Feedback for the EU Commission Inception Impact Assessment towards a "Proposal for a Regulation of the European Parliament and the Council laying down requirements for Artificial Intelligence"

Denise Amram – Giovanni Comandé[1]

*LIDER Lab, DIRPOLIS Institute, Scuola Superiore Sant'Anna (Pisa- Italy)

Table of contents: 1. Introduction. 2. Selecting Option 4 with option 3.c as a baseline (*Option 4+3.c*) 3. A combined approach towards AI Regulation. 4. A role for the GDPR.

1. Introduction.

This feedback is provided considering the ongoing studies undertaken within the research lines RIGHTS in the classifying society (https://www.lider-lab.sssup.it/lider/rights/) and ETHOS (EThics & law witH and fOr reSearch, https://www.lider-lab.sssup.it/lider/ricerca/linee/ethos-ethics-law-with-and-for-research/) developed within the LIDER Lab research activities (www.lider-lab.eu)[2] at Scuola Superiore Sant'Anna (SSSA; www.santannapisa.it).

Our remarks focus on two main issues: 1) providing operational tools to link the ethics and the legal dimension of a Trustworthy AI avoiding risks of ethics washing; 2) the role that the EU Regulation 2016/679 (General Data Protection Regulation, hereinafter "GDPR") may play to achieve the purposes of the EU Strategy on Artificial Intelligence, providing a multilevel legal

[1] Denise Amram, PhD (SSSA, is currently Affiliate Researcher at Scuola Superiore Sant'Anna and Data Protection Officer (denise.amram@santannapisa.it) – Giovanni Comandé, PhD (SSSA) LLM (Harvard), is Full Professor of Private Comparative Law at Scuola Superiore Sant'Anna and Director of LIDER – Lab (giovanni.comande@santannapisa.it).

[2] This position paper has been developed within the "SoBigData Plus Plus: European Integrated Infrastructure for Social Mining and Big Data Analytics" Project, funded by the EU Commission under the H2020 INFRAIA-1-2019 programme (GA 871042), starting from the ideas developed in Denise Amram. *The Role of the GDPR in Designing the European Strategy on Artificial Intelligence: Law-Making Potentialities of a Recurrent Synecdoche*. Opinio Juris in Comparatione, [S.l.], jul. 2020. ISSN 2281-5147. Available at: http://www.opiniojurisincomparatione.org/opinio/article/view/145/153 and Giovanni Comandé. *Unfolding the legal component of trustworthy AI: a must to avoid ethics washing*. In *Annuario di diritto comparato*, ESI, 2020, forthcoming (available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3690633).

framework that may include a General Regulation on Artificial Intelligence and specific safeguards both in terms of national and domain legislation, as well as in terms of soft law.

2. Selecting Option 4 with option 3.c as a baseline (Option 4+3.c)

Both addressed issues lead to suggest **selecting option 4 (combination of the other options) using as a starting point option 3 (we call this Option 4+3.c)**.

Indeed, *establishing general principles on mandatory requirements on issues such as training data, record-keeping about datasets and algorithms, information to be provided, robustness and accuracy and human oversight would offer an horizontal light set of rules enabling guidance for a reliable uptake on the market of AI solutions (option 4 based on 3.c as a baseline) on which to innervate eventual specific (high-risk or category specific AI applications).* Such a combination would a) offer a clear framework, easily linked to the overarching protection of fundamental rights and flexible enough to enable malleable degrees of self-regulation to be adapted according to the specificities (risks, typology of AI, sector of deployment,…).

Option 4+3.c would help to determine whether existing and applicable legal rules suffice according to the contexts and manners in which the AI is developed and deployed, enabling the unfolding of the legal leg of trustworthy AI as a scalable and flexible approach that ranges effectively between soft-law, labeling and mandatory requirements. These efforts need to consider that only some "requirements are already reflected in existing legal or regulatory regimes while most of those regarding transparency, traceability and human oversight are not addressed by the legislation in many economic sectors"[3]. *Thus, **Option 4+3.c** would downplay requirements for mere industrial applications to optimize processes with little or no human involvement/impact.* Also, it should take into account the relationships between different intelligent agents and between different players to avoid the compartmentation of tasks or ownerships which could (intentionally or inadvertently) hide high human impacts in a setting of Chinese walls. Such an approach would at the same

[3] White paper at 11.

time avoid the costs of an unclear and/or fragmented legal framework while reaping the benefits of avoiding unnecessary regulatory costs (e.g. to an AI system where trust is not relevant and thus a human impact assessment results are low).

In fact, different domains of AI application require more or less different rules, but also different applications in the same domain might need dissimilar rules. From the end-users perspective, AI could exploit known as well as new vulnerabilities with a different impact on fundamental rights. While a one rule fits all is not advisable, **the absence of a general set of rules able to operationalize general principles might lead to further fragmentation both in legal rules with ensuing market fragmentations and barriers**. There are intuitive differences in the deployment of an AI based solution for illness screening and diagnosis or of an AI for designing the appropriate treatment for individual patients. Similarly, there are differences if AI is deployed in a high-level institution or at smaller institution / general practitioner level. The implications in terms of automation and translational biases are diverse[4], as well as the implications in terms of liability. Conversely, in other instances, it is different if the AI exploits a known vulnerability[5] or not.

Relying only on ethical rules, voluntary labelling or soft-law, can end up in many problems for all stakeholders. First of all, it gives the impression that fulfilling certain ethical guidelines absolves one of regulatory obligations, leading to a potential legal disaster for businesses that might find themselves exposed to liability or fines, for instance. A very simple example is offered by possible liabilities that emerge in the use of certain personal datasets to produce an AI system, a use that needs to adhere to the GDPR to avoid triggering its art. 82 (liability) and art. 83 (fines). Of course, indiscriminate data use leaves data subjects exposed and limits the trustworthiness of AI and its uptake.

Conversely, several players can establish their own ethical guidelines on the pretext of a loose and unclear legal framework which, even when followed, might not produce any

[4] G. COMANDÉ, *Intelligenza artificiale e responsabilità tra «liability» e «accountability». Il carattere trasformativo dell'IA e il problema della responsabilità* in *Analisi Giuridica dell'Economia*, 1, 2019, pp 169-188

[5] This is very well illustrated in the automated decision making context for which the WP29 clarified that an important factor in the art. 22 GDPR analysis is whether "knowledge of the vulnerabilities of the data subjects targeted" is used. Article 29 Working Party, Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, WP 251 rev.01, 22.

actual constraints to abuses, while giving end users, customers, consumers, and regulators the impression of a trustworthy solution. Such an illusion causes a general societal spillover in case of litigation related to the AI solutions, and fueling micro-harms to individuals and groups.

In any cases, inaction and uncertainty are inefficient as they leave to ex post legal rules, to case law and litigation, the solution of individual cases in order to offer guidelines for the future. Incidentally, it is worth noting that the establishment of legal parameters helps to level the regulatory playing field, thus enabling a better governance of competition.

We stress the urgent need to give content to the legal dimension of Trustworthy AI and thus avoid risks of ethics washing for a much required regulatory framework for the entire industry. To date the debate has been swinging between downplaying the role of legal rules up to ethics washing[6] (an exercise on which the HLEGAI Guidelines are accused as well[7]) and stressing the threats versus the opportunities of AI[8]. The initiative "Proposal for a legal act of the European Parliament and the Council laying down requirements for Artificial Intelligence" has the opportunity to overcome these criticalities.

A very large number of initiatives, both public and private, are targeting ethical principles for AI. Any list can only be incomplete[9]. On the other hand, to maintain a sufficiently high

[6] See B. WAGNER, *Ethics as an Escape from Regulation: From ethics-washing to ethics-shopping?*, https://www.privacylab.at/wp-content/uploads/2018/07/Ben_Wagner_Ethics-as-an-Escape-from-Regulation_2018_BW9.pdf; G. COMANDÉ, *Unfolding the legal component of trustworthy AI: a must to avoid ethics washing*, in *Annurio di diritto comparato*, 2020 , forthcoming .

