CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

ETHICS IN THE DESIGN OF INTELLIGENT ARTIFACTS

GORDANA DODIG-CRNKOVIC
CHALMERS UNIVERSITY OF TECHNOLOGY
GOTHENBURG, SWEDEN

AI for Health and Healthy AI conference
Gothenburg, 28th August 2019

http://gordana.se/Presentations

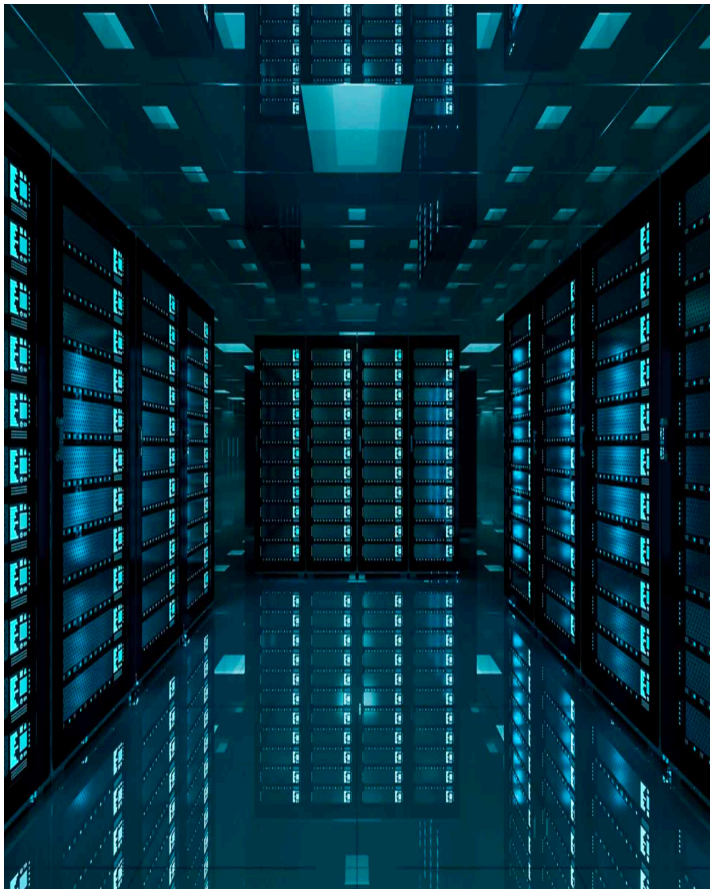
PROCESS OF DIGITALIZATION & COGNITIZATION

INTELLIGENCE & ARTIFICIAL INTELLIGENCE

- Digitalization process is followed by the introduction of cognitive properties and intelligence into artifacts
- Ethical implications of the digitization & cognitization of our society: new technologies followed by **regulatory and legal vacuums**
- My background: theoretical physics, computer science, philosophy of computing and ethics, computational models of cognition, recent interest: interaction design – Interdisciplinary elucidation of AI.

DESIGN OF INTELLIGENT ARTIFACTS

- Ambient intelligence
- Intelligent robots & softbots
- Intelligent transportation systems
- Intelligent cities, Intelligent IoT
- Decision making algorithms
- AI for health: used for intelligent health-care systems with new diagnostic technologies, disease prediction, treatment plans, clinical trials, drug discovery, personalized medicine, and much more



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

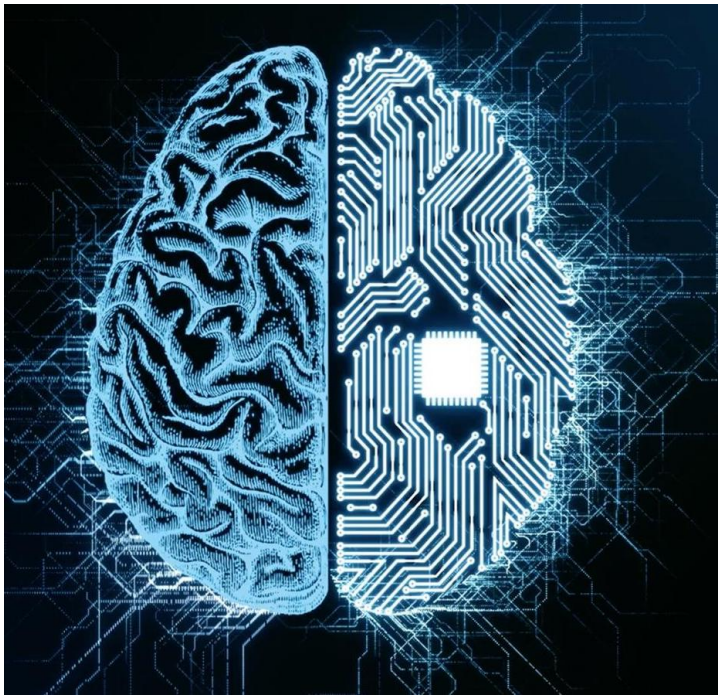
DESIGN OF INTELLIGENT ARTIFACTS

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

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



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

DESIGN OF INTELLIGENT ARTIFACTS

- “Mind reading” technologies
  Massachusetts Institute of Technology, 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
- AI mimicking perception, reasoning and planning
- Delegating decision-making to the intelligent programs
- AI can be autonomous, unpredictable and opaque/inscrutable unlike classical technologies

