



ETHICS OF ARTIFICIAL INTELLIGENCE

GORDANA DODIG-CRNKOVIC CHALMERS UNIVERSITY OF TECHNOLOGY GOTHENBURG, SWEDEN

Management of Al projects course

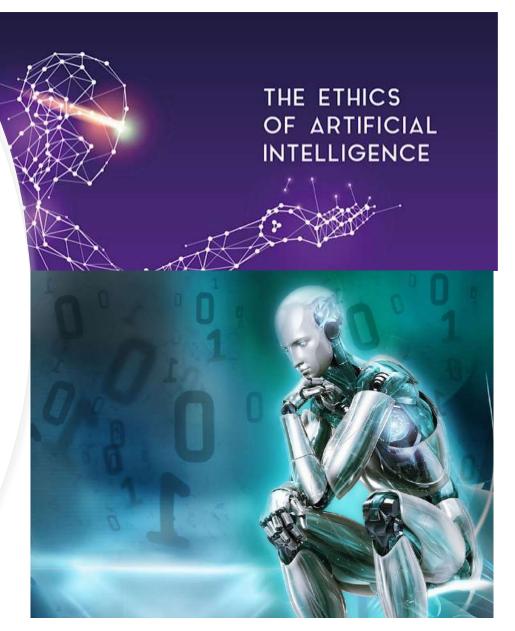
25th November 2020

http://www.gordana.se/work/presentations.html http://gordana.se Zoom meeting URL: <u>https://chalmers.zoom.us/j/69543540316</u> Meeting ID: 695 4354 0316 Passcode: 387834

LECTURE PLAN

14:00 - 14:30 – Ethical aspects of Al 14:30 - 14:45 – Group work 9 in break-out rooms 14:45 – 15:00 – Feedback from the group work 15:00 - 15:15 – Coffee break

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





EU ICT ETHICS WEEK - CURRENTLY ONGOING



ARTIFICIAL INTELLIGENCE & NATURAL INTELLIGENCE

TYPES OF INTELLIGENCE

EMBODIED: HUMAN, CYBORG, ROBOT "DISEMBODIED": SOFTWARE, INFRASTRUCTURE

GENERAL (STRONG): HUMAN LEVEL AND ABOVE NARROW (WEAK): PRESENT INTELLIGENT ARTIFACTS

The most important capacity: learning and meta-learning (learning to learn)



CHALMERS

- Digitalization happens in parallel with introduction of cognitive properties and intelligence into artifacts
- Ethical implications of the digitization of our society: new technologies are 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



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

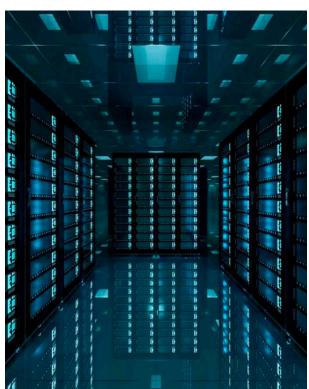


INTELLIGENT ARTIFACTS BEING DEVELOPED

- Ambient intelligence
- Autonomous Intelligent Systems

UNIVERSITY OF GOTHENBURG

- 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-becomingconcern-us-cities/

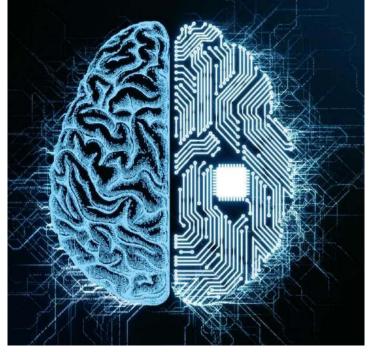
DESIGN OF INTELLIGENT ARTIFACTS

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

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

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

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



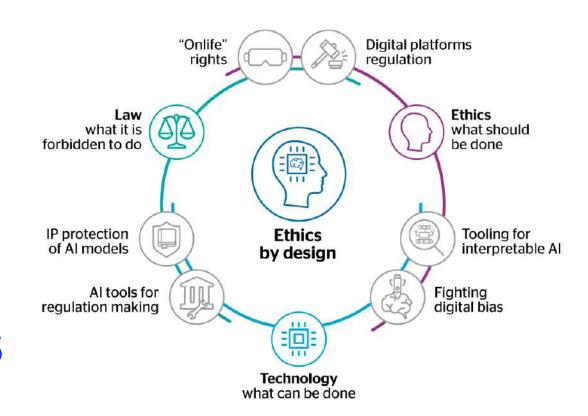
https://www.pbs.org/wgbh/nova/video/nova-wonderscan-we-build-a-brain/





DO WE WANT INTELLIGENT EVERYTHING?

