ETHICS OF ARTIFICIAL INTELLIGENCE

GORDANA DODIG-CRNKOVIC

MÄLARDALEN UNIVERSITY &
CHALMERS UNIVERSITY OF TECHNOLOGY

26 02 2024

Erasmus Mundus Joint Master Degree Programme
on the Engineering of Data-intensive Intelligent Software Systems

https://www.youtube.com/watch?v=GboOXAjGevA&feature=emb_logo Hewlett Packard enterprise - Moral Code: The Ethics of AI

WHAT IS ARTIFICIAL INTELLIGENCE?

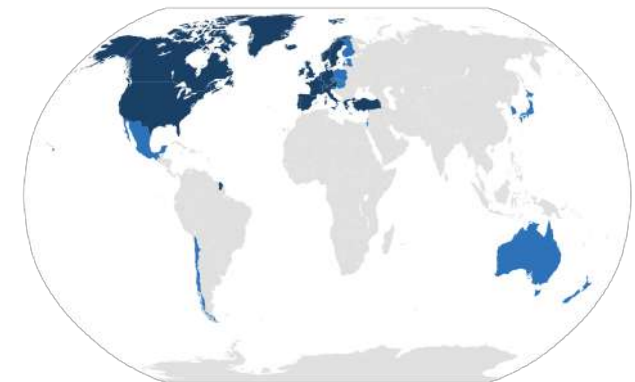Definition of AI from European AI Act:

"'Artificial intelligence system' (AI system) means "a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments"

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

This coincides with the OECD (Organisation for Economic Co-operation and Development definition:

https://mneguidelines.oecd.org/RBC-and-artificial-intelligence.pdf
https://www.biometricupdate.com/202303/eu-ai-act-definition-of-ai-aligns-with-oecd-definition-biometric-risk-updated

WHAT IS NATURAL 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."

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

TYPES OF INTELLIGENCE:
NATURAL, ARTIFICIAL, HYBRID (CYBORGS)
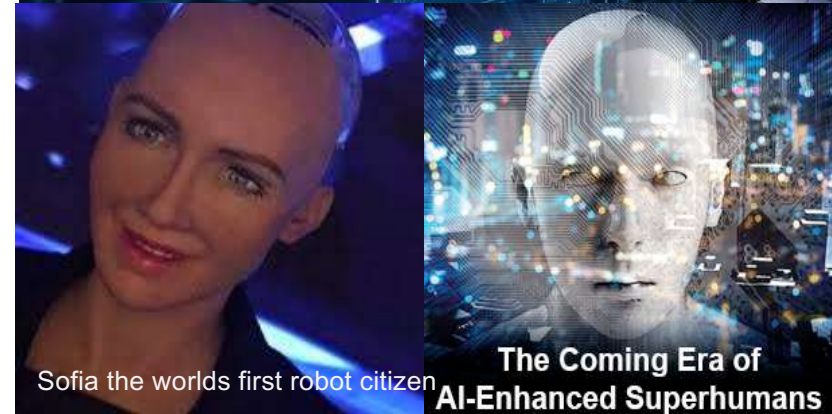
Sofia the worlds first robot citizen

The Coming Era of AI-Enhanced Superhumans

ANCIENT ROOTS OF AI IDEA

TALOS OF CRETE- the first intelligent robot &


https://taloscrafts.com/who-is-talos/

THE ANTIKYTHERA MECHANISM
ancient analog computer

Talos was a mythical bronze age (3200 to 1200 B.C.E.) giant, the first robot in history, which protected Minoan Crete from invaders. Talos was not born but made, either by Zeus himself or, according to other versions of the myth, by the Hephaestus, god of fire and iron, on Zeus's order.

Antikythera was 1st century BC analog computer, designed to calculate astronomical positions of stars and planets.
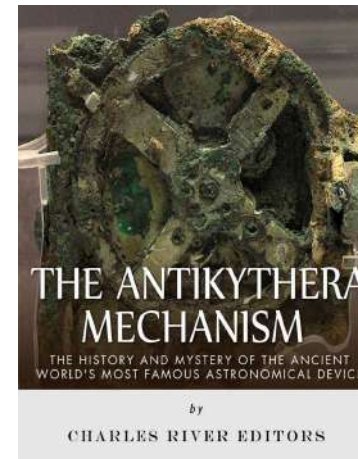
https://en.wikipedia.org/wiki/Antikythera_mechanism

TYPES OF INTELLIGENCE

EMBODIED: HUMAN, CYBORG, ROBOT

DISEMBODIED: SOFTWARE, INFRASTRUCTURE

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

NARROW/ WEAK: INTELLIGENT ARTIFACTS FOR
SPECIFIC PURPOSES

DIGITALIZATION & COGNITIZATION/INTELIGIZATION

- Digitalization happens in parallel with introduction of cognitive properties (sensors and actuators) and intelligence into artifacts

- Ethical implications of the digitization of our society: new technologies are followed by **regulatory and legal vacuums**

- The background from which I study this phenomenon: theoretical physics, computer science, philosophy of computing and ethics, computational models of cognition, with recent interest in interaction design. It has been increasingly recognized how important is the interdisciplinary approach to AI.

