Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Internationales Zentrum für Ethik in den Wissenschaften (IZEW)

Eberhard Karls Universität Tübingen

Responsible Data Science

Session 1: 10.04.2024, 13.30 – 15.00 h
MA Seminar, summer semester 2024,
Hasso-Plattner Institute, Potsdam

S. Hirsbrunner

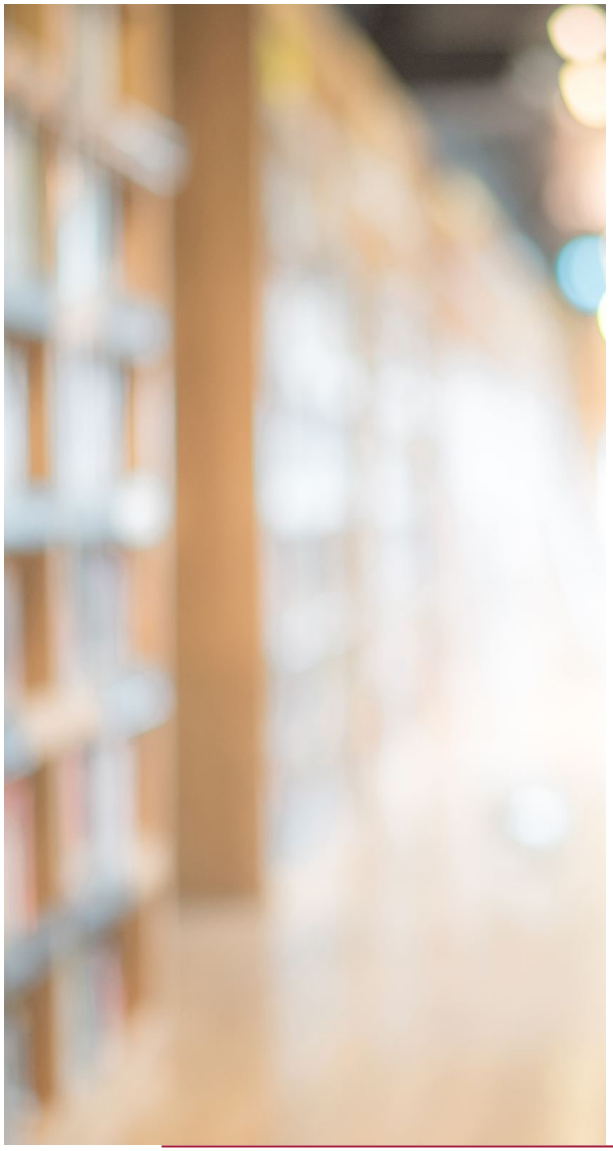
Today

1. Introduction of the lecturer and course concept

2. Plan and themes of the course sessions

3. Seminar outputs and examination

4. Q & A



International Center for Ethics in the Sciences (IZEW) at the University of Tübingen

Inter- and transdisciplinary research center at the University Tübingen

- Philosophers, social and humanities scholars

Ethical questions linked to the sciences and humanities.

Research areas

- Ethics and education
- Nature and sustainable development
- Security ethics
- Media ethics
- **AI and data ethics**

Examplary research projects in AI ethics

VIKING

Trustworthy AI for police investigations (industry, computer science, law, police)

KITQAR

KITQAR – AI training data quality from the perspective of computer science, standardization, law and ethics (HPI, VDI, University of Köln, IZEW)

Dedicated tools for addressing ethical aspects in integrated technology development

Dr. Simon David Hirsbrunner

Research Team Leader at IZEW, University of Tübingen

- Academic background in International Relations, Media Studies and Science & Technology Studies (STS)
- Policy consultant in the area of climate diplomacy, environmental politics and international development
- Scientist at the Human-Centered Computing research group of Freie Universität Berlin, Potsdam Institute for Climate Impact Research and Wikimedia Foundation.
- Researcher in the area of HCI, critical algorithm and data studies, social media analysis, interactive ML, now applied AI, data and security ethics
- Advisor on AI ethics for various institutions

Why ‚responsible data science'?

