AI for Humanity and Society 2024, WASP-HS conference
https://wasp-hs.org/event/ai-for-humanity-and-society-2024
https://tinyurl.com/ypss9epd Unfolding Ethics in Research and Society: Beyond Ethical Principles and Guidelines

ETHICS: COMPLIANCE AND BEYOND

AN INTERDISCIPLINARY, MULTI-LEVEL, MULTI-ASPECT APPROACH

GORDANA DODIG-CRNKOVIC

CHALMERS UNIVERSITY OF TECHNOLOGY & MÄLARDALEN UNIVERSITY

19 11 2024

CONTENT

Affiliations

School of Innovation, Design and Engineering

Division of Computer Science and Software Engineering

Research groups:
Artificial Intelligence and Intelligent Systems
Ubiquitous Computing

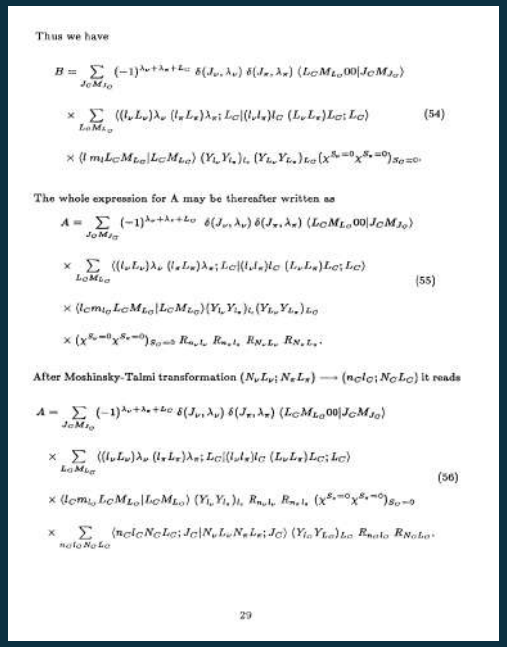
AI & Society-AI@MDU Research Fellow

Department of Computer Science and Engineering

Computer Science and Software Engineering Division

Research groups:
Interaction Design and Software Engineering
Critical Robotics

Background



Thus we have

$$B = \sum_{J_C M_{J_C}} (-1)^{\lambda_\nu+\lambda_\pi+L_C}\ \delta(J_\nu,\lambda_\nu)\ \delta(J_\pi,\lambda_\pi)\ \langle L_C M_{L_C} 00 | J_C M_{J_C}\rangle$$
$$\times \sum_{L_\nu M_{L_\nu}} \langle (l_\nu L_\nu)\lambda_\nu\ (l_\pi L_\pi)\lambda_\pi; L_C | (l_\nu l_\pi) l_C\ (L_\nu L_\pi) L_C; L_C\rangle \qquad (54)$$
$$\times \langle l\ m_l L_C M_{L_C} | L_C M_{L_C}\rangle\ \langle Y_{l_\nu} Y_{l_\pi}\rangle_{l_\nu}\ \langle Y_{L_\nu} Y_{L_\pi}\rangle_{L_C} \langle \chi^{S_\nu=0}\chi^{S_\pi=0}\rangle_{S_C=0}.$$

The whole expression for A may be thereafter written as

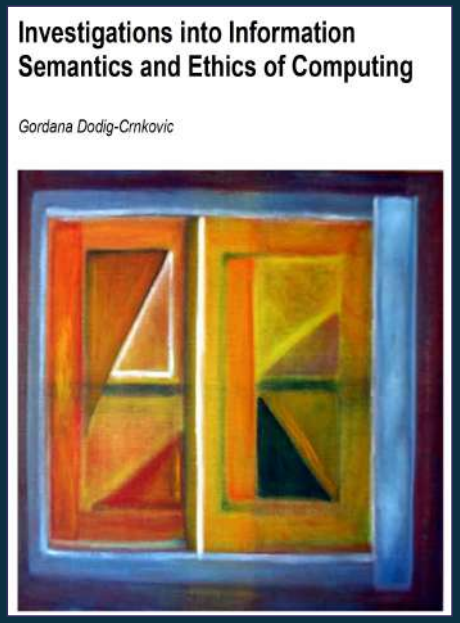
$$A = \sum_{J_C M_{J_C}} (-1)^{\lambda_\nu+\lambda_\pi+L_C}\ \delta(J_\nu,\lambda_\nu)\ \delta(J_\pi,\lambda_\pi)\ \langle L_C M_{L_C} 00 | J_C M_{J_C}\rangle$$
$$\times \sum_{L_C M_{L_C}} \langle (l_\nu L_\nu)\lambda_\nu\ (l_\pi L_\pi)\lambda_\pi; L_C | (l_\nu l_\pi) l_C\ (L_\nu L_\pi) L_C; L_C\rangle \qquad (55)$$
$$\times \langle l_C m_{l_C} L_C M_{L_C} | L_C M_{L_C}\rangle \{Y_{l_\nu} Y_{l_\pi}\}_{l_C} \{Y_{L_\nu} Y_{L_\pi}\}_{L_C}$$
$$\times \langle \chi^{S_\nu=0}\chi^{S_\pi=0}\rangle_{S_C=0}\ R_{n_\nu l_\nu}\ R_{n_\pi l_\pi}\ R_{N_\nu L_\nu}\ R_{N_\pi L_\pi}.$$

After Moshinsky-Talmi transformation $(N_\nu L_\nu; N_\pi L_\pi) \longrightarrow (n_C l_C; N_C L_C)$ it reads

$$A = \sum_{J_C M_{J_C}} (-1)^{\lambda_\nu+\lambda_\pi+L_C}\ \delta(J_\nu,\lambda_\nu)\ \delta(J_\pi,\lambda_\pi)\ \langle L_C M_{L_C} 00 | J_C M_{J_C}\rangle$$
$$\times \sum_{L_C M_{L_C}} \langle (l_\nu L_\nu)\lambda_\nu\ (l_\pi L_\pi)\lambda_\pi; L_C | (l_\nu l_\pi) l_C\ (L_\nu L_\pi) L_C; L_C\rangle \qquad (56)$$
$$\times \langle l_C m_{l_C} L_C M_{L_C} | L_C M_{L_C}\rangle\ \langle Y_{l_\nu} Y_{l_\pi}\rangle_{l_C}\ R_{n_\nu l_\nu}\ R_{n_\pi l_\pi}\ \langle \chi^{S_\nu=0}\chi^{S_\pi=0}\rangle_{S_C=0}$$
$$\times \sum_{n_C l_C N_C L_C} \langle n_C l_C N_C L_C; J_C | N_\nu L_\nu N_\pi L_\pi; J_C\rangle\ \langle Y_{l_C} Y_{L_C}\rangle_{L_C}\ R_{n_C l_C}\ R_{N_C L_C}.$$

29

PhD in Physics, 1988
On Alpha-decay, Department of Physics, University of Zagreb



PhD in Computing, 2006
Computer Science, Mälardalen University

Current: Morphological Computing and Cognition
AI Ethics, Digital Ethics, Digital Humanism

AN EXAMPLE: AI ETHICS

WHAT IS ARTIFICIAL INTELLIGENCE?

Definition of AI from European AI Act:

'Artificial intelligence system' (AI system) means a system that is designed to operate with a certain level of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of human-defined objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions , influencing the environments with which the AI system interacts.

https://www.artificial-intelligence-act.com/

This coincides with the OECD (Organisation for Economic Co-operation and Development definition:
https://mneguidelines.oecd.org/RBC-and-artificial-intelligence.pdf
https://www.oecd.org/en/topics/artificial-intelligence.html

ARTIFICIAL INTELLIGENCE

Definition of AI from the English Oxford Living Dictionary:

"The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages."

TYPES OF INTELLIGENCE

EMBODIED: HUMAN, CYBORG, ROBOT

“DISEMBODIED”: SOFTWARE, INFRASTRUCTURE

NARROW/ WEAK: INTELLIGENT ARTIFACTS FOR SPECIFIC PURPOSES

GENERAL (AGI)/STRONG: HUMAN - LEVEL AND ABOVE

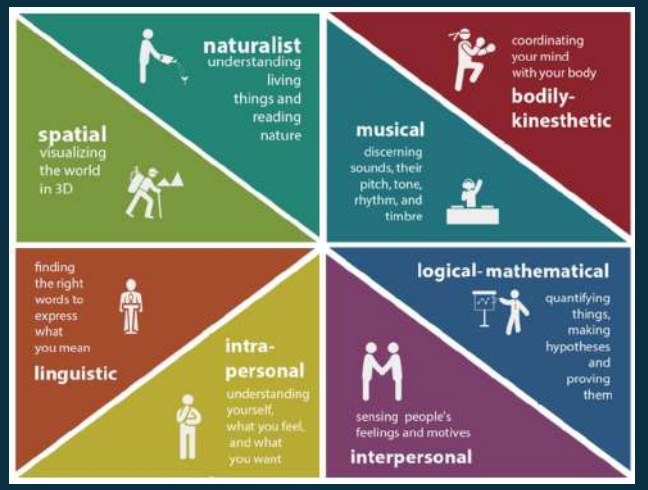
LLMs SUCH AS GPT MADE A SUDDEN DISRUPTIVE LEAP FROM NARROW AI TOWARDS AGI

WHAT IS NATURAL INTELLIGENCE?

Intelligence has been defined in many ways: the capacity for abstraction, logic, understanding, self-awareness, learning, emotional knowledge, reasoning, planning, creativity, critical thinking, and problem-solving. It can be described as the ability to perceive or infer information; and to retain it as knowledge to be applied to adaptive behaviors within an environment or context.
https://en.wikipedia.org/wiki/Intelligence

Also, most importantly:
learning and meta-learning (learning to learn)

Dimensions of natural intelligence:
Logical-Mathematical, Interpersonal, Intra-personal,
Linguistic, Spatial, Naturalist, Bodily-Kinesthetic, Musical



INTELLIGENT ARTIFACTS

- Intelligent Assistants & Co-pilots
- Intelligent Infrastructure
- Intelligent Robots & Softbots
- Intelligent Transportation
- Intelligent Cities
- Ambient Intelligence
- Intelligent IoT
- Intelligent Decision-Making Algorithms
  (introduced into particular technologies as self-driving vehicles but also into democratic institutions of governance, law, etc.)

https://bitcoinist.com/crypto-mining-becoming-concern-us-cities/

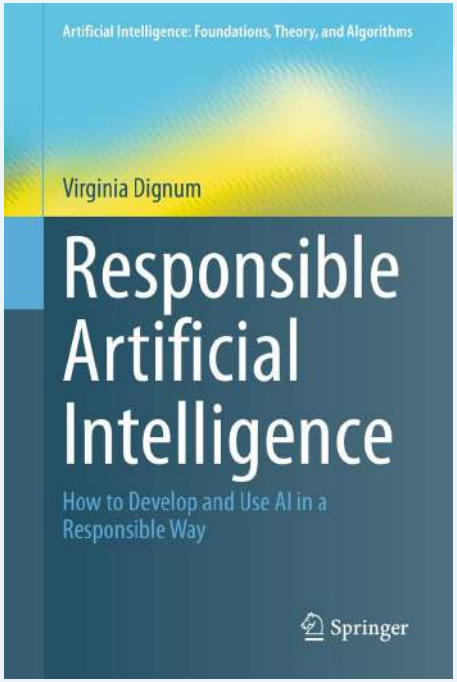
RESPONSIBLE AI: WHY CARE?

