Der Bundesbeauftragte
für den Datenschutz und
die Informationsfreiheit

Keynote

of The Federal Commissioner for Data Protection and Freedom of Information

Prof. Ulrich Kelber

„Artificial Intelligence and AI Act“

at the Data Privacy Day 2024

Helsinki
26 January 2024 10.45 – 11.30

The spoken word shall prevail.

Dear Ms. Fornaciari,

Ladies and Gentlemen,

thank you very much for the invitation.

“Data protection and Artificial Intelligence (AI) are two crucial aspects that need to be considered hand-in-hand. AI algorithms heavily rely on collecting and processing vast amounts of data to provide accurate insights and make intelligent decisions. However, ensuring data protection and privacy is of paramount importance to maintain user trust and ethical practices.”

I could not have said this any better myself, but the previous statement was in fact generated by an AI system, highlighting the improvements that have been made in the field of AI in recent years.

I. Introduction

The rapid development and deployment of increasingly powerful AI systems poses a challenge to regulators. While they desire to promote innovation to reap the potential benefits of this technology, they also strive to address the potential risks and protect fundamental rights and safety.

Existing regulation that protects fundamental rights and safety already applies to AI. For example, the data protection principles in the GDPR are applicable in a technology-neutral manner, in particular to AI.

However, there are some AI-specific challenges. A lack of explainability and therefore transparency, but also the problem of deletion from AI models, to name just a few. Given such AI-specific challenges, additional principles and specific requirements for AI systems are needed to ensure and support compliance with existing principles.

Additionally, a predictable and reliable environment for the development and application of human-centered AI fosters innovation. This requires legal certainty as well.

That is why I am pleased that the EU institutions were able to reach a compromise on the AI Act[1].

[1] https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai

II. Generative AI Systems

A particularly controversial aspect were generative AI systems, like the one that generated my initial statement. Initially, generative AI systems, such as ChatGPT, were not covered by the commission's proposal. However, the disruptive character of such systems led the parliament to include certain requirements for those as well. Hence, the widespread public interest in ChatGPT sparked intense discussions concerning the AI Act.

Eventually, transparency requirements for certain general-purpose AI models and systems, for example ChatGPT, were included in the AI Act. These are also necessary for data subjects to exercise their data protection rights effectively. If it is unknown, which data was used to train an AI model, access is not possible, deletion requests are futile and inaccurate data cannot be corrected. I thus welcome transparency requirements for general-purpose AI.

In case of "systemic risks" that could stem from general-purpose AI, the AI Act specifies additional requirements for the respective models and systems to mitigate those risks.

Companies using general-purpose AI, in particular generative AI systems, should apply further measures, even if ultimately not required in the AI Act. For example they need to take into consideration that the large language models, at the core of the generative AI systems they employ might have been trained with user data.

In order to comply with the GDPR, such generative AI systems must not be fed with arbitrary personal data. The integration of generative AI systems into everyday tools, such as text processing programs or search engines, lowers the barrier to interact with them like with any other software. Thus, employers are obliged to train and sensitise their employees accordingly. The same caution should be applied with respect to company internal data. Otherwise, confidential data might be extractable from some AI system.

III. Ambivalence of AI

While generative AI receives the most public attention, it is not the only type of AI. Notably, AI has enabled advances in many scientific fields, for example by predicting the three-dimensional structures of proteins. The technology also promises progress in the medical field, for instance by analysing medical images and detecting illnesses.

In many ways, AI has the potential to make our lives easier. Be it in medical diagnostics, automated administrative decisions, intelligent traffic guidance systems or search engines. AI systems can do tedious routine work, and thus enable us humans to concentrate on more creative tasks that bring more value to our lives.

However, there are less meaningful and valuable applications of AI as well, some present obvious risks to personal rights. The controversial facial recognition technology “Clearview AI” being one of them. The company gathered billions of facial images, by extracting photos from social media platforms and websites. Those are then used to enable customers to identify individuals. One problem there is that users, whose private photos were collected, are unaware that they are used in this way.

Another example that comes to mind is predictive policing. There, plenty of data are analysed to try to predict when, where or by whom future crimes might be committed. This can and in fact often does place entire communities under general suspicion and harm the rehabilitation of offenders, if they continue to be accused of criminal intentions.

In general, the expectation that AI is more objective than humans is misguided. AI models are often trained on vast amounts of data, which contain human bias and existing prejudices. Those are then transferred to and reproduced by the respective AI models. This can lead to a vicious cycle where bias is reinforced, solidified or enhanced in an AI model, as AI technologies are highly capable at detecting patterns in data.