See V. Dignum, https://sigai.acm.org/static/aimatters/5-4/AIMatters-5-4-07-Dignum.pdf
https://www.youtube.com/watch?v=PBDnXAIX864
https://wasp-hs.org/blogposts/why-we-shouldnt-pause-research-on-ai-but-instead-prioritize-multidisciplinary-research-and-ai-governance/
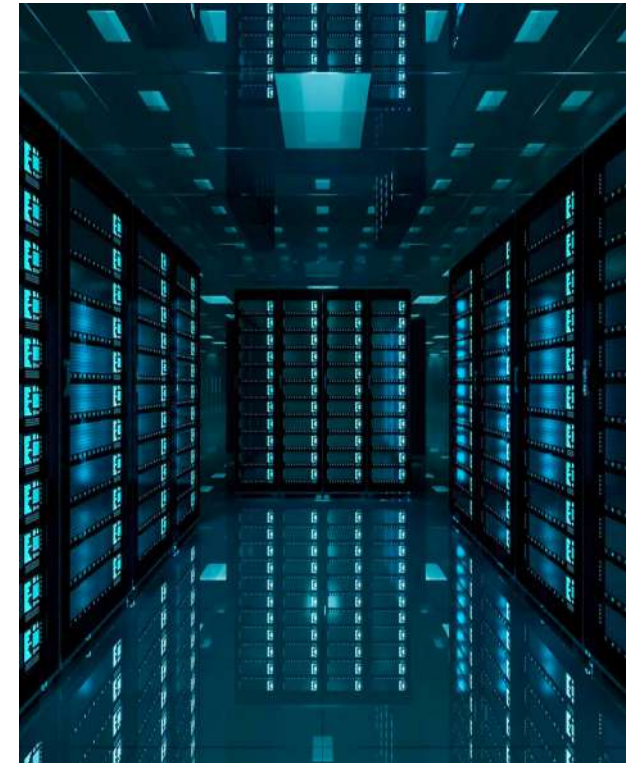
DESIGN OF INTELLIGENT ARTIFACTS

INTELLIGENT ARTIFACTS BEING DEVELOPED

- Intelligent Assistants & Co-pilots

- Ambient Intelligence

- Intelligent Infrastructure

- Intelligent Robots & Softbots

- Intelligent Transportation

- Intelligent Cities

- Intelligent IoT

- Intelligent Decision-Making Algorithms
  (introduced into particular technologies as self-driving vehicles but also into democratic institutions of governance, law, etc.)
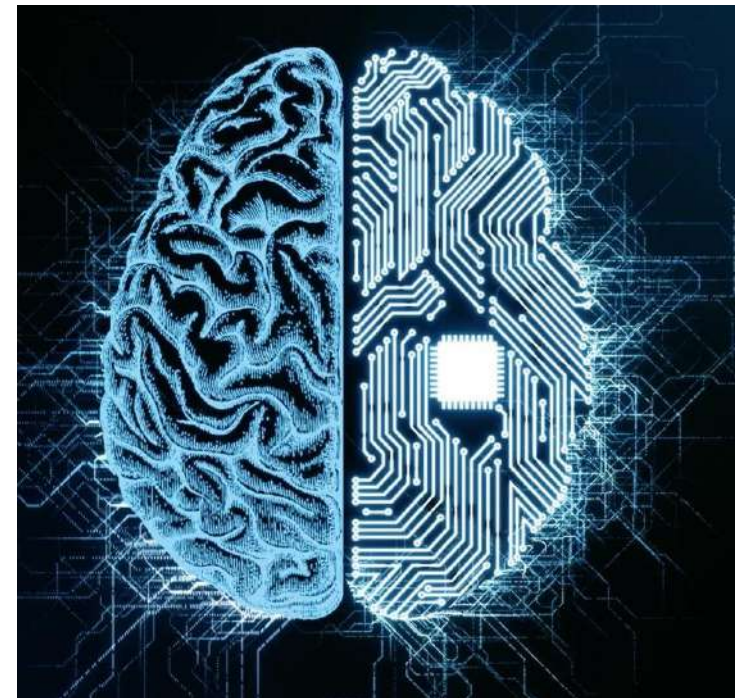
https://bitcoinist.com/crypto-mining-becoming-concern-us-cities/