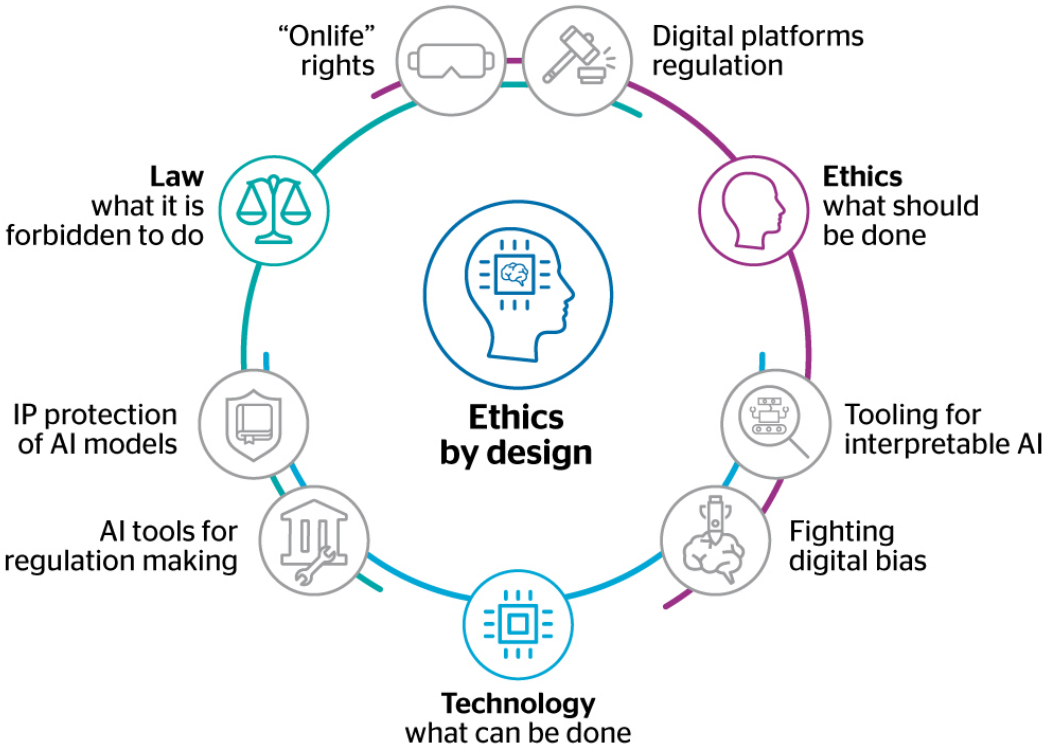
RECURRENT PATTERN: INSPIRATION FROM NATURE (INTELLIGENCE)

QUESTION: HOW MUCH DO WE KNOW ABOUT NATURE?

DO WE WANT INTELLIGENT EVERYTHING?

DO WE WANT TO CONTROL EVERYTHING?

WHY?
QUESTION OF VALUES & ETHICS



https://atos.net/content/mini-sites/journey-2022/human-centric-ai/

VALUE-SENSITIVE DESIGN

Value-sensitive design (VSD) is based on the insight that artefacts are value-laden and design is value-sensitive. We need to identify early implicit values embedded in new technologies by focusing on the use of technology.

"Value" is defined broadly as property that a person or a group considers important in life, and designers can intentionally or unintentionally 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

ETHICAL AI DESIGN EXPECTATIONS

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

TRANSPARENT ETHICAL GOVERNANCE

- DEMOCRATIC LEGITIMACY

- CURRENT CONCENTRATION OF POWER IN THE HANDS OF THE FEW (GOOGLE, APPLE, AMAZON, MICROSOFT)

- FOLLOWING RESPONSIBLE RESEARCH & INNOVATION (EUROPE)*

- IEEE INITIATIVE – EDUCATION – ETHICS IN THE WHOLE PROCESS OF ENGINEERING

- NEW TYPES OF PROBLEMS IN “AUTOMATED PUBLIC SPHERE”

- STAKEHOLDERS INVOLVEMENT IS ESSENTIAL

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

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			
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