[7] See M. VEALE, *A Critical Take on the Policy Recommendations of the EU High-Level Expert Group on Artificial Intelligence*, in *European Journal of Risk Regulation*, 2020 doi:10/ggjdjs.

[8] GIUSEPPE CONTISSA, FRANCESCA LAGIOIA, MARCO LIPPI, HANS-WOLFGANG MICKLITZ PRZEMYSŁAW PAŁKA, GIOVANNI SARTOR AND PAOLO TORRONI, *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, IJCAI-18.

[9] The following is just a first incomplete list: AI Now Institute (AINI), Association for Computing Machinery (ACM) with its Committee on Professional Ethics (https://ethics.acm.org/2018-code-draft-2/), the Public Policy Council (https://acm.org/public-policy/usacm), the so called the Asilomar Principles of the Future of Life Institute (https://futureoflife.org/ai-principles/), the Foundation for Responsible Robotics (http://responsiblerobotics.org), Googles AI Principles (https://www.blog.google/technology/ai/ai-principles), The Institute of Electrical and Electronics Engineers (IEEE) Global Initiative on Ethics of Autonomous and Intelligent Systems (*IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*, Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2, http://standards.ieee.org/develop/indconn/ec/autonomous_ systems.html), OpenAI (https://openai.com/), Partnership on AI (a partnership on AI industry-led set up by Google, Apple, Facebook, Amazon, IBM, and Microsoft https://www.partnershiponai.org/), Software and Information Industry Association (SIIS) (*SIIS*, Ethical Principles for Artificial Intelligence and Data Analytics, 2017, http://www.siia.net/LinkClick.aspx?fileticket=b46tNqJuiJA%3d&tabid=577&portalid=0&mid=17113), The

level of normativity, some declarations/guidelines anchor them to the normativity of fundamental rights as received in legally binding documents. This is the case for the Council of Europe (CoE)[10], the HLEGAI, and the European Group on Ethics in Science and New Technologies[11]. In these cases, however, there is a risk of collapsing different levels of normativity (ethics and law). It is also for this reason that the debate and the national/international actions still lack a sufficient regulatory grip. By regulatory grip we do not mean strings and laces for businesses but clear guidance. An abstract declaration of a principle, even if it can ideally find almost unanimous agreement, does not provide stakeholders with certainties to act upon, leading either to inaction – waiting for a clear-cut framework – or to abuses, and the exploitation of an apparent regulatory gap.

Asking to be "ethically correct" or merely rely on soft-law and/or labelling schemes at this stage of AI development is like asking car drivers to be attentive in an era in which speed limits, traffic lights and rules were not in place, and driver license or ban on drunk-driving were not enacted[12]. Can an active and trusted AI industry unfold in such a vacuum? Is it not this regulatory gap the real roadblock to innovation and uptake?

With no attempt to operationalize in legal terms the aspiration to technical correctness and legal fairness of data production and use, such a demand is toothless. This operationalization is much demanded and can come only through the legal system.

Any labelling or soft law tool will be more effective and procompetitive if deployed against a solid bedrock of legal rules.

Last but not least, uncertainty has a cost, and trial and error is not a very efficient and reassuring solution for businesses or individuals. Indeed, a cloudy framework does not offer individuals the required awareness of their rights, nor of something bad that might happen to them (micro-

World Economic Forum's Center for the Fourth Industrial Revolution (https://www.weforum.org/center-for-the-fourth-industrial-revolution/areas-of-focus). See also https://ethicsinaction.ieee.org/.

