Fact Sheets Sustainability

Fair algorithms: Legal framework for artificial intelligence

MANAGEMENT SUMMARY

There are many uses for Artificial intelligence (AI). Like any other technology, it is often value-neutral at first but it can be harmful to humans and the environment if misused. The European Union is working on regulations for the development and use of AI to address this issue. This has the potential - similar to what has already happened in the area of data protection - to become the worldwide de facto standard for ethical AI due to the global network and its adoption in countries outside the EU. We present a compact summary of the current status (March 2022).

FACTS

Development of AI legislation in the EU

While AI is being driven forward in the USA by large corporations, and in China by the state, Europe has lagged behind in this area. There are many reasons for this. Due to ethical reservations, legal concerns, and the resulting mistrust of the technology, public and private sector initiatives are making little to no progress. AI is only being incorporated into products and processes very tentatively.

The EU wants to promote the development of AI, while also protecting the public's interests and building trust in AI. To do this, the EU is defining standards to provide legal and ethical clarity and give developers a sensible framework. An international, independent expert group developed and published the Guideline for Trusted AI in

2019, which served as the foundation for an official draft of the European AI law[1]. After revision and final adoption, the current draft law from April 2021 is expected to form the legal framework on AI in the EU from 2023 (as part of the "Coordinated Plan for AI"[2]). The draft includes prohibitions on certain AI applications, requirements and obligations for high-risk AI systems, and transparency obligations for AI systems. In addition, environmental and social sustainability aspects were strengthened against the backdrop of the EU's Green Deal. The proposed AI bill is supplemented by a Machinery Directive[3].

The OECD also published similar guidelines for AI in 2019 as recommendations for governments, investors, and developers of AI. However, these OECD principles are not legally binding. The member states of the EU will be obliged to implement much of the European law. Due to the economic importance of Europe and the global interconnectedness of markets, the future EU law for AI has the potential to become the global standard (similar to what has happened, for example, in data protection through the GDPR).

Basic requirements for AI that can be trusted

The EU guideline and draft law lay out some basic requirements. For example, AI systems must comply with all applicable laws, as well as ethical principles, respect human rights, and avoid harm - especially to humans. They must be reliably secured against adverse effects, whether these unintentional system errors or from external attacks. Safety and resilience to errors also include aspects such as accuracy, reliability, and reproducibility.

The fairness requirement and the prohibition of discrimination may sound self-evident, but they have already been the subject of public discussion in recent years because human biases have already crept into some AI. Algorithms learn from the data input into the system. If these are biased or based on real discrimination in the analog world, then this leads to the widely discussed AI bias.

The demand for AI to be human-centered is also reflected in the requirement that AI be explainable. It is essential that recommendations and decisions are transparent, traceable, verifiable, and comprehensible. Appropriate communication, e.g., with authorities and affected parties, must be ensured. This also applies to the occurrence and redress of damage (accountability).

In addition to the requirements made of the algorithms, the law also imposes requirements on the quality and security of the underlying data. Of course, the principles of personal data protection (data protection under GDPR) apply in full. Ideally, only

[1] https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:52021PC0206&from=DE
[2] https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review
[3] https://ec.europa.eu/docsroom/documents/45508?locale=de

anonymous data should be used, and where personal data is required, it should be kept to a minimum, used only for the pre-agreed purpose, and deleted once this purpose has been fulfilled.

However, for an AI system to function properly and legally, it must also be ensured that the data is correct and that poor data quality does not result in faulty AI models and decisions. The topic of data quality is known to have many facets beyond simple errors[4]. Problems are not always immediately apparent due to the high level of automation within AI systems. Intensive testing measures may be necessary, especially during ongoing operations.

The requirements above apply to all phases of the lifecycle of an AI system, i.e., during conception and development, during productive use, but also afterward (e.g., in archives). Once the law is passed, it will be binding for all users and providers of AI solutions within the EU, regardless of the origin and company location. And this represents the potentially global leverage of EU legislation.

What does this mean in the real world?

To make the general principles viable, the EU directive and draft law classify the various AI systems according to risk.

Specific applications are classified as an unacceptable risk. These include social scoring applications, systems that subconsciously manipulate human behavior, and remote biometric recognition (e.g., face, voice, gait) in real-time. These will be prohibited within the scope of the EU law in the future[5].

Precise requirements exist for AI systems that, while fundamentally useful, are also considered particularly risky. The list of such high-risk applications is long: critical infrastructure (including autonomous driving) and security components within products (e.g., med-tech). Applications that evaluate people and can disadvantage them through this evaluation (for example, credit risk evaluation or automated scoring of applications or exams) also belong to this category. Areas of governmental authority (e.g., law enforcement or border control) and permitted applications for biometric identification solutions are also classified as high-risk. All of these applications are subject to stringent regulations regarding robustness, security, risk management, transparency, data quality, and other criteria. These applications must be supervised by humans to minimize risks. Violations of these requirements should be prosecutable as a breach of law.

For low-risk AI systems, transparency is a key imperative. Users of chatbots, for example, must be able to recognize that they are interacting with a machine. The use of

[4] https://de.wikipedia.org/wiki/Informationsqualit%C3%A4t
[5] Social and legal debate still surrounds the use of facial recognition.

AI systems with minimal risk (e.g., spam filters and many other AI-supported functions within larger applications) should also be largely unrestricted. Standard product liability applies here too.

CONCLUSION

Despite some unresolved issues in many application areas, e.g., regarding the rights to the underlying data, liability issues, or even the specific civil and criminal consequences of violations, the future legal framework provides a solid foundation for providers and users of AI. It would be good if the legislative process would conclude fairly quickly so that the funding laid out in the “Coordinated Plan for AI’ could be implemented, and Europe could move swiftly to catch up in this race.

Disclaimer

Due to the ongoing legislative process, much of the information in this document should be understood as the status of facts and debate as of March 2022. Various changes are to be expected in the future.

Author

Dr. Marcus Dill is a successful serial entrepreneur and long-time management consultant with a focus on data, analytics, and sustainability.

Contact: Ingdilligenz GmbH

www.ingdilligenz.com

sustainability@ingdilligenz.com