"Responsibility refers to the **role of people** as they develop, manufacture, sell, and use [AI] systems, but also to the **capability of [AI] systems to answer for their decisions** and identify errors or unexpected results." (Dignum et al. 2018, p. 26)

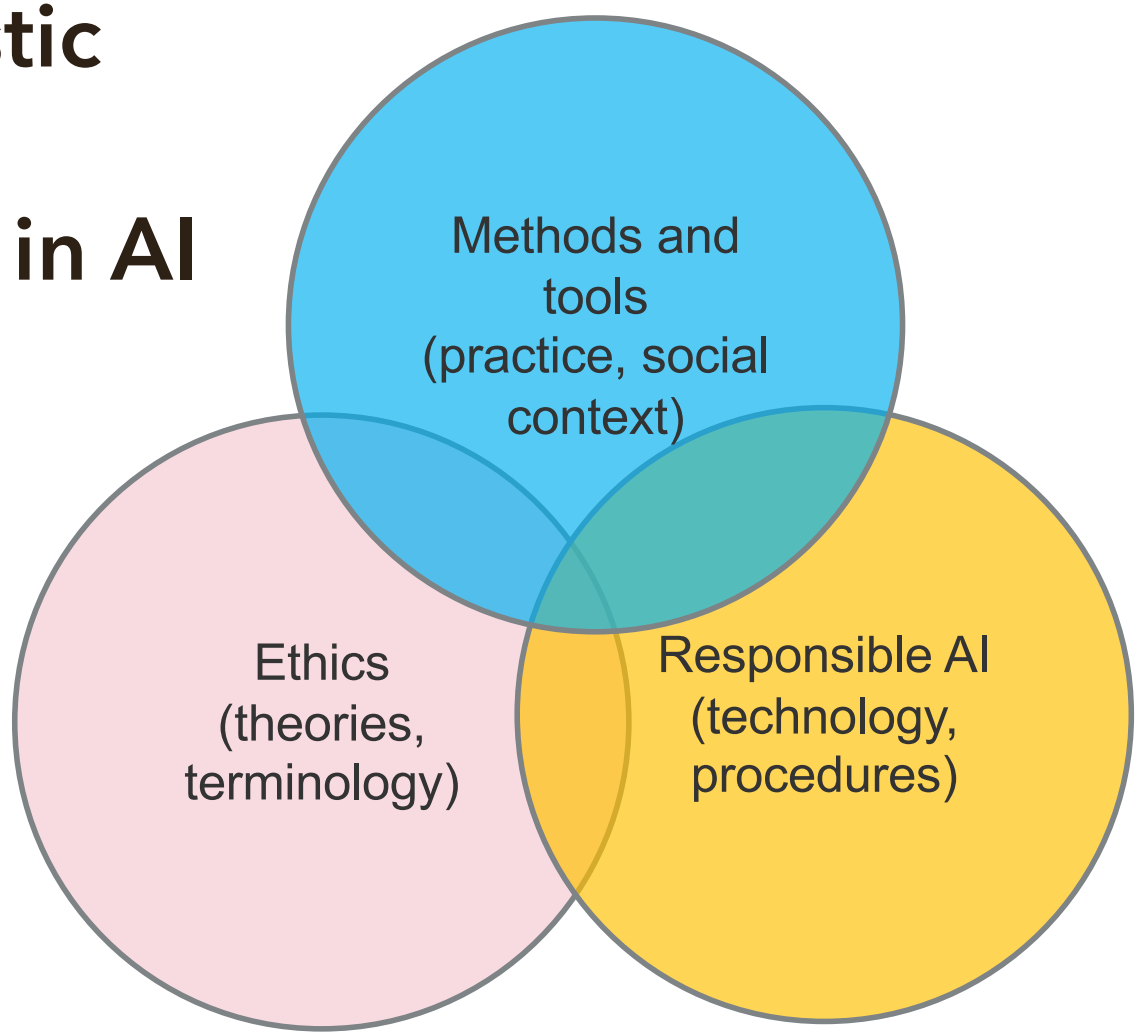
Why ‚responsible data science'?

- Problem of ethicists: challenge to get **from distant ethical critique to responsible engineering** (AI Ethics Impact Group 2019)

- Problem of AI engineers: ‚Responsible AI' as family of technologies may be insufficient as ‚responsibility' of **technology is always sensible to social context** – i.e. there cannot be ‚responsible technology'

> "technology is neither good nor bad; nor is it neutral" (Kranzberg, 1986, 545)

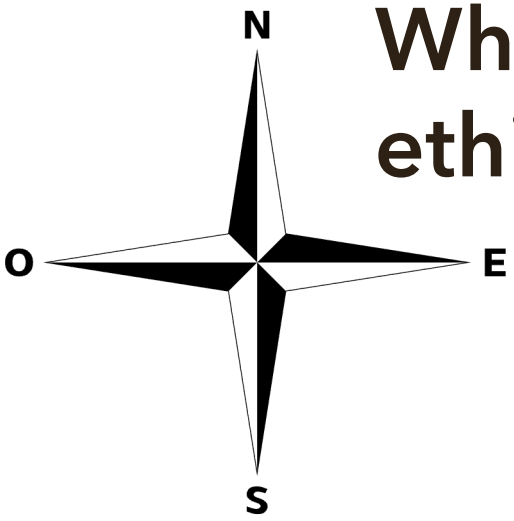
Taking a holistic view on responsibility in AI



