Response to the EU Commission White Paper

"On Artificial Intelligence – A European approach to excellence and trust "

Authors and disclaimer

Anny Pinto has been working on data protection, privacy and IT law for over 16 years. She is the Chief Privacy Officer and & Legal Head Group IT at the Adecco Group. Siddhartha Singh is a machine learning engineer at Auterion AG.

This document expresses the authors' sole opinions and by no means should be regarded as their companies' views. Should you have any queries on this document, the authors can be reached out on annypz@hotmail.com and siddhartha.singh@protonmail.ch.

Introduction

Over the last few years interest in artificial intelligence (AI) has had an exponential surge by academics, private and public sectors. This has been mostly propelled by new machine learning techniques combined with cheapening of computational resources and availability of large amounts of data. This sudden growth in interest towards AI has also forced policymakers around the world to act fast because they have also realised the potential harms such technology is capable of. The White Paper *"A White Paper on Artificial Intelligence A European Approach to Excellence and Trust"* by the European Commission is a laudable attempt in that direction.

We, however, feel that the White Paper fails to address certain notions and ideas in machine learning accurately. In this document, we would like to flag these points and recommend amendments that, we believe, would strengthen the overall vision of the White Paper in the future of the AI technology and its regulation.

There can be AI without data

Contrary to what the White Paper sets forth (*"without data, there is no AI")*, there can be AI without data. Modern systems which mostly involve techniques like deep learning are based on data. However, this is not a necessary condition to achieve AI agents. AI agents can be based on abstract algorithms that are not based on data. The White Paper self alludes to this fact by referring to symbolic systems that can either be combined with data-based machine learning models or can singularly also be used to create AI agents.

Artificial Intelligence happens to be a term with a lot of historical baggage and the current understanding of what AI means encompasses many different fields like machine learning, reinforcement learning, symbolic systems and leaves out others like *artificial general intelligence*. We feel that defining *Artificial Intelligence* correctly is important to create focus in the vision for its future in the EU. In this regard, we recommend the following definition: Designing agents capable of perceiving their environment and take action that maximizes their chance of success.

Risk based approach throughout the entire AI life cycle

Whereas we welcome the risk- based approach, we believe that defining ex-ante what high risk means and the prior conformity assessment are too restrictive and not adapted to the rapidly evolving context of AI.

Firstly, we believe that the risk assessment should be performed throughout the entire AI lifecycle which means that AI should be continuously assessed and adjusted. An AI technology which may be initially regarded as low risk or no risky at all for the individuals, may become high risk throughout the deployment and further use. An example that illustrates this idea is the recommendation engine in Youtube or Netflix that may be used to identify political beliefs, sexual orientation and other personal and sensitive attributes of the consumers which can potentially cause serious harm to their fundamental rights. Notwithstanding, it may happen that an initial high-risk AI technology may be categorised as low risk in the latter use by consumers. For example, when self-driving cars are ubiquitous enough that AI driven cars starts decreasing the overall mortality rate on the roads, the risk assessment must follow the same trajectory.

Considering the above, an ex-ante determination of high risk as the paper sets forth (i.e. i) a specified sector in combination with the significant impact of the AI, or ii) pre-determined, by default high risk AI applications) is too prescriptive.

Moreover, we believe that the determination of high risk should be performed through impact assessments as GDPR contemplates for high risk data processing activities. Companies should be encouraged to have a flexible accountability approach where they consider the overall context and concrete impact of the AI technology on the individuals and society in general, and the companies determine the risks and the respective mitigation measures, with recurrent and interactive AI compliance checks throughout all AI lifecycle.

Validation and Test Data

The White Paper is correct in mentioning that training data should be the focus when thinking about regulations on AI. However, so does *validation* and *test* data sets. We can – very broadly – think of the three data sets *i.e. training, validation, test* data sets as the following. Training set is the data from which the model *learns*, validation set is the data on which the model is *tuned*, and test set is the data on which the model is (finally) *evaluated*. More often than not, a given large data set is divided into the three above mentioned sets (with some constraints). However, that may not always be the case. For example, there may be situations where the data sets used to solve a given business case is unlabeled or partially labeled. In these circumstances, it is not uncommon to use a training set that does not belong to the data set collected for solving the business case. We could potentially use an open data set for training and use the partially labeled (collected) data for validation and testing only.