[10] EUROPEAN COMMISSION FOR THE ADMINISTRATION OF JUSTICE (CEPEJ). (2018). European ethical charter on the use of Artificial Intelligence in judicial systems and their environment. Retrieved from: https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c

[11] EUROPEAN GROUP ON ETHICS IN SCIENCE AND NEW TECHNOLOGIES. (2018). Statement on Artificial Intelligence, robotics and 'autonomous systems'. Retrieved from: https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf Electronic copy available at: https://ssrn.com/abstract=3414805 proposed 'a set of basic principles and democratic prerequisites, based on the fundamental values laid down in the EU Treaties and in the EU Charter of Fundamental Rights

[12] Incidentally to have all these "common sense" rules took decades in the 20th century.

harms hardly are perceived, but they accumulate). Thus, Option 0, as a scenario, delays the process of regulating by *ex post* litigation and benefits the least compliant players on the market, while disadvantaging both the most compliant ones and SMEs and start-ups that either incur higher costs or do not have sufficient resources to compete in a fully compliant way.

3. A combined approach towards AI Regulation.

The Inception Impact Assessment lists the EU initiatives in terms of soft law regulation that helped to define the purposes of the legislative initiatives of AI: namely to build up a "trustworthy" ecosystem, considering the need to protect fundamental rights both within the development and the application of the given AI technology.

As known, the High-Level Expert Group on Artificial Intelligence (HLEGAI) distinguished the three pillars 1) lawful, 2) ethical and 3) robust that are clearly interconnected, following an overall approach grounded on fundamental rights protection.

The option we suggested above (lines 24 and ff) should not be a straightjacket, but an enabler of innovation and fast uptake against a background of clear protection for fundamental rights and liberties. It is misleading to think legal rules in terms of roadblocks to innovation. To the contrary, a weak or unclear legal framework can very easily backfire against the development and deployment of AI-based products and services by undermining trust in them.

A clear example supporting our considerations is offered by the new vulnerability related to AI based solutions. Vulnerability of individuals and groups is often evoked in ethical charters and declarations, and the HLEGAI Guidelines are no exception, requiring to "pay particular attention to situations involving more vulnerable groups such as children, persons with disabilities and others that have historically been disadvantaged or are at risk of exclusion, and to situations which are characterised by asymmetries of power or information, such as between employers and workers, or between businesses and consumers" and to acknowledge that "AI systems also pose certain risks and may have a negative impact, including impacts which may be difficult to anticipate, identify or measure (e.g. on democracy, the rule of law and distributive justice, or on the human mind itself.)".

Once again, the ethical call can be seen as a means to distract attention from regulatory needs, implying the sufficiency of abiding by ethical principles. On the contrary, the mentioned claims for ethics can offer interpretative guidance for legal rules applicable to AI-based solutions turning the risk of ethics washing into a path to unleash the evolution of legal rules requiring only to frame the general principles applicable to all AI solutions and leave to specific regulatory or soft-law interventions special situations (high-risks or specific sectors).

Indeed, more in general we should ask ourselves why AI should require a different (allegedly harsher) treatment than other technologies which "benefitted" from years of loose legal frameworks. The reason maybe because, as anticipated, it enables the exploitation of a number of unseen vulnerabilities that have not (yet) triggered a legal reaction, and it does so in a massive, personalized way[13], silently clearing activities and results that are ethically questionable but not yet framed by the law, at least because their specific impact as such might not trigger existing laws, while its effect on individuals accumulate. In this sense, ethics and law are mutually complementary, but this complementarity needs to be operationalized, otherwise it will only lead to potential ethics washing of the regulatory needs.

Thus, if AI offers the opportunity to exploit specific biases, individual and even transient vulnerabilities to establish contingent and lasting asymmetries of powers, a set of reactions need to be put in place to prevent these risks and to reap all the immense benefits Trustworthy AI can bring to humanity

A similar interplay of the "legal leg" of Trustworthy AI with its "ethics leg" can be identified with its robustness component "both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm".

The call for robustness of AI is more or less a call for security from internal flaws and external attacks[14] to avoid expanding the unacceptable reach to old and new vulnerabilities.

