Human-Centered Computing
Research Group

Freie Universität Berlin

«Human-Centered Data Science»

Discussing Human Values and their Impact on Your Data Science Practice

Prof. Dr. Claudia Müller-Birn

Human-Centered Computing, Institute of Computer Science

Freie Universität Berlin

May 05, 2022

Lecture Overview

Recap

Motivation for Considering Ethics (Defining Ethics, Naming Ethical Theories, Examples of Applied Ethics, Deriving Principles of a Ethical Practice, Risks of Applying Principles)

Case Study Reflection

Break

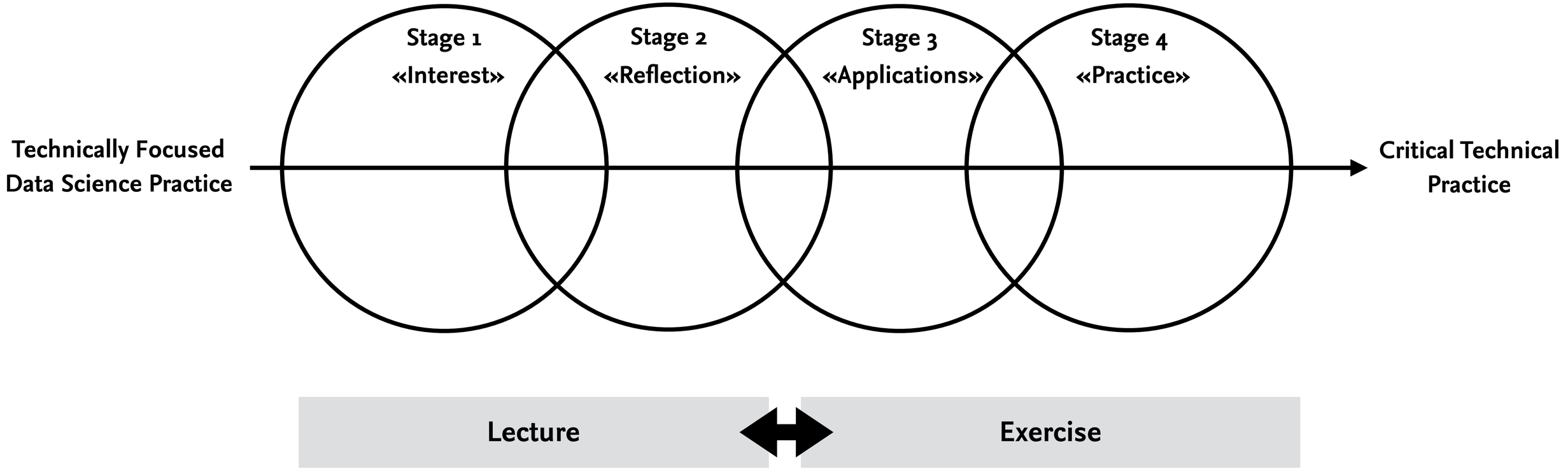
Incorporating Principles in Your Data Science Practice (Principles, Ethos, Values, Value Levers, Remaining Challenge)

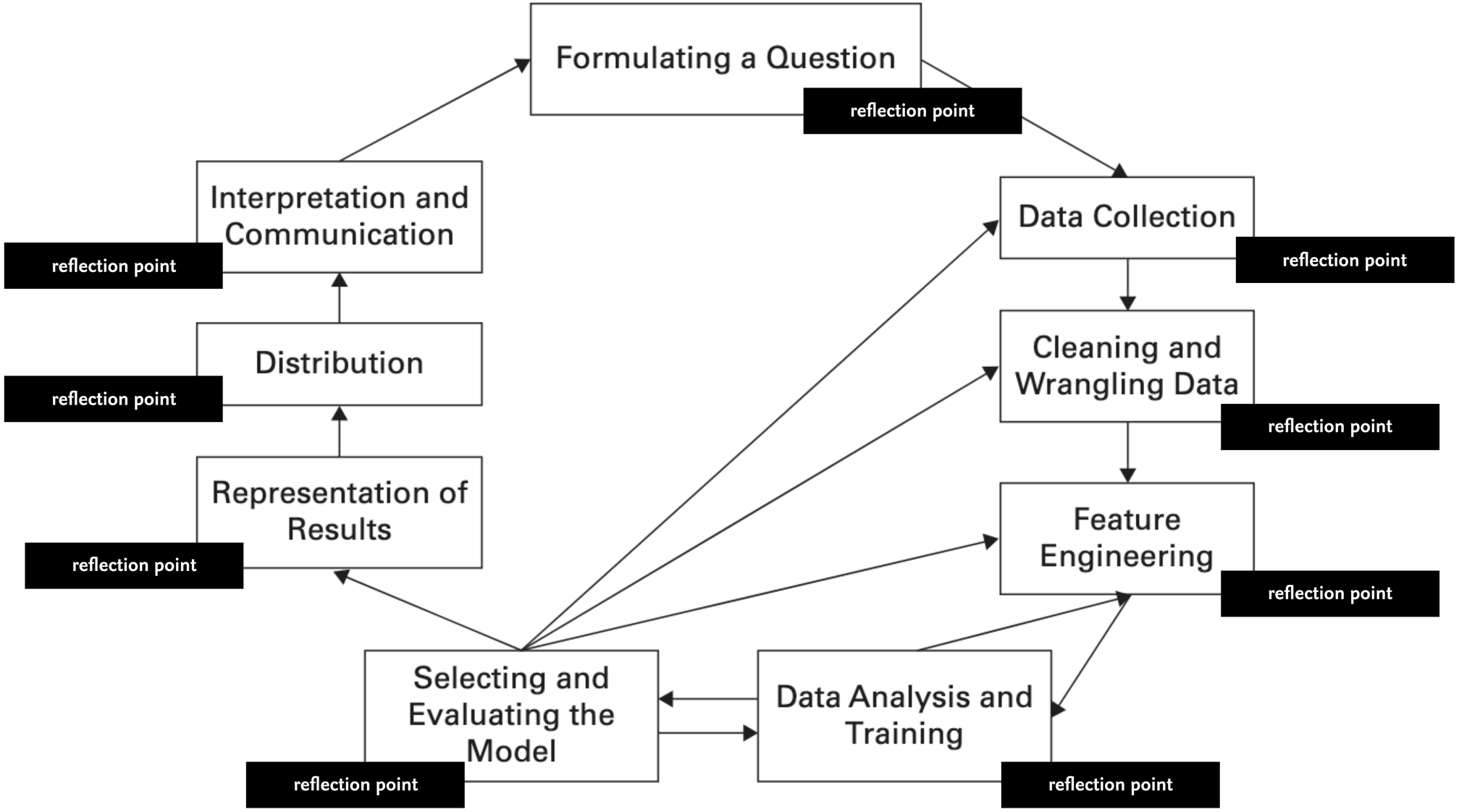
Recap

> Human-Centered Data Science (HCDS) is the study of data based on a human-centered perspective which requires methods that both harness the automated power of computational techniques and account for the highly situated and nuanced nature of social activity.

Kogan, M., Halfaker, A., Guha, S., Aragon, C., Muller, M., & Geiger, S. (2020). Mapping Out Human-Centered Data Science (pp. 151–156). GROUP '20: The 2020 ACM International Conference on Supporting Group Work, New York, NY, USA: ACM.

Concept: Establish a Critical-Reflexive Cycle





Aragon, Cecilia, Shion Guha, Marina Kogan, Michael Muller, und Gina Neff. Human-Centered Data Science: An Introduction. Cambridge, MA, USA: MIT Press, 2022.

Motivation

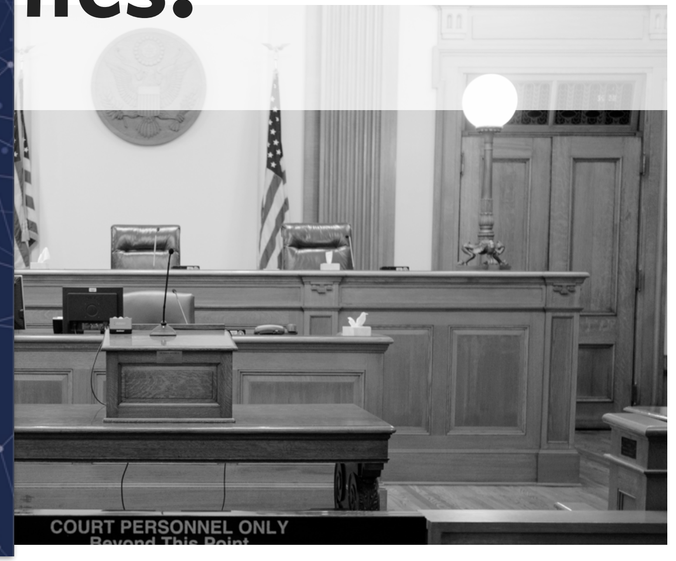

NYPD
COUNTERTERRORISM


SCHOOL BUS
EMERGENCY DOOR
LK21251
STARCRAFT
522


ID : 254876592
MALE
BROWN HAIR
CAUCASIAN
STRESSED
ID : 548765942
MALE
GREY HAIR
CAUCASIAN
RELAXED
BAG
SYSTEM
RECOGNITION
IN PROGRESS ...
27%
ID : 458625943
MALE
CAUCASIAN
RELAXED
BAG


COURT PERSONNEL ONLY

We n hics!



