Radboud University

IN·DEI·NOMINE·FELICITER

GENERAL RESPONSE TO THE EUROPEAN COMMISSION'S WHITE PAPER "ON ARITIFICIAL INTELLIGENCE – A EUROPEAN APPROACH TO EXCELLENCE AND TRUST"

Radboud AI
Radboud University, Nijmegen, The Netherlands

12 June 2020

Hereunder we present a general response to the European Commission's (EC) White Paper on an Ecosystem of Excellence, an Ecosystem of Trust, and regarding data infrastructure incl. GDPR:

Ecosystem of Excellence

Fundamental research

The White Paper gives the impression that AI is "finished" and can now be deployed and applied to different sectors. We urge the EC to not lose sight of the challenges that still lie ahead of us in the foundational (algorithmic) dimension of AI, including machine learning (for a good overview, we refer to the Dutch AI Manifesto: http://ii.tudelft.nl/bnvki/wp-content/uploads/2019/09/Dutch-AI-Manifesto-2019.pdf). Parallel to ensuring take up of AI technologies by the public and private sector, it is paramount to ensure dedicated, long-term investment in fundamental research into the theoretical foundations of AI. Only if the EU aims to be a leader in fundamental AI research can we lead the development and application of ethical, human-centred AI. Europe is well positioned to take this lead (compared to the USA and China), conditional upon the dedicated investment of the EC in fundamental AI research.

Uptake of AI by the public and private sector

The White Paper is very industrial oriented to overcome the well-known barrier between academic research and application. In that sense the setting up of a public-private partnership is relevant. However, including the human rights perspective prominently is challenging in this setting. For this the public sector would be an interesting testing place for the take up of AI applications.

Skills and education

- There is very little mention of how work practices will change and what kind of de-skilling and re-skilling should be expected. This is addressed very briefly, p. 6, but more from the perspective of teaching the workforce new skills. For the highly skilled workforce, like doctors, there will also be changes in what they do, as they work with/alongside/complementarily with AI. This comes back a little in the section on recommendations for human oversight (p. 21). However, the EC should pay a lot more attention to how work will change, also for white collar professions. Proper training/re-skilling for affected professionals should aim at keeping up with these changes, ensuring continued fulfilment of work and high quality of life.
- To counter the risk of exclusion and access to technology, it is important to decrease the gap between the digital literate and illiterate and the associated differences in health and socioeconomic status (although aspects of diversity, non-discrimination and fairness are mentioned). Here Language and Speech Technology (LST) can play an important role in online and offline education programs. LST focusses on atypical communication processing (language learners and people with language disorders/pathologies) which inherently leads to a focus on inclusiveness.

Ecosystem of Trust

Regulatory framework

- The white paper adopts a perspective that aims to support AI developers/producers (the industrial perspective of AI) while neglecting the social dimension (for what kind of socially desirable goals and aims would we need AI?). As such, it is more oriented towards the first goal (ecosystem of excellence) than the second (ecosystem of trust). It is up to the producers to demonstrate their applications may have socially added value.
- Missing: the possible need for a sectorial approach to regulation concerning discrimination and fairness that Prof. Frederik Zuiderveen Borgesius (Radboud University) recommended in his report (this is addressed in other terms on p. 17, where they speak of a targeted approach, but based on risk level; see: https://www.coe.int/en/web/artificial-intelligence/-/news-of-the-european-commission-against-racism-and-intolerance-ecri-)
- The white paper focuses upon AI as a kind of product that requires the usual liability, safety and consumer protection approach. This misses the approach of AI as part of a more encompassing socio-technical system in which it not only functions as a product/service that has an input and an output but that meanwhile changes the system in which it operates: social norms may become impacted on what is deemed acceptable (in whatever sense) as well. How to understand liability, safety and consumer protection for AI if we do not understand how AI will impact upon these categories?
- The relevance of AI for the public sector is hardly elaborated upon as a critical topic. Of course, it is sketched in the white paper but when it comes to trust the way AI is used in public sector domains (taxation, social domain, health domain) is terribly relevant, for obvious reasons. In promoting a trustworthy use of AI one would expect much more attention for what it means in terms of regulating use of AI in the public domain. Now, the common rules (ranging from strict human intervention to loose human intervention) may be followed at will.
- We endorse the analysis made by ALLAI on the white paper, as it makes some very good points (see https://allai.nl/first-analysis-of-the-eu-whitepaper-on-ai/):
  - “Suggesting that we can remove all biases in (or even with) AI is wishful thinking at best and an error of language at worst. In either case, for the purposes of any regulatory framework we should not merely focus on technical solutions at data set level, but devise socio-technical processes that help us:
    a. understand the potential legal, ethical and social effects of the AI-system and improve our design and implementation choices based on that understanding;
    b. audit our algorithms and their output to make any biases transparent; and
    c. continuously monitor the workings of the systems to mitigate the ill effects of any biases.”
  - “Trustworthy AI needs a continuous, systematic socio-technical approach, looking at the technology from all perspectives and through various lenses. For policy making, this requires a multidisciplinary approach where policy makers, academics from a variety of fields (AI, data-science, law, ethics, philosophy, social sciences, psychology, economics, cybersecurity), social partners, businesses and NGO’s work together on an ongoing basis.”
  - “When considering governance, the focus should not just be on the technology, but more on the social structures around it: the organizations, people and institutions that create, develop, deploy, use, and control it, and the people that are affected by it, such as citizens in their relation to governments, consumers, workers or even society as a whole.” (First analysis of the EU Whitepaper on AI by ALLAI)

Data infrastructure incl. GDPR

We miss the following aspects considering data infrastructure and GDPR-related issues in the White Paper:

- The concept of FAIR data, connection to European Open Science Cloud and automated quality and integrity assessment procedures for data;
- safe and GDPR-compliant access to LST resources needed to train tools for educational and health related purposes;
- the concept of federated learning, up to individual's cell phones;
- the concept of machine readable data licenses that specify what can be done with what data and by whom;
- possible limitations of the current regulation (GDPR) where data can only be collected with a specific purpose, which does not consider advancements in AI algorithms.

