

«Human-Centered Data Science»

Introducing Human-Centered Data Science and the Course Concept

Prof. Dr. Claudia Müller-Birn

Human-Centered Computing, Institute of Computer Science

Freie Universität Berlin

Lecture Overview

Motivating the Need for Human-Centered Data Science (existing challenges in our society caused by data science projects, reflecting about the role of technology in our society, looking into the data science cycle, defining human-centered data science)



Course Overview and Concept (scope, learning objective, underlying concept, assignments and active participation, class schedule)

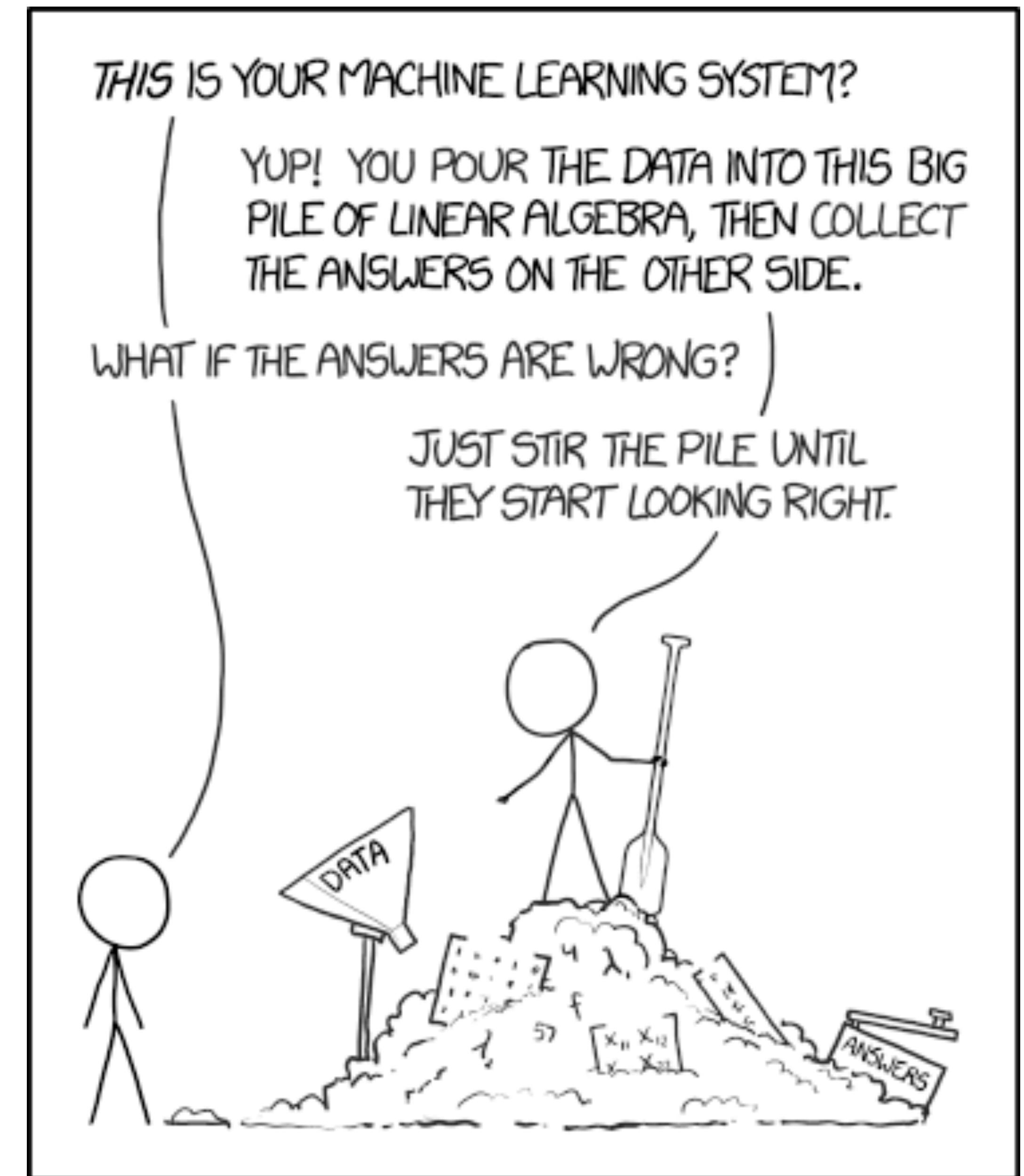
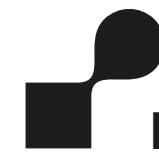


Motivation



This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from linguistics to sociology.[...] Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.

Chris Anderson - WIRED
(<https://www.wired.com/2008/06/pb-theory/>)



<https://xkcd.com/1838/>



The Assumption: Big Data is Objective

This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from linguistics to sociology.[...] Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.

Chris Anderson - WIRED
(<https://www.wired.com/2008/06/pb-theory/>)



WEAPONS OF MATH DESTRUCTION



HOW BIG DATA INCREASES INEQUALITY
AND THREATENS DEMOCRACY

CATHY O'NEIL

'Fascinating and deeply disturbing'
YUVAL NOAH HARARI, GUARDIAN BOOKS OF THE YEAR



Lack of Robustness

“

The Tesla struck the left side of the semi trailer. The roof of the Tesla was sheared off [...] The 50-year-old male Tesla driver died as a result of the crash. [...]



