to machine legibility)<sup>17</sup> covering all types of data (including news media) in Europe,<sup>18</sup> the Fair Learning principle in the USA<sup>19</sup> and the establishment of an online clearinghouse for machine learning training datasets. Each solution promotes the urgently needed freedom to operate and removes roadblocks for accelerated AI-infused innovation.

### *Three solutions*

The TDM exceptions were originally not created with machine learning training datasets in mind. Prominent scholars advocating the introduction of robust TDM provisions to make Europe fit for the digital age and more competitive vis-a-vis the United States and China are Bernt Hugenholtz and Christophe Geiger. The '[Joint Comment to WIPO on Copyright and Artificial Intelligence](#)' addresses - *inter alia*- challenges related to machine learning and the much needed freedom to use training *corpora*. This 'amicus brief' discusses solutions such as individual and collective TDM licenses/exceptions, whether for commercial or scientific objectives.

On the other side of the Ocean, Mark Lemley and Bryan Casey introduced the concept of Fair Learning.<sup>20</sup> The authors contend that AI systems should generally be able to use databases for training whether or not the contents of that database are copyrighted. Permitting copying of works for non-expressive purposes will be -in most cases- a properly balanced, elegant policy-option to remove IP obstacles for training machine learning models and is in line with the idea/expression dichotomy.

A third solution could be the establishment of an online clearinghouse for machine learning training datasets. An ex ante or ex post one-stop-shop resembling a collective rights society, however on the basis of a sui generis compulsory licensing system. A framework that would include a right of

---

<sup>15</sup> Whether for research purposes or for commercial product development purposes.

<sup>16</sup> Bernt Hugenholtz, The New Copyright Directive: Text and Data Mining (Articles 3 and 4), Kluwer Copyright Blog (July 24, 2019), <http://copyrightblog.kluweriplaw.com/2019/07/24/the-newcopyright-directive-textand-data-mining-articles-3-and-4/?print=print>. Article 4 CDSM allows right holders to opt out of the TDM exemption.

<sup>17</sup> Ducato, Rossana and Strowel, Alain M., 'Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to Machine Legibility' (October 31, 2018). CRIDES Working Paper Series, 2018. Available at SSRN: <https://ssrn.com/abstract=3278901>

<sup>18</sup> Geiger, Christophe and Frosio, Giancarlo and Bulayenko, Oleksandr, 'The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects' (March 2, 2018). Centre for International Intellectual Property Studies (CEIPI) Research Paper No. 2018-02.

<sup>19</sup> Lemley, Mark A. and Casey, Bryan, Fair Learning (January 30, 2020). Available at SSRN:

<https://ssrn.com/abstract=3528447>

<sup>20</sup> *Ibid.* (*supra* note 19)

remuneration for rights holders, but without the right to prohibit data usage for commercial and scientific machine learning purposes.<sup>21</sup> With a focus on permitted, free flow of interoperable data.

### *Public versus private data*

Another legal dimension that we can distinguish is on the one hand public (in the hands of the government) machine generated (non) personal data, and private (in the hands of the business community) machine generated (non) personal data. By machine generated data, we mean in particular information and data that are continuously generated by edge devices in the Internet of Things (IoT).<sup>22</sup> These edge devices are connected via edge (or fog) nodes (transmitters) to data centers that together with edge servers form the cloud. This architecture is known as [edge computing](#).

### *Legal reform*

Mandatory TDM exceptions are a sine qua non for machine learning in Europe.<sup>23</sup> A right of fair, remunerated text and data use to train an AI system needs to be mandatory and without opt outs. Would a broadly scoped TDM exception be an optional limitation, with room for Member States to implement their own rules, the Digital Single Market will become fragmented instead of harmonized. A right to machine legibility that drastically improves access to data, will greatly benefit the growth of the European AI-ecosystem.<sup>24</sup>

Besides implementing broader scoped TDM exceptions, it is opportune that the EU Database Directive 96/9/EC shall be reformed by the EU Commission to prevent that data generated by connected edge devices qualifies for *sui generis* database right protection. Edge computing data must not be monopolized.<sup>25</sup>

## 2. Technical dimensions of data in machine learning

Most AI models need centralized data. In the current, dynamic field of machine learning<sup>26</sup>, hand-labelled training datasets are a sine qua non for supervised machine learning, which uses regression and classification techniques to solve its prediction and optimization problems. This process mimics biological cognition. In contrast, unsupervised machine learning, which utilizes association and clustering (pattern recognition) techniques, uses unlabelled (unstructured) datasets as an input to train its algorithms to discover valuable regularities in digital information. Semi-supervised learning employs a combination of structured and unstructured training datasets to feed our thinking machines.

---

<sup>21</sup> See also WIPO (*supra* note 6)

<sup>22</sup> Such as in smart cities, smart energy meters, Wi-Fi lamps and user gadgets including smart wearables, televisions, smart cameras, smartphones, game controllers and music players.