- AI systems are designed to act autonomously in our world
- Eventually, AI systems will make *better* decisions than humans in specific well-defined domains

  AI is designed by humans (currently), it is an artefact

- "We need to be sure that the purpose put into the machine is the purpose which we really want" (Alignment problem)

  Norbert Wiener, 1960 (Stuart Russell) King Midas, c540 BCE



Based on: Responsible Artificial Intelligence, Virginia Dignum, https://www.youtube.com/watch?v=BqwVRzKVz30

ETHICS & DESIGN

Ethics in Design (doing it right)

- Ensuring that development processes take into account ethical and societal implications of AI as it integrates and replaces traditional systems and social structures

Ethics by Design (doing right thing)

- Integration of ethical reasoning abilities as part of the behaviour of artificial autonomous systems

Ethics for Design(ers)

- Research integrity of researchers and manufacturers, and certification mechanisms

Based on: Responsible Artificial Intelligence, Virginia Dignum, https://www.youtube.com/watch?v=BqwVRzKVz30

TAKE AWAY MESSAGE ON RESPONSIBLE AI

- AI systems are artefacts built by us for our purposes
  - Our decision, our responsibility (currently)
- AI influences and is influenced by our social systems
  - Design in never value-neutral
  - Society shapes and is shaped by design
- Knowing ethics is not being ethical
  - Not for us and not for machines
  - Different ethics – different decisions
- Artificial Intelligence needs ART (Accountability, Responsibility, Transparency)
  - Be explicit!

Technology Policy Council Chair, Virginia Dignum

Virginia Dignum has been appointed Chair of ACM's Technology Policy Council, which addresses global technology policy. Dignum is a Professor of Computing Science / Responsible Artificial Intelligence and the Director of the AI Policy Lab at Umeå University, Sweden. She is also a member of the United Nations High Level Advisory Body on AI, whose report was released during the AI with Purpose Summit as part of the Science Summit at the UN General Assembly. She is also Senior Advisor to the Wallenberg Foundations.

Based on: Responsible Artificial Intelligence, Virginia Dignum, https://www.youtube.com/watch?v=BqwVRzKVz30

ASSIGNMENT OF RESPONSIBILITY

Time perspective

- Short-term perspective
  We, humans, decide
- Middle-term perspective
  AGI & We co-decide
- Long-term perspective
  Superintelligence? Who decides?

Levels of AI

- ANI (Narrow AI)
- AGI (Artificial General Intelligence)
- ASI (Artificial Super Intelligence)

Stakeholders

- Politicians
- Legislators
- Businesses
- Requirements engineers
- Designers, Developers
- Programmers
- Deployment engineers, testers
- Maintenance engineers

https://tinyurl.com/pjbdyn95 Global AI governance: barriers and pathways forward- Huw Roberts, Emmie Hine, Mariarosaria Taddeo, Luciano Floridi

ACM (Association of Computer Machinery) Code of Ethics

1. GENERAL ETHICAL PRINCIPLES

1.1 Contribute to society and to human well-being, acknowledgin that all people are stakeholders in computing.

1.2 Avoid harm.

1.3 Be honest and trustworthy.

1.4 Be fair and take action not to discriminate.

1.5 Respect the work required to produce new ideas, inventions, creative works, and computing artifacts.

1.6 Respect privacy.

1.7 Honor confidentiality.

…

https://www.acm.org/code-of-ethics

Responsible Research and Innovation*

Global challenges and opportunities prompted Responsible Research and Innovation (RRI), defined as:

"*a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society).*"

Von Schomberg

Education of future engineers should follow!

*Responsible Research and Innovation (RRI) is a term used by the European Union's Framework Programmes to describe scientific research and technological development processes that take into account effects and potential impacts on the environment and society.

Facing Grand Challenges of the Future with Tripple Helix Approach

The transformation of “ivory tower” context-independent to socially-aware paradigm in increasingly information-rich knowledge-based societies.

The triple helix model connects:

– ACADEMIA

– INDUSTRY/BUSINESS

– GOVERNMENT

– (+ NGO in 4-HELIX*)

https://inquiryumn.files.wordpress.com/2014/09/triple-helix.png

Inspired by biology: THE TRIPLE HELIX - Gene, Organism, and Environment by Richard Lewontin

*https://en.wikipedia.org/wiki/Quadruple_and_quintuple_innovation_helix_framework Quadruple and quintuple innovation helix framework

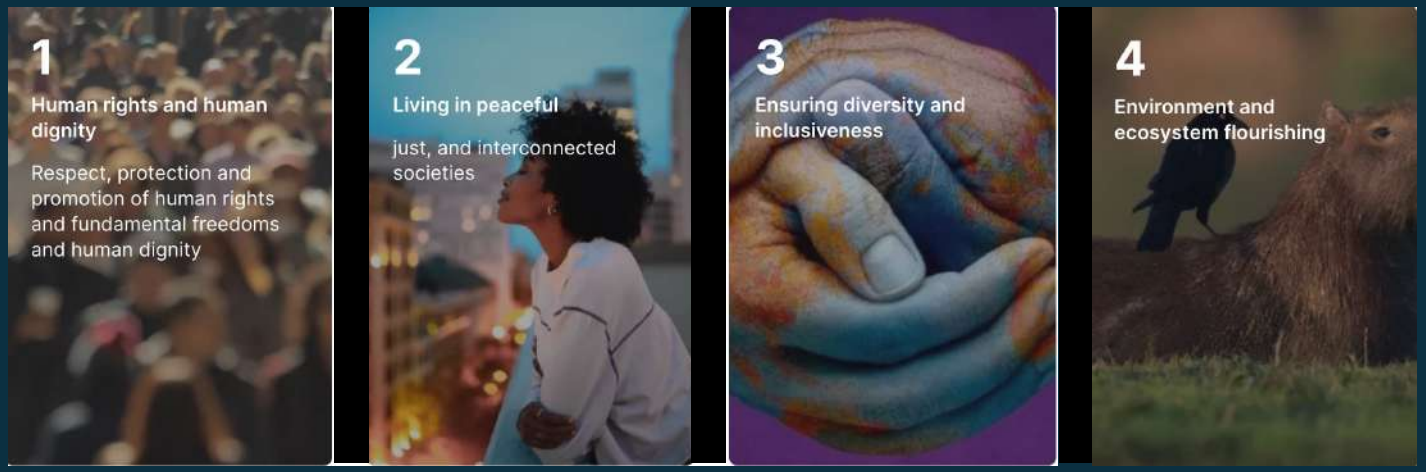
SYSTEMS OF RULES/LEGISLATION
Recent work on AI regulation

United Nations report (2023)
"Governing AI for Humanity"

https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf

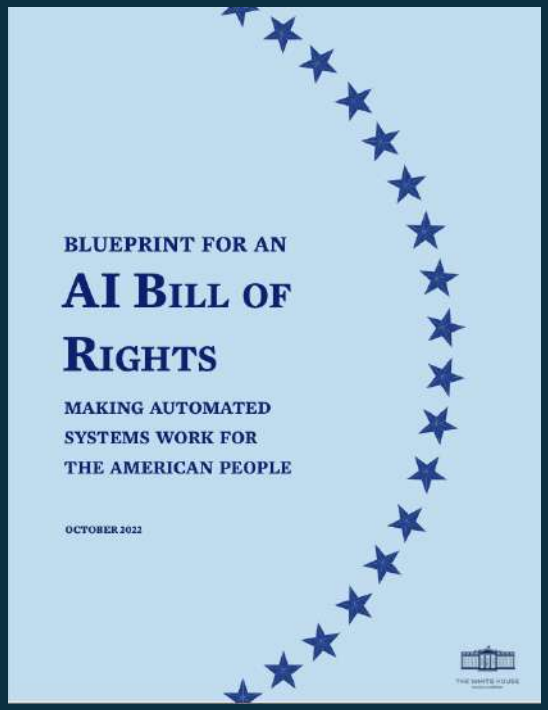
UNESCO 2022
‘Recommendation on the Ethics of Artificial Intelligence’

UNESCO's first-ever global standard on AI ethics



UNESCO: The United Nations Educational, Scientific and Cultural Organization
https://www.unesco.org/en/artificial-intelligence/recommendation-ethics
https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence

USA "AI Bill Of Rights" (2022) - principles



https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf

The US AI Bill of Rights outlines principles, including that people have a **right to control how their data is used** and to **not be discriminated against by unfair algorithms.**

It is a white paper, which does not have the force of law. It's primarily aimed at the federal government and could influence **which technologies government agencies acquire**, or help parents, workers, policymakers, and designers **ask tough questions about artificial intelligence systems**.

However, it can't constrain large tech companies, which arguably play a bigger role in shaping future applications of AI.

EU “AI Act” (2024)

The world’s first AI legislation

AI Act, European Commission. Shaping Europe’s digital future

https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

https://artificialintelligenceact.eu/ai-act-explorer/

Source: ISACA

The European Parliament granted final approval of the EU Artificial Intelligence Act on March 13, 2024, by a vote of 523 for passage, 46 against, and 49 abstaining. The Act faces a final step – approval by EU member states – as its provisions gradually take effect.

PROFESSIONAL ORGANISATIONS
IEEE* (Institute of Electrical and Electronics Engineers) on future intelligent autonomous systems

ETHICALLY ALIGNED DESIGN

The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems

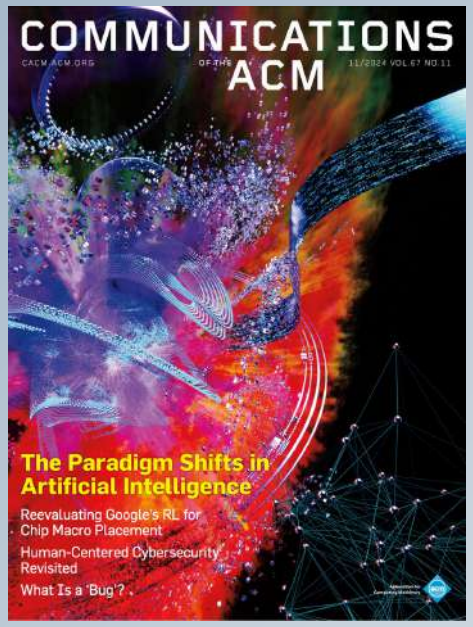
https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

Prioritizing human well being in the age of artificial intelligence: https://youtu.be/z5yZU8tp9W8 (5:56)

https://techethics.ieee.org/

*IEEE (USA) is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.

PROFESSIONAL ORGANISATIONS
Communications of the ACM (CACM)
ACM = Association for Computing Machinery



https://tinyurl.com/4bjr4vaz

https://vimeo.com/1021039731

PROFESSONAL LITERATURE defining norms

HANDBOOK OF AUTOMATION: AUTOMATION AND ETHICS

809

47. Automation and Ethics

Srinivasan Ramaswamy, Hemant Joshi

Should we trust automation? Can automation cause harm to individuals and to society? Can individuals apply automation to harm other individuals? The answers are yes; hence, ethical issues are deeply associated with automation. The purpose of this chapter is to provide some ethical background and guidance to automation professionals and students. Governmental action and economic factors are increasingly resulting in more global interactions and competition for jobs requiring lower-end skills as well as those that are higher-end endeavors such as research. Moreover, as the Internet continually eliminates geographic boundaries, the concept of doing business within a single country is giving way to companies and organizations focusing on serving and competing in international frameworks and a global marketplace. Coupled with the superfluous nature of an Internet-driven social culture, the globally-distributed digitalization of work, services and products, and the reorganization of work processes across many organizations have resulted in ethically challenging questions that are not just economically, or socially sensitive, but also highly culturally sensitive. Like the shifting of commodity manufacturing jobs in the late 1900s, standardization of information technology and engineering jobs have also accelerated the prospect of services and jobs more easily moved across the globe, thereby driving a need for innovation in design, and in the creation of higher-skill jobs. In this chapter, we review the fundamental concepts of ethics as it relates to automation, and then focus on the impacts of automation and their significance in both education and research.

Part E | 47

Ramaswamy S., Joshi H. (2009)

Automation and Ethics.

In: Nof S. (eds) Springer

Handbook of Automation.

Springer, Berlin, Heidelberg

Navigating our 'WHITE WATER WORLD' – complex & dynamic by design

"We are forcing the past as a solution set. But the past as a solution set is not a viable option. We need a new toolset."

Design Unbound presents a new tool set for having agency in the world today, which we characterize as a 'white water world' – one that is rapidly changing, hyperconnected and radically contingent.