https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

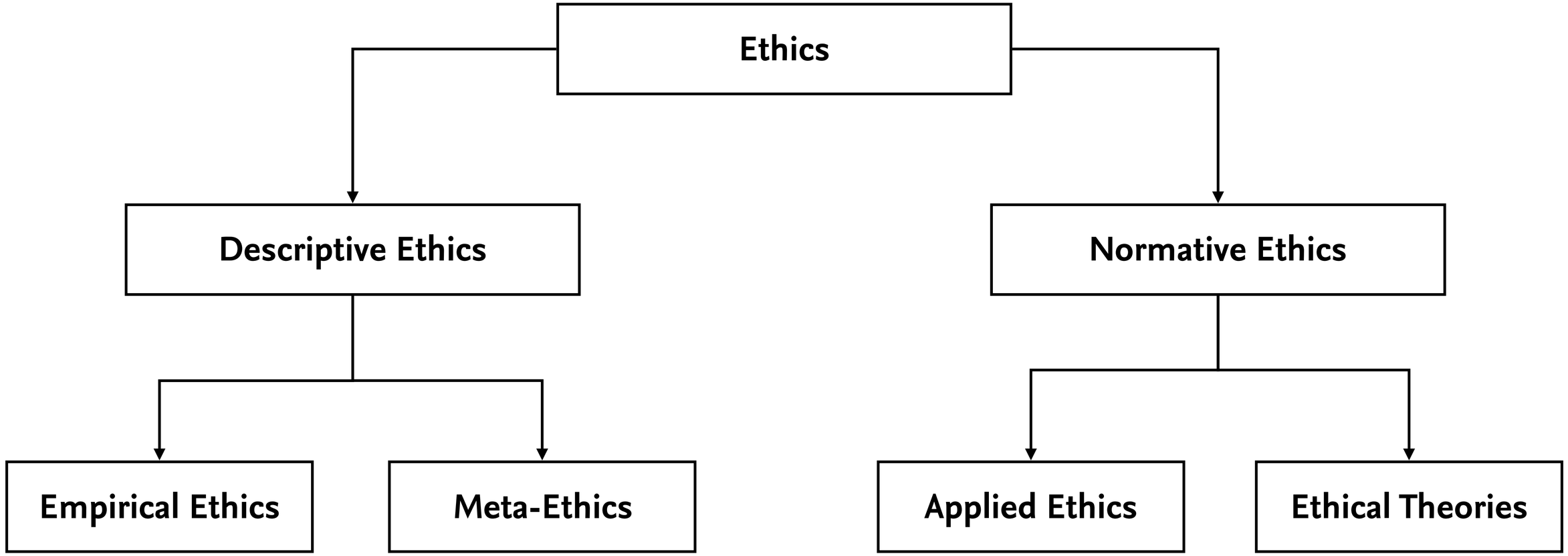
What is Ethics?

"Ethics as a discipline is the philosophical and systematic study of morality which is expressed in enquiries into the nature of the good life, how we should live, what kind of society we want to live in, and how we should treat others."

Ethics is a philosophical discipline.

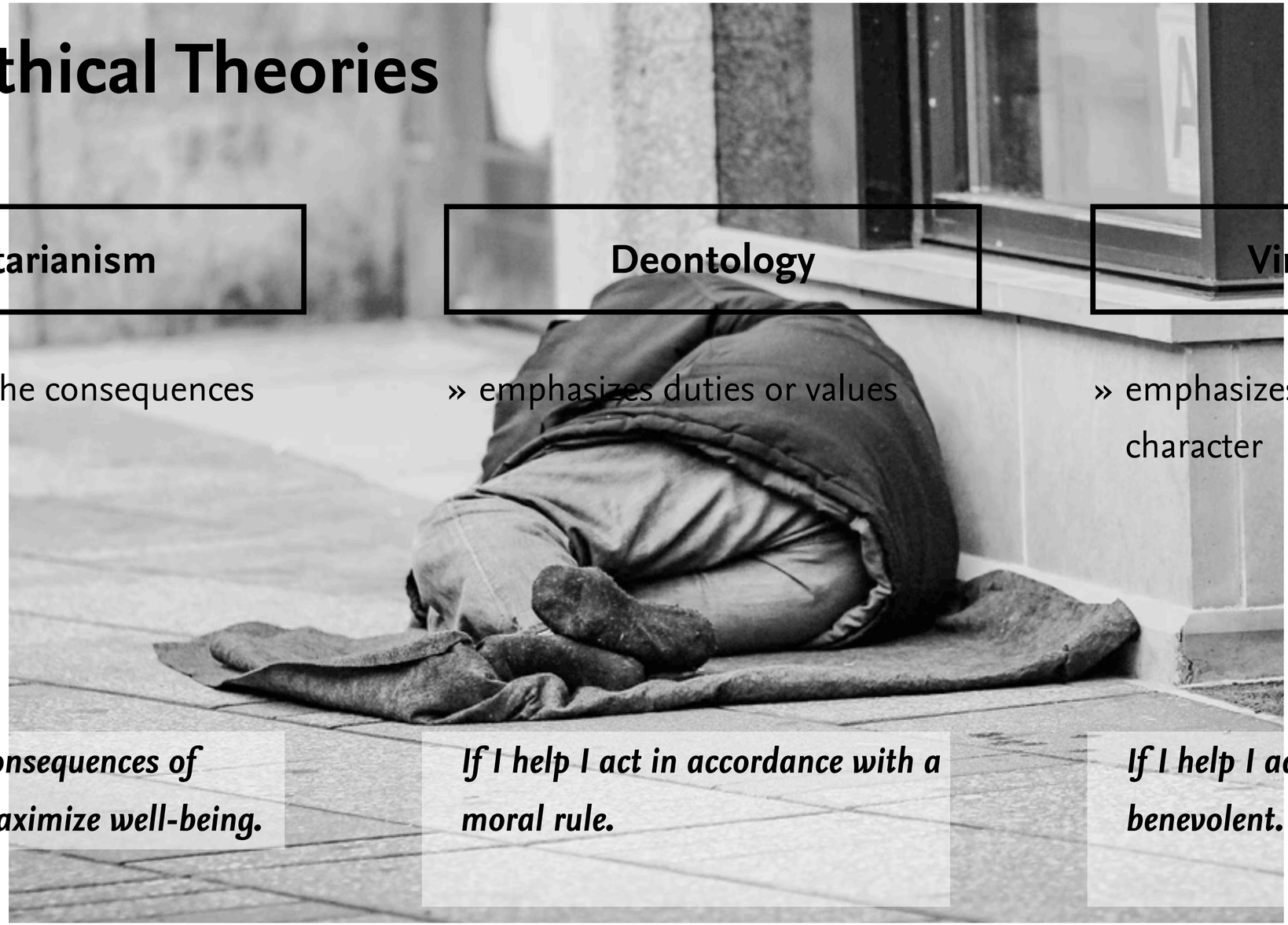
Morality is a set of moral norms, feelings, attitudes, or actions.

Frauenberger, Christopher, Marjo Rauhala, und Geraldine Fitzpatrick. „In-Action Ethics". _Interacting with Computers_ 29, Nr. 2 (8. März 2017): 220–36.



Misselhorn, Catrin. Grundfragen der Maschinenethik. 4., Durchgesehene und Überarbeitete Auflage. Reclams Universal-Bibliothek, Nr. 19583. Ditzingen: Reclam, 2019.

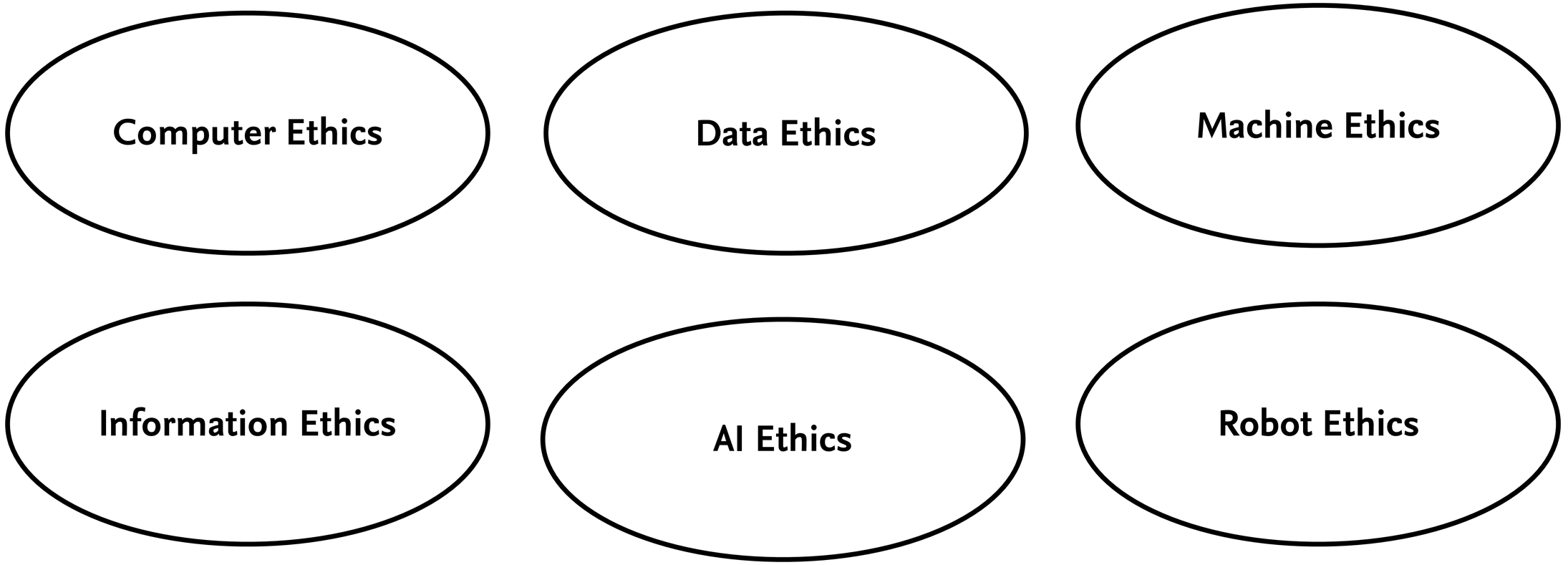
Major Ethical Theories



Utilitarianism	Deontology	Virtue Ethics
» emphasizes the consequences of actions	» emphasizes duties or values	» emphasizes virtues, or moral character
If I help, the consequences of helping will maximize well-being.	*If I help I act in accordance with a moral rule.*	*If I help I act charitable or benevolent.*

Misselhorn, Catrin. Grundfragen der Maschinenethik. 4., Durchgesehene und Überarbeitete Auflage. Reclams Universal-Bibliothek, Nr. 19583. Ditzingen: Reclam, 2019.

