



UNIVERSITY OF GOTHENBURG



ETHICS OF ARTIFICIAL INTELLIGENCE

GORDANA DODIG-CRNKOVIĆ

CHALMERS UNIVERSITY OF TECHNOLOGY

GOTHENBURG, SWEDEN

<http://www.gordana.se>

<https://www.chalmers.se/en/staff/Pages/gordana-dodig-crnkovic.aspx>

http://www.es.mdh.se/staff/37-Gordana_Dodig_Crnkovic

Lecture in the course MVE370 Matematik och samhälle (Mathematics and Society), 26th April 2022



MVE370 MATEMATIK OCH SAMHÄLLE. FROM SYLLABUS

- Kursen vänder sig till studenter som har intresse i att utveckla sin förmåga att stödja människor i lärandeprocesser samt att driva samhällsnyttiga processer som inte primärt motiveras av företagsekonomisk egennytta (som kallas samhällsentreprenör), utan snarare av behov av meningsfull delaktighet i samhällsutveckling och av personlig utveckling. Kursdeltagaren ska ha läst minst 30 högskolepoäng.
- Kursen riktar sig till studenter med påbörjat och aktivt ideellt engagemang i ideell förening eller motsvarande med samhällsentreprenöriell uppgift, som exempelvis Intize.



CHALMERS
UNIVERSITY OF TECHNOLOGY

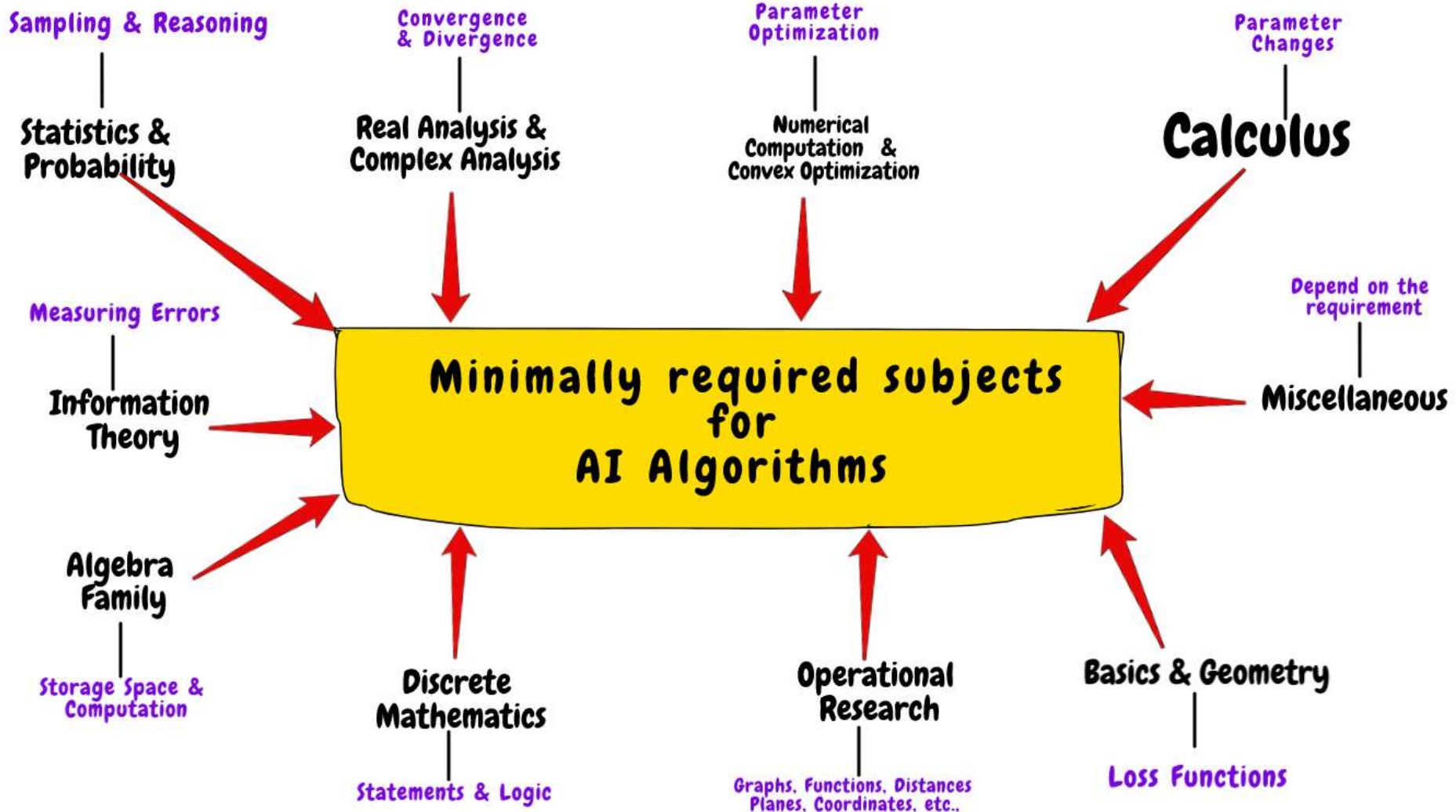


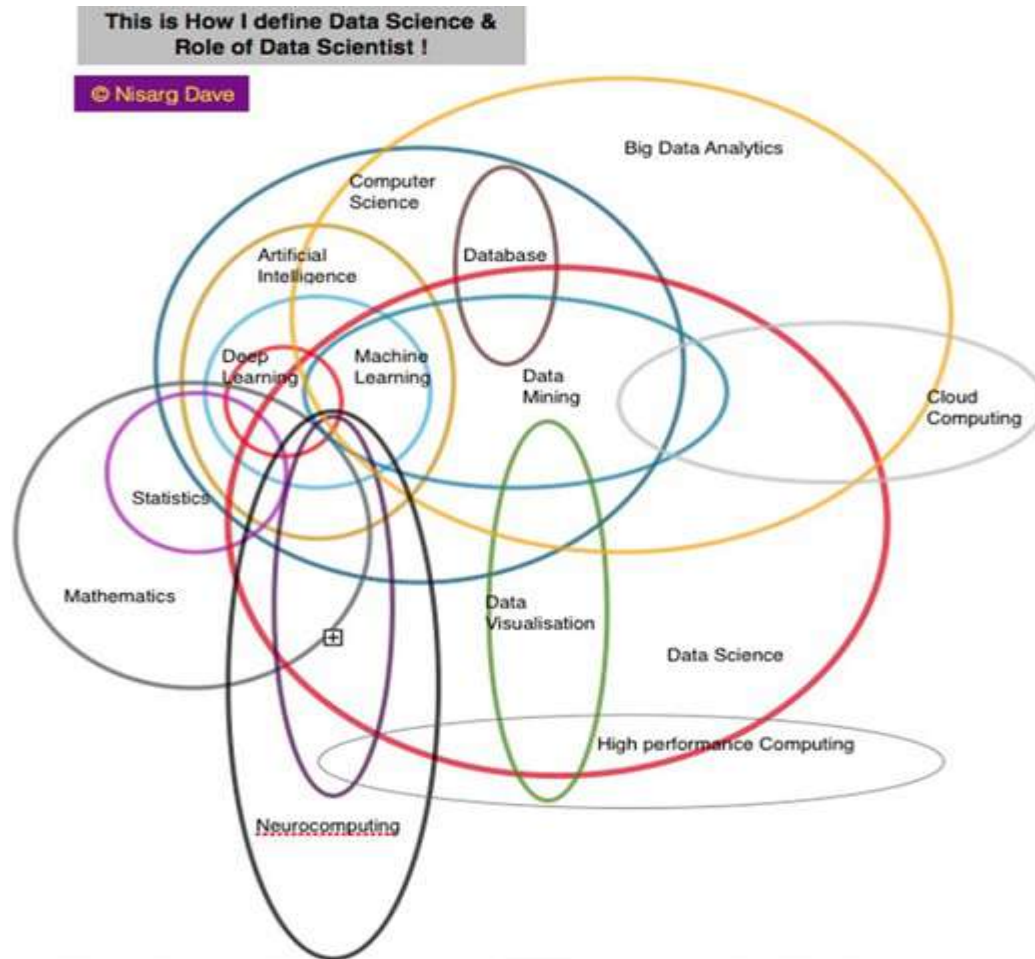
UNIVERSITY OF GOTHENBURG

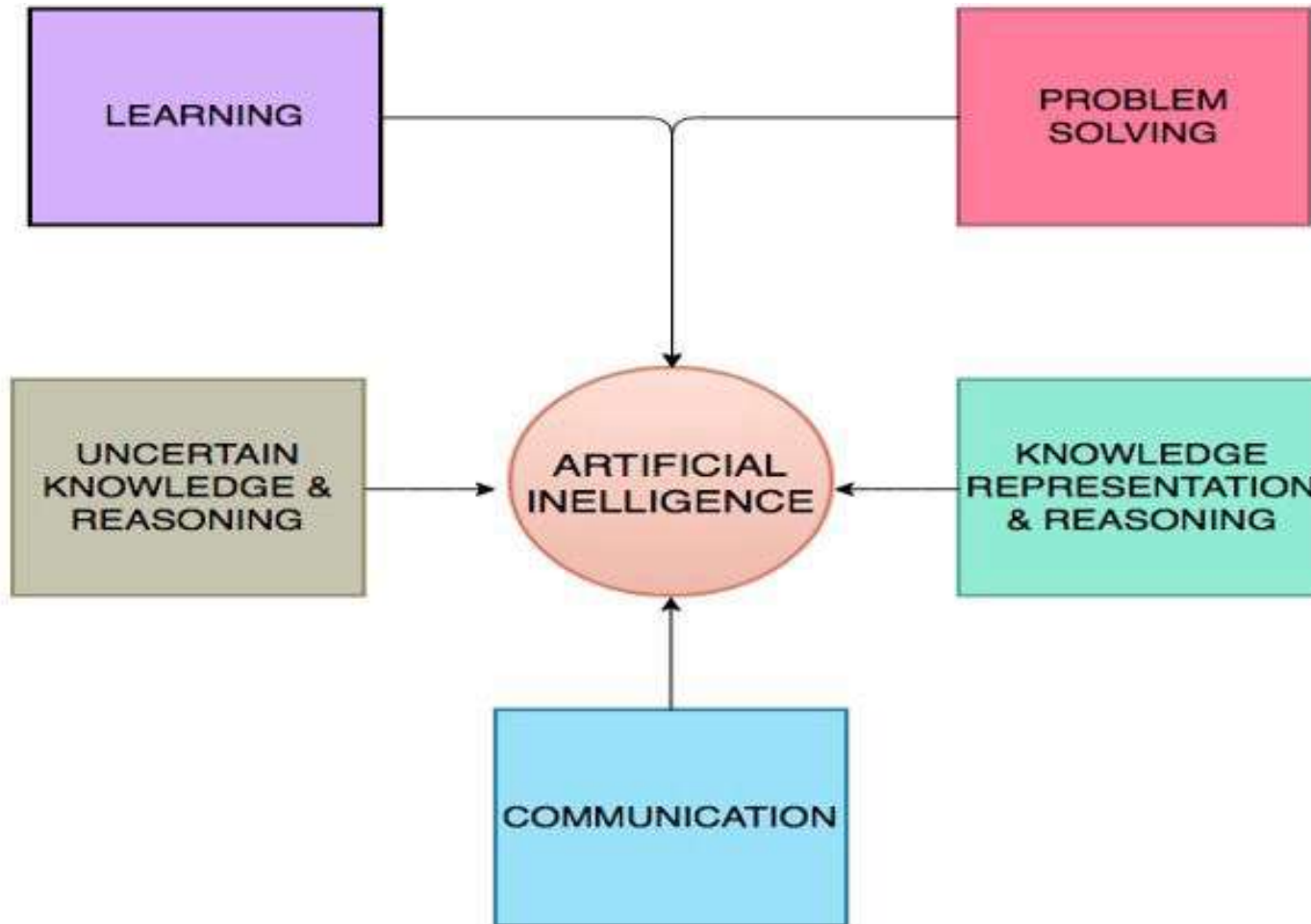


RELATIONS MATHEMATICS TO SOCIETY, MATHEMATICS TO AI AND AI TO SOCIETY (AND ITS BUILDING BLOCKS, INDIVIDUALS)

Mathematics and society stand in mutually dependent relationship – they both influence each other and depend on each other.









ARTIFICIAL INTELLIGENCE & NATURAL INTELLIGENCE

Definition of AI:

“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.”

The [English Oxford Living Dictionary](#)

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

TYPES OF INTELLIGENCE: NATURAL, ARTIFICIAL,
MIX (LIKE CYBORGS)





LECTURE PLAN FOR TODAY

- Introduction on intelligence, natural and artificial
