The European path towards Data Quality and its standardisation in AI: a legal perspective

Nadia Maccabiani*

THE EUROPEAN PATH TOWARDS DATA QUALITY AND ITS STANDARDISATION IN AI: A LEGAL PERSPECTIVE

ABSTRACT: In a data – driven period, with Machine Learning (ML) systems that thrive, owing to the huge data availability (Big Data), and affect people with assessments, predictions and decisions, our focus rests upon some prerequisites which must be met if ML is ever to operate fairly, i.e. data quality and its standardisation. In reference to the underlying (apparently mere) technical procedures, the paper rests on the relevant legal implications in terms of both fundamental rights and regulatory techniques. In this respect, it is the constitutional recovery of the EU through its recently launched Strategies (on Artificial Intelligence and Standardisation) that comes into play, paving the path towards a steering and monitoring role by the European institutions that supports an improving rights-oriented approach and a re-framing of regulatory techniques.

KEYWORDS: Artificial intelligence; big data; data quality; harmonised standard; machine learning

SUMMARY: 1. Premise – 2. A brief outline about Artificial Intelligence, Machine Learning and Big Data – 3. Technical and legal implications of data quality – 4. The European normative approach to data quality – 5. Data quality standards – 6. Legal implications of standardisation – 7. Conflicting rationales: is the EU "climbing back to the top?" – 8. Conclusions.

1. Premise

We are living a period where some "buzzwords", such as Big Data and Artificial Intelligence, not only have become of common use but they also underlie a complex system that is affecting many aspects of our daily life.[1] From this perspective, the EU has proved itself to be aware of the salient importance of the issue, and like the leadership it displayed at the global level

* *Professore Associato di Diritto Pubblico – Dipartimento di Economia e Management – Università degli Studi di Brescia. Mail: nadia.maccabiani@unibs.it.* *The article was subject to a double-blind peer review process.*

[1] L. FLORIDI, *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo, (The Fourth Revolution, How the Infospere is Reshaping Human Reality),* Milan, 2017, 7, names this period the "iperstoria": a society where ICT and their ability to process data are not only important but essential conditions to promote social well – being, individual and collective development, as well as economic growth grounded on intangible assets.

ISSN 2284 – 4503

with the GDPR,[2] it is currently striving[3] to take the lead with its Strategy on a "human – centric" approach to Artificial Intelligence.[4] Bearing this in mind, our research aims at focusing on one component of this wider process, more specifically on data quality and its standardisation, not only because of the well – known legal challenges stemming from poor quality and biased data that feed into Machine Learning algorithms, but also because of its far – reaching legal implications in terms of EU governance and legal system. In this respect, it is not only the EU proposal on Artificial Intelligence (AI Act) that comes into play but also – although less noticed – the EU Strategy on standardisation.[5]
After a brief introduction on the essential features of the technical aspects implied by Artificial Intelligence and data quality, the paper tests "whether and how" the path paved by the European Union adequately copes with the underlying legal challenges. More specifically, "whether and how" this path carries out a re – balance in both directions, a more "rights – oriented" approach to data quality and a "renewed" regulatory approach to its standardization.[6]

2. A brief outline about Artificial Intelligence, Machine Learning and Big Data

Studies about Artificial Intelligence (AI) date back to the '50s[7] but a shared definition of AI is still lacking. However, as stressed by doctrine, commonly "Artificial Intelligence (AI) refers to machines or agents

[2] S. CALZOLAIO, *Protezione dei dati personali,* in *Digesto delle Discipline Pubblicistiche – Aggiornamento,* Milano, 2017, 612, underlines how the GDPR represents the EU starting point to shape the data – society.
[3] Borrowing the words by EC's President Ursula von der Leyen, in https://ec.europa.eu/info/sites/default/files/political – guidelines – next – commission_en_0.pdf. (last visited 25/09/2022)
[4] Many communications, opinions and declarations by the European Commission (COM(2018) 237 final – *Artificial Intelligence for Europe*; COM(2018) 795 final – *Coordinated Plan on Artificial Intelligence*; COM(2019) 168 final – *Building Trust in Human – Centric Artificial Intelligence*; COM(2020) 65 final – *White Paper On Artificial Intelligence – A European approach to excellence and trust*) have paved the way to the recent proposal *Laying down harmonised rules on Artificial Intelligence – Artificial Intelligence Act* (COM(2021) 206 final).
[5] Taking stock of the evidence of the last Report from the European Commission on the implementation of the Regulation (EU) No. 1025/2012, *An EU Strategy on Standardisation* – COM(2022) 31 final – has been drafted in a more consistent way with a «resilient, green and digital EU single market» and a consequent proposal for amending the existing EU Regulation has been submitted (*Proposal for a Regulation of the European Parliament and of the Council amending Regulation (EU) No 1025/2012 as regards the decisions of European standardisation organisations concerning European standards and European standardisation deliverables* – COM(2022) 32 final).
[6] Borrowing and paraphrasing the wordings of the Resolution (85/C 136/01) on a New Approach to Technical Harmonisation and Standards, adopted by the Council on 7th May 1985.
[7] The mathematician commonly considered the founding father of Artificial Intelligence is Alan Turing, with its test addressed to cope with the question «Can machines think?» (A. TURING, *Computing machinery and intelligence,* in *Mind*, LIX, 236, 1950, 433 – 460). This test underlies the conformity of the machine's way of action to the human performance and, as such, this perspective has triggered – in philosophy, neuroscience, cognitive science and psychology – the debate about "mind" and "consciousness" of machines (H. PUTMAN, *Robots: Machines or Artificially created life?*, in *The Journal of Philosophy* , 61, 21, 1964, 668 – 691). The lack of success of this initial approach was due to its Boolean logic and general problem – solving methods that led to elementary reasoning steps (weak methods that gave rise to the so called Good Old Fashioned AI – GOFAI). In the '70s – '80s, a different perspective took place by means of expert systems: they imply "more powerful, domain – specific knowledge that allows larger reasoning steps", and later on (in the '90s), the incorporation of the probabilistic reasoning led to Bayesian networks (that include the formalistic representation of uncertain knowledge and practical algorithmic for probabilistic reasoning, i.e. the representation of "any full joint probability distribution"

that are capable of observing their environment, learning, and based on the knowledge and experience gained, taking intelligent action or proposing decisions".[8] Similarly, a strong AI is still lacking, indeed "human level AI or general AI – programs that can solve an arbitrarily wide variety of tasks, including novel ones, and do so as well as a human"[9] has not been realized yet. As a consequence, we are dealing with weak AI systems, in reference to which the philosophical question whether machines are actually conscious agents (consciousness of itself and its surroundings) is far from being concrete.[10]

Against this background, recent breakthrough developments have brought about a new "hype" for AI: computing power, data availability and new algorithms.[11] Consequently, the EU legislator has tried to draft a legal but flexible definition of AI,[12] listing its different and rapidly evolving techniques in Annex I of its proposal and providing for its periodical updating by the European Commission's delegated power.[13] The first technique set out in Annex I is for Machine Learning that includes "supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning".[14]

Thus, Machine Learning (ML), a subset of AI systems, that has spread since the year 2000[15] differs from an expert – system; it does not need a base of knowledge previously provided by a scientist, but it autonomously acquires its knowledge by means of inference from data[16]: "a computer observes some data, builds a model based on the data, and uses the model as both a hypothesis about the world and a piece of software that can solve problems".[17]In the case of supervised learning, the system is fed with input and output data from which the agent learns a function and, given an input, it predicts the appropriate label (output). Unsupervised learning differs because the agent is fed only with input data

doing so very concisely, see S. RUSSELL, P. NORVIG, *Artificial intelligence – A Modern Approach*, Hoboken, 2021, 22 – 24 and 412.

[8] M. CRAGLIA (eds.), *Artificial Intelligence – A European Perspective*, Report of the Joint Research Centre of the European Commission (JRC113826), Luxembourg, 2018, 19.

[9] S. RUSSELL, P. NORVIG, *Artificial intelligence – A Modern Approach,* cit., 981.

[10] H. PUTMAN, *Robots: Machines or Artificially created life?*, in *The Journal of Philosophy* , 61, 21, 1964, 668 ff.

[11] According to the European Commission White Paper on *Artificial Intelligence – a European approach to excellence and trust* – COM(2020) 65 final, 2: "AI is a collection of technologies that combine data, algorithms and computing power. Advances in computing and the increasing availability of data are therefore key drivers of the current upsurge of AI". Moreover, "AI systems have become more and more statistical and probabilistic and are increasingly powered by a growing variety of data types": *OECD Framework for the classification of AI systems*, in *OECD Digital Economy Papers,* 323, 2022, 35.

[12] According to Article 3, point 1 of the *Artificial Intelligence Act* – COM(2021) 206 final: "artificial intelligence system (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human – defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with".

[13] As stressed by C. CASONATO, G. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione europea in materia di intelligenza artificiale,* in *BioLaw Journal,* 3, 2021, 419, the flexibile approach of the proposal is future – proof.

[14] *Artificial Intelligence Act* – COM(2021)206 final –, Annex I, (a).

[15] S. D'ACQUISTO, *Intelligenza Artificiale – Elementi,* Turin, 2021, 127. As stressed by G.F. ITALIANO, S. CIVITARESE MATTEUCCI, A. PERRUCCI, *L'Intelligenza artificiale: dalla ricerca scientifica alle sue applicazioni. Una introduzione di contesto,* in A. PAJNO, F. DONATI, A. PERRUCCI (eds.), *Intelligenza Artificiale e Diritto: una rivoluzione?,* I, Bologna, 2022, 49 ff., early academic studies on machine learning date back to the '60s, however ML has started to enter our daily life only recently due to the increasing computer power able to process a huge amount of data.

[16] L. PORTINALE, *Intelligenza artificiale: storia, progressi e sviluppi tra speranze e timori,* in *MediaLaws,* 3, 2021, 26.

[17] S. RUSSELL, P. NORVIG, *Artificial intelligence – A Modern Approach,* cit., 651.

and it learns patterns (usually detecting clusters) without any explicit feedback (label). Moreover, in reinforced learning, the agent learns by its own experience and improves its performance by means of a series of rewards and punishments.[18] Lastly, in deep learning, the computation paths from inputs to outputs are organized into many layers (often called neural networks) and, as such, the several intermediate computations performed by the system before producing the output remain "hidden" to the programmer as well.[19]

Similarly, in order for ML to thrive the availability of data is essential. This is the reason why its methodologies have been fostered by the "datafication" process,[20] i.e. the conversion of analogue signals (texts, images, videos, audio) into digital formats.[21] Moreover, the presence of networks that link together different devices and in turn become multiple sources of data detection and collection, a reduction in storage costs,[22] the increasing growth of computational and storage power of Information Technology systems and the availability of new complex algorithmic methods for data analysis are all "ingredients" that are improving the performance of ML systems.[23]

Data is usually defined as "the raw material produced by abstracting the world into categories, measures and other representational forms[...]that constitute the building blocks from which information and knowledge are created".[24] It is categorised in different types according to its provenance, the ways of collection, its nature and scale[25] and it can assume different states: structured, semi – structured, unstructured.[26] Not only AI, but also Big Data are still lacking an accepted uniform definition,[27] however the term "has gained traction[...]becoming a buzzword".[28]

---

[18] According to the definition given by S. RUSSELL, P. NORVIG, *Artificial intelligence – A Modern Approach,* cit., 653.

[19] *Ibidem*, 750.

[20] As stressed by R. KITCHIN, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences,* London, 2014, 98, "the production of big data has been facilitated by the confluence of five technological innovations from the start of the new millennium onwards – growing computational power, dense internetworking, pervasive and ubiquitous computing, indexical and machine – readable identification, and massive distributed storage".

[21] A. DE MAURO, M. GRECO, M. GRIMALDI, *A formal definition of Big Data based on its essential features,* in *Library Review*, 65, 3, 2016, 123.

[22] M. MIRTI, *Il cyberspace – Caratteri e riflessi sulla Comunità internazionale,* Naples, 2021, 46 ff.

[23] A. DE MAURO, M. GRECO, M. GRIMALDI, *op. cit.*, 125.

[24] R. KITCHIN, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences,* cit., 1.

[25] *OECD Framework for the classification of AI systems*, in *OECD Digital Economy Papers,* 323, 2022, 35 – 37, underlines that data can be detected and collected by humans (that usually provide subjective evaluations) or by automated sensors (devices that automatically monitor and record data); moreover, depending on their provenance, data sources are: experts (human knowledge codified into rules and structures such as ontologies, knowledge graphs, analytical functions), individuals or organisations that voluntarily provide data; observed by humans or sensors; synthetic data (usually generated by computer simulation of real life events); derived data (from other data). According to their nature, data may be static (data does not change after it is collected) or dynamic (updated from time – to – time or in real – time). Lastly data can be described according to its scale (volume).

[26] Report of the Data Governance Working Group of the Global Partnership of AI – *The Role of Data in AI* – November 2020, 3. In a similar way, according to the ISO 2015, digital data means the "representation of information".

[27] A. DE MAURO, M. GRECO, M. GRIMALDI, *op. cit.*

[28] R. KITCHIN, *Big Data,* in *International Encyclopedia of Geography,* 2017, 1.