Imagination is a 'muscle that must be exercised' (John Seely Brown)

Hyperconnectivity causes a transition from equilibrium to constant non-equilibrium.
The need for adaptivity, anticipation, and resilience.

Complexity science gives us a new lens through which to view the world as one that is entangled and emerging.

'Wicked problems': As soon as you start to solve them, they morph.
"Computational irreducibility": You must run the model to see the outcome. Computation takes the same time as the process itself.

Design Unbound. Designing for emergence in a 'white water world'.

Ann Pendleton-Jullian and John Seely Brown, MIT Press 2018

(1) Designing for Emergence &

(2) Ecologies of Change

https://www.desunbound.com/
https://www.youtube.com/watch?v=-U8h4wNBfCQ
https://www.youtube.com/watch?v=tFPvK1mO6Sg
https://www.youtube.com/watch?v=Lto8szGvPfM
https://www.desunbound.com/assets/DesUnbound_chapter_8.pdf

Richard Buchanan (1992) Wicked Problems in Design Thinking. Design Issues, Vol. 8, No. 2, pp. 5-21. The MIT Press http://www.jstor.org/stable/1511637.

Speculative Everything – Antony Dunne and Fiona Raby

"what if" questions

https://www.youtube.com/watch?v=kmibm20UsoA

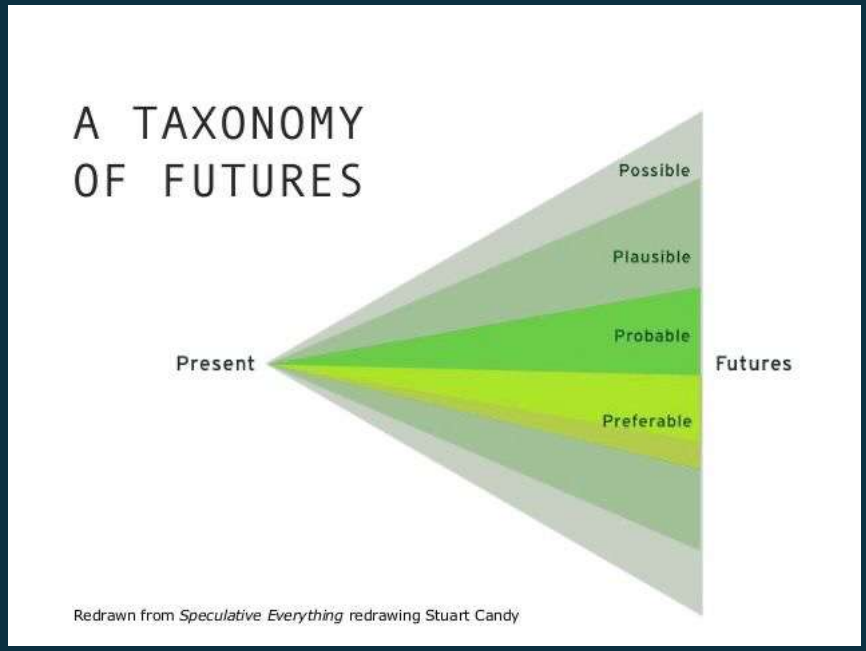
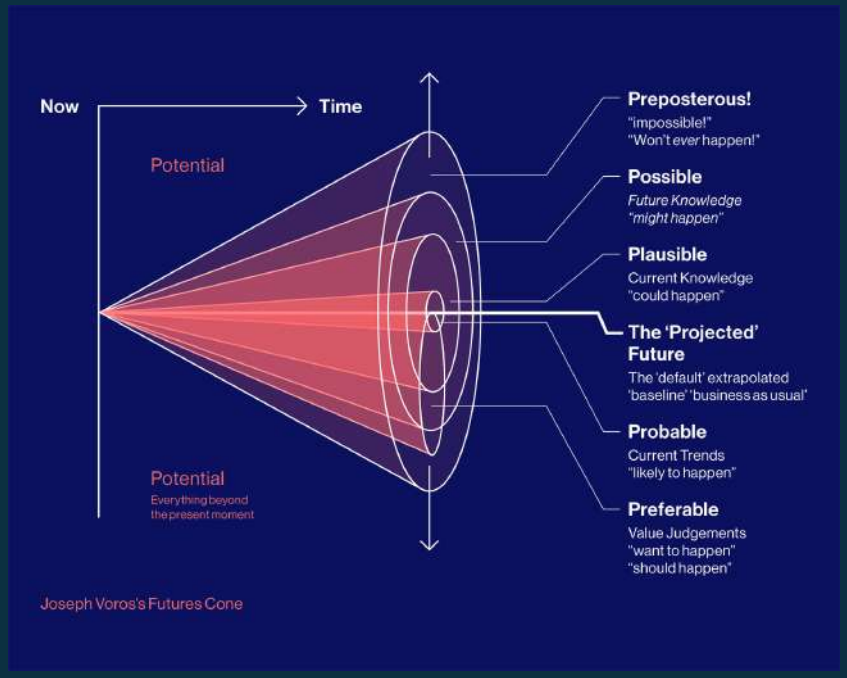


Table of Contents:
Beyond radical design?
A map of unreality
Design as critique
Consuming monsters: big, perfect, infectious
A methodological playground: fictional worlds and thought experiments
Physical fictions: invitations to make believe
Aesthetics of unreality
Between reality and the impossible
Speculative everything.

Speculative Design creates space to…



Arrange emerging (not yet available) technological 'elements' to hypothesize future, products and artifacts.

Apply alternative plans, motivations, or ideas to those currently driving technological development, in order to facilitate new arrangements of existing elements.

Develop new perspectives on big systems.

VALUE-BASED HUMAN-CENTRIC DESIGN

Values

Values serve as a guide to action and knowledge.

They are relevant to all aspects of scientific and engineering practice, including discovery, analysis, and application.

TUANA. COMMUNICATIONS OF THE ACM | DECEMBER 2015 | VOL. 58 | NO. 12

VALUE-SENSITIVE DESIGN

- Value-sensitive design (VSD) holds that artefacts are value-laden and design can be value-sensitive. The approach refers to the need to identify early implicit values embedded in new technologies by focusing on the usage situations of technology.
- "Value" is defined broadly as property that a person or a group considers important in life, and designers can intentionally inscribe their values in the design objects thus shaping them.
- The design is carried out iteratively by combining the following approaches supporting the values:

- conceptual (conceptions of values for users and stakeholders)
- empirical (how values are realized in practice)
- technical (design of technology),
- research all of which is followed by
- assessment

Luciano Floridi, Josh Cowls, Thomas C. King, Mariarosaria Taddeo (2020) How to Design AI for Social Good: Seven Essential Factors. Science and Engineering Ethics. https://doi.org/10.1007/s11948-020-00213-5

Sarah Spiekermann (2015) Ethical IT Innovation: A Value-Based System Design Approach, https://www.amazon.com/Ethical-Innovation-Value-Based-System-Approach/dp/1482226359

https://www.researchgate.net/publication/318993631_IEEE_P7000 The First Global Standard Process for Addressing Ethical Concerns in System Design

FIVE CAPITALS (VALUES)	
NATURAL	Natural world. Landscapes, ecologies, animals, plants and other life. Raw materials. Any stock or flow of energy and matter that yields valuable goods and services
HUMAN	Health, knowledge, skills and motivation, Intellectual, cultural, cognitive & emotional (well-being, happiness)
SOCIAL	Structures, institutions, networks and relationships which enable individuals to maintain and develop their human capital in partnership with others. Organized in circles of proximity, cognitively sustainable.
PRODUCTIVE	Production means of material culture - Material goods – tools, machines, buildings and other forms of infrastructure – which contribute to the production process but do not become embodied in output.
FINANCIAL	Monetary capital and relationships

Dodig-Crnkovic, G. and Sapienza, G.*, Ethical Aspects of Technology in the Multi-Criteria Decision Analysis. IACAP conference, Ferrara, June 14-17, 2016.

Implementing Value-sensitive approach presupposes ETHICS AS A PARTICIPATORY AND ITERATIVE PROCESS

Ethics involves a participatory and iterative process of ethical reflection, inquiry, and deliberation. Combining action and reflection is crucial.

It is instructive to go back and forth between zooming out and zooming in on the problem.

In this process, we consult different ethical approaches (Consequentialism, Duty ethics, Virtue ethics, Relational ethics, etc.)

Methods from Human-Centered Design (HCD) organizing participatory and iterative processes, Value Sensitive Design (VSD), bringing different stakeholder values, and Responsible Innovation (RI) with a focus on inclusion, participation, and diversity.

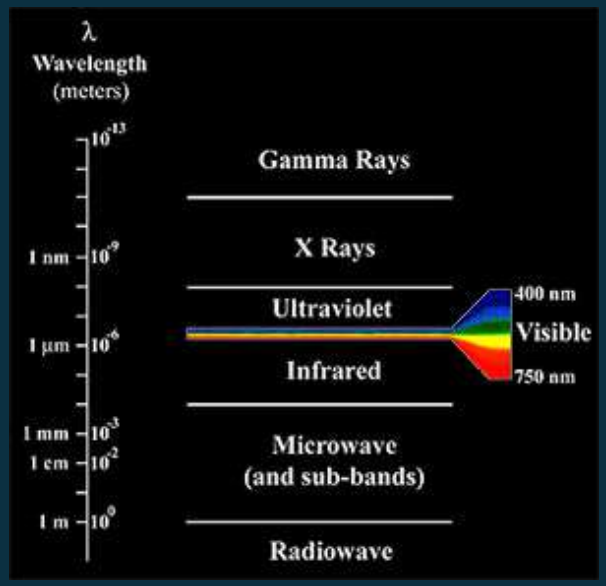
We face complex, interdisciplinary, and global challenges: climate crisis, political polarization, and inequalities. These are all **wicked problems**, which require diverse disciplines, both to better understand the problem and to envision and create solutions.

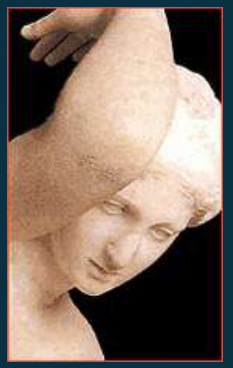
Doing ethics is not always easy or pleasant. It can involve asking uneasy questions, creating awkward situations, and tolerating tension and uncertainty.

https://dl.acm.org/doi/pdf/10.1145/3550069

Marc Steen

DIVERSITY OF VIEWS: World seen in different light



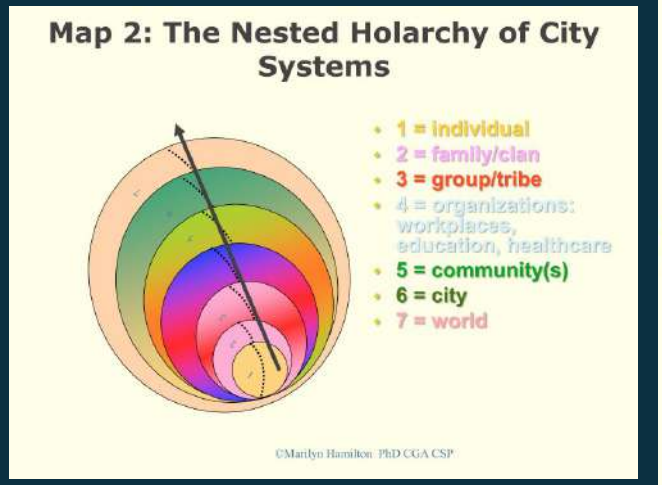


What if we could see in any wavelength of the electromagnetic spectrum, from gamma-rays to radio waves? How would the world appear to us?

