ENHANCE workshop

Artificial Intelligence in Higher Education

June 22, 2023

Roman Z. Morawski

Warsaw University of Technology

roman.morawski@pw.edu.pl

ETHICAL & EDUCATIONAL IMPLICATIONS OF THE DEVELOPMENT OF AI-SUPPORTED TECHNOLOGIES

EXECUTIVE SUMMARY
or the essence of this talk

- AI-supported systems will more and more frequently:
  - generate knowledge
    (not only acquire data and extract information from them)
  - make morally significant decisions
    (not only support us in making such decisions)
- *Ergo*, the their designers should be equipped with some intellectual skills in the areas of:
  - methodology of technoscientific knowledge generation
    (philosophy of science)
  - methodology of making morally significant decisions
    (ethics of research and engineering)
- *Ergo*, the educational programmes should be enhanced with relevant contents.
- The programme enhancement should be implemented at both graduate and undergraduate level of studies in all branches of engineering because AI tools, already today, are applied there
  (not only in the field of computer engineering)

AS TO AVOID MISUNDERSTANDING

Mal nommer un objet c'est ajouter au malheur de ce monde

(Albert Camus, 1944)

- **Morality** = a social phenomenon, *viz*. a system of principles, values and judgments concerning individual-individual relationships and individual-group relationships, developed by a community to control those relationships

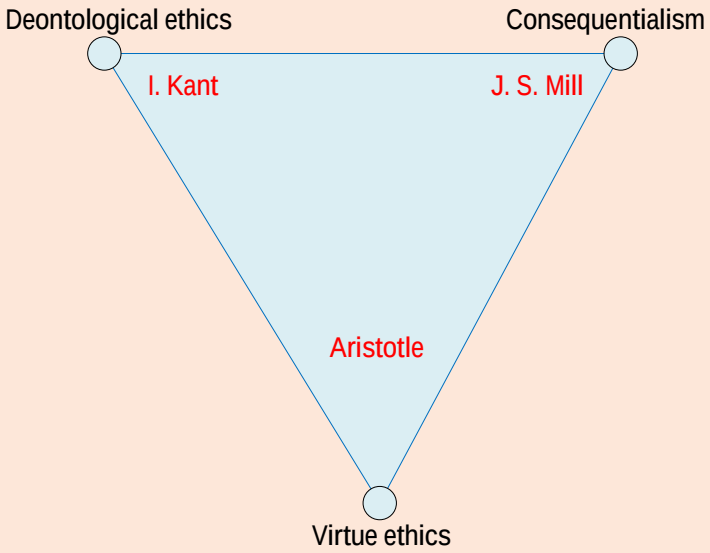
- **Ethics** = a philosophical discipline oriented on development of theories of morality:
  - general ethics (developed without indication of any specific field of application)
  - applied ethics (developed with indication of a specific field of application)

- Ethics is neither morality nor law, neither moralisation nor indoctrination…

MORAL ASPECT OF DECISION-MAKING PROCESS

Moral evaluation of actions

- Three paradigms of moral evaluation of actions:
  - *Virtue ethics* is emphasizing the virtues or the moral character of a person; it is focusing on what makes a good person rather than on what makes a good action.
  - *Deontological ethics* is holding that decisions should be made solely or primarily by considering one's duties and the rights of others.
  - *Consequentialism* is holding that the consequences of a particular action form the basis for any valid moral judgment about that action; a morally right action is an action which implies good consequences.



- Every ethical system refers to the above paradigms of moral evaluation of actions

MORAL ASPECT OF DECISION-MAKING PROCESS
General pattern of making morally significant decisions

- Engineering approach inspired by optimisation practice:
  - a measure of decision consequences ~ optimisation criterion
  - moral obligations ~ optimisation constraints
  - virtues of decision-maker ~ performance parameters of optimisation tools

MORAL ASPECT OF DECISION-MAKING PROCESS
Moral dilemmas and perverse effects of decisions

https://pl.freepik.com/zdjecia/bipolar

- **Dilemma** = a problem whose solution must breach one of the important values (or principles of conduct) because of the conflict of values (or principles of conduct)