WHY? QUESTION OF VALUES & ETHICS



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

- 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 Al 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

Asilomar Al 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
Al4People (2018) [8]	Beneficence Non-maleficence Autonomy Justice Explicability	An ethical framework for a good AI society	AI



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.	



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.	

8°____



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	Al 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.	

1 Sept



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.	
Responsible & beneficial 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

- Friedman, B.; Kahn, P.H., Jr. (2003) Human values, ethics, and design. In The Human-Computer Interaction Handbook, Fundamentals, Evolving Technologies and Emerging Applications; Jacko, J.A., Sears, A., Eds.; Lawrence Erlbaum: Mahwah, NJ, USA; pp. 1177–1201.
- Friedman, B.; Kahn, P.H., Jr.; Borning, A. (2006) Value sensitive design and information systems. In Human-Computer Interaction in Management Information Systems: Applications; M.E. Sharpe, Inc.: New York, NY, USA; Volume 6, pp. 348– 372.
- 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 <u>https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v1.pd</u>f
- 4. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Ethically Aligned Design, Version 2 for Public Discussion (2017) A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems. Available online: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf



REFERENCES

- 5. Asilomar Conference 2017. Asilomar AI Principles. Available online: <u>https://futureoflife.org/ai-principles/?cn-reloaded=1</u>
- 6. European Group on Ethics in Science and New Technologies (2018) Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems. Available online: <u>https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf</u>
- 7. European Commission's High-Level Expert Group on Artificial Intelligence. Draft Ethics Guidelines for Trustworthy AI (2019) Available online: <u>https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-a</u>i
- Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; et al. (2018) Al4People—An Ethical Framework for a Good Al Society. Minds Mach. 28, 689–707. https://link.springer.com/article/10.1007%2Fs11023-018-9482-5

All links accessed on 16 June 2019





RESPONSIBLE AI

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



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



RESPONSIBLE AI: WHY CARE?

- Al 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

Al 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

ETHICS & DESIGN

Ethics <u>in</u> 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 <u>reasoning</u> abilities as part of the behaviour of artificial autonomous systems

Ethics <u>for</u> Design(ers)

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



ETHICS IN DESIGN: AI – DOING IT RIGHT

Principles for Responsible AI = ART

<u>A</u>ccountability

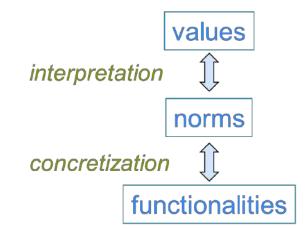
- Explanation and justification
- Design for values

Responsibility

- Autonomy
- Chain of responsible actors
- Human-like Al

<u>Transparency</u>

- Data and processes
- Algorithms
- Choices and decisions





ETHICS **BY** DESIGN

- Can AI artefacts be build to be ethical?
- What does that mean?
- What is needed?
- Understanding ethics
- Using ethics
- Being ethical

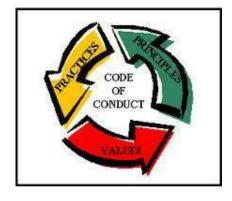




ETHICS <u>FOR</u> DESIGN(ERS) – REGULATION, CODES OF CONDUCT

- A code of conduct clarifies mission, values and principles, linking them with standards and regulations
 - Compliance
 - Risk mitigation
 - \circ Marketing
- Many professional groups have regulations
 - Medicine / Pharmacy
 - Accountants
 - Architects
 - o Military
- Regulation, accreditation: when society relies on a profession!







ETHICAL REASONING IS <u>OPEN-ENDED</u>

Normative reasoning

Utilitarian/Consequestialist car

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

Deontologic/Kantian car

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

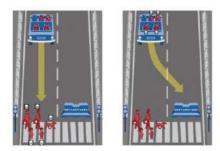
Virtuous/Aristotelian car

Aristotle's ethics is about how to live the good life (eudaimonia) based on virtues.

Act as a virtuous person. For Aristotle, there are three primary moral virtues: courage, temperance, and justice.







AI FOR GOOD, AI FOR PEOPLE, ...

