Hasso Plattner Institut
Digital Engineering · Universität Potsdam

Internationales Zentrum für Ethik in den Wissenschaften (IZEW)

Eberhard Karls Universität Tübingen

Responsible Data Science

Session 1: 19.04.2023, 15.15 – 16.45 h
MA Seminar, SoSe 2023, Hasso-Plattner Institut

S. Hirsbrunner

Today

- Introduction of the lecturer and course concept
- Plan and themes of the course sessions
- Seminar outputs and examination
- Q & A

Dr. Simon David Hirsbrunner

Senior researcher and project 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

International Center for Ethics in the Sciences (IZEW)

- 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
  - **Society, culture and technical transformations**
    - Security ethics
    - Social cohesion
    - **Media ethics and information systems, including AI**

Current research projects in AI ethics

- KITQAR – AI training data quality from the perspective of computer science, standardization, law and ethics (HPI, VDI, University of Köln, IZEW)
- Analysis of heterogenous mass data for police investigations (industry, computer science, law, police)
- Trustworthy AI for police investigations (industry, computer science, law, police)

Take aways from the seminar

Get familiar with …

- main ethical concerns regarding AI development and use;
- mainstream strategies to cope 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.

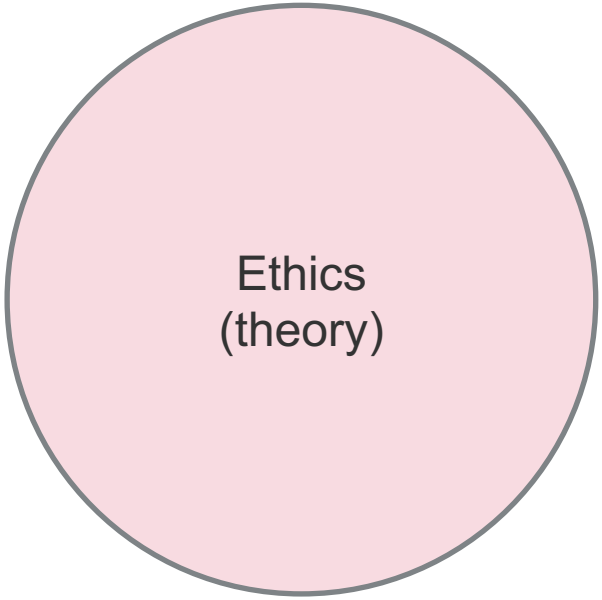
Why ‚responsible data science'?

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)

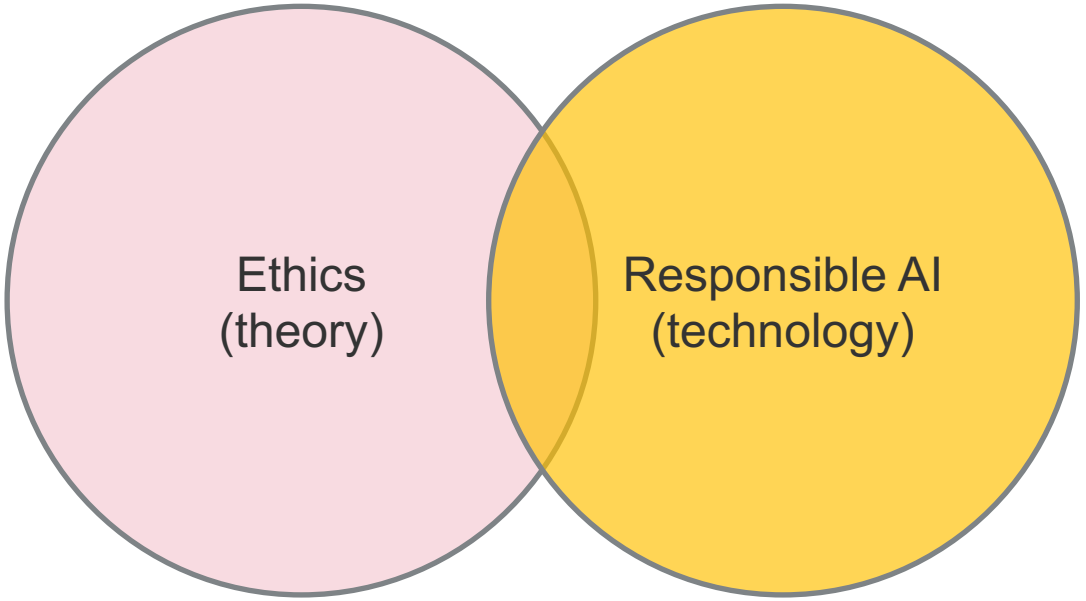
- Challenge to get from distant ethical critique to responsible engineering (AI Ethics Impact Group 2019)

- ‚Responsible AI' as family of technologies may be insufficent 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).