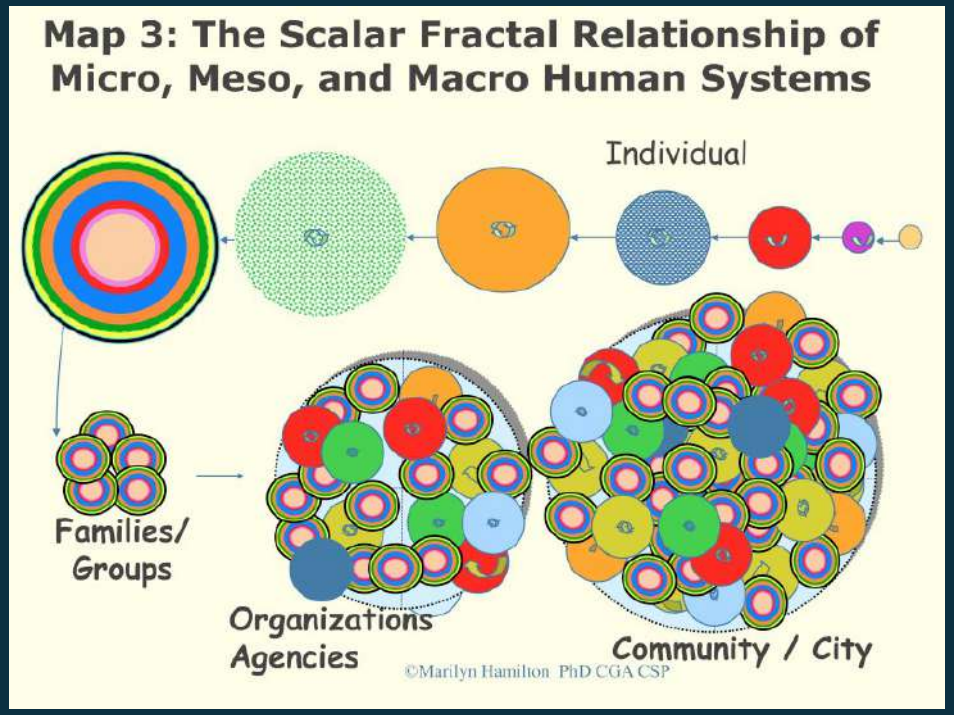
STAKEHOLDERS IN AN ACADEMIC RESEARCH PROJECT

Nature

Complexity Aspects Relating Micro – Meso – Exo – Macro Levels of Analysis –Example of City

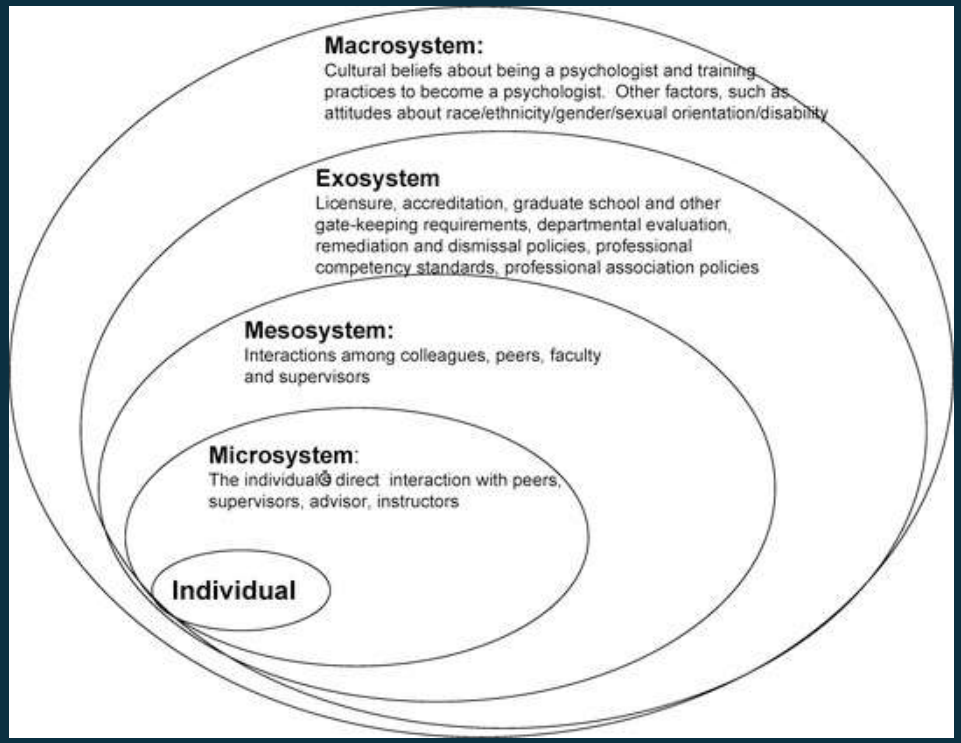


A holarchy, in the terminology of Arthur Koestler, is a connection between holons, where a holon is both a part and a whole. The term was coined in Koestler's 1967 book The Ghost in the Machine.



http://www.newsociety.com/Books/I/Integral-City

Micro – Meso – Exo – Macro Domains



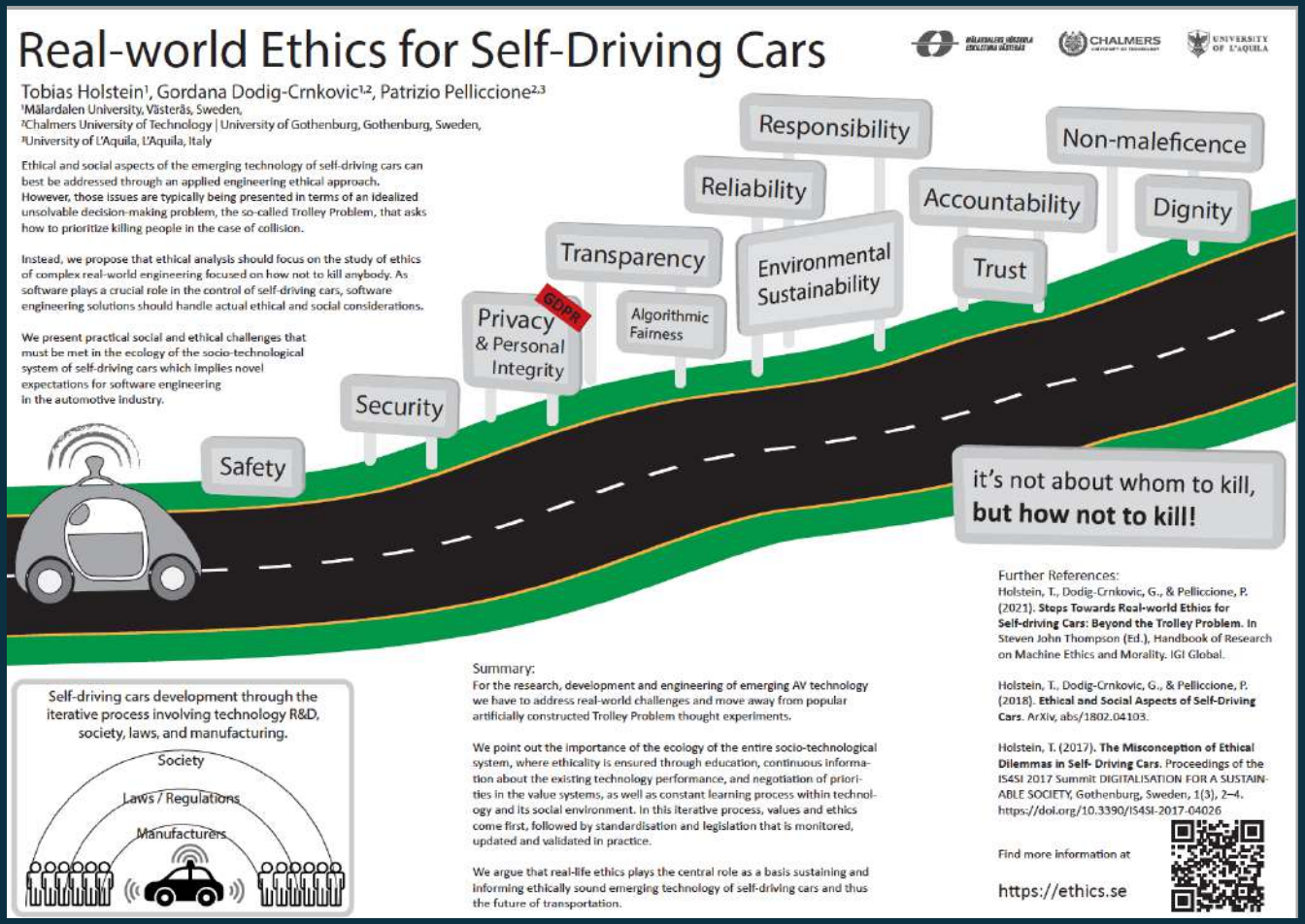
You will recognize this domain-based view in the analysis of many different types of problems – organization of society, sustainability of cities, ecology, economics, ethical aspects etc.

Source: American Psychological Association website

TAXONOMY OF RISKS

Categorisation of risk type	Definition	Probability typically	Researchers' obligations
Unaddressed	Consequences are known to the researchers but are not examined because of the predefined scope of the research.	High	Public disclosure and engagement with stakeholders to achieve consensus on research scope and objectives.
Unintended	Consequences are known to the researchers, but with low enough probabilities that associated risks can be ignored.	Low	Include in the scope of the research measures to avoid negative consequences, despite the low probability.
Unexpected	Consequences are known to the researchers, but no anticipated risk because of assumed zero probability of occurrence.	Zero	Disclosure and consultation within the professional and academic community to corroborate zero probability of occurrence.
Unforeseen	Consequences are unknown to the researchers, but could reasonably have been foreseen if researchers had anticipated risk.	Unknown	Anticipate possible negative consequences, and their associated probabilities and impacts.
Unforeseeable	Consequences could not have been reasonably identified by researchers at the given moment.	Undefined	Remain open-minded to possible negative consequences, and seek input on these, no matter which stakeholder(s) recognize(s) the risks.

EXAMPLE OF ETHICAL ANALYSIS: EHICS OF SELF-DRIVING VEHICLES



Presented as poster at ICSE2020
Extended version to appear as a chapter in the
Handbook of Research on Machine Ethics and Morality
| IGI Global 2021

Basis for speculative design. Assumes value sustem.

POSSIBLE FUTURES: AI UTOPIAN VS. DYSTOPIAN

Utopian Scenarios

- End of Poverty and Scarcity
- Universal Access to Knowledge and Education
- Improved Healthcare and Longevity
- Reduction of Human Labor and Increased Creativity
- Personal AI assistants (agents)
- Global Cooperation and Problem-Solving

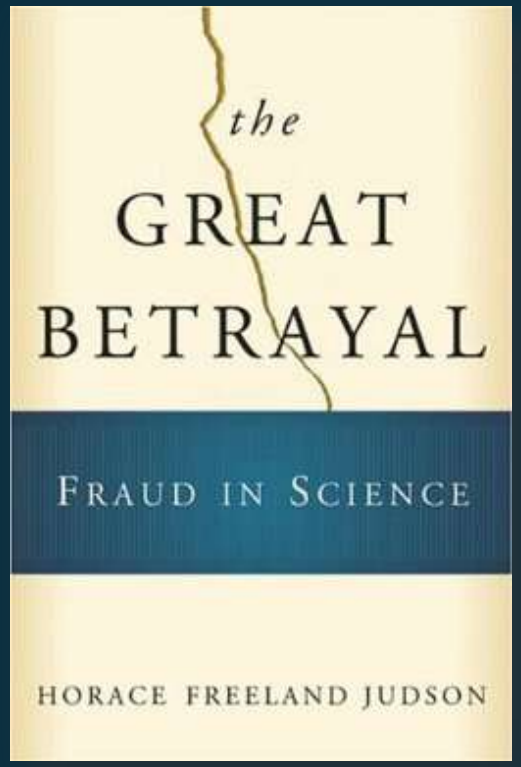
Dystopian Scenarios

- Mass Unemployment and Economic Inequality
- Mass Surveillance and Loss of Privacy
- AI-Driven Authoritarianism
- Weaponization of AI
- Existential Risk (AI Overreach)
- Erosion of Human Autonomy and Agency

