Ref. Ares(2020)3359883 - 26/06/2020

Feedback on the EC's white paper on Artificial Intelligence

OdiseIA

The European Commission published its long-awaited white paper on Artificial Intelligence on February 19, 2020 and has an open consultation for feedback until June 14, 2020. The white paper centers around two main pillars: excellence and trust, but given the earlier statements of von der Leyen about launching an AI regulation during her first 100 days, expectations were mostly focused on the approach for regulation (trust).

Much feedback has already been published by many different organizations. In this note, we briefly summarize the main messages of the white paper, express our support for some of the feedback published by others and provide our additional feedback that we haven't seen elsewhere.

The white paper

The white paper focuses on **Excellence** so that Europe becomes more competitive in the global AI ecosystem, and on Trust so that AI will be adopted at a large scale without causing negative impacts. We agree that Excellence is important and endorse the proposed actions to increase investments, skills and retention of world-class researchers and engineers. We also endorse the creation of a European center of excellence. However, we recommend the European Commission to let excellence prevail over politics, and to avoid the typical fragmentation. We need member states to think "Europe", and don't need another European Institute of Technology (EIT) that alludes to the world-renowned MIT, but due to its fragmentation is not even a shadow of it. We also applaud the focus on technology transfer from research to market, something where Europe has systematically failed in the past two decades. The Commission is creating local ecosystems called Digital Innovation Hubs that aim to bring AI technology to SMEs and the local market. We give the EC the benefit of the doubt and hope that this approach will make a difference compared to the current situation.

Regarding **trust**, the focus is on the avoidance of the potential negative impacts of AI, in particular related to autonomous decisions such as liability, explainability and transparency, bias and discrimination, and privacy. While not explicitly mentioned, the focus is on the risks of "narrow AI" powered mostly by deep learning and other machine learning techniques. The Commission aims to increase trust by distinguishing between high-risk and low-risk AI applications, and proposes regulation in case of high risk. An AI application is considered high risk if it belongs to a high-risk sector and at the same time concerns a high-risk use case. The Commission proposes to define an exhaustive list of high-risk sectors and use cases with periodical review. Examples of high-risk sectors would be health, transport, police, and legal. A typical example of a high-risk use case would be facial recognition.

Regulation reduces legal uncertainty, and this is important for companies to invest in AI technologies.

Common feedback on the AI white paper

Many organizations already have published their feedback. Here we will summarize the points which we agree with.

Definition of AI too broad

The white paper states that for defining AI the concepts *data* and *algorithms* are essential. However, any computer program uses data and an algorithm, and therefore such a definition is too broad as it classifies almost any computer program as AI. We recommend that the EC maintains the AI definition provided by the HLEG on AI published in April 2019. Moreover, while the white paper states that potential regulation should not only apply to AI of today, but also to that of the future, it seems mostly applicable to "narrow AI" based on only advanced Machine Learning.

Definition of high-risk

By classifying all AI applications into high risk or low risk, where regulation would apply only to high-risk, it is important to make explicit what defines a sector and use case as high-risk. Companies have a natural preference not to be regulated and will try to prove (lobby) their sector not to be of high risk. As yet, it is unclear who would decide on this. What seems to be needed at least is a transparent definition of what defines high risk. Others say that two levels of risk is too coarse-grained and prefer the 5-level approach of the German Data Ethics Commission.

Specific AI regulation versus existing regulations

There are already many regulations in place that closely relate to the motivations for a potential specific AI regulation. Privacy, use of personal data and algorithmic transparency are regulated under the GDPR. Liability of products falls under the PLD. Safety of products is regulated under the GPSD. Consumer rights are protected in the CRD. And discrimination based on sensitive attributes is already forbidden by law. The intuition behind a potential regulation is that, with AI, enforcing and monitoring those existing regulation is becoming impossible. For example, what happens with liability if products continue to "learn" new behavior once they have been released in the market? Or, how can an organization be sure that it is not discriminating against race when using algorithmic decisions for e.g. credit scoring? The feedback is that the EC should carefully investigate where existing regulation is insufficient and avoid any overlap as this would increase legal uncertainty. The Commission should also investigate the alternative of extending existing regulations to include artificial intelligence. We also advocate that the EC uses the instrument of regulatory sandboxes to experience with various options, and based on the outcomes, take evidence-based decisions. Such approach would also allow to decide better on what needs to be regulated ex ante and ex post, which is especially relevant for AI products produced outside the European Union.

