Erasmus Mundus Joint Master Degree Programme
on the Engineering of Data-intensive Intelligent Software Systems
Winter School 2024

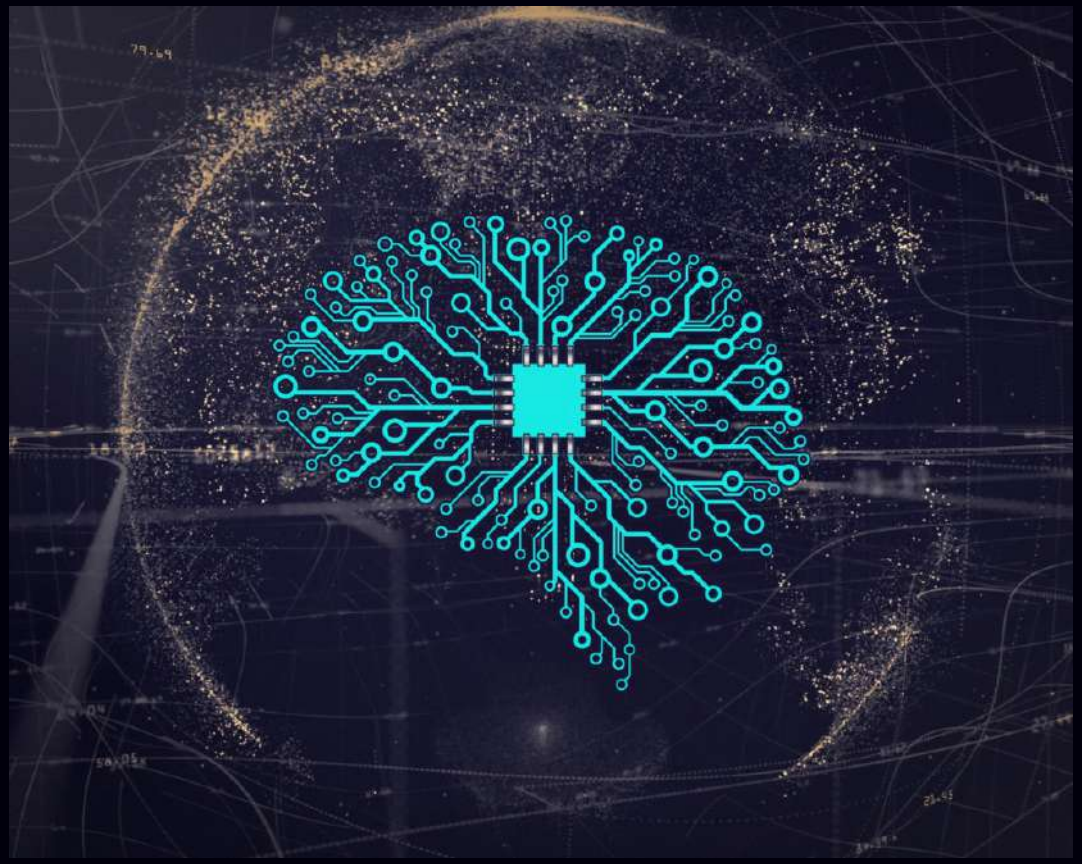
AI ETHICS

Multidisciplinarity, Responsibility, Guidelines

GORDANA DODIG-CRNKOVIC
MÄLARDALEN UNIVERSITY &
CHALMERS UNIVERSITY OF TECHNOLOGY

Västerås, 2024 02 26

http://gordana.se/Presentations



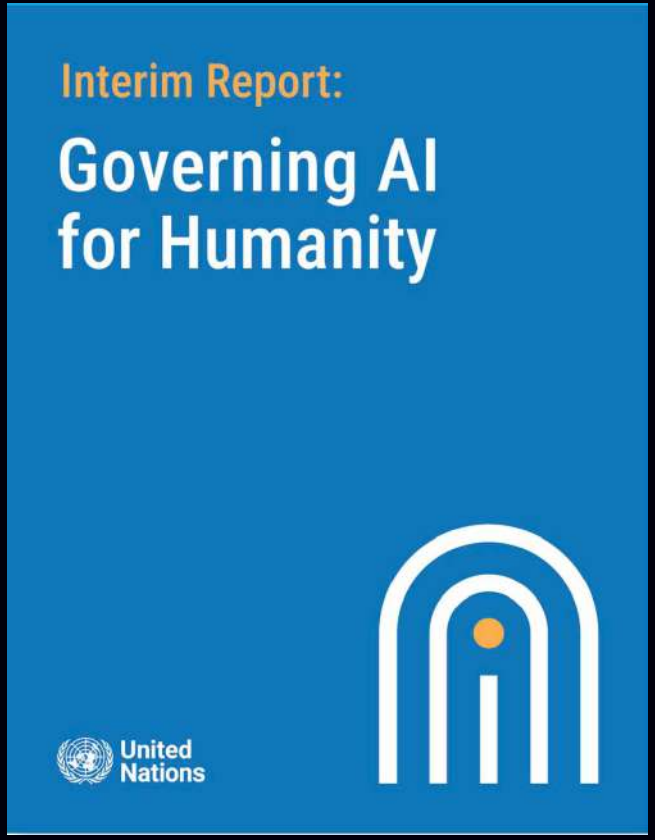
ARTIFICIAL INTELLIGENCE: RESPONSIBILITY

https://www.quantamagazine.org/artificial-intelligence-will-do-what-we-ask-thats-a-problem-20200130/Ai genie in a bottle

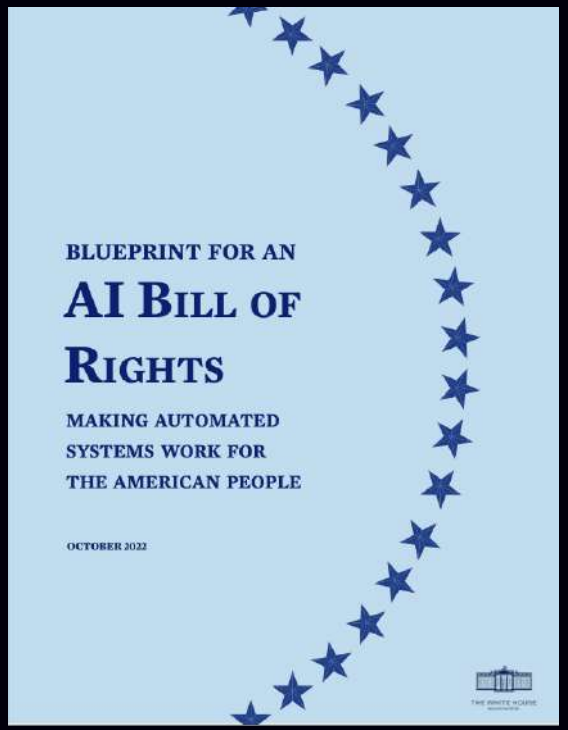
United Nations report “Governing AI for Humanity”

United Nations report (2023)
“Governing AI for Humanity”

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



US “AI Bill of Righhts”



The 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

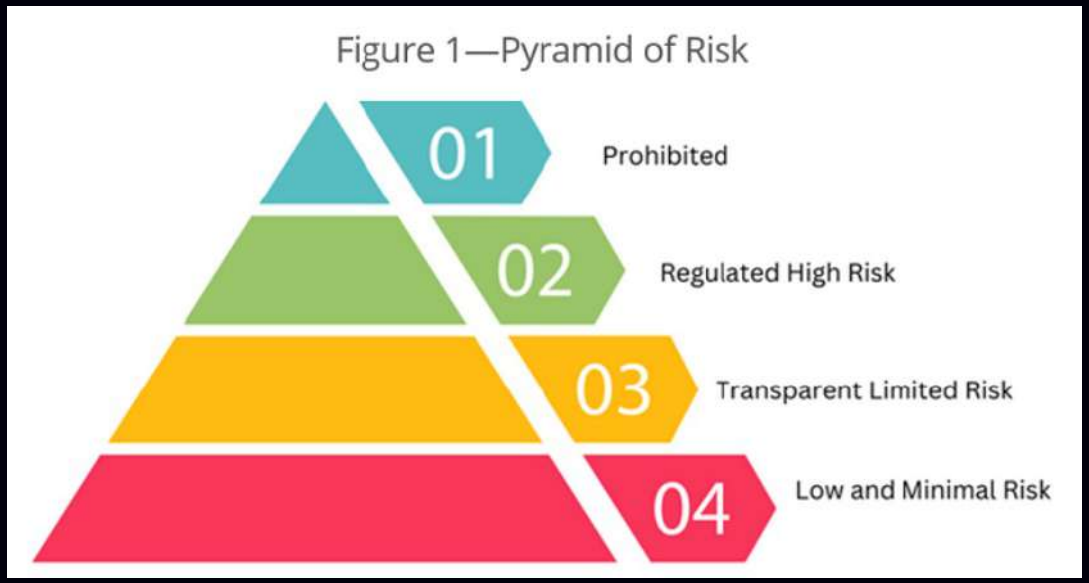
European Union’s “AI Act”

The World’s First AI Legislation

EU’s “AI Act” (2024)

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

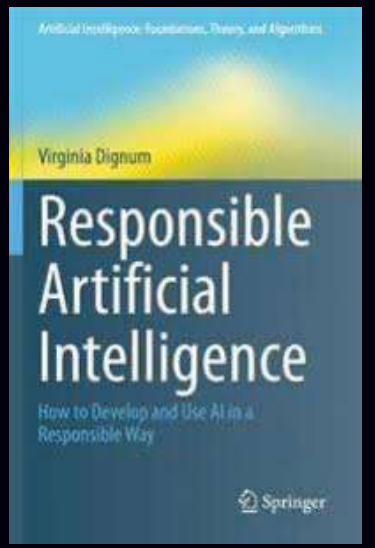
https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai



Source ISACA

RESPONSIBLE ARTIFICIAL INTELLIGENCE

"(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".

The European AI Act (EU 2024) defines an AI system as…

"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"

This coincides with the OECD's latest definition, (OECD.AI 2024).

DISTRIBUTED RESPONSIBILITY IN A SOCIO-TECHNOLOGICAL SYSTEM WITH NATURAL AND ARTIFICIAL INTELLIGENCE

The analysis of distributed responsibility in socio-technological a network of agents as presented by (Taddeo and Floridi 2018) builds on the following argument pointed out by (Basti and Vitiello 2023):

"The effects of decisions or actions based on AI are often the result of countless interactions among many actors, including designers, developers, users, software, and hardware. This is known as a distributed agency. With distributed agency comes distributed responsibility.[1, p. 751]"

Basti makes an important observation (Basti 2020) about the difference "between the slow responsibility of conscious ethical agents such as humans, and the fast responsiveness of unconscious skilled moral agents such as machines with respect to the ethical constraints from the shared social environment". (Basti and Vitiello 2023)

DISTRIBUTED RESPONSIBILITY

- Delegating tasks to machines means delegating responsibility. If an AI learns, makes decisions, and handles tasks autonomously, then it makes sense to consider it "responsible" for those tasks, similar to how we call it "intelligent." As technology affects our lives, so task "responsibility" can have moral implications. This responsibility, then, becomes "moral responsibility" worthy of consideration. (Dodig-Crnkovic 2008)

- As Dignum says, “Obviously, errors will be made, disasters will happen. More than assigning blame for these failures, we need to learn from them and try again, try better.” (Dignum 2019) The focus shouldn't be on blame, but on ensuring good behavior in the future. Moral responsibility as a "regulation mechanism" can guide the development and use of AI for societal benefit.

AI ETHICS GUIDELINES IN A GLOBAL LANDSCAPE

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

https://www.nature.com/articles/s42256-019-0088-2.pdf

ARTIFICIAL INTELLIGENCE: MULTIDISCIPLINARITY

https://www.quantamagazine.org/artificial-intelligence-will-do-what-we-ask-thats-a-problem-20200130/Ai genie in a bottle

AI & HUMAN INTELLIGENCE ENHANCEMENT TECHNOLOGIES

Artificial general intelligence approached through construction (by engineering)

Human level intelligence: Starting from the human brain, scientific approach. Cognitive enhancements. Restoring & enhancing memory. Creating artificial memories.