Yet, while technical robustness relates to the appropriateness of technical safeguards "in a given context, such as the application domain or life cycle phase", and links to the legal dimension in terms of safety and security-related rules, the "social perspective" of robustness – by relating to the context and environment in which the system operates – is open again to

[13] R. CALO, *Digital Market Manipulation*, in *George Washington Law Review*, 82, 2014, 995.
[14] "3. it should be **robust**, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm".

law but in a different way, requesting to consider closely the actual deployment of the AI solution: calls for legal solutions for AI in the context. Robustness relates to how the operational context may render an AI solution "unfair", how it may lead to asymmetries worth rebalancing, and how it may replicate past biases and discriminations or create new ones[15]. It reflects the characteristic of AI solutions of creating/exploiting context-specific vulnerabilities or undiscovered ones.

Reference to old and new vulnerabilities recalls a dynamic but slowly evolving divide between those vulnerabilities that have emerged in the domain of legal relevance over time (gender, religion, political opinions, age, status, ethnic origins, etc.), gaining different levels of sectoral/vertical legal recognition and protection for instance in terms of antidiscrimination law, consumer law, and the like. On the other hand, it recalls the "new" vulnerabilities enhanced by AI-led solutions (e.g. the exploitation of transient mood statuses, specific situations distinguishing purchase assumed to be for personal or business reasons, the perils of mass personalization for individual autonomy, the massive personalized exploitation of cognitive biases via dark patterns[16], etc.) that would not be relevant under colour of law as such.

Only against a set of general principles it is possible to enable the coordinated reading of existing applicable legal rules. Indeed, in the interplay between fundamental rights enshrined in constitutions, treaties or charters, and material rules, there is a large scope for reinterpretation and balancing; but this balancing approach cannot be left to simple ethical considerations, since this may harm at the same time a rising industry for lack of clear red tapes and individual rights for lack of enforceability.

4. A role for the GDPR

To deal with the lawfulness component shall not be limited to the compliance with the applicable data protection regulation or to the existing framework, even if the GDPR approach

[15] Especially on exposing the discrimination risks see S. RUGGIERI, D. PEDRESCHI, F. TURINI, *Data mining for discrimination discovery. ACM Transactions on Knowledge Discovery from Data*, 4(2):9:1–9:40, May 2010.
[16] Dark patterns are tricks used mostly in websites and apps that make the user do things that she didn't mean to, like buying or signing up for something.

might be considered as the first ally in the setting of a fundamental rights-based legal framework. Yet, the GDPR can offer some useful hints.

In this regard, the structure of the GDPR may influence the new legislative process, as follows.

- AI solutions are based on information: collection, generation, understanding, and exploitation are the main processing activities.
- The risk-based approach aiming at reaching an acceptable level of fundamental rights protection is functional to induce the development of trustworthy AI solutions by design and by default.
- It would be useful to identify roles and responsibilities while designing the AI system, including who is responsible, accountable, and liable to assess risks that may occur in the given AI solution; who is monitoring its development, and who has to comply with the domain regulatory framework according to the current standards and scientific knowledge.
- While the GDPR assigns obligations to the one who determines "means and purposes of the personal data processing", within the AI context, an "AI controller" could be identified as the one who determines *methods for data acquisition* (*i.e.* what function/algorithm is chosen and which data train the algorithm), *actions required* (*i.e.* what task shall the AI perform), and *goals* (*i.e.* which is the final purpose of the automated decision making/reasoning activity).
- In addition, a series of roles could be identified to support the AI controller in the assessment and monitoring activities: likewise the GDPR, an internal distribution of roles and responsibilities, a so-called RACI matrix - that identifies who is Responsible, Accountable, Consultable, and Informed of the AI processing.
- Synergies between other roles could be included (i.e. collaboration between the AI controller and the data protection officer as well as the identification of "AI officer/advisor" and the IT manager).

- As a consequence, an AI external governance could be defined in case more than one AI controller is involved in determining methods, actions, and goals[17].

These features might be included in a binding harmonised regulation, while possible safeguards might be introduced (as well as confirmed) at national level, according to specific domains (e.g. workplace, healthcare, etc.), as well as in terms of soft law regulation.