<sup>23</sup> Countries with more room in their legal frameworks i.e. less legal barriers to train machine learning models are Switzerland, Canada, Israel, Japan and China.

<sup>24</sup> Ducato and Strowel (*supra* note 17)

<sup>25</sup> Such an innovation friendly reform directly impacts the Digital Single Market. It is to be hoped that the necessary policy space to realize these much needed revisions exists in Brussels.

<sup>26</sup> For the latest scientific breakthrough in machine learning methods see: Matthew Vollrath, 'New machine learning method from Stanford, with Toyota researchers, could supercharge battery development for electric vehicles', February 19, 2020 <https://news.stanford.edu/2020/02/19/machine-learning-speed-arrival-ultra-fast-charging-electric-car/> According to Stanford professors Stefano Ermon and William Chueh the machine isn't biased by human intuition. The researcher's ultimate goal is to optimize the process of scientific discovery itself.



Data in machine learning can be discrete or continuous, numerical and categorical. AI systems that utilize deep learning techniques for predictive analysis and optimization, contain deep layers of artificial neural networks, with representation learning.<sup>27</sup> Artificial deep neural networks (ANN's and DNN's) rudimentarily mimic the architecture of human biological brains and are comprised of simplified, artificial neuron layers. Anno 2020 DNN's do not yet have axon's, soma, dendrites, neurotransmitters, plasticity, cerebral cortices and synaptic cores. In the field of AI, data mining, statistics, engineering and neuroscience [converge](#).

### *Deep reinforcement learning*

Reinforcement learning does not require existing input datasets. Instead, the model learns from data from simulations and games using a reward system based on continuous feedback. Deep reinforcement learning systems, such as AlphaGo, are not easy to train. Too many correlations in the data interfere with its goal-oriented algorithms' stable learning process. [Inference](#) applies the capabilities of a pre-trained deep learning system to new datasets, to predict its output in the form of new, useful real-world values and information.

Transfer learning is a machine learning method that seeks to apply a certain solution model for a particular problem to another, different problem. Applying a pre-trained model to new (and smaller) datasets can turn a one trick pony into the ultimate synthetic multitasker.

Evolutionary computing uses genetic optimization algorithms inspired by neo-Darwinian evolution theory.<sup>28</sup> Genetic algorithms can be used standalone<sup>29</sup>, or to train ANN's and DNN's and to identify suitable training *corpora*.

The approaches described above are all centralized machine learning techniques. [Federated learning](#), in contrast, trains algorithms that are distributed over multiple decentralized edge devices in the Internet of Things. These mobile devices -such as your smartphone- contain local data samples, without exchanging their data samples. The interconnected IoT devices collaboratively train a model under a central server.<sup>30</sup> Federated Learning is a scalable, distributed machine learning approach which enables model training on a large *corpus* of decentralized data.<sup>31</sup> "*Federated learning embodies the principles of focused data collection and minimization, and can mitigate many of the*

---

<sup>27</sup> An example of such an AI system is a generative adversarial network, which consists of two different neural networks competing in a game.

<sup>28</sup> Drexl, Josef and Hilty, Reto and Beneke, Francisco and Desautettes, Luc and Finck, Michèle and Globocnik, Jure and Gonzalez Otero, Begoña and Hoffmann, Jörg and Hollander, Leonard and Kim, Daria and Richter, Heiko and Scheuerer, Stefan and Slowinski, Peter R. and Thonemann, Jannick, Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective (October 8, 2019). Max Planck Institute for Innovation & Competition Research Paper No. 19-13. Available at SSRN: <https://ssrn.com/abstract=3465577>

<sup>29</sup> For example in NASA Antenna. See: Hornby, Greg & Globus, AI & Linden, Derek & Lohn, Jason. (2006), 'Automated Antenna Design with Evolutionary Algorithms', Collection of Technical Papers - Space 2006 Conference. 1. 10.2514/6.2006-7242. [https://ti.arc.nasa.gov/m/pub-archive/1244h/1244%20\(Hornby\).pdf](https://ti.arc.nasa.gov/m/pub-archive/1244h/1244%20(Hornby).pdf)

<sup>30</sup> Kairouz, Peter & McMahan, H. & Avent, Brendan & Bellet, Aurélien & Bennis, Mehdi & Bhagoji, Arjun & Bonawitz, Keith & Charles, Zachary & Cormode, Graham & Cummings, Rachel & D'Oliveira, Rafael & El Rouayheb, Salim & Evans, David & Gardner, Josh & Garrett, Zachary & Gascón, Adrià & Ghazi, Badih & Gibbons, Phillip & Gruteser, Marco & Zhao, Sen. (2019). 'Advances and Open Problems in Federated Learning', <https://arxiv.org/pdf/1912.04977.pdf>