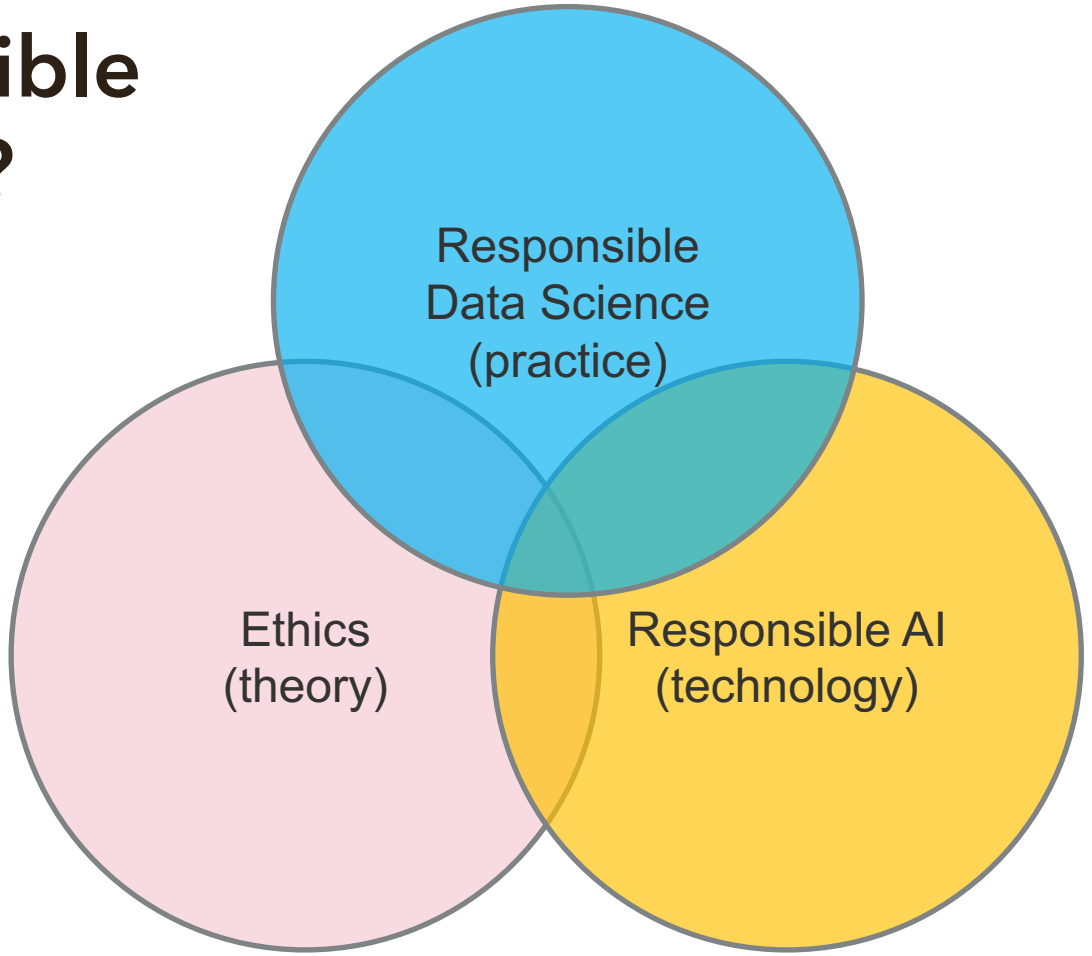
Why ‚responsible data science'?



Why ‚responsible data science'?



Why ‚responsible data science'?



Why care about ethics in AI?

- Create technology that supports the well-being and prospering of humans;
- Imagine the socio-technical embedding of your technology;
- Anticipate future risks, problems, transformations;
- Faciliate compliance with future legislation.

What you will not learn in this seminar

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')

How to be compliant with enacted regulation (- or maybe a little bit).

Plan of the seminar

19.4., 15.15 – 16.45 h
Session 1: Introductory session with assignment of presentation topics

26.04., 15.15 – 18.30 h
Session 2: Applied Ethics and Responsible Data Science as socio-technical challenges