When focusing on its attributes, "The most popular description of big data[...]is the '3V' model, where '3V' refers to volume, variety, and velocity…In addition to the '3V' model, '4V' and '5V' models are emerging as researchers attempt to redefine big data",[29] consequently, a fourth "V" has been added for "veracity", and other "Vs" have surfaced, describing features of Big Data such as value, variability, and visualization.[30] However, as stressed by doctrine, Big Data definitions often go beyond the characteristics of its object and encompass its use and value and, as such, a cross – cutting definition has been advanced: "Big Data is the information asset characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value".[31]
In similar vein, Big data impacts not only on traditional data collection, but also on data processing, and data analysis. Firstly, "the data collection approach has been transformed from traditional methods (e.g., questionnaires and interviews) into a fast and powerful ICT – based method"; secondly "the methods and procedures to process[...]big data must have the capability to handle high volume and real – time data and serve as a filter to decrease data errors and data noise"; lastly, data analysis "requires new approaches and tools that can accommodate big data with different data structures and can process data with different spatial and temporal scales".[32] Accordingly, it "becomes possible to interlink diverse sets of data – personal, transactional, interactional, social, financial, spatial, temporal, and so on – and to analyse them on an individual and collective basis for relationships and patterns"[33]
Bearing this in mind, it goes without saying that, from an economic perspective, investing in high – quality datasets for training ML algorithms is costly and time – consuming and, as such, economies of scale and scope support a cost – effective activity. This is the reason why "data – driven firms are so data – hungry",[34] even if from a legal perspective it is challenging to qualify real data – ownership because of its non – rivalry nature that makes it comparable to common goods, except for the *sui generis* right provided by Directive 96/9/EC with reference to datasets.[35]
Consequently, among the three mentioned enablers (data, computer power and algorithms), our focus is on data, because of its being at the very roots of the algorithms that run ML systems: data is the basic building block and its features become significant not only for the technical performance of the system but also for its legal compliance. On the one hand, as regards technical performance, "the more

[29] J. LIU, J. LI, W. LI, J. WU, *Rethinking big data: A review on the data quality and usage issues,* in *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 2016, 135.
[30] Other "Vs" have been added, such as "vocabulary", "venue", "vagueness", as recalled by C.W. TSAI, C.F. LAI, H.C. CHAO, A.V. VASILAKOS, *Big data analytics: a survey,* in *Journal of Big Data,* 2, 21, 2015, 2. As for a further definition of the features of Big Data, focusing on its difference with small data systems (Exhaustivity, Resolution and Indexicability, Relationality, Flexibility), see R. KITCHIN, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences, cit*., 68 ff.
[31] A. DE MAURO, M. GRECO, M. GRIMALDI*, op cit*., 134.
[32] J. LIU, J. LI, W. LI, J. WU, *op. cit.*, 136.
[33] R. KITCHIN, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences,* cit., 75.
[34] M. CRAGLIA, *op. cit.*, 104, see also L. AMMANNATI, *I 'signori' nell'era dell'algoritmo,* in *Diritto Pubblico,* 2, 2021, 381 ff.
[35] Consequently, the real value is not in data *per se* but for who owns the tools to deploy it, in this sense, see V. ZENO – ZENCOVICH, *Dati, grandi dati, dati granulari e la nuova epistemologia del giurista,* in *MediaLaws*, 2, 2018, 34. On the *sui generis* database right, and Big Data as falling under its scope of application, see M. BOGNI, A. DEFANT, *Big Data: diritti IP e problemi della privacy,* in *Il Diritto Industriale,* 2, 2015, 119.

data is available to a learning algorithm, the more it can learn"[36] since "ML algorithms enable computers to learn from 'training data', and even improve themselves without being explicitly programmed".[37] However, data quantity is not enough:[38] it needs to be complemented by data quality with the implied request to overcome "concerns as to how clean (error – and gap – free), objective (bias – free), and consistent (few discrepancies) the data are, and as to their veracity and the extent to which they accurately (precision) and faithfully (fidelity, reliability) represent what they are meant to".[39] On the other hand, as regards the legal implications of data, since AI technologies run by ML algorithms and empowered by Big Data are influencing "nearly every aspect of our lives",[40] from market functioning to public – authority decision – making as well as individual rights to privacy and self – determination,[41] assessments, previsions and decisions taken on the basis of poor – quality data can infringe upon fundamental rights,[42] as the next paragraph will try to prove.

3. Technical and Legal implications of Data Quality

Data quality can be assumed as the starting point of the Big Data Analytics process. Therefore, it deserves the utmost attention in order to avoid the data mining process falling into the well – known joint trap of "GIGO" (Garbage in Garbage out), and failing to give relevant information: in this case, the outputs of the systems are commonly considered "biased". [43] In other words, the data quality assessment is a *prius* of data analytics and data mining.

When data analytics is run by Machine Learning algorithms that are fed and trained by Big Data, the "input phase" is composed of different operations: first and foremost, gathering and selection of data from different data sources according to the pursued target (the output); secondly, pre – processing of the selected data in order to remove the "noise", i.e. detecting, cleaning and filtering the unnecessary, inconsistent and incomplete data; lastly, should the data be represented in different formats, the transformation into a suited format aimed at feeding data analysis.[44]

[36] M. EBERS, *Regulating AI and Robotics: Ethical and Legal Challenges,* in M. EBERS, S. NAVAS NAVARRO (eds.), *Algorithms and Law,* Cambridge, 2019, 61.

[37] M. EBERS, *Regulating AI and Robotics: Ethical and Legal Challenges*, cit., 42.

[38] C.W. TSAI, C.F. LAI, H.C. CHAO, A.V. VASILAKOS, *op. cit*., 10, observe that "although it seems that big data makes it possible for us to collect more data to find more useful information, the truth is that more data does not necessarily mean more useful information. It may contain more ambiguous or abnormal data".

[39] R. KITCHIN, *Big Data,* cit., 3.

[40] M. EBERS, *Regulating AI and Robotics: Ethical and Legal Challenges*, cit., 37.

[41] For an in – depth and broad overview of all the mentioned aspects, see R. GIORDANO, A. PANZAROLA, A. POLICE, S. PREZIOSI, M. PROTO (eds.), *Il diritto nell'era digitale – Persona, Mercato, Amministrazione, Giustizia,* Milan, 2022; for a specific focus on of the "algorithmic way of action" of the public administration, see R. CAVALLO PERIN, D.U. GALETTA (eds.), *Il Diritto dell'amministrazione pubblica digitale,* Torino, 2020.

[42] European Union Agency for Fundamental Rights, *Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights*, FRA Focus, Luxembourg, 2019, 1.

[43] As stressed by the Report of the Data Governance Working Group of the Global Partnership of AI, *The Role of Data in AI*, November 2020, 10 – 11, "quantity does not necessarily equal quality» and «quality and quantity of training data is important as any deficiencies present in training data may result in unreliable outcomes, decisions, output data".

[44] C.W. TSAI, C.F. LAI, H.C. CHAO, A.V. VASILAKOS, *op. cit.*, 4.

ISSN 2284 – 4503

While the procedural steps to reach data quality seem clear, its content is less certain, i.e. what is meant by data quality. Indeed, from a technical perspective, data quality remains rather hard to define,[45] not only for its variance according to the different involved disciplines, but also because of its dependency on the addressed purposes;[46] moreover, traditional quality assessments need to be retuned *vis à vis* Big Data's new challenges.[47]

Research on data quality within the "machine learning/artificial intelligence realm[...]has largely been neglected in order to focus more specifically on the learning algorithms and methods themselves" and only in relatively recent times has it gained relevance, starting from the demonstration of "how important data quality is to the outcomes of these algorithms and how severely they are affected by low quality data".[48] This has led to a shift in research "from a model centric to a data – centric approach for building AI systems".[49]

Despite many variances, most of the studies on data quality distinguish a set of essential dimensions (assessed by relevant metrics) that includes: accuracy (proximity to a known reference value), completeness (inclusion of all expected data for the purpose), timeliness (to what extent the data is up – to – date) and consistency (equivalent semantic use).[50] Additionally, in reference to the peculiarities of Big Data, other dimensions have been taken into consideration, such as uniqueness (de – duplication

[45] For the criteria usually adopted in order to assess data quality in AI systems (appropriateness, sample representativeness, adequate sample size, completeness and coherence of sample, low data «noise»), as well as for reference to their structure (unstructured, semi – structured, structured, complex structured data) and formats, see *OECD Framework for the classification of AI systems*, cit., 39 – 40.

[46] As stressed by L. CAI, Y. ZHU, *The challenges of Data Quality and Data Quality Assessment in the Big Data Era,* in *Data Science Journal,* 2, 2015, 4: "academia hasn't made a uniform definition of its data quality and quality criteria. Literature differs on a definition of data quality, but one thing is certain: data quality depends not only on its own features but also on the business environment using the data. Only the data that conform to the relevant uses and meet the requirements can be considered qualified (or good quality) data". Consistently, D. ARDAGNA, C. CAPPIELLO, W. SAMÁ, M. VITALI, *Context – aware data quality assessment for big data*, in *Future Generation Computer Systems*, 89, 2018, 548 ff., in reference to a big data quality assessment that is fit for purpose, the Authors maintain that "dealing with heterogeneous sources requires an adaptive approach able to trigger the suitable quality assessment methods on the basis of the data type and context in which data have to be used". It is worthwhile remembering that the ISO standards on quality (ISO 9000 series) define quality as "fitness for purpose", i.e. the compliance of a product with the users' needs.

[47] As for the new challenges stemming from Big Data, see L. CAI, Y. ZHU, *op. cit.*, 4. To sum up the Authors have found out the following challenges: the diversity of data sources that brings about different data types and data structures (unstructured, semi – structured, structured) and increases the difficulty of data integration; the timeliness of certain data is very short due to the rapid changes in big data; it is difficult to collect, clean, integrate, and finally obtain the necessary high – quality data within a reasonable time frame due to the huge amount of data; data producers do not necessarily coincide with data users, this is a further reason that fuels different perspectives about data quality standards; traditional data quality assessments usually address structured data while Big Data is mainly composed of unstructured data; traditionally, data consumers were either direct or indirect data producers that ensured the quality of data, but in the age of Big Data, data users are not necessarily data producers, thus it is harder for the former to measure data quality.

[48] V. SESSIONS, M. VALTORTA, *The effects of data quality on machine learning algorithms,* in *ICIQ,* 6, 2006, 485.

[49] As recalled by L. BUDACH et al., *The effects of Data Quality on Machine Learning Performance,* in arXiv:2207.14529, Cornell University.

[50] For the distinction between these data quality dimensions, see, *ex plurimis*, D. ARDAGNA, C. CAPPIELLO, W. SAMÁ, M. VITALI, *op. cit.*, 549.

of redundant data)[51] and credibility (data coming from qualified sources).[52] Conversely, two common sources of error are usually identified: representation errors and measurement errors.[53]

From a legal and constitutional perspective too, data quality implies multiple challenges. Automated decision – making, as named by the EU Regulation on Data Protection (GDPR), run by Machine Learning algorithms fed with Big Data is not negative *per se,*[54] but – as is likely with every technological development[55] – its beneficial or maleficent outputs depend on various circumstances, data quality included.[56]

In this respect, it has been stressed that "the use of data to inform decisions[...] [could be] considered a positive development, as it potentially allows for more objective and informed decisions, in comparison to decisions that do not take into account available data. It also has the potential to limit discriminatory treatment based on human decision – making that is derived from existing prejudices. While the limits of data and data analysis need to be taken into account, decisions supported by data are potentially better decisions than those without any empirical support. Therefore, big data also presents opportunities for assessing fundamental rights compliance"[57]

However, Big Data directly challenges some specific data protection principles, such as transparency, data minimisation, purpose limitation.[58] In addition to this individual data protection stance,[59] Big Data

[51] L. BUDACH et al., *op. cit.*, 5.

[52] L. CAI, Y. ZHU, *op. cit.*, 5.

[53] Measurement errors are linked to data incompleteness, inauthentic and unverified data and data noise, while representation errors are due to the sample coverage and its consequent representativeness, indeed "the large size and volume of big data do not necessarily mean that the data is random and representative", for a deeper insight see J. LIU, J. LI, W. LI, J. WU, *op. cit.,* 137 – 138.

[54] According to V. MOLASCHI, *Algoritmi e nuove schiavitù,* in *Federalismi.it,* 18, 2021, 205, data and automated decision processes have subverted the traditional relationship between humans and machines in a twofold direction. On the one hand, machines are no longer mere means deployed in support of human decisions but subjects able to take decisions concerning humans. On the other hand, humans become things, «res» that are computed by machines. Furthermore, doctrine has underscored the possible non – constitutional compliance of automated decision – making, in this sense A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e future delle libertà,* in *BioLaw Journal,* 1, 2019, 63 ff.

[55] According to Kranzberg's first law, technology is neither good nor bad; nor is it neutral, see M. KRANZBERG, *Technology and History: «Kranzberg's Laws»*, in *Technology and Culture*, 27, 3, 1986, 544 ff. For a broad overview of opportunities and risks of IT, including automated processing of data, see G. SARTOR, *Human Rights and Information Technologies,* in R. BROWNSWORD, E. SCOTFORD, K. YEUNG (eds.), *The Oxford Handbook of Law, Regulation and Technology,* Oxford, 2017, 424 ff.

[56] When Big Data comes into play, it is not only a question of quantity but – first and foremost – of improving the qualitative dimension of data, as underlined by O. POLLICINO, *Big Data e Diritto Costituzionale Europeo,* in G. DEMURO, G. COINU, R. MONTALDO (eds.), *Governance dei Big Data e politiche pubbliche,* Naples, 2021, 76.

[57] European Union Agency for Fundamental Rights – Report, *#BigData: Discrimination in Data Supported Decision – Making*, Luxembourg, 2018, 3.