Take aways from the seminar

Get familiar with …

- main **ethical concerns** regarding AI development and use;
- mainstream approaches for **coping with these challenges;**
- **theoretical concepts** and interdisciplinary approaches for the understanding and consideration of ethical concerns in software development;
- **methods** to imagine the socio-technical fabric of your technology early in a design process.



Why care about ethics in AI?

- Create technology that supports the well-being and prospering of humans;
- Understand the socio-technical embedding of your technology to create better systems;
- Anticipate risks, problems, transformations to make your technology future proof;
- Facilitate compliance with current and future legislation.

What you will not learn in this seminar

1. Technically designing solutions for
- AI fairness (e.g. evaluating bias, de-biasing);
- explanations for ML models (e.g. LIME, Shap, …);
- data protection (e.g. differential privacy, anonymization, system security).
  (what might be referred to as 'responsible AI')

2. How to be compliant with enacted regulation (- though, maybe a little bit).

Plan of the seminar

10.04.2024, 13.30 – 15.00 h
Session 1: Introductory Session

24.04.2024, 13.30 – 16.45 h
Session 2: Applied Ethics and Value-Sensitive Design

15.05.2024, 13.30 – 16.45 h
Session 3: Discrimination, Fairness and Diversity

22.05.2024, 13.30 – 16.45 h
Session 4: Privacy and Informational Self-Determination

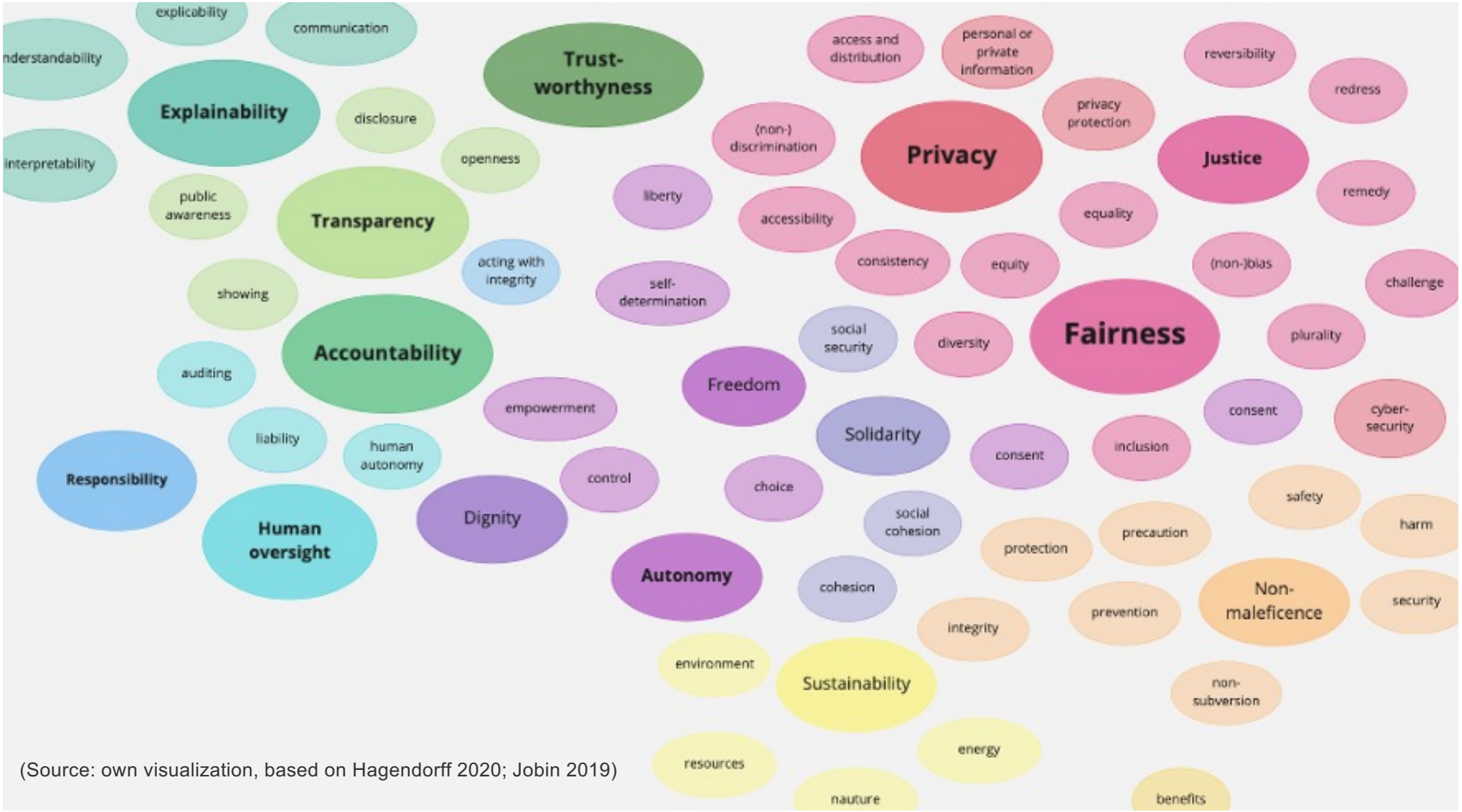
29.05.2024, 13.30 – 16.45 h
Session 5: Deep Fakes and Disinformation

