Comments on White Paper on AI (EC)

Mireille Hildebrandt[1]

General remarks

The White Paper opens with ‘It will change our lives by improving healthcare (e.g. making diagnosis more precise, enabling better prevention of diseases), increasing the efficiency of farming, contributing to climate change mitigation and adaptation, improving the efficiency of production systems through predictive maintenance, increasing the security of Europeans, and in many other ways that we can only begin to imagine.’

- **I object to the term ‘improving’.** We don’t know that. It depends. It would be great if these kind of policy papers would stop assuming what has to be verified on a case by case basis.
- This feeds into the objective of **‘promoting the uptake of AI’,** as if AI as such should be promoted without qualification, though with some amendments for ethical concerns. Whether AI uptake must be promoted depends
  1. on its **reliability,** i.e. on whether it does what it is claimed to do, and
  2. on whether adverse effects can be foreseen, which in turn depends on taking a **precautionary approach**.

The latter is not equivalent with risk aversion but – on the contrary – equivalent with **taking uncertainty seriously** and therefore investing time and money in foresight studies (which may include *computational simulation*, but foremost require *participation by those who will be affected*). My take is that the risk approach taken in the GDPR actually concerns this type of precautionary approach.[2]

The following declaration of intent seems crucial to me (at p. 1-2 of the White Paper): ‘Today most data are related to consumers and are stored and processed on central cloud-based infrastructure. By contrast a large share of tomorrow’s far more abundant data will come from industry, business and the public sector, and will be stored on a variety of systems, notably on computing devices working at the edge of the network. This opens up new opportunities for Europe, which has a strong position in digitised industry and business-to-business applications, but a relatively weak position in consumer platforms.’

- The added value created by generating and manipulating consumer data mostly consists of money flows toward advertising intermediaries (mostly owned by big tech platforms). There is **very little real value** for either advertisers, publishers or consumers in all this,[3] whereas our information ecosystem has been in a

[1] Research Professor of ‘Interfacing Law and Technology’ at Vrije Universiteit Brussels, appointed by the VUB Research Council, co-director of the Research Group for Law, Science, Technology and Society studies (LSTS). Part-time Full Professor of ‘Smart Environments, Data Protection and the Rule of Law’ at the Institute of Computing and Information Sciences (iCIS) at Radboud University, the Netherlands. PI of ERC Advanced Grant ERC-2017-ADG No 788734, see www.cohubicol.com.

[2] Katerina Demetzou, ‘Data Protection Impact Assessment: A Tool for Accountability and the Unclarified Concept of “High Risk” in the General Data Protection Regulation’, *Computer Law & Security Review* 35, no. 6 (1 November 2019): 105342, https://doi.org/10.1016/j.clsr.2019.105342.

[3] ‘The New Dot Com Bubble Is Here: It’s Called Online Advertising - The Correspondent’, accessed 17 November 2019, https://thecorrespondent.com/100/the-new-dot-com-bubble-is-here-its-called-online-advertising/13228924500-22d5fd24; ‘Targeted Ads Offer Little Extra Value for Online Publishers, Study Suggests’, *TechCrunch* (blog), accessed 29 April 2020,

constant state of turbulence over fake news and the more, precisely due to ***treating political opinions in the same vein as consumer preferences***.

- Industry, business and the public sector could indeed make a difference when integrating testable and contestable 'AI' into their operations, **based on verifiable claims and proven reliability**.[4]

On p. 3 the White Paper contends: 'Given the major impact that AI can have on our society and the need to build trust, it is vital that European AI is grounded in our values and fundamental rights such as human dignity and privacy protection.'

- This suggests that we ground European AI in our values and fundamental rights to build ***trust***. It should be the other way round: because we believe in specific values and fundamental rights, we aim to build ***trustworthy*** AI. The point is not to do whatever it takes to gain trust, but to ensure trustworthiness.
- Trust can be gained by nudging constituents into behaviour that public administration believes to be desirable. In a constitutional democracy we use law to clarify what is expected in terms of lawful interaction, we don't want to be coaxed behind our backs.
- Trustworthy AI means testable, tested and contestable applications and infrastructure, coupled with a proper liability regime that ensure that those who take hazardous risk **with other people's interests** will pay the price and therefore think twice.