[58] Big data challenges privacy and data protection compliance because of the difficulty to satisfy the duty of *ex – ante* explanation of the purposes of data collection, the principles of purpose limitation and data minimisation (since further data are deducted or inferred – by means of algorithmic systems – from initial input data): in this sense, see M.E. GONÇALVES , *The EU data protection reform and the challenges of big data: remaining uncertainties and ways forward*, in *Information & Communications Technology Law,* 26, 2, 2017, 90 ff.

[59] As observed by K. CRAWFORD, J. SCHULTZ, *Big Data and Due process: Towards a Framework to Redress Predictive Privacy Harms,* in *Boston College Law Review,* 55, 1, 2014, 94, "Big Data has radically expanded the range of data that can be personally identifying"; by means of the algorithmic analysis and cross – referenced combination of

triggers some other "collective" risks,[60] more specifically when data leads to "predictions based on patterns and correlations [...][that] affect numerous aspects of our lives".[61] Beyond the legal concern for the deployment of personal data for economic purposes and the underlying disputed issue about their tradability,[62] if data – whether personal or otherwise – are not accurately collected, structured, labelled and cleaned, they may lead to incorrect assessments and decisions by private or public powers that in turn can infringe upon fundamental rights[63] and the equality principle.[64] Depending on which

different data sets, the prediction of previously undisclosed personally identifiable information (PII), i.e. in the absence of any prior direct collection of this personal data, is now possible. In this respect, also antidiscrimination enforcement mechanisms can be circumvented by «isolating correlative attributes… as a proxy for traits such as race or gender» (100).

[60] According to L. TAYLOR, *What is data justice? The case for connecting digital rights and freedoms globally,* in *Big Data & Society,* 2017, 4, large – scale data collection and its inevitable engagement with the global data market, "raise fundamental questions about… whether individual rights should be the only instrument to combat data harms" since this individual level approach "is rendered problematic by… the fact that many of the negative impacts of data occur on the group as much as the individual level". Similarly, as underscored by A.C. DI LANDRO, *Big data – Rischi e tutele nel trattamento dei dati personali,* Naples, 2020*,* 189 ff., with regard to Big Data deployment by algorithms in ML systems, the risk of subjective harm goes beyond the individual dimension and conversely encompasses a collective dimension (discrimination, human dignity, manipulation of public opinion, mass surveillance).

[61] M.S. GAL, D.L. RUBINFELD*, Data Standardisation,* in *NYU Law Review, 4, 2019,* 738.

[62] Doctrine is divided between who considers personal data exclusively from the perspective of its fundamental rights feature (*ex plurimis*, S. RODOTÀ, *Tecnologie e diritti,* Bologna, 1995), who compares the right of the data subject over his/her data to the right of property (J. LITMAN, *Information Privacy/Information Property*, in *Stanford Law Review,* 52, 5, 2000), and who assumes a «third way» that splits the fundamental right over personal data, envisaged as a right to digital identity, from its object (M. MURSIA, C. TROVATO, *The commodification of our digital identity: limits on monetizing personal data in the European context*, in *MediaLaws*, 2, 2021). For a synthesis of the debate, see A.C. DI LANDRO, *op. cit.*, 151 ff.

[63] As stated by par. 3.5 of the *Proposal for Regulation of the European Parliament and of the Council Laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts* – COM(2021) 206 final, "The use of AI with its specific characteristics (e.g. opacity, complexity, dependency on data, autonomous behaviour) can adversely affect a number of fundamental rights enshrined in the EU Charter of Fundamental Rights", more specifically "the right to human dignity (Article 1), respect for private life and protection of personal data (Articles 7 and 8), non – discrimination (Article 21) and equality between women and men (Article 23)… the rights to freedom of expression (Article 11) and freedom of assembly (Article 12), […] the right to an effective remedy and to a fair trial, the rights of defence and the presumption of innocence (Articles 47 and 48), […] the general principle of good administration […] workers' rights to fair and just working conditions (Article 31), a high level of consumer protection (Article 28), the rights of the child (Article 24) […] the integration of persons with disabilities (Article 26). The right to a high level of environmental protection and the improvement of the quality of the environment (Article 37) […] including in relation to the health and safety of people". For an in – depth and critical assessment of the fundamental rights involved in AI systems (ML methodologies included), see F. DONATI, *Diritti fondamentali e algoritmi nella Proposta di Regolamento sull'intelligenza Artificiale*; A. ADINOLFI, *L'intelligenza artificiale tra rischi di violazione dei diritti fondamentali e sostegno alla loro promozione: considerazioni sulla (difficile) costruzione di un quadro normativo dell'Unione*; A. ODDENINO, *Intelligenza artificiale e tutela dei diritti fondamentali: alcune notazioni critiche sulla recente Proposta di Regolamento della UE, con particolare riferimento all'approccio basato sul rischio e al pericolo di discriminazione algoritmica,* in A. PAJNO, F. DONATI, A. PERRUCCI (eds.), *Intelligenza Artificiale e Diritto: una rivoluzione?,* cit., 111 ff.

[64] As stated by Recital 44 of the *Artificial Intelligence Act* – COM(2021) 206 final –, "High data quality is essential for the performance of many AI systems, especially when techniques involving the training of models are used, with a view to ensure that the high – risk AI system performs as intended and safely and it does not become the source of discrimination prohibited by Union law".

domain is affected by automated predictions or assessments, not only the right to information and the right to self – determination,[65] but also the right to good administration, the right to a fair trial,[66] the right to work, the right to an equal access to services[67] or – broadly speaking – the right to equal treatment and the underlying respect for human dignity could be threatened[68] as well as the democratic principle.[69]

Moreover, this not only occurs when the training data is of poor quality, but also when the data – even if accurate and complete – mirrors existing societal prejudices.[70] In this last respect, "The question [...] is whether we've eliminated human bias or simply camouflaged it with technology".[71] Indeed, as pointed out in the Opinion issued by the European Economic and Social Committee on Artificial Intelligence: "There is a general tendency to believe that data is by definition objective; however, this is a misconception. Data …, may be biased, may reflect cultural, gender and other prejudices and preferences".[72] In these cases, it is not the algorithm *per se* that pursues discriminatory goals, it suffices to let statistical regularities work by reproducing inequalities and discriminations already embedded

[65] Particular attention to «untrue information» and its direct infringement upon human dignity (beyond the infringement of the right to personal data protection) because of its producing an alteration of personal identity, has been paid by the recent opinion issued by the Advocate General G. PITRUZZELLA (C – 460/20 – Advocate General Conclusions, 7th April 2022, 30 – 34) that strikes a balance between the right to information, the freedom of expression, the right to de – referencing.

[66] With regard to the risks implied in predictive justice and policing for fundamental rights, see the *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment*, adopted at the 31st plenary meeting of the CEPEJ, December 2018. As stressed by B. PEREGO, *Predicting policing: trasparenza degli algoritmi, impatto sulla privacy e risvolti discriminatori,* in *BioLaw Journal,* 2, 2020, 452 ff., one of the major concerns of these ML systems falls on the difficulty to collect clean and accurate data.

[67] See the «use cases» in European Union Agency for Fundamental Rights – Report, *Getting the Future – Artificial Intelligence and Fundamental Rights*, Luxembourg, 2020, 25 ff. Moreover, as already stressed by C. O'NEIL, *Weapons of Math Destruction – How Big Data increases inequality and threatens democracy,* New York, 2016, 63 ff., people mainly affected by the risk of discrimination as a consequence of Big Data analytics are the more vulnerable people.

[68] European Union Agency for Fundamental Rights – Report, *Getting the Future Right – Artificial Intelligence and Fundamental Rights*, Luxembourg, 2020, 58 ff.

[69] As for this relation between the individual and collective dimension of the risks stemming from Artificial Intelligence systems and the underlying entanglement among fundamental rights, the democratic principle and the rule of law, see C. SCHEPISI, *Diritti fondamentali, principi democratici e rule of law: quale ruolo e quale responsabilità per gli Stati nella regolazione dell'intelligenza artificiale,* in A. PAJNO, F. DONATI, A. PERRUCCI (eds.), *Intelligenza Artificiale e Diritto: una rivoluzione?,* cit., 209.

[70] The Compas – Loomis case in the U.S. as well as other cases of ML algorithms trained on the basis of biases data, as such delivering predictions or assessments that discriminate a certain group of people (cases of predictive policing, or of detection of social welfare frauds as occurred in the Netherlands, or the Apple Card case, or the Amazon recruiting tool case, or the recommendation system of LinkedIn case), are only some examples of situations where societal existing prejudices feed into ML training systems and deliver amplified discrimination. On the issue, see A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e future delle libertà,* cit.

[71] C. O'NEIL, *op. cit*., 29.

[72] Opinion of the European Economic and Social Committee on Artificial intelligence, *The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society*, in OJ 2017/C 288/01, par. 3.5. In the same way, see also A. ODDENINO, *Intelligenza artificiale e tutela dei diritti fondamentali: alcune notazioni critiche sulla recente Proposta di Regolamento della UE, con particolare riferimento all'approccio basato sul rischio e al pericolo di discriminazione algoritmica,* cit., 193, that consequently considers these biases as inherently affecting data and thus to be qualified as a peril rather than a risk.

within societal structure,[73] bringing about their exponential growth when the AI systems deliver their outcomes.[74]

Under such conditions, the assurance of data quality represents a technical requisite to be monitored within an *ex – ante* risk management approach (when AI systems are still "in the lab")[75] as well as a legal requirement designed to safeguard constitutional rights and principles. Thus, coping with this purpose (data quality) means lifting the veil of Maya of the involved technicalities in order to delve into the underlying fundamental values, principles and rights. Bearing this in mind, our focus is not deemed to deepen the specific dimensions and metrics that underpin data quality, but it rather addresses the regulatory techniques that can support data quality more effectively for the protection of the underlying fundamental rights.[76]

4. The European normative approach to data quality

Regulation (EU) 2016/679 (GDPR) deals with "automated decision – making"[77] and underscores the implied challenges, poor data quality included. More specifically, it tackles the threats involved in the

[73] D. CARDON, *Che cosa sognano gli algoritmi*, Mondadori, Milano, 2016, 71. Discriminatory biases are present in the society and its institutions even in the absence of any conscious animus or discriminatory intent, as shown by studies in evolutionary psychology, in this regard see R. BRADLEY KAR, J. LINDO, *Race and the Law in the Genomic Age – A problem of equal treatment under the law,* in R. BROWNSWORD, E. SCOTFORD, K. YEUNG (eds.), *The Oxford Handbook of Law, Regulation and Technology,* Oxford, 2017, 875. The Authors, firstly observe that "many racial perceptions engage 'folk – biological' module of human psychology" and secondly that "folk biological inferences often occur automatically, unconsciously", furthermore – and as a consequence – today "popular biological misconceptions still affect legal officials' understanding", conclusively "tendencies like these can distort sound moral, legal and factual judgement by regularly causing the disparate treatment of persons" (902).

[74] As recalled by A. ODDENINO, *Intelligenza artificiale e tutela dei diritti fondamentali: alcune notazioni critiche sulla recente Proposta di Regolamento della UE, con particolare riferimento all'approccio basato sul rischio e al pericolo di discriminazione algoritmica,* cit., 175.

[75] Concerning this matter, see the OECD *Framework for the classification of AI systems*, in *OECD Digital Economy Papers,* 323, 2022, 21.

[76] In this sense, the approach might be deemed as a subset of the broader perspective addressed by literature by means of the concept of data justice in the era of datafication and algorithmic processing of data, that in turn underlies choices pertaining to the political and governance domain, see L. TAYLOR, *What is data justice? The case for connecting digital rights and freedoms globally,* in *Big Data & Society,* July – December 2017, 1 ff. Indeed, literature has begun to speak about data justice in a data – driven society in order "to recognise not only how data, its collection and use, increasingly impacts on society, but also that datafication is enabled by particular forms of political and economic organisation that advance a normative vision of how social issues should be understood and resolved. That is, data is both a matter *in* and *of* justice; datafication embodies not only processes and outcomes of (in)justice, but also its own justifications": L. DENCIK, J. SANCHEZ – MONEDERO, *Data justice,* in *Internet Policy Review – Journal on internet regulation,* 11, 1, 2022. As stressed by the 2021's Report of GPAI Working Group on Data Governance, cit., 7, "beyond understanding data governance narrowly as a compliance matter of individualised privacy and ethical design" it is also a question of having "to include considerations of equity and justice specifically as it relates to redressing the uneven distribution of opportunities and harms associated with AI and ML".