Selected Areas of Applied Ethics



Misselhorn, Catrin. Grundfragen der Maschinenethik. 4., Durchgesehene und Überarbeitete Auflage. Reclams Universal-Bibliothek, Nr. 19583. Ditzingen: Reclam, 2019.

Human-
Researc

ersität Berlin

Ethical Guidelines of the German Info

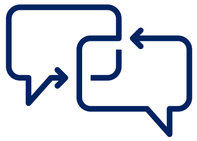
Professional accountability deals with accountability relations between a professional and their peer group.

SECTION 1: PROFESSIONAL COMPETENCE

GI members stay abreast of the current state of science and technology in their respective areas of specialization; they take new developments into account and provide constructive criticism.
GI members are constantly working to improve their professional competence.

SECTION 2: EXPERTISE AND COMMUNICATIVE COMPETENCE

GI members are constantly improving their levels of expertise and communicative competencies in order to meet the demands relevant to their duties in the design, manufacture, operation and use of IT systems and to understand the surrounding professional and technical contexts. In order to assess the consequences of IT-systems in the application environment and to propose suitable solutions, there must be a willingness to understand and take into account the rights, needs and interests of those parties who are impacted by them.



SECTION 3: LEGAL EXPERTISE

GI members are familiar with and observant of pertinent legal regulations concerning the design, manufacture, operation and use of IT systems. GI members, in conjunction with their expertise and professional competencies, participate actively in drafting legislative regulations.

SECTION 4: POWERS OF DISCERNMENT

GI members sharpen their powers of discernment to render themselves better equipped to contribute to design processes with individual and collective accountability. This presupposes not only a willingness to call into question and to make judgments about individual and collective actions in public discourse, but also the ability to acknowledge the limits of one's own powers of discernment.

SECTION 5: CONDITIONS OF EMPLOYMENT

GI members are active proponents of socially equitable contractual agreements concerning terms of employment, inclusive of opportunities for professional [illegible] and [illegible].

SECTION 6: ORGANIZATIONAL STRUCTURES

GI members advocate for organizational structures which foster and facilitate socially equitable contractual agreements concerning terms of employment.

SECTION 7: TEACHING AND LEARNING

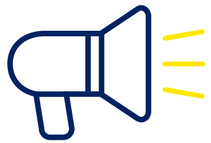
GI members who are computer science instructors foster in their students the capacity for critical thinking; they prepare learners to accept their own individual and collective responsibility, and they act as role models in this regard.

SECTION 8: RESEARCH

GI members who conduct research in the field of computer science adhere to the rules of best practices in scientific research. Of particular importance in this regard is openness and transparency in dealing with criticism and conflicts of interest, the ability to express and to accept criticism as well as the willingness to allow the impact of one's own scientific work in the research process to become the subject of discussion. Scientific research breaches boundaries. These must be clearly articulated.

SECTION 9: COURAGE OF CONVICTIONS

GI members staunchly advocate for the protection and safeguarding of human dignity, even when this is not explicitly mandated by laws, contracts or other norms, or when these stand in direct opposition to the protection and safeguarding of human dignity. This applies even in situations in which GI members' obligations to clients conflict with their responsibility to third-party stakeholders.



SECTION 10: SOCIAL ACCOUNTABILITY

In the development, manufacture, operation and use of IT systems, GI members should contribute to the betterment of local and global living conditions. GI members are responsible for the social and societal consequences of their work. Their influence on positioning, marketing and further development of IT systems should contribute to the socially acceptable and sustainable application of these technologies.

SECTION 11: FACILITATING SELF-DETERMINATION

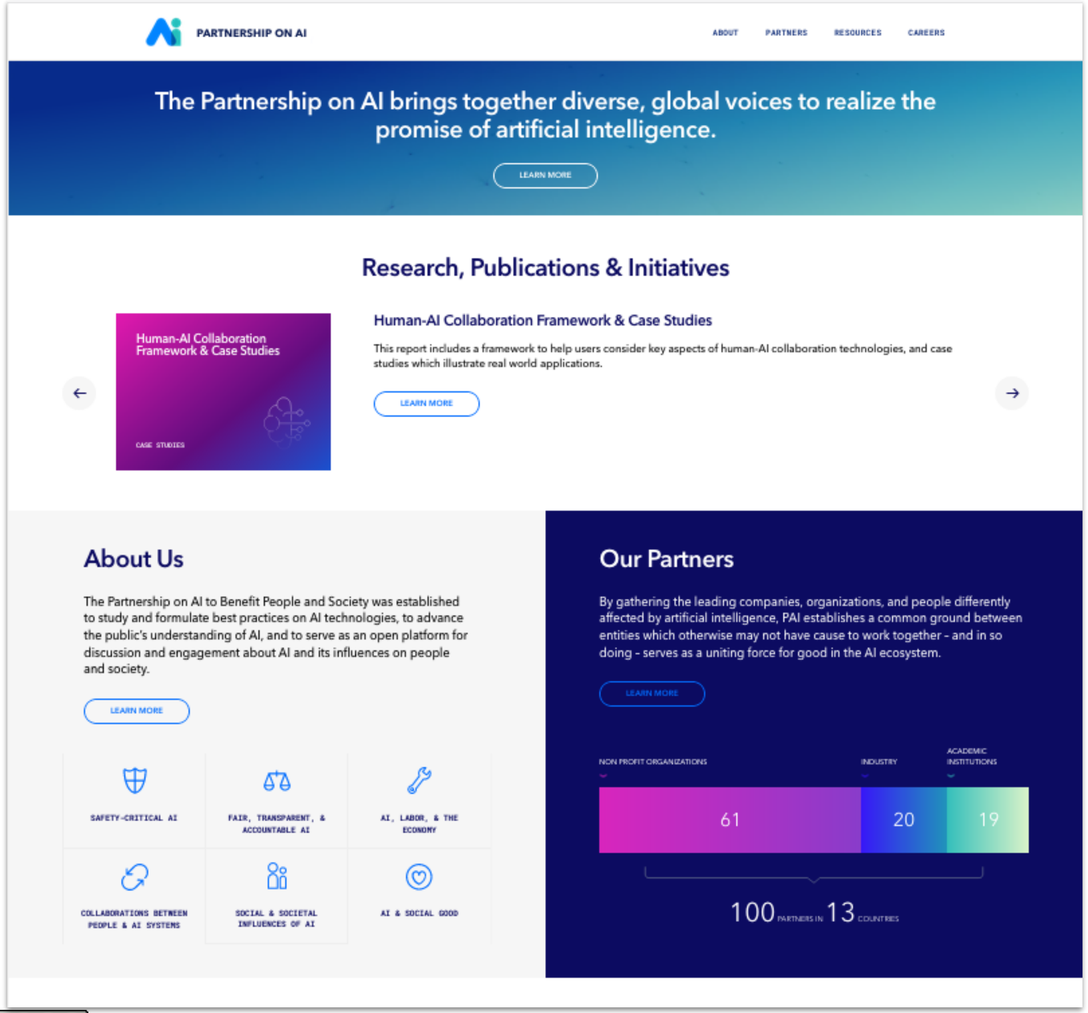
GI members work toward ensuring that those people impacted by the usage and conditions of use of IT systems are granted adequate opportunity to participate in the design of these systems. This is especially pertinent with regard to systems whose application involves the exerting influence over, monitoring, or surveillance of said populations.

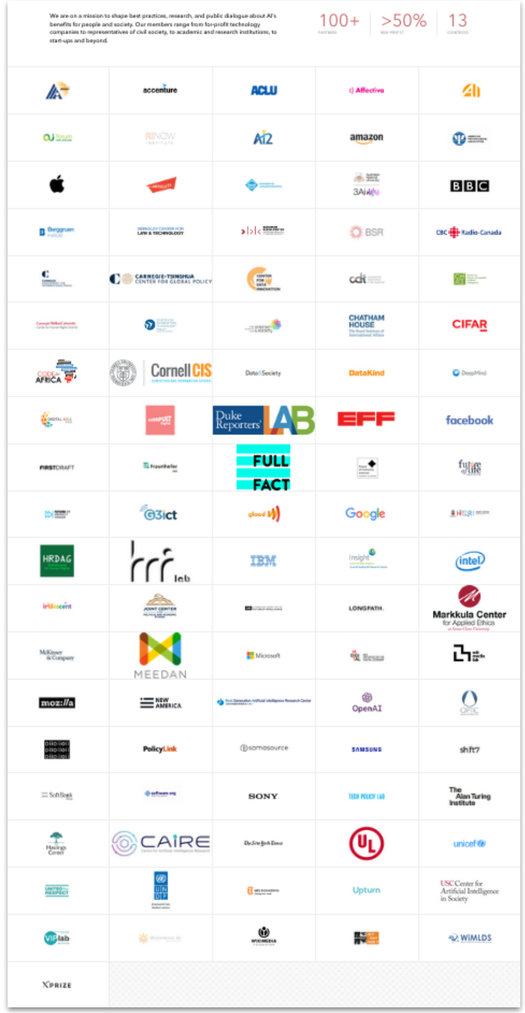
SECTION 12: THE GERMAN INFORMATICS SOCIETY

The German Informatics Society encourages its members to adhere to these guidelines at all times. The GI shall attempt to mediate between parties in situations in which conflicts arise.

Guidelines taken from https://gi.de/fileadmin/GI/Allgemein/PDF/GI_Poster_Ethical_Guidlines.pdf

Ethical Guidelines between Companies





Our Goals

1 Develop and share best practices

Support research, discussions, identification, sharing, and recommendation of best practices in the research, development, testing, and fielding of AI technologies. Address such areas as fairness and inclusivity, explanation and transparency, security and privacy, values and ethics, collaboration between people and AI systems, interoperability of systems, and of the trustworthiness, reliability, containment, safety, and robustness of the technology.

2 Advance public understanding

Advance wider public understanding and awareness of AI by sharing insights into AI's core technologies, potential benefits – and costs. We will act as a trusted experts on AI for society and their leaders, and will work to increase public understanding of how AI is progressing.

3 Provide an open and inclusive platform for discussion & engagement

Create and support opportunities for AI researchers and key stakeholders, including people in technology, law, policy, government, civil liberties, and the greater public, to communicate directly and openly with each other about relevant issues to AI and its influences on people and society. Ensure that key stakeholders have the knowledge, resources, and overall capacity to participate fully.