() UNIVERSITY OF GOTHENBURG

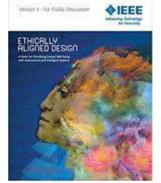
- Harness the positive potential outcomes of AI in society, the economy
- Ensure inclusion, diversity, universal benefits

CHALMERS

- 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/highlevel-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

TAKE AWAY MESSAGE ON **RESPONSIBLE AI**

- (Currently) AI systems are artefacts built by us for our own purposes
 - Our decision, our responsibility
- Al influences and is influenced by our social systems
 - Design in never value-neutral
 - $_{\circ}$ $\,$ Society shapes and is shaped by design
- Knowing ethics is not being ethical
 - $_{\circ}$ $\,$ 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 <u>ways</u> in different <u>roles</u>



DIVISION/ASSIGNMENT OF RESPONSIBILITY

Time perspective

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

<u>Stakeholder's roles</u>

- 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). <u>tiny.cc/1tnvez</u>

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

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

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

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

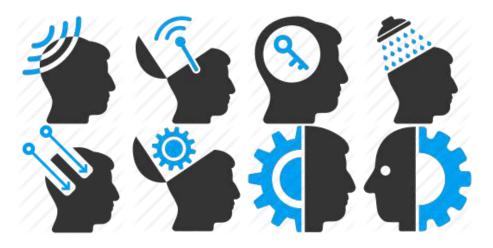
Rafael Capurro (2019) THE AGE OF ARTIFICIAL INTELLIGENCES. Contribution to the AI, Ethics and Society Conference, University of Alberta, Edmonton (Canada), May 8-10, 2019. <u>http://www.capurro.de/edmonton2019.html</u>

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





SUMMARY: ETHICS OF AI



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

HIGHLIGHTS OF ETHICAL ISSUES OF AI

- REGULATION value-based, codes of ethics & laws
- RESPONSIBILITY
- TRANSPARENCY
- PRIVACY & INTEGRITY. GDPR <u>https://www.gdprexplained.eu</u> General Data Protection Regulation – data protection by design
- BIAS/FAIRNESS in machine classification systems (algorithmic bias) & decision-making
- Al-guided weapon Systems Lethal Autonomous Weapons security & responsibility
- Agency and moral status of Al
- Future of work & end of employment job replacement and redistribution
- Human dependency on technology and loss of skills
- Value-misalignment
- Unintended consequences of goals and decisions



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

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

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

Autonomy: the power to decide (whether to decide)

Justice/Fairness: promoting prosperity and preserving solidarity

Explicability: enabling the other principles through intelligibility and accountability

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



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



RESOURCES

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

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

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

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

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

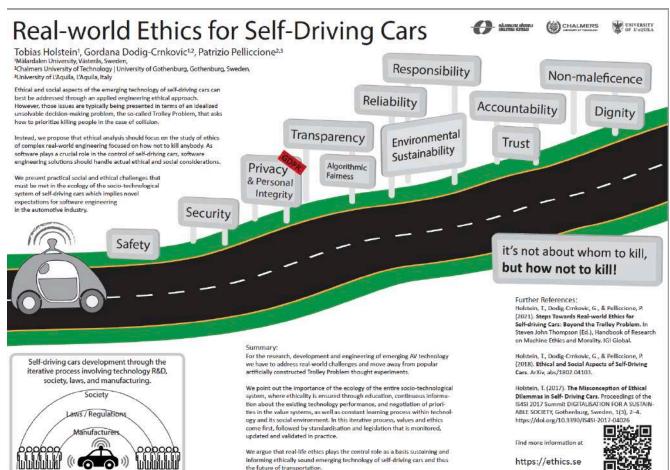
https://www.microsoft.com/en-us/ai/ai-for-good AI for Earth, AccessibilityHumanitarian Action, Cultural Heritage

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

CHALMERS

(E) UNIVERSITY OF GOTHENBURG

EXAMPLE: EHICS OF SELF-DRIVING VEHICLES



https://ethics.se

Premier Reference Source

Machine Law, Ethics, and Morality in the Age of Artificial Intelligence



Presented as poster at ICSE2020

Extended version is a book chapter: Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). <u>Steps Towards</u> <u>Real-world Ethics for Self-driving</u> <u>Cars: Beyond the Trolley Problem</u>. In Steven John Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global

https://www.igi-global.com/book/machine-law-ethics-morality-age/244609#table-of-contents

Please provide examples of ethical questions in each category

Challenge/ Theme	Ethical Questions
Project Related Challenge	 What Harm Can the Project Do? Who May Be Harmed by the Project? Do we know the Consequences of relevant Laws & Regs? Is Ethical Accountability in Place? Is Ethical Oversight Sufficient?
Data Related Challenges	 Are the Data Sources Ethically Sound? Are Individuals' Privacy and Anonymity Impinged? Can Data Quality Issues Cause Unethical Outcomes?
Model Related Challenges	 Is the Model Fair? Is the Model Transparent? Are Project Objectives Structured for Objectivity? Can the Result be Misinterpreted? Should the Model Operate Autonomously?

https://www.youtube.com/watch?v=C5JElgliTeE&t=2467s&ab_channel=OpenDataScience

QUESTIONS FOR YOUR DISCUSSION







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