National Transportation Safety Board PRELIMINARY Report Highway HWY 19FH008 <https://www.ntsb.gov/investigations/AccidentReports/Reports/HWY19FH008-preliminary.pdf>

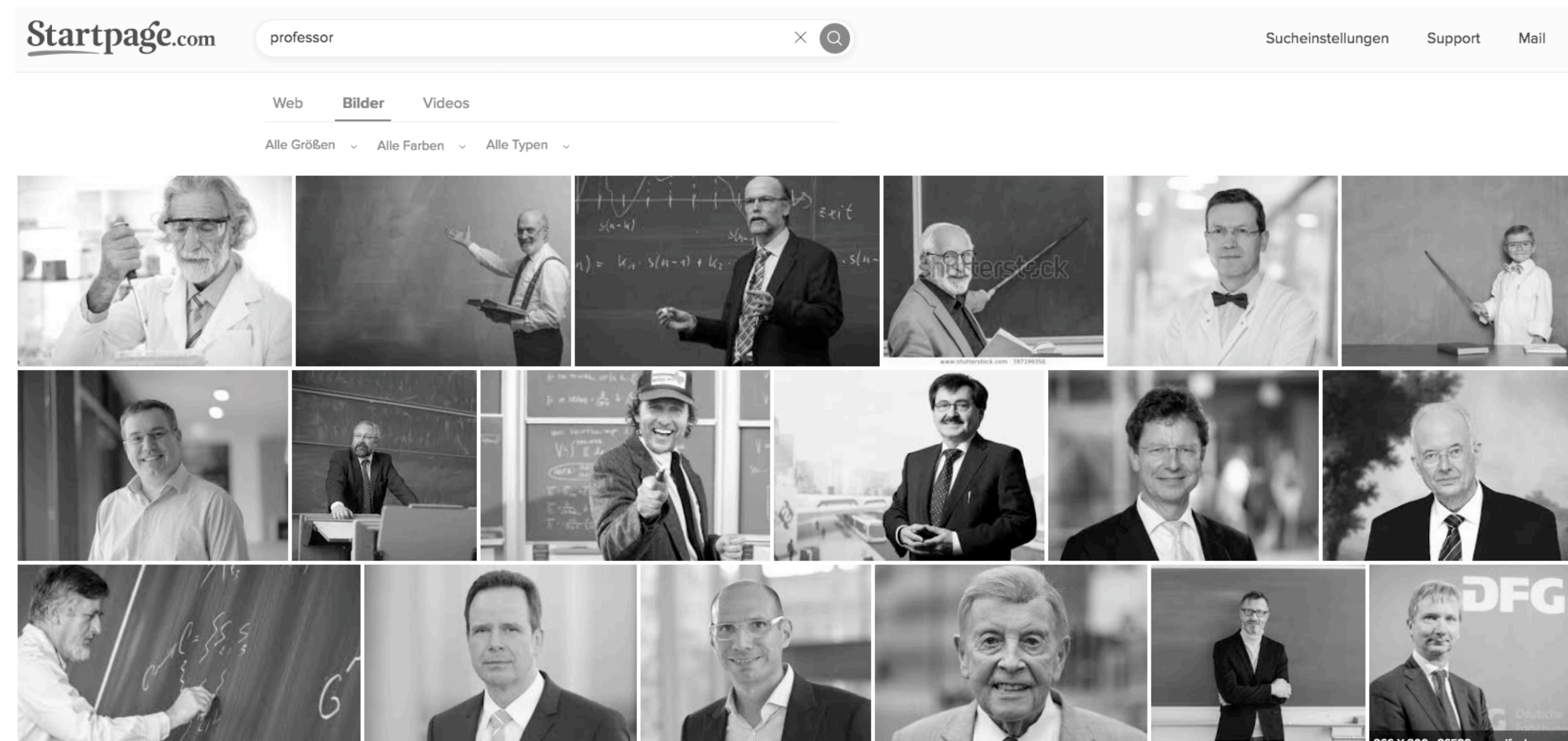


Limited Transferability of Domain Specific Data Sets

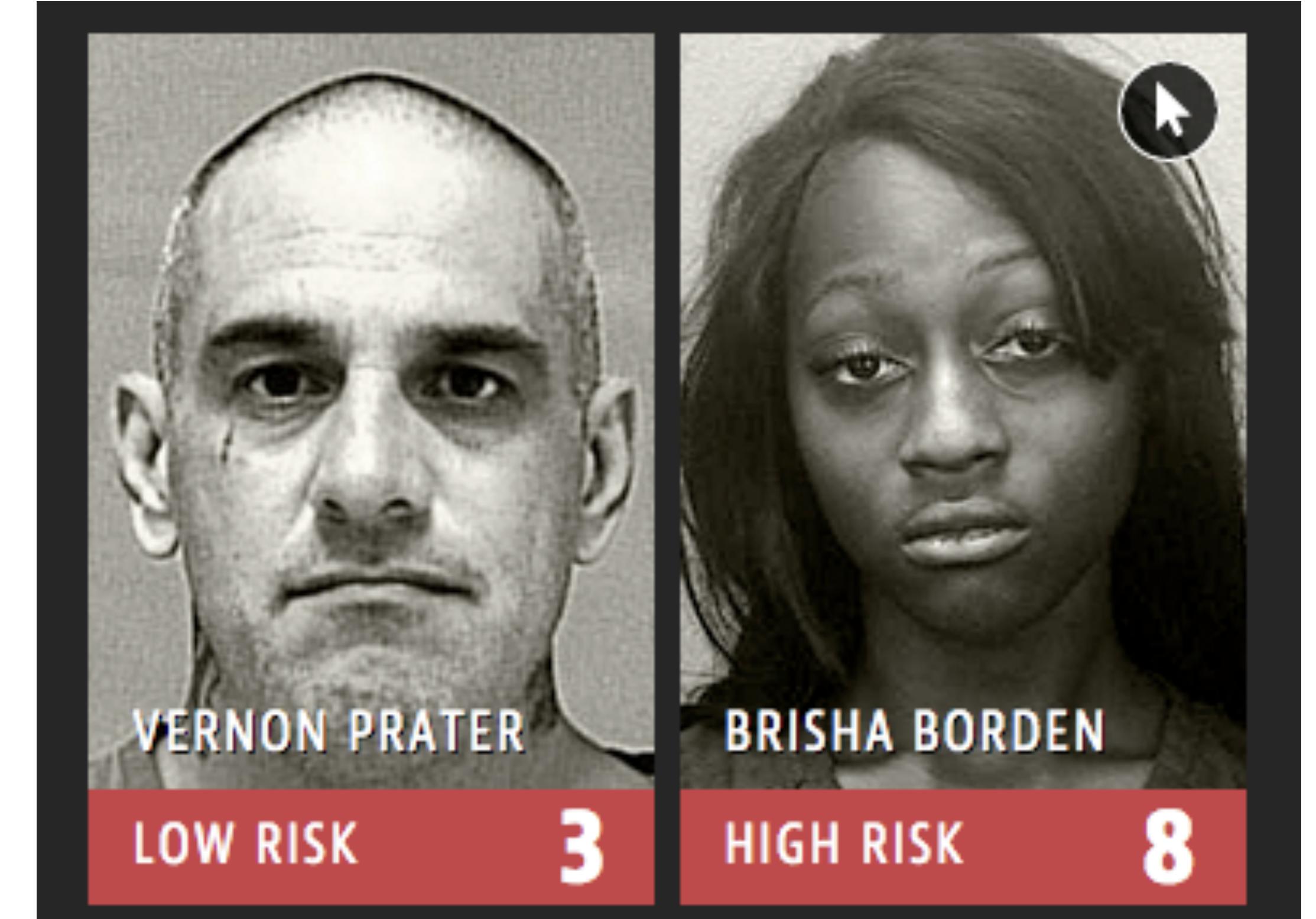


Manifestation of Pre-existing Bias of Historical Data

Search Term: professor



Underestimation of the Influence of Feedback-Loops



Images by <https://www.flickr.com/photos/kneoh/14931652922/>, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>



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

Challenges in Data Science

Inputs to the Algorithm

- » Poorly selected data
- » Incomplete, incorrect, or outdated data
- » Unintentional continuation and promotion of historical biases
- » ...

Inner Workings of the Algorithm

- » Personalization and recommendation services that narrow user options
- » Decision-making systems that assume correlation necessarily implies causation
- » ...

Adopted from Executive Office of the President, et al. Big data: A report on algorithmic systems, opportunity, and civil rights.
Executive Office of the President, 2016.



“

“technology’s interaction with the social ecology is such that technical developments frequently have environmental, social, and human consequences that go far beyond the immediate purposes of the technical devices and practices themselves, and the same technology can have quite different results when introduced into different contexts or under different circumstances.”

Kranzberg’s First Law: “Technology is neither good nor bad; nor is it neutral.”