ETHICS IN RESEARCH - ETHICAL REVIEW

Peer Review, by AJC1, Flickr.com

RESEARCH MISCONDUCT SEEN AS FFP



Fabrication
Making up data

Falsification
Manipulating data

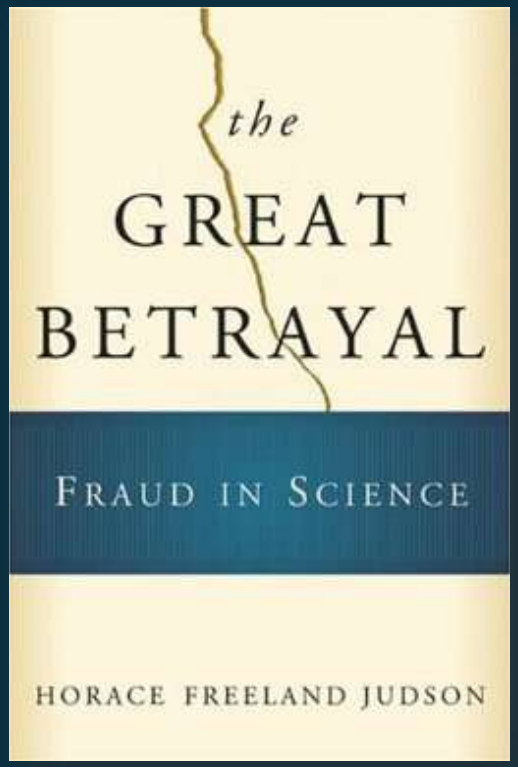
Plagiarism

Threat to Reliability

Taking words & ideas
Armstrong, J.D. (1993)

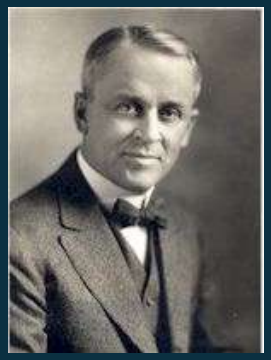
Lack of Justice

Famous “Sinners” in Science

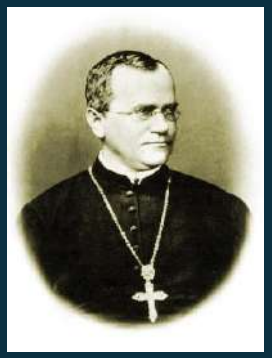


Newton

Kepler

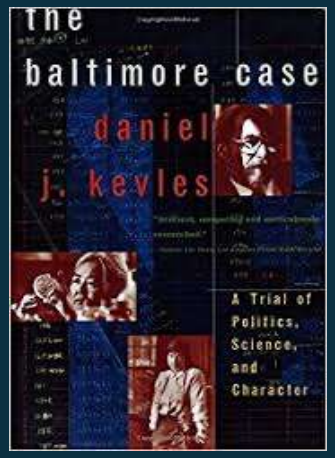


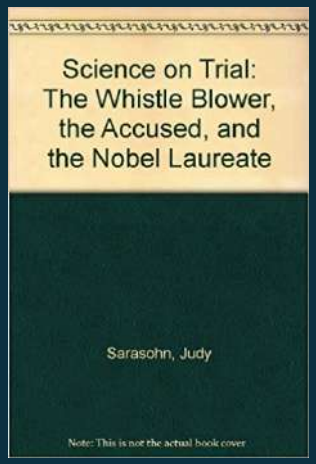
Millikan



Mendel

The The “Baltimore case”- The Baltimore scandal





David Baltimore, best known to the public not for his Nobel prize but for his defense of a research collaborator who was accused of misconduct but officially exonerated after a decade of government inquiries.

The Baltimore Case (W. W. Norton, 1998) by Daniel J. Kevles

Science on trial: the whistle blower, the accused, and the Nobel laureate (New York : St. Martin's Press, 1993) by Judy Sarasohn

A recent tragic offer of a scandal of research misconduct accusation in stem cell research

http://america.aljazeera.com/articles/2014/8/5/japan-stem-suicide.html
Yoshiki Sasai, 52, was the co-author of the high-profile research that had seemed to offer hope for replacing damaged cells or even growing new human organs. , 2014 08 05

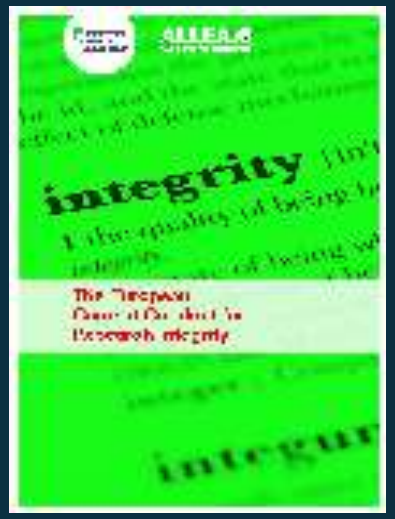
"As deputy director of the RIKEN Center for Developmental Biology, Sasai supervised the work of lead author Haruko Obokata. The work took the world of of molecular *biology* by storm when it was published in the British journal Nature in January."

"Last week, Japan's prestigious Riken institute said the 30-year-old Dr. Obokata's research at Riken, which had seemed to offer a groundbreaking way of making stem cells easily, contained basic errors and wasn't backed up by laboratory notes. Dr. Obokata rejected the conclusion, saying the errors were made without ill intent, and said she planned to appeal the findings."

http://blogs.wsj.com/japanrealtime/2014/04/07/japan-stem-cell-researcher-obokata-is-hospitalized/

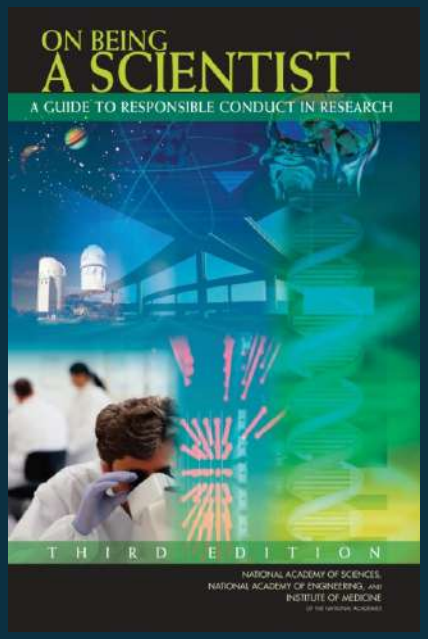
"Sasai's team retracted the research papers from the journal Nature over Obokata's alleged malpractice, which she has contested. Retractions of papers in major scientific journals are extremely rare."

RESEARCH ETHICS - ESF PUBLICATIONS



- The European Science Foundations Code of Conduct for Research Integrity http://www.esf.org/
- European Peer Review Guide – Integrating Policies and Practices into Coherent Procedures
- Fostering Research Integrity in Europe http://www.oeawi.at/downloads/ESF-research-integrity-report.pdf
- http://www.oeawi.at/en/links.asp

ON BEING A SCIENTIST

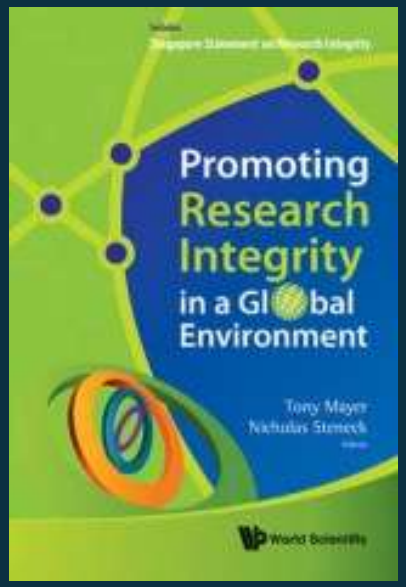


ON BEING A SCIENTIST
A GUIDE TO RESPONSIBLE CONDUCT IN RESEARCH

Committee on Science, Engineering, and Public Policy
NATIONAL ACADEMY OF SCIENCES, NATIONAL ACADEMY OF ENGINEERING, AND INSTITUTE OF MEDICINE OF THE NATIONAL ACADEMIES
THE NATIONAL ACADEMIES PRESS
Washington, D.C.

http://www.nap.edu/catalog.php?record_id=12192

PROMOTING RESEARCH INTEGRITY IN A GLOBAL ENVIRONMENT



http://www.worldscientific.com/worldscibooks/10.1142/8102 Promoting Research Integrity in a Global Environment

Sample Chapters:

Introduction

http://www.worldscientific.com/doi/suppl/10.1142/8102/suppl_file/8102_intro.pdf

Section II: Research Integrity Structures

http://www.beck-shop.de/fachbuch/leseprobe/9789814340977_Excerpt_001.pdf

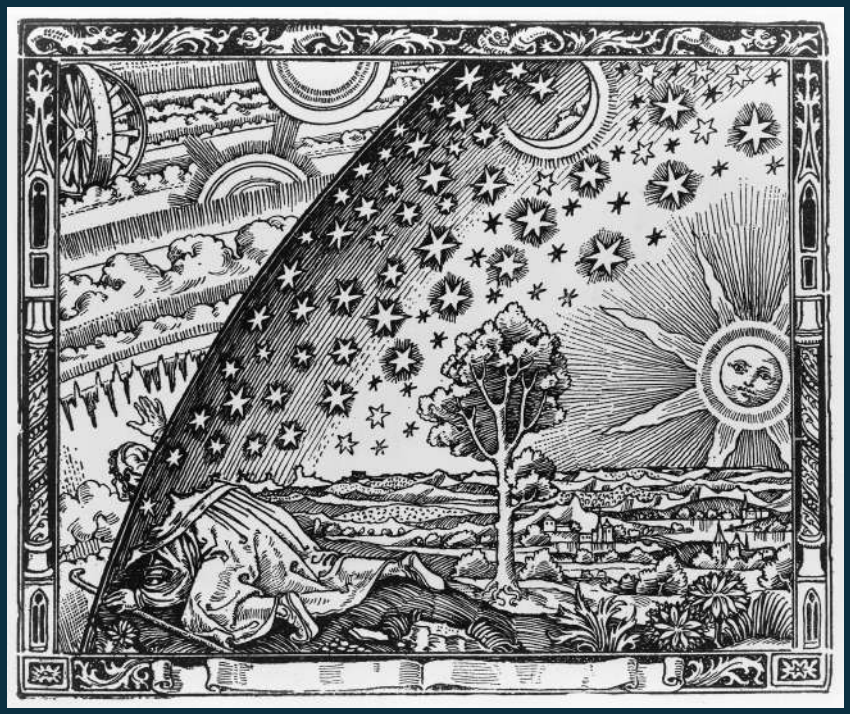
Section III: Research Misconduct

http://www.worldscientific.com/doi/suppl/10.1142/8102/suppl_file/8102_chap14.pdf

http://books.google.de/books?id=q3TZu6sovJMC&printsec=frontcover&hl=de&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Promoting Research Integrity in a Global Environment

IS THERE ANYTHING BEYOND COMPLIANCE?

