

Can code be responsible? - How to study and communicate algorithmic accountability

Henrike Weinert, Lisa-Marie Eckardt, Christina Elmer, Lisa Goldschmidtböing, Katja Ickstadt and Kai-Robin Lange

TU Dortmund University, Germany
henrike.weinert@tu-dortmund.de

Before we can trust algorithms, we must learn how to question and understand them. Based on this principle, we explore how algorithmic accountability can be taught through interdisciplinary, project-based learning as a part of the Digital Humanities Ruhr project. We present a seminar involving students of journalism, statistics, computer science, and data science. In interdisciplinary teams, they investigated real-world algorithmic systems and presented their findings in accessible public formats. The seminar demonstrates how algorithmic accountability can become an integral part of interdisciplinary education in the humanities and beyond. To accomplish this, we reflect on lessons learned and outline future improvements, such as expanding participation and enhancing mentoring to support both the selection of algorithmic cases and the choice of communicative formats. We integrate our improved seminar into the broader Data Literacy Certificate at TU Dortmund University, contributing to the development of micro-credentials in Digital Humanities across the University Alliance Ruhr.

INTRODUCTION

In recent years, algorithmic systems have become integral to decision-making processes in many areas of society, from hiring and policing to education and media consumption. These developments raise questions about the fairness, transparency, and accountability of automated systems. Algorithms are no longer confined to the back end of software applications but are actively shaping human experiences and institutional practices. Their reach includes health diagnoses, financial scoring, criminal risk assessments, and more. As a result, understanding algorithmic decision-making is a pressing societal concern. As part of their responsibility to prepare students for these challenges, universities must take on the task of addressing algorithmic accountability through research and education.

The interdisciplinary project "Digital Humanities Ruhr" addresses this issue by exploring how we can investigate and communicate algorithmic accountability, especially in educational settings. This paper presents our project with a particular focus on a seminar developed at TU Dortmund University that brings together students and teachers of journalism, statistics, data science, and computer science, fostering mutual understanding and collaboration across disciplinary boundaries. In the future, additional academic disciplines will be included, particularly political science and philosophy, to further broaden the perspectives on algorithmic accountability and to reach more fields within the humanities. The central question guiding the course was: Can code be responsible? If not, who bears responsibility for the outcomes produced by automated processes – and how can this be meaningfully examined and communicated?

DATA LITERACY EDUCATION AT TU DORTMUND UNIVERSITY: FOUNDATIONS FOR INTERDISCIPLINARY COMPETENCE

Data literacy has increasingly been recognized as a foundational competence across all areas of education and society (Ridsdale et al., 2015; Schüller, 2020; Data Literacy Charter). It includes the ability to understand, interpret, and critically engage with data and data-driven systems. It involves more than statistical reasoning; it requires an understanding of the data lifecycle, algorithmic logic, ethical implications, and communication skills. TU Dortmund University has placed a strategic focus on fostering data literacy across all disciplines, establishing the 'Data Competence Network' in 2020 to coordinate and advance these efforts. This network is based at the TU Dortmund Center for Data Science and Simulation (DoDaS), a central academic unit dedicated to supporting interdisciplinary collaboration in data-related fields.

DoDaS functions as a central hub that supports interdisciplinary exchange and infrastructure for data-driven research and education across the university. It connects faculties involved in data

generation, methodological innovation, and applied data analysis, offering a space for joint projects, shared services, and coordinated teaching strategies. It coordinates collaborative teaching formats, research infrastructures, and interdisciplinary training programs. Data Literacy Education is anchored here as a key element of digitization across the university. It is designed as a flexible framework that enriches existing curricula – such as introductory statistics courses, courses on data ethics, or subject-specific seminars in the humanities. It does so both by bringing more data-related thinking into discipline-specific contexts, and by making data literacy modules or (introductory) statistics modules more responsive to the needs and perspectives of the respective fields. This bidirectional enrichment is pursued by encouraging statistics courses to focus more on data processing and critical thinking, while ensuring that data literacy offerings stay closely aligned with the disciplinary contexts. These modules are implemented as self-contained micro-learning units worth 1 CP. These units can be integrated into various study programs and aim to deepen students' ability to critically process data.

The program also includes a structured Data Literacy Certificate, which comprises at least 10 CP and consists of the following parts:

- A. Introduction (1 CP): Module A covers basic concepts like data awareness, ethical perspectives, and the role of data in everyday and scientific contexts.
- B & C. Basic and Supplementary Modules: These modules cover foundational aspects of data analysis and critical evaluation. Topics include different types of data (quantitative, qualitative, structured, unstructured), descriptive statistics, data visualization techniques, and the basics of inferential statistics. Additionally, students learn to identify typical errors and pitfalls in data-based reasoning, such as biased samples, misleading graphs, or inappropriate statistical conclusions. These modules also introduce basic tools for data analysis and visualization, such as spreadsheet software or beginner-friendly coding environments, depending on the disciplinary context.
- D. Advanced Module: This module requires students to engage in an in-depth exploration of at least one specific area of data literacy. This can take the form of a seminar, case-based project work, or an interdisciplinary study focused on topics such as text mining, data ethics, or algorithmic accountability. The goal is to deepen both technical understanding and critical reflection within a chosen thematic field, fostering a more specialized and application-oriented competence.