27.04., 15.15 – 18.30 h
Session 3: discrimination, fairness and diversity

17.05., 15.15 – 19.30 h
Session 4: privacy and informational self-determination

25.05., 15.15 – 19.30 h
Session 5: data quality, reliability and safety

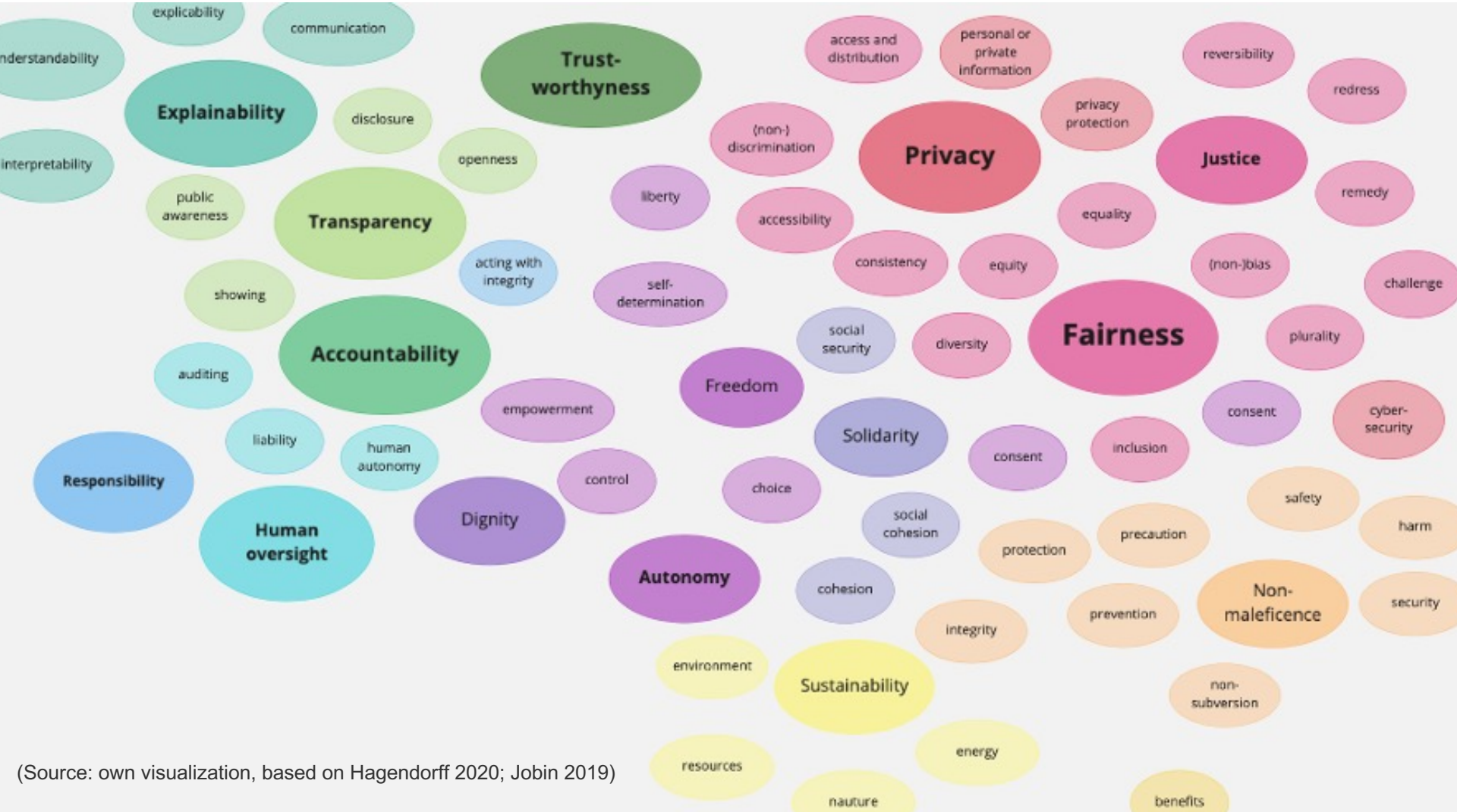
31.05., 15.15 – 19.30 h
Session 6: people and planet

01.06., 15.15 – 19.30 h
Session 7: transparency and accountability

Example plan of a session

15h15 Flashlight and recap
15h30 Input from lecturer 1 (theory, concepts)
16h00 Student presentation
16h20 Discussion
16h50 –––––––– Break ––––––––––
17h00 Input from lecturer 2 (practice-approach)
17h30 Exercise in small groups
18h30 Reporting from groups
19h00 Project work (towards seminar papers)
19h30 End

Session 2: Applied ethics and Responsible Data Science as socio-technical challenges



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

Session 2: Applied ethics and Responsible Data Science as socio-technical challenges

- Concepts
  - Social norms, ethical norms
  - Human values, ethical values
  - Socio-technical systems
  - Responsibility
  - […]
- Methods
  - Value-Sensitive Design
  - Stakeholder analysis
  - Risk analysis
  - Scenario-Based Design / Value Scenarios
  - […]

Session 2: Applied ethics and Responsible Data Science as socio-technical challenges

- Concepts
  - Social norms, ethical norms
  - Human values, ethical values
  - Socio-technical systems
  - Responsibility
  - […]
- Methods
  - Value-Sensitive Design
  - Stakeholder analysis
  - Risk analysis
  - Scenario-Based Design / Value Scenarios
  - […]

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

Envisioning cards

(Friedman et al. 2011)