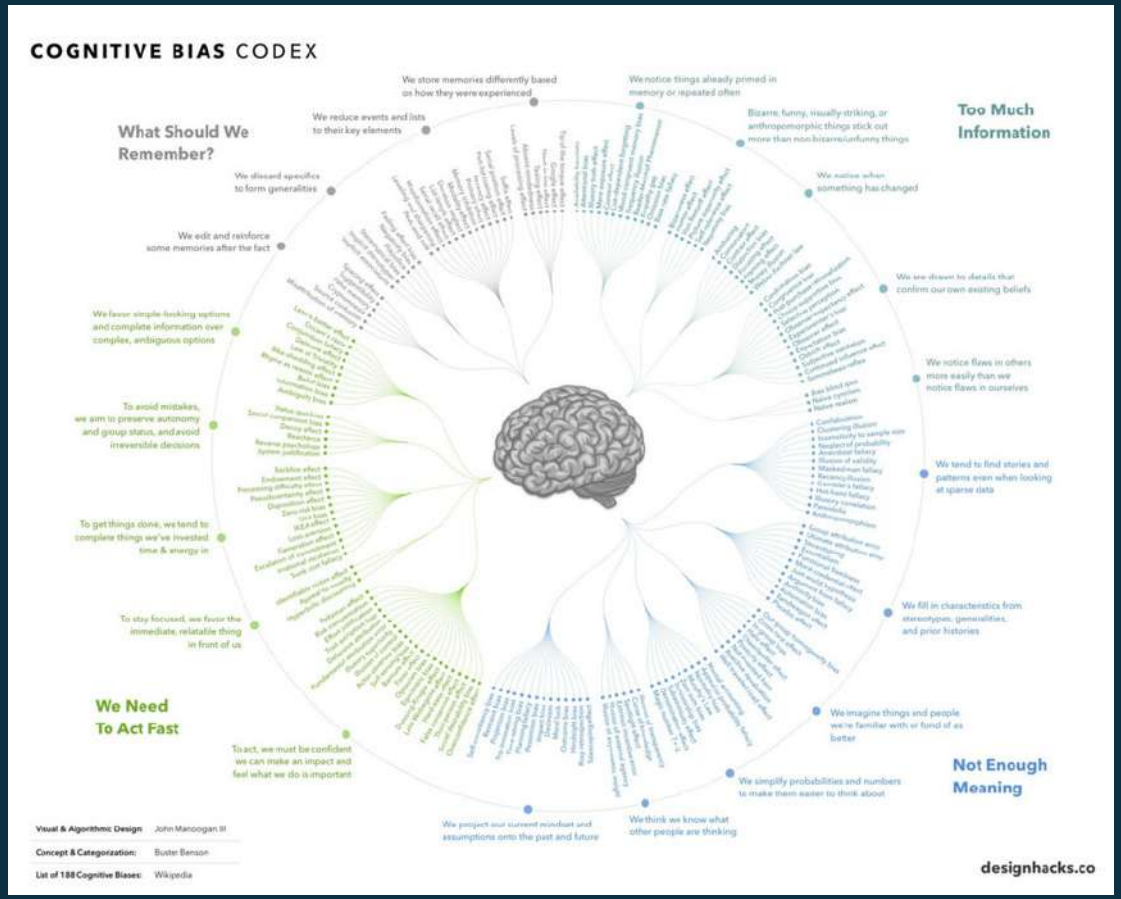


A traveller peers through an opening in the firmament in this illustration from Camille Flammarion's L'atmosphère : météorologie populaire (Paris: Hachette, 1888), p. 163

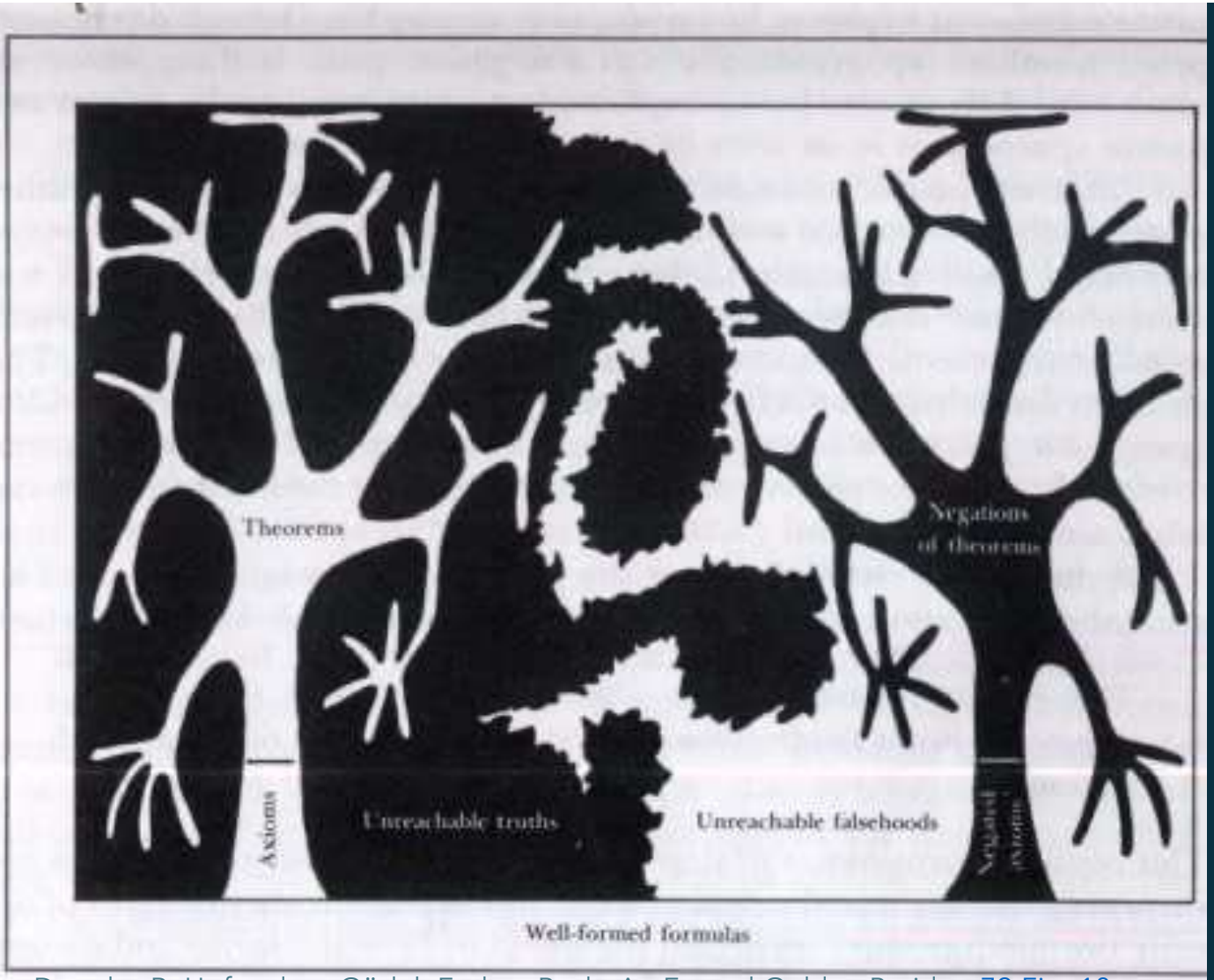
HOW IS ETHICS DONE: STEPS IN ETHICAL ANALYSIS

1. WHAT IS THE CASE?
   Taking into account that our knowledge is limited/imperfect.
   How can we interpret it in the context?
2. STAKEHOLDERS & OBLIGATIONS
3. RELATED NORMATIVE SYSTEMS
4. OPTIONS & CONSEQUENCES
5. ASSESSMENT OF THE DECISION & ACTION

HUMAN COGNITIVE BIASES



https://www.visualcapitalist.com/wp-content/uploads/2021/08/all-188-cognitive-biases.html



Douglas R. Hofstadter: Gödel, Escher, Bach: An Eternal Golden Braid, p.79 Fig. 18

LIMITS OF NORMATIVE/ AXIOMATIC SYSTEMS

Normative (rule-based) systems have a finite range
Distinguishing true from false is not always simple.

CONTRIBUTIONS - BEYOND COMPLIANCE

(1) THE MEANINGS AND SCOPES OF ETHICS

Ethics of a specific domains (AI Ethics, Healthcare Ethics, Business Ethics, Design Ethics, Engineering Ethics, Research Ethics, etc.

Ethics at a specific level (Theoreticsl Ethics, Practical Ethics, Applied Ethics, Meta-Ethics,…)

Aspects of ethics: Cognitive, Logical, Organizational, Socio-political…

ETHICS AS A PHENOMENON

"ETHICS THEATHER" = simulacrum of "ethicity" as a ritual without genuine interest or engagement. Often used for entirely different purposes. Related to:

"Ethical whitewashing" (or "ethical window dressing") - the practice of making 'misleading claims … or implementing superficial measures' in order to defend one's practice and thereby make it appear more ethical than it is in reality (Floridi 2019: 186).

Floridi, L. (2019). Translating principles into practices of digital ethics: Five risks of being unethical. Philosophy & Technology 32:2, 185-193. https://doi.org/10.1007/s13347-019-00354-x

(2) PROACTIVE RESEARCH ETHICS
ANTICIPATION, LEARNING, RESPECT & TRUST

- Ethics of different research fields – what research fields are addressing and how they are doing that – Information Ethics, Computing ethics, Bioethics, Medical ethics
- Ethics of researchers – codes of conduct – virtues and utility http://ec.europa.eu/research/participants/data/ref/fp7/89888/ethics-for-researchers_en.pdf
- Study of research ethics itself – can cognitive science help us understand ethics better? http://www.iep.utm.edu/m-cog-sc/
- Can Ethics be defended by unethical means?

(3) BEYOND HUMAN DECISION-MAKING:
DELEGATING RESPONSIBILITIES TO AUTONOMOUS AIs

- AI assistants as autonomous intelligent agents
- With autonomy and intelligence comes the necessity of ethics by design.*
- Artificial Intelligence needs ART (Accountability, Responsibility, Transparency)
- Two-step process proposal sub-symbolic level based on deep learning which in the next step is analyzed by a symbolic-level AI that can compare it with the regulations, rules and best practices.

Dodig-Crnkovic G. and Çürüklü B., Robots - Ethical by Design, Ethics and Information Technology 2011, Volume 14, Number 1, pp. 61-71.
https://dl.acm.org/doi/abs/10.1007/s10676-011-9278-2, https://link.springer.com/article/10.1007/s10676-011-9278-2,

4) AI-GENERATED WEB CONTENT:
AUTONOMOUS AI DECISION-MAKING IN THE BACKGROUND

- Articles and Blog Posts
- Product Descriptions
- News Articles, Automated Journalism, Data Journalism
- Social Media Posts
- Video and Audio Scripts and Captions
- Chatbots and Customer Support

US elections: "Perplexity dove into real-time election tracking" – with additional guardrails to avoid fabrication ("hallucination") and partial information/misinformation. ChatGPT refused to comment elections.

In the first step, AI is trained on the content produced by humans.

In the next step, an increasing amount of content will be produced by AI and used to train AI.

AI ethics – beyond piecemeal evaluations: comprehensive threat surface analysis for advanced AI systems – Upcoming Chalmers AI Ethics Seminar 26/11

"Current AI risk assessment methods often fall short in addressing the full spectrum of potential threats posed by advanced AI systems. This talk will explore the need for a paradigm shift towards comprehensive threat surface analysis, drawing on techniques like Probabilistic Risk Assessment (PRA)."

Anna Katariina Wisakanto, the Center for AI Risk Management & Alignment (CARMA)

Reference: https://link.springer.com/content/pdf/10.1038/s41598-023-34622-w.pdf?pdf=core Bernd Carsten Stahl- Embedding responsibility in intelligent systems: from AI ethics to responsible AI ecosystems

Thoughts related to the contributions

CULTURE, EDUCATION, RESPONSIBLE INNOVATION

EMPOWERMENT OF STAKEHOLDERS

UNCERTAINTIES, INCOMPLETE KNOWLEDGE, FINITE RESOURCES

WOKEISM AND “ETHICAL THEATHER”

EXERCISING IMAGINATION MUSCLE

LOOKING BEYOND WHAT WE CAN SEE IN FRONT OF OUR EYES IN THIS MOMENT

THE FASCINATION OF TESTIMONY

Thoughts related to the contributions

MISCONCEPTIONS ABOUT AUTONOMOUS CARS ETHICS - Trolley problem irrelevance

CENTRAL REFERENCE: Awad, E., Dsouza, S., Kim, R. et al. The Moral Machine experiment. Nature 563, 59–64 (2018). https://doi.org/10.1038/s41586-018-0637-6

“decentre the human and take the nonhuman seriously [4].”

COMMON MISCONCEPTION OF HUMANISM: IT IS NOT HUMAN VS. NATURE,

IT IS HUMAN (AS NATURAL BEING) VS. SUPERNATURAL

PERSPECTIVE, SCALE

Thoughts related to the contributions

WOKEISM

The term is used to refer to anything which is perceived to be closely connected to left-wing politics, with emphasis on actions or ideas believed to threaten freedom of speech. The term was first popularized in 2021 when Elon Musk began using it on Twitter.