<sup>31</sup> Bonawitz, Keith & Eichner, Hubert & Grieskamp, Wolfgang & Huba, Dzmitry & Ingerman, Alex & Ivanov, Vladimir & Kiddon, Chloe & Konečný, Jakub & Mazzocchi, Stefano & McMahan, H. & Overveldt, Timon & Petrou, David & Ramage, Daniel & Roselander, Jason. (2019), 'Towards Federated Learning at Scale: System Design', <https://arxiv.org/pdf/1902.01046.pdf>

*systemic privacy risks and costs resulting from traditional, centralized machine learning and data science approaches.*<sup>32</sup> It brings the code to the data, instead of bringing the data to the code.<sup>33</sup> In other words, there is no need for sharing data.

### 3. Data: contracts, property law and trade secrets

IP on training data and data management systems is subject to both property law aspects and proprietary rights *in rem* that are enforceable against everyone. Data is not a purely immaterial, non-physical object in the legal (not the natural-scientific) meaning of the word. However, if a party to a dataset transaction has acquired a contractual claim right in exchange for material benefits provided by him, there is a proprietary right. This proprietary right *in rem* is subject to transfer, license and delivery.

The attitude of the parties, and their legal consequence-oriented behaviour when concluding contracts about datasets and their proprietary aspects may perhaps prevail over the absence of a clear legal qualification of data<sup>34</sup> (or information) in the law. In this case, party intentions go beyond the legal void.<sup>35</sup> In other words, legislative gaps can be remedied by contracts.<sup>36</sup>

Legal ownership, or property, is different from an IP right. IP is a proprietary right *in rem*. An IP right can entail a right to use data, in the form of a license.

#### *Extra layers of rights will not bring more innovation*

Raw non personal machine generated data are not protected by IP rights.<sup>37</sup> Introducing an absolute data property right or a (neighboring) data producer right for augmented machine learning training datasets, or other classes of data, is not opportune. Economic literature has made clear that there are no convincing economic, or innovation policy arguments for the introduction of a new layer of rights, especially due to the absence of an incentive and reward problem for the production and analysis of datasets.<sup>38</sup>

---

<sup>32</sup> *Ibid.* (supra note 30)

<sup>33</sup> *Ibid.* (supra note 31)

<sup>34</sup> Tjong Tjin Tai, Eric, 'Een goederenrechtelijke benadering van databestanden', *Nederlands Juristenblad*, 93(25), 1799 - 1804. Wolters Kluwer, ISSN 0165-0483. The author contends that data files should be treated analogous to property of tangible objects within the meaning of Book 3 and 5 of the Dutch Civil Code, as this solves several issues regarding data files.

<sup>35</sup> Until new European legislation creates clarity, gaps and uncertainties will have to be filled by the courts.

<sup>36</sup> Unfortunately, licensing large datasets commercially almost never works out in practice.

<sup>37</sup> For further reading about IP and property rights vested in private data see Begonia Otero, 'Evaluating the EC Private Data Sharing Principles: Setting a Mantra for Artificial Intelligence Nirvana?', 10 (2019) JIPITEC 87 para 1. <https://www.jipitec.eu/issues/jipitec-10-1-2019/4878>. For non-personal machine generated data see P. Bernd Hugenholtz, 'Data Property: Unwelcome Guest in the House of IP (25 August 2017), <http://copyrightblog.kluweriplaw.com/2017/08/25/data-producers-right-unwelcome-guest-house-ip/> and Ana Ramalho, 'Data Producer's Right: Power, Perils & Pitfalls' (Paper presented at Better Regulation for Copyright, Brussels, Belgium 2017)

<sup>38</sup> Kerber, Wolfgang, 'A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis' (October 24, 2016). *Gewerblicher Rechtsschutz und Urheberrecht, Internationaler Teil (GRUR Int)*, 11/2016, 989-999. See also Landes, William M., and Richard A. Posner. "An Economic Analysis of Copyright Law." *The Journal of Legal Studies*, vol. 18, no. 2, 1989, pp. 325–363. JSTOR, [www.jstor.org/stable/3085624](http://www.jstor.org/stable/3085624)



Moreover, additional exclusive rights will not automatically bring more innovation. Instead, it will result in overlapping IP rights and database right thickets.<sup>39</sup> The introduction of a sui generis system of protection for AI-generated Creations & Inventions is -in most industrial sectors- not necessary since machines do not need incentives to create or invent.<sup>40</sup> Where incentives are needed, IP alternatives exist. Finally, there are sufficient IP instruments to protect the various components of the AI systems that process data, create and invent.<sup>41</sup> Because of theoretical cumulation of copyrights, patents, trade secrets and database rights, protection overlaps may even exist.<sup>42</sup>

### *Public Property from the Machine*

Non-personal data that is autonomously generated by an AI system and where upstream and downstream no significant human contribution is made to its creation, should fall into the public domain.<sup>43</sup> It should be open data, excluded from protection by the Database Directive, the Copyright Directive<sup>44</sup> and the Trade Secrets Directive.

