Comment piece

March 2020

Lack of Vision

A Comment on the EU's White Paper on Artificial Intelligence

Emre Kazim, Adriano Koshiyama
Department of Computer Science, University College London, UK
Contact: ekazim@cs.ucl.ac.uk a.koshiyama@cs.ucl.ac.uk

Abstract

In February 2020 the EU published its white paper on 'Artificial Intelligence: A European approach to excellence and trust'. This is likely to form the core of future policy and legislation within the EU and as such will have global impact on standards and norms. In this comment piece we survey the five sections of the white paper and then critically examine three themes, namely, i. regulatory signalling, ii. the risk-based approach, and, iii. the auditing styles. The key takeaway is that the white paper, and the EU's strategy at large, is ambiguous and lacks vision, which, if unchecked, is likely to have a negative impact on EU competitiveness in the development of AI solutions and services.

1. Introduction

In February 2020 the EU published its white paper on 'Artificial Intelligence: A European approach to excellence and trust'. This is likely to form the core of future policy and legislation within the EU and as such will have global impact on standards and norms. In the following we summaries each section and then critically analyse three themes, namely:

i. regulatory signalling,
ii. the risk-based approach
iii. the auditing styles.

The key takeaway is that the white paper, and the EU's strategy at large, is ambiguous, which, if unchecked, is likely to have a negative impact on EU competitiveness.

2. Section Summaries

The EU White paper has five sections. Below we have surveyed and summarised these sections by grouping sections 1-3 together and treating sections 4 and 5 separately, in the form of tables 1, 2 and 3 respectively.

2.1 Framework and Strategic Landscape (Sections 1-3)

EU political guidelines regarding AI is focused on utilising the benefits of AI through digital transformation and uptake, and mitigation of risk though appropriate legal and institutional frameworks [2]. We read the former in economic and geopolitical terms (where adoption of AI can aid in public service delivery and have significant economic impact) and the legal and ethical concerns that encapsulate the EU's assertion that 'new technologies are based on values' ([2] p. 2) and as such that EU values (ex. digital inclusion, respect for human rights, privacy, sustainability, efficiency, security) should be reflected in the development and adoption of AI within the EU. In other words, '[deriving the benefit of] AI based on European values and rules' ([1] p. 3).

The white paper also recognises that the EU must act in an appropriate manner to ensure that i. trust in the governing structures is maintained and ii. avoiding fragmentation of the single market due to a lack of a common and scaled European approach ([1] p. 2, 15). With this the two principles that the white paper pivots upon are:

- **Ecosystem of excellence**: along the entire value chain (from research and innovation), and incentive mechanisms that accelerate the adoption of AI solutions (including by small and medium-sized enterprises (SMEs)).
- **Ecosystem of trust**: compliance with EU rules and laws; trust as a policy objective in itself; legal certainty; a human-centric approach [3, 4, section 1].

The paper then surveys the strengths of the EU and areas where further funding/focus should be directed (section 2) and this is followed by the identification of key strategic areas - referred to as 'The next data wave' – (section 3).

A summary of the first three section of the paper can be found below (Table 1):

TABLE 1. FRAMEWORK AND STRATEGIC LANDSCAPE (sections 1-3)	
REPORT FRAMEWORK (section 1)	Ecosystem of excellence: along the entire value chain (from research and innovation), and incentive mechanisms that accelerate the adoption of AI solutions (including by small and medium-sized enterprises (SMEs))
	Ecosystem of trust: compliance with EU rules and laws; trust as a policy objective in itself; legal certainty; a human-centric approach [2, 3]
CAPITALISING ON STRENGTHS IN INDUSTRIAL AND PROFESSIONAL MARKETS (section 2)	EU strengths and capacities: a strong computing infrastructure (e.g. high-performance computers); holds large volumes of public and industrial data; well recognised industrial strengths in safe and secure digital systems with low-power consumption
	Targets for Investment: digital literacy; creating European data pools; expanding its position in the ecosystems and along the value chain (from hardware, to software, to services); investment levels to match North America and Asia
SEIZING THE OPPORTUNITIES AHEAD: THE NEXT DATA WAVE (section 3)	Value and re-use of data
	Data-agile economy
	Build upon Europe's lead on neuromorphic solutions (that can improve energy efficiency)
	Quantum computing
	Algorithmic foundations of AI
	Towards an explainable AI

These three sections can be thought of in terms of i. values and strategic vision, ii. landscape and capacity assessment, and iii. identifying future trends and opportunities. What is noteworthy is the skeletal nature of these sections – perhaps commissioning of a separate report on sections 2 and 3 is itself in order.