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

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 IS	
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 IS	
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 IS	
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. 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

REFERENCES

6. 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

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

11. Virginia Dignum (forthcoming) Responsible Artificial Intelligence, Springer

12. 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)

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

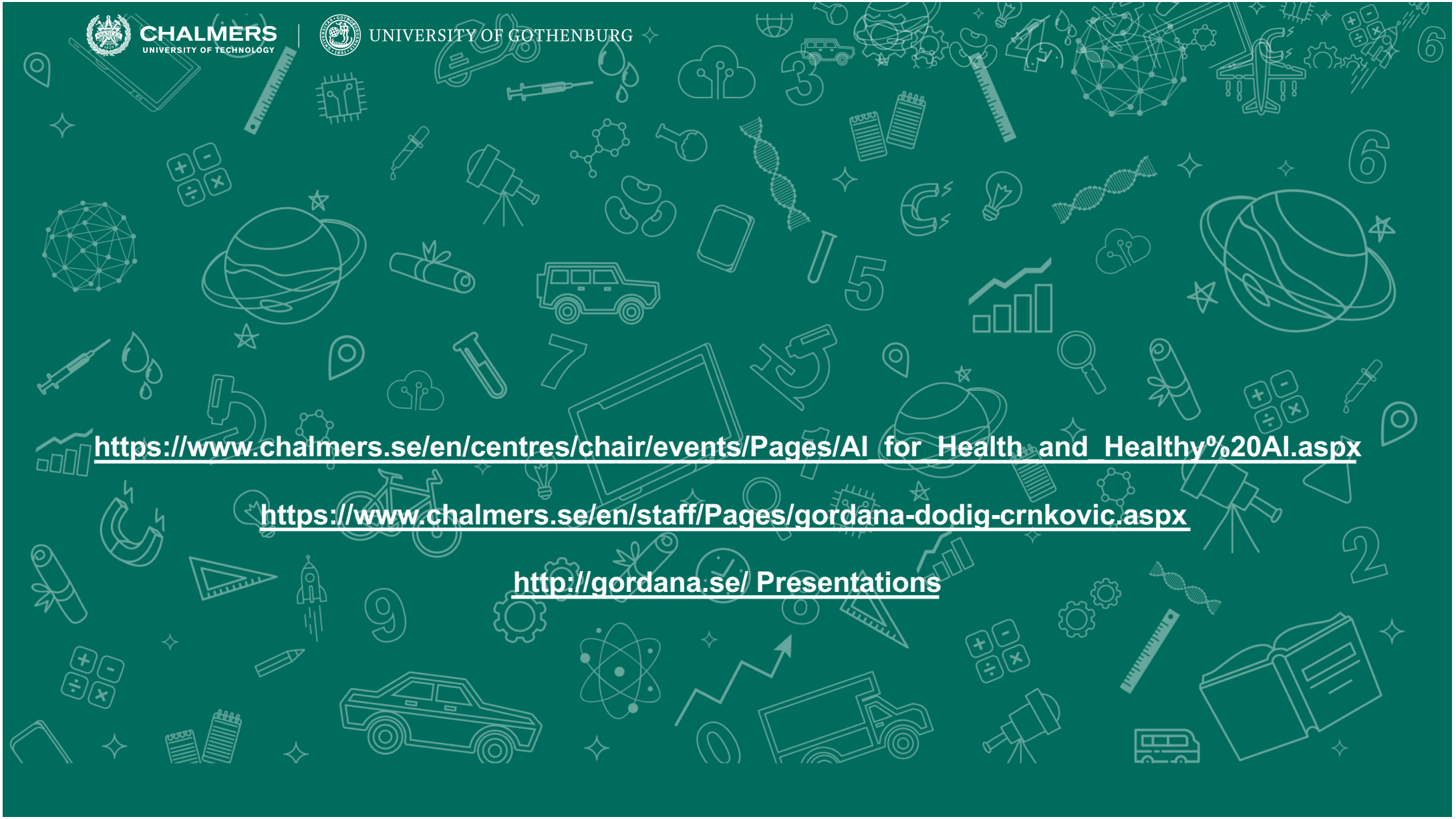
IEEE P7001™ - IEEE Standards Project for Transparency of Autonomous SystemsInspired by the General PrinciplesCommittee, and supported by IEEE Vehicular Technology Society https://standards.ieee.org/project/7001.html

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 August 2019


CHALMERS
UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
https://www.chalmers.se/en/centres/chair/events/Pages/AI_for_Health_and_Healthy%20AI.aspx
https://www.chalmers.se/en/staff/Pages/gordana-dodig-crnkovic.aspx
http://gordana.se/ Presentations

CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG