

ETHICAL PROBLEMS IN THE USE OF ARTIFICIAL INTELLIGENCE TOOLS IN STATISTICAL RESEARCH

Krzysztof Jajuga
**Wrocław University of Economics and
Business**

ETHICS IN SCIENTIFIC RESEARCH

A set of norms setting out specific moral obligations related to the profession and social professional relations

Two fundamental principles:

Value for user of scientific research

Professionalism of researcher

BASIC ETHICAL RULES

Compliance with the law

Independence and objectivity

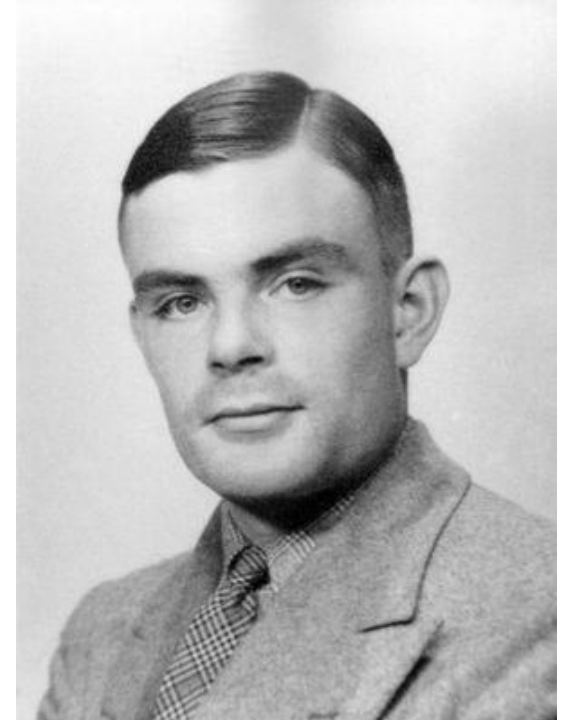
Professionalism on the market

Stakeholder relations

WHAT IS AI ?

Alan Turing:

**AI is the science and engineering
of making intelligent machines,
especially
intelligent computer programs**



NOTIONS

Artificial intelligence – area of science dealing with solving cognitive problems using available methods on available data sets

Statistical learning – framework for machine learning drawing from the fields of statistics and functional analysis

Machine learning – automated detection of meaningful patterns in data

STATISTICAL LEARNING AND MACHINE LEARNING

Conglomerate of two research streams:

- **Stream developed for several decades in the field of statistics and econometrics, in the area of confirmatory and exploratory methods**
- **Stream trend developed by combining the efforts of scientists from applied mathematics and statistics with scientists dealing with information technologies, in the area of exploratory methods (e.g. artificial neural networks)**

TWO PILLARS OF AI

Data + method → AI tool

DATA

Numerical data

Textual data

Image data

Audio data

Video data

Sensor data

METHODS

Descriptive methods

Confirmatory methods

Exploratory methods:

- **Discrimination (classification)**
- **Clustering**
- **Regression**
- **Dimensionality reduction**

GENERATIVE AI, DISCRIMINATIVE AI, PREDICTIVE AI

Generative AI (e.g. Generative Pretrained Transformer – GPT) – tools to create content in the form of text, audio, video, image, etc. Detect patterns in the available data and generate new content with similar characteristics

Discriminative AI – tools that assign content in the form of numbers, text, etc. to predefined classes

Predictive AI – tools that use datasets to recognize patterns over time and suggest future trends

VALUE TRANSFER THROUGH AI

AI Creator →

Black Box →

AI End User

ETHICAL GUIDELINES FOR STATISTICAL PRACTICE

Committee on Professional Ethics of the American Statistical Association

Professional Integrity and Accountability

Integrity of Data and Methods

Responsibilities to Stakeholders

Responsibilities to Research Subjects, Data Subjects, or Those Directly Affected by Statistical Practices

Responsibilities to Members of Multidisciplinary Teams

Responsibilities to Fellow Statistical Practitioners and the Profession

Responsibilities of Leaders, Supervisors, and Mentors in Statistical Practice

Responsibilities Regarding Potential Misconduct

INTEGRITY OF DATA AND METHODS

Communicating data sources

Communicating data transformation methods

Transparency about the assumptions and limitations of the methods used

Disclosure of information about corrections in the conducted research

Sharing data for possible replication

Protection of proprietary and confidential data

Quick correction of errors revealed during the research process

Validation of the methods used

DATA

FAIR

Findable

Accessible

Interoperable

Reusable

AI ACT – EU 2024 – AI SYSTEMS

Four levels of risk

- **Low-risk systems**
- **Limited or minimal risk systems**
- **High-risk systems – systems that can have a significant impact on the life chances of a user**
- **Systems with unacceptable risk**

ETHICAL PRINCIPLES FROM AI ACT

The principle of respect for human autonomy

Principle of prevention of harm

The principle of fairness

The principle of explainability

AI IN STATISTICAL RESEARCH – ETHICAL STANDARDS – PROPOSAL

Reliability of data

Adequacy of methods to the problem being solved

Transparency of the tool (data and methods) for the end user

Adapting the tool to the user's goal

SOME TAKEAWAYS

AI solutions used in research must be tailored to the problem (there is no such thing as "one size fits all")

The more automatically we apply AI, the more Naturally Stupid we are

THANK YOU!