2.2 An Ecosystem of Excellence (Section 4)

Introduced in the start as a key pivot of the report 'An Ecosystem of Excellence' is fleshed out in section 4 (Table 2). Here the discussion revolves around who the key stakeholders are and in what manner they currently operate and should be structured strategically. There appears to be a tension between the centralised role and vision of the EU as a coordinator and director of agenda (for example, we can see how the EU would drive the skills agenda) and the need to have a vibrant SME community.

TABLE 2. AN ECOSYSTEM OF EXCELLENCE (Section 4)	
WORKING WITH MEMBER STATES	Delivering on its strategy on AI adopted in April 2018 [2]
	To attract over €20 billion of total investment in the EU per year in AI over the next decade
	Address societal and environmental well-being as a key principle for AI
FOCUSING THE EFFORTS OF THE RESEARCH AND INNOVATION COMMUNITY	Cohere current fragmented landscape of centres of competence
	A lighthouse centre of research, innovation and expertise that would coordinate these efforts
	Create testing and experimentation sites to support the development and subsequent deployment of novel AI applications
SKILLS	Promote a skills agenda
	Support sectorial regulators
	Updated Digital Education Action Plan
	Increase awareness of AI
	May include transforming assessment list of the ethical guidelines into an indicative 'curriculum' for developers of AI to be used in training
FOCUS ON SMEs	Access and use of AI
	Access to finance (c.f. InvestEU)
PARTNERSHIP WITH PRIVATE SECTOR	Co-investment
	Public-private partnerships
	Cooperation with top-management companies
PROMOTING THE ADOPTION OF AI BY THE PUBLIC SECTOR	Focus on healthcare and transport
SECURING ACCESS TO DATA AND COMPUTING INFRASTRUCTURES	European data strategy
	Improved access and management to data
INTERNATIONAL ASPECTS	EU is well positioned to exercise global leadership in ethical AI
	International cooperation
	Fundamental rights (human dignity, pluralism, inclusion, non-discrimination, privacy)
	Exporting of values
	Achieving Sustainable Development Goals and 2030 agenda

2.3 An Ecosystem of Trust (section 5)

'An Ecosystem of trust' was introduced as the second principle of the white paper and section 5 seeks to flesh this out by addressing the regulatory framework for AI in the EU. Seven key requirements are set out as a non-binding guideline. These are: Human agency and oversight; Technical robustness and safety; Privacy and data governance; Transparency; Diversity, non-discrimination and fairness; Societal and environmental well-being; and, Accountability.

An immediate concern is raised with respect to human oversight, where the following quote is germane: 'A key result of the feedback process is that while a number of the requirements are already reflected in existing legal or regulatory regimes, those regarding transparency, traceability and human oversight are not specifically covered under current legislation in many economic sectors' ([1] p.9). This can be considered an issue of governance. More generally, ethics is central, 'unintended effects', 'malicious purpose', 'lack of trust' ([1] p. 9) are all mentioned in the context of promoting Europe's innovation capacity.

TABLE 3. AN ECOSYSTEM OF TRUST: REGULATORY FRAMEWORK FOR AI (Section 5)		
		NOTES
PROBLEM DEFINITION	Risks for fundamental rights, including personal data and privacy protection and non-discrimination ([1] p.10)	Also includes - right to an effective judicial remedy and fair trial and consumer protection. Mitigate risks of; tracking; mass surveillance; retrace and de-anonymization of data; bias and discrimination and AI may be more efficiently propagated; opacity (black-box); unpredictability; compliance
	Risks for safety and the effective functioning of the liability regime ([1] p. 12)	Legal uncertainty; flaws in design related to availability and quality of data; difficulty in tracing back potentially problematic decisions
POSSIBLE ADJUSTMENTS TO EXISTING EU LEGISLATIVE FRAMEWORK RELATING TO AI	Effective application and enforcement of existing EU and national legislation	Transparency (opaqueness) makes it difficult to identify and prove possible breaches of laws, protection of fundamental rights, attribute liability and meet conditions to claim compensation
	Limitations of scope of existing EU legislation	General EU safety legislation applies to products but not to services, and therefore in principle not to services based on AI technology either
	Changing functionality of AI systems	Changing risk, perhaps not adequately addressed in the existing legislation which predominantly focuses on safety risks present at the time of placing on the market
	Uncertainty as regards the allocation of responsibilities between different economic operators in the supply chain	Ex. rules can become unclear if AI is added after the product is placed on the market by a party that is not the producer
	Changes to the concept of safety: risks may be present at the time of placing products on the market or arise as a result of software updates or self-learning	The EU should make full use of the tools at its disposal to assess risk
SCOPE OF A FUTURE EU REGULATORY FRAMEWORK	Regulation should be effective but not excessive such that it begets a disproportionate burden on SMEs	
	Risk based approach, that determines level of risk	High-risk is defined in terms of sectors (healthcare; transport; energy; and parts of public sector (ex. asylum, migration, social security and employment services) AND where use means that significant risk is likely to arise (risk of injury, death or significant material or immaterial damage) ([1] p. 17)