4 Identify and foster aspirational efforts in AI for socially beneficial purposes

Seek out, support, celebrate, and highlight aspirational efforts in AI for socially benevolent applications. Identify areas of untapped opportunity, including promising technologies and applications not being explored by academia and industry R&D.

https://www.partnershiponai.org/

Ethical Guidelines within Companies

Oversight Board

APPEALS PROCESS NEWS DECISIONS

THE BOARD

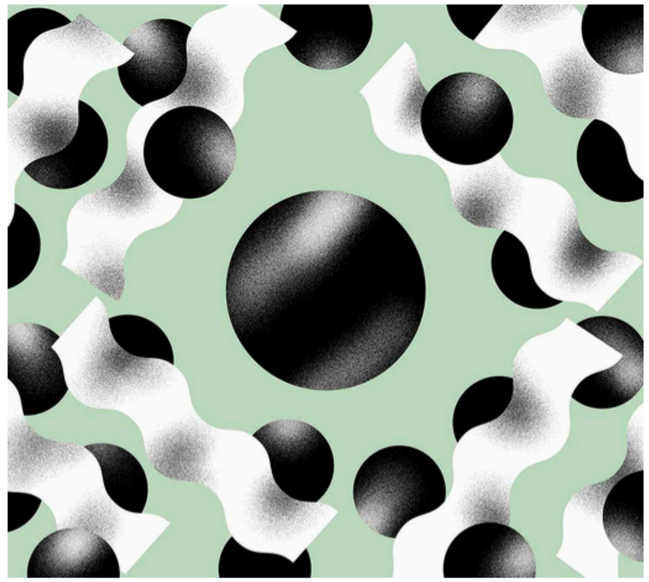
The Purpose of the Board

The purpose of the board is to promote free expression by making principled, independent decisions regarding content on Facebook and Instagram and by issuing recommendations on the relevant Facebook company content policy.

When fully staffed, the board will consist of 40 members from around the world that represent a diverse set of disciplines and backgrounds. These members will be empowered to select content cases for review and to uphold or reverse Facebook's content decisions.

The board is not designed to be a simple extension of Facebook's existing content review process. Rather, it will review a select number of highly emblematic cases and determine if decisions were made in accordance with Facebook's stated values and policies.

MEET THE BOARD



https://oversightboard.com/

Codes of Ethics for Responsible AI

Table 1 | Ethics guidelines for AI by country of issuer (Australia–UK)

Name of document/website	Issuer	Country of issuer
Artificial Intelligence. Australia's Ethics Framework: A Discussion Paper	Department of Industry Innovation and Science	Australia
Montréal Declaration: Responsible AI	Université de Montréal	Canada
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
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
Discussion Paper: National Strategy for Artificial Intelligence	National Institution for Transforming India (NITI Aayog)	India
L'intelligenzia Artificiale al Servizio del Cittadino	Agenzia per l'Italia Digitale (AGID)	Italy
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
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
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
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

Table 2 | Ethics guidelines for AI by country of issuer (USA, international, EU and N/A)

Name of document/website	Issuer	Country of issuer
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	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	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	USA
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
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

Continued

Table 2 | Ethics guidelines for AI by country of issuer (USA, international, EU and N/A) (Continued)

Name of document/website	Issuer	Country of issuer
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
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

Jobin, Anna, Marcello Ienca, und Effy Vayena. „The global landscape of AI ethics guidelines". Nature Machine Intelligence 1, Nr. 9 (1. September 2019): 389–99. https://doi.org/10.1038/s42256-019-0088-2.

Codes of Ethics for Responsible AI

Table 1 | Ethics guidelines for AI by country of issuer (Australia–UK)

Name of document/website	Issuer	Country of issuer
Artificial Intelligence. Australia's Ethics Framework: A Discussion Paper	Department of Industry Innovation and Science	Australia
Montréal Declaration: Responsible AI	Université de Montréal	Canada
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
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
Discussion Paper: National Strategy for Artificial Intelligence	National Institution for Transforming India (NITI Aayog)	India
L'intelligenzia Artificiale al Servizio del Cittadino	Agenzia per l'Italia Digitale (AGID)	Italy
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 by 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
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
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
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

Table 2 | Ethics guidelines for AI by country of issuer (USA, international, EU and N/A)

Name of document/website	Issuer	Country of issuer
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	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
[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]
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	USA
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
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

Continued

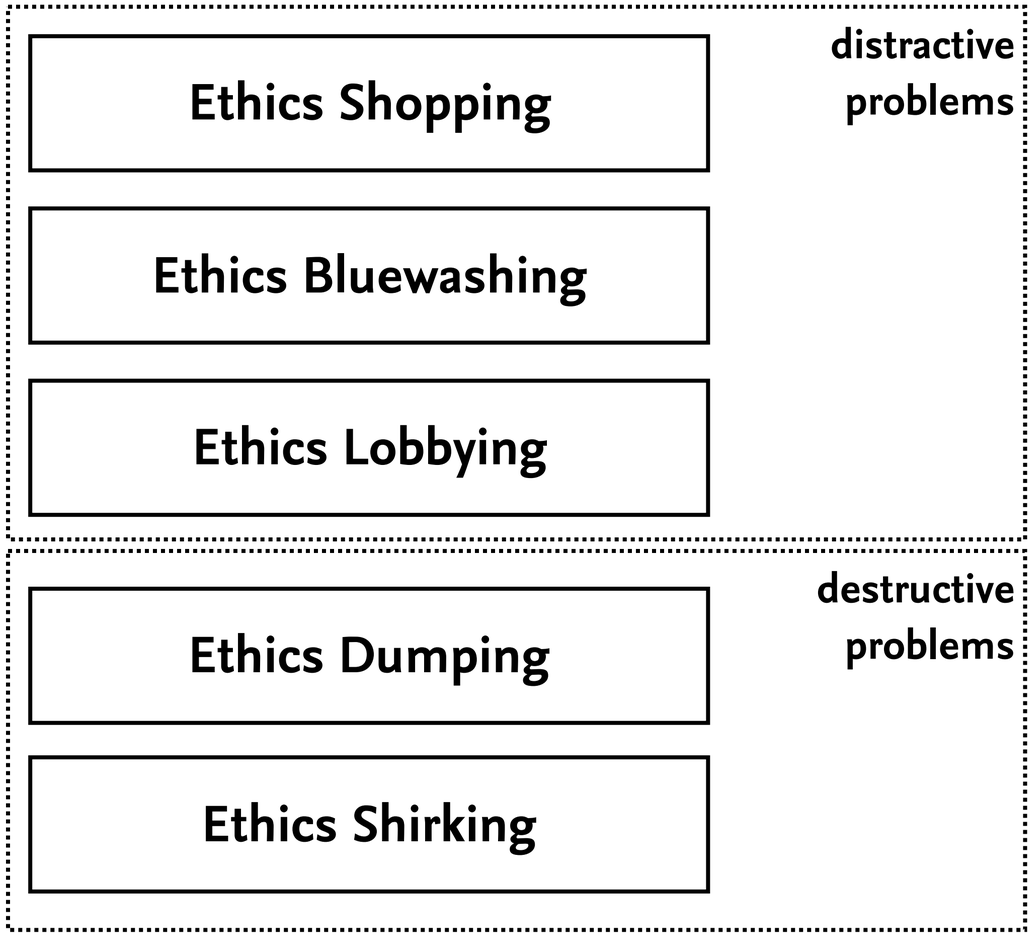
Table 2 | Ethics guidelines for AI by country of issuer (USA, international, EU and N/A) (Continued)

Name of document/website	Issuer	Country of issuer
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
Asilomar AI Principles	Future of Life Institute	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

Does serve Ethics as Fig Leaf?

Jobin, Anna, Marcello Ienca, und Effy Vayena. „The global landscape of AI ethics guidelines". Nature Machine Intelligence 1, Nr. 9 (1. September 2019): 389–99. https://doi.org/10.1038/s42256-019-0088-2.

Translating Principles to Practice - What are the Risks?

