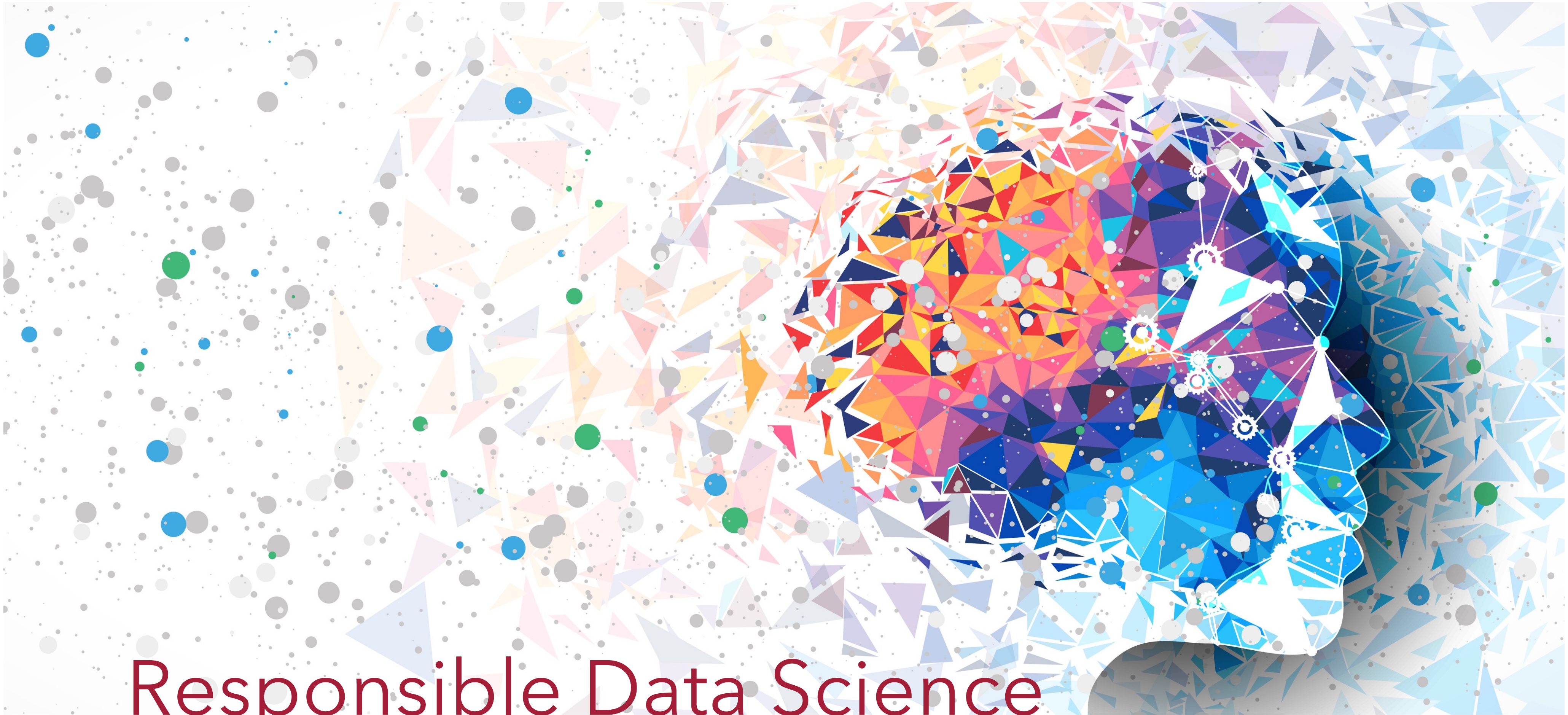
HPI Hasso Plattner Institut
Digital Engineering • Universität Potsdam

Internationales Zentrum für Ethik in den Wissenschaften (IZEW)

