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 has the potential to be harmful to humans and the environment if misused. The European Union therefore passed the first-ever comprehensive legal framework on AI worldwide in the summer of 2024. Similar to regulations in the field of data protection, this law has the potential – through global networking and adoption in countries outside the EU – to become the worldwide de facto standard for ethical AI. We provide a concise summary of the current status (March 2025).

FACTS

Development of AI legislation in the EU

While AI is driven by large corporations in the USA and by the government in China, Europe has so far lagged behind in this field. The reasons for this are diverse. Moral reservations, legal concerns, and the resulting mistrust of the technology certainly contribute to the fact that both public and private sector initiatives are either stalled or not progressing at all, and AI is only slowly being integrated into products and processes.

The EU aims to promote the development of AI while also protecting public interests and building trust in AI. To this end, standards are to be defined that provide legal and ethical clarity and offer meaningful frameworks for AI developers. For this reason, an international, independent group of experts developed guidelines for trustworthy AI, which were published in 2019 and served as the basis for the official draft of a

European AI Act[1]. The AI Act, which was adopted by EU member states in May 2024, will form the legal framework for artificial intelligence in the EU (as part of the so-called 'Coordinated Plan for AI'[2]). The law includes bans on certain AI applications (with the first such bans taking effect in February 2025[3]), requirements and obligations for high-risk AI systems, as well as transparency obligations for AI systems. Additionally, ecological and social sustainability aspects were strengthened in light of the EU Green Deal. The AI Act will come into full effect on August 2, 2026.

Similar principles for artificial intelligence were also published by the OECD in 2019 as recommendations for governments, investors, and AI developers. However, these OECD principles are not legally binding. In contrast, the European AI Act largely obliges EU member states to implement corresponding frameworks. Due to Europe's economic significance and the global interconnection of markets, EU law on AI has the potential to become a global standard (similar to what has already happened in the field of data protection through GDPR).

Basic requirements for AI that can be trusted

The guidelines and the EU law set out some fundamental requirements. AI systems must comply with all applicable laws, but also adhere to ethical principles, respect human rights, and avoid harm—especially to individuals. They must be secure against adverse effects, whether these result from unintended system errors or external attacks. Safety and robustness against errors also include aspects such as accuracy, reliability, and reproducibility.

The principle of fairness and the prohibition of discrimination may seem self-evident, but they have already been part of public discussions in recent years, as human biases have crept into some AI systems. Algorithms learn based on input data. If these are unbalanced or contain correlations that stem from factual discrimination in the analog world, this leads to the much-discussed AI bias.

The required human-centered approach for AI is also reflected in the fact that AI must be explainable. Transparent traceability, comprehensibility, verifiability, and clarity of recommendations and decisions are essential. Corresponding communication, for example with authorities and affected individuals, must be ensured. This also applies to the occurrence and compensation of damage (accountability).

In addition to the algorithms themselves, the law also sets requirements for the quality and security of the underlying data. Naturally, the principles of personal data protection (data protection according to GDPR) apply in full. Ideally, only anonymized data should

[1] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689
[2] https://digital-strategy.ec.europa.eu/en/policies/plan-ai
[3] https://germany.representation.ec.europa.eu/news/ki-gesetz-erste-vorschriften-kraft-getreten-2025-02-03_de

be used, and where personal data is necessary, it must be limited to a minimum and the purpose previously agreed upon, and deleted once this purpose has been fulfilled.

For the proper and legal functioning of an AI system, it must also be ensured that the data is accurate and that faulty AI models and decisions do not arise from poor data quality. Data quality, as we know, has many facets beyond trivial errors. Problems may not always be immediately apparent due to the high level of automation within AI systems. Intensive inspection measures - especially and precisely during ongoing operations - may be required.

The aforementioned requirements apply to all phases of the AI system lifecycle, including conception and development, during productive use, and even afterward (e.g., in archives). After the law is enacted, they will be binding for all users and providers of AI solutions within the EU, regardless of origin and company headquarters. This is where the potentially global leverage of EU legislation lies.

What does this mean in the real world?

To make the above general principles practical, the guidelines and EU law classify the various AI systems into four risk levels based on their risk.

Applications that pose a threat to the safety, livelihood, or rights of individuals fall into the category of unacceptable risk. This includes, among other things, social scoring applications, systems that subconsciously manipulate human behavior, and biometric remote identification (e.g., face, voice, gait) in real-time. All of these are prohibited under the scope of the EU law[4].

High-risk applications are those that pose significant risks to health, safety, or fundamental rights. The list of such high-risk applications is long: critical infrastructure (including autonomous driving), security components within (e.g., medical) products, and any applications that assess and potentially discriminate against individuals (e.g., credit risk assessment or automatic scoring of applications or exams). Sovereign areas (e.g., law enforcement or border control) and permitted applications for biometric identification solutions are also classified as high risk. For all these applications, particularly strict regulations apply concerning robustness, security, risk management, transparency, data quality, and other criteria. To minimize risks, these applications must be subject to appropriate human oversight.

For AI systems with low risk, transparency is the key requirement. For example, users of chatbots must be able to recognize that they are interacting with a machine, or providers of generative AI must ensure that AI-generated content is clearly identifiable. The use of AI systems with minimal or no risk (e.g., spam filters and many other AI-

[4] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689

supported functions within larger applications) should also be largely unrestricted in this regard. In these cases, normal product liability applies.

CONCLUSION

Despite some still open issues in many application areas, such as the question of rights to the underlying data, liability issues, or the specific civil and criminal consequences of violations, the legal framework provides a secure foundation for providers and users of AI. With the completion of the legislative process, it is hoped that the systematic promotion outlined in the 'Coordinated Plan for AI' will be implemented promptly, and that Europe's catch-up effort in this field will proceed with full momentum.

Author

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

Contact: Ingdilligenz GmbH

www.ingdilligenz.com

sustainability@ingdilligenz.com