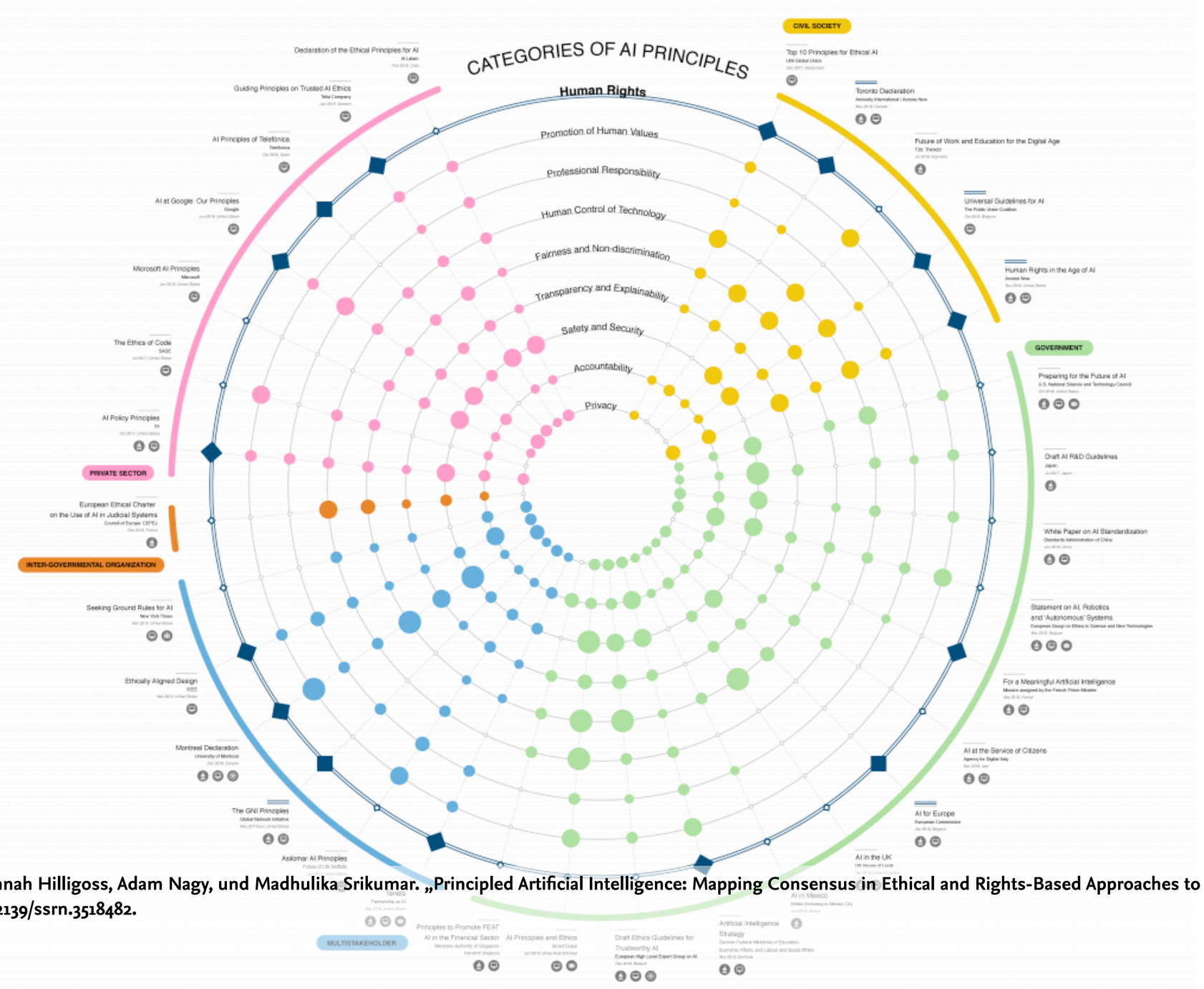


Floridi, Luciano. „Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical“. _Philosophy & Technology_ 32, Nr. 2 (1. Juni 2019): 185–93. https://doi.org/10.1007/s13347-019-00354-x.

Translating Principles into Practice - What are the Risks?

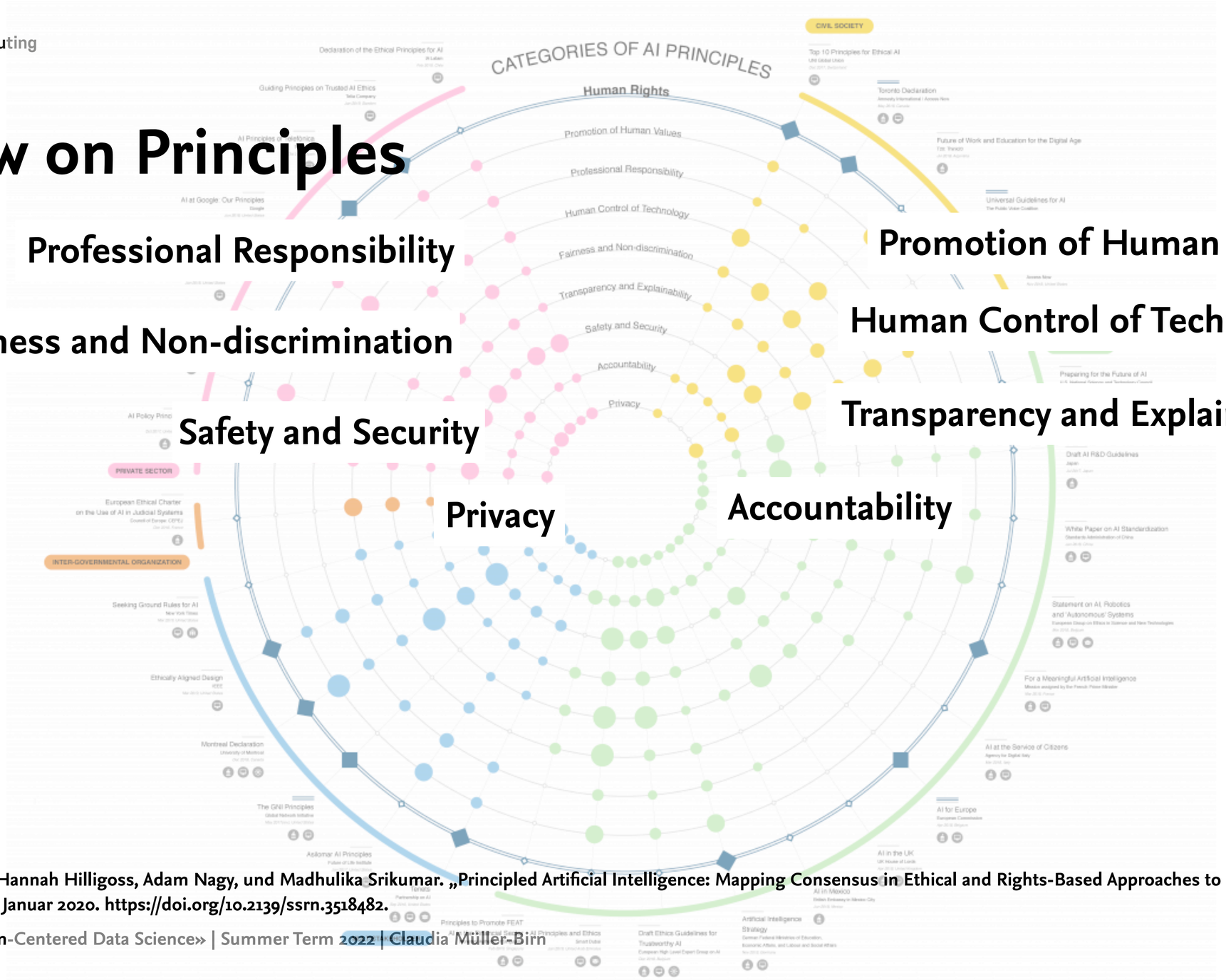
Ethics Shopping	… the malpractice of choosing, adapting, or revising ("mixing and matching") ethical principles, guidelines, etc.
Ethics Bluewashing	… the malpractice of making unsubstantiated or misleading claims about, or implementing superficial measures.
Ethics Lobbying	… the malpractice of exploiting digital ethics to delay, revise, replace, or avoid good and necessary legislation.
Ethics Dumping	… the malpractice of exporting AI research activities in other contexts in ways that would be ethically unacceptable in the original context.
Ethics Shirking	… the malpractice of doing increasingly less ethical work in a given context the lower the return of such ethical work is mistakenly perceived to be.

Floridi, Luciano. „Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical". _Philosophy & Technology_ 32, Nr. 2 (1. Juni 2019): 185–93. https://doi.org/10.1007/s13347-019-00354-x.



Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, und Madhulika Srikumar. „Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI“. SSRN, 2020. https://doi.org/10.2139/ssrn.3518482.

Overview on Principles



Professional Responsibility

Promotion of Human Values

Fairness and Non-discrimination

Human Control of Technology

Safety and Security

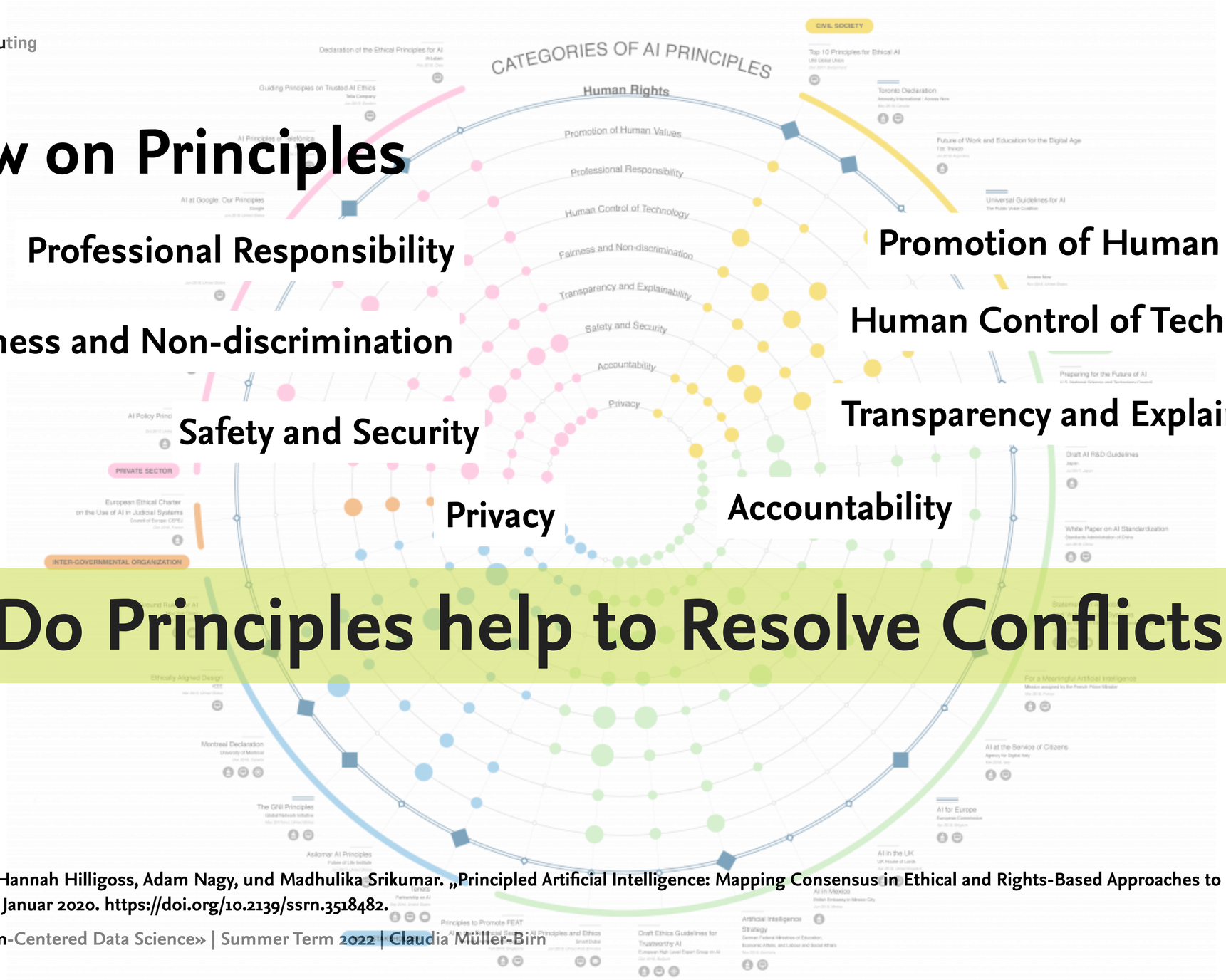
Transparency and Explainability

Privacy

Accountability

Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, und Madhulika Srikumar. „Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI“. SSRN Electronic Journal, 15. Januar 2020. https://doi.org/10.2139/ssrn.3518482.