This certificate is designed to be interdisciplinary, flexible, and integrable into existing degree programs. It is recognized across the University Alliance Ruhr and includes open educational resources (OER) to support reusability. As algorithmic technologies become more prevalent, discussions are emerging about the need to expand data literacy toward AI literacy. This emerging concept includes understanding how artificial intelligence systems are developed, how they function, and what ethical and societal implications they entail. The Data Literacy program at TU Dortmund is currently being expanded in this direction, with the goal of systematically integrating AI-related competencies into its certificate structure and teaching modules. Algorithmic Accountability forms one of the thematic pillars of the advanced stage of the certificate.

THE PROJECT "DIGITAL HUMANITIES RUHR"

The seminar and teaching concept presented in this paper is embedded in the broader project "Digital Humanities Ruhr: Algorithmic Accountability," which is part of TU Dortmund University's strategic efforts in the field of data literacy. Funded by the Freiraum 2023 program of Stiftung Innovation in der Hochschullehre, the project is a collaboration between the universities of the University Alliance Ruhr, UAR (TU Dortmund University, Ruhr University Bochum, and University of Duisburg-Essen). It focuses on strengthening digital competencies in the humanities and social sciences by integrating data literacy into teaching, especially through interdisciplinary and practice-oriented formats.

At TU Dortmund University, the project is coordinated by DoDaS and implemented jointly by the departments of Statistics and Journalism. It builds on the existing Data Literacy Certificate program.

The project follows three main goals:

- To understand and investigate the impact of socially relevant algorithms and make them accessible for critical discussion.

- To report on algorithmic systems and their implications in a way that is comprehensible and engaging for a broad audience.
- To raise awareness of ethical standards, data protection, transparency, and accountability – both from the perspective of users and developers.

These aims are supported through the development of reusable teaching materials, practical case studies, and train-the-trainer formats that help integrate algorithmic accountability sustainably into various curricula.

Within the University Alliance Ruhr, the participating institutions develop coordinated learning materials and offer modules that can be mutually recognized. At the University of Duisburg-Essen, the project focuses on the integration of corpus linguistic and text technological methods into digital humanities teaching. This involves revising the curriculum of the German Studies program and creating a dedicated module to introduce students to computational text analysis. At Ruhr University Bochum, the emphasis is on embedding future skills into humanities curricula. Here, interdisciplinary digital competencies – especially for working with large text corpora – are taught and integrated into the elective areas of two-subject Bachelor programs across 14 faculties. Teaching formats at all locations are adapted to disciplinary contexts while sharing a common methodological foundation. The project also contributes to the development of a UAR-wide micro-credential in digital humanities.

THE SEMINAR "CAN CODE TAKE RESPONSIBILITY?" AS AN EXAMPLE OF INTERDISCIPLINARY TEACHING PRACTICE

As a practical example of how algorithmic accountability can be addressed in interdisciplinary teaching, we developed the seminar "Can Code be Responsible? – Investigating and Communicating Algorithmic Accountability." This seminar demonstrates how students from various disciplines can collaborate to understand, question, and communicate the impact of algorithmic systems on society.

Our teaching concept builds on the idea that understanding the technical foundations of algorithms is crucial for meaningful accountability discussions. Simultaneously, we emphasize the need for interdisciplinary collaboration, particularly with disciplines that traditionally do not engage deeply with data science, such as journalism and other humanities.

The seminar aims to create an inclusive and experimental learning environment in which students from different academic backgrounds explore the social impact of algorithms. Participants work in interdisciplinary teams to investigate specific cases of algorithmic decision-making. These cases are chosen to highlight the multifaceted nature of algorithmic accountability, including technical, ethical, legal, and communicative dimensions. The teaching approach combines lectures, project-based learning, and public communication training. The interdisciplinary setting allows students to confront different epistemologies and methodologies, fostering a deeper understanding of the complexity of algorithmic systems. One major goal of the seminar is to empower students not only to analyze algorithmic systems critically but also to communicate their findings in accessible and impactful formats.