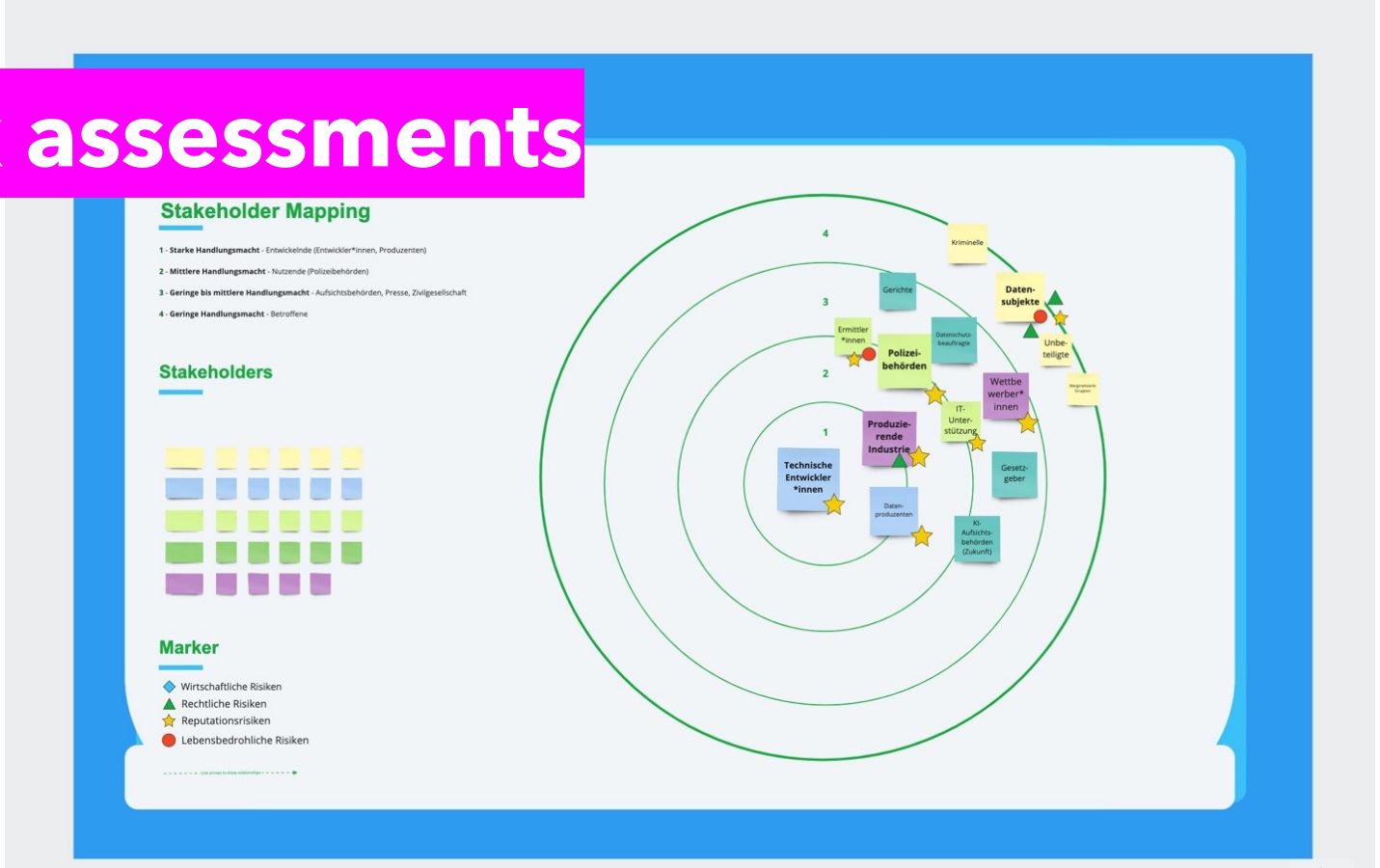
Tools (2)

Stakeholder tokens

(Yoo 2017)

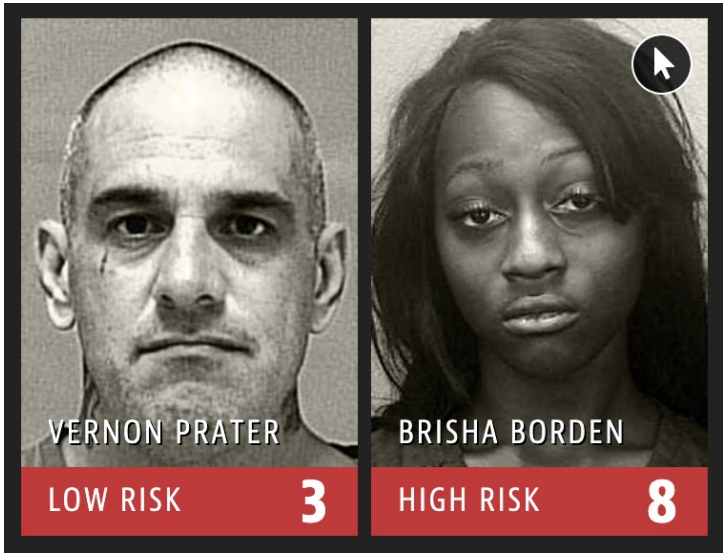
Tools (3)

Risk assessments



Session 3: discrimination, fairness and diversity

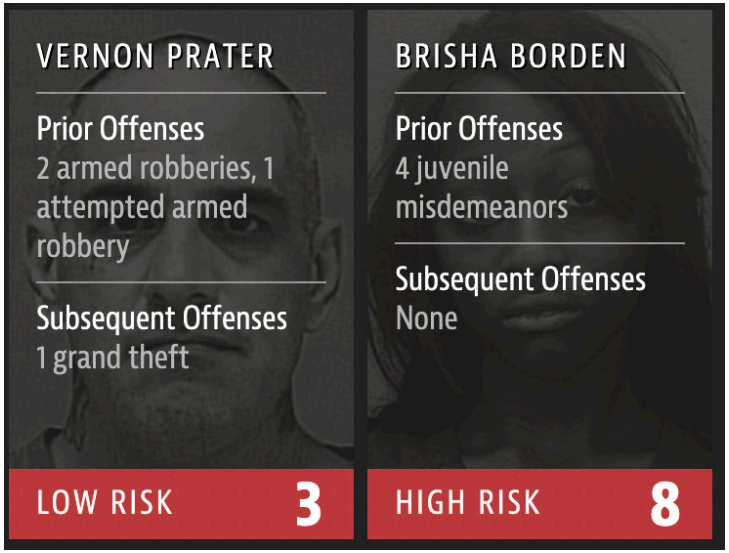
- COMPAS software calculates the probability (risk) that an arrested person will commit crimes in the future.