IV. Risk-based Approach

Not all possible application scenarios for AI entail the same level of risk. Intelligent traffic guidance systems are less critical than self-driving cars. AI systems that recommend potential partners in a dating app have less impact on people's lives than AI systems that decide which job applicant is hired or whether someone is going to jail.

The AI Act follows a risk-based approach. AI systems that do not pose a risk for the health, safety of fundamental rights of people face no additional obligations under this regulation, whereas applications that are associated with a high risk are required to meet certain quality requirements.

Both in the area of data protection and in the development and use of AI systems, regulators face a very heterogeneous field of stakeholders. My experience with the application of the risk-based approach in the GDPR shows that in such a rapidly changing and heterogeneous environment such an approach can lead to positive results in terms of practically achievable regulatory requirements. I strongly support this regulatory approach as it offers a good balance between innovation-friendliness and a high level of proactive protection of fundamental rights.

V. Prohibited AI Applications

Some AI systems and applications entail an unacceptable risk and are incompatible with our European values. The AI Act strives to prohibit such systems and applications.

The untargeted scraping of facial images from the internet to create facial recognition databases will be prohibited. This means that the approach of Clearview AI I mentioned before will likely not be lawful under the AI Act. Regardless of this, the approach is already highly problematic under the GDPR. However, I welcome the explicit clarification of this matter.

Less pleasant are the exceptions for real-time remote biometric identification that are included in the AI Act. Remote biometric identification of individuals in publicly accessible spaces poses a high-risk of intrusion into their private lives. This would have severe effects on the population’s expectation of being anonymous in public spaces. For these reasons, the EU data protection authorities called for a general ban on any use of AI for an automated recognition of human features in publicly accessible space in any context. I myself advocated for such a ban relentlessly as well. Unfortunately, this is not included in the AI Act, despite the parliament’s demands. Regrettably, the compromise contains plenty of exceptions for the application of this technology.

At least the co-legislators agreed to prohibit biometric categorisation systems that use sensitive characteristics, like political or religious beliefs, sexual orientation, and race[2]. The EDPB and the EDPS had already recommended a ban on such AI systems in a joint opinion in May 2021. This ban aims to mitigate the risk to cluster individuals according to ethnicity, gender, as well as political or sexual orientation, or other grounds for discrimination under Article 21 of the Charter.

Given the great risk of discrimination, the AI Act also prohibits social scoring based on social behaviour or personal characteristics. I thoroughly welcome a complete ban on social scoring, be it by public authorities or private companies.

Furthermore, the EDPB and the EDPS consider that the use of AI to infer emotions of a natural person is highly undesirable and should be prohibited. In the AI Act at least emotion recognition in the workplace and educational institutions is prohibited. This means, that call center operators are not allowed to detect their employee's emotions using AI for example to evaluate their work performance. Likewise, it is not allowed to identify employees, who might be more likely to quit, by using an emotion recognition AI system.

[2] Nicht alle Quellen sind da einheitlich, leider liegt uns kein Kompromisstext vor.

VI. High Risk AI Systems

Other AI systems that are generally compatible with our European values, but could still cause significant potential harm, are classified as high-risk. The AI Act contains certain obligations for high-risk AI systems.

In addition, when high-risk AI systems are based on the processing of personal data or process personal data, compliance with the GDPR is required.

I will focus today on three key aspects:

- fundamental rights impact assessments;
- measures against discrimination; and
- rights of affected persons.

VII. Fundamental Rights Impact Assessments

Fundamental rights impact assessments are a risk assessment tool inspired by data protection impact assessments (DPIAs) from the GDPR. They will be mandatory for high-risk AI systems.

Furthermore, AI is a “new technology” within the meaning of Article 35 GDPR. Thus, if personal data are affected by an AI model or system, a DPIA is often required. A careful risk assessment should always be carried out when processing personal data using AI. If a DPIA and a fundamental rights impact assessment are required for a high-risk AI system, both can be conducted in conjunction.

A special focus should be placed on dealing with the identified risks. The impact assessment should be accompanied by a detailed plan describing the measures to mitigating those risks. Based on my experience with DPIAs, I welcome the obligation to conduct a fundamental rights impact assessment. Structured considerations about potential risks and ways to mitigate these, help to prevent harm to individuals. This will also increase the public’s trust in AI systems.

VIII. Discrimination and Human Oversight

In particular, when AI systems are used to assist with decisions affecting human beings, there is a high risk of discrimination. This risk must be mitigated during the entire life cycle of an AI system.