DESIGN OF INTELLIGENT ARTIFACTS

- Intelligent Health-Care Systems
- Intelligent Personalized Medicine
- Cognitive Enhancements

Theodore Berger (University of Southern California, L.A.) Engineering Memories: A Cognitive Neural Prosthesis for Restoring and Enhancing Memory Function

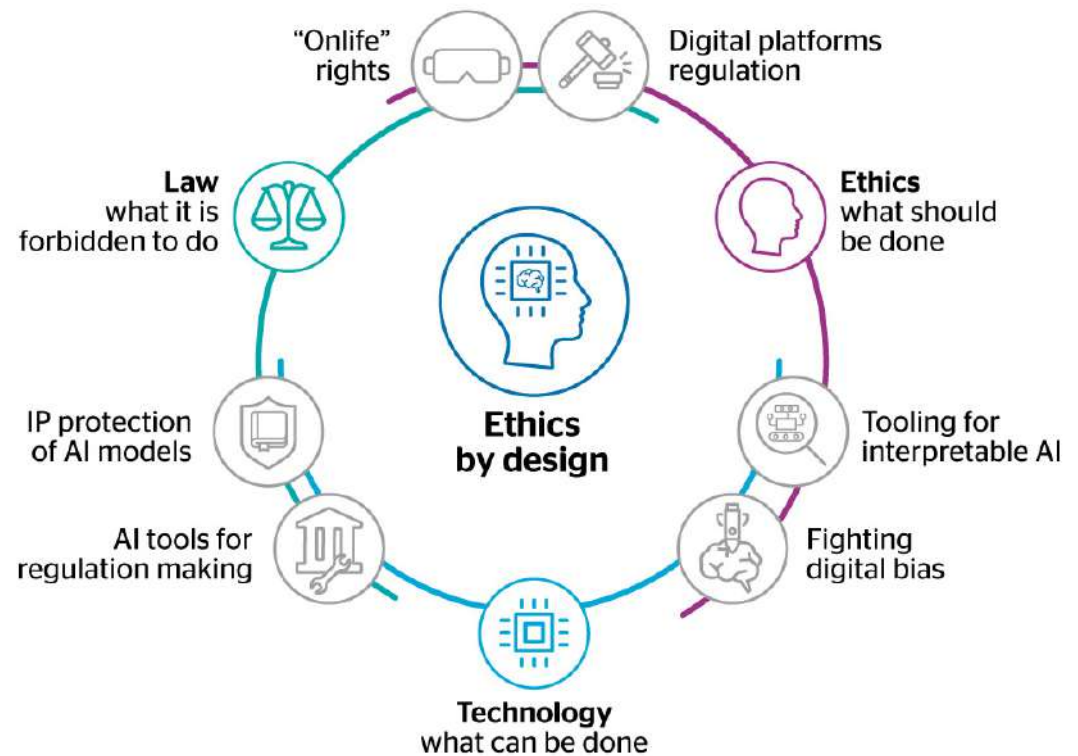https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3141091/

https://www.technologyreview.com/s/513681/memory-implants/

https://www.pbs.org/wgbh/nova/video/nova-wonders-can-we-build-a-brain/

DO WE WANT INTELLIGENT EVERYTHING?

WHY?
QUESTION OF VALUES & ETHICS

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

VALUE-SENSITIVE DESIGN – TOWARDS COMMON PRINCIPLES

Ethical values and principles in European discussion			
Expert Group/ Publication	**Ethical Value/Principle**	**Context**	**Technology**
Friedman et al. (2003; 2006) [1,2]	Human welfare Ownership and property Freedom from bias Universal usability Courtesy Identity Calmness Accountability (Environmental) sustainability	Value-sensitive design	ICT
Ethically Aligned Design (EAD) IEEE Global initiative (2016, 2017) [3,4]	Human benefit Responsibility Transparency Education and Awareness	Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems: Insights and recommendations for the AI/AS technologists and for IEEE standards	AI/AS

VALUE-SENSITIVE DESIGN – TOWARDS COMMON PRINCIPLES

Ethical values and principles in European discussion			
Asilomar AI Principles (2017) [5]	Safety Failure and juridical transparency Responsibility Values, alignment Privacy and liberty Shared benefit and prosperity Human control Non-supervision Avoiding arms race	Beneficial AI to guide the development of AI	AI
The European Group on Ethics in Science and New Technologies (EGE) (2017) [6]	Human dignity Autonomy Responsibility, Accountability Security, Safety Justice, Equality and solidarity Democracy Bodily and mental integrity Data protection and privacy Sustainability	Statement on Artificial Intelligence, Robotics and Autonomous Systems	AI, Robotics, AS