These open, public domain datasets can then be shared freely without having to pay compensation and without the need for a license. No monopoly can be established on this specific type of database. I would like to call these AI Creations "*Res Publicae ex Machina*"<sup>45</sup> (Public Property from the Machine). Their classification can be clarified by means of an official public domain status stamp or marking (PD Mark status).<sup>46</sup> Freedom of expression and information are core democratic values that -together with proportionality- should be internalized in our IP framework. Reconceptualizing and strengthening the public domain paradigm within the context of AI, data and IP is an important area for future research.<sup>47</sup>

### *Data as trade secret*

In practise however, to safeguard investments and monetize AI applications, companies will try hard either to keep the data a trade secret or to protect the overall database, whether it was hand-coded

---

<sup>39</sup> James Boyle, *The Public Domain: Enclosing the Commons of the Mind*, (Orange Grove Books 2008) 236

<sup>40</sup> Kop, Mauritz, *AI & Intellectual Property: Towards an Articulated Public Domain* (June 12, 2019). Forthcoming *Texas Intellectual Property Law Journal* 2020, Vol. 28. Available at SSRN: <https://ssrn.com/abstract=3409715>  
The legal concept of *Res Publicae ex Machina* is a catch-all solution.

<sup>41</sup> Exhaustion of certain IP rights may apply, see note 11. See also Shubha Ghosh and Irene Calbol, '*Exhausting Intellectual Property Rights: A Comparative Law and Policy Analysis*', (CUP 2018), 101

<sup>42</sup> *Ibid.* Kop (*supra* note 40). See also Deltorn, Jean-Marc and Macrez, Franck, *Authorship in the Age of Machine learning and Artificial Intelligence* (August 1, 2018). In: Sean M. O'Connor (ed.), *The Oxford Handbook of Music Law and Policy*, Oxford University Press, 2019 (Forthcoming) ; Centre for International Intellectual Property Studies (CEIPI) Research Paper No. 2018-10. Available at SSRN: <https://ssrn.com/abstract=3261329>

<sup>43</sup> This means that there should be no sui generis database right vested in such datasets in Europe. No contract or license will be required for the consent of the right holders for analysis, use or processing of the data.

<sup>44</sup> Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (CDSM Directive), <https://eur-lex.europa.eu/eli/dir/2019/790/oj>

<sup>45</sup> Kop (*supra* note 40). The legal concept of *Res Publicae ex Machina* is a catch-all solution.

<sup>46</sup> Autonomously generated non personal datasets should be public domain.

<sup>47</sup> Hilty, Reto and Hoffmann, Jörg and Scheuerer, Stefan, *Intellectual Property Justification for Artificial Intelligence* (February 11, 2020). Draft chapter. Forthcoming in: J.-A. Lee, K.-C. Liu, R. M. Hilty (eds.), *Artificial Intelligence & Intellectual Property*, Oxford, Oxford University Press, 2020, Forthcoming; Max Planck Institute for Innovation & Competition Research Paper No. 20-02. Available at SSRN: <https://ssrn.com/abstract=3539406>  
The article debates the question of justification of IP rights for both AI as a tool and AI-generated output in light of the theoretical foundations of IP protection, from both legal embedded deontological and utilitarian economic positions.

or machine generated. From an AI perspective, the various strategies to maximize the quality and value of a company's IP portfolio can differ for database rights, patents and trade secrets on the input and output of an AI system. Moreover, this strategy can differ per sector and industry (e.g. software, energy, art, finance, defence).

As legal uncertainty about the patentability of AI systems<sup>48</sup> is causing a shift towards trade secrets, legal uncertainty about the protection and exclusive use of machine generated databases is causing a similar shift towards trade secrets. Although it is not written with the data driven economy in mind, the large scope of the definition of a trade secret in the EU means that derived and inferred data can in theory be classified under the Trade Secrets Directive.<sup>49</sup> This general shift towards trade secrets to keep competitive advantages results in a disincentive to disclose information and impedes on data sharing.<sup>50</sup>

In an era of exponential innovation, it is urgent and opportune that both the Trade Secrets Directive, the Copyright Directive and the Database Directive shall be reformed by the EU legislature with the data-driven economy in mind.

#### 4. EU open data sharing initiatives

Data can be shared between Government, Businesses, Institutions and Consumers. Within an industry sector or cross-sectoral.