OdiseIA's feedback

The "GDPR" for Artificial Intelligence?

Many European policymakers claim that the GDPR is now considered a world-wide standard, and that the same ambition is appropriate for a future AI regulation. However, we should not forget that the GDPR came into force when the use of personal data was already pervasive, and much was understood about the use (and abuse) of personal data, and it was clear what needed to be regulated. On the contrary, AI is not yet so widespread (in contrast to what one might think when reading the press …) While extensively used by the Tech Giant, still many organizations are struggling to make effective use of AI. It is therefore much less clear what and how to regulate. One could argue that if the GDPR had come into force several years earlier, the opaque use of personal data would not have been so common, and that therefore AI regulation is needed now. But we believe that regulation of a technology that is still in fast development has its risks. Again, we advocate the use of regulatory sandboxes to explore alternatives and take data-driven decisions.

Is it regulating the right topics?

Looking at the majority of the examples in the white paper and at the requirements on high-risk AI (training data, data and record keeping, robustness and accuracy, human oversight), it looks like AI regulation is mainly aimed at *autonomous decision making*, leading to typical problems such as black box algorithms (explainability & transparency) and bias (undesired discrimination). AI is not perfect and makes errors (false positives and false negatives), hence the importance of robustness and accuracy. Maybe we should come up with criteria for when autonomous decisions require HITL, HOTL or HOOTL. What types of decisions require a human IN the loop (HITL), approving every decision a machine makes? What types of decisions require a human ON the loop (HOTL), monitoring the AI outcome and correcting when wrong? And what types of decisions can be taken without any human intervention (H Out Of TL)?

Explainability and bias are among the most popular research areas in AI, receiving a huge amount of attention. So, in the coming five years we may expect significant advances. Maybe any deep learning algorithm (or whatever new AI paradigm will appear) can be made transparent in the future. And, probably, in the coming years, bias can be detected and corrected in any data set, even when they do not contain sensitive attributes. A regulation that focuses largely on transparency and bias might become obsolete. Should regulation, therefore, include other significant implications of AI, such as the future of work (or the work of the future), the relation between people and AI systems, or avoiding the ignorance of powerful AI technology for solving big societal problems such as the SDGs?

Make a reality check of current autonomous systems

The white paper speaks about high risk in the context of: "from the viewpoint of protection of safety, consumer rights and fundamental rights" and "that produce legal or similarly significant effects for the rights of an individual or a company; that pose risk of injury, death or significant material or immaterial damage; that produce effects that cannot reasonably be avoided by individuals or legal entities." Let's make a reality check and have a look at

algorithmic trading. Estimations are that about 70% of all stock market transactions are performed by algorithms, some of which are opaque and only (barely) understood by quants. Several "flash crashes" have occurred because of such black box trading algorithms, where the stock market falls sharply in a very short amount of time. Such flash crashes can have huge impact on many different groups of people, from rich to poor (e.g. through social organizations that invest).

Therefore, we recommend the Commission makes a thorough assessment of what currently operating AI systems would need to stop if regulated, and what would be the impact on societies and economies.

An Accountability Scorecard

In case the Commission decides to introduce AI regulation, it becomes important to define how organizations will be audited (ex-ante or ex-post) for compliance. The white paper, however, only mentions "audit" once, without any further elaboration. Such auditing process would need to adequately cover the possible regulation (robustness, bias, transparency, liability, etc.) in a flexible framework that allows externalization to designated, competent auditing organizations. The objective would be to come up with a scorecard with relevant metrics which evaluate the trustworthiness of individual AI products.

Conclusion

In conclusion, we welcome the AI white paper, especially because it is generating an essential debate on the future of AI (regulation). And we are confident that the Commission will come up with an improved version based on the public consultation. However, we should not forget that even if the use of AI implies significant risks, it could still be better or generate less risk than when only humans are making decisions as it happens today. And we end with an open question: would the world be a worse or better place today if AI had been available 120 years ago?