Eberhard Karls Universität Tübingen



Responsible Data Science

Session 5: 24.05.2023, 15.15 – 19.30 h
MA Seminar, SoSe 2023, Hasso-Plattner Institut

Dr. Simon David Hirsbrunner

Today

topic	time
Introduction	15h15
Input: data quality and its ethical implications	15h30
Guest input und discussion: data diversity (Lisa Köritz)	16h00
—— Break ——	17h00
Student presentation and discussion: Data trusts (Oliver Derwisch)	17h15
—— Break ——	18h15
Exercise: developing an ethical code for a data trust in a specific domain	18h30
Sharing and discussing insights from groups	19h00
Seminar paper brainstorming in small groups	19h15
End	19h30

Your favorite learning from last (privacy-related) session?

-- Illustrating picture removed for copyright reasons ---

What does quality mean for you?

-- Illustrating picture removed for copyright reasons ---

Characterizing data quality

Data quality can be understood as "data that are fit for use by data consumers" (Wang and Strong 1996: 6)

A data quality dimension is referred to as "a set of data quality attributes that represent a single aspect or construct of data quality." (Wang and Strong 1996: 6)

Characterizing data quality

Data quality can be understood as
"data that are fit for use by data consumers"
(Wang and Strong 1996: 6)

A data quality dimension is referred to as "a set of data quality attributes that represent a single aspect or construct of data quality."
(Wang and Strong 1996: 6)

→ Quality is often dependent on context

Data quality in the EU AI act

Art. 44
Training, and where applicable, validation and testing data sets, including the labels, should be **sufficiently relevant, representative, appropriately vetted for errors** and **as complete as possible** in view of the **intended purpose of the system.** They should also have the appropriate **statistical properties,** including as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used, with specific attention to the **mitigation of possible biases** in the datasets, that might lead to **risks to fundamental rights or discriminatory outcomes** for the persons affected by the high-risk AI system.

(EU AIA proposal, EU parliament compromise by 16.05.23, highlighting added)

AI data quality dimensions

Timeliness, reputation, balance, ease of manipulation, diversity, documentation, efficiency, uniqueness, fairness, precision, credibility, consistent representation, consistency, accuracy, cost, portability, privacy, relevancy, representativity, traceability, security, transparency, concise representation, appropriate amount of data, reliability, understandability, completeness, value-added, accessibility.

(DQ glossar of Kitqar project https://www.kitqar.de/de/veroeffentlichungen/dq-glossar accessed 22/05/2023)

Ethical dimensions of AI data quality (I)

Many dimensions of DQ in AI have an ethical component. Examples:

Accuracy
Accuracy describes the correspondence between a phenomenon in the world and its description as data. When comparing the data value with the empirically observable value, the difference can be determined either binary (equal or unequal) or the degree of difference can be determined using a similarity measure (e.g., as similarity between 0 and 1). The quality of accuracy is particularly relevant for data whose factual correctness can be conclusively clarified and whose meaning is not ambiguous.
(Kitqar DQ glossar, translated from German)

Ethical dimensions of AI data quality (I)

Many dimensions of DQ in AI have an ethical component.
Examples:

Accuracy

Accuracy describes the correspondence between a phenomenon in the world and its description as data. When comparing the data value with the empirically observable value, the difference can be determined either binary (equal or unequal) or the degree of difference can be determined using a similarity measure (e.g., as similarity between 0 and 1). The quality of accuracy is particularly relevant for data whose factual correctness can be conclusively clarified and whose meaning is not ambiguous.
(Kitqar DQ glossar, translated from German)

Ethical considerations

The quality of labels in annotated data for supervised learning varies considerably. Faulty or biased labels may underrepresent or misrepresent subgroups of the population and may lead to negative consequences for these groups.
Also, faulty data may lead to faulty AI systems and can therefore lead to safety risks or other negative impact for users and stakeholders.