“There was also an awareness that ethical issues can sometimes be used competitively, with researchers trying to demonstrate greater ethical awareness or "wokeness" than others.”

Thoughts related to the contributions

"**Security theater**" has been defined as an effort to "provide the feeling of security instead of the reality."

"a parallel category of "**health theater**" picks out a set of practices in medical screening and health care delivery that provide a mere simulacrum of protection against medical risk, rather than providing genuine medical benefit."

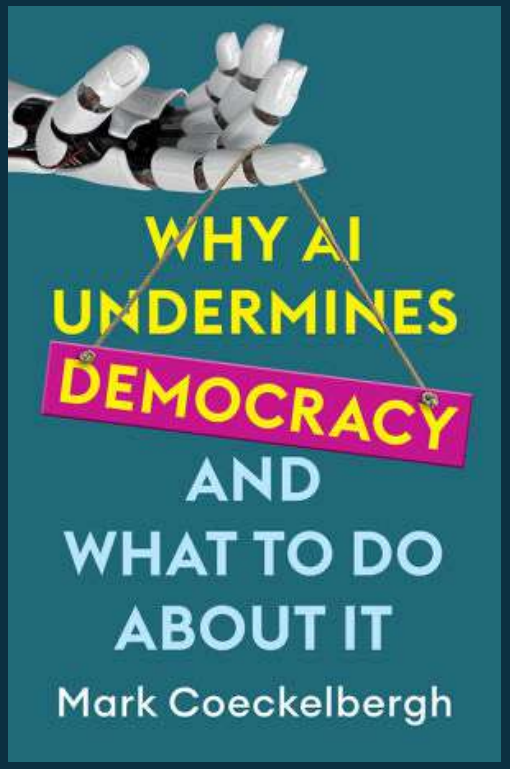
https://lawecommons.luc.edu/luclj/vol48/iss2/11/ Health Theater- Govind Persad, Loyola University School of Law

One more parallel: "**Ethics theater**"

CULTURE VS. STRATEGY

LITERATURE

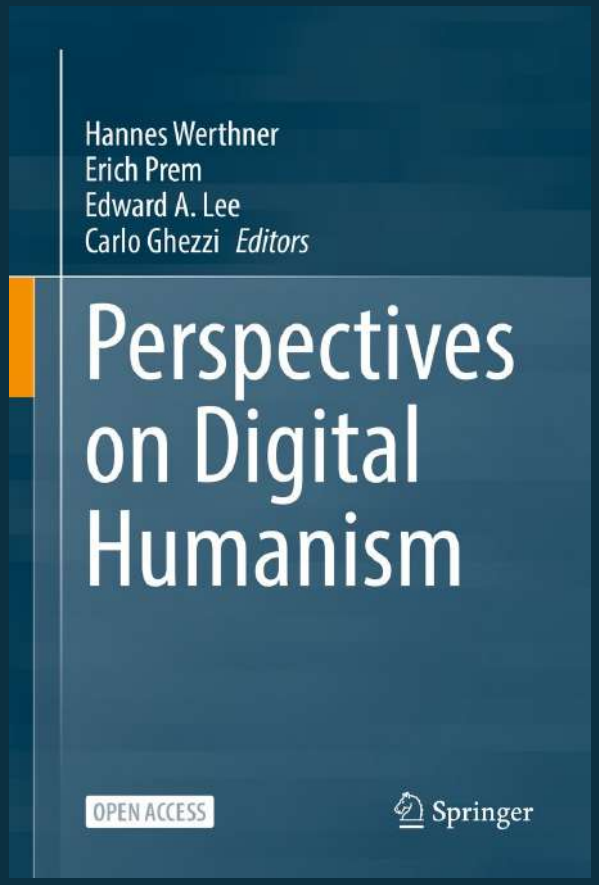
The Political Philosophy of AI: An Introduction



Mark Coeckelbergh (2022) The Political Philosophy of AI: An Introduction
ISBN: 978-1-509-54855-2

Table of Contents

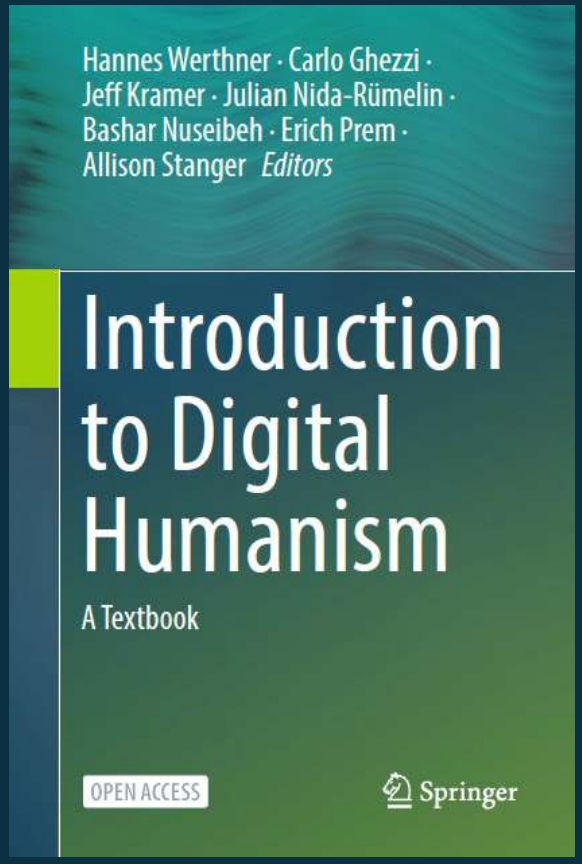
Perspectives on Digital Humanism – Open Access



Hannes Werthner, Erich Prem, Edward A. Lee, and Carlo Ghezzi (eds): **Perspectives on Digital Humanism**, Springer, 2022.
https://link.springer.com/book/10.1007/978-3-030-86144-5

Introduction to Digital Humanism – A Textbook

Open Access



Hannes Werthner, Carlo Ghezzi, Jeff Kramer, Julian Nida-Rümelin, Bashar Nuseibeh, Erich Prem, and Allison Stanger (eds): **Introduction to Digital Humanism**, Springer, 2024.
https://link.springer.com/book/10.1007/978-3-030-86144-5

REFERENCES

https://deepmind.com/about/ethics-and-society GOOGLE DEEP MIND Ethics & SOCIETY

https://framtidsprao.trr.se/documents/Framtidens_arbetsliv_rapport_WEB.pdf

https://www.youtube.com/watch?v=RXCqKwMHpb0 Ethics of AI @ NYU: Opening & General Issues (1:23:30 - Yann LeCun "Should We Fear Future AI Systems?")

https://www.youtube.com/watch?v=1oeoosMrJz4 AI ethics and AI risk - Ten challenges

https://futureoflife.org/ai-principles/ Asilomar Principles

https://www.microsoft.com/en-us/ai/ai-for-good AI for Earth, AccessibilityHumanitarian Action, Cultural Heritage

https://www.partnershiponai.org PARTNERSHIP ON AI to benefit humanity
Started by Microsoft, Amazon, Google, Facebook, IBM, and Google-owned DeepMind. 2019: 90+ partners, >50% non-profit, 13 countries

https://link.springer.com/book/10.1007/978-3-030-69978-9 Bernd Carsten Stahl- Artificial Intelligence for a Better Future – Open access

https://link.springer.com/content/pdf/10.1038/s41598-023-34622-w.pdf?pdf=core Bernd Carsten Stahl- Embedding responsibility in intelligent systems: from AI ethics to responsible AI ecosystems

REFERENCES

Friedman, B.; Kahn, P.H., Jr. (2003) Human values, ethics, and design. In The Human-Computer Interaction Handbook, Fundamentals, Evolving Technologies and Emerging Applications; Jacko, J.A., Sears, A., Eds.; Lawrence Erlbaum: Mahwah, NJ, USA; pp. 1177–1201.

Friedman, B.; Kahn, P.H., Jr.; Borning, A. (2006) Value sensitive design and information systems. In Human-Computer Interaction in Management Information Systems: Applications; M.E. Sharpe, Inc.: New York, NY, USA; Volume 6, pp. 348–372.

IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. Ethically Aligned Design, Version One – For Public Discussion (2016) A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v1.pdf

IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design, Version 2 for Public Discussion (2017) A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems. Available online: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf

REFERENCES

Asilomar Conference 2017. Asilomar AI Principles. Available online: https://futureoflife.org/ai-principles/?cn-reloaded=1

European Group on Ethics in Science and New Technologies (2018) Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems. Available online: https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

European Commission's High-Level Expert Group on Artificial Intelligence. Draft Ethics Guidelines for Trustworthy AI (2019) Available online: https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai

Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. (2018) AI4People—An Ethical Framework for a Good AI Society. Minds Mach. 28, 689–707. https://link.springer.com/article/10.1007%2Fs11023-018-9482-5

All links accessed on 16 June 2024

http://www.gordana.se/work/presentations.html

BEYOND COMPLIANCE

https://www.ercim.eu/beyond-compliance Beyond Compliance, conference
Forum on Digital Ethics in Research, October 17/18, 2022
Institute Imagine, Paris and online

Organizers

Inria - National Institute for Research in Digital Science and Technology https://www.inria.fr/fr)

ERCIM - the European Research Consortium for Informatics and Mathematics https://www.ercim.eu/

CCNE - the National Ethical Consultative Committee for Life Sciences and Health
https://www.ccne-ethique.fr/en

BEYOND COMPLIANCE: TRANSFORMING RESEARCH CULTURE

Emily E. Anderson, PhD, MPH

Associate Professor

Neiswanger Institute for Bioethics

Loyola University Chicago Stritch School of Medicine

Director, Regulatory & Bioethics, University of Illinois at Chicago, Center for Clinical and Translational Science

https://www.youtube.com/watch?v=Fu6xpzBBkFI

COMPLIANCE: NECESSARY BUT NOT SUFFICIENT

- No rules
- Gray areas
- Rapidly changing science
- Changing norms
- Unpredictability

WHAT CONTRIBUTES TO NON- COMPLIANCE? (BAD DECISIONS)

Myths

- It's just the "bad guys"
- Doing more is always better
- Knowing the science is enough
- Regulations are a punishment for all due to the mistakes of a few (FASEB)

DuBois, et al., 2016 "Lessons from Researcher Rehab"
https://www.nature.com/articles/534173a

Reality

- New, unfamiliar territory
- Complex situations
- Emotions and stress - overload!

DuBois, Anderson, et al., 2013
"Understanding Research Misconduct"
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3805450/pdf/gacr20_320.pdf

COMPLIANCE

INCREASING BURDEN, LOWER PRIORITY, “RITUAL PERFORMANCE"?

National Science Bord, 2014 – Reducing investigators’ administrative workload for federally funded research:

“The past two decades have witnessed increasing recognition that the administrative workload placed on federally funded researchers at U.S. institutions is interfering with the conduct of science in a form and to an extent substantially out of proportion to the well-justified need to ensure accountability, transparency, and safety.”

RESEARCH CULTURE

How is research incentivized?

What is prioritized? What is valued?

What is rewarded? What is ignored? What is punished?

What support is provided? What infrastructure exists?

Are resources limited? Is access to resources equitable?

Do researchers feel comfortable admitting mistakes?

Saying “I don’t know”? Asking for help?

Antes, A. (2018) “First law of leadership: be human first, scientist second,” Nature 563, 601
https://www.nature.com/articles/d41586-018-07530-7

HOW DO WE TRANSFORM RESEARCH CULTURE?