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

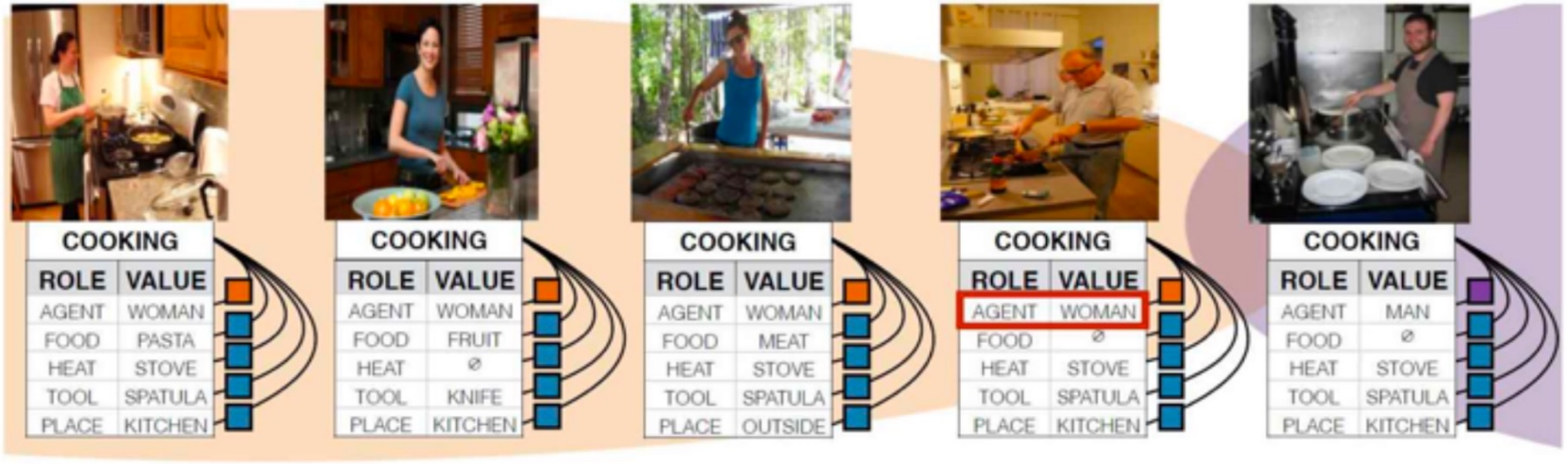
Topic 1: discrimination, fairness and diversity

- COMPAS software calculates the probability (risk) that an arrested person will commit crimes in the future.
- 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).



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

Topic 1: discrimination, fairness and diversity



- Data set: 67% of people cooking are women
- Algorithm predicts: 84% of people cooking are women

(Source: Zhao et al. 2017)

Session 3: discrimination, fairness and diversity

Main concepts

- Discrimination
- Bias
- Discriminatory bias
- Fairness
  - Equalized odds
  - Demographic parity
  - etc.
- Limits to AI fairness, diversity-awareness

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

Topic 2: privacy and informational self-determination

(Source: Alexander Kirch / Shutterstock.com)

Topic 2: privacy and informational self-determination

Main concepts

- Private sphere
- Decisional privacy
- Informational privacy
- Group privacy
- Predictive privacy
- […]