Ethical dimensions of AI data quality (II)

Completeness

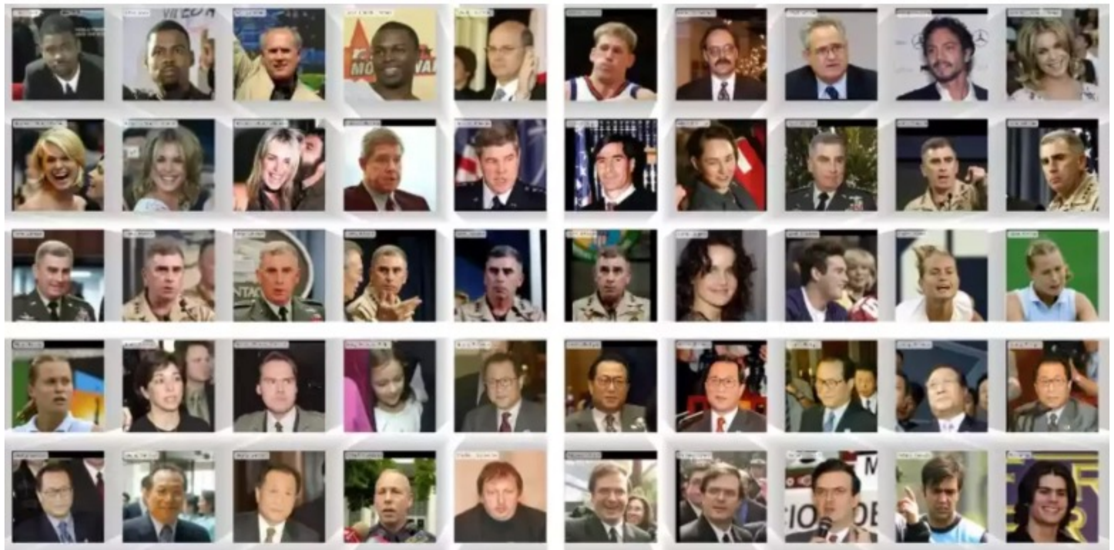
Completeness is the relationship between the amount of represented data and the amount of data to be represented. While the former can be counted (number of rows, number of non-null values), the latter is often only estimated. A dataset (e.g., a table) is considered complete with respect to a domain if every entity in the domain is represented in the dataset. A record (e.g., a row) is complete if it has a value for every attribute (column).
(Kitqar DQ glossar, translated from German)

Ethical considerations

Who gets to define what constitutes 'complete data' ? Compared to more traditional data sources with a well-defined domain and population, Big Data is rarely 'complete'. It's composition may considerably vary in time and depending on collection methods. Completeness is often a matter of interpretation and therefore subjective.

Example: datasets for facial recognition

LFW [Labeled Faces in the Wild] , a dataset composed of celebrity faces which has served as a gold standard benchmark for face recognition, was estimated to be 77.5% male and 83.5% White (Buolamwini and Gebru 2018; Han and Jain, 2014).



Presentation on data diversity by Lisa Köritz

Presentation and discussion

O'Hara K (2019) Data Trusts. Ethics, Architecture and Governance for Trustworthy Data Stewardship.

(presentation by Oliver Derwisch)

Scenario-based design (I): definition

Scenario-based design is a family of techniques in which the **use of a future system is concretely described at an early point in the development process.** Narrative descriptions of envisioned usage episodes are then employed in a variety of ways to guide the development of the system that will enable these use experiences.