Human-Centered Data Science

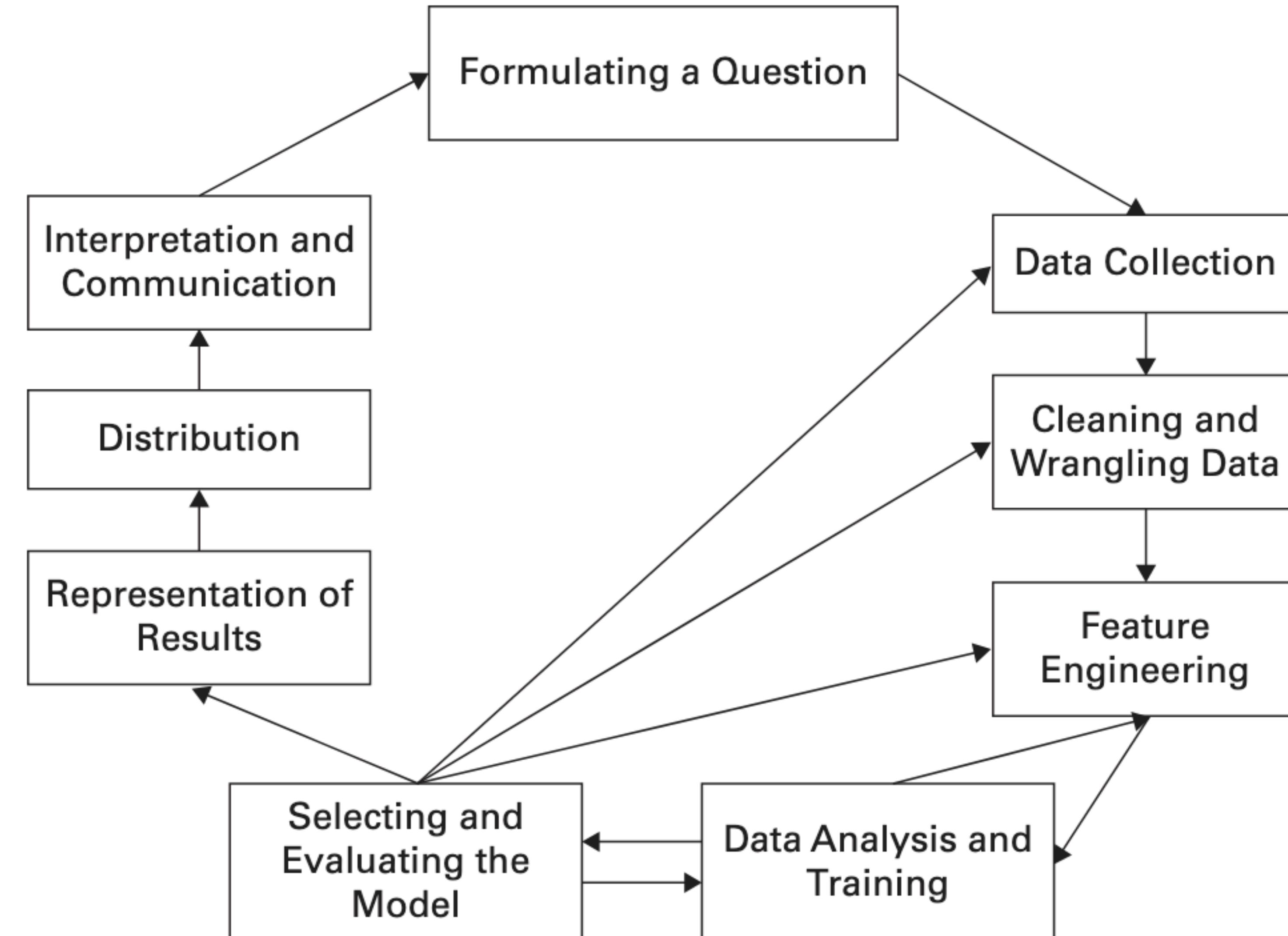
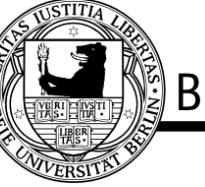
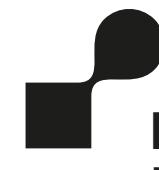