- Design of intelligent artifacts
- Value-sensitive design
- Responsible AI
- Different initiatives for ethical AI
- Summary: Ethics of AI

Break

- Group work: Discussions in breakout rooms
- Getting together with short accounts from group work



INTRODUCTION



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



TYPES OF INTELLIGENCE

EMBODIED INTELLIGENCE - HUMAN, CYBORG, ROBOT

(Direct interaction with a body of intelligent agent)

“DISEMBODIED” INTELLIGENCE - SOFTWARE, SENSING
AND REASONING, DECISIONMAKING INFRASTRUCTURE

(Intelligent agent (software) does not take shape of a body we interact with)

GENERAL/STRONG INTELLIGENCE- HUMAN LEVEL AND
ABOVE

NARROW/ WEAK INTELLIGENCE: TODAY'S INTELLIGENT
ARTIFACTS (Intelligent cars, intelligent cities etc.)



DIGITALIZATION & COGNITIZATION

INTELLIGENCE & ARTIFICIAL INTELLIGENCE

- The process of digitalization happens in parallel with the introduction of cognitive properties and intelligence into artifacts
- Ethical implications of the digitalization of our society: new technologies are followed by **regulatory and legal vacuums – we lack regulations and laws for emerging technologies** (James Moore)
- My background: theoretical physics, computer science, philosophy of computing and ethics, computational models of cognition, recent interest: interaction design – Interdisciplinary approaches to AI.



DESIGN OF INTELLIGENT ARTIFACTS

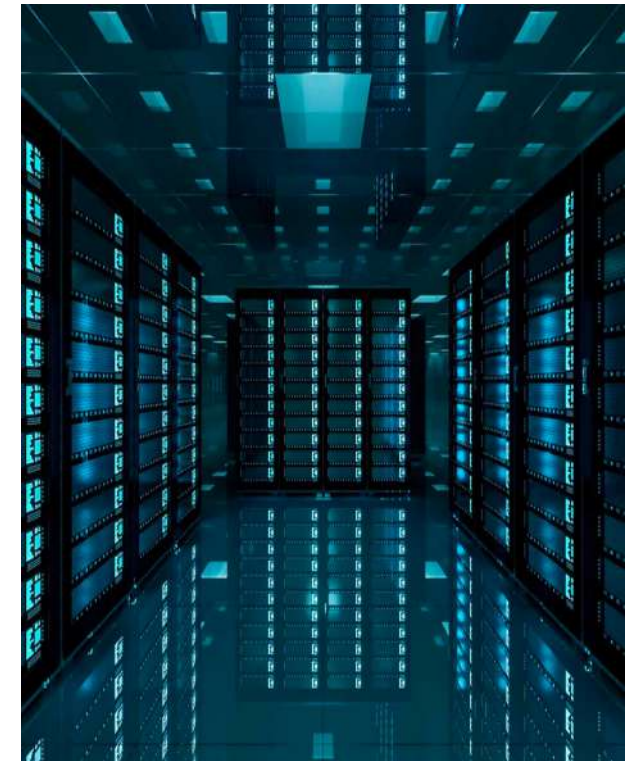


<https://www.iconfinder.com/iconsets/brain-service-2>



INTELLIGENT ARTIFACTS BEING DEVELOPED

- Ambient intelligence
- Autonomous Intelligent Systems
- Intelligent Robots & Softbots
- Intelligent transportation
- Intelligent Cities
- Intelligent IoT
- 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/>