VALUE-SENSITIVE DESIGN – TOWARDS COMMON PRINCIPLES

Ethical values and principles in European discussion			
European Commission's High-Level Expert Group on Artificial Intelligence (AI HLEG) (2018) [7]	Respect for human dignity Freedom of the individual Respect for democracy, justice and the rule of law Equality, non-discrimination and solidarity Citizens rightsBeneficence: "Do Good" Non maleficence: "Do no Harm" Autonomy: "Preserve Human Agency" Justice: "Be Fair" Explicability: "Operate transparently"	Trustworthy AI made in Europe	AI
AI4People (2018) [8]	Beneficence Non-maleficence Autonomy Justice Explicability	An ethical framework for a good AI society	AI

VALUE-SENSITIVE DESIGN – TOWARDS COMMON PRINCIPLES

A preliminary set of ethical values for the context of Autonomous Intelligent Systems	
Integrity and Human Dignity	Individuals should be respected, and AIS solutions should not violate their dignity as human beings, their rights, freedoms and cultural diversity. AIS (Autonomous Intelligent Systems) should not threaten a user's physical or mental health.
Autonomy	Individual freedom and choice. Users should have the ability to control, cope with and make personal decisions about how to live on a day-to-day basis, according to one's own rules and preferences.
Human control	Humans should choose how or whether to delegate decisions to AIS, to accomplish human-chosen objectives.
Responsibility	Concerns the role of people and the capability of AIS to answer for the decisions and to identify errors or unexpected results. AIS should be designed so that their affects align with a plurality of fundamental human values and rights.

VALUE-SENSITIVE DESIGN – TOWARDS COMMON PRINCIPLES

A preliminary set of ethical values for the context of Autonomous Intelligent Systems	
Justice, equality, fairness and solidarity	AIS should contribute to global justice and equal access. Services should be accessible to all user groups irrespective any physical or mental deficiencies. This principle of (social) justice goes hand in hand with the principle of beneficence: AIS should benefit and empower as many people as possible.
Transparency	If an AIS causes harm, it should be possible to ascertain why. The mechanisms through which the AIS makes decisions and learns to adapt to its environment should be described, inspected and reproduced. Key decision processes should be transparent and decisions should be the result of democratic debate and public engagement.
Privacy	People should have the right to access, manage and control the data they generate.

VALUE-SENSITIVE DESIGN – TOWARDS COMMON PRINCIPLES

A preliminary set of ethical values for the context of Autonomous Intelligent Systems	
Reliability	AIS solutions should be sufficiently reliable for the purposes for which they are being used. Users need to be confident that the collected data is reliable, and that the system does not forward the data to anyone who should not have it.
Safety	Safety is an emerging property of a socio-technical system, which is created daily by decisions and activities. Safety of a system should be verified where applicable and feasible. Need to consider possible liability and insurance implications.
Security	AI should be secure in terms of malicious acts and intentional violations (unauthorized access, illegal transfer, sabotage, terrorism, etc.). Security of a system should be verified where applicable and feasible.
Accountability	Decisions and actions should be explained and justified to users and other stakeholders with whom the system interacts.

VALUE-SENSITIVE DESIGN – TOWARDS COMMON PRINCIPLES

A preliminary set of ethical values for the context of Autonomous Intelligent Systems	
Explicability	Also 'explainability'; necessary in building and maintaining citizen's trust (captures the need for accountability and transparency), and the precondition for achieving informed consent from individuals.
Sustainability	The risks of AIS being misused should be minimized: Awareness and education. "Precautionary principle": Scientific uncertainty of risk or danger should not hinder to start actions of protecting the environment or to stop usage of harmful technology.
Role of technology in society	Governance: Society should use AIS in a way that increases the quality of life and does not cause harm to anyone. Depending on what type of theory of justice a society is committed to, it may stress e.g., the principle of social justice (equality and solidarity), or the principle of autonomy (and values of individual freedom and choice).

REFERENCES

1. 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.

2. 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.