05.06.2024, 13.30 – 16.45 h
Session 6: Human Oversight and Contestability

12.06.2024, 13.30 – 16.45 h
Session 7: Transparency, Documentation and Accountability

26.06.2024, 13.30 – 16.45 h
Session 8: Paper Review Session and Wrap Up

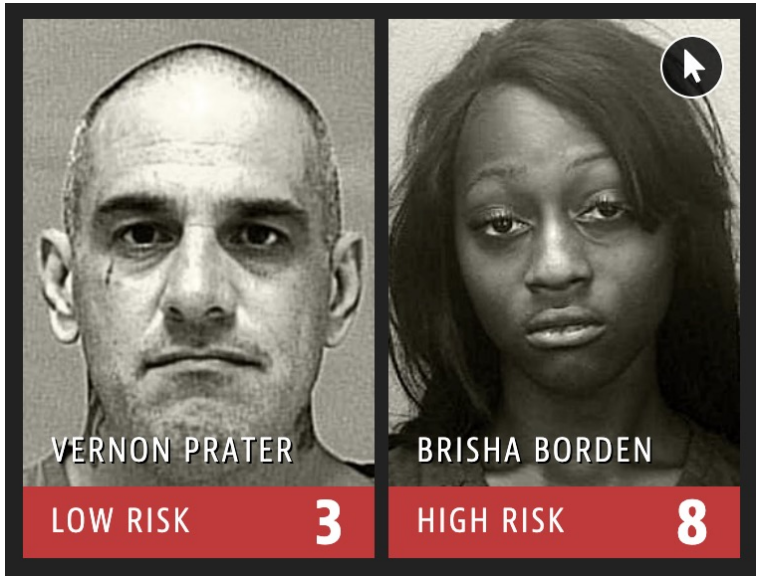
Ethical issues and principles



(Source: own visualization, based on Hagendorff 2020; Jobin 2019)

Example topic (1) discrimination, fairness and diversity

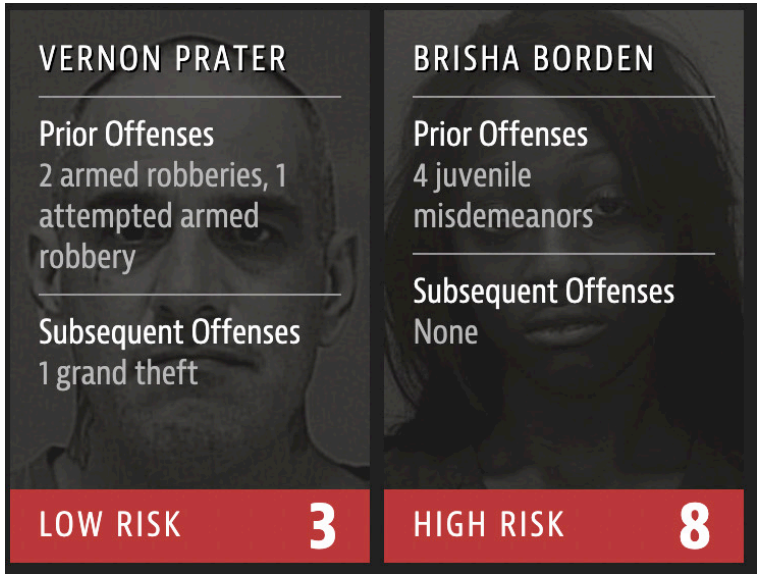
- Based on biographic data, COMPAS software calculates the probability (risk) that an arrested person will commit crimes in the future.
- This score can be used, for instance, to inform the determination of bail bonds.