[77] As noted by M. BRKAN, G. BONNET, *Legal and Technical Feasibility of the GDPR's quest for Explanation of Algorithmic Decisions: of Black Boxes, White Boxes and Fata Morganas,* in *European Journal of Risk Regulation,* 11, 2019, 23 – 24: "While the GDPR refers to automated decision – making", computer scientists "distinguish between automated, autonomous and algorithmic decision – making and processes […] An automated process is a

deployment of data by means of ML algorithms, states the protective aim pursued (data accuracy and non – discrimination) and places the underlying commitment to safeguard against inaccuracies and discrimination on mutual cooperation between the data controller and the data subject.[78] The GDPR does not, in its provisions but in Recital No. 71, enshrine the duty of the data controller to "use appropriate mathematical or statistical procedures [...] implement technical and organisational measures appropriate to ensure [...] that factors which result in inaccuracies in personal data are corrected and [...] prevent, inter alia, discriminatory effects on natural persons".[79] In addition, the GDPR focuses (Article 5, par. 1 (d)) on data accuracy providing for the right of the data subject to rectification of inaccurate personal data (Article 16), the right to erasure (Article 17) and the right to restriction of processing in case of contested accuracy of the personal data (Article 18, pr. 1(a)), but the effectiveness of these rights is limited *vis à vis* Big Data. But that is not all: "accuracy" is only one of the multiple dimensions assuring data quality,[80] as proved – since their inception in the '90s – by researches on data quality,[81] but it may also not be very simple for the data subject to monitor the movement of its data in order to prevent inaccuracy, since in Big Data analytics the original provided data results in inferred and derived data (classifications, clusters, associations).[82] In this respect, a strong commitment on the part of the data controller is required, specifically, as stressed by the EDPB in its guidelines on automated decision – making and profiling, "Controllers need to introduce robust measures to verify and ensure on an

software or hardware process which executes a predefined sequence of actions without human intervention […] Therefore […] automated decisions are fully predictable [and] fully explainable, as far as the system's specifications and the situation in which the decision was made are known.[…] Differently, autonomous decision – making entails that the algorithmic procedure behind the decision is computed by the agent and relies only on a high – level goal defined by a human. Hence, autonomy emphases the capacity to compute which decisions must be made in order to satisfy a formalised goal, and therefore autonomy is the central notion in the design of artificial agents […] Algorithmic decision – making, for its part, is an overarching notion, encompassing both automated and autonomous decision – making. It means that a given decision is made (partly or completely) with the help of an algorithm; this algorithm may be either automated or autonomous and based or not based on AI techniques […] The analysis above demonstrates that the GDPR, by referring only to automated decision – making, does not take stock of these definitions. From a computer science perspective, it would be more sensible if this legal act deployed the overarching term algorithmic, rather than automated, decision – making".

[78] As for the change of perspective of Regulation (EU) 2016/679 towards an *ex – ante* approach grounded on the principle of accountability charging the data controller with the adoption of techniques by design and by default (Article 25), the data impact assessment (Article 35) and the appointment of a data protection officer (Articles 37 – 39), see L. CALIFANO, *Regolamento UE 2016/679 e la costruzione di un modello uniforme di diritto europeo alla riservatezza e alla protezione dei dati personali* (34 ff.) and S. CALZOLAIO, L. FEROLA, V. FIORILLO, E.A. ROSSI, M. TIMIANI, *La responsabilità e la sicurezza del trattamento* (137 ff.), both in L. CALIFANO, C. COLAPIETRO (eds.), *Innovazione tecnologica e valore della persona – Il diritto alla protezione dei dati personali nel Regolamento UE 2016/679,* Naples, 2017.

[79] As highlighted by E. CELESTE, G. DE GREGORIO, *Digital Humanism: The Constitutional Message of the GDPR,* in *Global Privacy Law Review,* 3, 1, 2022, 9, the GDPR enshrines not only the right to personal data, but it also encompasses – with specific reference to its Article 22 – the broader aim to protect the constitutional value of human dignity.

[80] As pointed out by V. SESSIONS, M. VALTORTA, *op. cit.*, 485, data accuracy "is only a small piece of the overall data quality of our datasets".

[81] R.Y. WANG, D.M. STRONG, *Beyond Accuracy: What Data Quality Means to Data Consumers*, in *Journal of Management Information Systems*, 12, 4, 1996, 5 ff.

[82] R. KITCHIN, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences,* cit., 104.

ongoing basis that data [...] is accurate and up to date. This reinforces the importance of providing clear information about the personal data being processed, so that the data subject can correct any inaccuracies and improve the quality of the data".[83] But a strong commitment is required from the data subject too, since – as stressed by the mentioned guidelines – the rights enshrined by Article 15, 16, 17, 18 of the GDPR are referred not only to the input but also to the output data that maybe re – used in connection with further data sets.[84] In addition, Article 19 sets out the data controller's duty to communicate the requested erasure or rectification and also tracks the boundaries of such a duty stating that it does not apply when the communication is impossible or involves disproportionate efforts. Such circumstances will easily be the case in respect of Big Data movement in today's digital market, moreover in case of secondary use, derived data, aggregated data[85] or when the data are transferred and stored in servers outside the EU borders and consequently control over further users is lost.[86]

Also in line with the accountability principle and the risk – based approach, protection against automated processing of personal data (including profiling)[87] leading to decisions that produce legal effects concerning the natural person or similarly significantly affect him/her (qualified as a high risk activity to personal right and freedoms), is provided by Article 35 of the GDPR. More specifically, it charges the data controller with an impact assessment duty which is the main step within the risk management process,[88] it could be deemed a pre – processing phase in respect of the following phase of automated processing by algorithms and it necessarily involves the evaluation of data quality. Against this backdrop and bearing in mind the abovementioned rights of erasure and rectification held by the data subject, it should make sense, when the impact assessment is carried out, for the data subject's participation to be underpinned by the data controller. This is a procedural step that will result in compliance with the principles of transparency and fairness (Article 5, par. 1) as well as with the duty of the data controller to inform about the "existence of automated decision – making, including profiling"

[83] Guidelines on Automated individual decision – making and Profiling for the purposes of Regulation 2016/679 (*WP251rev.01*), adopted by Article 29 Data Protection Working Party on 3rd October 2017, and endorsed by the EDPB in its first plenary meeting (25th May 2018), 12.

[84] *WP251rev.01,* cit., 17.

[85] As for this subject – matter, see G.M. RICCIO, G. GIANNONE CODIGLIONE, *La rilevanza delle basi giuridiche per il trattamento di dati personali mediante sistemi di intelligenza artificiale*, in A. PAJNO, F. DONATI, A. PERRUCCI (eds.), *Intelligenza Artificiale e Diritto: una rivoluzione?,* cit., 295.

[86] In this respect, it is enough to recall the Schrems (I and II) judgements of the European Court of Justice (C – 362/14 and C – 311/18).

[87] According to Article 4 (4) of Regulation (EU) 2016/679, profiling means "any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements". More specifically, following the definition given by Article 29 WP (*WP251rev.01*, cit.), profiling consists of "a procedure which may involve a series of statistical deductions. It is often used to make predictions about people, using data from various sources to infer something about an individual, based on the qualities of others who appear statistically similar".

[88] Risk assessment differs from risk management since not only the former is a *prius* for the latter, but also because the former is guided by a scientific approach while the latter also includes political evaluation about the measure to be taken in order to tackle the assessed risks, see J. BLACK, *The role of risk in regulatory processes,* in R. BALDWIN, M. CAVE, M. LODGE (eds.), *The Oxford Handbook of Regulation,* Oxford, 2010, 314 ff.

providing "meaningful information about the logic involved" (Article 13, par. 2(f); Article 14, par. 2(g)). In this last respect (the logic involved), complete information should not only cover the explainability issue (in reference to the algorithm's way of functioning),[89] but also the data that has been fed into the process and the underlying data quality assessment (because it is "meaningful information", it is surely useful for the understanding and clarification of the "logic involved" in the whole algorithmic process)[90] and, as such, it would allow the data subject to help with correcting any possible inaccuracies.[91] This conclusion is endorsed by the EDPB guidelines on automated decision – making and data profiling, when it extends the scope of the right to access (Article 15), rectification (Article 16) and erasure (Article 17) not only to the input data but also to the output (derived data).[92]

However, while certain principles of the GDPR (Article 5) recall some dimensions of data quality, such as transparency, accuracy and integrity of data processing, other principles result at odds with the dimensions commonly used to define data quality, *a fortiori* when Big Data comes into question.[93] In this last respect, the "data minimisation" and the "purpose limitation" principles, as stressed by doctrine, do not comply with what is requested for improving the training performance of ML algorithms.[94]

[89] As for the different subsets of the right to explanation according to the moment (*ex – ante* or *ex – post*) and the included information to be given as well as an overview of the existing European legal provisions on explainability (GDPR, consumer protection, platform – to – business regulation, financial regulation), see A. BIBAL, M. LOGNOUL, A. DE STREEL, B. FRÉNAY, *Legal Requirements on Explainability in Machine Learning,* in *Artificial Intelligence and Law,* 29, 2, 2021, 129 ff. These Authors shape a four – level taxonomy for explainability from weakest to strongest obligations: level 1 (main features involved in the algorithmic process); level 2 (all features involved); level 3 (how the features are combined to reach a decision); level 4 (complete knowledge of the model). Moreover, in respect of the right to explanation according to adjudication n. 8472/2019 of the Italian Council of State and the different doctrinal stances on the degrees of comprehensiveness of this right according to the GDPR provisions, see the description set forth by F. LAVIOLA, *Algoritmico, troppo algoritmico: decisioni amministrative automatizzate, protezione dei dati personali e tutela delle libertà dei cittadini alla luce della più recente giurisprudenza amministrativa,* in *BioLaw Journal,* 2, 2020, 5. As for an interdisciplinary perspective that combines philosophical, legal and computer science approaches to the metrics and mechanisms that quantitatively assess the quality of explainability of artificial intelligence testing them in terms of their compliance with the Artificial Intelligence Act, see F. SOVRANO, S. SAPIENZA, M. PALMIRANI, F. VITALI, *Metrics, Explainability and the European AI Act Proposal,* in *Multidisciplinary Scientific Journal,* 5, 2022, 126 ff.

[90] As stressed by E. LONGO, *I processi decisionali automatizzati e il diritto alla spiegazione,* in A. PAJNO, F. DONATI, A. PERRUCCI (eds.), *Intelligenza Artificiale e Diritto: una rivoluzione?,* cit., 354, the mere reference to the logic deployed by algorithms might not be enough for the full comprehension and the consequent compliance with the right to explainability: it should also be complemented by the knowledge of the involved data.

[91] As observed by Article 29 WP (*WP251rev.01*, cit.) "If the data used in an automated decision – making or profiling process is inaccurate, any resultant decision or profile will be flawed".

[92] *WP251rev.01*, cit., 17.

[93] As stressed by R. KITCHIN, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences,* cit., 178, data minimisation is "clearly antithetical to the rationale of big data and the functioning of data markets".

[94] M. EBERS, *Regulating AI and Robotics: Ethical and Legal Challenges*, cit., 24: in reference to the principle of purpose limitation, the Author states that "personal data must be collected for specified, explicit, and legitimate purposes and not further processed in a way incompatible with those purposes. However, analyzing Big Data quite often involves methods and usage patterns which neither the entity collecting the data nor the data subject considered or even imagined at the time of collection. Additionally, when it comes to ML algorithms it may be difficult to define the purpose of processing already at the stage of data collection because it is not possible to predict what the algorithm will learn. To inform the data subjects of the future forms of processing might prove

The completeness of data – i.e. the more detailed the data is the more accurate ML – algorithm predictions and outcomes will be[95] – as well as any secondary uses, for purposes not known at the time of the consent expressed by the data subject or when the information statement was given to him/her (Articles 13 – 14), and any derived (inferred) data are not in line with the data minimisation principle and the duty of communication of all the purposes and personal data involved in processing.[96] Furthermore, the mentioned provisions of the GDPR referable to data quality, leave a loophole open, since "as a rule, ML models[...]contain[...]information about groups and classes of persons" and "although algorithmically designed group profiles may have big impact on a person, (ad hoc) groups are not recognized as holders of privacy rights".[97]

Against this backdrop and in consideration of the peculiarities of Big Data analytics, a further step seems necessary in order to address the underlying issue of data quality, more specifically a step that aims at coping with a more collective perspective and in a more harmonised way throughout the whole EU legal system. Indeed, on the one hand, the GDPR, as seen, adopts an individual and limited perspective, in the interest of the data subject *uti singuli* and, as such, it lacks a collective perspective, typically involved in Big Data analytics methods,[98] on the other hand, the GDPR has a further limited scope since it does not apply in cases of anonymization[99] of data that usually occurs for Big Data.[100] In

costly, difficult, and even impossible". In reference to the principle of data minimization, the Author stresses that "Both Big Data and ML algorithms need a large amount of data to produce useful results. Arguably, the principle of data minimization does not mean that data controllers shall always collect as little data as possible, but only that the quantity must be related to the purpose provided that the data are adequate. Nevertheless, this principle potentially undermines the utility and benefits of Big Data analyses".

[95] S. D'ACQUISTO, M. NALDI, *Big Data e privacy by design. Anonimizzazione. Pseudononimizzazione. Sicurezza,* Torino, 2017, 3, the Authors state that the more attributes are linked to a datum the bigger it becomes, the more descriptors a datum is endowed with the more correlations and associations will be found out. As recalled by M. EBERS, *Regulating AI and Robotics: Ethical and Legal Challenges*, cit., 22, "the amount of data used to train ML algorithms has a greater effect on prediction accuracy than the type of ML method used", and, more specifically, "AI technologies create a strong incentive to collect and store as much additional data as possible in order to gain meaningful new insights".

[96] In respect of Big Data, as stressed by R. KITCHIN, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences,* cit., 100, "much of big data is generated with no specific question in mind or is a by – product of another activity".

[97] M. EBERS, *Regulating AI and Robotics: Ethical and Legal Challenges*, cit., 24.

[98] G. DE MINICO, *op. cit.*, 89 ff. that points out some criticalities of the GDPR provisions in respect of Big Data, first and foremost, also in the case of compliance with the Regulation because of the processing of essentially anonymous data, big data deployment by algorithmic systems is able to produce discriminations, thus the Author has accordingly qualified this activity as a "dangerous activity" with relevant consequence in terms of strict liability.

[99] On the privacy enhancing technologies designed to anonymise personal data see S. D'ACQUISTO, M. NALDI, *op. cit.,* 41 ff.