- **Ethical (moral) dilemma** = a dilemma referring to the values (principles) if at least one of them is of ethical (moral) nature

  Examples:
  - to extend the testing of an aircraft prototype or to pay higher bonuses to its developers
  - to design a durable product or to build in a mechanism of planned obsolescence
  - to join a well-paid arms project or a low-budget biomedical project

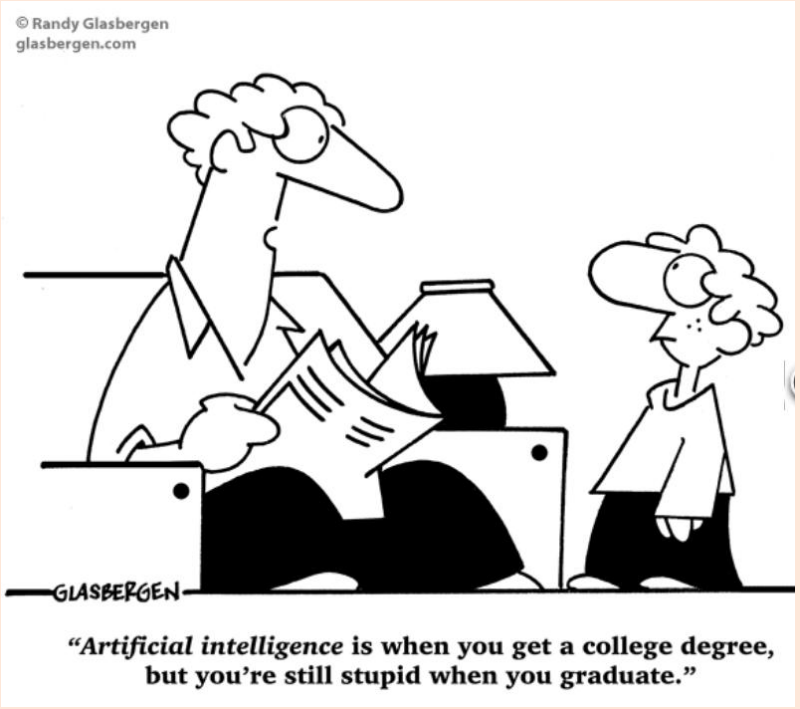
- **Perverse effects of decisions** = side effects of decisions, contrary to the purpose of making them

  Examples:
  - the evaluation of classes by students is aimed at improving teaching quality,
    its side effect – the elimination of the most demanding academic teachers
  - the evaluation of research performance is aimed at improving research quality badan,
    its side effect – the elimination of the most creative researchers who hate bureaucracy

INTELLIGENT DATA PROCESSING
Sources of moral anxiety

- The quality of data processing causally affects the quality of decisions made:
  - in management practice
  - in medical practice
  - in research practice
  - in everyday exploitation of the products of modern civilisation, *etc.*

- Among the algorithms of data processing, the most problematic (from ethical point of view) are algorithms that refer to AI tools



INTELLIGENT DATA PROCESSING
Approaching definition of AI

- **Artificial intelligence (AI)** = a branch of information technology referring to the most sophisticated methods of data processing (*sic*: lacking criterion of demarcation)

- Four types of AI-supported systems:

  Type 1 = systems designed for solving specific tasks (e.g. playing chess) – the systems which, when solving a new task, do not use the experience acquired during solving similar tasks in the past

  Type 2 = systems designed for solving specific tasks (e.g. driving cars) – the systems which, when solving a new task, use the experience acquired during solving similar tasks in the past