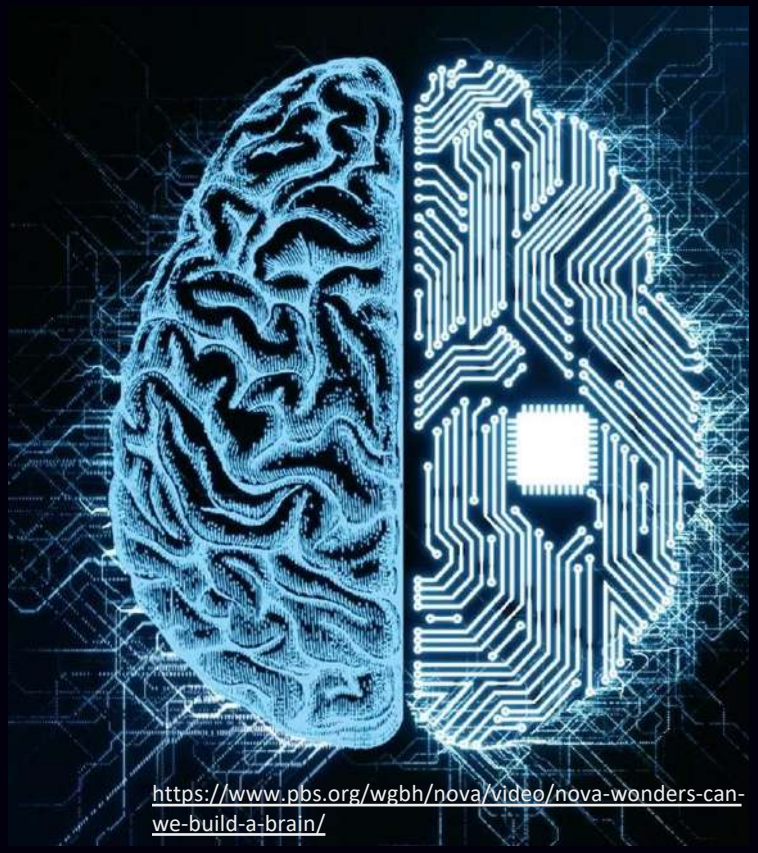
Theodore Berger (University of Southern California, L.A.) Engineering Memories: A Cognitive Neural Prosthesis for Restoring and Enhancing Memory Function
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3141091/
https://viterbischool.usc.edu/news/2018/03/prosthetic-memory-system-successful-in-humans-study-finds/

https://www.youtube.com/watch?v=r8bix3d0tCs A Hippocampal Neural Prosthesis for Human Memory (2019)

Developing a hippocampal neural prosthetic to facilitate human memory encoding and recall, Robert E Hampson et al 2018 J. Neural Eng. 15 036014
https://iopscience.iop.org/article/10.1088/1741-2552/aaaed7/pdf

A Successful Artificial Memory Has Been Created. Scientific American https://tiny.cc/obuxbz The growing science of memory manipulation raises social and ethical questions. By Robert Martone on August 27, 2019

Ethical aspects of memory manipulation – affecting human identity and autonomy



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

AI MIND READING & MIND CONTROL
HUMAN INTELLIGENCE ENHANCEMENT & REPLACEMENT TECHNOLOGY

"Mind reading" technologies. MIT, The University of California in San Francisco, Elon Mask and Facebook are "in the race to read minds with computers"- in order to enable merging the human brain with the computer https://www.youtube.com/watch?v=R3G5fzz76lQ

"The ability to control electrical activity in brain circuits has the potential to do for brain disorders what electrical stimulation has accomplished in treating cardiac disorders. By beaming electrical or magnetic pulses through the scalp, and by implanting electrodes in the brain, researchers and doctors can treat a vast array of neurological and psychiatric disorders, from Parkinson's disease to chronic depression."

Fields, R. D. (2020) Mind Reading and Mind Control Technologies Are Coming. We need to figure out the ethical implications before they arrive. https://blogs.scientificamerican.com/observations/mind-reading-and-mind-control-technologies-are-coming/

ELON MUSK'S NEURALINK BRAIN CHIP: WHAT SCIENTISTS THINK OF FIRST HUMAN TRIAL

Neuralink, the company through which entrepreneur Elon Musk hopes to revolutionize brain–computer interfaces (BCIs), has implanted a 'brain-reading' device into a person for the first time, according to a tweet posted by Musk on 29 January.

BCIs record and decode brain activity, to allow a person with severe paralysis to control a computer, robotic arm, wheelchair or other device through thought alone. Apart from Neuralink's device, others are under development and some have already been tested in people.

The trial is not registered at ClinicalTrials.gov, an online repository curated by the US National Institutes of Health. Many universities require that researchers register a trial and its protocol in a public repository of this type before study participants are enrolled. Additionally, many medical journals make such registration a condition of publication of results, in line with ethical principles designed to protect people who volunteer for clinical trials.

https://www.nature.com/articles/d41586-024-00304-4

REFERENCES

1. Philosophy of artificial intelligence https://en.wikipedia.org/wiki/Philosophy_of_artificial_intelligence
2. David Deutsch (2012) Philosophy will be the key that unlocks artificial intelligence. https://www.theguardian.com/science/2012/oct/03/philosophy-artificial-intelligence?CMP=share_btn_tw
3. Dwivedi, Y. K. et al., (2019) Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, https://doi.org/10.1016/j.ijinfomgt.2019.08.002
4. The multidisciplinary nature of machine intelligence https://www.nature.com/collections/csgqqsrfxh
5. Benderskaya E.N., Zhukova S.V. (2013) Multidisciplinary Trends in Modern Artificial Intelligence: Turing's Way. In: Yang XS. (eds) Artificial Intelligence, Evolutionary Computing and Metaheuristics. Studies in Computational Intelligence, vol 427. Springer, Berlin, Heidelberg
6. Robert E Hampson et al (2018) Developing a Hippocampal neural prosthetic to facilitate human memory encoding and recall, J. Neural Eng. 15 036014 https://iopscience.iop.org/article/10.1088/1741-2552/aaaed7/pdf
7. Robert Martone (2019) A Successful Artificial Memory Has Been Created. Scientific American https://tiny.cc/obuxbz The growing science of memory manipulation raises social and ethical questions.
8. Fields, R. D. (2020) Mind Reading and Mind Control Technologies Are Coming. We need to figure out the ethical implications before they arrive. https://blogs.scientificamerican.com/observations/mind-reading-and-mind-control-technologies-are-coming/

AI, TECHNOLOGY AND VALUES

https://www.quantamagazine.org/artificial-intelligence-will-do-what-we-ask-thats-a-problem-20200130/Ai genie in a bottle

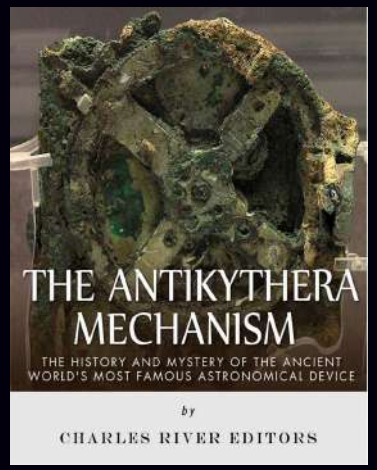
ANCIENT ROOTS OF AI IDEA

TALOS OF CRETE- FIRST INTELLIGENT ROBOT
&
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://taloscrafts.com/who-is-talos/



https://en.wikipedia.org/wiki/Antikythera_mechanism

INTELLIGENT ARTIFACTS TODAY

WITH DIFFERENT INTELLIGENT PROPERTIES

- GenAI, LLMs
- Ambient intelligence
- Intelligent robots & softbots
- Intelligent transportation systems
- Intelligent cities, Intelligent IoT
- Decision making algorithms
- AI for health
- Scientific AI
- AI for software, etc.

Leading companies

Open AI

Google

Microsoft

Nvidia

Amazon

Apple

Facebook

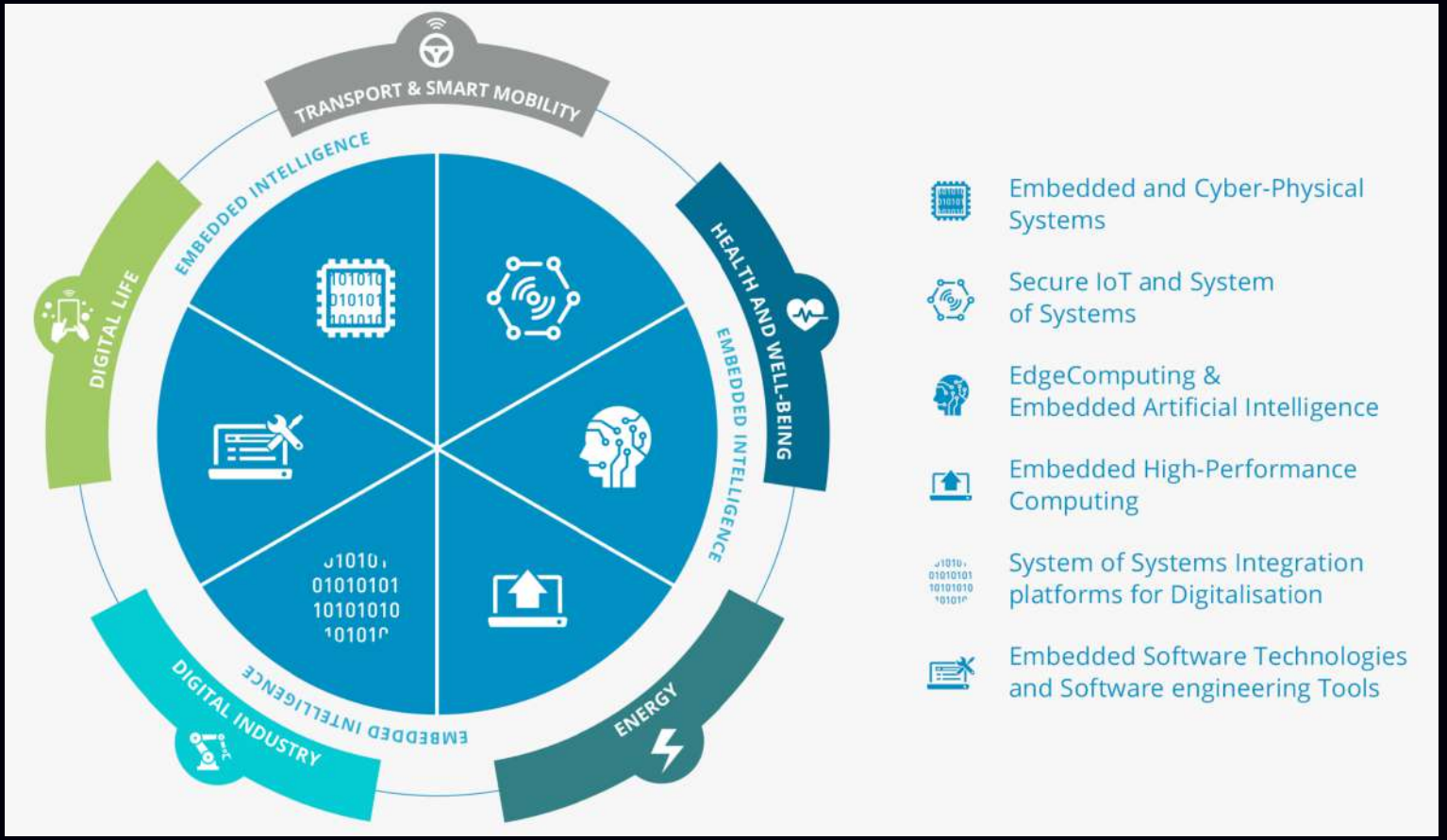
IBM

Intel

Twitter

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

EMBEDDED INTELLIGENCE TECHNOLOGIES



ALGORITHMIC AUTHORITY

Authority is increasingly expressed algorithmically according to the book "The Black Box Society,"* by Frank Pasquale (2016)

Because of information overload is the reason why we depend on algorithms to *sort, categorize, and prioritize* which information we will read and in which way.

Our reality is increasingly shaped by algorithms** . Power of platforms (contrary to the illusion of "technical neutrality and progressive openness")

Algorithmic biases affecting personal integrity (invisible discrimination driven by particular interests and social power)

Algorithmic culture - Humans defined by algorithms

* https://www.hup.harvard.edu/catalog.php?isbn=9780674970847

** David Beer (2017) The social power of algorithms, Information, Communication & Society, 20:1, 1-13, DOI: 10.1080/1369118X.2016.1216147 https://doi.org/10.1080/1369118X.2016.1216147

ALGORITHMIC AUTHORITY

Example: Netflix company that shapes the *media industry with recommendation engine- shaping both audience habits*

Algorithmic authority: algorithmically curating news and social media feeds, evaluating job performance, matching dates, algorithmic management – hiring, supervision, firing employees. (e.g., Facebook's friend feed and Google's search algorithms, Uber's algorithms and Amazon Mechanical Turk.

Algorithms are increasingly used for social, economical and political governance.***

***Katzenbach, C. & Ulbricht, L. (2019). Algorithmic governance. Internet Policy Review, 8(4). DOI: 10.14763/2019.4.1424. https://policyreview.info/node/1424/pdf

AI ETHICS & VALUE-SENSITIVE DESIGN

To connect ethics with AI technology, we need to be able to decide what is right and good in technology, its use and its effect on individuals, society and environment.

We have seen examples of intelligent technologies where AI has important ethical consequences such as *algorithmic governance* or *algorithmic decision-making*.

As we learned in the first part of this lecture, ethics is a branch of Philosophy dealing with the question "what is right?" or "what is good?"

As technology provides us with artifacts that are result of the process of design, our expectation is that they are intended for good of people.