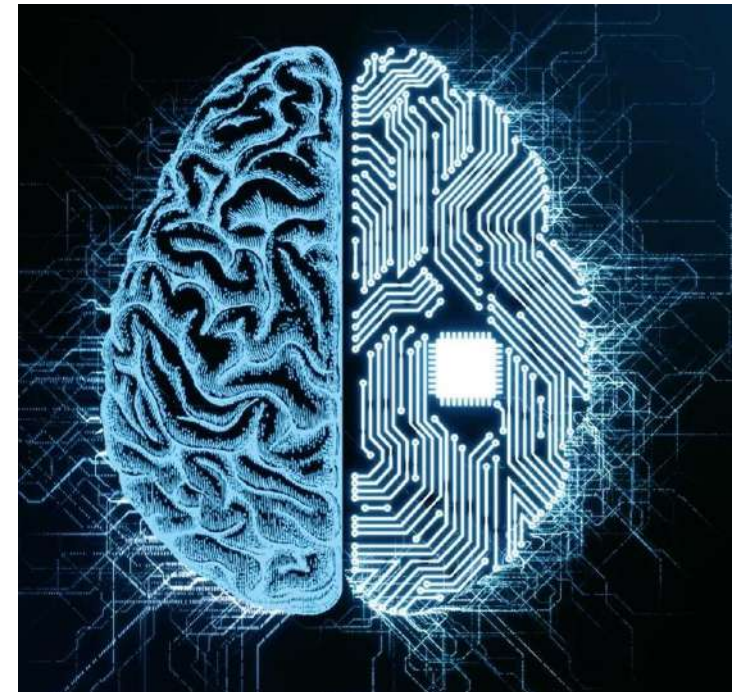
DESIGN OF INTELLIGENT ARTIFACTS

- 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/>



CHALMERS
UNIVERSITY OF TECHNOLOGY



UNIVERSITY OF GOTHENBURG



RECURRENT
PATTERN:
Inspiration FROM
NATURE
(INTELLIGENCE)
QUESTION: HOW
MUCH DO WE
KNOW ABOUT
NATURE?

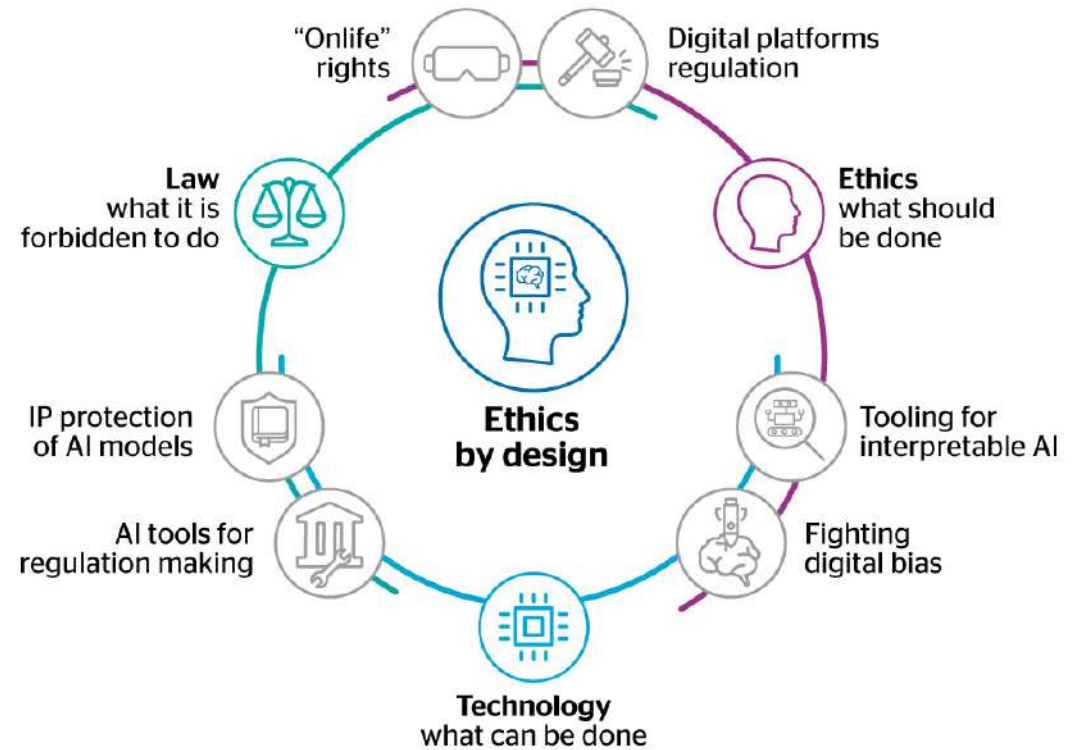
Similar to the way antibodies are able to modify the responses of blood cells to
smart services will adjust their behavior according to their ecosystem's operatin



DO WE WANT INTELLIGENT EVERYTHING?

WHY?

QUESTION OF VALUES & ETHICS





VALUE-SENSITIVE DESIGN



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



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>



VALUE-SENSITIVE DESIGN

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

Ethical values and principles in European discussion

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



VALUE-SENSITIVE DESIGN

Ethical values and principles in European discussion

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

<https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence#Coordinated-EU-Plan-on-Artificial-Intelligence>



VALUE-SENSITIVE DESIGN

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

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

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

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
5. [AlgorithmsWatch. AI Ethics Guidelines Global Inventory. https://inventory.algorithmwatch.org 2020.](#)



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.; Cows, 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>
9. Spiekermann, Sarah and Winkler, Till, Value-based Engineering for Ethics by Design (May 12, 2020). Available at SSRN: <https://ssrn.com/abstract=3598911> or <http://dx.doi.org/10.2139/ssrn.3598911>
10. Sarah Spiekermann (2015) *Ethical IT Innovation: A Value-Based System Design Approach* CRC Press <https://www.amazon.com/Ethical-Innovation-Value-Based-System-Approach/dp/1482226359>

All links accessed on 16 September 2021



RESPONSIBLE AI

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

Dignum, Virginia. Responsible artificial intelligence:
How to develop and use AI in a responsible way.
Springer Nature, 2019. (Book)



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



RESPONSIBLE AI: WHY CARE?

- AI systems are designed to act autonomously in our world (in the future)
- 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**

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

- 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

- 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

- (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 is 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!

RESPONSIBLE ARTIFICIAL INTELLIGENCE



WE ALL ARE RESPONSIBLE
STAKEHOLDERS
in different ways
in different roles

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



DIVISION/ASSIGNMENT OF RESPONSIBILITY

Time perspective

- Short-term perspective
(We decide)
- Middle-term perspective
(AGI – We co-decide)
- Long-term perspective
(Superintelligence? Who decides?)

Stakeholder's roles

- 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](https://plato.stanford.edu/entries/ethics-of-artificial-intelligence/)

Judith Simon

<https://www.youtube.com/watch?v=cvLtFoJme0> Judith Simon - Big data & machine learning

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

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

Luciano Floridi & Cows, 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



CHALMERS
UNIVERSITY OF TECHNOLOGY



UNIVERSITY OF GOTHENBURG



DIFFERENT INITIATIVES FOR ETHICAL AI



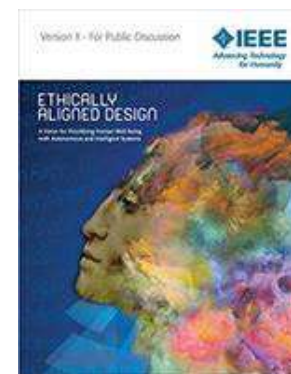
<https://www.iconfinder.com/iconsets/brain-service-2>

AI FOR GOOD, AI FOR PEOPLE, ...

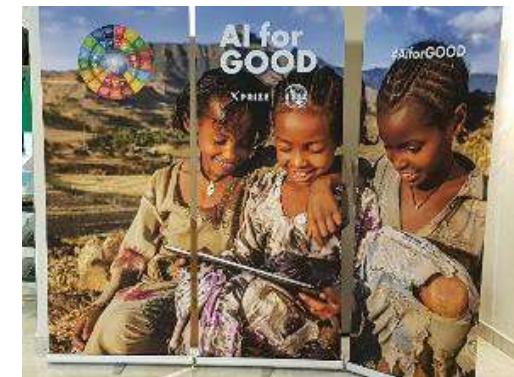
- Harness the positive potential outcomes of AI in society, the economy
- Ensure inclusion, diversity, universal benefits
- Prioritize UN2020 Sustainable Development Goals
- The objective of the AI system is to maximize the realization of human values



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



<https://ethicsinaction.ieee.org/>



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



<http://www.ai4people.eu>

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



CHALMERS
UNIVERSITY OF TECHNOLOGY



UNIVERSITY OF GOTHENBURG



<https://aiforgood.itu.int/> 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



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/>



RESOURCES

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

https://framtidsp rao.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, Accessibility Humanitarian 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



EXAMPLE: ETHICS OF SELF-DRIVING VEHICLES

Real-world Ethics for Self-Driving Cars

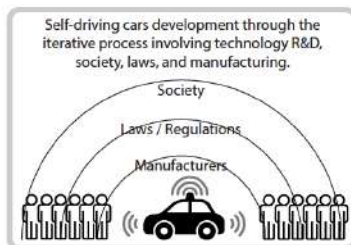
Tobias Holstein¹, Gordana Dodig-Crnkovic^{1,2}, Patrizio Pelliccione^{2,3}

¹Mälardalen University, Västerås, Sweden,
²Chalmers University of Technology | University of Gothenburg, Gothenburg, Sweden,
³University of L'Aquila, L'Aquila, Italy

Ethical and social aspects of the emerging technology of self-driving cars can best be addressed through an applied engineering ethical approach. However, those issues are typically being presented in terms of an idealized unsolvable decision-making problem, the so-called Trolley Problem, that asks how to prioritize killing people in the case of collision.

Instead, we propose that ethical analysis should focus on the study of ethics of complex real-world engineering focused on how not to kill anybody. As software plays a crucial role in the control of self-driving cars, software engineering solutions should handle actual ethical and social considerations.

We present practical social and ethical challenges that must be met in the ecology of the socio-technological system of self-driving cars which implies novel expectations for software engineering in the automotive industry.



Summary:

For the research, development and engineering of emerging AV technology we have to address real-world challenges and move away from popular artificially constructed Trolley Problem thought experiments.

We point out the importance of the ecology of the entire socio-technological system, where ethicality is ensured through education, continuous information about the existing technology performance, and negotiation of priorities in the value systems, as well as constant learning process within technology and its social environment. In this iterative process, values and ethics come first, followed by standardisation and legislation that is monitored, updated and validated in practice.

We argue that real-life ethics plays the central role as a basis sustaining and informing ethically sound emerging technology of self-driving cars and thus the future of transportation.



Further References:
Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). **Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem**. In Steven John Thompson (Ed.), *Handbook of Research on Machine Ethics and Morality*. IGI Global.

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2018). **Ethical and Social Aspects of Self-Driving Cars**. *ArXiv*, abs/1802.04103.