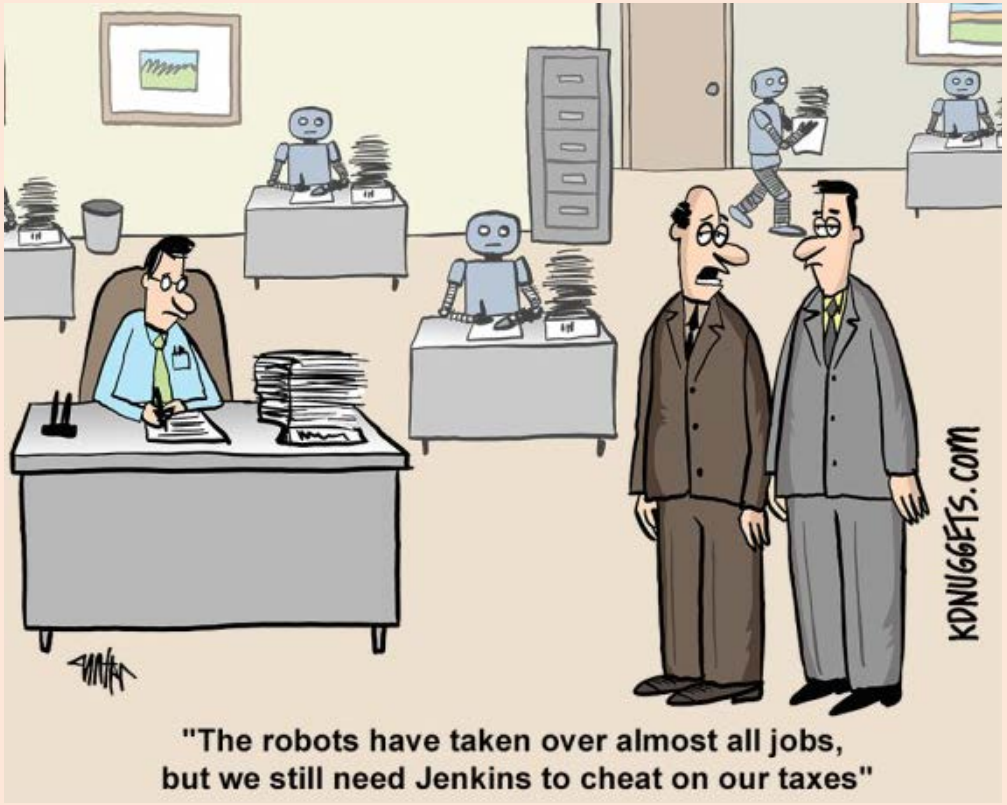
  Type 3 = systems equipped with so-called social intelligence – the systems which are able to read the intentions and predict the behaviour of people (e.g., the behaviour of elderly persons they "take care of")

  Type 4 = systems equipped with self-awareness – the systems of the future which are able to understand their internal processes imitating human thinking and emotions, as well as intentions and consequences of their actions.

C. Ategeka, *The Unintended Consequences of Technology: Solutions, Breakthroughs, and the Restart We Need*, 2022, s. 62–63.

INTELLIGENT DATA PROCESSING
Threats implied by dissemination of AI applications

- discriminatory actions
  e.g. on the labour market
- disinformation
  e.g. political or economic
- dominance of corporations
  e.g. Google, Microsoft, Apple
- negative impact on natural environment
  e.g. by controlling energy systems
- infringement of privacy
  e.g. in the field of healthcare
- military applications
  e.g. in combat drones
- operational uncertainty
  e.g. resulting from "opacity" of algorithms
- support for cybercrime
  e.g. by breaking security measures
- support for cheating
  e.g. in the field of education
- unemployment
  e.g. due to the automation of production or services



https://www.kdnuggets.com/images/cartoon-robots-human-tax-cheating.jp

INTELLIGENT DATA PROCESSING
Normative initiatives concerning ethics of intelligent robots

- Selected documents:
  - Principles (called *laws of robots*) proposed by Isaac Asimov (1942)
  - "Asilomar AI Principles" – a document promoted by Future of Life Institute (2017)
  - "Montréal Declaration for Responsible Development of Artificial Intelligence" – a document launched by Université de Montréal (2018)
  - "Ethics Guidelines for Trustworthy AI" – a document issued by European Commission (2019)
  - "Pause Giant AI Experiments" – an open letter, signed by numerous AI experts, published on the website of Future of Life Institute (March 2023)
  - "Proposal for a regulation on a European approach for artificial intelligence" – a document issued by European Parliament (recent amendment: June 2023)
- Common requirement: every AI system of Types 2–4 should be equipped with "moral intelligence"
- Key obstacles:
  - *of technical nature*: underdevelopment of the methods of implementing abductive reasoning, methods of identifying causal relationships and the methodology of resolving moral dilemmas and avoiding the negative effects of decisions
  - *of non-technical nature*: imperfection of human nature and business spirit of our times

INTELLIGENT DATA PROCESSING
Main obstacles...