Topic 3: ethics and data quality

Representativity

Balance

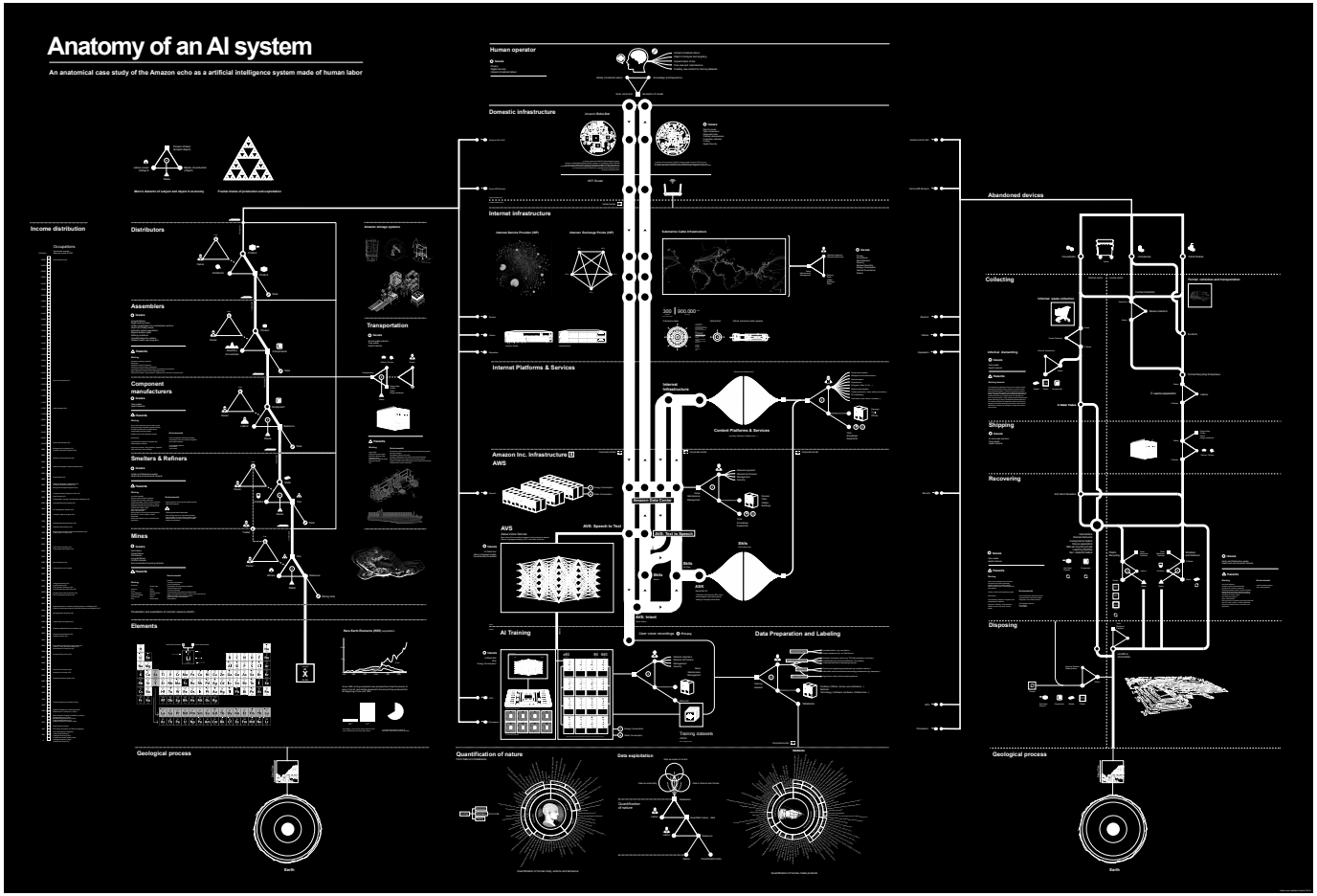
Diversity

Completeness

Coverage

Variety

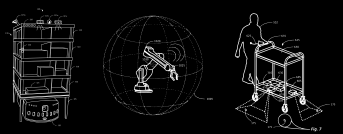
Topic 4: people and planet

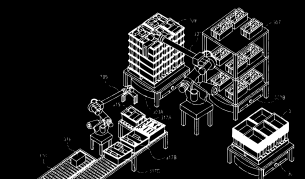


(Source: https://anatomyof.ai, Crawford und Joler 2018)

uman labor

Amazon storage systems





Human operator

Issues

Privacy
Digital security
Unpaid immaterial labour

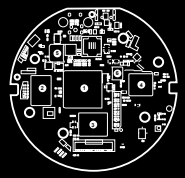
- Unpaid immaterial labour
- Object of analysis and targeting
- Unpaid trainer of AIs
- Free care and maintainance
- Creating new content for training datasets

Mostly immaterial Labour — Knowledge and Experience

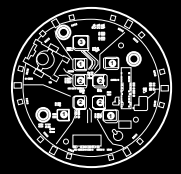
Voice command — Reception of results

Domestic infrastructure

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

Amazon Echo Dot

WiFi Router

Home Wifi Routers

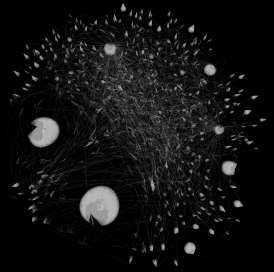
Visible Infrastructure

Invisible Infrastructure

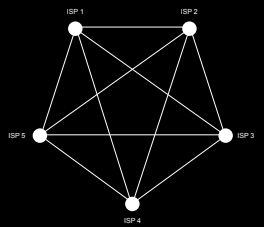
Home border

Internet infrastructure

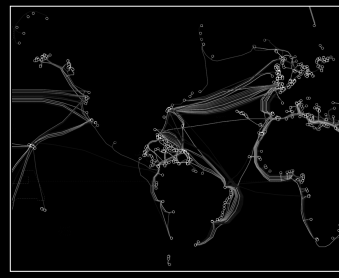
Internet Service Provider (ISP)



Internet Exchange Points (IXP)