Holstein, T. (2017). **The Misconception of Ethical Dilemmas in Self-Driving Cars**. *Proceedings of the IS4SI 2017 Summit DIGITALISATION FOR A SUSTAINABLE SOCIETY*, Gothenburg, Sweden, 1(3), 2-4. <https://doi.org/10.3390/IS4SI-2017-04026>

Find more information at

<https://ethics.se>



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



References

- A. Jobin, M. Ienca, and E. Vayena. The global landscape of ai ethics guidelines. *Nature Machine Intelligence*, 1(9):389–399, 2019
- Feldt, R., de Oliveira Neto, F.G., Torkar, R.: Ways Of Applying Artificial Intelligence In Software Engineering. In: *International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (RAISE)*, pp. 35–41. IEEE (2018)
- Tantithamthavorn, Chakkrit, Jirayus Jiarpakdee, and John Grundy. "Explainable AI for software engineering." *IEEE Computer* (2020).
- Google. Responsible AI practices. <https://ai.google/responsibilities/responsible-ai-practices/>. Accessed 2021-01-20.
- D.Hoffman and R. Masucci. Intel's AI privacy policy white paper. <https://blogs.intel.com/policy/files/2018/10/Intels-AI-Privacy-Policy-White-Paper-2018.pdf> , 2018. Accessed 2021-07-06
- Telia Company AB. AI Ethics. <https://www.teliacompany.com/en/about-the-company/public-policy/ai-ethics/> , 2019. Accessed 2021-07-06.
- A. Belloni, A. Berger, O. Boissier, G. Bonnet, G. Bourgne, P.-A. Chardel, J.-P. Cotton, N. Evreux, J.-G. Ganascia, P. Jaillon, et al. Dealing with ethical conflicts in autonomous agents and multi-agent systems. In *AAAI'15 Workshops*, 2015.
- Underspecification Presents Challenges for Credibility in Modern Machine Learning <https://arxiv.org/abs/2011.03395>



References

R. T. Vought. Guidance for Regulation of Artificial Intelligence Applications. <https://www.whitehouse.gov/wp-content/uploads/2020/01/Draft-OMB-Memo-on-Regulation-of-AI-1-7-19.pdf>, 2019. Accessed 2021-07-06.

Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission (AI HLEG). Ethics Guidelines for Trust- worthy AI. <https://ec.europa.eu/futurium/en/ai-alliance-consultation>, 2019. Accessed 2021-01-20.

G20. Principles for responsible stewardship of trustworthy AI. <https://www.g20-insights.org/wp-content/uploads/2019/07/G20-Japan-AI-Principles.pdf>, 2019. Section 1. Accessed 2021-01-20.

Floridi, Luciano, et al. "AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations." *Minds and Machines* 28.4 (2018): 689-707.

Dignum, Virginia. *Responsible artificial intelligence: how to develop and use AI in a responsible way*. Springer Nature, 2019.

Vakkuri, Ville, et al. "ECCOLA-a method for implementing ethically aligned AI systems." *Journal of Systems and Software* (2021): 111067.

AlgorithmsWatch. AI Ethics Guidelines Global Inventory. <https://inventory.algorithmwatch.org> 2020. Accessed 2021-07-06.



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



CHALMERS
UNIVERSITY OF TECHNOLOGY



UNIVERSITY OF GOTHENBURG



DISCUSSIONS IN BREAKOUT ROOMS



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



DISCUSSIONS IN BREAKOUT ROOMS

- VALUE-SENSITIVE DESIGN
- RESPONSIBLE AI
- DIFFERENT INITIATIVES FOR ETHICAL AI
- ETHICS OF SELF DRIVING CARS
- AlgorithmsWatch. AI Ethics Guidelines Global Inventory.
<https://inventory.algorithmwatch.org> 2020.
- <https://www.youtube.com/watch?v=5pM6NFb4tqU> Artificial Intelligence: The Ethical and Legal Debate, European Parliament



CHALMERS
UNIVERSITY OF TECHNOLOGY



UNIVERSITY OF GOTHENBURG

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