Errors can be introduced in the design of the system or during programming. The result of an AI system also depends to a large extent on how it was trained. If the system was trained with erroneous or biased data, this will also be reflected in its decisions. The data used and the design choices shall be captured in a technical documentation.

Whether an AI system is potentially discriminatory could already be apparent from the conception of the system and thus the technical documentation. Additionally, the AI Act requires testing of high-risk AI systems against defined metrics to detect unwanted bias and an appropriate level of accuracy for its intended purpose. This also aligns with existing nondiscrimination law: Deployers of a model can reasonably be expected to take available bias metrics into account. The use of a known flawed model would in turn be unacceptable.

Inadequate training or testing data or an unsuitable design are not the only factors that can lead to discriminatory results. The data used for training might become outdated over time, thus the accuracy of an AI system and the occurrence of unwanted bias have to be monitored continuously. This can be done by establishing a risk management system.

Another important factor that contributes to discrimination is that AI models are to some extent black boxes. Despite this property, an adequate level of comprehensibility and traceability shall be ensured. It is not necessary to understand all aspects in detail. However, data flows and considerations when designing an AI system should be comprehensible on the basis of the technical documentation. Record keeping while the high-risk AI systems is operating and instructions for use are also required according to the AI Act.

Using an AI system for a task it is not fit for, can lead to discriminatory outcomes and undermine the rule of law as well. This is a serious problem, as it could have severe consequences. Especially since the AI system was not designed for the task and its accuracy is therefore unclear. In particular, this means that generative AI systems should not be used to assist decision-making.

Last but not least, when using AI to assist with decisions affecting human beings, qualified human oversight is required. A qualified human must be able, allowed and even encouraged to disregard, override or reverse the output of a high-risk AI System, if necessary. This is the only way we can respect and guarantee the rights of affected persons. Only in this way, can we avoid negative consequences for individuals and facilitate acceptance of the respective AI system.

Compliance with high data protection standards is also a key factor in building trust in AI systems. Notably the right not to be subject to a decision based solely on automated processing, including profiling, already protects people from algorithmic discrimination in specific situations. Requirements for human oversight when using AI systems to assist in decision-making complement and extend the protection the GDPR offers in this regard.

Particular attention should be paid to the concept of automation bias though. This refers to the human tendency to favour suggestions from automated decision-making systems and to ignore contradictory information made without automation, even if it is correct. This might cause even qualified people to make mistakes when overseeing the application of AI for decision-making.

IX.

X. Rights of Affected Persons

Thus, persons affected by an AI system should receive information about it, to be able to recognize algorithmic discrimination. For this purpose, the AI Act requires transparency for AI systems, similar to the transparency requirements and the data subjects' right of access in the GDPR. Transparency is essential to ensure the protection of fundamental rights and democratic control. In a sense, it is a prerequisite for the meaningful application of all other data subjects' rights.

Affected persons also have the possibility to defend themselves against discrimination by AI systems. The AI Act provides complaint mechanisms and the possibility to take appropriate legal action for people whose fundamental rights have been violated by such systems. Citizens also have the right to lodge complaints with the AI supervisory authorities and receive explanations about decisions based on high-risk AI systems that impact their rights. This is of paramount importance in order to identify and counteract discrimination.

The GDPR also stipulates the possibility of presenting one's own point of view and contesting the decision in the exceptional case of automated decision-making. In general, AI systems must comply with applicable data protection regulations. That means data protection principles like purpose limitation and data minimisation apply to AI.

The requirements formulated in the AI Act complement the data protection requirements for AI systems. The interplay of the GDPR and the AI Act will strengthen the fundamental rights and data subjects' rights of EU citizens. At the same time, developers, providers and deployers of AI systems can use the synergies between the requirements from both regulations.

XI. Sandboxes

Ensuring the practical implementation of the requirements will still be a challenge. This requires well-equipped and expert supervisory authorities. These should already advise and support companies and authorities in the development phase of AI systems, in order to make them legally compliant. For this purpose, the AI Act provides for regulatory sandboxes.

Data protection authorities can also advise on the lawful development and use of AI systems and are available as points of contact.

XII. Conclusion

With the GDPR the EU has an innovative legal framework for data protection that serves as a reference and model for other regulations around the world. I appreciate that the AI Act will complement and concretize the GDPR by establishing core principles for the development and use of AI in the EU. I call on developers, providers and deployers of AI systems to respect these principles and to use AI in a responsible and trustworthy way for the benefit of all.

Consider the AI Act as a chance. For a Europe as an innovation-friendly business location with a stable framework for action. But first and foremost for a Europe that respects and protects of the fundamental rights of its citizens.

Thank you for your attention.