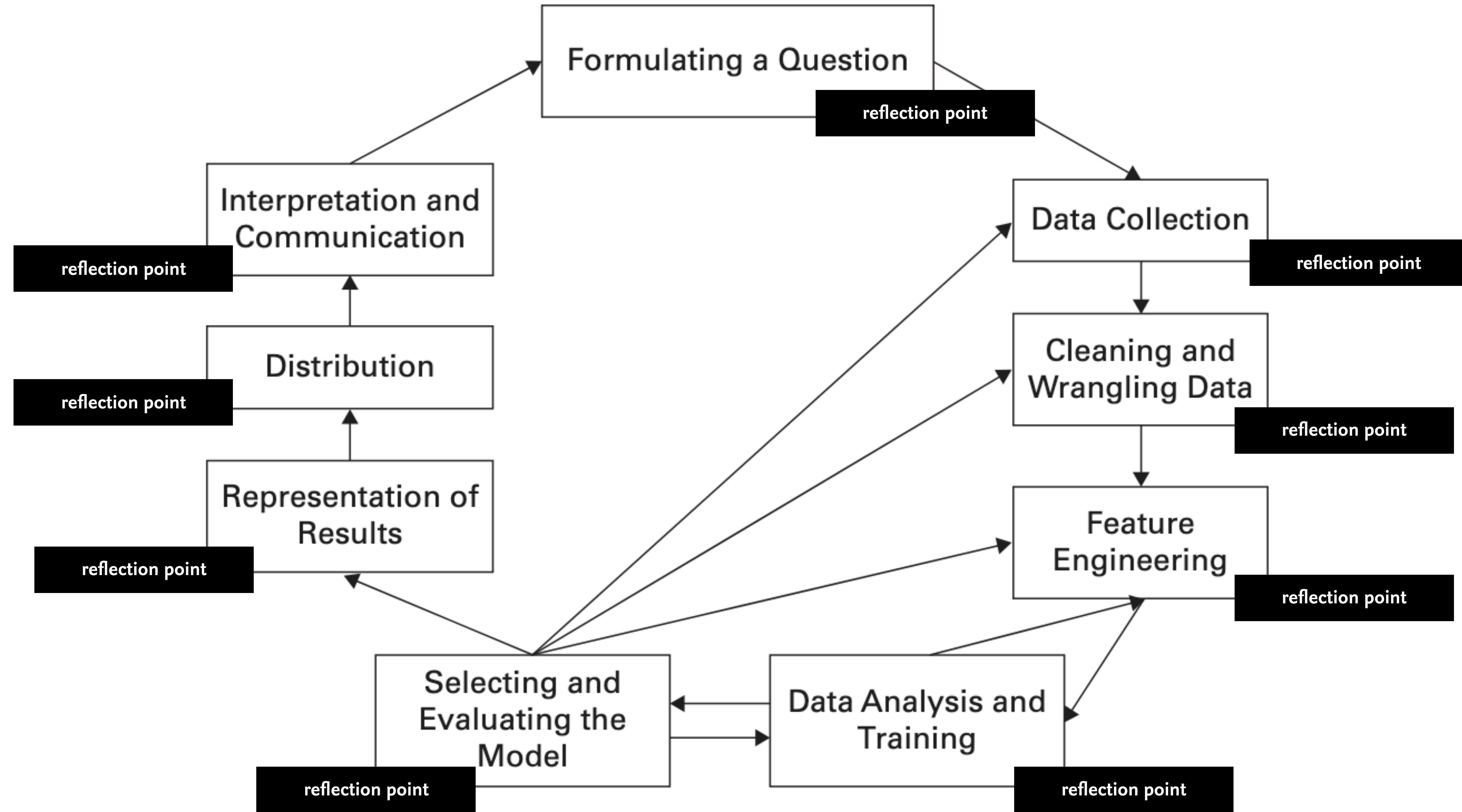
“

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.









**Does it really matter for you?
What are arguments against it?**



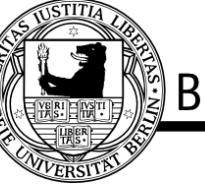


Argument 1: "I'm just an engineer."

Green, Ben. „Data Science as Political Action: Grounding Data Science in a Politics of Justice“. *Journal of Social Computing* 2, Nr. 3 (September 2021): 249–65. <https://doi.org/10.23919/JSC.2021.0029>.



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



Argument 2: “Our job isn't to take political stances. Technology is neutral”

Green, Ben. „Data Science as Political Action: Grounding Data Science in a Politics of Justice“. *Journal of Social Computing* 2, Nr. 3 (September 2021): 249–65. <https://doi.org/10.23919/JSC.2021.0029>.



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



Argument 3: “We should not let the perfect be the enemy of the good.”

Green, Ben. „Data Science as Political Action: Grounding Data Science in a Politics of Justice“. *Journal of Social Computing* 2, Nr. 3 (September 2021): 249–65. <https://doi.org/10.23919/JSC.2021.0029>.