[100] From a computer science perspective, M. BRKAN, G. BONNET, *op. cit.*, 19, ML algorithms deploy mixed datasets of personal and non – personal data and the separation is often not easy. Indeed, as observed by L. AMMANNATI, *I 'signori' nell'era dell'algoritmo,* cit., 398, the distinction between personal and non – personal data will quickly become obsolete in a world where data are gathered by IoT and M2M. Conversely, as denounced by A. MACERATINI, *New Technologies, Big Data and Human Rights: An Overview,* in A. CALIGURI (eds.), *Legal technology transformation – A practical Assessment,* Naples, 2020, 11 ff., even if Big Data tends to work on anonymous data, one cannot exclude the possibility that after appropriate correlations it becomes referable to very specific people. In the same way, S. CALZOLAIO, *Protezione dei dati personali,* in *Digesto delle Discipline Pubblicistiche* –

addition, the accountability principle and the underlying risk – based approach adopted by the GDPR could lead to a heterogeneous solution among data controllers and across Member States, but, due to the level of the interests involved in data quality and the consequent risks for fundamental rights and freedoms,[101] a more "convergent" approach is requested.

More specifically, in the age of "datafication" where Big Data feeds into ML algorithms that deliver assessments and decisions that are capable of infringing upon a huge set of fundamental rights, public authorities should undertake a major commitment for the assurance of data quality. Thus, if the GDPR, owing to the period of adoption, undertook the aforementioned individual perspective,[102] a different approach should be expected in more recent proposals, with specific reference to those on AI and data governance.

The Proposal of Regulation of the European Commission "Laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts",[103] expressly complements the GDPR provisions with specific reference to algorithmic discrimination stemming from training data.[104] In order to underpin data quality, when data is deployed to train algorithmic models for AI high – risk systems, the AI proposal follows a twofold path: on the one hand, it strengthens the scope of the transparency principle already enshrined by the GDPR; on the other hand, it integrates the risk assessment and management provisions already adopted by the GDPR with express reference to a data governance system. In the first respect, the principle of transparency and the connected rights to information and explainability is expressly extended – for high – risk AI systems – in order to include the specifications of the input data, or any other relevant information in terms of the training, validation and testing data sets used.[105] Secondly, in compliance with the "para – constitutional"

Aggiornamento, cit., 606, stresses that, in a data – driven reality, it is the possible personal identification stemming from the data deployment, rather than the possible identification at the time of data collection, that gains relevance.

[101] As stressed by G. MOBILIO, *L'Intelligenza Artificiale e I rischi di una «disruption» della regolamentazione giuridica,* in *BioLaw Journal,* 2, 2020, 294, the GDPR fails to offer an adequate protection against discrimination stemming from algorithmic decisions within ML systems. Taking a similar approach, see C. NAPOLI, *Algoritmi, Intelligenza artificiale e formazione della volontà pubblica: la decisione amministrativa e quella giudiziaria,* in *Rivista AIC,* 3, 2020, 329. To sum up, as stated by M. EBERS, *Regulating AI and Robotics: Ethical and Legal Challenges*, cit., 25 "all these considerations show how little the new GDPR is compatible with big data analysis and AI products".

[102] Doctrine calls for a collective dimension of privacy and data protection in the era of Big Data Analytics, see O. POLLICINO, *Big Data e Diritto Costituzionale Europeo,* in G. DEMURO, G. COINU, R. MONTALDO (eds.), *op. cit*., 75, that speaks about a "third dimension" of data protection; M.F. DE TULLIO, *La privacy e I big data verso una dimensione costituzionale collettiva,* in *Politica del Diritto,* 4, 2016, 641, that speaks about a "collective dimension of the right", to be entitled to a whole category of data subjects that has undergone predictions and decisions based on Big Data Analytics.

[103] *Artificial Intelligence Act* – COM(2021) 206 final, 21st April 2021.

[104] *Artificial Intelligence Act* – COM(2021) 206 final –, par. 1.2., states that the proposal complements "the […] existing Union law on non – discrimination with specific requirements that aim to minimise the risk of algorithmic discrimination, in particular in relation to the design and the quality of data sets used for the development of AI systems". As confirmed by the *European Declaration on Digital Rights and Principles for the Digital Decade* – COM(2022) 28 final, Chapter III, algorithmic systems need to be "based on suitable datasets to avoid unlawful discrimination".

[105] Article 13, par. 3, b), (v), *Artificial Intelligence Act* – COM(2021) 206 final. As maintained by B. PEREGO, *op. cit.*, 460, transparency is surely useful in order to protect against discrimination, but it is not enough, indeed the

perspective of the proposal,[106] after stating that high quality data is "strictly necessary to mitigate the risks to fundamental rights and safety posed by AI"[107], the proposal provides for the implementation of "appropriate data governance and management practices"[108] with due account to the addressed purposes and the underlying appropriate statistical properties. On this basis, Article 10 of the proposal sets out quality criteria for training, validation and testing data sets,[109] and lays down some procedural steps to be undertaken in order to improve data quality such as collection, annotation, labelling, cleaning, enrichment and aggregation, as well as the formulation of relevant assumptions, notably with respect to the information that data is supposed to measure and represent. Moreover, a prior assessment of the availability, quantity, suitability and possible biases of the data sets, in addition to the identification and tackling of any possible data gaps or shortcomings, are requested.[110]

Not only does the proposal concerning AI deal with data quality, the recent Data Act proposal[111] also delves into the matter. This proposal, according to the addresses laid down by the European Data Strategy, aims at fostering free movement of data[112] between the manufacturer of a product or a supplier of a service and the user that generates data, or between the data holder (different from the manufacturer or supplier of services) and the data recipient, but it also applies to data requested by the public sector institutions or bodies for exceptional public interest needs (Article 1). With specific reference to data quality, it requires the data holder to make data available with the same quality as

correction of dirty data is all the more necessary. Moreover, doctrine has stressed the "paradox" of the transparency principle because of the involved economic interest of the "holder" of the algorithmic model, see C. COLAPIETRO, *Circolazione dei dati, automatizzazione e regolazione,* in *osservatoriosullefonti.it,* 2, 2021, 837 – 838. Against this background, in a Big Data system, even in case of explicit consent of the data subject, this latter often lacks freedom and self – determination, as stressed by A.C. DI LANDRO, *op. cit.*, 170 ff.

[106] A. PAJNO, *Introduzione allo studio della proposta della Commissione europea di Regolamento sull'intelligenza artificiale,* in *Astridonline.it.*

[107] *Artificial Intelligence Act* – COM(2021) 206 final – par. 2.3.

[108] *Artificial Intelligence Act* – COM(2021) 206 final – Recital 44.

[109] *Artificial Intelligence Act* – COM(2021) 206 final – Article 10, par. 3 – 5 listed the following dimensions and properties of data, firstly (par. 3), they shall be "relevant, representative, free of errors and complete. They shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons on which the high – risk AI system is intended to be used". Moreover (par. 4) they are requested to be fit for purpose: they "shall take into account, to the extent required by the intended purpose, the characteristics or elements that are particular to the specific geographical, behavioural or functional setting within which the high – risk AI system is intended to be used"; lastly (par. 5), in order to correct data bias, the processing of special categories of personal data is allowed provided that safeguards for fundamental rights and freedoms are adopted: "To the extent that it is strictly necessary for the purposes of ensuring bias monitoring, detection and correction in relation to the high – risk AI systems, the providers of such systems may process special categories of personal data referred to in Article 9(1) of Regulation (EU) 2016/679, Article 10 of Directive (EU) 2016/680 and Article 10(1) of Regulation (EU) 2018/1725, subject to appropriate safeguards for the fundamental rights and freedoms of natural persons, including technical limitations on the re – use and use of state – of – the – art security and privacy – preserving measures, such as pseudonymisation, or encryption where anonymisation may significantly affect the purpose pursued".

[110] *Artificial Intelligence Act* – COM(2021) 206 final –, Article 10, par. 2.

[111] *Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)* – COM(2022) 68 final.

[112] As stressed by L. AMMANNATI, *I 'signori' nell'era dell'algoritmo*, cit., 381, free movement of data has become the "fifth fundamental freedom" in the European single market.

those in his possession, protecting against manipulation or alteration of data.[113] This proposal complements Article 10 of the Artificial Intelligence proposal and rather than focusing on the "inherent" features of data quality (already dealt with by the latter), deals with the "system – dependent features"[114] aimed at assuring the interoperability of data and data sharing.[115]
Both these proposals surely deserve consideration due to their "collective" approach to data quality and their efforts (first and foremost the proposal on AI) to "grasp" some of its technical dimensions, but the vagueness and flexibility used to pursue the task, risk being reduced to mere "ethics bluewashing".[116] Indeed, in order to address the data quality issue on a harmonised basis, both these proposals of Regulation (AI Act and Data Act) call on relevant standards (Article 40, AI Act and Article 28, par. 4, Data Act), but this reference to standards opens further legal challenges that will be dealt with in the following paragraphs.

5. Data quality standards

Well before the two abovementioned proposals, different EU bodies called for standards aimed at improving data quality in order to achieve "high – quality data sets for developing and training AI systems".[117] More specifically, data quality has been targeted as "a first ethical constraint" and "a responsible practice for the very first step in the lifecycle of AI systems"[118]; as such a "data hygiene certificate" has been envisaged, in order to "ensure the quality of the data being used to train the algorithm, where quality is measured according to its sourcing, acquisition, diversity, and labelling".[119] Furthermore, the Council of the EU addressed the issue of "quality standards of data sets" envisioning "quality seals" in order to offer adequate guarantees to this end.[120]
Scholars too, have stressed the importance of standardizing the features of data quality from its very first collection in order to avoid errors, minimize further management costs[121] and reduce the risk of infringement of fundamental rights.[122] Some scholars have underscored this need, referring to a certification system[123] or an independent audit system in order to assess which data is collected, and how

[113] *Data Act* – COM(2022) 68 final, Articles 4 – 5.
[114] According to the distinction between inherent and system – dependent data quality characteristics set out by the ISO/IEC 25012 standard.
[115] *Data Act* – COM(2022) 68 final, Article 28, par. 1, (a).
[116] Borrowing the wording by L. FLORIDI, *Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical,* in *Philosophy & Technology,* 32, 2019, 187.
[117] Opinion of the European Economic and Social Committee on *Artificial Intelligence – The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society*, 26th Plenary session 31 May – 1 June 2017, par. 1.9.
[118] *Artificial Intelligence: from ethics to policy* – European Parliamentary Research Service, PE 641.507, June 2020, 27.
[119] *Ibidem*, II.
[120] *Berlin Declaration on Digital Society and Value – Based Digital Government*, 8th December 2020.
[121] M.C. CARROZZA, C. ODDO, S. ORVIETO, A. DI MININ, G. MONTEMAGNI, *AI: profili tecnologici. Automazione e Autonomia: dalla definizione alle possibili applicazioni dell'intelligenza artificiale,* in *BioLaw Journal,* 3, 2019, 237.
[122] A.C. AMATO MANGIAMELI, *Algoritmi e big data. Dalla carta sulla robotica,* in *Rivista di filosofia del diritto,* 1, 2019, 107 ff., stresses the need to set out rules for the collection, classification, analysis, and synthesis of data.
[123] F. LAVIOLA, *op. cit.,* 50.

data is tracked and fed into the algorithmic model.[124] Finally a direct and explicit call for technical standards to assess and manage data quality has been raised.[125]

In this respect, it is worth remembering that to some extent data quality is already dealt with by the European regulatory framework developed by the European Statistical System (ESS),[126] but not with specific regard to the measurement of the quality of the Big Data deployed to train and pilot ML systems for which a common European approach continues to be lacking.[127]

At the international level, Standards Development Organisations (SDOs) have established a data quality model for structured data recorded into ITs (ISO/IEC 25012) consisting of fifteen metrics to be considered[128] but it does not specifically address Big Data that feeds into ML algorithms (that is mainly unstructured data).[129] Recently, a standard was developed with specific regard to AI systems, including data biases, namely the ISO/IEC TR 24027:2021 (Information technology — Artificial intelligence (AI) — Bias in AI systems and AI – aided decision – making). This standard, among other goals, aims to tackle the biases "inherent in the datasets used to train the system". [130] In addition, with reference to the specific issue of data quality for analytics and machine learning algorithms, a further standard is currently under development by the ISO Technical Committee on Artificial Intelligence.[131]

Beyond this international intervention on data quality standards, a European focus on "Data quality requirements for inclusive, non – biased and trustworthy AI"[132] is requested, since not only "to date there is still not a common agreed methodology" despite the efforts undertaken by "researchers in academia and industry[...]to evaluate and mitigate bias present in the different AI components",[133] but also because – as pointed out by a CEN – CENELEC Report – "international standards being developed might not take into account sufficiently or protect adequately the European values, principles, or

[124] B. PEREGO, *op. cit.*, 459 – 460.

[125] C. COLAPIETRO, *Circolazione dei dati, automatizzazione e regolazione,* in *Osservatoriosullefonti.it,* 2, 2021, 840.

[126] See Regulation (CE) 223/2009 and the relevant Code of Practice (adopted on the basis of Article 2, par. 1 and Article 11 of Regulation (CE) 223/2009.

[127] As proven by the Workshop on 8th June 2022 organised by CEN and CENELEC, together with the European Commission's Joint Research Centre (JRC), focusing on the topic *Data quality requirements for inclusive, non – biased, and trustworthy artificial intelligence*.

[128] Namely, "accuracy, completeness, consistency, credibility, currentness, accessibility, compliance, confidentiality, efficiency, precision, traceability, understandability, availability, portability, recoverability". Most of these requirements pertain to inherent features of data, others (such as availability, portability and recoverability) pertain to the IT system, see D. NATALE, *La Qualità dei Dati e la ISO/IEC 25012,* in *U&C*, 2, 2009, 19 – 20.

[129] M. TALHA et al., *Big data: Trade – off between Data Quality and Data Security,* in *Procedia Computer Science,* 151, 2019, 918.