Three domains:

- Researcher education, socialization, and ongoing professional development
- Institutional structures
- Across science

TRANSFORMING EDUCATIONAL FRAMING: CHANGING HOW RESEARCHERS ARE SOCIALIZED

"Compliance helps me stay out of trouble":

- "knowing the rules is enough
- "ethics is arbitrary"
- "others will make policy, decisions about ethics"

"Compliance is a virtue of any good researcher"

"RCR* is life-long learning"

- "Developing skills that help me make decisions in the gray areas is as important as any data analysis technique"
- "I will engage in oversight locally and policy-making and developing ethical guidance for my discipline/field"

*RCR = Responsible Conduct of Research

TRANSFORMING EDUCATION: ADDITIONAL CONTENT

- Leadership skills/styles and relationship building
- Stress management, Emotional control
- Reflection on personality traits and personal biases

Antes, A. (2018) "First law of leadership: be human first, scientist second," Nature 563, 601
https://www.nature.com/articles/d41586-018-07530-7
DuBois, et al. (2016) "Lessons from Researcher Rehab" https://www.nature.com/articles/534173a
DuBois, Anderson, et al. (2013) "Understanding Research Misconduct"
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3805450/pdf/gacr20_320.pdf

TRANSFORMING INSTITUTIONAL STRUCTURES

"When researchers lack knowledge of technical matters they frequently turn to colleagues or the literature to find answers – why do they not do the same with questions about compliance? Why is the investigator not taking time to pay attention to the details?" (DuBois, Chibnall, & Gibbs, 2016, Science & Engineering Ethics)

"Pause and discussion" are evidence-based strategies

Can be operationalized: Research ethics consultation services

"Embedded" ethics/ethicists

DuBois, Chibnall, & Gibbs (2016) Compliance Disengagement in Research: Development and Validation of a New Measure Science & Engineering Ethics, pp. 965 – 988 https://link.springer.com/journal/11948/volumes-and-issues/22-4

DuBois, James M. PhD, DSc; Kraus, Elena M.; Mikulec, Anthony A. MD, MBA; Cruz-Flores, Salvador MD, MPH; Bakanas, Erin MD, MA. A Humble Task: Restoring Virtue in an Age of Conflicted Interests. Academic Medicine: July 2013 - Volume 88 - Issue 7 - p 924-928 doi: 10.1097/ACM.0b013e318294fd5b

https://journals.lww.com/academicmedicine/Fulltext/2013/07000/A_Humble_Task__Restoring_Virtue_in_an_Age_of.17.aspx

TRANSFORMING SCIENCE: WHAT DO WE VALUE ?

The Hong Kong Principles for Assessing Researchers: Fostering research integrity

- Implement more responsible metrics
- Value complete reporting
- Reward openness
- Acknowledge a broader range of research activities (dissemination)
- Recognize essential tasks (mentoring, peer review)

Moher D, Bouter L, Kleinert S, Glasziou P, Sham MH, Barbour V, et al. (2020) The Hong Kong Principles for assessing researchers: Fostering research integrity. PLoS Biol 18(7): e3000737. https://doi.org/10.1371/journal.pbio.3000737 https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000737

TRANSFORMING WORKING TOOLS - EXAMPLES - Gordana Dodig-Crnkovic

Automated Proofreader and Plagiarism Checker

- http://www.grammarly.com/?q=plagiarism&utm_source=google&utm_medium=cpc&utm_campaign=SCM&utm_content=28349967846&utm_term=&matchtype=&placement=www.univie.ac.at&network=d&gclid=CJSkgN7io70CFa3LtAod-FgAAQ

- GRAMMARLY is an automated proofreader and plagiarism checker. It corrects up to 10 times as many mistakes as other word processors.

DEVELOPING PEER REVIEW

- http://journal.frontiersin.org/Journal/10.3389/fncom.2012.00079/full
- Open evaluation: a vision for entirely transparent post-publication peer review and rating for science

- Nikolaus Kriegeskorte
- Medical Research Council, Cognition and Brain Sciences Unit, Cambridge, UK
- - See more at: http://journal.frontiersin.org/Journal/10.3389/fncom.2012.00079/full#sthash.2nhOvvq0.dpuf

ARTICLE GENERATORS

SciNote Can Write a Draft of Your Scientific Manuscript Using Artificial Intelligence

- **SciNote Can Write a Draft of Your Scientific Manuscript Using Artificial Intelligence**
  https://www.scinote.net

Copymatic

Generate Content & Copy In Seconds with AI
Use AI to boost your traffic and save hours of work. Automatically write unique, engaging and high-quality copy or content: from long-form blog posts or landing pages to digital ads in seconds.

https://copymatic.ai/?gclid=Cj0KCQiAveebBhD_ARIsAFaAvrF_UcAETiXu7gM8U7iKFQCeZ9IFt14uxBeoKTlV--SN9K4RapSJgXsaAsKbEALw_wcB

#1 AI Writer That Writes Articles, Essays and Long-Form Content In Seconds
https://writesonic.com/ai-article-writer-generator

Gpt Text Generator Online
https://chatgpt.com/g/g-OolQ7FMzJ-ai-text-generator-gpt

Bernd Carsten Stahl (2021) Artificial Intelligence for a Better Future,

An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies

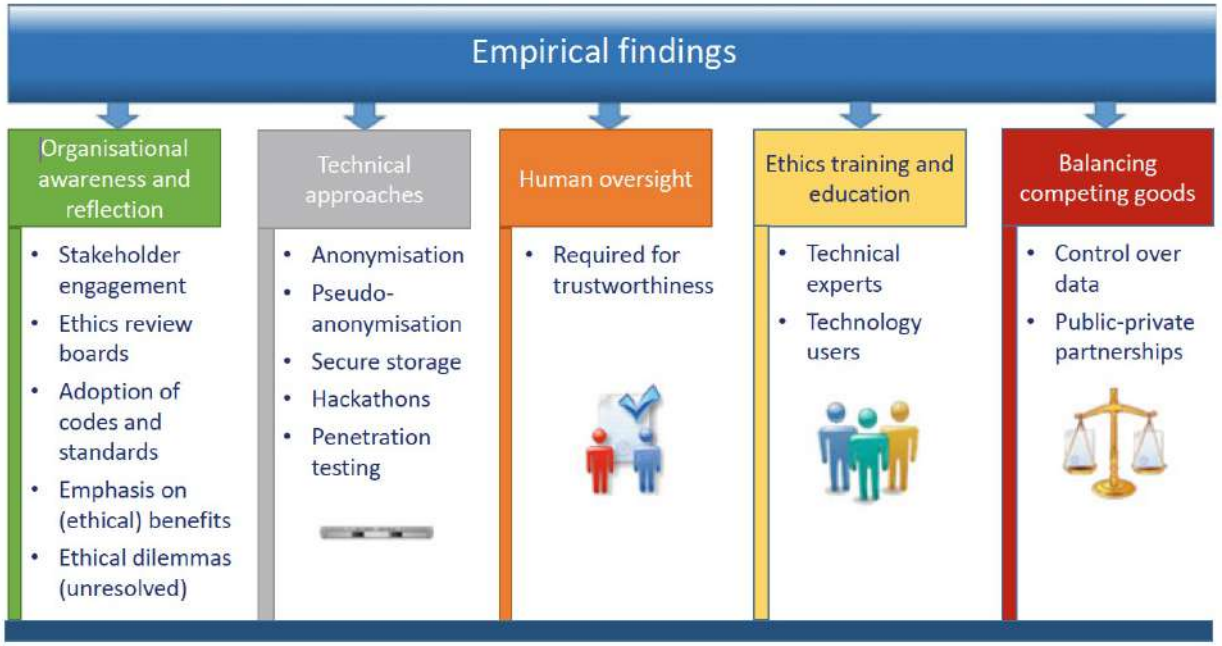


Fig. 5.2 How case study organisations address ethical issues of AI: empirical findings

Bernd Carsten Stahl (2021) Artificial Intelligence for a Better Future, An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies https://link.springer.com/book/10.1007%2F978-3-030-69978-9

Overview of AI stakeholders

Bernd Carsten Stahl (2021) Artificial Intelligence for a Better Future, https://link.springer.com/book/10.1007%2F978-3-030-69978-9 p.73

Key Challenges of Ethical Governance of AI

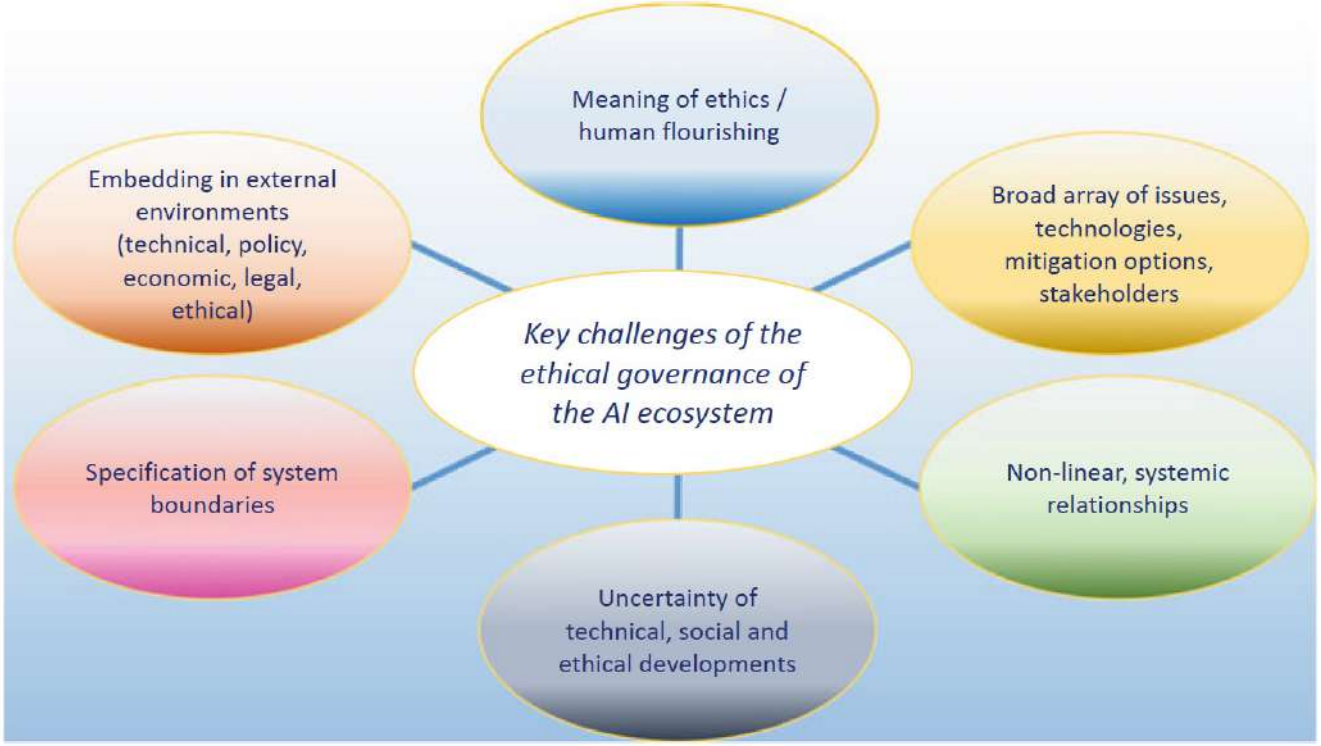


Fig. 7.1 Key challenges of ethical governance of AI ecosystems

Bernd Carsten Stahl (2021) Artificial Intelligence for a Better Future, https://link.springer.com/book/10.1007%2F978-3-030-69978-9

Practical Use of the Proposed Ethical Program for Intelligent Emergent Technologies -
Importance of Transdisciplinarity and Transversal Knowledge

Ethical requirements must be fulfilled in all phases in the life-cycle of technology, in the context of:

- Conceptualization/Design/Prototyping/ Construction/Development/Testing/Production
- Deployment/Application/
- Maintenance/Support
- Oversight/Regulation

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). In Steven John Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global

https://www.indiamart.com/kaynes-tecnology-india-pvt-ltd/product-life-cycle-management.html

A comparison: Scientists using tools for computation vs. tools for text/program/video/audio etc. generation

"It is unworthy of excellent men to lose hours like slaves in the labour of calculation which could safely be relegated to anyone else if machines were used.

(Describing, in 1685, the value to astronomers of the hand-cranked calculating machine he had invented in 1673.)"

Gottfried Wilhelm Leibniz