AI ETHICS & VALUE-SENSITIVE DESIGN

One way to introduce ethics into the AI is through Value-sensitive design which is based on the insight that artefacts are value-loaded. We identify values embedded in technologies by studying its use.

"Value" is defined broadly as property that a person or a group considers important, and designers can intentionally or unintentionally inscribe their values in the design objects thus shaping them accordingly.

For AI technology some of important values are: *safety, security, privacy, autonomy, trust, fairness, non-maleficence, beneficence, reliability, responsibility, sustainability.*

The design is typically carried out iteratively and values are being assimilated by combining the following approaches : conceptual, technical – empirical and research, with a continuous assessment and learning process within the ecology of socio-technological system.

ETHICAL AI DESIGN - EXPECTATIONS

- EXPLAINABLE & ACCOUNTABLE AI
- PROMOTING HUMAN RIGHTS
- PROTECTING PRIVACY, PERSONAL INTEGRITY (RESPECTING GDPR*)
- FAIR, TRANSDPARENT, ACCOUNTABLE SYSTEMS
- SAFETY CRITICAL AI MUST BE REGULATED, CERTIFIED,

  WITH REGULATORY OVERSIGHT

*General Data Protection Regulation

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

5. Spiekermann S. (2015) Ethical IT Innovation: A Value-Based System Design Approach. Taylor & Francis

6. Lustig, C., Pine, K., Nardi, B., Irani, L., Lee, M. K., Nafus, D., & Sandvig, C. (2016). Algorithmic authority: The ethics, politics, and economics of algorithms that interpret, decide, and manage. In CHI EA 2016: #chi4good - Extended Abstracts, 34th Annual CHI Conference on Human Factors in Computing Systems (Vol. 07-12-May-2016, pp. 1057-1062). Association for Computing Machinery. https://doi.org/10.1145/2851581.2886426

ETHICALLY ALIGNED DESIGN STANDARDS

The IEEE P7000™ series of standards projects under development addresses specific issues at the intersection of technological and ethical considerations. Like its technical standards counterparts, the IEEE P7000 series empowers innovation across borders and enables societal benefit.

The IEEE P7000™ - IEEE Standards Project Model Process for Addressing Ethical Concerns During System DesignInspired by Methodologies to Guide Ethical Research and Design Committee, and supported by IEEE Computer Society https:// standards.ieee.org/project/7000.html

IEEE P7001™ - IEEE Standards Project for Transparency of Autonomous SystemsInspired by the General PrinciplesCommittee, and supported by IEEE Vehicular Technology Society https://standards.ieee.org/project/7001.html

IEEE P7002™ - IEEE Standards Project for Data Privacy Process Inspired by The Personal Data and Individual Agency Control Committee, and supported by IEEE Computer Society https://standards.ieee.org/project/7002.html

IEEE P7003™ - IEEE Standards Project for Algorithmic Bias ConsiderationsSupported by IEEE Computer Society https://standards.ieee.org/project/7003.html

IEEE P7004™ - IEEE Standards Project for Child and Student Data GovernanceInspired by The Personal Data and Individual Agency Control Committee, and supported by IEEE Computer Society https://standards.ieee.org/project/7004.htm

All links accessed on 15 March 2020

AI ETHICS – GUIDELINES AND POLICIES

ETHICAL GOVERNANCE OF AI

Requirement of transparency and democratic legitimacy

Following responsible research & innovation (Europe)*

IEEE initiative – starting with education, ethics in the whole process of engineering

New types of problems emerge in "automated public sphere"

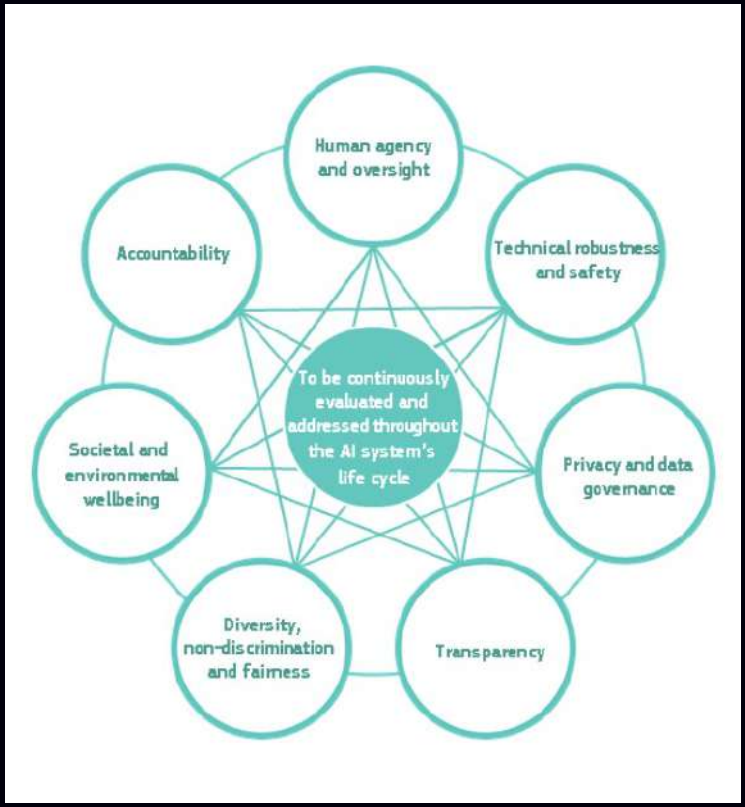
Stakeholders involvement is essential

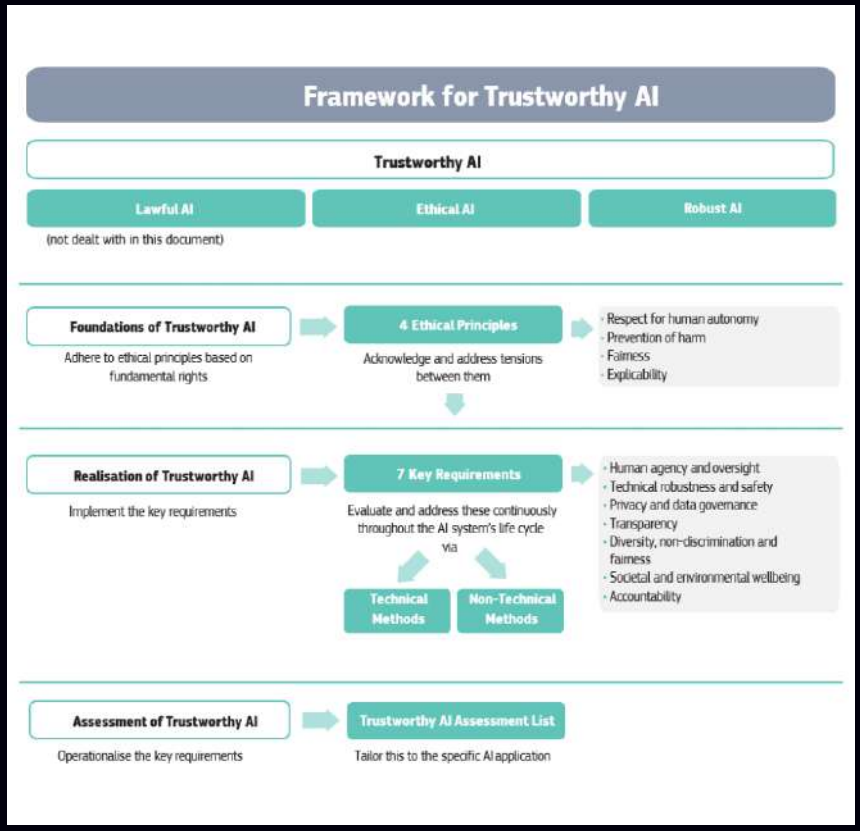
Currently: governmental bodies are working on both legislative and guidelines

- Various groups and initiatives are proposing policies and guidelines
- Concentration of power in few companies (Google, Apple, Amazon, Microsoft) who provide their proposals for ethical policies, guidelines and practices

*https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation

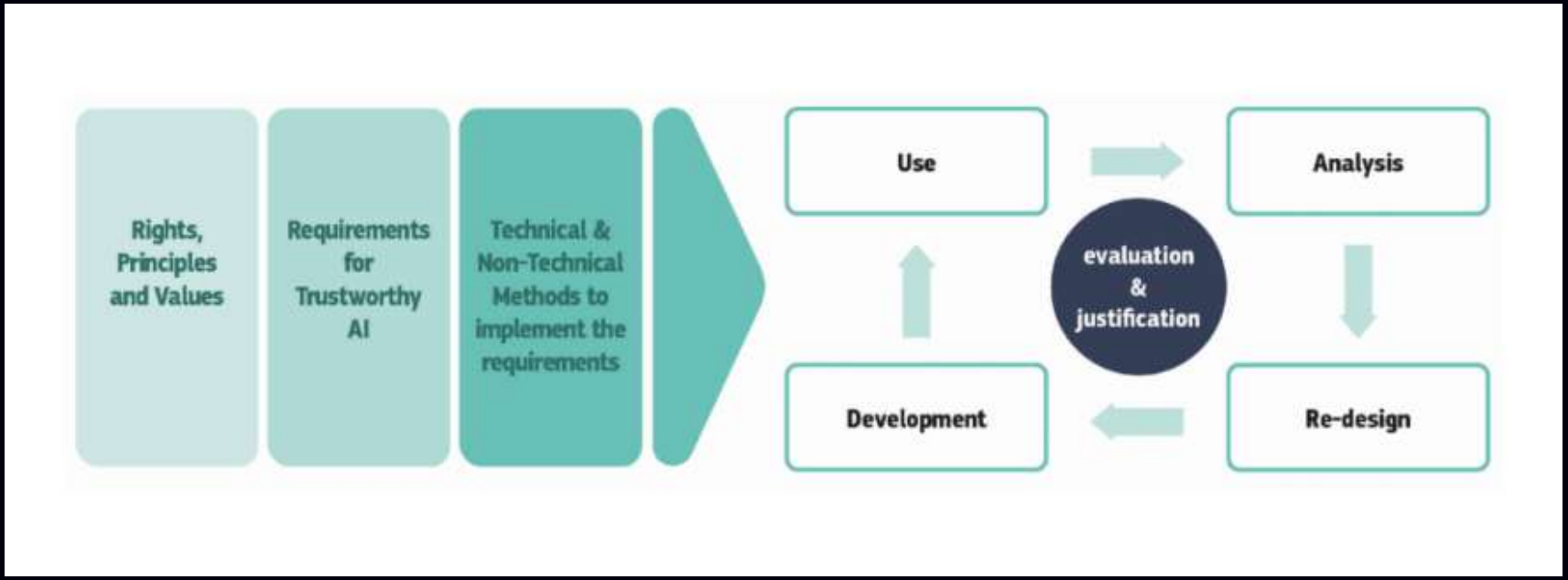
TRUSTWORTHY AI, EU VIEW



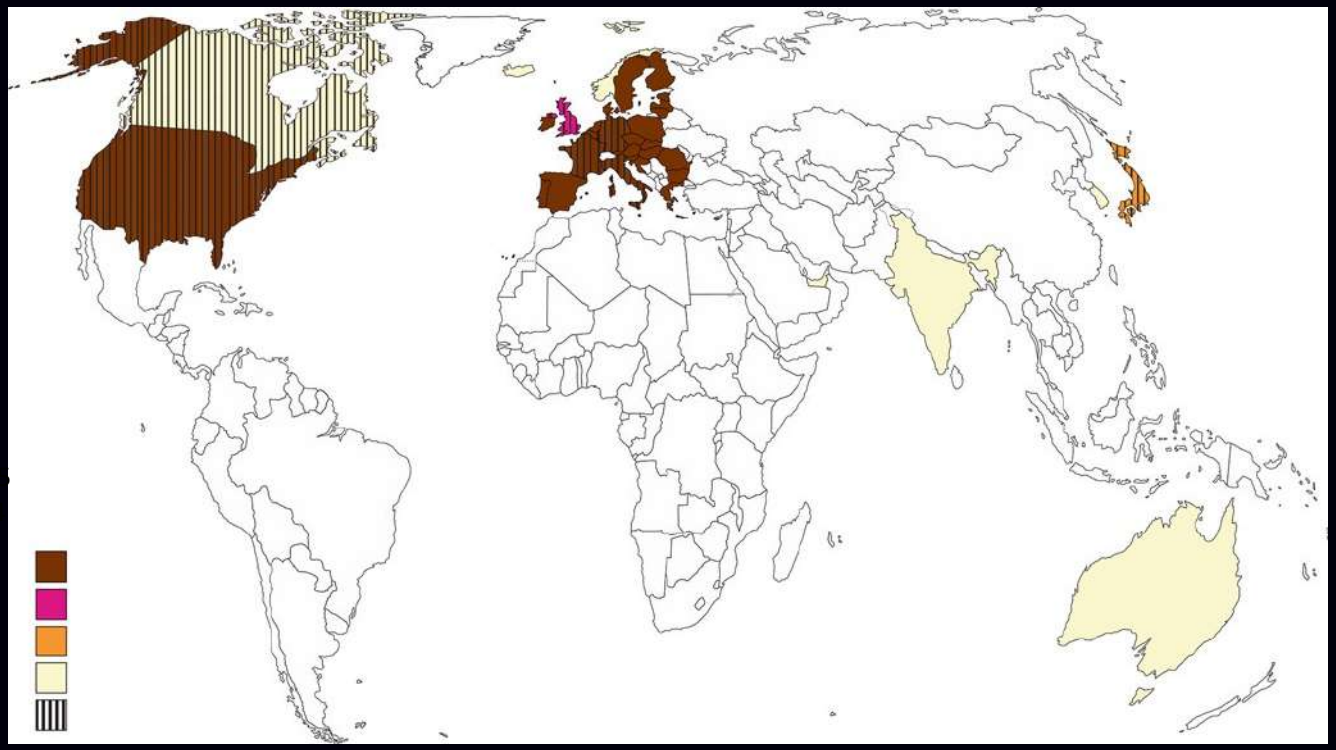


Ethics Guidelines for Trustworthy AI, EU Commission
https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