T. Leśniak & R. Skarżycki, *http://bi.gazeta.pl/im/05/fe/15/z23062533V,Komiks-z-cyklu---Polska-mistrzem-Polski--.jpg*

PRACTICAL CONCLUSIONS

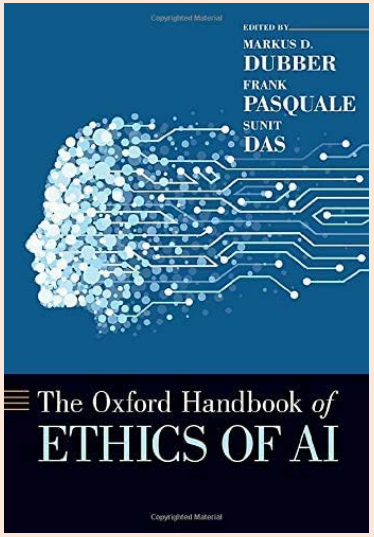
- AI research should be intensified rather than constrained; especially, if weak points of today's AI are concerned:
  - abductive reasoning and identification of causal relationships
  - implementation of the "ethical triangle"
  - resolving moral dilemmas and avoiding perverse effects of decisions

- The answer to the convergence of technoscientific disciplines, progressing with involvement of AI, should be the convergence of technoscience and humanities at all stages of education

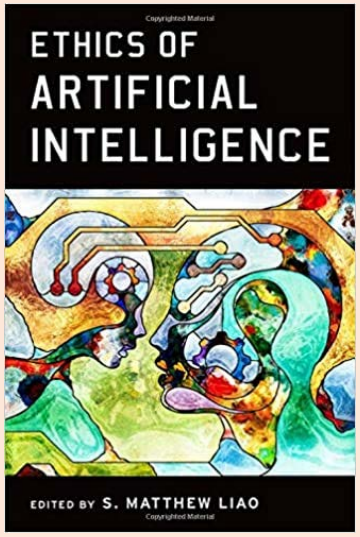
- It is necessary, therefore:
  - to introduce elements of ethics and philosophy of technoscience into the curricula of technoscience faculties
  - to introduce elements of knowledge about technoscience into the curricula of humanities
  - to give the highest priority to teaching methodologies which foster critical thinking and reflective attitudes towards social phenomena