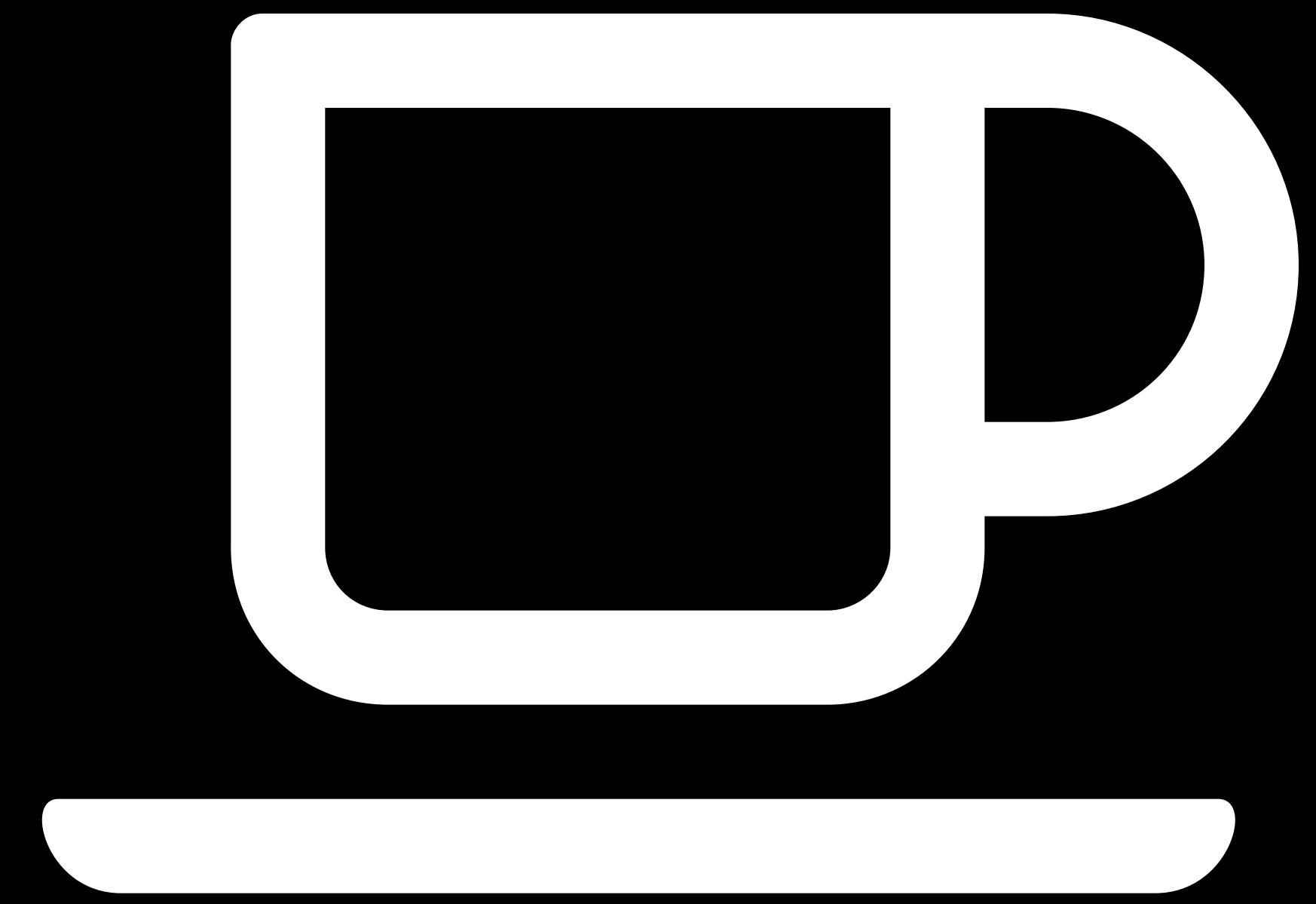
Course «Human-Centred Data Science» | Summer Term 2022 | Claudia Müller-Birn

Methodological Directions in HCDS

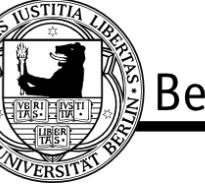
- » **Incorporate rich social context** (e.g., interviews) into data science and computational workflows or scale up qualitative analyses (e.g., based on topic modeling).
- » **Highlight the commonalities** in the inductive nature of data science methods and traditionally qualitative approaches (e.g., machine learning and grounded theory).
- » Combine ethnographic methods with computational data science techniques to **complement their strengths**.
- » **Provide interpretable representations** for describing decision-making process within algorithms.

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.





5 minutes break



Course Concept and Overview



Scope

This course is designed as part of the master program Computer Science and Data Science but open for all interested students from other subjects.

We expect students to have a basic understanding and experience of data science practices. They should know most of the existing tools and frameworks. Students should be capable of applying their expertise to specific problem settings.

The class material is provided in English and we communicate in English as well.

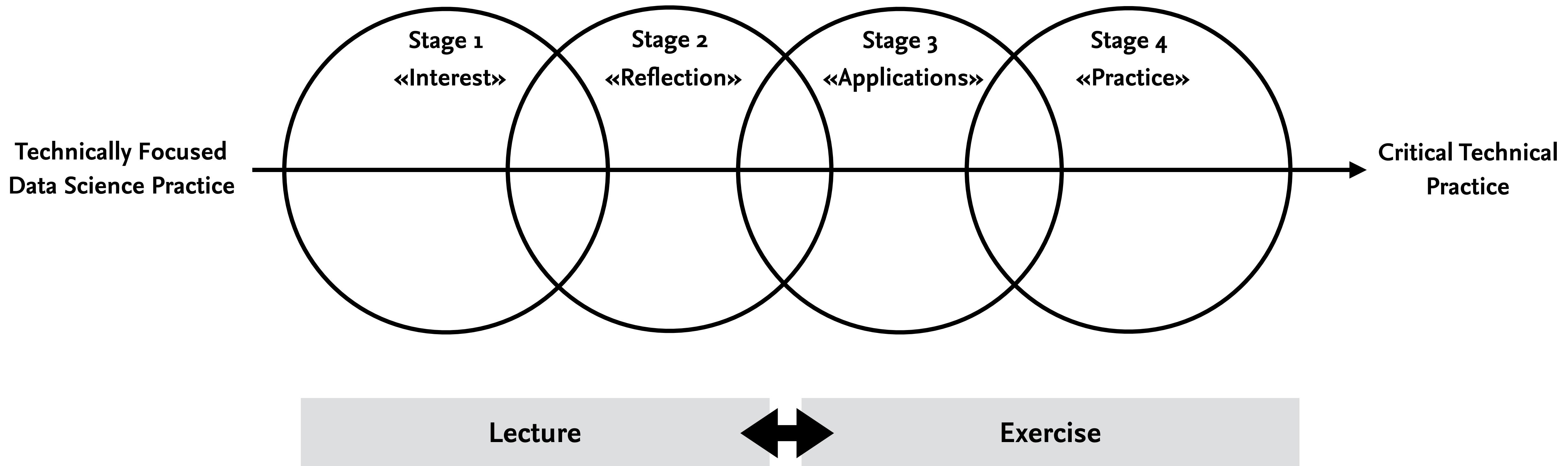
We will adapt this course continuously to your needs, thus, your feedback is appreciated!