SOCIETAL AND ENVIRONMENTAL ASPECTS IN THE DESIGN OF AI



NUMBER OF CURRENT AI GUIDELINES, GLOBALLY



Geographic distribution of issuers of ethical AI guidelines by number of documents released. Most ethics guidelines are released in the United States (*n* = 21) and within the European Union (19), followed by the United Kingdom (13) and Japan (4). Canada, Iceland, Norway, the United Arab Emirates, India, Singapore, South Korea and Australia are represented with 1 document each. Having endorsed a distinct G7 statement, member states of the G7 countries are highlighted separately. https://www.nature.com/articles/s42256-019-0088-2.pdf

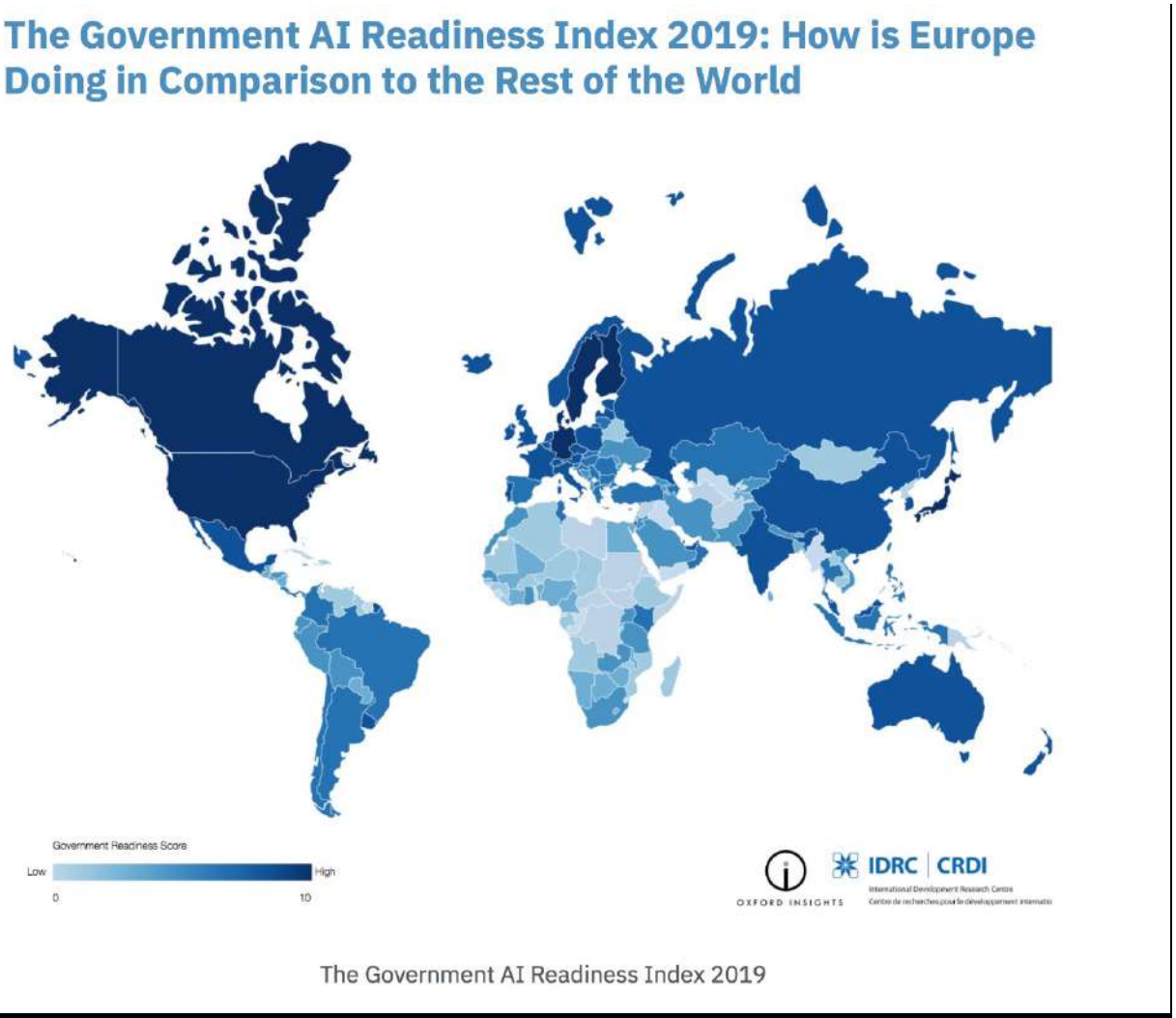
"AI READINESS"

"Government AI Readiness Index" captures the current capacity of governments to exploit the innovative potential of AI.

The overall score is comprised of governance;
infrastructure and data;
skills and education; and
government and public services.

REGULATION

State-Based Regulation
Industry-based Regulation
Citizen-based Regulation (crowdsourced)
Academic or Expert-Based Regulation



https://www.humane-ai.eu/ai-readiness-index-2019/

Ethics guidelines for AI by country of issuer, European
(Colors: Red = governmental + academia, Blue=companies, Black= groups)

Name of document/website	Issuer	Country
Position on Robotics and Artificial Intelligence	The Greens (Green Working Group Robots)	EU
Report with Recommendations to the Commission on Civil Law Rules on Robotics	European Parliament	EU
Ethics Guidelines for Trustworthy AI	High-Level Expert Group on Artificial Intelligence	EU
AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations	AI4People	EU
European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment	Council of Europe: European Commission for the Efficiency of Justice (CEPEJ)	EU
Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems	European Commission, European Group on Ethics in Science and New Technologies	EU
Work in the Age of Artificial Intelligence. Four Perspectives on the Economy, Employment, Skills and Ethics	Ministry of Economic Affairs and Employment	Finland
Tieto's AI Ethics Guidelines	Tieto	Finland
Commitments and Principles	OP Group	Finland

Ethics guidelines for AI by country of issuer, European		
Name of document/website	Issuer	Country
How Can Humans Keep the Upper Hand? Report on the Ethical Matters Raised by AI Algorithms	French Data Protection Authority (CNIL)	France
For a Meaningful Artificial Intelligence. Towards a French and European Strategy	Mission Villani	France
Ethique de la Recherche en Robotique	CERNA (Allistene)	France
AI Guidelines	Deutsche Telekom	Germany
SAP's Guiding Principles for Artificial Intelligence	SAP	Germany
Automated and Connected Driving: Report	Federal Ministry of Transport and Digital Infrastructure, Ethics Commission	Germany
Ethics Policy	Icelandic Institute for Intelligent Machines (IIIM)	Iceland
L'intelligenzia Artificiale al Servizio del Cittadino	Agenzia per l'Italia Digitale (AGID)	Italy
Human Rights in the Robot Age Report	The Rathenau Institute	Netherlands
Dutch Artificial Intelligence Manifesto	Special Interest Group on Artificial Intelligence (SIGAI), ICT Platform Netherlands (IPN)	Netherlands
Artificial Intelligence and Privacy	The Norwegian Data Protection Authority	Norway

Ethics guidelines for AI by country of issuer, Europe		
Name of document/website	Issuer	Country
Principles of robotics	Engineering and Physical Sciences Research Council UK (EPSRC)	UK
The Ethics of Code: Developing AI for Business with Five Core Principles	Sage	UK
Big Data, Artificial Intelligence, Machine Learning and Data Protection	Information Commissioner's Office	UK
DeepMind Ethics & Society Principles	DeepMind Ethics & Society	UK
Business Ethics and Artificial Intelligence	Institute of Business Ethics	UK
AI in the UK: Ready, Willing and Able?	UK House of Lords, Select Committee on Artificial Intelligence	UK
Artificial Intelligence (AI) in Health	Royal College of Physicians	UK
Initial Code of Conduct for Data-Driven Health and Care Technology	UK Department of Health & Social Care	UK
Ethics Framework: Responsible AI	Machine Intelligence Garage Ethics Committee	UK
The Responsible AI Framework	PricewaterhouseCoopers UK	UK
Responsible AI and Robotics. An Ethical Framework.	Accenture UK	UK
Machine Learning: The Power and Promise of Computers that Learn by Example	The Royal Society	UK
Ethical, Social, and Political Challenges of Artificial Intelligence in Health	Future Advocacy	UK

Artificial Intelligence and Machine Learning: Policy Paper	Internet Society	International
Report of COMEST on Robotics Ethics	COMEST/UNESCO	International
Ethical Principles for Artificial Intelligence and Data Analytics	Software & Information Industry Association (SIIA), Public Policy Division	International
ITI AI Policy Principles	Information Technology Industry Council (ITI)	International
Ethically Aligned Design. A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2	Institute of Electrical and Electronics Engineers (IEEE), The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	International
Top 10 Principles for Ethical Artificial Intelligence	UNI Global Union	International
The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation	Future of Humanity Institute; University of Oxford; Centre for the Study of Existential Risk; University of Cambridge; Center for a New American Security; Electronic Frontier Foundation; OpenAI	International
White Paper: How to Prevent Discriminatory Outcomes in Machine Learning	WEF, Global Future Council on Human Rights 2016-2018	International
Privacy and Freedom of Expression in the Age of Artificial Intelligence	Privacy International & Article 19	International

The Toronto Declaration: Protecting the Right to Equality and Non-discrimination in Machine Learning Systems	Access Now; Amnesty International	International
Charlevoix Common Vision for the Future of Artificial Intelligence	Leaders of the G7	International
Artificial Intelligence: Open Questions About Gender Inclusion	W20	International
Declaration on Ethics and Data Protection in Artificial Intelligence	ICDPPC	International
Universal Guidelines for Artificial Intelligence	The Public Voice	International
Ethics of AI in Radiology: European and North American Multisociety Statement	American College of Radiology; European Society of Radiology; Radiology Society of North America; Society for Imaging Informatics in Medicine; European Society of Medical Imaging Informatics; Canadian Association of Radiologists; American Association of Physicists in Medicine	International
Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition (EAD1e)	Institute of Electrical and Electronics Engineers (IEEE), The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems	International

Privacy and Freedom of Expression in the Age of Artificial Intelligence	Department of Industry Innovation and Science	Australia
Montréal Declaration: Responsible AI	Université de Montréal	Canada
Discussion Paper: National Strategy for Artificial Intelligence	National Institution for Transforming India (NITI Aayog)	India
The Japanese Society for Artificial Intelligence Ethical Guidelines	Japanese Society for Artificial Intelligence	Japan
Report on Artificial Intelligence and Human Society (unofficial translation)	Advisory Board on Artificial Intelligence and Human Society (initiative of the Minister of State for Science and Technology Policy)	Japan
Draft AI R&D Guidelines for International Discussions	Institute for Information and Communications Policy (IICP), The Conference toward AI Network Society	Japan
Sony Group AI Ethics Guidelines	Sony	Japan
Discussion Paper on Artificial Intelligence (AI) and Personal Data—Fostering Responsible Development and Adoption of AI	Personal Data Protection Commission Singapore	Singapore
Mid- to Long-Term Master Plan in Preparation for the Intelligent Information Society	Government of the Republic of Korea	South Korea
AI Principles of Telefónica	Telefónica	Spain
AI Principles & Ethics	Smart Dubai	UAE
Tenets	Partnership on AI	N/A
Principles for Accountable Algorithms and a Social Impact Statement for Algorithms	Fairness, Accountability, and Transparency in Machine Learning (FATML)	N/A
10 Principles of Responsible AI	Women Leading in AI	N/A