3. 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

4. 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

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

6. 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

7. 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

8. 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 2019

RESPONSIBLE AI

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

https://www.iconfinder.com/iconsets/brain-service-2

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, 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
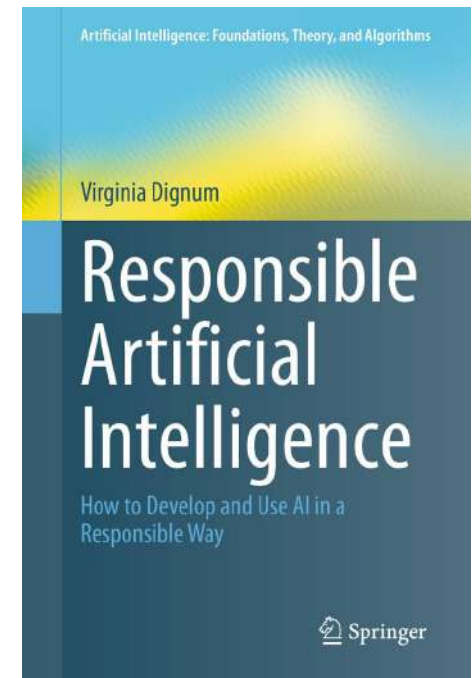
ETHICS & DESIGN

Ethics in Design

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

Ethics by Design

- o Integration of ethical <u>reasoning</u> abilities as part of the behaviour of artificial autonomous systems

Ethics for Design(ers)

- o Research integrity of <u>researchers</u> and manufacturers, and certification mechanisms

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

ETHICS IN DESIGN: AI – DOING IT RIGHT

Principles for Responsible AI = ART

Accountability

Responsibility

Transparency

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

ETHICS <u>IN</u> DESIGN: AI – DOING IT RIGHT

Principles for Responsible AI = ART

<u>A</u>ccountability
- Explanation and justification
- Design for values

<u>R</u>esponsibility

<u>T</u>ransparency

values

interpretation

norms

concretization

functionalities

Based on: Responsible Artificial Intelligence, Virginia Dignum, https://www.youtube.com/watch?v=BqwVRzKVz30
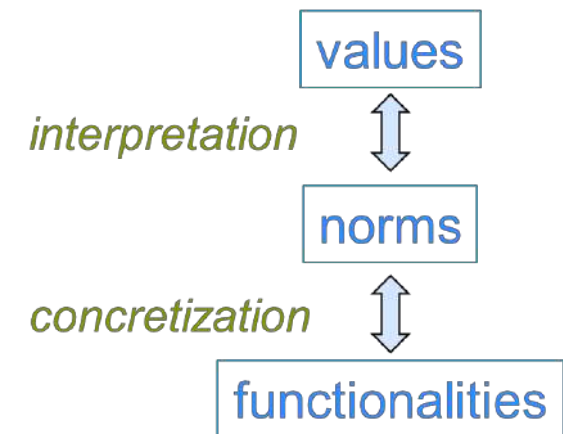
ETHICS IN DESIGN: AI – DOING IT RIGHT

Principles for Responsible AI = ART

Accountability
- Explanation and justification
- Design for values

Responsibility
- Autonomy
- Chain of responsible actors
- Human-like AI

Transparency

ETHICS <u>IN</u> DESIGN: AI – DOING IT RIGHT

Principles for Responsible AI = ART

<u>A</u>ccountability
- Explanation and justification
- Design for values

<u>R</u>esponsibility
- Autonomy
- Chain of responsible actors
- Human-like AI

<u>Transparency</u>
- Data and processes
- Algorithms
- Choices and decisions

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

ETHICS <u>BY</u> DESIGN (DOING RIGHT THING)

- Can AI artefacts be build to be ethical?

- What does that mean?

- What is needed?

- Understanding ethics

- Using ethics

- Being ethical

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

ETHICAL REASONING IS OPEN-ENDED

Normative reasoning (Trolley Problem/Moral Machine)

Utilitarian/Consequestialist car

Consequentialism in ethics is the view that whether or not an action is good or bad depends solely on what **effects** that action has on the world. "The greatest amount of good for the greatest amount of people". The only valuable consequence is pleasure, and the only disvaluable consequence is pain. The best for the most; results matter

Deontologic/Kantian car

Human-centric, Duty ethics. Good and evil reside in the individual's **intentions** rather than in the **consequences** of the act

Aristotelian car

Aristotle's ethics is about how to live the good life (eudaimonia) based on virtues. Act as a virtuous person. For Aristotle, there are three primary moral virtues: courage, temperance, and justice.
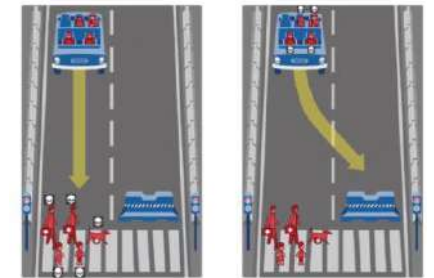
ETHICS _FOR_ DESIGN(ERS) – REGULATION, CODES OF CONDUCT

- A code of conduct clarifies mission, values and principles, linking them with standards and regulations
  - Compliance
  - Risk mitigation
  - Marketing

- Many professional groups have regulations
  - Medicine / Pharmacy
  - Accountants
  - Architects
  - Military

- Regulation, accreditation: when society relies on a profession!