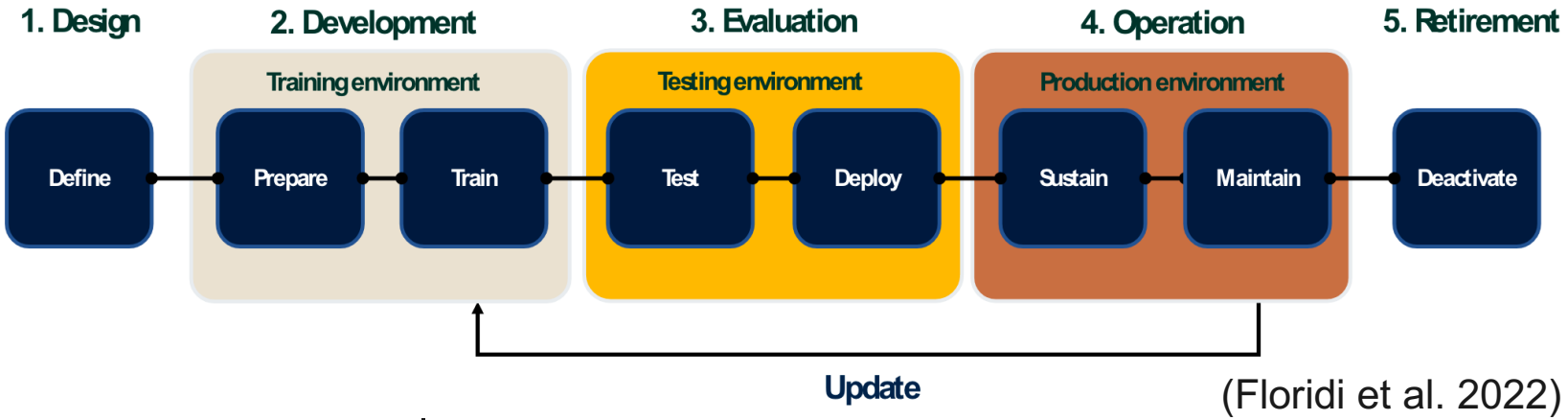


Submarine Cable Infrastructure



Topic 5: transparency and accountability

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

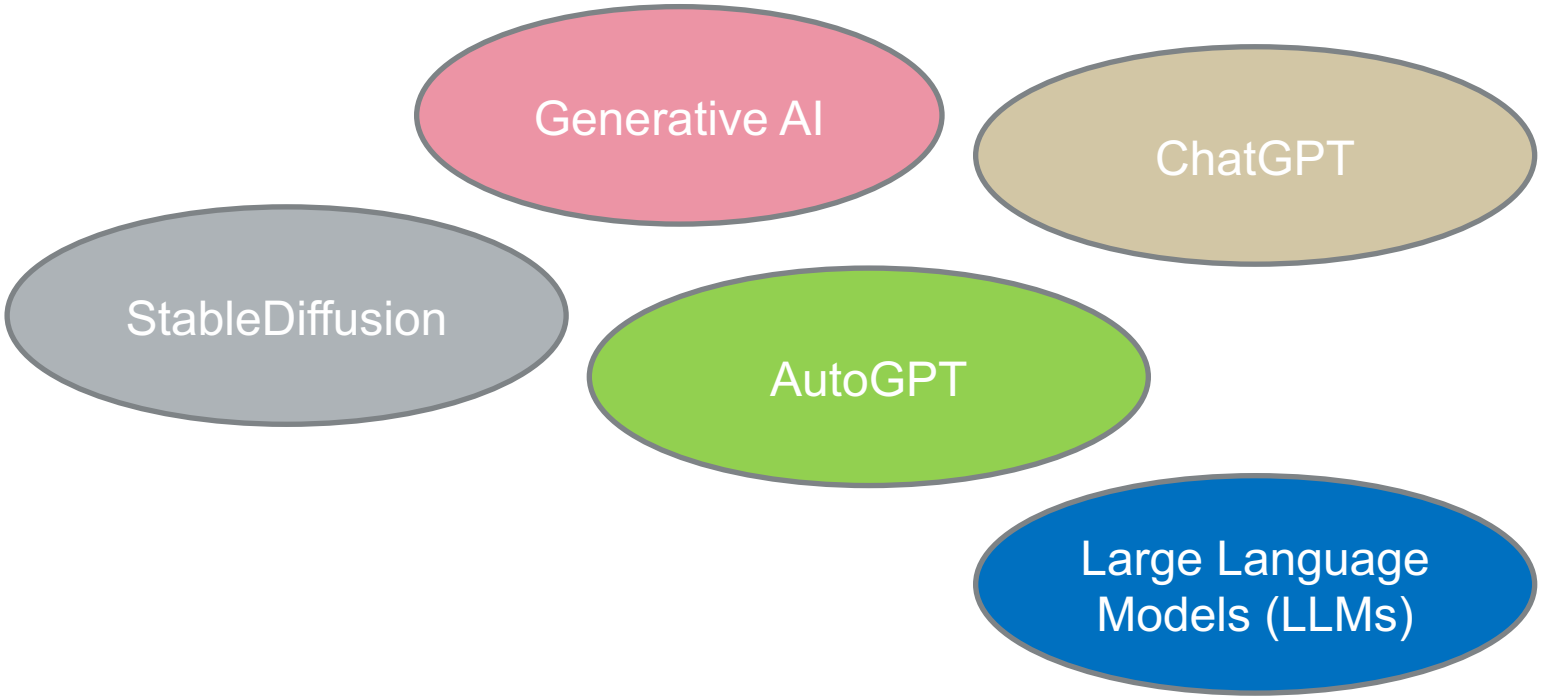


(Floridi et al. 2022)

Transparency tools

- Data sheets
- Model cards
- Impact assessments

The elephant(s) in the room…



→ may be treated as use cases among others

Examination

The grade is based on

- a **group presentation of 20 min incl. handout** (40% of the grade). The group presentations (2-3 students) will each address one specific literature piece and related ethical issues.

- a **written analysis and concept for a responsible AI system à 6 pages per person** (60% of the grade). The paper should apply a VSD-based approach 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 section. markers for the individual contributions).

Examination

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
- Propose 3 questions for the discussion (and send it to lecturer in advance)
- Handout (2 pages text per person)

Next steps

- If you like to participate in the seminar, **register per Google Forms** via https://tinyurl.com/rds-registration stating a) your name, b) study program and c) semester until **Sunday (23.4.) evening**. I will put together an email-list with all the students and send you updates via this channel.

- **Read the literature** for the next two sessions (26.4. or 27.4.).