Unified Ethical Frame for Big Data Analysis. IAF Big Data Ethics Initiative, Part A	The Information Accountability Foundation	USA
The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term	AI Now Institute	USA
Statement on Algorithmic Transparency and Accountability	Association for Computing Machinery (ACM)	USA
AI Principles	Future of Life Institute	USA
AI—Our Approach	Microsoft	USA
Artificial Intelligence. The Public Policy Opportunity	Intel Corporation	USA
IBM's Principles for Trust and Transparency	IBM	USA
OpenAI Charter	OpenAI (a research laboratory based in San Francisco, California. investors include Microsoft, Reid Hoffman's charitable foundation, and Khosla Ventures)	USA
Our Principles	Google	USA

Policy Recommendations on Augmented Intelligence in Health Care H-480.940	American Medical Association (AMA)	USA
Everyday Ethics for Artificial Intelligence. A Practical Guide for Designers and Developers	IBM	USA
Governing Artificial Intelligence. Upholding Human Rights & Dignity	Data & Society	USA
Intel's AI Privacy Policy White Paper. Protecting Individuals' Privacy and Data in the Artificial Intelligence World	Intel Corporation	USA
Introducing Unity's Guiding Principles for Ethical AI—Unity Blog	Unity Technologies	USA
Digital Decisions	Center for Democracy & Technology	USA
Science, Law and Society (SLS) Initiative	The Future Society	USA
AI Now 2018 Report	AI Now Institute (at New York University)	USA
Responsible Bots: 10 Guidelines for Developers of Conversational AI	Microsoft	USA
Preparing for the Future of Artificial Intelligence	Executive Office of the President; National Science and Technology Council; Committee on Technology	USA
The National Artificial Intelligence Research and Development Strategic Plan	National Science and Technology Council; Networking and Information Technology Research and Development Subcommittee	USA
AI Now 2017 Report	AI Now Institute (at New York University - an interdisciplinary research center dedicated to understanding the social implications of artificial intelligence)	USA

ETHICAL PRINCIPLES IDENTIFIED IN EXISTING AI GUIDELINES

From: The global landscape of AI ethics guidelines (Anna Jobin, Marcello Ienca and Effy Vayena) https://www.nature.com/articles/s42256-019-0088-2.pdf

Ethical principle	Number of documents	Included values
Transparency	73/84	Transparency, explainability, explicability, understandability, interpretability, communication, disclosure
Justice and fairness	68/84	Justice, fairness, consistency, inclusion, equality, equity, (non-) bias, (non-)discrimination, diversity, plurality, accessibility, reversibility, remedy, redress, challenge, access and distribution
Non-maleficence	60/84	Non-maleficence, security, safety, harm protection, precaution, prevention, integrity (bodily or mental), non-subversion
Responsibility	60/84	Responsibility, accountability, liability, acting with integrity
Privacy	47/84	Privacy, personal or private information, Integrity
Beneficence	41/84	Benefits, beneficence, well-being, peace, social good, common good
Freedom and autonomy	34/84	Freedom, autonomy, consent, choice, self-determination, liberty, empowerment
Trust	28/84	Trust
Sustainability*	14/84	Sustainability, environment (nature), energy, resources (energy)
Dignity	13/84	Dignity
Solidarity	6/84	Solidarity, social security, cohesion

*https://www.aisustainability.org/ AI SUSTAINABILITY CENTER Stockholm

REFERENCES

1. 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

2. Jobin, A., Ienca, M. and Vayena, E. (2019) The global landscape of AI ethics guidelines. Nature Machine Intelligence | VOL 1 | SEPTEMBER 2019 | 389–399 | www.nature.com/natmachintell https://www.nature.com/articles/s42256-019-0088-2.pdf

3. WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust. Brussels, 19.2.2020. COM(2020) 65 final https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

4. Spiekermann S. (2015) Ethical IT Innovation: A Value-Based System Design Approach. Taylor & Francis

5. Virginia Dignum (2019) Responsible Artificial Intelligence. How to Develop and Use AI in a Responsible Way. Springer Nature Switzerland AG

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

7. 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

REFERENCES

6. Floridi, L., Cowls, J., King, T.C. et al. How to Design AI for Social Good: Seven Essential Factors. Sci Eng Ethics (2020). https://doi.org/10.1007/s11948-020-00213-5

7. 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

8. Floridi, L. (2019) Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. Philosophy & Technology. https://doi.org/10.1007/s13347-019-00354-x

9. Morley, J., Floridi, L., Kinsey, L., Elhalal, A. (2019) From What to How: An Overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices. arXiv:1905.06876

10. Wachter S, Mittelstadt B, Floridi L (2017) Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. International Data Privacy Law, vol. 7, issue 2 (2017) pp. 76-99 Published by Oxford University Press (OUP)

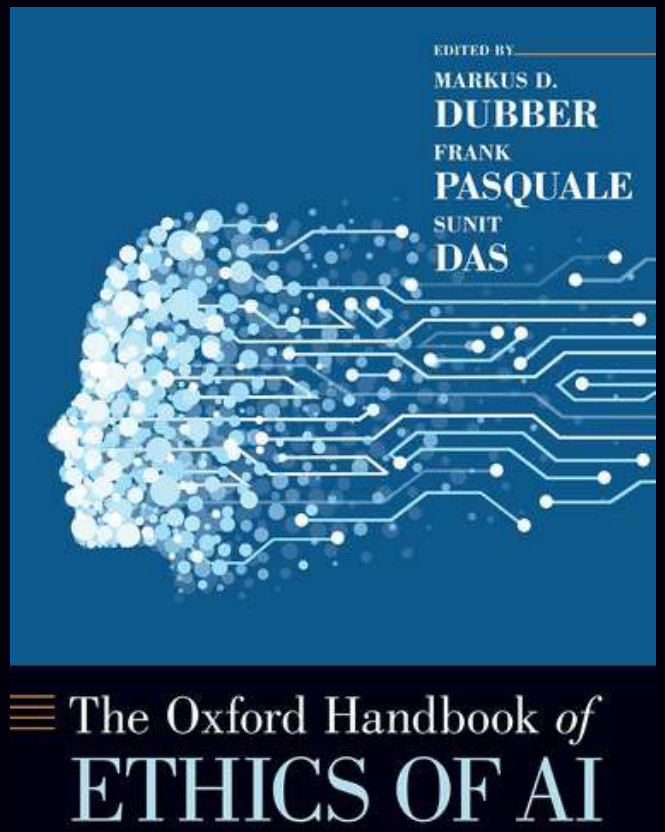
11. Dodig-Crnkovic G. and Çürüklü B., Robots - Ethical by Design, Ethics and Information Technology 2011, Volume 14, Number 1, pp. 61-71. http://www.springerlink.com/content/f432g33181787u63/fulltext.html

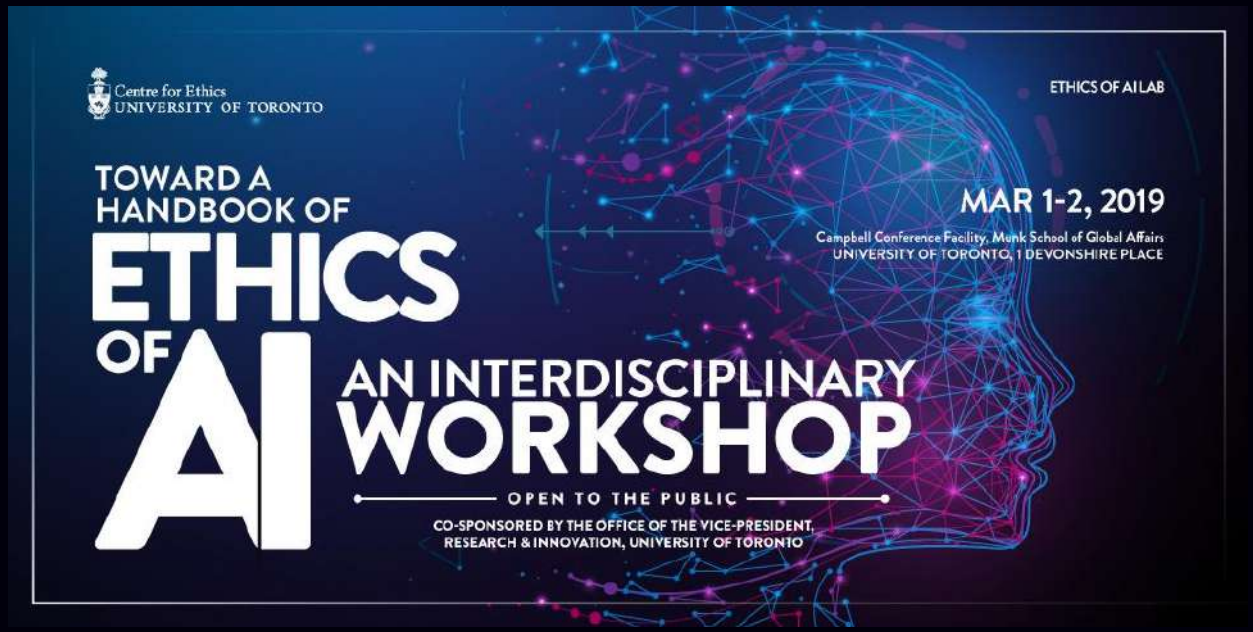
AI ETHICS INITIATIVES & LAWS

https://www.quantamagazine.org/artificial-intelligence-will-do-what-we-ask-thats-a-problem-20200130/Ai genie in a bottle

ONGOING INITIATIVES

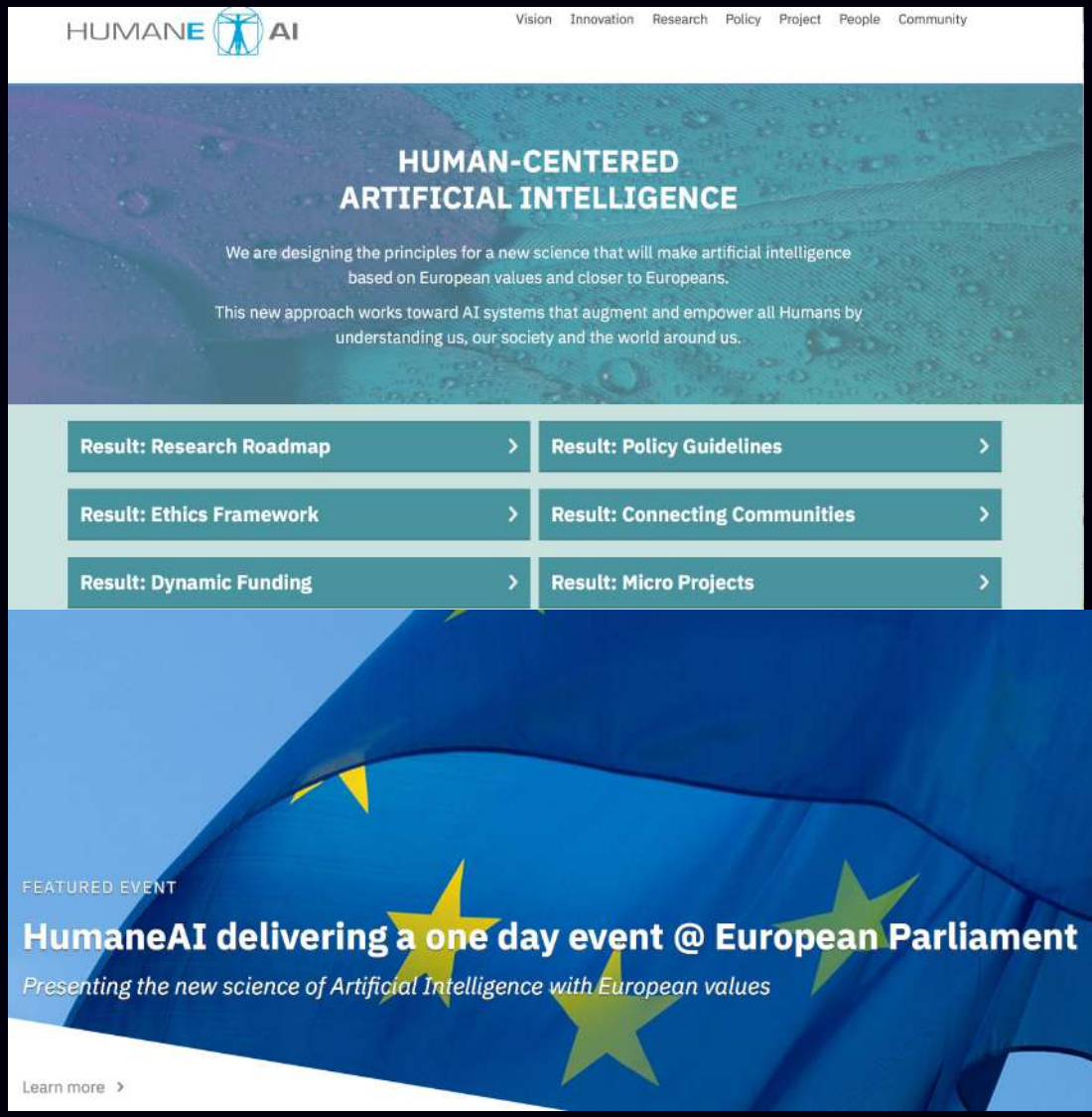
The Oxford Handbook of Ethics of AI





https://global.oup.com/academic/product/the-oxford-handbook-of-ethics-of-ai-9780190067397?cc=ca&lang=en&#

The Oxford Handbook of Ethics of AI



HUMAN-CENTRIC HUMANE AI

"The goal of Humane AI is to harness the emergence of enabling technologies for human-level interaction to empower individuals and society, by providing new abilities to perceive and understand complex phenomena, to individually and collectively solve problems, and to empower individuals with new abilities for creativity and experience."