[130] See https://www.iso.org/obp/ui/#iso:std:iso – iec:tr:24027:ed – 1:v1:en (last access 25/09/2022).

[131] The ISO/IEC Joint Technical Committee subcommittee 42 (JTC 1/SC 42) on Artificial Intelligence was established by the International Standardisation Organisations in 2017, see the relevant website https://www.iso.org/committee/6794475.html (last access 25/09/2022).

[132] This is indeed the title of the 2022's workshop held by the European Commission's Joint Research Centre (JRC) and the European Standardisation Organization (specifically CEN and CENELEC) that aimed "at identifying emerging science and technology areas that could benefit from standardisation activities" bringing together regulators and the standardisation community as well as businesses and scientific experts.

[133] Workshop on Data Quality requirements for inclusive, non – biased and trustworthy AI, 8th – 9th June 2022, in https://www.cencenelec.eu/media/CEN – CENELEC/Events/Events/2022/20220608_PSIS/psis_ai_brochure.pdf, (last access 25/09/2022).

specificities, thus requiring specific regional developments in Europe",[134] as indeed proved by the US dominance within the ISO Technical Committee on AI.[135] Consequently, since it is of paramount importance "to alleviate the root causes of AI biases [...] proving that data on which models are trained fulfil quality standards",[136] CEN – CENELEC, according to Article 8 of Regulation 1025/2012, called on the European Commission to introduce in its Annual Union Work Programme (AUWP) the development of harmonized standards in support of the Artificial Intelligence Act.[137] Moreover, in 2020, CEN and CENELEC established their own Joint Technical Committee CEN – CENELEC/JTC 21 "Artificial Intelligence" in order to produce standardization deliverables in this field against the belief that "it is crucial to create the most relevant architecture of AI standards to support Europe's needs and to ensure a smooth alignment between the European and international standardization frameworks".[138]
However, this call for data quality standards,[139] supported, as mentioned in par. 4, by the AI proposal and the Data Act proposal, is not neutral from a legal perspective. Not only because (as dealt with in the previous paragraphs) choices and assessments on quality features of the data that feeds into ML algorithms may result in outcomes that could infringe upon fundamental rights and principles, but – first and foremost – it is the delivery of this task to the standardisation process itself that underlies a debated balance between conflicting rationales, as the next paragraph will try to explain.

[134] CEN – CENELEC Focus Group Report: Road Map on Artificial Intelligence (AI), cit., 7.
[135] M. EBERS, *Standardizing AI – The Case of the European Commission's Proposal for an Artificial Intelligence Act*, in L. DI MATTEO, C. PONCIBÒ, M. CANNARSA (eds.) *The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics,* Cambridge, 2022, 326.
[136] Workshop on Data Quality requirements for inclusive, non – biased and trustworthy AI, cit.
[137] See CEN – CENELEC Position Paper *Proposal for a Regulation laying down harmonised rules on Artificial Intelligence – Artificial Intelligence Act* – COM(2021) 206, October 2021, 2.
[138] *Ibidem*, 3.
[139] For an overview of the current standardisation activities (already delivered standards or under – development standards) in the field of AI, both at international and European level, see M. EBERS, *Standardizing AI – The Case of the European Commission's Proposal for an Artificial Intelligence Act,* cit., 330.

6. Legal implications of standardisation

Reference to private technical standards[140] represents a regulatory choice[141] consistent with the EU broader approach to governance and better regulation[142] and a way to address – borrowing a recent scholarship assumption – contemporary inter – legality issues typical of the global legal space.[143] Thus, reference to technical standards mirrors the broader hybridisation process between public and private spheres,[144] due to the complexities of a globalised world and the consequent crisis of the traditional

[140] Research on regulation has underscored the "variety of ways in which standards may be expressed", more specifically "the term regulatory standards is often deployed in a narrow sense as referring to the standards developed by technical standardisation bodies such as the International Organisation for Standardisation (ISO) and its sectoral, regional and national equivalents. A broader conception of standards defines them as instruments which encourage the pursuit or achievement of a value, a goal or an outcome without specifying the action(s) required to achieve this, in contrast with a legal rule, which is prescriptive as to what its subject must or must not do[…] Accordingly technical standards are an important sub – set of the larger group of regulatory standards", C. SCOTT, *Standard – setting in Regulatory Regimes,* in R. BALDWIN, M. CAVE, M. LODGE (eds.), *The Oxford Handbook of Regulation,* Oxford, 2020, 105. Within the EU, it is worthwhile recalling the Resolution (85/C 136/01) on a New Approach to Technical Harmonisation and Standards, adopted by the Council on 7th May 1985. According to it, the new legislative approach rests on the following principles: – legislative harmonisation is limited to essential safety requirements (or other requirements in the general interest); – the task of drawing up technical production specifications is entrusted to the European Standardisation Organisations (ESOs); – these technical specifications are not mandatory and maintain their status of voluntary standards; – but the authorities are obliged to recognise that products manufactured in conformity with harmonised standards are presumed to conform to the essential requirements established by EU legislation.

[141] As maintained by A. ZEI, *Shifting the boundaries or breaking the branches? On some problems arising with the regulation of technology,* in E. PALMERINI, E. STRADELLA (eds.), *op. cit.*, 179, standardisation of technologies "play[s] a major role within any market regulation policy".

[142] In this respect see L. SENDEN, *Towards a More Holistic Legitimacy Approach to Technical Standardisation in the EU,* in M. ELIANTONIO, C. CAUFFMAN (eds.), *The Legitimacy of Standardisation as a Regulatory Technique in the EU – A Cross – disciplinary and Multi – level Analysis,* Cheltenham – Northampton, 2020, 20. At the European level, this "New approach" started in the mid – 1980s with reference to product safety giving rise to a sort of division of border between European rule – making and technical standardisation as occurred in the German legal system. Moreover, as stated by the ECJ in the Elliott case (C – 613/14, parr. 34 – 40), harmonised technical standards adopted by ESOs under mandate given by the EC, monitored and managed by it and lastly published in the C series of the Official Journal (as currently provided by Article 10 of Regulation (EU) No. 1025/2012), as such giving rise to the presumption of conformity by means of compliance with the standards, are considered "part of the EU law" because of "their nature [of] measures implementing or applying an act of EU law" and notwithstanding their voluntary and not binding effect. As for technical standards and their relationship with the traditional constitutional features of the modern State in terms of national sovereignty, democratic legitimation and the relevant legal system, see A. IANNUZZI, *Il diritto capovolto – Regolazione a contenuto tecnico – scientifico e costituzione,* Naples, 2018, 65 ff. Standard – setting is one of the manifestations of those private powers that have been deemed to challenge State sovereignty giving rise to a sort of "counter – sovereign" that lacks the democratic foundation featuring the nation state: M. LUCIANI, *L'Antisovrano e la crisi delle costituzioni,* in *Rivista di diritto costituzionale,* 1, 1996, 124 ff.

[143] E. CHITI, A. DI MARTINO, G. PALOMBELLA, *Nel mondo delle legalità al plurale e dell'interconnessione,* in E. CHITI, A. DI MARTINO, G. PALOMBELLA (eds.), *L'era dell'interlegalità,* Bologna, 2022, 10 ff.

[144] With regard to the increasing spread of legal phenomena that cut through traditional conceptual frameworks giving rise to forms of "multiple hybridization" such as transnational legal systems, see K. TUORI, *On Legal Hybrids,* in *EUI Working Papers – A self – sufficient European private law – A viable concept?,* 31, 2012, 67 ff. Regulatory techniques such as self – regulation and co – regulation too, as well as – more recently – the so called technological management, carry out other forms of "hybridization". As for the former, the EU definition regarding the

categories implied by the modern sovereign State,[145] as well as the underlying attempt to reach transnational "convergence"[146] by means of new governance patterns.[147]
Technical standards are flexible, "undated reference[s] [that] [do] not generate any obligation[...]they are not mandatory" but they may generate legal effects as, for instance, the reversal of the burden of proof.[148] They are issued at international level by ISO, IEC and ITU, while at the European level they are issued by European Standard Organizations – ESOs (CEN, CENELEC, ETSI).[149] These are organisations falling within private law, composed of experts and representatives of the involved sectors but open to multi – stakeholder participation.[150]

---

difference between self – regulation and co – regulation has been laid down in the Inter – institutional Agreement on better law – making (2003/C 321/01), but doctrine has highlighted the fact that the difficulty to define such a form of regulation «stems in part from the lack of a generally accepted framework for categorising any regulatory instrument»: C. COGLIANESE, E. MENDELSON, *Meta – Regulation and Self – Regulation,* in R. BALDWIN, M. CAVE, M. LODGE (eds.), *op. cit.*, 148. According to these Authors, in self – regulation, the regulator and its target (regulated bodies) coincide; while in meta – regulation, there is an outside regulator whose task is to "seek to induce targets" (i.e., regulated entities) "to develop their own internal, self – regulatory responses to public problems" (150). As for the latter (technological management), it represents the possibility to embed the rule in the architecture of the technology itself (deploying Lessig's stance in its *Code is Law)* as underscored by R. BROWNSWORD, *Law, Technology and Society – Re – imagining the regulatory environment,* New York, 2019, 160 ff., with specific regard to its implications in terms of personal liberty.

[145] More specifically, the traditional categories of the modern State infringed by globalisation process are: territorial boundaries, separation of powers, hierarchical rules adopted by the political power through legislation, in these terms see, M.R. FERRARESE, *Globalizzazione giuridica,* in *Enciclopedia del Diritto,* Annali IV, 2011, 547 ff.

[146] On the concept of convergence see R. BROWNSWORD, *Convergence: What, Why and Why Not,* in *EUI Working Papers – A self – sufficient European private law – A viable concept?,* 31, 2012, 77 ff. With specific regard to convergence as "one of the underlying values on which the standardisation process is based" and for the underlying distinction between convergence through rules harmonisation at the European level and convergence through standardisation, see B. VAN LEEUWEN, *European Standardisation in Healthcare: towards convergence through self – regulation,* in *EUI Working Papers – A self – sufficient European private law – A viable concept?,* 31, 2012, 141 ff.

[147] M.R. FERRRARESE, *La governance tra politica e diritto*, Bologna, 2008, 52 and 190. More specifically, new modes of governance encompass self – regulation and co – regulation that limit State intervention; as such a subset of this process is represented by the 1985's New Approach on Technical Standards and Regulation "combining mandatory framework regulation with voluntary rule making in standard bodies": H.W. MICKLITZ, *A self – sufficient European Private Law – A Viable concept?* in *EUI Working Papers – A self – sufficient European private law – A viable concept?,* 31, 2012, 24. Broadly speaking, as stressed by S. RODOTÀ, *Technology and regulation: a two – way discourse,* in E. PALMERINI, E. STRADELLA (eds.), *Law and Technology – The Challenge of Regulating Technological Development*, Pisa, 2013, 30, this process implies "a juridification with a low formal intensity and a high effectiveness impact".

[148] A. ZEI, *Shifting the boundaries or breaking the branches? On some problems arising with the regulation of technology,* in E. PALMERINI, E. STRADELLA (eds.), *op. cit.*, 182.

[149] For an in – depth description of the standardisation bodies at international, European and national level, as well as the difference between legal rules and technical standards (their being a consensual, voluntary expression of private autonomy and self – regulation by the regulated entities in respect of the imperative, hierarchical, binding nature of legal rules), see A. IANNUZZI, *op. cit.*, 29 ff.

[150] F. CAFAGGI, *New foundations of transnational private regulation,* in E. PALMERINI, E. STRADELLA (eds.), *op. cit.*, 94 and 97. It is also worth recalling that part of the doctrine prefers to qualify these transnational phenomena as "non – law", see R. TARCHI, *Diritto transnazionale o diritti transnazionali? Il carattere enigmatico di una categoria giuridica debole ancora alla ricerca di un proprio statuto,* in *osservatorisullefonti.it,* 1, 2021, 16.

However, the legislator's reliance on standard – setters involves a huge and controversial debate about the legitimacy of this "alternative" mode of regulation.[151] Indeed, this process is mainly led by private actors and transnational enterprises, also because – moreover in the new technologies domain – "they are normally the only ones to hold the know – how, necessary to the formulation of the standards themselves".[152] Consequently, it is an interest – driven process with no openness to public debate.[153] Comments on the draft standards published on the website of the standardisation body are possible but the decision – making within the technical committee of National or European Standard Organisations is confidential and not made publicly available and the final standard is delivered against payment, since it is protected by copyright.[154]

Public interference in the organisation of these private bodies, their way of action, their funding and the legal status recognized to their outcomes is provided by the EU legislation on standardisation in order to bolster transparency and the legitimacy of the process (Regulation (EU) No. 1025/2012); as such ESOs have been defined a "public – private partnership[...]entrusted with public tasks".[155] However, beyond the ESOs composition with representatives from the national standardisation bodies, experts and representatives from the relevant business sector, participation by the Annex III organisations (representing SMEs, consumers, environmental interests, social interests) in the activities of ESOs does not involve having any voting rights. Moreover, it is impossible for an ordinary citizen to have information on when the Technical Committees of ESOs meet, who participates in them, which discussions take place, and which positions the participants have held.[156] In this respect, representativeness

[151] "A topical, yet still open question surrounding its use [is]: the legitimacy of standardisation as a regulatory technique in the European Union", as pointed out by M. ELIANTONIO, C. CAUFFMAN, *The Legitimacy of Standardisation as a Regulatory Technique in the EU – A Cross – disciplinary and Multi – level Analysis: An Introduction,* in M. ELIANTONIO, C. CAUFFMAN (eds.), *op. cit.*, 4, that also underscores the recent public law scholars' attention towards the phenomenon of standardisation, while it was previously studied from an economic and political science perspective.