Important European initiatives in the field of open data<sup>51</sup> and data sharing are: the [Support Centre for Data Sharing](#) (focused on data sharing practices), the [European Data Portal](#) (EDP, data pooling per industry i.e. sharing open datasets from the public sector), the Open Data Europe Portal (ODP, sharing data from European institutions), the [Free flow of non-personal data](#) initiative (including the FFD-Regulation, cyber security and self-regulation) and the [EU Blockchain Observatory and Forum](#).

A European initiative in the strongly related field of AI is the European AI Alliance, established by the EU Commission. An international project on AI and -inter alia- training data is the "[AI and Data Commons](#)" of the ITU (International Telecommunication Union).

#### *EU Data Strategy*

On February 19 2020 The EU Commission published its 'EU Data Strategy'.<sup>52</sup> The EU aims to become a leading role model for a society empowered by data and will to that end create Common European Data Spaces in verticals such as Industrial Manufacturing, Health, Energy, Mobility, Finance,

---

<sup>48</sup> Kop (*supra* note 40). Not opting for the patent route poses the risk of (bona fide) independent invention by someone else who does opt for the patent route instead of the trade secret strategy.

<sup>49</sup> Wachter, Sandra and Mittelstadt, Brent, 'A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI' (October 05, 2018). Columbia Business Law Review, 2019(1). [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3248829](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3248829)

<sup>50</sup> Kop (*supra* note 40). Besides that, uncertainty about the scope of the TDM exceptions leads to litigation.

<sup>51</sup> For certain AI systems, open data should be required for safety reasons.

<sup>52</sup> European Commission, 'A European strategy for data', Brussels, 19.2.2020 COM(2020) 66 final, [https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020\\_en.pdf](https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf) & [https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy\\_en#documents](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en#documents)

Agriculture and Science. An industrial package to further stimulate data sharing follows in March 2020.

In addition, the EU Commission has appointed an [Expert Group](#) to advise on Business-to-Government Data Sharing (B2G).<sup>53</sup> In its final report, the Expert Group recommends the creation of a recognized data steward function in both public and private sectors, the organization of B2G data-sharing collaborations and the implementation of national governance structures by Member States.<sup>54</sup> The aim of B2G data sharing is to improve public service, deploy evidence-based policy and advise the EU Commission on the development of B2G data sharing policy.

In its 2019 Policy & Investment Recommendations, the High-Level Expert Group on Artificial Intelligence (AI-HLEG) also devoted an entire section to fostering a European data economy, including data sharing recommendations, data infrastructure and data trusts.<sup>55</sup> Finally, in a recent report, the [German Opinion of the Data Ethics Commission](#) made 75 authoritative recommendations on general ethical and legal principles concerning the use of data and data technology.

Given that data are generated by such a vast and varied array of devices and activities, and used across so many different economic sectors and industries, it is not easy to picture an all-inclusive single policy framework for data.<sup>56</sup>

#### *Dutch vision on B2B data sharing*

At the beginning of this year, the Dutch government published a booklet about the [Dutch Digitization Strategy](#), in which it sets out its vision on data sharing between companies. This vision consists of 3 principles:

Principle 1: Data sharing is preferably voluntary.

Principle 2: Data sharing is mandatory if necessary.

Principle 3: People and companies keep a grip on data.

The Dutch Ministry of Economic Affairs is currently exploring the possibilities of encouraging the use of internationally accepted FAIR principles in sharing private data for AI applications. FAIR stands for (Findable, Accessible, Interoperable, Reusable). The [Personal Health Train](#) initiative builds on FAIR data principles.<sup>57</sup>

---

<sup>53</sup> Towards a European strategy on business-to-government data sharing for the public interest. Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing, Brussels, European Union, February 2020, doi:10.2759/731415 <https://www.euractiv.com/wp-content/uploads/sites/2/2020/02/B2GDataSharingExpertGroupReport-1.pdf> The report provides a detailed overview of B2G data sharing barriers and proposes a comprehensive framework of policy, legal and funding recommendations to enable scalable, responsible and sustainable B2G data sharing for the public interest.

<sup>54</sup> *Ibid.*

<sup>55</sup> High-Level Expert Group on Artificial Intelligence, 'Policy and Investment Recommendations for Trustworthy AI' (European Commission, 26 June 2019). <https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence>

<sup>56</sup> *Ibid.* (*supra* note 6)

<sup>57</sup> Johan van Soest, Chang Sun, Ole Mussmann, Marco Puts, Bob van den Berg, Alexander Malic, Claudia van Oppen, David Towend, Andre Dekker, Michel Dumontier, 'Using the Personal Health Train for Automated and Privacy-Preserving Analytics on Vertically Partitioned Data', *Studies in Health Technology and Informatics* 2018, 247: 581-585

Recent Dutch initiatives in the field of data sharing are the [Dutch Data Coalition](#) (self-sovereignty of data), aimed at cross-sectoral data sharing between companies and institutions, the [Dutch AI Coalition](#) (NL AIC) as well as some hands-on Data Platform and Data Portal projects from leading academic hospitals, Universities of Technology and frontrunning companies.