I have added some thoughts on the 7 key requirements of the Guidelines of the High-Level Expert Group (between brackets):

- Human agency and oversight (please think in terms of 'machine in the loop' instead of 'human in the loop')
- Technical robustness and safety (focus should be on 'methodological integrity', see above and below on mathematical and empirical testability)
- Privacy and data governance (note that within the EU this is called 'data protection', which explicitly aims to protect against violations of all human rights; while data governance must build on art. 5 GDPR if personal data)
- Transparency (public administration should provide its systems as free software by default, and require open source software by default when tenders are assigned)
- Diversity, non-discrimination and fairness (here again there is a clear connection with data protection, and please let's not assume that unlawful bias can be technologically fixed)
- Societal and environmental wellbeing (connection with surveillance and 'new economy')
- Accountability (this should not be paper dragon, meaning we need 'real' testability before market entry and strict liability for those who stand to profit)

https://social.techcrunch.com/2019/05/31/targeted-ads-offer-little-extra-value-for-online-publishers-study-suggests/.

[4] For this we need both mathematical verification and empirical testing, see e.g. Jie M. Zhang et al., 'Machine Learning Testing: Survey, Landscapes and Horizons', *ArXiv:1906.10742 [Cs, Stat]*, 21 December 2019, http://arxiv.org/abs/1906.10742; Miles Brundage et al., 'Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims', *ArXiv:2004.07213 [Cs]*, 20 April 2020, http://arxiv.org/abs/2004.07213.

It is important to briefly mark out to the role of consent in relation to the Single European Data Space (as part of the European Data Strategy), as this seems to be preconditional for the ecosystem of excellence and trust.

- There are incompatibilities between the Digital Content Directive and the GDPR as to the role played by personal data. In the first, consent refers to the acceptance of the terms of a contract, where consent must be understood in terms of the private law of obligations (where notions such as duress and fraud play out). In the second, consent refers to the legal ground for the processing of personal data, which is part of the fundamental right of data protection (where it is e.g. not possible to provide valid consent beyond an explicit, specific and legitimate purpose; and where art. 7.4 jo recital 43 stipulate that consent to process personal data that is not necessary for the provision of a service is not valid if the provision of that service is withheld if consent is not given).
- There are tensions between consent required in the case of the GDPR (data protection), Police DPD (public security), the ePrivacy Directive (confidentiality of communication) and the Open Data Directive (data sharing in the public interest) and the Clinical Trial Regulation (confidentiality of medical data).
- The different roles played by consent in the context of employment, tax or social security fraud detection, justice authorities, education, medical interventions, scientific research, Big Tech, Big Pharma, food chains, etc. raise issues around the idea of a Single European Data Space. The different legislative regimes demonstrate that such a Big Data Space is **distributed** in terms of purpose, access, deployment, protection, legality and lawfulness.[5]

- On top of that, the other ***5 legal grounds*** will play their role in the Single European Data Space, raising even more questions around the safeguards, interoperability, portability, and repurposing of the data.
- Finally, ***data is often incorrect (e.g. outdated), incomplete, biased, or irrelevant*** and without rigorous methodological constraints, the Single European Data Space will generate far more problems than it solves.[6]
-

Regulatory framework for AI

White Paper takes a risk approach, based on two types of risk:

1. Risks to fundamental rights
2. Risks to safety

Please note that this particular risk approach is core to the GDPR, notably in art. 25 (Data Protection by Design and Default) and 35 (Data Protection Impact Assessment) which both aim to assess and mitigate risks **to fundamental rights and freedoms** (not merely data protection or privacy).

[5] Mireille Hildebrandt, *Smart Technologies and the End(s) of Law. Novel Entanglements of Law and Technology* (Cheltenham: Edward Elgar, 2015), section 2.3, p. 30-40.

[6] E.g. with regard to patient data Federico Cabitza, Davide Ciucci, and Raffaele Rasoini, 'A Giant with Feet of Clay: On the Validity of the Data That Feed Machine Learning in Medicine', *Lecture Notes in Information Systems and Organisation*, 1 January 2019, 121–36, https://doi.org/10.1007/978-3-319-90503-7_10.

The section on effective application and enforcement is crucial, please note the issue of budget of the national DPAs that should apply and enforce the GDPR. Action is needed here, notably with regard to **underbudgeting of e.g. the Irish DPA**,[7] and attention must be paid to the resulting **erosion of the the one-stop-shop mechanism** (as it was gamed by Big Tech).[8]

The section on limitations of the scope in EU law should better address the lack of European private law and the need for autonomous interpretation of private law concepts in EU legislation, e.g. art. 79 (injunctions) and 82 (tort liability) GDPR.