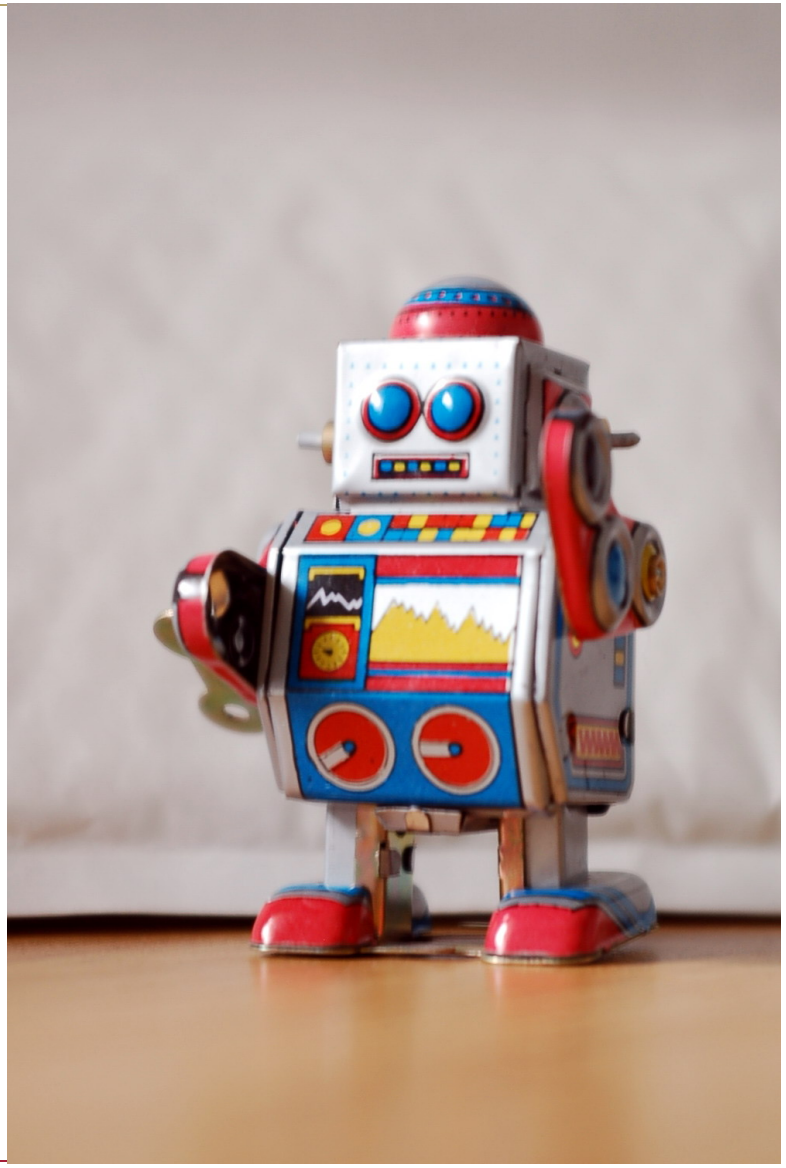
- **Choose one of the literature sources for your presentation** in https://tinyurl.com/rds-hpi and put your name/email into the document (until 26.4. evening)

- If you have the opportunity to **hold a presention** during one of the **next two sessions** (26.4. or 27.4.), please let me know as soon as possible per email: simon.hirsbrunner@uni-tuebingen.de (Friday 21.4.2023 evening).

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

Your questions, specific wishes?

simon.hirsbrunner@uni-tuebingen.de



Source: image by genewolf CC BY-ND 2.0

Bibliography

Angwin J, Larson J, Mattu S, et al. (2016) Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. Propublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Crawford, K., & Joler, V. (2019). Anatomy of an AI System. Virtual Creativity, 9(1), 117–120. https://doi.org/10.1386/vcr_00008_7

Dignum, V., Baldoni, M., Baroglio, C., Caon, M., Chatila, R., Dennis, L., Génova, G., Haim, G., Kließ, M. S., Lopez-Sanchez, M., Micalizio, R., Pavón, J., Slavkovik, M., Smakman, M., van Steenbergen, M., Tedeschi, S., van der Toree, L., Villata, S., & de Wildt, T. (2018). Ethics by Design: Necessity or Curse? Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 60–66. https://doi.org/10.1145/3278721.3278745

Floridi L, Holweg M, Taddeo M, et al. (2022) capAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act. 4064091, SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.4064091.

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

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86–92. https://doi.org/10.1145/3458723

Hagendorff T (2020) The Ethics of AI Ethics: An Evaluation of Guidelines. Minds and Machines 30(1): 99–120. https://link.springer.com/article/10.1007/s11023-020-09517-8

Kranzberg, M. (1986). Technology and History: "Kranzberg's Laws." Technology and Culture, 27(3), 544–560. https://doi.org/10.2307/3105385

Larson J, Mattu S, Kirchner L, et al. (2016) How we analyzed the COMPAS recidivism algorithm. ProPublica (5 2016) 9(1): 3–3. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys, 54(6), 1–35. https://dl.acm.org/doi/10.1145/3457607

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220–229. https://doi.org/10.1145/3287560.3287596

Image sources

Most sources are cited on the relevant slide. Slide 1: © Adobe Stock / kras99