[152] E. FOSCH VILLARONGA, A. GOLIA, *Robot, Standards and the Law,* in *Computer Law&Security Review,* 35, 2, 2019, 132. A further rationale supporting private standards is the effectiveness of the public – private partnership in addressing social objectives or environmental sustainability: M. MATAIJA, *Leveraging Trade Law for Governance Reform: The Impact of the WTO Agreement on Technical Barriers to Trade on Private Standard – Setting,* in *European Review of Private Law,* 220, 19, 293. Similarly, A. ZEI, *Shifting the boundaries or breaking the branches? On some problems arising with the regulation of technology,* in E. PALMERINI, E. STRADELLA (eds.), *op. cit.*, 174, observes that "the organizational resources and expertise necessary to address the challenges related to new technologies can be more easily found in the sphere of private autonomy. Due to that, legal provisions often merely set out general terms" leaving space to "generally speaking, private standards".

[153] Indeed, "frequently the capture by industry dilutes [the] neutrality and objectivity" of technical standards: F. CAFAGGI, *op. cit.*, 94. In the same way: "Such private bodies are not sufficiently accountable, representative, and transparent, and their procedures may be biased towards certain interests at the expense of others. Even though legal systems commonly rely on the output of standards bodies, they lack mechanisms of influencing or even scrutinizing the way those bodies act and the way they are organized. These problems may be exacerbated with transnational standards bodies which are not easily 'caught' by national legal systems": M. MATAIJA, *op. cit.*, 294

[154] B. VAN LEEUWEN, *European Standardisation in Healthcare: towards convergence through self – regulation,* in *EUI Working Papers – A self – sufficient European private law – A viable concept?,* 31, 2012, 155 – 156.

[155] M. ELIANTONIO, *Private Actors, Public Authorities and the Relevance of Public Law in the Process of European Standardisation,* in *European Public Law,* 24, 3, 2018, 477 – 478.

[156] As clearly stressed by M. ELIANTONIO, *Private Actors, Public Authorities and the Relevance of Public Law in the Process of European Standardisation,* cit., 481.

and decision – making are lacking adequate participation and transparency, thus essential features of public law relevance are challenged when rule – making power is delegated to standard – setters.[157] This is moreover the case when, beyond the apparent technical scope and neutral feature, standards underlie political choices[158] and the balancing of assessments referring to fundamental rights and values,[159] that are the traditional domain of competence of public institutions (directly or indirectly) endowed with democratic legitimacy.[160]

In these circumstances, reference to standards made by the EU proposal on Artificial Intelligence (as well as the Data Act proposal), in particular when data quality issues are at stake, surely calls into question the mentioned legal concerns. However, the "clause" that the EU has enshrined in its proposals deserves consideration, since it reveals a sort of recovery of awareness about the non – neutral nature of standards and, as such, it lays down a major steering and monitoring role by public power when fundamental rights and principles come into play.

7. Conflicting rationales: is the EU "climbing back to the top?"

The economic rationale that has underpinned the development of the European Union is well known[161] as well as the related progressive emerging of fundamental rights.[162] This European path mirrors the

[157] The regulatory relationship between the EC and the relevant standardisation bodies, when the former entrusts the latter with the task of elaborating technical standards, has been differently qualified by doctrine as described by A. ZEI, *Shifting the boundaries or breaking the branches? On some problems arising with the regulation of technology,* cit., 197 and 202 (hidden attribution of regulatory power to private subjects, material delegation, *de facto* delegation, dissembled allocation of regulative responsibilities, *munus publicum*, concession).

[158] As for the non – neutral nature of technical standards and their underlying political rationale, see M.R. FERRARESE, *Privatizzazioni, poteri invisibili e infrastrutture giuridiche globali,* in *Diritto Pubblico,* 3, 2021, 888.

[159] E. STRADELLA, *La regolazione della Robotica e dell'Intelligenza artificiale: il dibattito, le proposte, le prospettive. Alcuni spunti di riflessione,* in *MediaLaws,* 1, 2019, 80. Similarly, L. AMMANNATI, *Per una Intelligenza Artificiale affidabile. Presupposti e sviluppi della sua regolazione,* in *AstridOnline – Rassegna,* 5, 2022, points out that the standardisation of AI systems is not merely a technical issue to be delivered to private organisations, since the underlying ethical and legal aspects call for a political and democratic debate.

[160] In this sense, L. SENDEN, *op. cit.,* 26, recalls: "given that traditional setting of rules is a power lying with government, legitimised by a democratic decision – making process, it can be convincingly argued that the 'outsourcing' of any rule – making and/or enforcement actions to private players should be subject to similar requirements and guarantees".

[161] M. LUCIANI, *La Costituzione italiana e gli ostacoli all'integrazione europea*, in *Politica del diritto*, 4, 1992, 579, the Author raises doubts about the compatibility of the Maastricht Treaty with fundamental constitutional values, since it enshrines economic growth as a goal, and not as a means that serves social welfare, with the implied risk to an equal balance between economic and social needs. The Ordoliberal approach of the EU has, as a result, been stressed by doctrine, see *ex plurimis*, C. JOERGES, *La Constitution économique européenne en processus et en procès,* in *Revue Internationale de Droit Économique,* 20, 3, 2006, 245 ff. Moreover, after the 2008's economic crisis, doctrine has underscored the "displacement of social Europe" in respect of its financial, economic and market imperatives, see. C. KILPATRICK, *The displacement of Social Europe: a productive lens of inquiry,* in *European Constitutional Law Review*, 14, 2018, 62 ff.; consequently, the protection of the weakest people under EU law is taken into consideration as far as it is functional to and supports market performances: A. SOMMA, *Scienza giuridica, politica ed economica nell'uniformazione del diritto privato,* in *osservatoriosullefonti.it,* 1, 2021, 306 – 307.

[162] J.H.H. WEILER, *The Transformation of Europe,* in J.H.H. WEILER, *The Constitution of Europe – «Do the new clothes have an Emperor» and other essays on European Integration,* Cambridge, 1999, 10 ff.

debate about the implications of regulation and regulatory choices at large, with regard to the underlying conflicting rationales (economic and private interests against public interest).[163] A quest for both alternative ways of regulation and the relevant governance systems[164] has been undertaken by the EU in order to strike a balance between market freedoms and the scope of public authority intervention aimed at protecting fundamental rights.[165] This quest has once again appeared when dealing with emerging disruptive technologies and the underlying issue of data quality for the training of algorithmic systems that lies at the very root of (some) Artificial Intelligence systems.[166]

In this last regard, the EU, with its Strategy on AI, is trying to strike a fair balance between conflicting rationales by means of a regulatory framework that pursues an AI that serves mankind and adheres to a "human – centric" approach in compliance with fundamental values and rights,[167] while boosting innovation, competitiveness and economic growth.[168] Within this general framework, data quality is a component of the process and its regulation not only requires the accountability principle (as already provided for by the GDPR) but also co – regulatory tools.[169] Against this backdrop, harmonised standards, standards and technical specifications come into play, but – differently from the past – the European Union seems to be following a reverse path in respect of the usual deference towards standard – setters.[170] Not only does Article 41 of the AI proposal support this direction, but the new European Strategy on standardisation as well[171]: both expressly show their awareness about the possible existence of fundamental rights and principles behind the façade of the technical framework of standard – setting activities and try to provide some "new" remedies.

[163] M. FEINTUCK, *Regulatory rationales beyond the economic: in the search of the public interest,* in R. BALDWIN, M. CAVE, M. LODGE (eds.), *op. cit.*, 39 ff. In order to legitimize the public interest rationale against the economic rationale, the Author makes reference to the underlying fundamental values of constitutional relevance, observing that "in the absence of the establishment of values, which can inform the regulatory endeavour[…]we are left with regulation in pursuit of that which can be measured in economic terms – we may end up exclusively valuing the measurable, rather than measuring, and regulating for, the valuable".

[164] *Ex plurimis*, O. DE SCHUTTER, J. LENOBLE (eds.), *Reflexive Governance: Redefining the Public Interest in a Pluralistic World,* Oxford, 2010.

[165] R. BALDWIN, M. CAVE, M. LODGE, *Introduction: Regulation – The field and the developing agenda,* in R. BALDWIN, M. CAVE, M. LODGE (eds.), *op. cit.*, 3 ff.

[166] With regard to the connection between data quality in AI systems and the European strategy on data governance, see L. AMMANNATI, *Per una Intelligenza Artificiale affidabile. Presupposti e sviluppi della sua regolazione,* cit.

[167] *Building Trust in Human – Centric Artificial Intelligence* – COM(2019) 168 final –, 2. As underlined by C. CASONATO, G. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione europea in materia di intelligenza artificiale,* in *BioLaw Journal,* 3, 2021, 437, it is not only a question of technically assessing new AI systems, it is rather the definition of the societal model that is at stake when drawing the framework of a sustainable way of cohabitation between human and non – human components according to a renewed subsidiarity principle.

[168] Communication from the Commission*, Artificial Intelligence for Europe* – COM(2018) 237 final, 5.

[169] As for the rationales fostering the European proposal of Regulation on AI, more specifically the risk – based approach and reliance on the "New Legislative Framework", qualified by the reference to essential safety requirements defined by EU legislations "whereas the task of giving these essential requirements a more concrete form is entrusted to the three European standardization organizations", see M. EBERS, *Standardizing AI – The Case of the European Commission's Proposal for an Artificial Intelligence Act,* cit., 335.

[170] With the consequent risk of «regulatory capture» as pointed out by A. SIMONCINI *Forum: Law and Artificial Intelligence,* cit., 500.

[171] *An EU Strategy on Standardisation – Setting global standards in support of a resilient, green and digital EU single market* – COM(2022) 31 final.

More specifically, if Article 40 sets out the usual presumption of conformity for AI systems compliant with the harmonised standards published in the Official Journal of the European Union,[172] in accordance with the general provisions of Regulation (EU) No 1025/2012 on European Standardisation, it is Article 41 that adds a peculiar clause: if consistent harmonised standards are unavailable or the Commission considers that the relevant harmonised standards are insufficient or that there is a need to address specific safety or fundamental right concerns, the Commission may, by means of implementing acts, adopt common specifications that produce the same presumption of conformity.[173] This monitoring role of the European Commission is even worthier if conceived with reference to data quality, due to a twofold intertwined matter: on the one hand, data quality – as many times repeated – lies at the very roots of Artificial Intelligence systems (more specifically those systems that work by means of Machine Learning algorithms) and its low quality or biases can bring about fundamental rights infringements; on the other hand, Article 10 of the AI proposal has drafted data quality features by means of very wide and flexible wordings, and therefore remits the effective implied choices to the broad margin of manoeuvre of the European standard – setters (ESOs).[174]Thus, the abovementioned European Commission tool of intervention lays down a reversal path: the remittance of discretionary power to the ESOs is indeed under condition; more specifically, should the European Commission consider a need to address fundamental rights concerns or that the relevant standards are insufficient, the discretionary power delivered by Article 40 is retrieved and it will be the Commission itself (under the comitology examination procedure) that will undertake the task by means of implementing acts. Against this backdrop, and taking stock of what has been said until now, this should very likely be the case for data quality standards.[175]

This recovery of "constitutional" awareness is further confirmed by the EU strategy on standardisation. The latter – *expressis verbis* – admits that "more than ever, standards do not only have to deal with

[172] *Artificial Intelligence Act* – COM(2021) 206 final –, Article 40. A. MONICA, *Regulating AI and the key – role of standard in the co – regulation of ICT: EU, Member States and private entities,* in *MediaLaws*, 3, 2021, 145 ff., stresses the importance of standardisation in the proposal on AI. As for a critical perspective in respect of "the real rule – making" role delivered by the AI proposal to ESOs, see M. VEALE, F. ZUIDERVEEN BORGESIUS*, Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach,* in *Computer Law Review International,* 4, 2021, 105.

[173] *Artificial Intelligence Act* – COM(2021) 206 final – Article 41. This implementing act shall be adopted under the "examination procedure" laid down by Article 5 of Regulation (EU) No. 182/2011, according to which the Committee, composed of the representatives of the Member States, delivers its opinion by majority vote (in compliance with the majority rules enshrined by the TFEU) and in the case of a negative opinion the Commission shall not adopt the implementing act (but there is provision for possible referral to the appeal Committee, under Article 6).

[174] As underlined by M. EBERS, *Standardizing AI – The Case of the European Commission's Proposal for an Artificial Intelligence Act,* cit., 336, "all of these requirements are worded in a rather broad way. Instead of formulating the requirements for high – risk AI systems itself, the regulation defines only the essential requirements, whereas the details are left to standards elaborated by the ESOs. For example, AIA states that training, validation and testing data should be 'relevant, representative, free of errors and complete' (Art. 10(3) AIA), to ensure that the AI system 'does not become the source of discrimination prohibited by Union law' (Recital (44) AIA), without indicating what forms of biases are prohibited under the existing framework".