Concerning new legislation specifically for AI

1. The definitions of AI in footnotes 46 and 47 are somewhat cumbersome. Serious AI is better described as a system capable of 'adapting to the environment while working with insufficient knowledge and resources'.[9]
2. It makes far more sense to develop legislation for all automated systems that have a potentially major impact on natural persons due to the implications of automation.
3. Automation implies that (1) effects scale, (2) remote control is enabled in time and space, (3) effects are more difficult to foresee, (4) effects are more difficult to redress.
4. Therefor it seems better to follow art. 22 GDPR that has extended previous protection (art. 15 DPD) from profiling only to also include more deterministic decision-making. Though some people find art. 22 unclear, I think it has the right level of abstraction, allowing fine tuning by the industry, public administration, EDPB, EDPS, DPAs and courts in line with further developments. Aligning the scope of AI-specific legislation with the scope of art. 22 GDPR will enhance legal certainty and create a level playing field.
5. Scope should concern (1) all applications that include automated behaviour, that (2) have a significant effect on natural persons. The scope should not be limited to data-driven AI but also include code-driven AI, precisely because of the effects of the inherent automation
6. High risk approach could miss out on distributive effects of low risk AI deployment (e.g. resulting in accumulation of myriad 'low' risks for what Rawls called the 'least advantaged')
7. I added some thoughts on the requirements for high risk deployment (between brackets):
    - training data (relevance, validity and completeness are huge issues; distribution of training/validation/test data is huge issue; relationship with machine readable task, GDPR purpose and e.g. legality principle play out here);
    - data and record-keeping (storage limitation for personal data; reliability of re-use as training data for other purpose is huge issue, from a methodological perspective; preregistration of research design including updates that specify

[7] Nicole Kobie, 'Germany Says GDPR Could Collapse as Ireland Dallies on Big Fines', *Wired UK*, 27 April 2020, https://www.wired.co.uk/article/gdpr-fines-google-facebook.

[8] 'One-Stop-Non?', thecybersolicitor, accessed 29 April 2020, https://www.thecybersolicitor.com/single-post/2019/05/29/One-Stop-Non.

[9] Quoting part of the definition of intelligence of Wang (2008): "The essence of intelligence is the principle of adapting to the environment while working with insufficient knowledge and resources. Accordingly, an intelligent system should rely on finite processing capacity, work in real time, open to unexpected tasks, and learn from experience. This working definition interprets "intelligence" as a form of "relative rationality"'. See p. 1 of Dagmar Monett et al., 'Special Issue "On Defining Artificial Intelligence"—Commentaries and Author's Response', *Journal of Artificial General Intelligence* 11, no. 2 (1 February 2020): 1–100, https://doi.org/10.2478/jagi-2020-0003.

design choices and clarify whether the design concerned exploratory or confirmatory research);

- information to be provided (verifiability of claims made about behaviour of a system);
- robustness and accuracy (accuracy concerns a specified performance metric on validation data; it does not necessarily reflect correctness in real world applications)[10];
- human oversight (we must start thinking in terms of 'machine in the loop', instead of 'human in the loop'; human oversight means that at granular level where AI affects human beings, decisions are made by humans that understand the AI and are competent to change a decision);
- specific requirements for certain particular AI applications, such as those used for purposes of remote biometric identification (highly relevant for crowd control apps such as contract tracking apps, repurposed for other types of surveillance).

As to **addressees** of legal obligations, liability should address those who profit from deployment, because they should be incentivised to reduce risk.

- The type of addressees should not be too large (invites gaming)
- Strict liability will simplify the decisional space for businesses

As to AI-assessments

I think we need both prior and post hoc **conformity-assessments** by those who stand to gain from deployment, and dedicated independent testing/auditing/supervisory bodies (e.g. the latter at p. 26).

Prior and post hoc **fundamental rights impact-assessments** by e.g. the controller (GDPR) or the entity that stands to gain from an application (Product Liability, Machinery Directive).

[10] On high accuracy that turned out completely incorrect, see e.g. Rich Caruana et al., 'Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-Day Readmission', in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15 (New York, NY, USA: ACM, 2015), 1721–1730, https://doi.org/10.1145/2783258.2788613. On the need for methodological integrity see e.g. Iris Van Rooij and Giosué Baggio, 'Theory before the Test: How to Build High-Verisimilitude Explanatory Theories in Psychological Science', 2020; Tal Yarkoni, 'The Generalizability Crisis', preprint (PsyArXiv, 22 November 2019), https://doi.org/10.31234/osf.io/jqw35.