Source: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Example topic (1) discrimination, fairness and diversity

- Created many ‚false positives' (people considered risky without being risky) and ‚false negatives' (people considered low risk who later committed crimes)
- Biased against People of Color (PoC).
- Was used for purposes the software had not been designed for.



Source: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

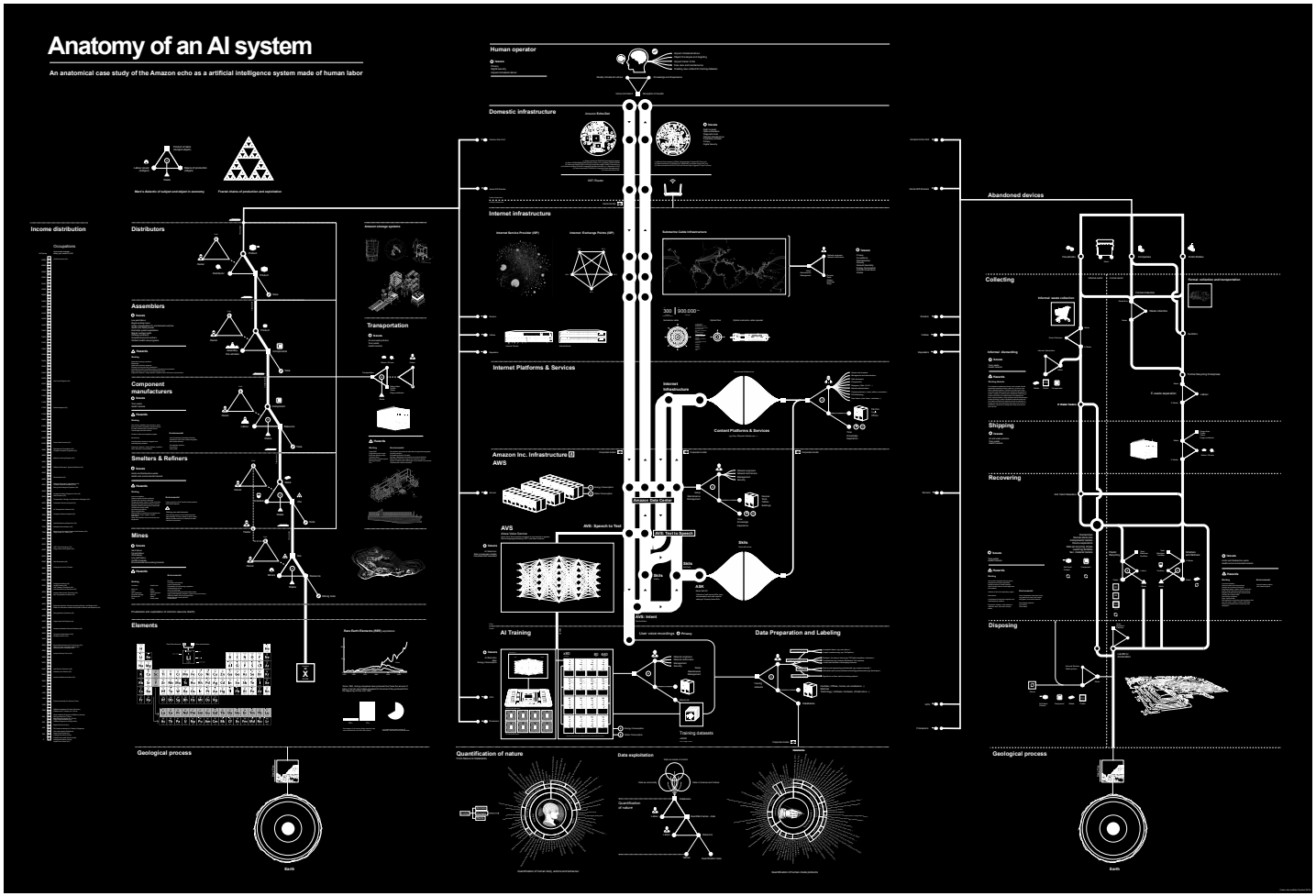
Example topic (1) discrimination, fairness and diversity

The question of (a lack of / exagerated) diversity in large language models

Sure, here is an illustration of a 1943 German soldier:

Image source: https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical

Example topic (2): people and planet



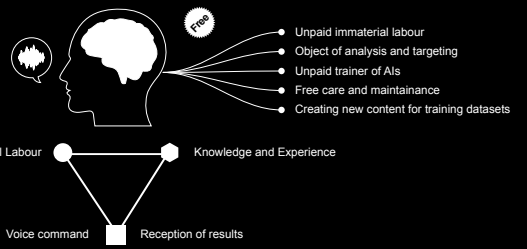
(Source: https://anatomyof.ai, Crawford und Joler 2018)

uman labor

Human operator

Issues

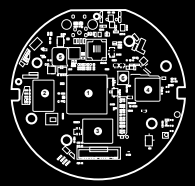
Privacy
Digital security
Unpaid immaterial labour



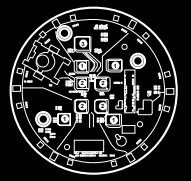
Domestic infrastructure

Amazon **Echo Dot**

Amazon Echo Dot



(1) Texas Instruments DM3725 Digital Media Processor
(2) Micron MT46H64M32LFBQ 256 MB (16 Meg x 32 x 4 Banks) LPDDR SDRAM
(3) Samsung KLM4G1FEPD 4GB High Performance eMMC NAND Flash Memory
(4) Qualcomm Atheros QCA6234 Integrated Dual-Band 2x2 802.11n + Bluetooth 4.0 SIP
(5) Texas Instruments TPS65910A1 Integrated Power Management IC
(6) Texas Instruments DAC



(1) National Semiconductor LP55231 Programmable 9-Output LED Driver (x4)
(2) Texas Instruments TLV320ADC3101 92dB SNR Low-Power Stereo ADC (x4)
(3) Texas Instruments SN74LVC74A Dual Positive-Edge-Triggered D-Type Flip-Flops

Issues

Right to repair
Open schematics
Diagnostic tools
Planned obsolescence
Proprietary software
Privacy
Digital Security

WiFi Router

Home Wifi Routers

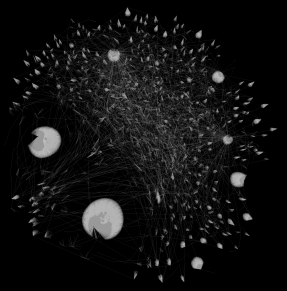
Visible Infrastructure

Invisible Infrastructure

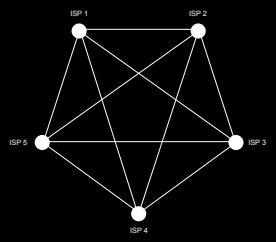
Home border

Internet infrastructure

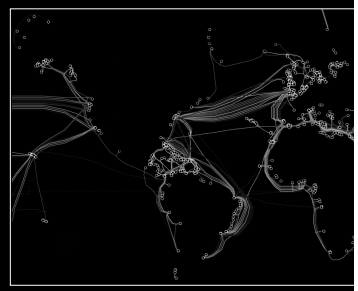
Internet Service Provider (ISP)



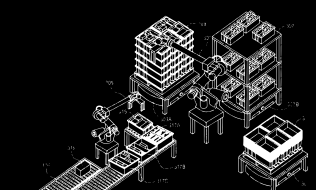
Internet Exchange Points (IXP)