TYPES OF REQUIREMENTS	Training data	Compliance with safety rules; broad and representative data sets; avoid outcomes entailing prohibited discrimination; protection of privacy and personal data
	Data and record-keeping	Compliance; traceability; records of data sets used and their characteristics; documentation on programming, methodologies, processes and techniques, testing and validation; records should be retained for a 'limited, reasonable time period' for effective enforcement of the relevant legislation; available for testing or inspection by competent authorities
	Information provision	Transparency to promote trust and facilitate redress; system capabilities and limitations; citizens should be informed when they are interacting with AI and not humans, in objective, concise and easily understandable manners; information given should reflect the context
	Robustness and accuracy	Trustworthiness; reliable; correct reflection of level of accuracy; reproducible; requirement that systems can deal with errors/inconsistencies; resilience against attack, manipulation
	Human oversight	Mitigated undermining of autonomy; degree of human oversight is context/risk dependent; output from AI does not become effective until human validation (ex. loan rejection); post-facto human review must be possible; ability to intervene in real time; imposition of operational constraints so that AI stops working in certain conditions
	Specific requirements for remote biometric identification	Prohibition on processing of biometric data to uniquely identify a natural person; strict necessity, proportionality and legal framework followed when done so; launch of general debate to alleviate public concerns
ADDRESSEES (in relation to high-risk Ai)	Distributions of obligations among economic operators (developer, deployer, importer, private user, etc.)	Obligations should be directed to those that are best placed to address any potential risk
	Geographic scope of legislative intervention	Any and all economic operators providing AI-enabled products or services in the EU
COMPLIANCE AND ENFORCEMENT ('prior conformity assessment')	Call for an objective assessment for high-risk applications are complied with.	Verification and ensuring mandatory requirements
	This prior conformity assessment may include	Attention should be paid to evolving nature of some AI systems, and thus repeated assessments over life-cycle

	procedures for testing, inspection or certification.	
	Validation of training data and methods/techniques of AI system	
	Re-training if/when system fails conformity assessment	
VOLUNTARY LABELLING FOR NO-HIGH RISK AI APPLICATIONS (Non-high-risk applications)	A quality label, signalling trustworthiness, will be awarded according to EU benchmarks	
GOVERNANCE (in form of overarching European governance structure)	Avoid fragmentation of responsibilities	
	Increase capacity in member states	
	Progressive equipping with capacity to test and certify AI products and services	
	Functions	Exchange information and best practice; identify emerging trends; advising on standardisation activity and certification; facilitate implementation of law through issuing of guidance, options and expertise
	Guarantee maximum stakeholder participation	Social partners, businesses, researchers, civil society organisations
	Conformity assessments could be entrusted to notified bodies designed by member states	Independent testing and auditing centres

Some noteworthy themes are a need to produce a framework for trustworthy AI, and a call to study and draw upon the German five-level risk-based system of regulation that goes from no regulation for innocuous AI to complete ban on most dangerous ([1] p. 10).

3. Analysis and Criticisms

3.1 Ambiguity in Regulatory Signalling

The first and perhaps most fundamental criticism we offer is that the white paper is **ambiguous** with respect to its signalling of regulation/legislation. Indeed, some parts indicate clear openness to regulation, other parts reiterate that legislation can be updated, amended and that other regulatory frameworks/statutes won't be incurred upon etc. like the ones directly quoted below:

i. *Given how fast AI is evolving, the regulatory framework must leave room to cater for further developments. Any changes should be limited to clearly identified problems for which feasible solutions exist. ([1] p. 10)*

ii. *Member States are pointing at the current absence of a common European framework. The German Data Ethics Commission has called for a five-level risk-based system of regulation that would go from no regulation for the most innocuous AI systems to a complete ban for the most dangerous ones. ([1] p. 10)*

iii. *While in EU product safety legislation software, when is part of the final product, must comply with the relevant product safety rules, it is an open question whether* *stand-alone software is covered by EU product safety legislation, outside some sectors with explicit rules ([1] p. 14)*

iv. *While the EU legislation remains in principle fully applicable irrespective of the involvement of AI, it is important to assess whether it can be enforced adequately to address the risks that AI systems create, or whether adjustments are needed to specific legal instruments.([1] p. 13)*

v. *Given already existing structures such as in finance, pharmaceuticals, aviation, medical devices, consumer protection, data protection, the proposed governance structure should not duplicate existing functions ([1] p. 25)*

vi. *The governance structure relating to AI and the possible conformity assessments at issue here would leave the powers and responsibilities under existing EU law of the relevant competent authorities in specific sectors or on specific issues (finance, pharmaceuticals, aviation, medical devices, consumer protection, data protection, etc.) unaffected ([1] p.25)*

Such ambiguity is likely to have impacts on planning, risk evaluation, innovation and investment.

