

**AI ethics** is a set of values, principles, and techniques that employ widely **accepted standards of right and wrong** to guide moral conduct in the development and use of AI technologies.

Leslie, D. (2019). Alan Turing Institute.

(has connection to **social responsibility** and **social psychology**)

**Ethics as an element of the Trustworthy AI** ([European Commission](#), HLEG)

**Ethics as a principle of social responsibility**, [ISO/IEC 26000](#)

**Trustworthiness** - ability to meet stakeholders' **expectations** in a verifiable way

This definition can be applied across the broad range of AI systems, technologies and application domains ([ISO/IEC TR 24028](#))

To implement ethical AI principles (the main steps among others):

- Conducting **human rights impact assessments** on emerging technologies;
- **observance of safety and security threats as well as an investigation of the legal, regulatory and ethical framework of the product** ([ISO/IEC TR 24028](#))

# **codification of AI ethics for subject-specific areas**

**code of AI ethics  
in medicine**

**code of AI ethics  
in education**

**code of AI ethics  
in transport sector**

**other  
sectors**

**Codification of ethical standards in the field of AI**

**Principle of informed consent; Precautionary principle; Codes of ethics in IT;  
technology ethics; research and innovation ethics; data ethics;  
professional ethics; other ethical principles and standards**

**general principles for AI development**

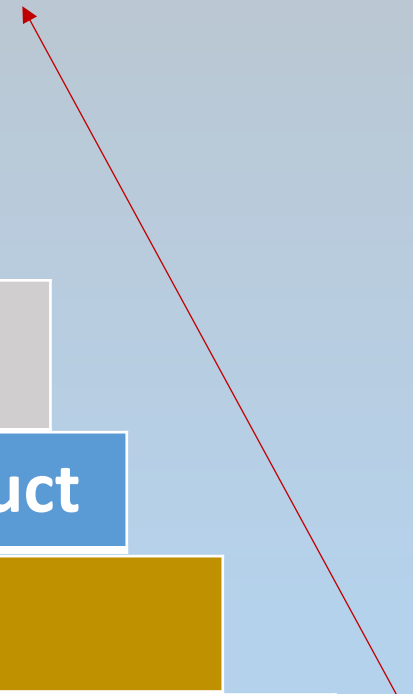
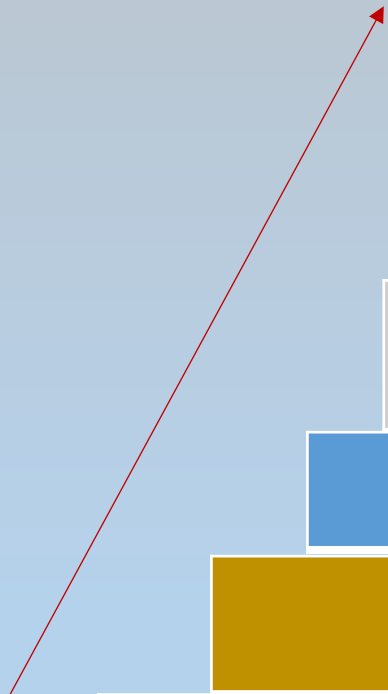
**social responsibility**