Response to the public consultation on the European Commission's White Paper on Artificial Intelligence – legal and regulatory issues

1. General remark: Artificial intelligence and the precautionary principle

Artificial intelligence (AI) is still at an emerging stage, and the risks it poses to individuals and society are still uncertain. For this reason, the European Commission should adopt a precautionary stance, as was recommended by the High-Level Expert Group on Artificial Intelligence.[1] The precautionary principle is enshrined in the Treaty on the Functioning of the European Union, and its scope has been extended from its original environmental domain, to issues of food safety, or nanotechnology. It appears quite logical that it should also apply to novel digital technologies that carry important and uncertain risks for the EU citizens' fundamental rights and freedoms. Furthermore, the precautionary principle is one of the elements of the EU's responsible research and innovation policy, which underpins its whole innovation and research agenda.[2] The precautionary principle should not be understood as stifling research and innovation. On the contrary, it is a call for erring on the side of EU citizens' fundamental rights in the face of uncertain risks, and for the active identification of knowledge gaps.[3] Crucially, the precautionary principle has also been associated with calls for broad public participation of all affected and concerned publics.[4] This appears crucial given the White Paper's ambition to address the weak level of citizens' trust, and to provide for maximal stakeholder's participation.[5] A precautionary approach would ensure that not only institutional and corporate stakeholders are represented but also ordinary citizens and all those affected by AI applications, and who have a claim to make in this regard.

[1] See High-Level Expert Group on Artificial Intelligence, 'Policy and Investment Recommendations for Trustworthy AI' (2019) 37–38.

[2] https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation

[3] René von Schomberg, 'A Vision of Responsible Research and Innovation' in R Owen, M Heintz and J Bessant (eds), *Responsible Innovation* (John Wiley 2013) 23.

[4] See, Ulrike Felt and others, 'Taking European Knowledge Society Seriously: Report of the Expert Group on Science and Governance to the Science, Economy and Society Directorate, Directorate-General for Research, European Commission' (2007).

[5] European Commission, 'White Paper on Artificial Intelligence: A European Approach to Excellence and Trust' (2020) 2, 25.

2. Risk-based approach to regulating AI

2.1 Number of risks

The White paper only foresees two types of risks: high-risks and non-high risks. This is too little. There should at least be 3 categories: high-risks, medium risks, and low risks. Only low risks should be minimally regulated with the type of voluntary self-regulation mechanisms that the European Commission proposes.

2.2 High-risk criteria

In order to determine whether an AI application presents a high risk, the White paper puts forth 2 cumulative criteria: the sector in which it is used (e.g., healthcare, transport), and the type of use (e.g., mortgage allocation or spam filter).[6] The author fears that such a system of high-risk criteria might not be sufficiently horizontal and complete, and therefore fail to adequately cover all AI applications. In this respect, some valuable lessons can be learned from the General Data Protection Regulation's (GDPR) own system of high-risk criteria. Art. 35 GDPR on data protection impact assessments adopts criteria that are based on the properties of the processing operation at stake, and for this reason are applicable in a horizontal/transversal way. These include the scope, context, nature, purpose, and type of processing, as well as the type of data, and the impacts on the data subjects' fundamental rights and freedoms. The advantage of such a general, horizontal approach is its inclusive nature and applicability to all types of AI applications. As in the case of the GDPR, the application of these general criteria should be further refined by European and national authorities' guidance documents.[7] They should also be complemented by additional secondary and sectoral legislation, which would then target specific sectors, or specific applications (e.g., relating to the individual's health or work situation).

2.3 Risk assessment

The White paper suggests that the assessment of risks could be based on the impact on the affected parties. This is a welcome suggestion, which could also take inspiration from the GDPR (which assesses the impacts on the data subject's fundamental rights and freedoms). In this regard it appears crucial to assess the impacts on a distributive rather than an aggregate way. This would allow to take into account specific vulnerable categories of persons. Further, the White Paper's call for the deployment of AI in the EU to produce optimal social, environmental, and economical outcomes, should lead to also take the social impacts of AI into account, which include the impact on labour rights (e.g., so-called AI sweatshops)[8], or the environmental impact of AI into account (e.g., its very important carbon footprint).[9]

3. Rights and requirements

The White Paper frames the substantive AI regulating provisions in terms of requirements.[10] Given the centrality of fundamental rights, it might be a good idea to reframe some of these requirements in terms of

[6] ibid 17.

[7] See, Art. 29 WP, 'Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679 - Adopted on 3 October 2017 As Last Revised and Adopted on 6 February 2018' (2018).

[8] https://www.theguardian.com/technology/2019/may/28/a-white-collar-sweatshop-google-assistant-contractors-allege-wage-theft, last consulted 10 June 2020.

[9] https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/, last consulted 10 June 2020.

[10] European Commission (n 5) 18 et s.

rights. For instance, the requirement to inform about the AI system's capabilities and limitations, could be reframed as the right to anticipate the type of AI/profile that may be constructed and applied to us.[11]

Finally, some fundamental rights should be adapted to the properties of AI. This is certainly the case for the right not to be discriminated against, which should allow to take into account the new proxies that AI uses a discrimination grounds and which the current legislation does not always cover (e.g., keystrokes in place of gender).

[11] See, Mireille Hildebrandt, 'Profiling: From Data to Knowledge. The Challenges of a Crucial Technology' (2006) 30 Datenschutz und Datensicherheit 548.