-- Illustrating picture removed for copyright reasons ---

(Rosson and Carroll 2007: 1)
image: https://unitid.nl/2014/02/scenario-based-design/

Scenario-Based Design (II): example

S Martin is a police officer at the organized crime unit of the federal police. He currently investigates the selling of fake COVID-19 vaccination passports by an alleged criminal organization named *The Medics*. The Medics offer the counterfeit certificates to their S customers via the Telegram messenger. Unknown to Martin yet, S Chris, S Carlos, and S Eggert are Medics members, also communicating with their colleagues and suppliers via group Telegram channels while using pseudonyms, sometimes coded language, and images. In their free time, they also communicate with several friends, including their girlfriends, S Sarah and S Marta, who are unaware of their business. Martin's police unit gathers much information about The Medics using traditional investigative methods. This information leads to the identification of the suspect, Chris, who seems to be a low-level member of The Medics. On one evening, Chris is found with blank vaccination certificates during a traffic stop. He is arrested, and his phone is seized by investigator Martin, who aims at using the information on the phone to track down the individuals pulling the strings. After calling judge S Robert to get a search warrant, which is granted, he then searches Chris's unprotected phone, finds the Telegram communication, and extracts it. He recalls that his superior, S Dr. D, asked him to try out the new AutoCommAnalyzer software, which was recently purchased from the multinational company AI-Tech Corp. The software purchase was part of a strong push by the government to digitally optimize work processes at the police forces. Martin looks at the training notes by the head developer S Molly, trying to remember how the machine learning-driven software — trained with texts by S Alf and S Bert — is supposed to direct him to the relevant communication. The software presents him with the most frequent contacts, with Sarah on top. He reads through this communication, as the software has flagged several words like package and hospital, discovering some explicit images but finding that the

(Fischer MT, Hirsbrunner SD, Jentner W, et al. (2022) Promoting Ethical Awareness in Communication Analysis: Investigating Potentials and Limits of Visual Analytics for Intelligence Applications. In: *2022 ACM Conference on Fairness, Accountability, and Transparency*, New York, NY, USA, 20 June 2022, pp. 877–889. FAccT '22. Association for Computing Machinery. DOI: 10.1145/3531146.3533151)

Scenario-Based Design (III): qualities

- Scenarios are concrete but rough;
- Scenarios maintain an orientation to people and their needs;
- Scenarios are evocative, raising questions at many levels.

(Rosson and Carroll 2007)

Scenario-Based Design (IV): ingredients

- **Actors** (direct and indirect stakeholders)
- **Setting** (application context, situation of use)
- **Tools and objects** (technologies, interfaces)

(Rosson and Carroll 2007)

Value scenario

- Stakeholders
- Pervasiveness
- Time
- Systemic effects
- **Value implications**

Value scenario = VSD + Scenario-Based Design (SBD)

(Nathan et al. 2007)

3 Exercice

- Watch the provided video representing a scenario of a comprehensive AI system.
- While watching, try to identify the value scenario that is promoted through the video.
- We discuss these value representations in the group.

Stakeholders

- Whose values should be taken into account?
  Stakeholder's values
- What are stakeholders?
  - "A stakeholder is anyone who will be affected, directly or indirectly, by the new system like the end users, the software staff, and the organization's clients." (Shneiderman and Rose 1996: 92)
  - Stakeholders "can be people, groups, neighborhoods, communities, organizations, institutions, or societies, and can also include past and future generations, nonhuman species, and other elements such as historic buildings or sacred mountaintops" (Friedman and Hendry 2019: 37)

Focus on roles, not entities

- Stakeholders are defined by and understood in relationship to their interaction with a technology or sociotechnical system.

- They are considered by role, rather than by "person" or other "entity."
  A "role" pertains to a stakeholder's duties, contextual identity, or particular circumstances.

(Friedman and Hendry 2019: 37)

Dr. Simon David Hirsbrunner

Methods to identify stakeholders

- Semi-structured interviews
- Participant observation
- Document analysis
- Scenario-Based Design
- Integrated Technology Development
- …

4 Exercice

- Form groups of 2-3 people.
- Identify 5+ stakeholders (roles) in the provided scenario.
- Discuss at least one stakeholder with multiple roles.
- Create a stakeholder mapping, placing stakeholders according to their agency in the system.

Dr. Simon David Hirsbrunner

Sources

See entire list of course references on Github:
https://github.com/simonsimson/responsible-data-science/tree/main/slides