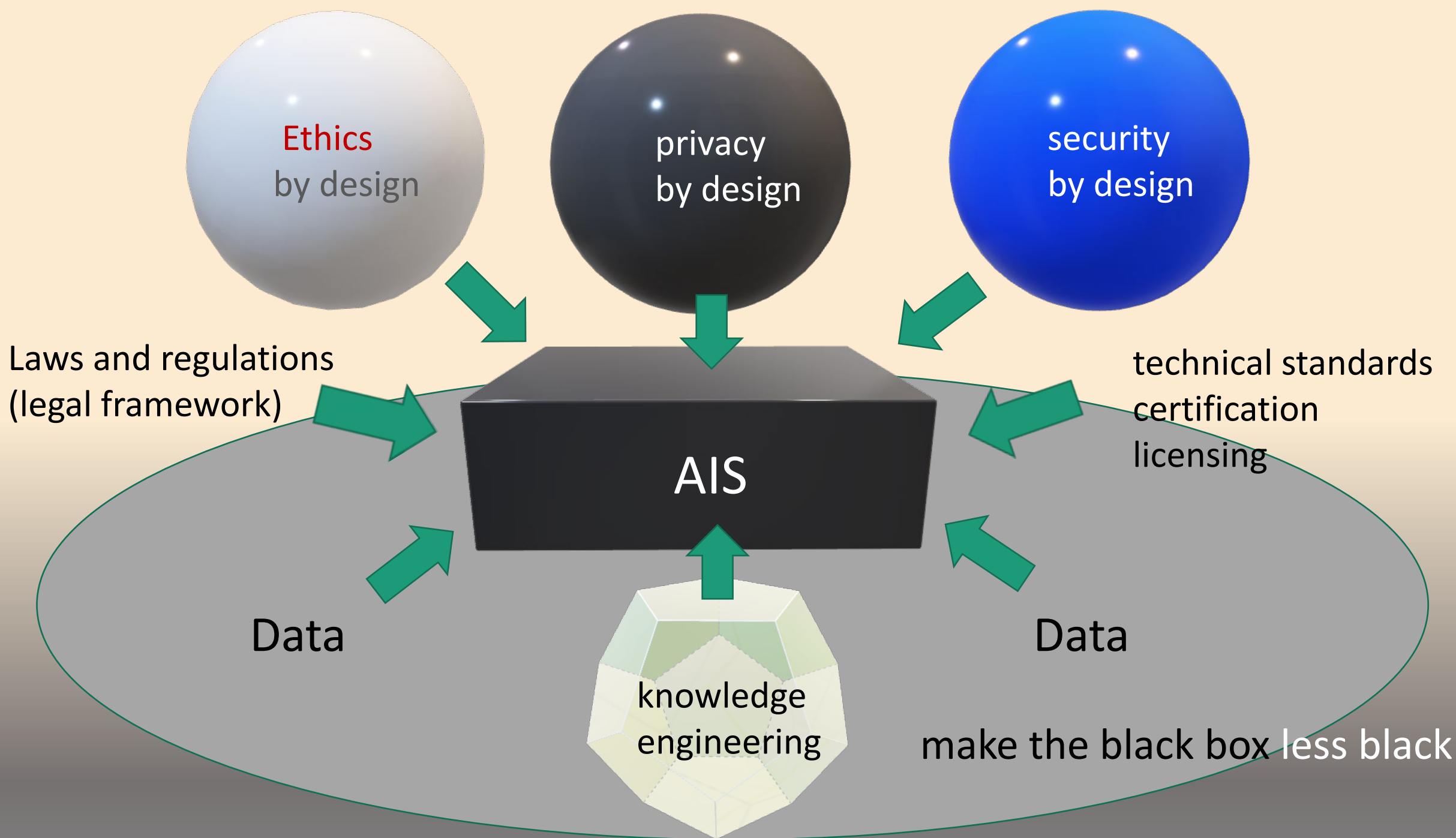
**responsible business conduct**

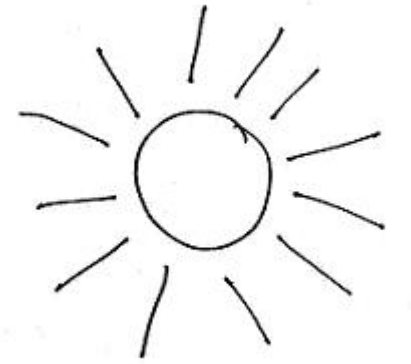
**human rights**

**human values**

**fairness  
(ideas of fairness)**







thank you for your attention!



Andrey Ignatyev

[a.ignatyev@odin.mgimo.ru](mailto:a.ignatyev@odin.mgimo.ru)

MGIMO University

Center for Global IT-Cooperation