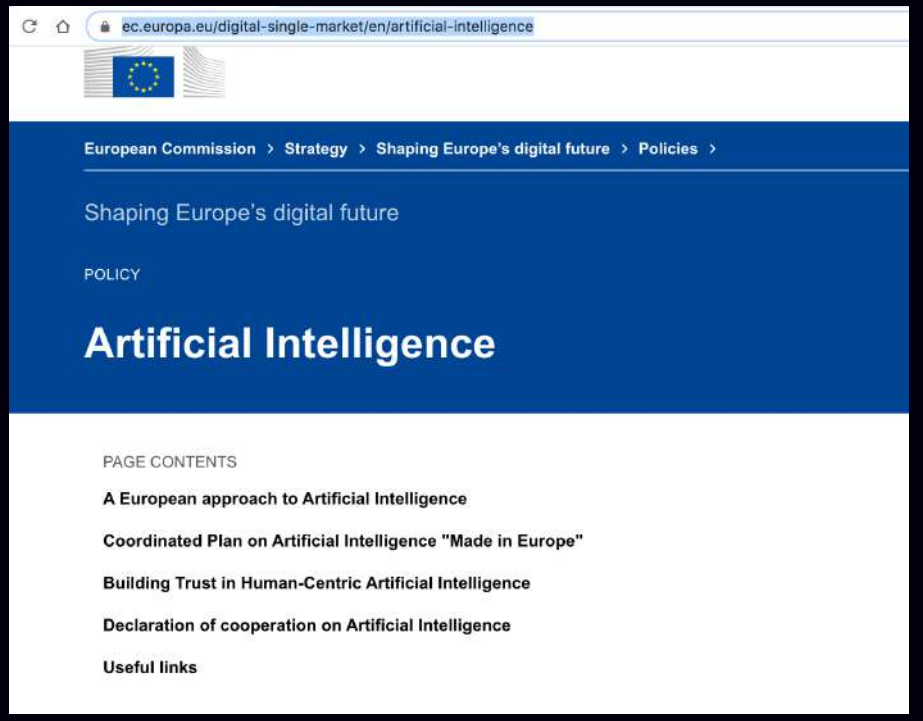
https://www.humane-ai.eu

European Strategy on AI

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

52 experts of a High-Level Expert Group on Artificial Intelligence, are comprising representatives from academia, civil society, as well as industry, supporting the implementation of the European Strategy on Artificial Intelligence.
Ethics Guidelines on Artificial Intelligence: The Guidelines put forward a human-centric approach on AI and list 7 key requirements that AI systems should meet in order to be trustworthy.

1. Human agency and oversight
2. Technical robustness and safety
3. Privacy and Data governance
4. Transparency
5. Diversity, non-discrimination and fairness
6. Societal and environmental well-being
7. Accountability



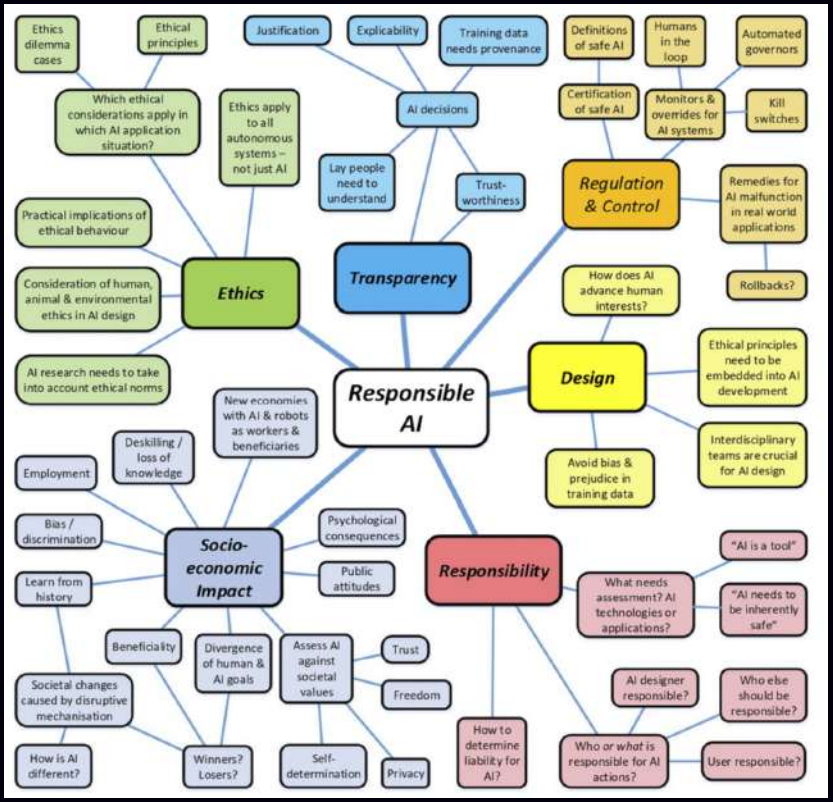
https://ec.europa.eu/digital-single-market/en/artificial-intelligence

RESPONSIBLE AI (NOT CAUSING HARM) & AI FOR GOOD (ACTIVELY CONTRIBUTING TO HUMANITY)

"Responsible AI – Key Themes, Concerns & Recommendations for European Research and Innovation" Report by cross-disciplinary experts.

"Responsible AI" is an umbrella term for investigations into legal, ethical and moral standpoints of autonomous algorithms or applications of AI whose actions may be safety-critical or impact the lives of citizens in significant and disruptive ways. It reports a summary of results from a consultation with cross-disciplinary experts in and around the subject ."

Steve Taylor, Brian Pickering, Michael Boniface, University of Southampton, UK
Michael Anderson, Uni. of Hartford, USA
David Danks, L.L., Carnegie Mellon USA
Dr Asbjørn Følstad, SINTEF, Norway
Dr. Matthias Leese, ETH Zurich, CH
Vincent C. Müller, University of Leeds, UK
Tom Sorell, University of Warwick, UK
Alan Winfield, Uni. of the West of England
Dr Fiona Woollard, Uni. Southampton, UK



https://www.ngi.eu/news/2018/07/23/responsible-ai/

https://ai4good.org/
https://en.wikipedia.org/wiki/AI_for_Good
https://www.microsoft.com/en-us/ai/ai-for-good

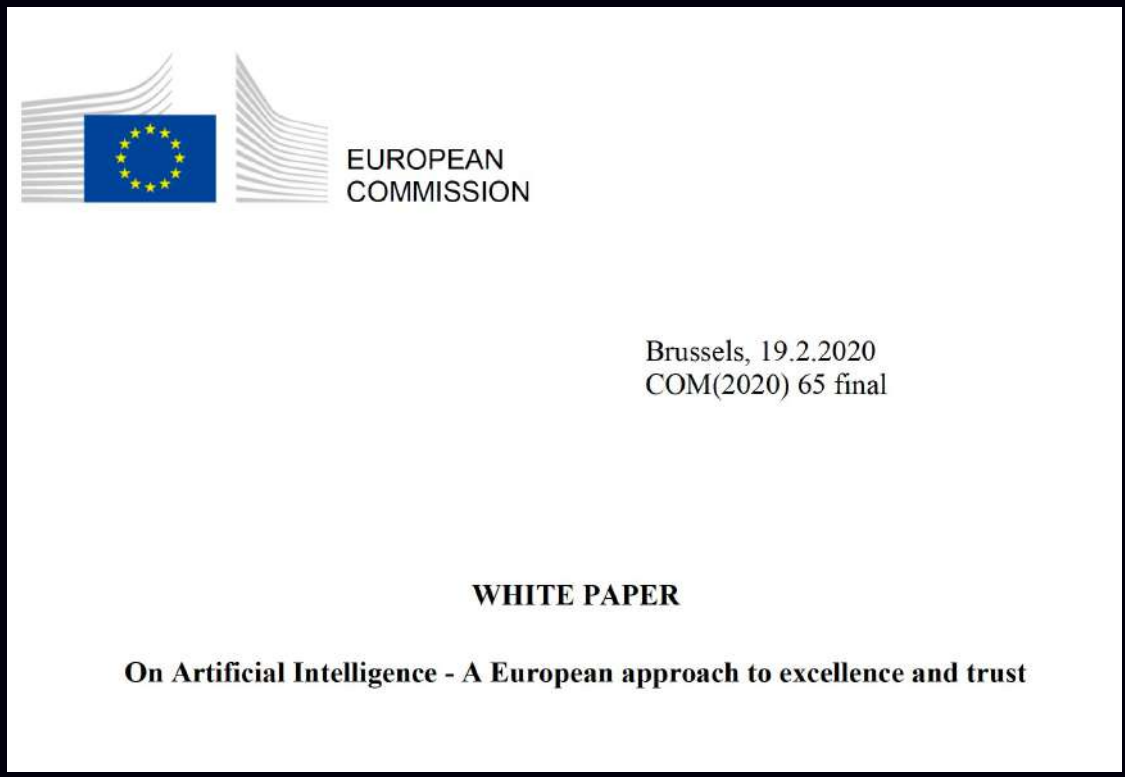
WHITE PAPER ON AI ECOSYSTEM

- AI definition
- AI context (socio-technological system)
- "AI Race" versus "AI Exploration"
- Trustworthy AI (cannot be a choice between an accurate black box AI-system or an explainable, but less accurate AI-system)
- Explainability
- Bias and transparency
- Traceability on the decisions made by the human actors related to the design, development, and deployment of a system
- Liability (requires adjustments to the existing safety and liability regimes)
- Assessment for high-risk AI based on Ethics Guidelines for Trustworthy AI, developed by the High-Level Expert Group on AI.

Comments invited until 19 May 2020

http://allai.nl/first-analysis-of-the-eu-whitepaper-on-ai/

First analysis of the EU Whitepaper on AI


EUROPEAN
COMMISSION

Brussels, 19.2.2020
COM(2020) 65 final

WHITE PAPER

On Artificial Intelligence - A European approach to excellence and trust

https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

LEGAL VIEW OF AI

- AI develops faster than laws
- Liability in case AI causes damage or loss of lives, like autonomous weapons
- When artificial intelligence is not enough but common sense is needed. Biases and errors.
- Privacy loss because of data-hungry AI
- Patents based on or produced by AI, intellectual Property
- Job loss and wealth inequality
- Robot rights. How should we treat AIs (when they get more intelligent)

LEGAL ASPECTS OF AI

Mireille Hilderbrandt

- Law as architecture
- The choice architecture of the Rule of law
- The GDPR and the Charter of Fundamental Rights
- The methodological integrity of computer science and the GDPR
- Legal protection by design

29/10/19 ECSS 2019 ROME Keynote Hildebrandt

Mireille Hildebrandt is a Dutch lawyer and philosopher who works at the intersection between law and computer science, the Research Professor on 'Interfacing Law and Technology' at the Vrije Universiteit Brussel. Principal investigator of the 'Counting as a Human Being in the Era of Computational Law' project. [Wiki]

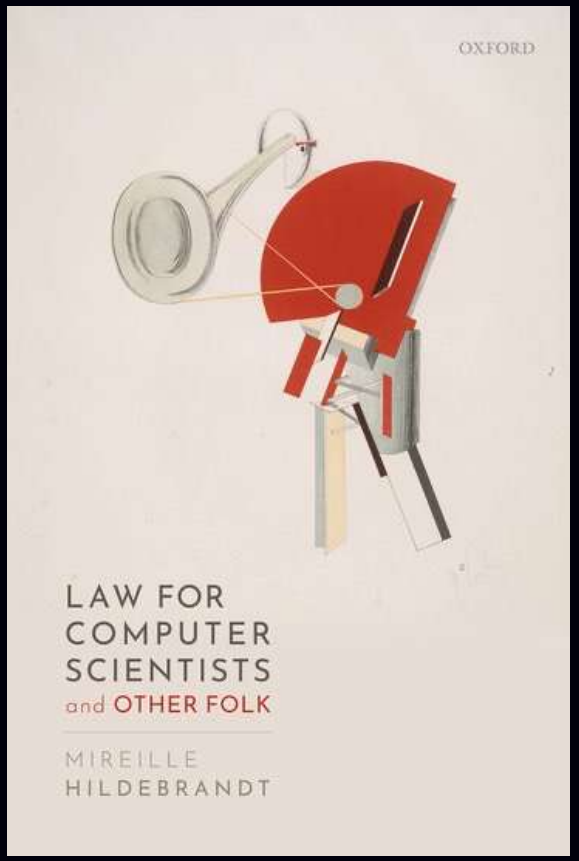
LAW FOR COMPUTER SCIENTISTS

Mireille Hilderbrandt

Forthcoming 2020 by Oxford University Press
Available online, also for comments
E-book open access

https://lawforcomputerscientists.pubpub.org/

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme on 'Counting as a Human Being in the Era of Computational Law'