[175] M. VEALE, F. ZUIDERVEEN, *op. cit.,* 105, observes that "the Draft AI Act's value – laden nature might plant a constitutional bomb under the New Legislative Framework. Even 'technical' safety standards entail value – laden choices".

technical components, but also incorporate core EU democratic values and interests, as well as green and social principles".[176] This is the reason why the Commission is fostering a "democratisation" of the representativeness and inclusiveness of the decision – making process of the European Standardisation Organisations (ESOs), by means of "addressing uneven and intransparent representation of industrial interests and increasing the involvement of SMEs, civil society and users"; it also calls upon the ESOs to allow free access to standards and other deliverables.[177] Moreover, the Strategy on Standardisation points to certain European "critical standardisation urgencies" with specific regard to new emerging technologies, including standardisation of data,[178] and tries to foster the need to (re)take the lead at international level in shaping "international standards in line with […] [EU] values and interests".[179] In a similar vein, the Strategy repeats – once more – that "as standards do not only regulate the technical aspect […], but can have an impact on people, workers and the environment, an inclusive and multi – stakeholder approach can bring important checks and balances to standards – making".[180] This is the reason why the European Commission envisages the possibility of its direct intervention in absence of an adequate governance reform carried out by the ESOs. Moreover, well in line with this steering and monitoring role, recent trends in EU legislation have provided the Commission with the power to adopt technical or common specifications, in place of harmonised standards adopted by the ESOs.[181]

In respect of these recent trends, recalled by the EU Strategy on standardisation, the possibility of "alternative intervention" is also reinforced by Article 41 of the AI proposal, since the power of the European Commission to adopt common specifications by means of implementing acts in substitution of the ESOs, is not only set out when harmonised standards do not exist or are deemed insufficient, but also when the Commission considers that "there is a need to address specific [...] fundamental rights concerns". In this last respect, the European Commission's margin of manoeuvre is broadened, well beyond the technical evaluation of the adequacy of the existing harmonised standards and far

[176] *An EU Strategy on Standardisation* – COM(2022) 31 final, 4. Indeed, this approach for standards that embed fundamental values was already sketched by the GDPR, as pointed out by E. CELESTE, G. DE GREGORIO, *op. cit.*, 4 ff.

[177] *An EU Strategy on Standardisation* – COM(2022) 31 final, 4. As underscored by the European Commission's Report on the implementation of Regulation (EU) No. 1025/2012, from 2016 to 2020 – COM(2022) 30 final, 2 – 3, concern has been expressed by Small and Medium Enterprise associations, civil society organisations and consumers organisations (see Annex III, Regulation (EU) 1025/2012) about the real inclusiveness of the decision – making process within ESOs. In this regard, as stressed by doctrine, the purpose of the European Commission to strengthen participation and inclusiveness may further be hindered with specific reference to the Artificial Intelligence field: "It is unclear whether limited existing efforts to include stakeholder representation will enable the deep and meaningful engagement needed from affected communities. The vast majority will have absolutely no experience of standardisation, and may lack EU – level representation", in these terms, M. VEALE, F. ZUIDERVEEN, *op. cit.,* 105.

[178] *An EU Strategy on Standardisation* – COM(2022) 31 final, 1 – 2.

[179] *Ibidem*, 5.

[180] *Ibidem*, 6.

[181] *Ibidem*, 5. Indeed, this "European Commission's alternative intervention" approach is spreading by means of the most recent regulations, i.e., the provision of common specifications adopted by the European Commission by means of implementing acts in place of harmonised standards adopted by ESOs not only when they are lacking but also when the Commission deems them to be insufficient or when there is a need to address public health concerns. Concerning this matter, see Article 9 of Regulation (EU) 2017/745 on medical devices and Article 9 of Regulation (EU) 2017/746 on *in vitro* diagnostic medical devices.

beyond the formal control usually carried out before the publication of a standard in the Official Journal or when it outlines the requirements in its request for standards.[182] More specifically, the European Commission is thus endowed with a more extensive and comprehensive assessment of constitutional relevance that could in turn result in a reduction of the regulatory role usually recognised to standard – setters,[183] as indeed proven by the concomitant criticisms raised by the ESOs in respect of the proposed Article 41.[184]

At this point in time, and to sum up, it is evident that the path followed by these EU proposals towards a curtailed role of EU standard – setters (moreover when standards underlie fundamental rights concerns) also supports the recovery of the boundaries originally tracked by the Meroni doctrine[185] in reference to the legitimacy of delegation of discretionary power, often blurred in favour of the ESOs under the veil of the apparent technical nature of standard – setting procedures.[186]

This EU effort could be – in turn – deemed as part of the broader attempt of the EU legislator to closely monitor private power,[187] as outlined by the Digital Service Act and the Digital Markets Act.[188] In doing

[182] As for the non – substantial role of the European Commission when, according to Article 10, parr. 2, 5,6, it issues the request for harmonised standards and sets out their requirements as well as when it assesses the harmonised standard compliance with EU legislation and requirements, see P. CUCCURU, *Regulation by Request: On the Role and Status of the 'Standardisation Mandate' under the New Approach,* in M. ELIANTONIO, C. CAUFFMAN (eds.), *op. cit.*, 63: "the ways in which mandates are drafted, monitored and enforced reveal that public powers may not necessarily take the lead in the 'implementation' of EU product harmonisation measures".

[183] As stressed by A. MONICA, *op. cit.*, 151, "this means that the Commission has to monitor the implementation of standard".

[184] Regarding this matter, CEN – CENELEC have expressed concerns with respect to Article 41 of the proposed Regulation on AI, by criticising the possibility of the European Commission to adopt an implementing act in place of harmonised standards: see the CEN – CENELEC Position Paper on the *Proposal for a Regulation laying down harmonised rules on Artificial Intelligence – Artificial Intelligence Act* – COM(2021) 206, October 2021, 3

[185] C – 9/56, *Meroni & Co., Industrie Metallurgiche spa v. High Authority of the European Coal and Steel Community.* According to the stance taken by the European Court of Justice, in order to respect the institutional balance principle, delegation of powers is possible only in reference to mere executive powers which shall be justiciable before a court.

[186] M. EBERS, *Standardizing AI – The Case of the European Commission's Proposal for an Artificial Intelligence Act,* cit., 342. On the matter of the softening over time of the Meroni requirements see also, P. CUCCURU, *Regulation by Request: On the Role and Status of the 'Standardisation Mandate' under the New Approach,* in M. ELIANTONIO, C. CAUFFMAN (eds.), *op. cit.*, 49.

[187] For a description of the scope and features of the dominant position gained by Big techs and the ecosystems they have structured, acting as real gatekeepers, see L. AMMANNATI, *I 'signori' nell'era dell'algoritmo,* cit. As pointed out by G. DE MINICO, *Big Data e la debole resistenza delle categorie giuridiche. Privacy e lex mercatoria,* in *Diritto Pubblico,* 1, 2019, 4, a real process of big tech power limitation should imply the re – framing of the traditional antitrust approach, in order to conceive data concentration and privacy standards as subsets of possible abuse of dominant position or unfair competition.

[188] As stressed by M. BETZU, *I poteri privati nella società digitale: oligopoli e antitrust,* in *Diritto Pubblico,* 3, 2021, 739 ff. Similarly, A. SOLA, *Primi cenni di regolazione europea nell'economia dei dati,* in *MediaLaws,* 3, 2021, 188 and 194, underscores the fact that by means of its Data Strategy, and the proposals on the DSA and DMA, the EU is striking a fair balance between the involved social interest and seeks to create a model of data management that is different from that of private operators, as well as from States such as China and the USA in order to become an international benchmark.

so, the EU seems to be trying to recover the "traditional path of constitutionalism"[189] by enhancing some tools of public power aimed at safeguarding fundamental rights while limiting the otherwise progressively increasing scope of decision of private powers.
Is the time perhaps ripe for starting to implement a "data quality due diligence"[190] strictly steered by public regulatory interventions because of its fundamental rights implications[191] and as a consequence to re – frame the traditional approach to standardisation?

8. Conclusions

At this point in time, the veil of Maya has therefore been cast aside: behind the technical features of data quality and its standardisation, it is an intertwined issue of constitutional relevance that bears, more specifically, the entanglement between fundamental rights and regulatory approaches. As usual, the result is that "the devil is in the details": because of the non – objectivity and non – neutrality of both the data (in particular when deployed by Machine Learning algorithms in order to find out correlations and patterns) and its underlying quality assessment, the consequent request for standardisation is similarly affected by discretionary choices that overcome the technical domain.
In this regard, from a general and broader perspective, this issue comes across with regulatory and governance questions, qualified by a hybridisation process that is typical of the global landscape.[192] It is indeed the relationship between law, technology and the market that is at stake and the underlying continuous struggle aimed at striking a fair balance between conflicting rationales.[193] On the one hand, the public power stands with its classical manifestations of authority (the legal system), while, on the other hand, the increasing effectiveness gained by transnational private regulations,[194] the consequent different regulatory spaces[195] and relevant regulatory techniques[196] clash with it.

[189] Borrowing from the wording of C. CASONATO, *Potenzialità e sfide dell'intelligenza artificiale,* in *BioLaw Journal,* 1, 2019, 178. This Author has also expressed his concern for the «individualistic» approach to consent adopted by the GDPR, denouncing it as insufficient in a period of Big data analytics and expressing the need to enhance it with a State intervention: C. CASONATO, *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro,* in *BioLaw Journal,* 2, 2019, 720.
[190] Borrowing and paraphrasing the "data due process" that qualifies the current digital constitutionalism, according to O. POLLICINO, *L'impatto dell'IA sul diritto e sui diritti*, in *BioLaw Journal*, 1, 2020, 492.
[191] The claim for a multi – stakeholder approach steered by public authorities in order to set out common standards for the protection of fundamental rights in respect of AI, is stated by A. PAJNO, M. BASSINI, G. DE GREGORIO, M. MACCHIA, F.P. PATTI, O. POLLICINO, *AI: profili giuridici. Intelligenza Artificiale: criticità emergenti e sfide per il giurista,* in *BioLaw Journal,* 3, 2019, 7. In similar direction, as for a co – regulation significantly surveilled by public authorities, see A. SIMONCINI *Forum: Law and Artificial Intelligence,* in *BioLaw Journal*, 1, 2020, 500; E. STRADELLA, *Approaches for regulating technologies: lessons learned and concluding remarks,* in E. PALMERINI, E. STRADELLA (eds.), *op. cit.*, 354; G. MOBILIO, *op. cit.*, 423. For a critical approach to this "privatisation process" of rules underpinned by States, see L. ANTONINI, *Globalizzazione e nuove sfide del costituzionalismo,* in *Diritto Pubblico,* 2, 2019, 323.
[192] M.R. FERRARESE, *Globalizzazione giuridica,* cit., 6.
[193] Among constitutional law scholars, it suffices to quote G. AZZARITI, *Diritto e conflitti – Lezioni di diritto costituzionale,* Roma – Bari, 2010, 198 ff. Among civil law scholars, it suffices to quote, N. IRTI, *Il diritto nell'età della tecnica,* Naples, 2007, 11 ff.
[194] F. CAFAGGI, *op. cit.*, 77 ff.
[195] S. CASSESE, *Chi governa il mondo?* Bologna, 2013, 15 ff.
[196] R. BALDWIN, M. CAVE, M. LODGE (eds.), *The Oxford Handbook of Regulation,* cit., 104 ff.

In this regard, the path that the EU is embarking upon by means of its Strategies on Artificial Intelligence and Standardisation can be deemed welcome. They are not only chronologically but also substantially intertwined and support the recovery of the margin of manoeuvre of public power *vis – à – vis* private power, whenever democratic values and fundamental rights are involved.

More specifically, in respect of the subject – matter addressed by this paper (data quality and standards), two EU requests and the connected conditions fit the purpose of a major rights – and public interest – oriented approach. On the one hand, the EU has urged for a reform of standard – setters' governance (calling for more transparency as well as representativeness in the standardisation process); on the other hand, it has drafted a data quality governance process that makes reference to harmonised standards. However, should there be a lack of an adequate approach to both, the European Commission will waive and replace the margin of manoeuvre previously left to standard – setters, by means of its direct intervention. On the one hand, should the ESOs not be able to modernise their governance system according to the EC's guidelines, it will be the European Commission itself that will undertake the relevant initiatives.[197] On the other hand, should harmonised standards fail to address "specific safety or fundamental rights" protection needs[198] (really relevant for data quality, as seen above) in high – risk Artificial Intelligence systems, it will, once again, be the European Commission itself that will implement consistent common specifications. Consequently, the foundation of a restriction of the margin of intervention previously granted to private powers (standard – setters) has begun to be established as well as the recovery of the original meaning of the Meroni doctrine.

This path assumes even more importance when it is put in relation to data quality and its convergence (i.e. standardisation) process through the EU, because of its standing at the very root of the current increasing deployment of Machine Learning methodologies, in order to avoid the giant (embodied by these new emerging technologies) rests on constitutional "clay feet", thus embedding more constitutional awareness within harmonised data quality requirements.

Is the provided European Union intervention a non – sufficient condition for achieving an adequate constitutional recovery because of the dominant role of the European Commission (and the relevant comitology procedures) when common specifications for data quality are adopted? This is a further question that delves into the never – ending issue of the EU democratic legitimacy trap.[199] Anyway, what is certain is that the provided possibility of intervention by the European Commission opens the door to a reframing of the traditional regulatory approach to standardisation, by means of the monitoring and steering role of the public power in respect of private powers and experts, that is deemed extremely relevant for the sake of the fundamental rights and principles involved in data quality. In consideration of the foregoing, why not extend this data quality standardisation beyond high – risk AI systems in order to lay down a stronger basis from the outset, in case the quickly changing technological scenario gives evidence to risks for fundamental rights implied in a previously qualified limited or minimal risk AI system?

[197] *An EU Strategy on Standardisation* – COM(2022) 31 final, 4.
[198] *Artificial Intelligence Act* – COM(2021) 206 final – Article 41.
[199] Literature on the issue of the EU democratic deficit is really wide, it suffices to quote J.H.H. WEILER, *European Democracy and its critics: polity and system,* in J.H.H. WEILER, *The Constitution of Europe – "Do the new clothes have an Emperor? And other essays on European integration,* Cambridge, 1999, 264 ff.