3.2 Risk Approach is Challenging

The call for an approach that determined the level of risk introduces two distinct notions of risk, both of which are challenging.

i. The first notion of risk is with respect to **sectors**, where high-risk is identified with respect to things such as healthcare, transport, energy, and, parts of public sector (ex. asylum, migration, social security and employment services).

We note that all these sectors have the commonality of human impact i.e. whether a service, instruction, decision, etc. impacts on a human user and citizen. We believe that this is a broad, abstracted and blanketed approach, that is highly likely to result in two things, i. **risk aversion**, and ii. **AI will be a high cost venture**. For example, a simple healthcare booking chat bot can become economically unfeasible to develop because it falls under health. Similarly, in the context of high-risk high-reward a risk-based approach based upon sector will **discourage potentially high-positive impact AI systems** (ex. medical AI has significant risk and lifesaving potential). As such we believe this will stifle innovation (which is what the EU white calls for).

ii. The second notion of risk introduced is that ‘where use means that significant risk is likely to arise (risk of injury, death or significant material or immaterial damage)’ ([1] p. 17).

We take issue with this categorisation of risk principally because it is unclear how unintended consequences can be assessed; indeed, how can organisations be liable in such circumstances.

In our own work on AI Impact Assessment [4] we envision **a declaration of interest as part of the deployment of a new algorithm**. This declaration will disclose qualitative information about the team that have built the system, as well as few scenarios that can happen if bad actors or the algorithm is used inappropriately. Hence, this can enable the algorithm’s designer to perform trade-off and worst-case analyses of Fairness, Robustness and Explainability. Indeed, this declaration and the trade-off analyses will also support which route the algorithm’s designer would take: a by Design or a post Assessment of their algorithm. By following this approach, risk-analysis is explicitly accounted for and built-in to the process and thinking of AI System design and deployment.

3.3 Auditing: Accounting not Process Based

The white paper outlines ‘types of requirements’ ([1] p. 20) and notes the following:

1. Data and record-keeping – verification of compliance;
2. Traceability;
3. Records of data sets used and their characteristics;

4. Documentation on programming, methodologies, processes and techniques, testing and validation;
5. Records should be retained for a 'limited, reasonable time period' for effective enforcement of the relevant legislation;
6. Available for testing or inspection by competent authorities.

We can surmise that points 1-3 about data, 4 is about the AI system and 5-6 can be construed as relating to compliance.

We note the similarity of this call to other publications that discuss issues of governance – we also note that in recent publication 'impact assessments' have been suggested [5, 6]. It is noteworthy that the EU's call in this space is not as advanced as sketching the outline of how points 1-6 could be integrated into an overarching governance structure and engineering practice [7].

We also note that the above would practically entail coherent and clear record keeping; this is not a process-based approach, where at each step economic and ethical impact could be assessed (leading to development and design integration *in situ*). This also can be related to the point raised above in section 3.2 regarding 'declaration of interest'. To be clear, we see EU paper as forwarding an accounting approach, rather than a governance and real-time intervention, which we believe is a better way to move forward. As a corollary, our real-time suggestion is likely to engender a culture of consciousness with respect to design ethics and impact, whereas the above approach is more a compliance culture.

4. Summary

The EU white paper 'Artificial Intelligence: A European approach to excellence and trust' is likely to form the core of future policy and legislation within the EU and as such will have global impact on standards and norms. It sets its own standards, namely creating an ecosystem of trust and excellence – we have therefore surveyed and critically committed upon the white paper with these standards and aims in mind. We can conclude that the EU the key takeaway is that the EU's strategy is ambiguous, which, if unchecked, is likely to have a negative impact of EU competitiveness and, fail to delivery 'an ecosystem of excellence'.

5. References

[1] 'White Paper on Artificial Intelligence: A European approach to excellence and trust', *European Commission* (February 2020)
[2] 'Communication on Building Trust in Human-Centric AI' COM (2019) 168, *European Commission*
[3] 'Ethics Guidelines High-Level Expert Group on AI', https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (April, 2019), accessed March 15th 2020
[4] Koshiyama, A; Engin, Z, 'Algorithmic Impact Assessment: Fairness, Robustness and Explainability in Automated Decision-Making', , *Zenodo* (June 2019)
[5] Kazim, Emre and Soares Koshiyama, Adriano, No AI Regulator: An Analysis of Artificial Intelligence and Public Standards Report (UK Government) (February 26, 2020). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3544871
[6] Kazim, E, Koshiyama, A, 'Impact Assessment Needed: An Analysis of Data Analytics and Algorithms in Policing (RUSI Report)' (March 4, 2020). Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3548552
[7] Calvo, R.A., Peters, D. & Cave, S, 'Advancing impact assessment for intelligent systems', *Nat Mach Intell* 2, 89–91 (2020)