Maxime Gennart*

Privacy, Ethics and Collaboration: the Roles of DPAs in AI development**

First of all, I would like to thank the personal data protection service of Georgia for this very interesting three days and the warm welcome in Batumi. I also would like to thank them for giving me the opportunity of presenting with such a distinguished panel of speakers on the topic of Privacy, Ethics, and Collaboration: the roles of DPAs in AI development.

For those of you whom I have not had the honor to meet, my name is Maxime Gennart, I am a legal advisor at the Belgian data protection authority and a member of its AI task force. In this role, I am involved in setting up the Belgian national framework for AI governance, and it is on this topic I wanted to talk to you today.

Keywords: *Privacy, Ethics, Data Protection Authorities, AI, Privacy-by-Design, Ethics-by-design.*

1. Introduction

This article explores how Data Protection Authorities (DPAs) can guide the development of AI in a way that respects both privacy and broader human rights. To do so, it first illustrates the type of interdisciplinary thinking that AI development will require, using a real-world use case. It then explains how the concept of privacy-by-design provides a valuable tool for the responsible development of technologies and why its expansion into a broader notion of ethics-by-design should be considered. Building on this, the article highlights how DPAs' experience in implementing privacy-by-design is crucial for advancing a framework such as ethics-by-design. Finally, it examines how certain provisions of the AI Act could foster private–public cooperation in the development of AI, thereby supporting the case for collaborative pre-market value assignments.

* Legal Advisor at the Belgian Data Protection Authority.

** The paper is the text of a keynote speech presented at the 33rd European Conference of Personal Data Protection Authorities ("Spring Conference"), hosted by the Personal Data Protection Service and held in Batumi. The information and views set out in this article are those of the author and do not necessarily reflect the official opinion of the European Commission.

2. Interdisciplinary Thinking & Cross-Sectoral Issues of AI

To exemplify the type of interdisciplinary thinking behind AI systems and the cross-sectoral issues they pose, I decided to talk briefly about Amazon's AI recruiting tool from 2018.

That experimental AI recruiting tool used a five-star rating system to score the likelihood of individuals getting software and technical job roles at Amazon.

It was trained for years using extensive datasets containing job applications, CV, cover letter, etc… The goal of the AI tool was to select the best candidates for technical and software jobs at Amazon.

To attain that goal, the tool spotted similarities across participant and sorted out the best ones. Now, it is a well-known fact that men are over-represented in software and technical jobs and the AI system was trained based on that fact.

Because of this, the tool began to skew reasoning towards this preference and quickly started showing signs of sexism, by lowering scores for resumes from women and steering preferences towards male candidates. In the end, and partly because of this discriminatory patterns, Amazon pulled this AI tool back.

Based on this example, we can already see how AI will affect a number of sectors and require the involvement of various disciplines. In this case, we can already see issues related to labor law, privacy & data protection, discrimination law, and the involvement of technical experts to adapt the tool and correct such biases.

3. Privacy-by-Design: The Collective Responsibility of Fundamental Rights

This example also shows that AI systems will affect individuals' life by having an impact on broader societal values that have to be implemented by technical experts.

This is not the first time we are seeing issues like this. We observed a similar fact with the rise of predictive technologies a few years ago. In fact, in the years following the Snowden revelations, the public started to become aware of the ways their personal data were being used online and how their fundamental right to data protection was being affected through the design of technologies.

This public awareness led to the realization that privacy and data protection could not solely rely anymore on individuals' decisions over their personal data. This led to a shift in the mind of people that personal data protection, although about individuals, became a collective responsibility that needed to be thought about holistically and implemented by technical profiles at the design phase. In addition, here came the concept privacy-by-design.

This shift was critical. It expressed the idea that compliance to common values, such as privacy and data protection, could be achieved or promoted at the design phase of a product more effectively than on an ex post basis through enforcement and corrective measures.

The rationale for privacy-by-design is thus both practical and philosophical. It rests on the understanding that the architecture of information systems can act as a form of regulation in itself.

4. Ethics-by-design: Embedding Universal Values at the Start

Today, this rationale remains entirely relevant. As exemplified with Amazon's AI recruiting tool, AI systems pose challenges to fundamental values that needs to be tackled across sectors.

However, privacy-by-design, as its name indicates, is about the safeguard of privacy, one fundamental rights among so many others. Privacy-by-design is thus in itself not sufficient anymore. That is where ethics-by-design could come in.

Ethics-by-design is about assigning values into the AI system's design that would guide it throughout its decision-making processes.

Given the breath of these challenges and the number of fundamental rights potentially affected, it may seem only logical to try and attribute significance or even moral worth to the decisions taken by AI systems.

While this is relatively easy on paper, how to technically instruct an AI system to reach the goal of presenting the best applicant while keeping in mind that this shouldn't be carried out based on gender is a complex and intricate task that would require the input of individuals from various backgrounds.

Practically, this shift would require us to shift from a reactive type of regulation to a more proactive one. Such as privacy-by-design require us to think about privacy when deciding the means and the way of a processing operation, ethics-by-design would require us to think about the values that will serve as guiding line for AI systems' decisions.

5. DPA's Experience to the Rescue

How can DPAs' experience implementing the concept of privacy-by-design be of help with the development and implementation of the ethics-by-design concept?