## 5. Mixed datasets: 2 laws (GDPR & FFD Regulation) in tandem

More and more datasets consist of both personal and non-personal machine generated data; both the General Data Protection Regulation (GDPR)<sup>58</sup> and the Regulation on the free flow of non-personal data (FFD)<sup>59</sup> apply to these "mixed datasets". The Commission has drawn up [guidelines](#) for these mixed datasets where both the FFD Regulation and the GDPR apply, including its right to data portability.<sup>60</sup> Based on these two Regulations, data can move freely within the European Union.<sup>61</sup>

### *Market barriers for early-stage AI-startups*

The GDPR thoroughly protects the personal data of EU citizens. In some cases however, GDPR legislation is also hampering the European internal market with regard to the rapid rollout of AI and data startups (SME's). This applies in particular to a smaller group of early-stage AI-startups who often lack sufficient resources to hire a specialized lawyer or a Data Protection Officer. Therefore, these companies are hesitant to do anything spectacular with personal data,<sup>62</sup> and otherwise in large public-private consortia in which one operates 'gründlich', but where it takes (too) long to create the necessary trust among the participants. This hinders the innovative performance of early-stage AI-startups. In that sense, complex data protection rules do not encourage ambitious moonshot thinking, creative, revolutionary AI and data field experiments and the design of clever products that solve real-world problems. It is paramount that the whole field has a good grasp on the legal

---

<sup>58</sup> Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). A new European ePrivacy Regulation is currently under negotiation. Data protection and privacy are two different things.

<sup>59</sup> Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union (FFD Regulation).

<sup>60</sup> Practical guidance for businesses on how to process mixed datasets: <https://ec.europa.eu/digital-single-market/en/news/practical-guidance-businesses-how-process-mixed-datasets>

<sup>61</sup> Besides the GDPR, the Law Enforcement Directive (LED) regulates requirements aimed at ensuring that privacy and personal data are adequately protected during the use of AI-enabled products and services. LED: Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA.

<sup>62</sup> I speak from personal experience in our law firm. This concerns especially European AI-startups who often do not have the necessary budget to be properly advised on how to navigate data protection and data sharing regulation. See for a first report that confirms this claim: OECD Report 'Enhancing Access to and Sharing of Data - Reconciling Risks and Benefits for Data Re-use across Societies', November 26, 2019, Chapter 4. <https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm>

dimensions of their data. And that there are no significant restrictions and market barriers in that important early stage.<sup>63</sup> Sharing data is simply a necessary condition for a [successful AI ecosystem](#).<sup>64</sup>

### *Precautionary principle*

A second axiom that has the potential to inhibit rapid scientific advances in the EU -in case of expected large risks or unknown risks- is the [precautionary principle](#). EU lawmakers have a tendency to minimize risk and prevent all possible negative scenarios ex ante via legislation. It doesn't make drafting directives and regulations faster. Rigid application of the precautionary principle in EU law promotes excessive caution and hinders progress. It remains at odds with accelerated technological innovation.<sup>65</sup>

## 6. California Consumer Privacy Act (CCPA 2020)

The GDPR also has some important advantages for European startups and scaleups. The advantage of the GDPR is that it is now the international standard in the field of the use of personal data when doing business internationally.<sup>66</sup> Partly for this reason, California has largely taken over the spirit/contents<sup>67</sup> of the GDPR, and implemented it -with a fundamental American approach- in its own regulations that better protect consumer data and safeguard the trade thereof.<sup>68</sup> The [California Consumer Privacy Act \(CCPA 2020\)](#), state-level privacy legislation, came into force on January 1, 2020.<sup>69</sup> If European startups and scaleups are completely GDPR-proof, there will be no privacy legislation anywhere in the world that will require major changes to their personal data protection policy, including the associated legal uncertainty and legal costs. This is a significant competitive advantage. From that lens, European tech startups and AI-scaleups have a head start on their competitors from outside the European Union.<sup>70</sup>

---

<sup>63</sup> A solution that takes away legal roadblocks and encourages market entry of early-stage AI-startups could be targeted government funding in the form of knowledge vouchers.

<sup>64</sup> From this point of view, innovation remains at odds with privacy.

<sup>65</sup> In certain domains, performing independent audits and conformity assessments by notified bodies might be a better option. Especially in a civil law legal tradition, where lawmakers draft concise statutes that are meant to be exhaustive.

<sup>66</sup> With 500 million consumers, Europe is the largest single market in the world.

<sup>67</sup> For a close comparison of the GDPR and California's privacy law, see Chander, Anupam and Kaminski, Margot E. and McGeeveran, William, 'Catalyzing Privacy Law' (August 7, 2019). U of Colorado Law Legal Studies Research Paper No. 19-25. Available at SSRN: <https://ssrn.com/abstract=3433922> The article contends that California has emerged as an alternate contender in the race to set the new standard for privacy (which, as mentioned in note 3, is not always the same as data protection).