Special attention shall be given to the fact that AI-based technologies enable the identification and exploitation of many conditions of weakness even *unknown* to the individuals themselves[18]. Vulnerability can derive from the development of AI (e.g. bias) or its application. This very simple fact makes AI-based vulnerabilities contextual and possibly endless, leading to situations of inferiority, dependency, and even *unidentified* subjugation[19]. Note also that AI operates on correlations without much reliance on causal explanations opening to further understudied vulnerabilities and or discriminations[20].

This topic recalls the interplay between the robustness requirement of AI, and the technical security from internal flaws and external attacks[21] (see above also). A hard law approach might seek the banning of non-explainable solutions[22], while a tailored reading of art. 22

[17] On the human control on algorithms, see STEFANO RODOTÀ, *Il mondo in rete* (Roma-Bari, 2017), REMO BODEI, *Dominio e sottomissione. Schiavi, animali, macchine e l'intelligenza artificiale* (Bologna, 2019), and GIUSEPPE ZACCARIA, 'Figure del giudicare: calcolabilità, precedenti, decisione robotica' (2020), *Riv. Dir. Civ.*, 277 ff.

[18] On the idea to identify layers of vulnerability see F. LUNA, *Elucidating the Concept of Vulnerability: Layers Not Labels*, in *International Journal of Feminist Approaches to Bioethics* II, no. 1, 2009, 121–39, https://doi.org/10.3138/ijfab.2.1.121. On the unaccountability of the possible exploitations of vulnerabilities see D. CITRON – F. PASQUALE, *The Scored Society: Due Process for Automated Predictions*, 2014, Faculty Scholarship https://digitalcommons.law.umaryland.edu/fac_pubs/1431.

[19] On the undetermined nature of vulnerability see R. E. GOODIN, *Protecting the Vulnerable: A Reanalysis of Our Social Responsibilities*, Chicago, 1985, 112. Also, examples of discriminations in access to services and goods are numerous V. EUBANKS, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*, New York: St Martin's Press, 2018; F. PASQUALE, *The Black Box Society: The Secret Algorithms That Control Money and In-formation.* Cambridge: Harvard University Press, 2015*;* J. ANGWIN – J. LARSON, *Bias in Criminal Risk Scores is Mathematically Inevitable, Researchers Say*, *ProPublica*, December 2016, https://www.propublica.org/article/bias- in- criminal-risk-scores-is-mathematically-inevitable- researchers-say.

[20] Giovanni Comandé; *"The Rotting Meat Error: From Galileo to Aristotle in Data Mining?",* in *European Data Protection Law Review (2018),* Volume 4, Issue 3, pages 270-277.

[21] "3. it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm".

[22] As argued in the data protection context by J. POWLES, *The Seductive Diversion of 'Solving' Bias in Artificial Intelligence*, 2018, https://medium.com/s/story/ the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53 .

of the GDPR[23] might balance things in a different way, activating for instance a right to oppose the data processing preceding the bidding itself or enabling forms of information capable of effectively reducing the impact on the individual decision-making context.

To sum up our feedback, we argued in favor of Option 4, using Option 3.c as a basis for a General AI Regulation able to bridge lawfulness, robustness and ethics concerns in AI while remaining flexible enough to avoid roadblocks to innovation and uptake for AI solutions. In addition, we briefly proposed some contents for a General AI Regulation, according to a risk-based oriented system of check and balance aimed at ensuring the development of any AI-solutions in light of fundamental rights protection. The Regulation shall consider the peculiarities emerging within the different domains and therefore provide the opportunity for each AI controller to allocate tasks, roles, and responsibilities. Under this system, independent authorities shall provide assistance and promote awareness and trustworthiness among data subjects/end-users/stakeholders.
To boost the cultural and inclusive challenge that the AI is driving, the interdisciplinary dialogue shall be maintained.

[23] G. COMANDÉ - G. MALGIERI, *Why a Right to Legibility of Automated Decision-Making Exists in the General Data Protection Regulation*, in *International Data Privacy Law*, cit.