Based on: Responsible Artificial Intelligence, Virginia Dignum, https://www.youtube.com/watch?v=BqwVRzKVz30
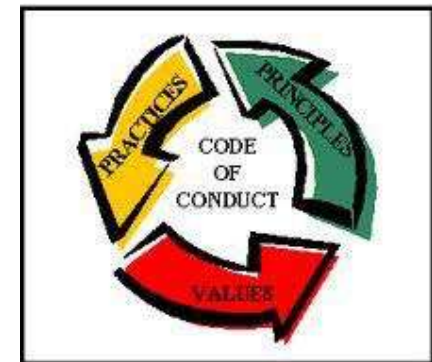
"(W)e need to ensure that we put in place the social and technical constructs that ensure responsibility and trust for the systems we develop and use in contexts that change and evolve. Obviously, the AI applications are not responsible, it is the socio-technical system of which the applications are part of that must bear responsibility and ensure trust. Ensuring ethically aligned AI systems requires more than designing systems whose results can be trusted. It is about the way we design them, why we design them, and who is involved in designing them. This is a work always in progress." (Dignum 2019) (emphasis added)

Even the United Nations Interim report, Governing AI for Humanity, published by the AI Advisory Body (UN 2023) emphasizes strongly the responsibility of humans for ethical AI, with the important first guiding principle: "AI should be governed inclusively, by and for the benefit of all".
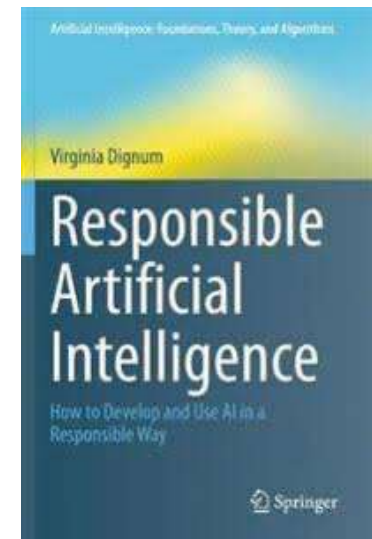
AI FOR GOOD, AI FOR PEOPLE, …

- Harness the positive potential outcomes of AI in society, the economy

- Ensure inclusion, diversity, universal benefits

- Prioritize UN2020 Sustainable Development Goals

- The objective of the AI system is to maximize the realization of human values
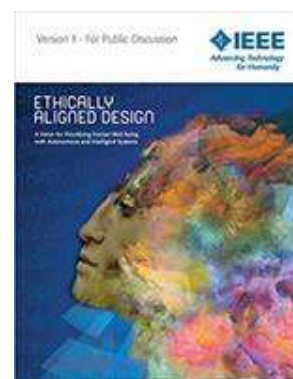
https://ai4good.org/
https://en.wikipedia.org/wiki/AI_for_Good
https://www.microsoft.com/en-us/ai/ai-for-good

https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

https://ethicsinaction.ieee.org/

http://www.ai4people.eu

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

TAKE AWAY MESSAGE ON **RESPONSIBLE AI**

- (Currently) AI systems are artefacts built by us for our own purposes
  - Our decision, our responsibility
- 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 (Stakeholders agreement needed)
- Artificial Intelligence needs ART
  - Accountability, Responsibility, Transparency
  - (Stakeholders must) Be explicit!

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

RESPONSIBLE ARTIFICIAL INTELLIGENCE

WE ALL ARE RESPONSIBLE STAKEHOLDERS
in different <u>ways</u>
in different <u>roles</u>