Submarine Cable Infrastructure

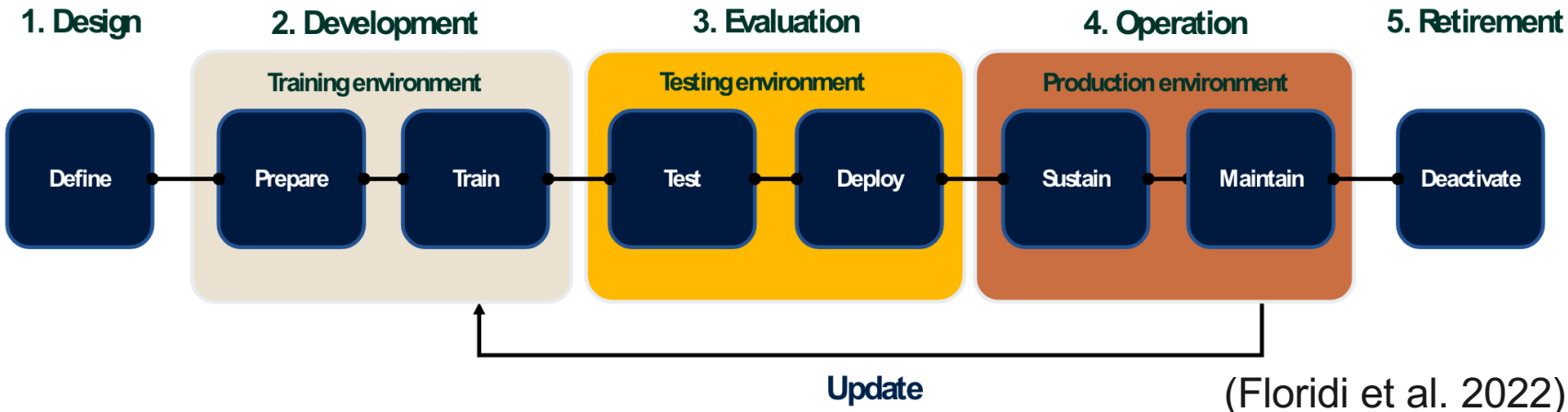


Amazon storage systems



Example topic (3): transparency and accountability

Ethical conformity assessments for the different stages of the AI life cycle



(Floridi et al. 2022)

Transparency tools

- Data sheets
- Model cards
- System cards

Example plan of a session

13h30	Flashlight and recap
13h35	Input from lecturer 1 (theory, concepts)
14h00	Student presentation
14h20	Discussion
14h50	–––––––– Break –––––––––
15h00	Input from lecturer 2 (practice-approach)
15h30	Exercise in small groups
16h15	Reporting from groups and discussion
16h45	End

Image source: Generated with BlueWillow v4

Value-Sensitive Design (VSD)

"Value sensitive design seeks to guide the shape of being with technology. It positions researchers, designers, engineers, policy makers, and anyone working at the intersection of technology and society to make insightful investigations into technological innovation in ways that foreground the well-being of human beings and the natural world."
(Friedman and Hendry 2019, 3)

Conceptualized by Prof. Dr. Batya Friedman (HCI Professor at the Information school of the University of Washington

VSD definition

„[VSD] provides theory, method, and practice to account for human values in a principled and systematic manner throughout the technical design process."
(Friedman and Hendry 2019, 3f)

This seminar:
Value-Sensitive Design (VSD) for data science and AI software engineering

Tools (1)

Envisioning cards

(Friedman et al. 2011)

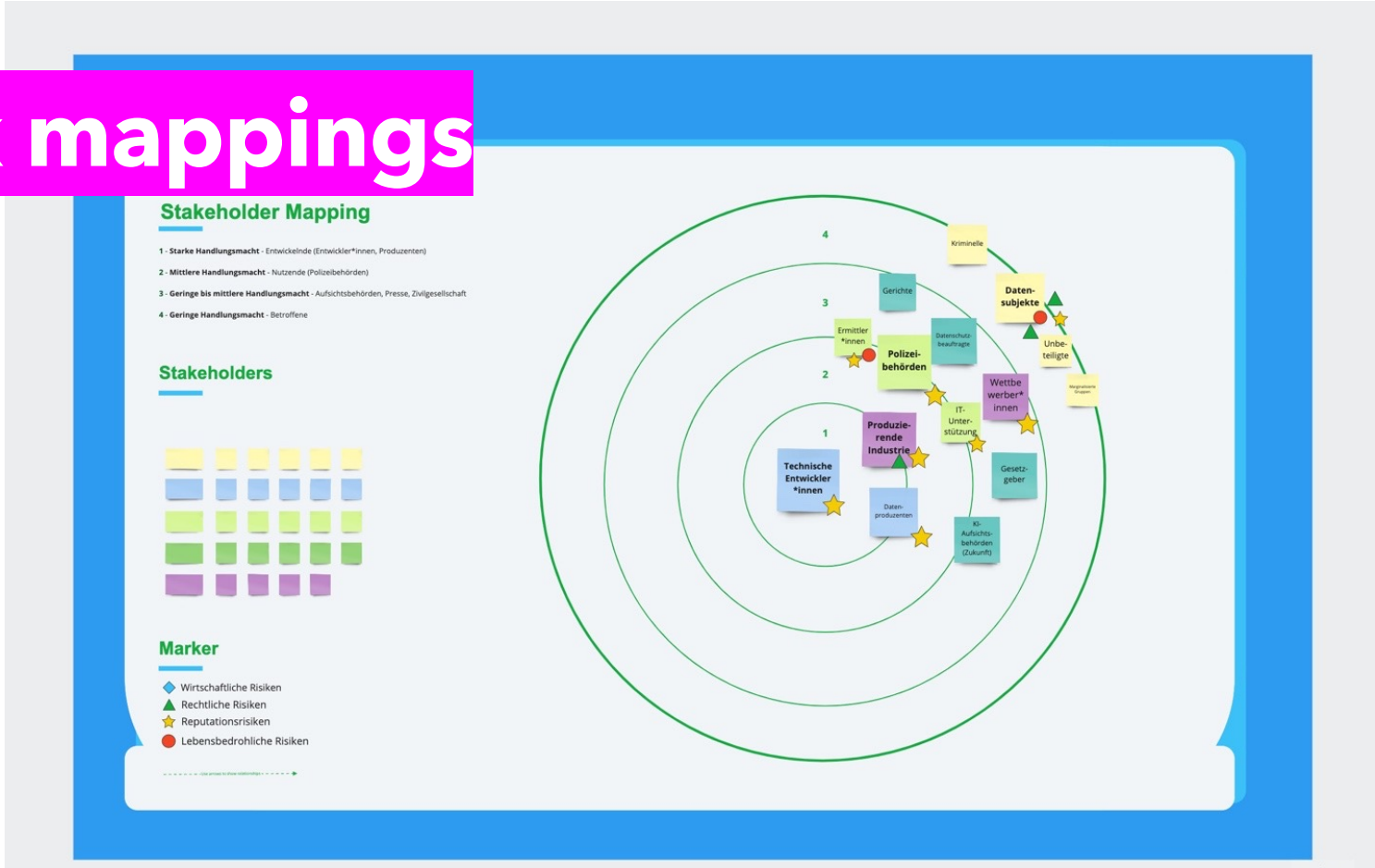
Tools (2)