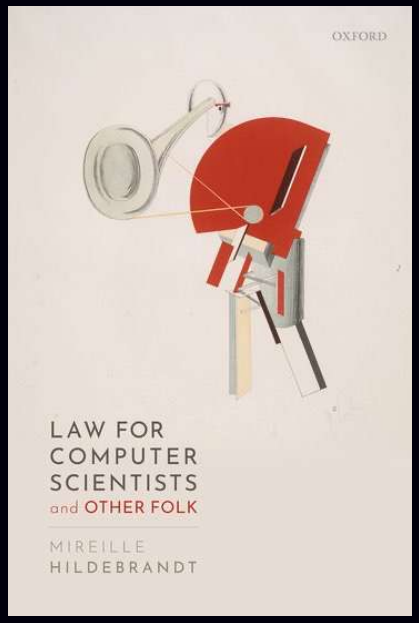
LAW FOR Computer SCIENTISTS

Mireille Hilderbrandt

Part I
What Law Does
1. Introduction: Textbook and Essay
2. Law, Democracy, and the Rule of Law
3. Domains of Law: Private, Public, and Criminal Law
4. International and Supranational Law

Part II
Domains of Cyberlaw
5. Privacy and Data Protection
6. Cybercrime
7. Copyright in Cyberspace
8. Private Law Liability for Faulty ICT

Part III
Frontiers of Law in an Onlife World
9. Legal Personhood for AI?
10. 'Legal by Design' or 'Legal Protection by Design'?
11. Closure: on ethics, code and law



https://lawforcomputerscientists.pubpub.org/

GDPR - General Data Protection Regulation

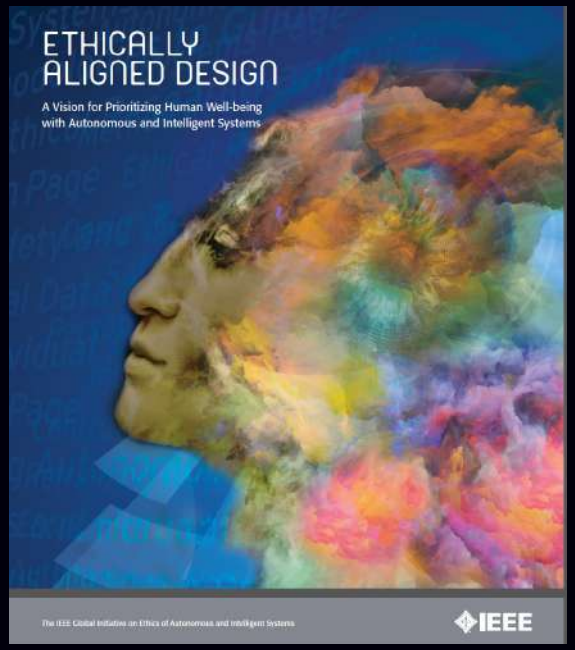
for citizens of the European Union

1. **Data consent:** A company that collects data on individuals must have "unambiguous" consent from those individuals — silence, pre-ticked boxes, or inactivity do not count as consent.
2. **Data portability:** Companies must be willing to move personal data to another location or company, even a direct competitor, if requested by the consumer.
3. **Data deletion:** Companies must delete personal data when requested by an individual.
4. **Consumer profiling:** Individuals can contest, object to, and request explanation for automated decisions or decisions made by algorithms.
5. **Data protection:** The GDPR has strict, specific data security requirements, and stronger enforcement. Data encryption is especially important.
6. **Data breach notification:** The GDPR has a specific definition for what constitutes a breach of "personal" data, along with strict requirements for notifying affected individuals if a breach occurs.
7. **Data Protection Officer (DPO):** All companies that store or process large amounts of personal data must appoint or hire a data protection officer (DPO), who will drive data security and oversee GDPR compliance.

There are two tiers of fines under the GDPR.
First tier: 2% of a company's annual revenue or €10 million, whichever is **larger**.
Second tier: 4% of a company's annual revenue or €20 million, whichever is **larger**.

https://gdpr-info.eu/

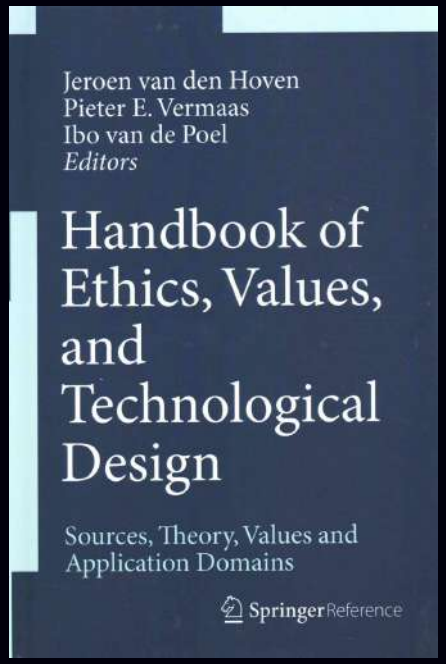
GUIDELINES, RECOMMENDATIONS, POLICIES

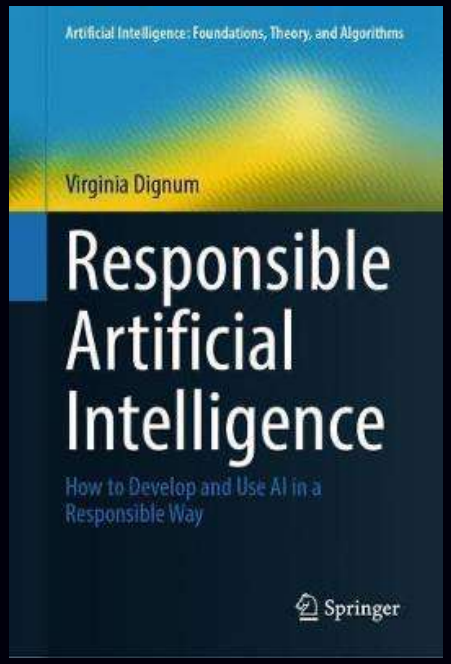


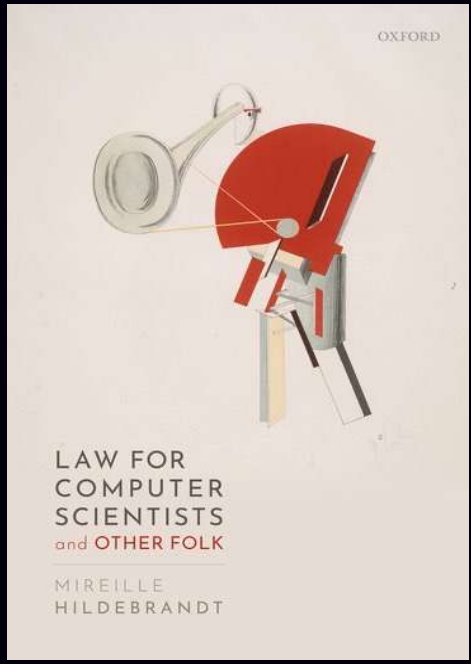
https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