<sup>68</sup> Mark A. Lemley, 'The Splinternet', Lange Lecture Duke Law School, January 22 2020, <https://www.youtube.com/watch?v=5MEI4c5BVCw>

<sup>69</sup> <https://oag.ca.gov/privacy/ccpa>

<sup>70</sup> Such as China, India, Japan, South Korea and Taiwan.

## 7. Future EU AI and Data Regulation: CAHAI & EU Commission Whitepaper

Transformative technology is not a zero sum game, but a [win-win strategy](#) that creates new value. The Fourth Industrial Revolution will create a world where anything imaginable to improve the human condition, could actually be built.<sup>71</sup>

The CAHAI ([Ad Hoc Committee on Artificial Intelligence](#)), established by the Committee of Ministers of the Council of Europe<sup>72</sup> is currently examining the possibility of a binding legal framework for the development, design and application of AI and data, based on the universal principles and standards of the Council of Europe on human rights, democracy and the rule of law. The CAHAI expects to be able to report by March 2020 on the possibilities and necessity of new legislation.

Both data sharing practices and AI-Regulation are high on the EU Commission's agenda. On February 19<sup>th</sup> 2020, the EU Commission published its 'White Paper On Artificial Intelligence - A European approach to excellence and trust'.<sup>73</sup> Fortunately, the White Paper uses a risk-based approach, not a precautionary principle-based approach. The Commission 'supports a regulatory and investment oriented approach with the twin objective of promoting the uptake of AI and of addressing the risks associated with certain uses of this new (data-driven) technology.'<sup>74</sup> In its White Paper, the Commission addresses issues concerning the scope of a future EU regulatory framework and -to ensure inclusiveness and legal certainty- discusses requirements for the use of training datasets.<sup>75</sup> In addition, the Commission contends that independent audits, certification and prior conformity assessments<sup>76</sup> for high risk areas like Health and Transportation, could be entrusted to notified bodies (instead of commercial parties) designated by Member States. The Commission concludes with the desire to become a global hub for data and to restore technological sovereignty.

### *Pareto optimum*

When developing informed transformative tech related policies, the starting point is to identify the desired outcome.<sup>77</sup> In the case of IP policy, that outcome would be to compose a regime that balances underprotection and overprotection of IP rights per economic sector. IP is supposed to serve as a regulatory system of stimulation of creation and innovation that uses market dynamisms to reach this objective.<sup>78</sup> The goal should be no less than a *Pareto optimum* and if possible a *Pareto*

---

<sup>71</sup> Autonomous AI agents that utilize data and deep learning techniques to continuously perform and improve at its tasks already exist. AI agents that autonomously invent novel technologies and create original art. These AI systems need data to mature.

<sup>72</sup> The Council of Europe, located in Strasbourg, France is not the same governing body as the European Commission. The Council of Europe is not part of the European Union. The European Court of Human Rights, which enforces the ECHR, is part of the Council of Europe.

<sup>73</sup> European Commission, White Paper on Artificial Intelligence - A European approach to excellence and trust, Brussels, 19.2.2020 COM(2020) 65 final, [https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020\\_en.pdf](https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf)

<sup>74</sup> *Ibid.*

<sup>75</sup> *Ibid.*

<sup>76</sup> Alternative Regulatory Instruments (ARIs) such as the AI Impact Assessment, see: <https://airecht.nl/blog/2018/ai-impact-assessment-netherlands> See also: Carl Vander Maelen, 'From opt-in to obligation? Examining the regulation of globally operating tech companies through alternative regulatory instruments from a material and territorial viewpoint', International Review of Law, Computers & Technology, 2020, DOI: [10.1080/13600869.2020.1733754](https://doi.org/10.1080/13600869.2020.1733754)

<sup>77</sup> See also WIPO (supra note 8). WIPO is comparing the main government instruments and strategies concerning AI and IP regulation and will create a dedicated website that collects these resources for the purpose of information sharing.

<sup>78</sup> Hilty (supra note 47)



*improvement* by incentivizing innovation, encouraging scientific progress and increasing overall prosperity.<sup>79</sup>

### *Modalities of AI-regulation*

Law is just one modality of AI-regulation.<sup>80</sup> Other important regulatory modalities to balance the societal effects of exponential innovation and digital transformation are the actual design of the AI system, social norms and the market.<sup>81</sup> Data [governance](#) should be less fixed on data ownership and more on rules for the usage of data.