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

https://w.un.org.techenvoy/files/ai_advisory_body_interim_report.pdf

BLUEPRINT FOR AN
**AI BILL OF
RIGHTS**

MAKING AUTOMATED
SYSTEMS WORK FOR
THE AMERICAN PEOPLE

OCTOBER 2022
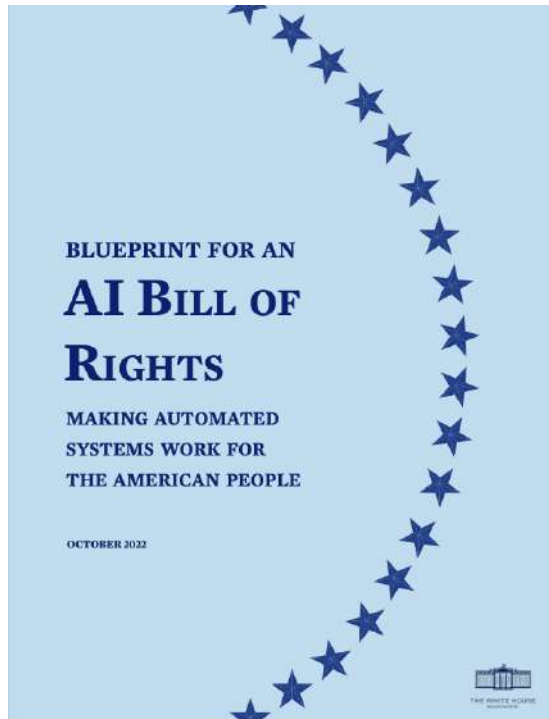
The US AI Bill of Rights outlines principles, including that people have a **right to control how their data is used** and to **not be discriminated against by unfair algorithms.**

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.

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

The world's first AI legislation

EU's "AI Act" (2024)

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

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

Figure 1—Pyramid of Risk

- 01 Prohibited
- 02 Regulated High Risk
- 03 Transparent Limited Risk
- 04 Low and Minimal Risk
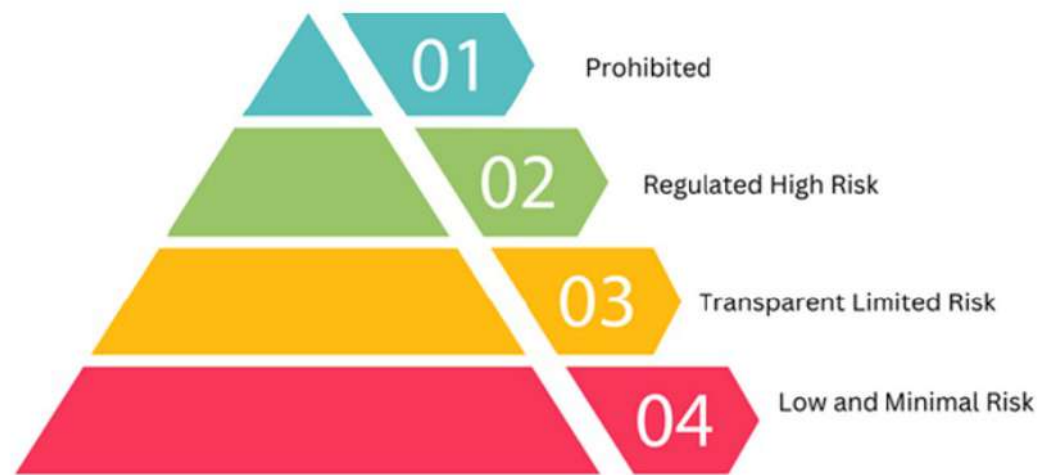
Source: ISACA

DIVISION/ASSIGNMENT OF RESPONSIBILITY

Time perspective

- Short-term perspective
  (We decide)

- Middle-term perspective
  (AGI – We co-decide)

- Long-term perspective
  (Superintelligence? Who decides?)

Stakeholders

- Politicians

- Legislators

- Business

- Developers, Designers

- Programmers

- Deployment, test

- Maintenance

- Learning from experience

- Feedback to development & design

PROFESSIONAL ETHICISTS ON AI ETHICS

Vincent C. Müller (forthcoming), 'Ethics of artificial intelligence and robotics', in Edward N. Zalta (ed.), Stanford Encyclopedia of Philosophy (Palo Alto: CSLI, Stanford University). tiny.cc/1tnvez

Judith Simon
https://www.youtube.com/watch?v=cvLltFoJme0 Judith Simon - Big data & machine learning

https://www.youtube.com/watch?v=mNhuriIZLcI Judith Simon - DataBust: Dissecting Big Data Practices and Imaginaries

Mark Coeckelbergh (2020) AI Ethics. MIT Press Essential Knowledge

Luciano Floridi & Cowls, Josh. (2019). A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review. 10.1162/99608f92.8cd550d1.
Floridi, L. The AI4People's Ethical Framework for a Good AI Society https://www.eismd.eu/wp-content/uploads/2019/03/AI4People%E2%80%99s-Ethical-Framework-for-a-Good-AI-Society.pdf

Rafael Capurro (2019) THE AGE OF ARTIFICIAL INTELLIGENCES. Contribution to the AI, Ethics and Society Conference, University of Alberta, Edmonton (Canada), May 8-10, 2019. http://www.capurro.de/edmonton2019.html

Nick Bostrom
https://intelligence.org/files/EthicsofAI.pdf The Ethics of Artificial Intelligence