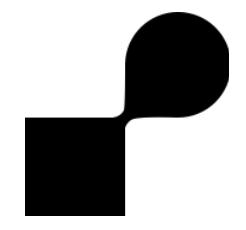


Learning Objectives

- » Apply human-centered design methods in the data science practice, including ethical concerns and privacy requirements
- » Build reproducible data science workflow
- » Differentiate the terms bias, fairness, accountability, transparency and explanations
- » Apply measures, techniques and frameworks for human-centered explainable AI (XAI)
- » Augment data science workflows by qualitative research approaches

Concept: Establish a Critical-Reflexive Cycle





HCDS Lecture

The lecture takes place in person each Thursday from 4 to 6 PM. We provide all slides in a PDF format in Whiteboard. In addition, each lesson comes with additional resources (articles, links, videos, etc.).

We designed the lecture not for "listening only"-mode. Instead, we want to **raise interest** and **provoke reflection**. Thus, four assignments accompany the lecture series. Each assignment consists of a case study describing a real-world situation. Each case study presents a hypothetical but realistic situation to initiate and promote discourses on ethical problems in data science. You need to read each case study in advance, answer the provided questions, and discuss your answer in class.

HCDS Exercise

We organize the exercise as a mix of plenum & group discussions, in-class activities, student presentations, feedback sessions, and peer reviews.

Within the exercise, we want to **translate insights into application** and **enable (participatory) data science practice**. First, we support students to incorporate the concept/perspective learned in the lecture into their data science practice. We want to encourage students' skepticism about and reinterpretation of their practice through five practical assignments. Second, we enable a participatory practice by supporting students to scrutinize assumptions that existing tools (e.g., data sets, trained models, data sets, software libraries, and services) might exhibit and learn about possible limitations in their applicability.



Assignments

There are two types of assignments:

- » **written reflections** relate to a case study and the provided questions.
- » **programming assignments** (e.g., analysis of a datasets or implementing measures/concepts presented during the lecture).

Assignments are provided via Whiteboard/github. Written reflections are submitted individually while programming assignments are submitted in groups.



Criteria for Active Participation

Your final grade is based on the result of your **written exam** only. But ...

In order to actively participate in this course, you need to fulfill the following requirements:

- » You need to submit **(n-1) written reflections**. [planned are 4]
- » You need to submit **(n-1) programming assignments**. [planned are 6]

Lecture Schedule

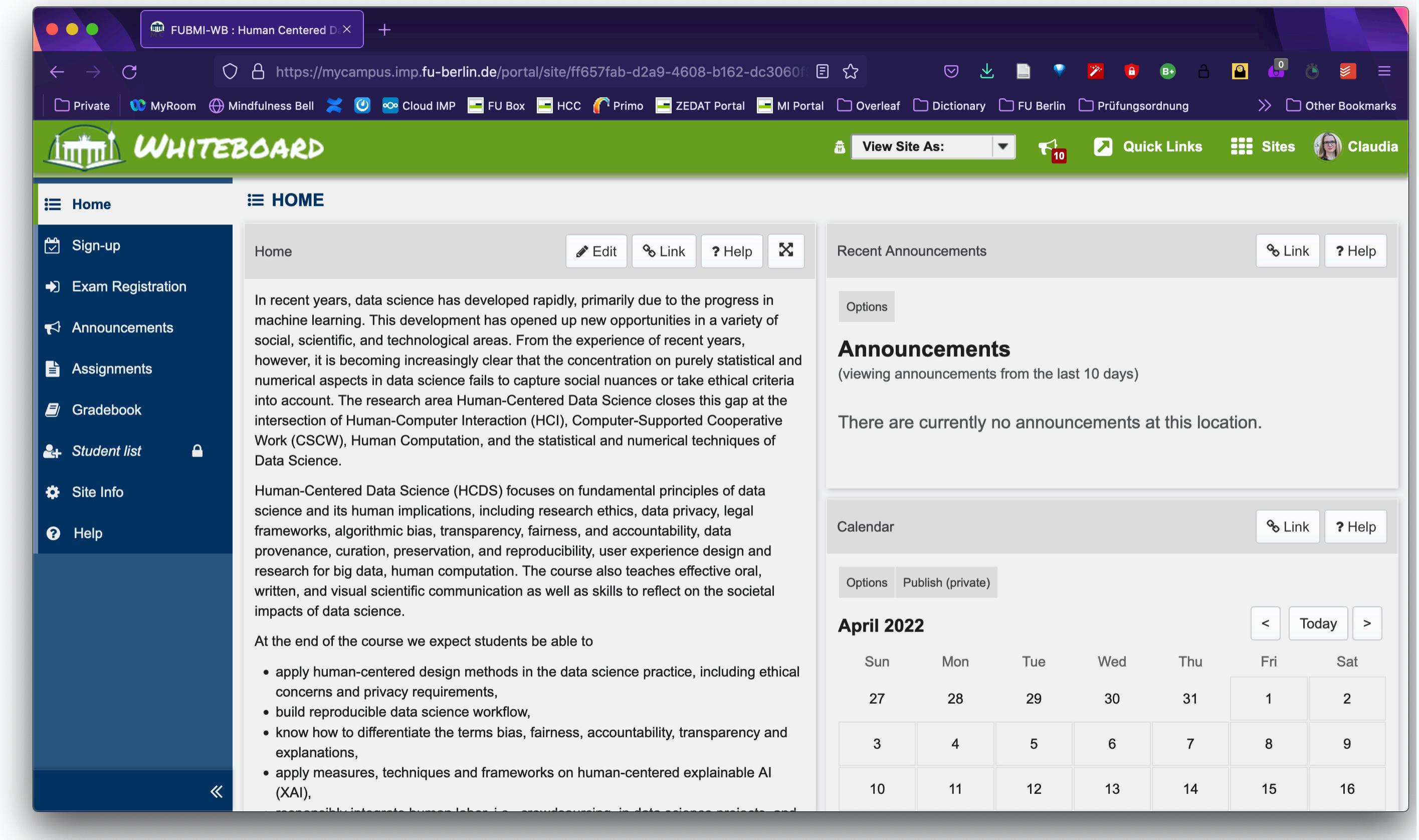
1	21.04.2022	-
2	28.04.2022	Introducing Human-Centered Data Science
3	05.05.2022	Discussing Human Values and their Impact on Your Data Science Practice
4	12.05.2022	Enabling the Reproducability of Your Data Science Practice
5	19.05.2022	Examining what becomes Data in Your Data Science Practice
6	02.06.2022	Understanding Approaches to Identify, Mitigate and Avoid Bias
7	09.06.2022	Dealing with the Complexity of Fairness