Overview on Principles



Professional Responsibility

Promotion of Human Values

Fairness and Non-discrimination

Human Control of Technology

Safety and Security

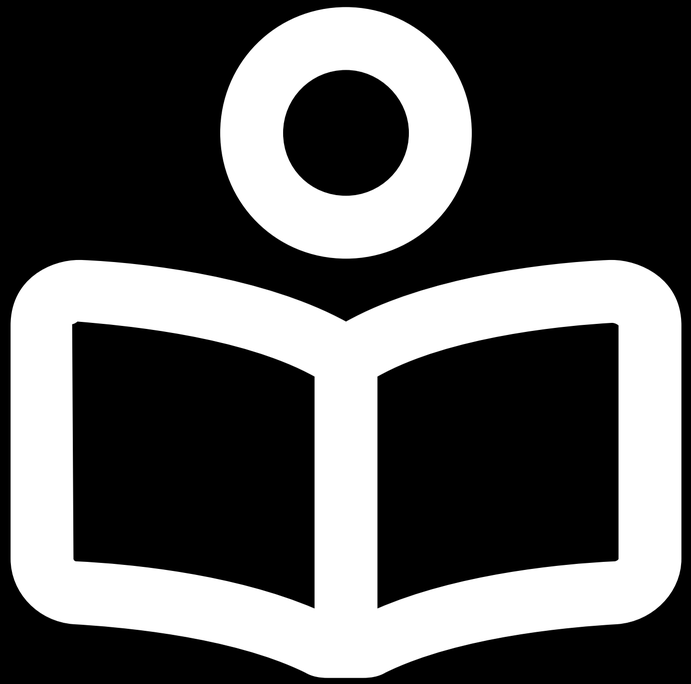
Transparency and Explainability

Privacy

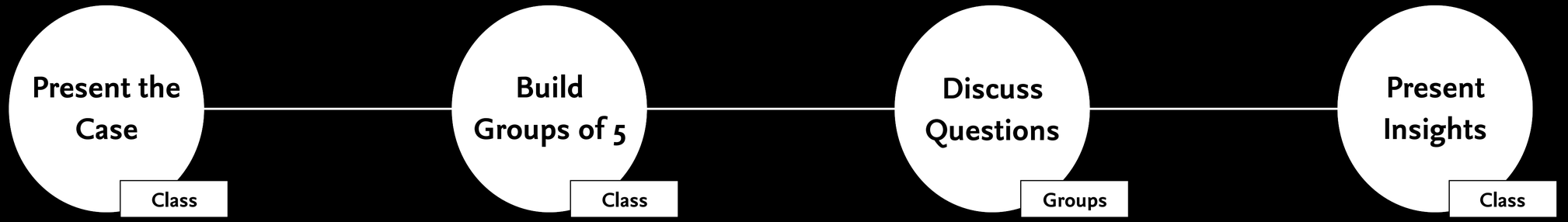
Accountability

Do Principles help to Resolve Conflicts?

Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, und Madhulika Srikumar. „Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI“. SSRN Electronic Journal, 15. Januar 2020. https://doi.org/10.2139/ssrn.3518482.



Case Study Reflection

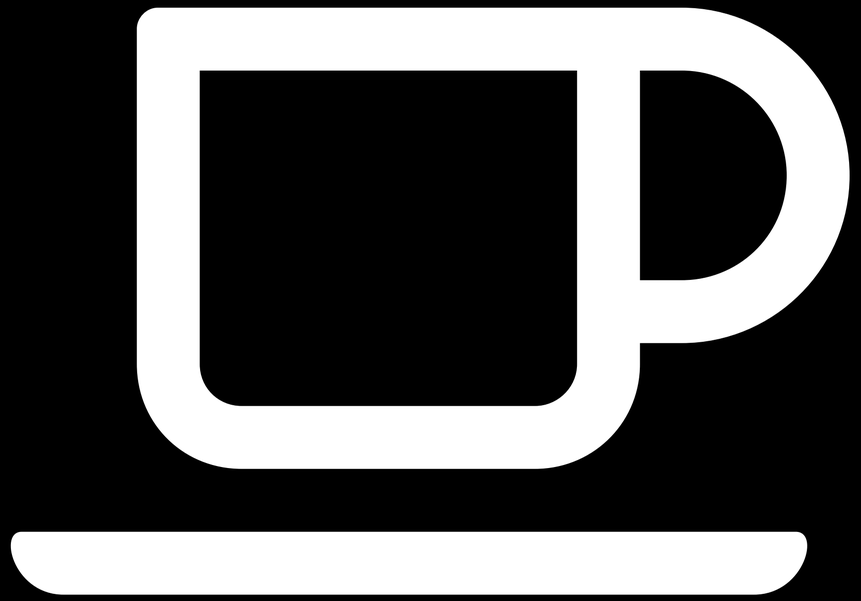


1. For which of the two focus topics (tracing app and triage assistance system) do partner countries support the unified requirements framework and also adherence to ethical principles?

2. Does a partner only support the ethical principles if they only adhere to the requirements framework for both focus topics?

3. How do you evaluate the lack of conviction of some partners to the ethical principles, if they nevertheless support and comply with them?

4. In which constellations can one clearly speak of "ethical whitewashing"?



5 minutes break

Course «Human-Centered Data Science» | Summer Term 2022

Case Study 1: Ethics Bluewashing

By Obert, Otto, und Carsten Trinitis. „Gewissensbits – wie würden Sie urteilen?“ *Informatik Spektrum* 43, Nr. 3 (1. Juni 2020): 227–29. https://doi.org/10.1007/s00287-020-01282-3.

Emma and Noah both studied medicine and landed a lucrative position at the application-oriented research institution EMPE3, specializing in the digital technology-based management of pandemic crisis scenarios. By its very nature, EMPE3 employs people with both medical and information technology training, such as Olivia and Ersoy, two computer scientists. In addition, most of them also have a degree in philosophy or ethics since ethical principles and specifics must always be taken into account when dealing with such sensitive topics as those handled by EMPE3.

For companies and institutions such as EMPE3, lucrative contracts and funding await them in the course of applied research in the context of pandemic crises; among other things, large sums of research funding were made available at short notice by the European Union due to the Corona pandemic.

A team of specialists has been formed from Emma, Noah, Ersoy and Olivia. This team is currently in the [illegible] an EU project called PANDEMRISK, in which numerous other partners from Germany and abroad are involved. The aim of the project is to develop systems for containment and decision support in the event of pandemics.

The main focal points here are the joint and unified development of a concrete AI solution on mobile devices for self-diagnosis of symptoms and automated notification of persons who have been in relevant contact with infected persons (tracing app) as well as the modeling of an AI-based assistance system for transparent and at all times fully comprehensible decision support in triage cases (triage assistance system for evidence-based initial assessment of treatment urgency and thus prioritization of e.g. competing intensive care and respiratory patients, especially in the event of resource bottlenecks).

In this context, the team's core competence is to agree on a standard framework of requirements, including compliance, agreed with all partners involved in the development, which will have an impact beyond the legal requirements (the most stringent national law in each case) and will in particular take into account value-based ethical principles.

In addition to EU countries, represented by EMPE3, other associated partner countries are the EU neighbors Asika, Eurasia and the United Islands.

Asika would like to assert its interests in particular in the tracing app, in order to monitor robust compliance with quarantine of infected persons and actively track down persons who have had relevant contact with infected persons, over and above the necessary scope of functions that also complies with data protection requirements.

With near-area surveillance based on cameras and facial recognition, as well as social scoring, Eurasia's ability to track chains of infection, as well as any necessary direction, does not present any particular additional requirements.

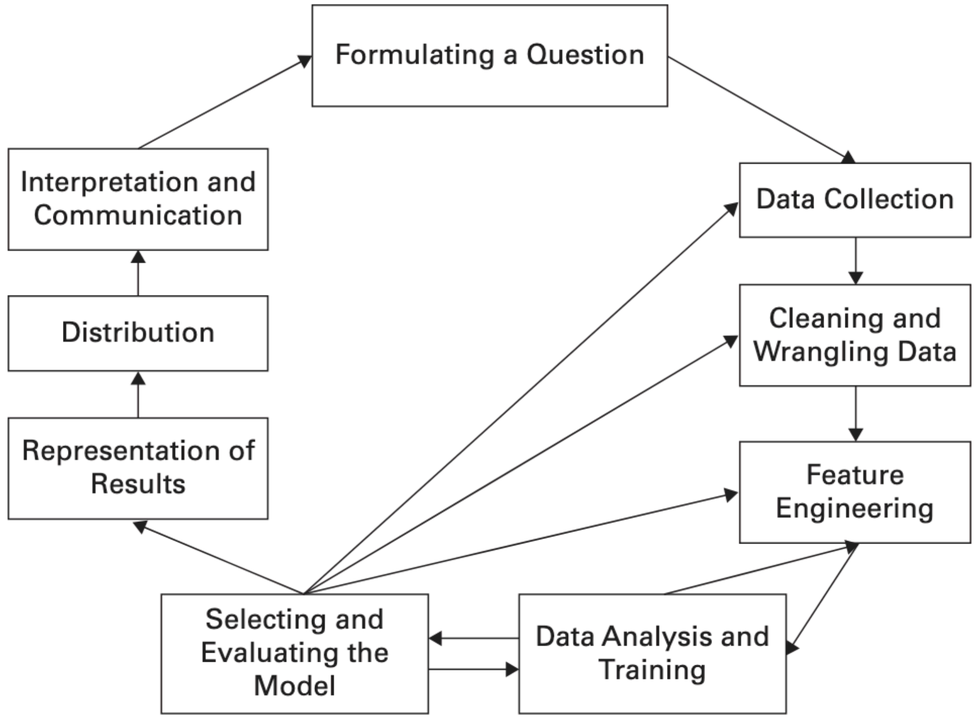
1

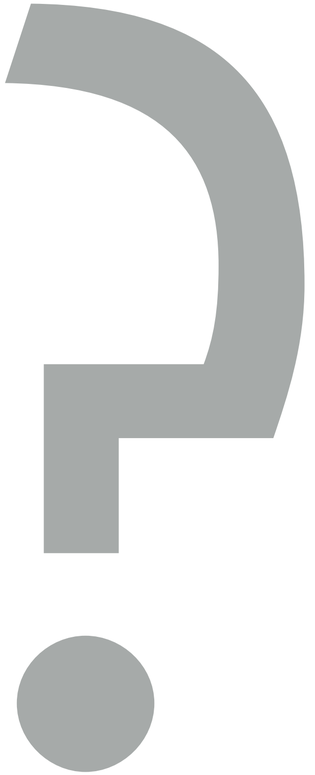
Most Problems arise from Conflicting Principles!