Stakeholder tokens

(Yoo 2017)

Tools (3)

Risk mappings



Examination

The grade is based on

- a **group presentation of 20-25 min incl. handout** (30% of the grade). The group presentations (2-3 students) will each address one specific literature piece and related ethical issues.

- a **written analysis of an ethical dilemma and possible solutions of 6 pages per person** (70% of the grade). The paper should apply VSD-based methods and discuss the ethical issues covered by the course. Students are also invited to submit (longer) papers developed in a group (e.g. 18 pages for 3 students with clear markers for the individual contributions).

Indications for presentation

- 20 min. for one person / 30 min. for two persons
- Focusing on one piece of literature (see xls-table)
- Content of the presentation
  - Context of the author and paper
  - Main ideas of the paper
  - Critique of the arguments
  - Links to further readings
- Propose 3 questions for the discussion
  (and send it to lecturer in advance)
- Handout (2 pages / 800 words text per person)

Indications for paper

- 6 pages per person, with the option of developing a paper in a group of two students (12 pages)
- Address one ethical dilemma in an application area of AI or data engineering using the concepts and methods learned in the seminar.

Indications for use of AI in the course

- We will occasionally use GenAI for brainstorming, gather information on theories and approaches and to generate and discuss ethical dilemma;

- While you are encouraged to use Generative AI applications (ChatGPT, Gemini, Midjourney, etc.) for the production of the paper (text and/or visualizations) while adhering to the following rules:

  - Use of AI has to be documented on a transparency sheet, whose form may be inspired by the literature addressed in the course (i.e. Data Sheets, Model and System Cards).
  - The transparency sheet has to specify a) what kind of AI has been used b) in what ways to c) produce which exact element of the study and/or paper.
  - An evaluation of the transparency sheet will also feed into your course grade.

Registration for the course

- If you like to participate in the seminar, **register** via https://tinyurl.com/rds-hpi-2024 (sheet ‚Seminar-participants_registration') until **Sunday 14.4. noon (12h00)**.

- On **Monday 15.04,** I will send you an update of the course schedule with the literature and dates for group presentations. You can then **choose your presentation topic and presentation date** in https://tinyurl.com/rds-hpi-2024 (sheet ‚literature-and-presenters') and enter your name into the document **until Fri 19.4. evening**.

- Please always check the up-to-date course information on **Github:** https://github.com/simonsimson/responsible-data-science (not the HPI seminar website, which will not be updated.)

Next session

24.04.2024, 13.30 – 16.45 h

Session literature

1: Jobin A, Ienca M and Vayena E (2019) The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1(9): 389–399. https://www.nature.com/articles/s42256-019-0088-2

2: Friedman B (1996) Value-sensitive design. Interactions 3(6): 16–23. https://dl.acm.org/doi/10.1145/242485.242493

Your questions or specific wishes for the course?

simon.hirsbrunner@uni-tuebingen.de

Hope to see you on 24.04!

Bibliography

The entire bibilography for the course will be uploaded on Github in another text file.

Image sources
Most sources are cited on the relevant slide. Slide 1: © Adobe Stock / kras99, slide 37: genewolf CC BY-ND 2.0