Ethics Briefing

Guidelines for a responsible development
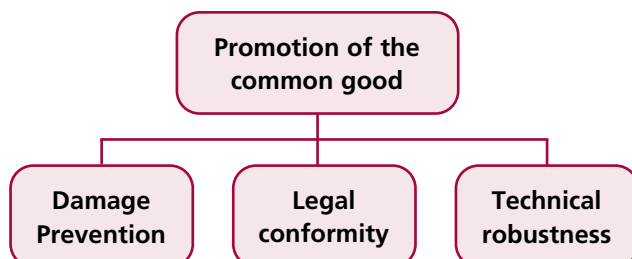and application of AI systems

White paper by Jessica Heesen et al.
Working Group IT-Security, Privacy, Legal
and Ethical Framework

Executive Summary

Ethical values and principles are important for many people, even in the eco-
nomic context. This is especially true for the development and application of
Artificial Intelligence (AI). Even if the possibilities offered using AI systems in
the various fields of application are viewed positively by many, there are still
some concerns regarding this technological innovation. A lack of trust is a
major reason for the lack of acceptance of AI systems. The working group IT
Security, Privacy, Legal and Ethical Framework of Plattform Lernende Systeme
has developed a guideline on the challenges that need to be considered when
developing and using AI systems responsibly. The guidelines also outline meas-
ures used to implement an ethically reflective technology development and
application process.

The development and use of AI systems should always aim to contribute to the
promotion of the common good. As a result, AI systems should be used in a
way that **avoids harm** to individuals, communities and the environment, en-
sures **legal compliance** of AI systems and actors, and provides the necessary
technical robustness so that the AI system used never poses an unacceptable
security risk.

Promotion of the common good		
Damage Prevention	Legal conformity	Technical robustness

Based on this, criteria for a responsible development and application of AI systems can be defined. Three superordinate categories of ethical values can be distinguished: **Self-determination**, **justice** and **protection of privacy and personality**. These values are followed by further ethical principles as well as necessary prerequisites for their realization.

| Ethical values |
| Further principles |
| Necessary requirements |
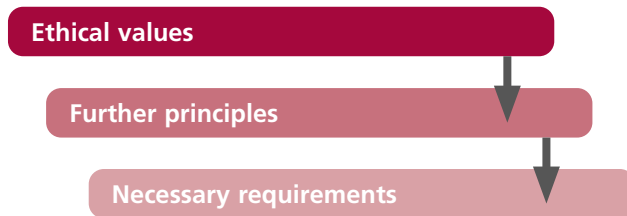
Self-determination justifies the demand that AI systems should support human actors in the decision-making process. Necessary prerequisites for implementation are **transparency** and **traceability** as well as **open interfaces** and **system operability** regarding self-determination on the market. The principle of **justice** demands **equality before the law**, **freedom from discrimination** and **equality of opportunity**. **Fairness** in distributive justice and **sustainability** as inter-generational-rational justice are concrete examples of application. Accountability is a prerequisite for its realization. **Privacy** refers to a space that is distinct from the public sphere and is protected from external interference. With increasing use of AI, the protection of privacy becomes a major challenge. **Data minimization** and **transparency** in data collection and processing are prerequisites for privacy protection. These ethical values can be realized through concrete measures by different actors, which are presented in the following guideline:

Options for action for developers of AI systems

Self-determination
- Development of AI systems as explainable AI
- Transparent communication of the use and degree of autonomy of the AI
- Development of AI-systems, which provide the human being as the final decision maker

Justice
- Avoidance of discriminating variables in data sets
- Establishment of control mechanisms

Privacy
- Use of anonymized data records
- Establishment of high security standards

Options for action for providers of AI-based products and services

Self-determination
- Open communication of the use of AI systems
- Provide training for employees

Justice
- Review of violations of principles of justice through forms of self-regulation
- Establishment of recognition and reaction patterns for damage

Privacy
- Obtaining declarations of consent for the collection and processing of data and offering easily understandable revocation options for these
- Use of anonymized data records

Options for action for users of AI systems

Self-determination
- Use of products that transparently mark the use of AI systems
- Selection of products that promote market plurality

Justice
- Avoiding the use of AI in areas with high potential for discrimination
- Random check of the plausibility of recommendations of AI systems

Privacy
- Obtaining declarations of consent for the collection and processing of data
- Use of anonymized data records

Options for action for those affected by AI systems

Self-determination
- Willingness to accept existing information offers on AI systems
- Active requesting of information offers

Justice
- Refrain from providing data that could lead to discrimination (if possible)

Privacy
- Provision of data only to reputable providers

This guideline is intended to serve as an orientation for developers, providers, users and affected persons who pursue an ethically reflected development and application process of AI systems.

It is therefore important that feedback is also given to the practice of companies in Germany. An ethically reflected design and application process is necessary, but not always sufficient to exclude all risks associated with AI for applications. Thus, human or technical failure as well as misuse of AI systems can lead to unintended consequences that cannot be foreseen despite a prudent development process.

The question of ethical criteria for the application and use of an AI system is closely related to the discussion about regulation and certification in general and the criticality of AI systems in particular. In general, the higher the risks associated with the use of a system, the more weight should be given to compliance with ethical principles as early as the development process. Nevertheless, these should ideally also be observed in applications with low damage potential.

SPONSORED BY THE

Federal Ministry
of Education
and Research

acatech

NATIONAL ACADEMY OF
SCIENCE AND ENGINEERING