Incorporating Principles in Your Data Science Practice

Principles

- Professional Responsibility
- Promotion of Human Values
- Fairness and Non-discrimination
- Human Control of Technology
- Safety and Security
- Transparency and Explainability
- Accountability
- Privacy





Incorporating Principles in Your Data Science Practice (2)

Principles

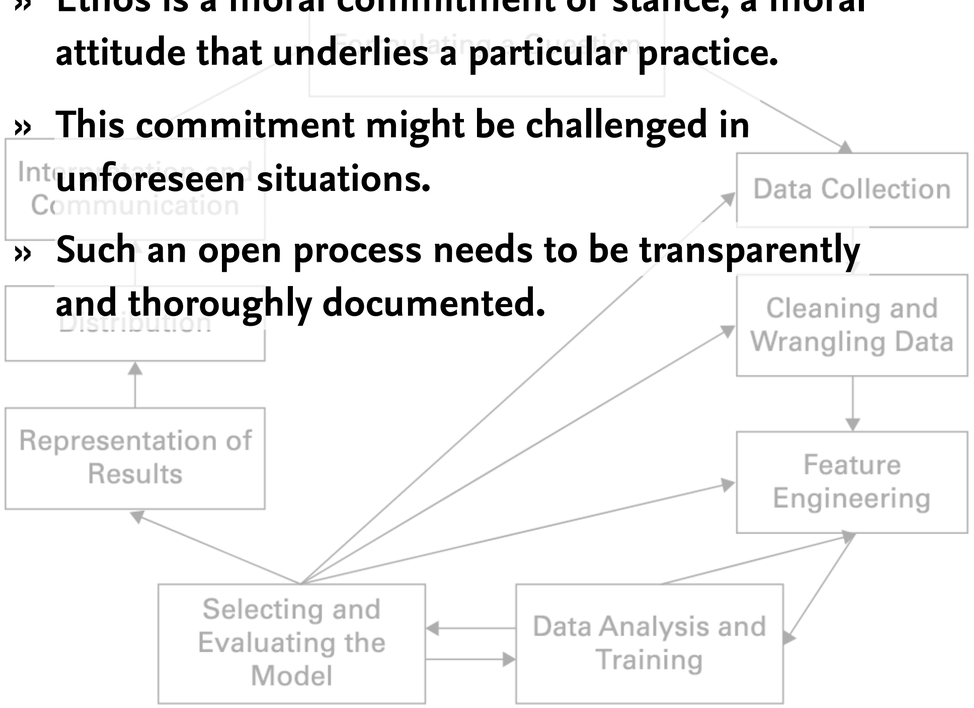
- Professional Responsibility
- Promotion of Human Values
- Fairness and Non-discrimination
- Human Control of Technology
- Safety and Security
- Transparency and Explainability
- Accountability
- Privacy

Development of a shared ethos

» Ethos is a moral commitment or stance, a moral attitude that underlies a particular practice.

» This commitment might be challenged in unforeseen situations.

» Such an open process needs to be transparently and thoroughly documented.



Frauenberger, Christopher, Marjo Rauhala, und Geraldine Fitzpatrick. „In-Action Ethics“. _Interacting with Computers_ 29, Nr. 2 (8. März 2017): 220–36.

Incorporating Principles in Your Data Science Practice (3)

Principles

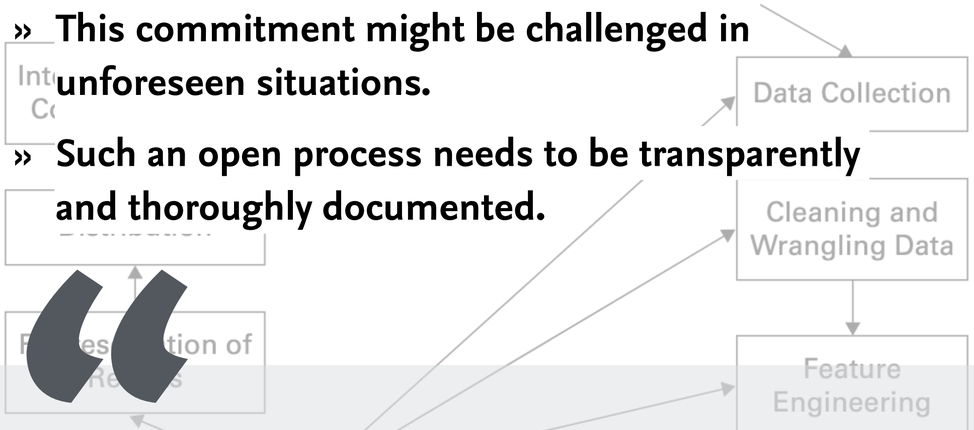
- Professional Responsibility
- Promotion of Human Values
- Fairness and Non-discrimination
- Human Control of Technology
- Safety and Security
- Transparency and Explainability
- Accountability
- Privacy

Development of a shared ethos

» Ethos is a moral commitment or stance, a moral attitude that underlies a particular practice.

» This commitment might be challenged in unforeseen situations.

» Such an open process needs to be transparently and thoroughly documented.



"Values [can be] defined as goals that are desirable, worthwhile, or good, and are perceived to transcend particular situations to be broadly applicable to social life"

Shilton, Katie. „Values and Ethics in Human-Computer Interaction". _Foundations and Trends® Human–Computer Interaction_ 12, Nr. 2 (16. Juli 2018): 107–71.
Aragon, Cecilia, Shion Guha, Marina Kogan, Michael Muller, und Gina Neff. Human-Centered Data Science: An Introduction. Cambridge, MA, USA: MIT Press, 2022.

Incorporating Principles in Your Data Science Practice (4)

Principles

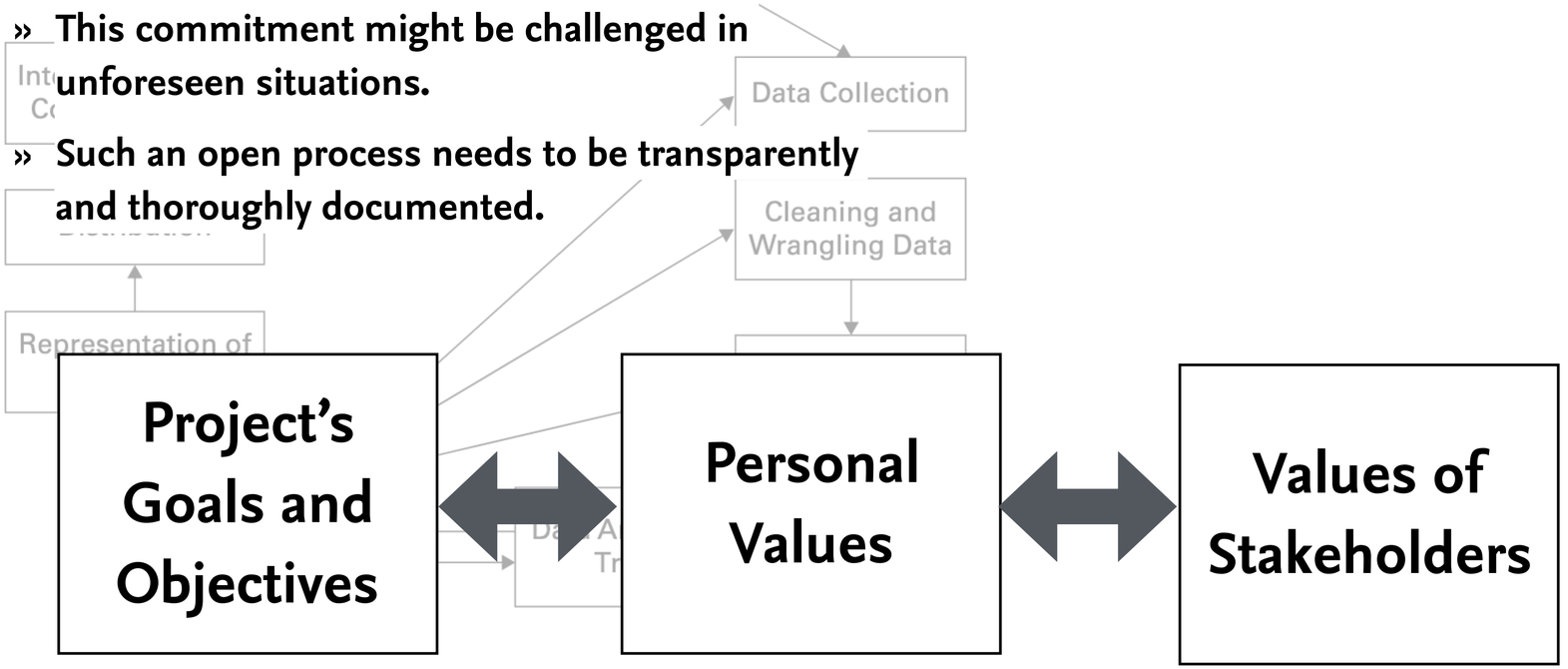
- Professional Responsibility
- Promotion of Human Values
- Fairness and Non-discrimination
- Human Control of Technology
- Safety and Security
- Transparency and Explainability
- Accountability
- Privacy

Development of a shared ethos

» Ethos is a moral commitment or stance, a moral attitude that underlies a particular practice.