The goal should be global open data sharing community with freedom to operate and healthy competition between firms, including unification of data exchange models so that they are interoperable and standardized in the IoT.<sup>82</sup> There is an urgent need for comprehensive, cross sectoral data reuse policies that include standards for interoperability<sup>83</sup>, compatibility, certification and standardization.<sup>84</sup>

Against this background, strengthening and articulation of competition law is more opportune than extending IP rights.<sup>85</sup> Within the context of AI-regulation and data sharing practices, there is no need for adding extra layers of copyrights, database rights, patent rights and trade secret rights.<sup>86</sup>

### *Technology shapes society, society shapes technology*

Society should actively shape technology for good. The alternative is that other societies, with social norms and democratic standards that perhaps differ from our own public values, impose their values on us through the design of their technology.

AI for Good norms, such as data protection by design and by default, as well as Accountability of controllers and processors, transparency, trust and control should be built in the architecture of AI

---

<sup>79</sup> Kop (*supra* note 40)

<sup>80</sup> Smuha, Nathalie A., From a 'Race to AI' to a 'Race to AI Regulation' - Regulatory Competition for Artificial Intelligence (November 10, 2019). Available at SSRN: <https://ssrn.com/abstract=3501410>. The author contends that AI applications will necessitate tailored policies on the one hand, and a holistic regulatory approach on the other, with due attention to the interaction of various legal domains that govern AI.

<sup>81</sup> Lawrence Lessig, The Law of the Horse: What Cyberlaw Might Teach, 113 Harvard Law Review 501-549 (1999)

<sup>82</sup> Otero (*supra* note 37). For user generated data see: Shkabatur, Jennifer, 'The Global Commons of Data' (October 9, 2018). Stanford Technology Law Review, Vol. 22, 2019; GigaNet: Global Internet Governance Academic Network, Annual Symposium 2018. Available at SSRN: <https://ssrn.com/abstract=3263466>

<sup>83</sup> For an example of interconnectivity and interoperability of databases in line with the fundamental rights standards enshrined in the EU Charter: Quintel, Teresa, Connecting Personal Data of Third Country Nationals: Interoperability of EU Databases in the Light of the CJEU's Case Law on Data Retention (March 1, 2018). University of Luxembourg Law Working Paper No. 002-2018. Available at SSRN: <https://ssrn.com/abstract=3132506>

<sup>84</sup> John Wilbanks; & Stephen H Friend, 'First, design for data sharing', (Nature, 2016)

<sup>85</sup> Drexl, (*supra* note 2). The Fourth Industrial Revolution may even require a complete redesign of our current IP regime.

<sup>86</sup> Kop (*supra* note 40). For non-IP policy tools that incentivize innovation, see: Hemel, Daniel Jacob and Ouellette, Lisa Larrimore, 'Innovation Policy Pluralism' (February 18, 2018). Yale Law Journal, Vol. 128, p. 544 (2019); Stanford Public Law Working Paper; Stanford Law and Economics Olin Working Paper No. 516; U of Chicago, Public Law Working Paper No. 664; University of Chicago Coase-Sandor Institute for Law & Economics Research Paper No. 849. Available at SSRN: <https://ssrn.com/abstract=3125784>. See also: Mauritz Kop, 'Beyond AI & Intellectual Property: Regulating Disruptive Innovation in Europe and the United States – A Comparative Analysis' (December 5 2019) <https://law.stanford.edu/projects/beyond-ai-intellectual-property-regulating-disruptive-innovation-in-europe-and-the-united-states-a-comparative-analysis/>

systems and high quality training datasets from the first line of code.<sup>87</sup> In practice, this can be accomplished through technological synergies such as a symbiosis of AI and blockchain technology. Crossovers can offer solutions for challenges concerning the AI-black box, algorithmic bias and unethical use of data.<sup>88</sup> That way, society can benefit from the benevolent side of AI.

Robust, collaborative AI framework development standards such as federated machine learning<sup>89</sup> models provide personalized AI and safeguard data privacy, data protection, data security and data access rights. Using Privacy by Design as a starting point, with built-in public values, the federated learning model is consistent with [Human-Centered AI](#) and the European Trustworthy AI paradigm.<sup>90</sup> As technology shapes society, society shapes technology.

---

<sup>87</sup> Kop (*supra* note 40)

<sup>88</sup> Combination is the key. Examples of potential unethical use of AI are facial recognition and predictive policing.

<sup>89</sup> See note 30 and 31.

<sup>90</sup> High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (European Commission, 8 April 2019). See [https://ec.europa.eu/newsroom/dae/document.cfm?doc\\_id=60419](https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419). See also Paul Opitz, 'European Commission Working on Ethical Standards for Artificial Intelligence (AI)', TTLF Newsletter on Transatlantic Antitrust and IPR Developments Stanford-Vienna Transatlantic Technology Law Forum, Stanford University, 2018 Volume 3-4, <https://ttrlnews.wordpress.com/2018/06/08/european-commission-working-on-ethical-standards-for-artificial-intelligence-ai/>