Lecture Schedule

8	16.06.2022	Considering Transparency in Your Data Science Practice
9	23.06.2022	Differentiating the Means of Interpretability
10	30.06.2022	Knowing Existing Explanation Methods and their Limitations
11	07.07.2022	Designing Human-Centered Explanation Interfaces
12	14.07.2022	Regarding Privacy and Sensitive Data in Your Data Science Practice
13	21.07.2022	Exam - we look for a new date :)

Course Coordination & Communication

Our main platform for providing the content, our communication and our coordination is Whiteboard.



The screenshot shows a web browser window for the course "FUBMI-WB : Human Centered". The URL is <https://mycampus.imp.fu-berlin.de/portal/site/ff657fab-d2a9-4608-b162-dc3060f>. The page title is "WHITEBOARD". The left sidebar has a dark blue background with white text and icons, listing "Home", "Sign-up", "Exam Registration", "Announcements", "Assignments", "Gradebook", "Student list" (locked), "Site Info", and "Help". The main content area has a light gray background with a green header bar. It displays the "HOME" page content. The content includes a brief introduction to data science, the focus of the course (Human-Centred Data Science), learning outcomes, and a list of expected skills. On the right side, there are sections for "Recent Announcements" (empty), "Announcements" (empty), and a "Calendar" for April 2022. The calendar shows days from 27 to 16, with the 1st and 2nd highlighted. The top of the browser window shows various tabs and icons, and the status bar indicates "View Site As:" and "Quick Links".

Check your Insights

- » At the top of the AI-hype, what kind of research was no longer considered to be necessary?
- » What type of problems have we briefly introduced that are caused by data-driven applications?
- » How can you differentiate these problems? Where do they originate?
- » What is Human-Centered Design?
- » How do you define Human-Centered Data Science?





«Human-Centered Data Science»

Next week: 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

References

- » Aragon, Cecilia, Shion Guha, Marina Kogan, Michael Muller, und Gina Neff. Human-Centered Data Science: An Introduction. Cambridge, MA, USA: MIT Press, 2022.
- » Aragon, Cecilia M, Clayton Hutto, Andy Echenique, Brittany Fiore-Gartland, Yun Huang, Jinyoung Kim, Gina Neff, Wanli Xing, und Joseph Bayer. „Developing a Research Agenda for Human-Centered Data Science.“, 529–35. New York, New York, USA: ACM Press, 2016. <https://doi.org/10.1145/2818052.2855518>.
- » Kogan, Marina, Aaron Halfaker, Shion Guha, Cecilia Aragon, Michael Muller, und Stuart Geiger. „Mapping Out Human-Centered Data Science“. In GROUP '20, 151–56. New York, NY, USA: SIGCHI ACM Special Interest Group on Computer-Human Interaction, 2020. <https://doi.org/10.1145/3323994.3369898>.
- » Kranzberg, Melvin. „Technology and History: ,Kranzberg's Laws“. Technology and Culture 27, Nr. 3 (1986): 544–60. <https://doi.org/10.2307/3105385>.
- » O'Neil, Cathy. Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy. Broadway Books, 2016.