» This commitment might be challenged in unforeseen situations.

» Such an open process needs to be transparently and thoroughly documented.



Shilton, Katie. „Values and Ethics in Human-Computer Interaction“. _Foundations and Trends® Human–Computer Interaction_ 12, Nr. 2 (16. Juli 2018): 107–71.
Aragon, Cecilia, Shion Guha, Marina Kogan, Michael Muller, und Gina Neff. Human-Centered Data Science: An Introduction. Cambridge, MA, USA: MIT Press, 2022.

Incorporating Principles in Your Data Science Practice (5)

Principles

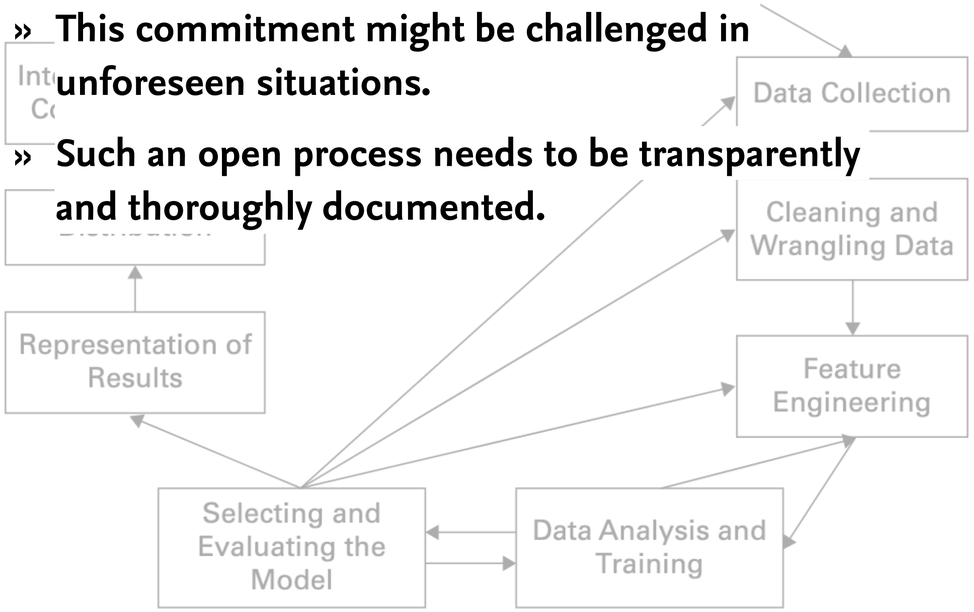
- Professional Responsibility
- Promotion of Human Values
- Fairness and Non-discrimination
- Human Control of Technology
- Safety and Security
- Transparency and Explainability
- Accountability
- Privacy

Development of a shared ethos

» Ethos is a moral commitment or stance, a moral attitude that underlies a particular practice.

» This commitment might be challenged in unforeseen situations.

» Such an open process needs to be transparently and thoroughly documented.



Values Levers

- working across disciplinary barriers
- experiencing self testing of tools
- designing around technical and policy constraints

Shilton, Katie. „Values and Ethics in Human-Computer Interaction“. _Foundations and Trends® Human–Computer Interaction_ 12, Nr. 2 (16. Juli 2018): 107–71.

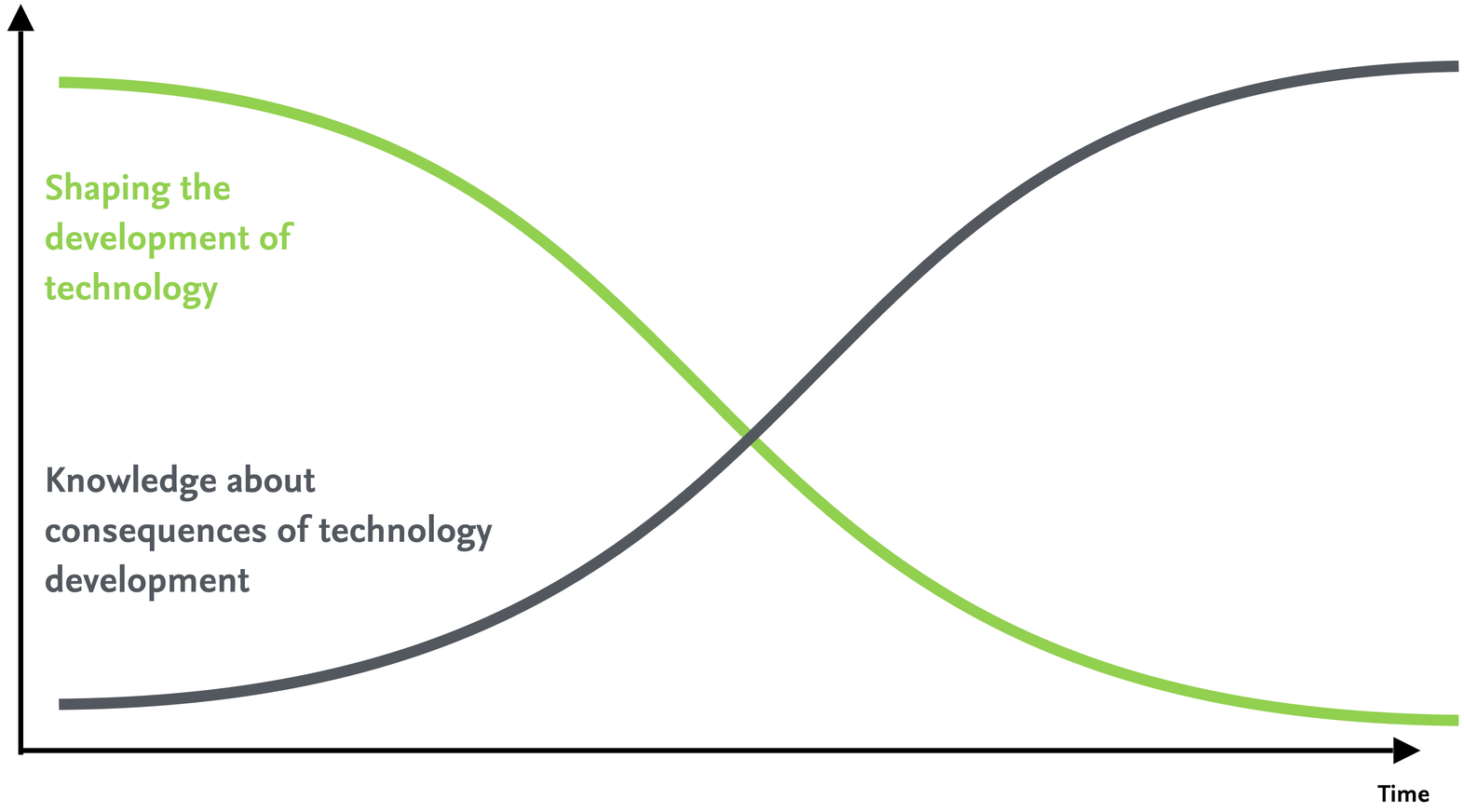
Incorporating Principles in Your Data Science Practice (6)

We need flexible, critical-reflexive approaches in data science to take into account:

1. **Procedural Aspects**, thus, no rigid alignment to abstract principles, but a continuous adjustment in value-sensitive processes.
2. **Deliberative Aspects**, especially in solving conflicting principles by considering multi-disciplinary perspectives from different stakeholders.
3. **Context sensitive aspects**, by considering qualitative and quantitative data of complex social contexts in technology development and deployment.
4. **Creative aspects**, imagine future developments - which are desirable and which are not, how can we promote the former and prevent the latter?

Slide adapted from Susanne Michl

A Challenge Remains



Slide adapted from Susanne Michl

Check your Insights

- What is the difference between ethics and morality?
- What are major ethical theories and how do they differ in evaluating certain situations?
- Where there is a need for applied ethics and what is the difference between applied ethics and ethical theories?
- What are the principles of Responsible AI?
- What is needed to incorporate these principles in your data science practice? What are the risks?

Human-Centered Computing
Research Group

Freie Universität Berlin

«Human-Centered Data Science»

Next week: Enabling the Reproducability of Your Data Science Practice

Prof. Dr. Claudia Müller-Birn

Human-Centered Computing, Institute of Computer Science

Freie Universität Berlin

References

» Aragon, Cecilia, Shion Guha, Marina Kogan, Michael Muller, und Gina Neff. Human-Centered Data Science: An Introduction. Cambridge, MA, USA: MIT Press, 2022.

» Fjeld, Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, und Madhulika Srikumar. „Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI“. SSRN, 2020. https://doi.org/10.2139/ssrn.3518482.

» Floridi, Luciano. „Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical“. _Philosophy & Technology_ 32, Nr. 2 (1. Juni 2019): 185–93. https://doi.org/10.1007/s13347-019-00354-x.

» Frauenberger, Christopher, Marjo Rauhala, und Geraldine Fitzpatrick. „In-Action Ethics“. _Interacting with Computers_ 29, Nr. 2 (8. März 2017): 220–36.

» Guidelines taken from https://gi.de/fileadmin/GI/Allgemein/PDF/GI_Poster_Ethical_Guidlines.pdf

» Jobin, Anna, Marcello Ienca, und Effy Vayena. „The global landscape of AI ethics guidelines“. Nature Machine Intelligence 1, Nr. 9 (1. September 2019): 389–99. https://doi.org/10.1038/s42256-019-0088-2.

» Misselhorn, Catrin. Grundfragen der Maschinenethik. 4., Durchgesehene und Überarbeitete Auflage. Reclams Universal-Bibliothek, Nr. 19583. Ditzingen: Reclam, 2019.

» Shilton, Katie. „Values and Ethics in Human-Computer Interaction“. _Foundations and Trends® Human–Computer Interaction_ 12, Nr. 2 (16. Juli 2018): 107–71.