The final performance of the trained models, therefore, will depend on the *validation* and *test* sets, even if these two data sets do not come from the same data set as the training data. Whether or not the models have certain underlying biases e.g. gender, race etc. also depends on these validation and test data sets. We, therefore, see that the regulations proposed on the training set should also apply to both validation and test sets, particularly when they do not belong to the same data sets.

Metric Specification

An important aspect of any machine learning models is the metric on which they are evaluated. In colloquial language, “accuracy” and “precision” are generally used as proxy to success and trustworthiness of these

models. However, *accuracy* has a specific definition in the machine learning field. Also, accuracy and precision have different meanings when training models.

The choice of this metric has a big impact on tuning certain aspects of the learned model. Not all metrics are made equal and different metric choices will inform an engineer differently about the model. Very often entirely new metrics may be created by the machine learning engineers in order to evaluate the model. There is a subjectivity to this selection. Therefore, success on a metric satisfactory to the business case does not guarantee that underlying biases and fairness of the model have been properly tested. We, therefore, recommend including in the EU Commission regulatory vision a mandate of transparency on the metrics that the models were evaluated upon. This, for example, could be in form of a *specification sheet* that every model may come with.

Setting standards thorough open-source software

Another element of the current machine learning eco-system that we feel the White Paper does not address is the role of the open-source software in the rapid evolution of the field. Apart from (and generally known) data and computing power, open source software has been an enormous boom to the machine learning community. Open source has not only propelled the field forward with a break neck speed but has also been important in establishing *standards* for this community. This is the main reason why most large technology companies like Google and Facebook either open source or contribute to other open-source machine learning software.

As of today, most companies and almost all startups in the AI space are dependent on these open-source frameworks from Google and Facebook. Although the software is open-source, these companies tightly control development and the direction the software takes, hence, controlling the standards as well. This not just includes standards established by core frameworks used to create modern machine learning algorithms like Tensorflow and Pytorch, but also standards in frameworks used in machine learning fairness, privacy and interpretability.

If the EU Commission truly wants to realise its vision of incorporating European values in the future AI technologies, it must encourage and help grow this open source community within the EU. Hence, we call upon the EU Commission to add this aspect of AI zeitgeist into the future AI regulation.

AI Literacy to be encouraged

As citizens with deep understanding on data protection and machine learning given our professions, we recognize that there is an urgent need for the society in general and the EU citizens in particular, to understand what AI truly means and how it may impact positively or negatively in their own life and fundamental rights. Only when knowledge on digital and concretely on AI is well-spread from primary schools to colleges and universities, the individuals will consciously exercise their rights and enforce *vis a vis* companies and authorities and the so-desired trustworthy AI will be completely materialized.

Additional Remark

A final point that we would like to bring forth is that even if all the recommendations in the White Paper, including our recommendations, are followed the negative consequences of certain types of AI (in the way it is done at the moment) cannot always be mitigated.

Today, most machine learning algorithms are used to capture the truly limited human resource – *attention*. The transformation into attention economy has meant that the algorithms are designed to addict and not to serve citizens. Technology companies try to capture attention in many different ways, some through the use of AI technology (e.g. News curation on Facebook Wall), and some through other simpler design choices (e.g. Snapchat Streaks). Use of AI algorithms for attention grabbing, however, has an element of adaptation built into the system. The AI can constantly adapt to ones changing tastes and even help nudge the person to acquire new ones. This is done solely to maximize attention capture. It is therefore important that AI regulatory vision that prioritises serving EU citizens take these facts into account and include measures that aim to mitigate the intended or unintended harm caused by such use of modern artificial intelligence technologies.

By Anny Pinto and Siddhartha Singh

19 May 2020