SUMMARY: ETHICS OF AI

https://www.iconfinder.com/iconsets/brain-service-2

HIGHLIGHTS OF ETHICAL ISSUES OF AI

- REGULATION – value-based, codes of ethics & laws

- RESPONSIBILITY

- TRANSPARENCY

- PRIVACY & INTEGRITY. GDPR https://www.gdprexplained.eu
  General Data Protection Regulation – data protection by design

- BIAS/FAIRNESS in machine classification systems (algorithmic bias) & decision-making

- AI-guided weapon Systems - Lethal Autonomous Weapons – security & responsibility

- Agency and moral status of AI

- Future of work & end of employment - job replacement and redistribution

- Human dependency on technology and loss of skills

- Value-misalignment

- Unintended consequences of goals and decisions

EXAMPLE: UNIFIED FRAMEWORK OF PRINCIPLES FOR AI IN SOCIETY (AI4PEOPLE)

Non-maleficence: privacy, security and "capability caution"

Beneficence: promoting well-being, preserving dignity, and sustaining the planet

Autonomy: the power to decide (whether to decide)

Justice/Fairness: promoting prosperity and preserving solidarity

Explicability: enabling the other principles through intelligibility and accountability

https://www.eismd.eu/ai4people-ethical-framework/

COST OF AI & LONG-TERM SUSTAINABILITY

Four pillars of sustainability

- Environmental sustainability

- Social sustainability

- Economic sustainability

- Human sustainability

https://sustainable-ai-conference.eu/ Bonn Sustainable AI conference 2023

AI for good, International Telecommunication Union (ITU)

"As the UN specialized agency for information and communication technologies, ITU is well placed to guide AI innovation towards the achievement of the UN Sustainable Development Goals. We are providing a neutral platform for international dialogue aimed at building a common understanding of the capabilities of emerging AI technologies."

- Houlin Zhao, Secretary General of ITU

RECENT DEVELOPMENTS OF AI

https://www.youtube.com/watch?v=C_78DM8fG6E

The Inside Story of ChatGPT's Astonishing Potential | Greg Brockman | TED (20 April 2023)

https://www.youtube.com/watch?v=hJP5GqnTrNo

Amazing AI Super Tutor for Students and Teachers | Sal Khan | TED (2 May 2023)

https://www.youtube.com/watch?v=eg9qDjws_bU (April 2023) Elon Musk's Opinion on ChatGPT (3:29)

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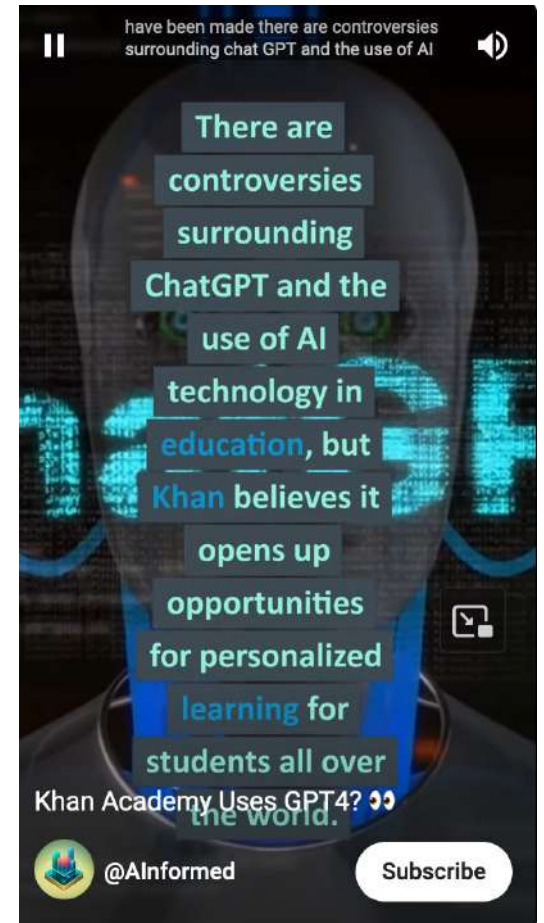
https://deepmind.com/about/ethics-and-society GOOGLE DEEP MIND Ethics & SOCIETY

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

EXAMPLE: 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
l IGI Global 2021

DISCUSSION

QUESTIONS

WHICH POSSIBLE DEVELOPMENTS OF AI CAN YOU ANTICIPATE IN THE FUTURE? SHORT TERM? LONG TERM?

WHAT ARE THE DEVELOPMENTS WITH THE BIGGEST ETHICAL CONSEQUENCES?

HOW DO YOU SEE YOUR ROLE AS AN ENGINEER IN THIS DEVELOPMENT?

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