Well, as data protection authorities, we have a wealth of experience conducting complex balancing exercises. We assess the necessity and proportionality of processing operations in a wide range of contexts by weighing data protection rights against freedom of expression, freedom of information, public interest, and more.

The balancing exercises that we carry out often requires us to combine interdisciplinary knowledge. We ask for the input of technical experts to fully comprehend the implications of a specific processing operation. We then come with our own legal expertise to identify the obligations applicable to the processing at hand. Thanks to this cooperation, we understand whether privacy was truly thought of at

the design phase of the processing operation or as an afterthought following an investigation or a complaint.

These skills makes us uniquely positioned to:

- Identify risks posed by AI systems,
- Think rationally and ethically in balancing rights, values and interests at stake, and
- Find a legitimate compromise between conflicting values.

In the regulation of AI systems and the potential shift from privacy by design to ethics-by-design, we, DPAs, therefore have a necessary experience to gather people with relevant knowledge to understand the intricacies of a specific technological environment and try to implement/assign a value that would guide AI systems throughout their lifecycle.

Now, how could we practically start to enable such a shift and how could that shift look like?

To answer that question, I will elaborate on five situations where we can see the shift appearing. Three are post-market monitoring practices, two others happen before entering the market.

6. From Post-Market Cooperation Mechanism

First, I believe the FRIA[1] is a first ex-ante assessment of high-risk AI Systems and is invaluable. It prompts designers to evaluate ethical trade-offs early in the design phase, not as an afterthought. There is therefore an opportunity to include an ethics-by-design thinking into AI development. The issue, if I may, is similar to DPIA, is that it is entirely carried out by private actors developing the assessed technology.

Second, we can look at the work carried out by international organization such as the EU, the OECD, UNESCO, etc. These initiatives are a great example of cooperation because they usually include the input of both the private and the public sector to work towards the translation of fundamental rights into values that AI systems could be asked to consider. However, these standards are not enforceable and their implementation is entirely left to the discretion of private actors.

You can already see here that the point I am trying to make is that there is still this huge gap between the private actors, technical experts and designers of technologies, and public authorities, experts in enforcing and implementing fundamental rights. In addition, here comes the AI Act.

Its article 79(2) already starts to close this gap. According to this article, Market surveillance authorities (MSA) which identify a risk to fundamental right(s), have to notify the relevant authority protecting that fundamental right. Together with the operator of the concerned AI system, they have to cooperate in remediating that risk.

[1] Disclaimer: the FRIA is here included under the post-market monitoring because it is written from a Litigation perspective. I am aware that this is an internal ex ante assessment carried out by private actors. However, as DPAs, they only appear when a complaint has been filed.

In this scenario, there is an obligation of cooperation between the private sector, MSA and DPAs to adapt the design of an AI system and ensure it respects the fundamental right to privacy and data protection. However, here we can say that the intervention arrives a bit late, as a risk has been identified and will potentially have materialized.

However, given the characteristics of AI systems and the real-world implications they already have, such cooperation mechanism involving ethical tradeoffs and value assignment exercises should be fostered to take place at the design phase and with the involvement of both the developer of the AI system and the authorities protecting fundamental rights.

To show you that this pre-market value assignment cooperation mechanism is feasible in practice, I wanted to draw your attention on two provisions of the AI Act.

7. To Pre-Market Value Assignment Cooperation

The first disposition relates to the regulatory sandboxes of art. 57 AIA. The article explains that member states should have at least one regulatory sandbox. The aim of these sandboxes is to identify possible risks, in particular to fundamental rights. A derived aim of these sandboxes is to promote innovation that adhere and respects fundamental rights, including the one to data protection.

Now what is also interesting is that, in its paragraph 4, it opens up the possibility for the authority responsible to operate the national regulatory sandbox to cooperate with other authorities in testing the AI system. From this disposition, we thus have an opening to operate some kind of pre-market monitoring of high-risk AI systems. Indeed, we, as public authorities protecting fundamental rights, could be asked to participate in the testing of high-risk AI systems. This would pave the way towards a cooperation between public and private actors involved in the testing of an AI system to adapt or modify it for it to respect fundamental rights, before its placing on the market.

Besides testing in controlled environment like regulatory sandboxes, the AI Act also offers the possibility to test high-risk AI systems in real-world conditions outside of regulatory sandboxes. This, of course has to be done, under a number of strict conditions among which the submission of a testing plan by the AI system operator to the national competent authority and the transmission of the final outcome of that real-world test to that same authority.

If that competent authority finds it necessary, it can monitor that testing by, among other, carry out onsite or remote inspection during the testing.

The national competent authority therefore has the possibility to either analyses the AI system live when it is being tested or analyses it after the test based on the final outcome transmitted by the operator.

Now, imagine that authority identifies a risk to the fundamental right to data protection, wouldn’t that trigger the notification and remediation mechanism of art. 79(2)? Hence, wouldn’t that trigger another type of pre-market cooperation

mechanism involving authorities protecting fundamental rights, market surveillance authorities and private actors?

The opportunities indeed seems to be present.

8. Conclusion

I would like to conclude by saying that although the challenges and risks posed by AI are numerous, similar concerned were raised with the rise of predictive technologies at a time where data protection was being strengthened.

The experience we, as DPA, have acquired in addressing intricate questions about fundamental values in complex technological environment is crucial to start thinking about the development of a pre-market ethics-by-design approaches to AI development and innovation.