ETHICAL ASPECTS OF TECHNOSCIENTIFIC APPLICATIONS OF AI

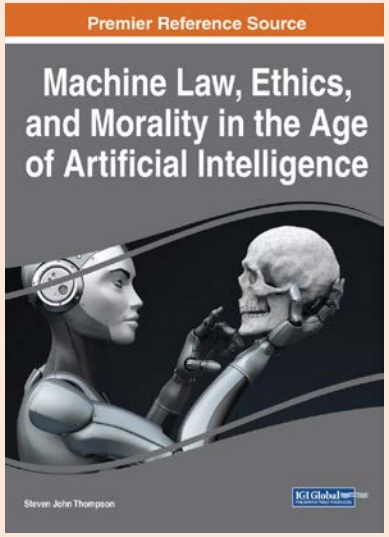
Books of inspiration



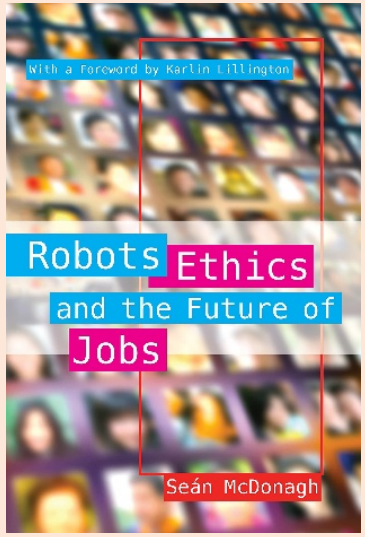
Oxford Univ. Press 2020



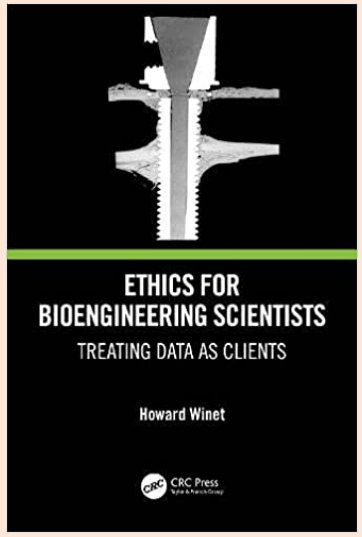
Oxford Univ. Press 2020



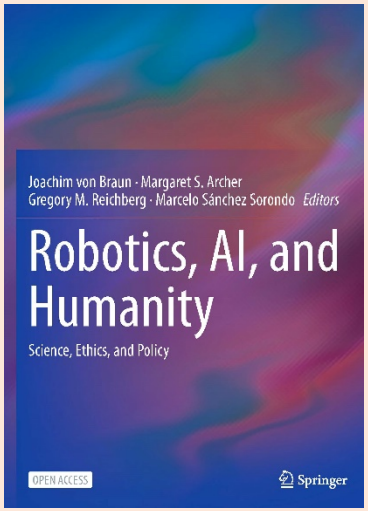
IGI Global 2020



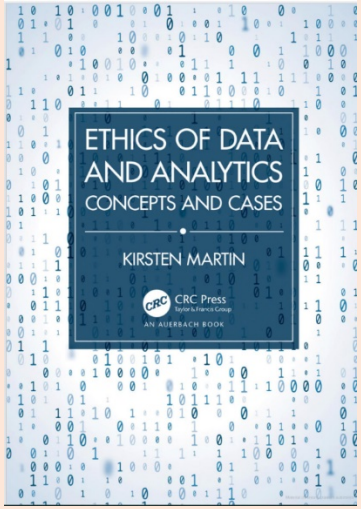
Messenger Pub. 2021



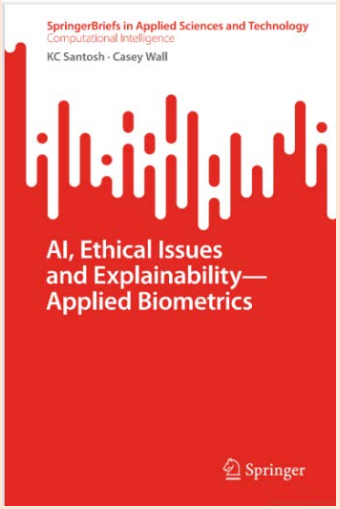
CRC Press 2021



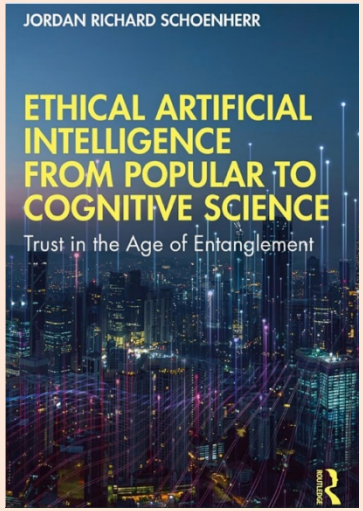
Springer Nature 2022



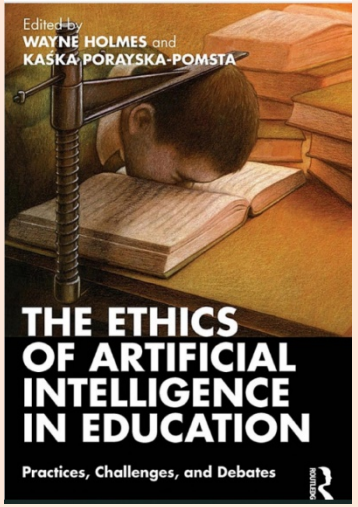
CRC Press 2022



Springer Nature 2022



Routledge 2022



Routledge 2022