https://ethicsinaction.ieee.org/

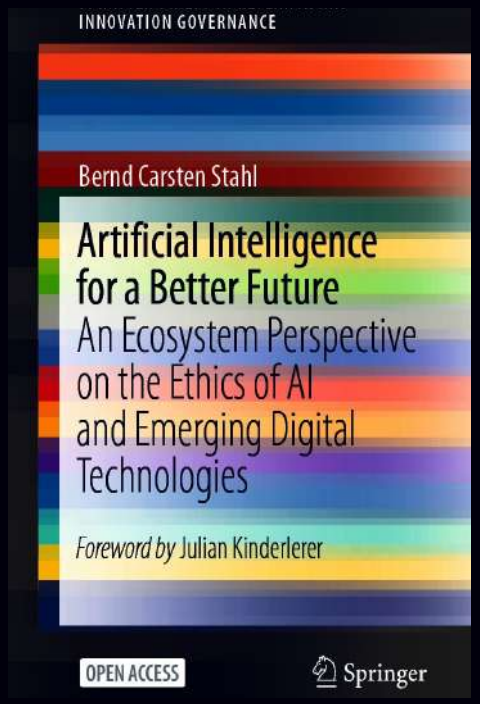
RECENT BOOKS

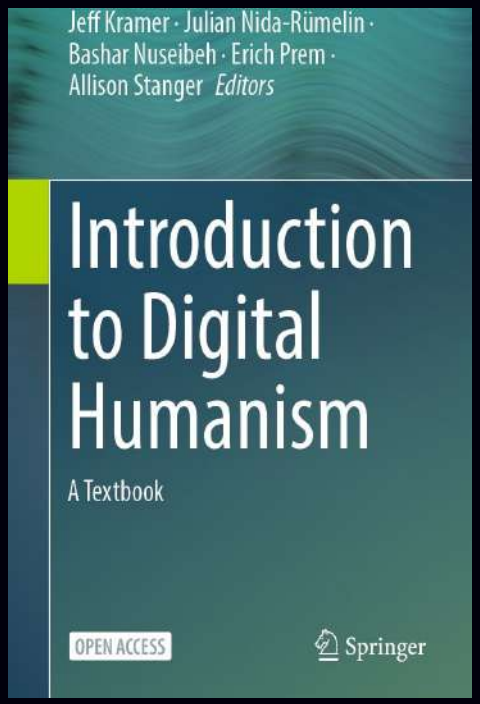


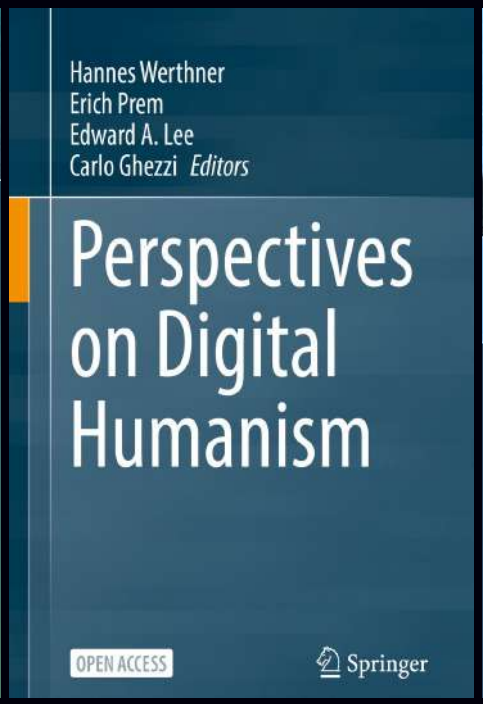


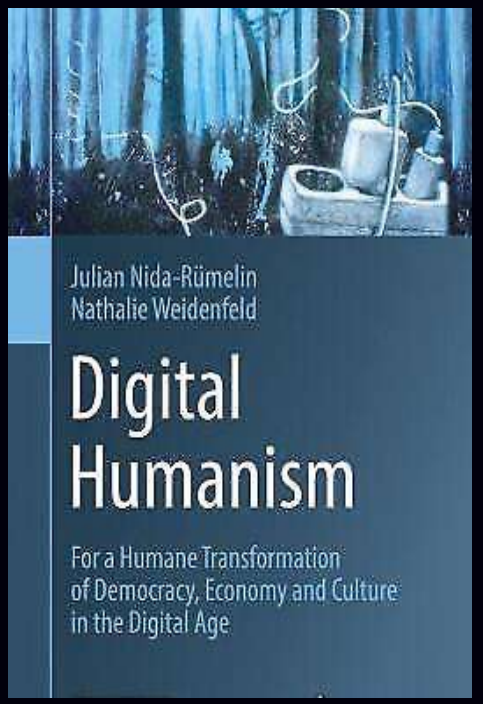


RECENT BOOKS

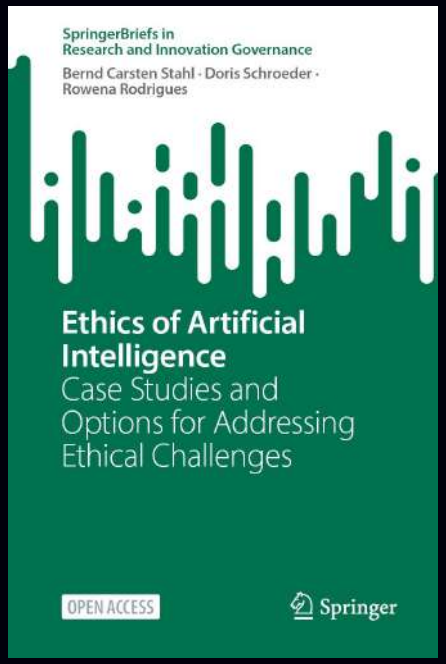


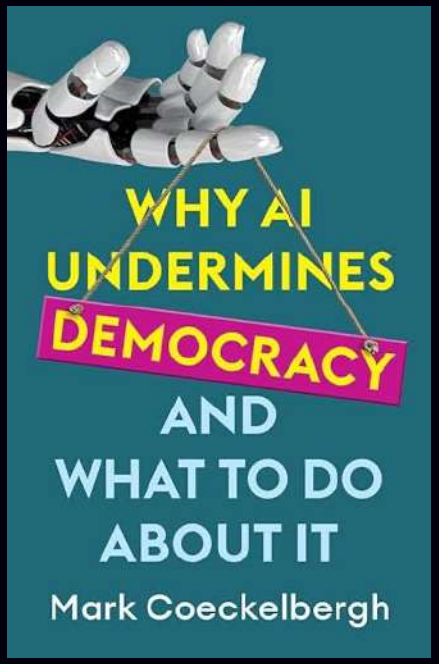


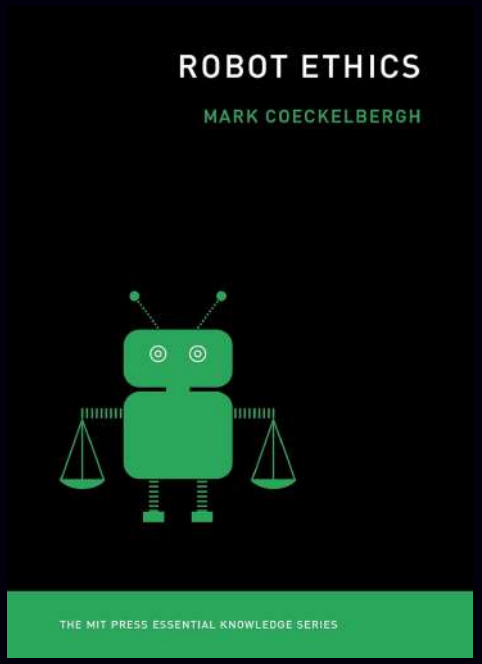


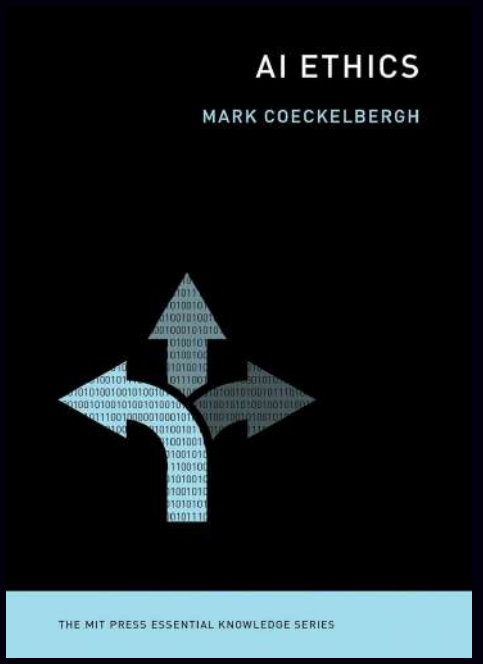


RECENT BOOKS

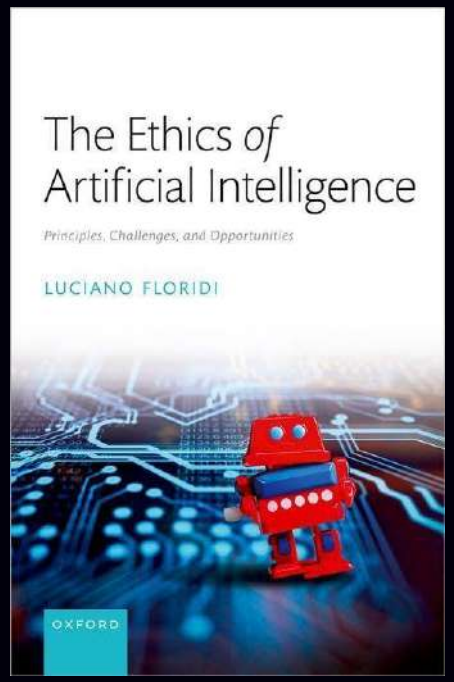


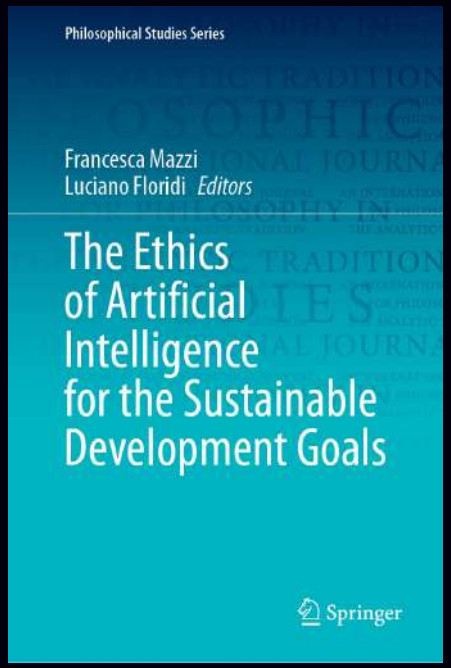


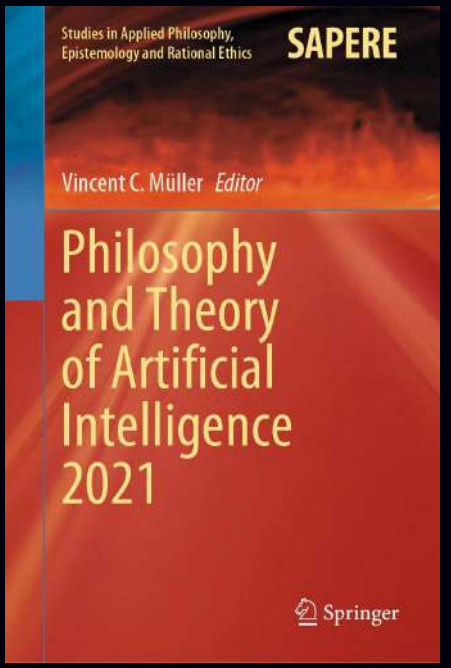




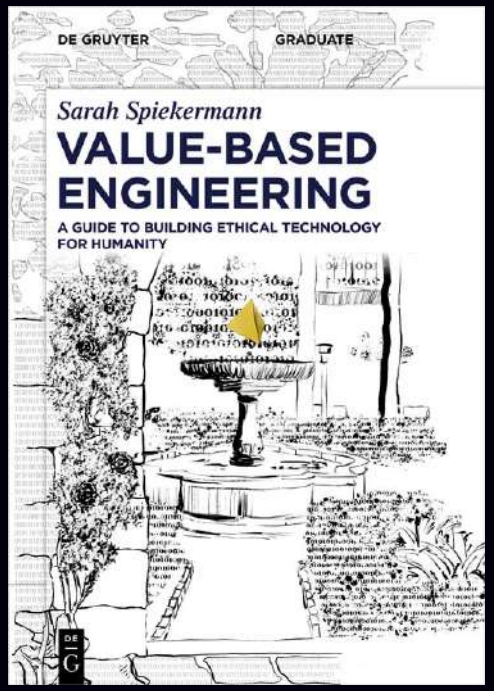
RECENT BOOKS

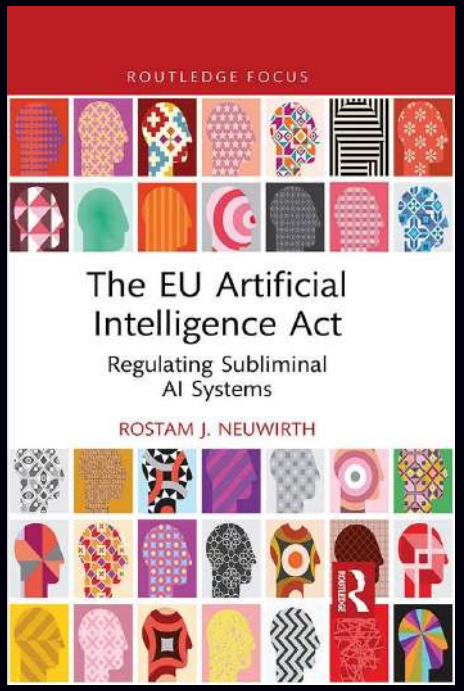


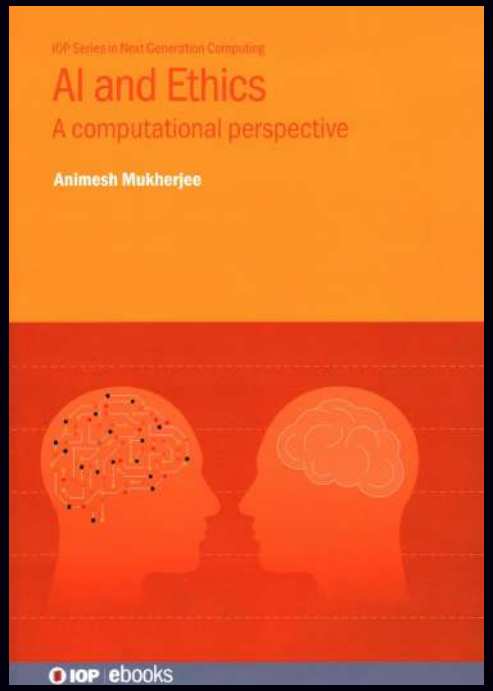


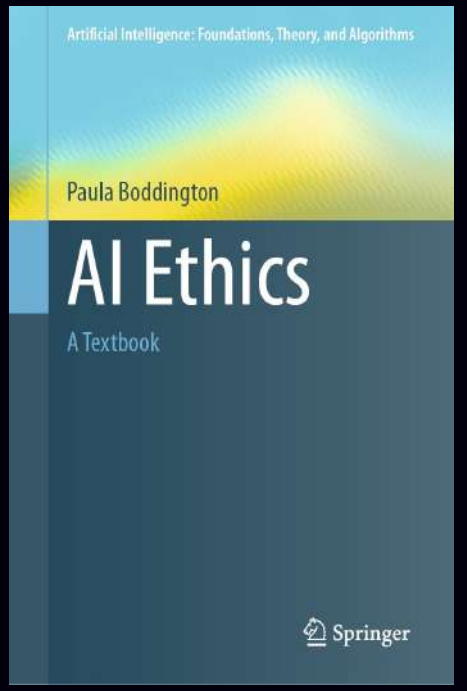


RECENT BOOKS









CONCLUSION

Currently AI is in the flow - very rapid development

Learning AI

Intelligent Assistants

Copilots

AI can be understood from its conceptual foundations - with questions about what AI is today and what it could be (and should be) developed into, to how AI affects our possibility to know, and ethical analysis of what is good AI.

Concurrently there is the development of various scientific domains relevant to AI, followed tightly by its technological applications.

Legal regulations come as soon as reasonably possible when technological applications become widespread and need regulation.

After practical experiences with current technology, the next step in the development is made, and the learning cycle (spiral) starts again.

REFERENCES

1. 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

2. Jobin, A., Ienca, M. and Vayena, E. (2019) The global landscape of AI ethics guidelines. Nature Machine Intelligence | VOL 1 | SEPTEMBER 2019 | 389–399 | www.nature.com/natmachintell https://www.nature.com/articles/s42256-019-0088-2.pdf

3. WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust. Brussels, 19.2.2020. COM(2020) 65 final https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

4. Spiekermann S. (2015) Ethical IT Innovation: A Value-Based System Design Approach. Taylor & Francis

5. Virginia Dignum (2019) Responsible Artificial Intelligence. How to Develop and Use AI in a Responsible Way. Springer Nature Switzerland AG

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

7. 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

REFERENCES

6. Floridi, L., Cowls, J., King, T.C. et al. How to Design AI for Social Good: Seven Essential Factors. Sci Eng Ethics (2020). https://doi.org/10.1007/s11948-020-00213-5

7. 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

8. Floridi, L. (2019) Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical. Philosophy & Technology. https://doi.org/10.1007/s13347-019-00354-x

9. Morley, J., Floridi, L., Kinsey, L., Elhalal, A. (2019) From What to How: An Overview of AI Ethics Tools, Methods and Research to Translate Principles into Practices. arXiv:1905.06876

10. Wachter S, Mittelstadt B, Floridi L (2017) Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. International Data Privacy Law, vol. 7, issue 2 (2017) pp. 76-99 Published by Oxford University Press (OUP)

11. Dodig-Crnkovic G. and Çürüklü B., Robots - Ethical by Design, Ethics and Information Technology 2011, Volume 14, Number 1, pp. 61-71. http://www.springerlink.com/content/f432g33181787u63/fulltext.html