The seminar was first offered in the winter semester 2024/25 and included weekly sessions combining theoretical input and group work. Early sessions provided examples of journalistic engagement with algorithmic systems and introduced students to core ethical and technical concepts. They also covered basic theoretical and technical foundations: what algorithms are, how they function, and what types of algorithms exist. The aim was to build a shared vocabulary and baseline understanding across all disciplines involved. Topics in the first sessions addressed foundational ethical questions, technical challenges, and public controversies surrounding algorithmic systems. These included algorithmic biases, the importance of transparency, and the difficulty of explaining how AI systems function. Case examples such as recommender systems in social media and predictive policing served to contextualize the discussion. Guest lectures from journalists and data scientists provided concrete perspectives and highlighted professional approaches to reporting on and working with algorithms.

Students were then asked to select a real-world algorithmic system and analyze its implications. To support this process, we provided a set of example cases: these included a deliberately biased algorithm we had designed for the purpose of critical deconstruction – challenging students to discover and explain the underlying mechanism of discrimination – as well as selected examples from journalistic

investigations that could serve as templates for replication or adaptation. Students were also encouraged to identify their own cases of interest, leading to a broad variety of project ideas. They were furthermore free to choose the format in which to present their results, allowing them to tailor the medium to the message and audience. The seminar culminated in six interdisciplinary group projects, each addressing a specific aspect of algorithmic accountability:

- **Spotify shuffle algorithm:** This group designed a strategy to explore how Spotify's shuffle algorithm behaves and whether any patterns or biases could be detected. They conducted two experiments. In the first, they created several new Spotify accounts that differed only in selected parameters such as user age, then recorded the first 100 songs delivered through the shuffle function. They collected associated track data such as genre and danceability using Spotify's API – later supplemented with web scraping after access was restricted. In a second experiment, they used the same five starting tracks for each account and tracked 45 subsequent songs. The group observed patterns such as persistent genre repetition and a tendency for certain artists to recur frequently, even in supposedly random order. Their results were presented in an interactive H5P learning module within the online learning platform Moodle. This included not only their findings, but also extensive background materials and quizzes – for example, on how the API works or providing context on featured artists. The students also critically reflected on their limited dataset and emphasized that, despite their elaborate methodology, their findings are not generalizable due to the inherent limitations of the approach.
- **Childcare allocation algorithm:** In Germany, there is often a shortage of daycare center places, which leads to dissatisfaction among parents when they do not receive their preferred placement. In addition, existing allocation procedures are not always transparent. The student group therefore developed a set of criteria for fair allocation and proposed a solution based on the Gale-Shapley algorithm. Citing a study by the Bertelsmann Foundation (Overdiek & Petersen, 2022) that shows low public trust in algorithmic fairness specifically when it comes to the allocation of childcare places, the group chose to present their results via an Instagram campaign. The posts explain the current situation and highlight the advantages of using the Gale-Shapley algorithm. The students also reflected on how to tailor the format to their target audience – young parents and daycare administrators – by creating "info posts of the week". The result is a visually appealing Instagram profile, though it still has room for improvement in explaining the algorithm in more depth and balancing informative content with promotional tone.
- **Grading algorithms in schools:** The project was based on an artificial grading process inspired by the situation in the UK during the COVID-19 pandemic – where final grades were algorithmically calculated and led to unfair outcomes (Bhopal and Myers, 2023). The students were given access to a simulation of such a system in which we purposely generated grades based on various influencing factors. The factors included clearly performance-related variables, such as previous grades, as well as background characteristics like parental income. The inclusion of such variables raises ethical concerns and requires careful justification. The students' task was to uncover and critically assess these distortions. They developed a design of experiments to identify which variables influenced the outcomes. Their findings were documented in two formats: a traditional academic seminar paper and a journalistic article. Not all programmed biases were discovered, which in itself became part of the reflection process.
- **ChatGPT and voice input:** This group investigated whether the use of ChatGPT via voice input could lead to discriminatory outcomes depending on speakers' accents or dialects. They recruited speakers with various linguistic backgrounds, including Kurdish, Turkish, Arabic, and French accents, as well as regional dialects such as Northern Hessian and Bavarian. As a control group, they included one speaker of standard German (Hochdeutsch). All participants asked ChatGPT identical questions – covering topics like current events or everyday situations – while deliberately using dialect-specific vocabulary (e.g., "Brötchen" vs. "Semmel"). The resulting responses were analyzed by two coders, using the categories of elaborateness, correctness, and comprehension, and were rated as better, worse, or similar compared to the control group. The group found indications that accent or dialect had an effect on the quality of responses. However, due to the small data set and the control group consisting of only one person, no definitive conclusions about discrimination

could be drawn. The group documented their approach and findings in a journalistic article that critically reflected on their methodology and preliminary results.

- Team composition algorithms: This group explored how algorithms are used to assemble teams in professional environments, considering factors such as productivity, diversity, and fairness. They conducted a case study comparing human resources software solutions – tools used for personnel management and team formation – and adopted a technical perspective to understand how different criteria could be modeled algorithmically. They made sure the algorithm incorporated the intended variables, but spent relatively little time evaluating which variables should be used or whether their use might be ethically questionable or discriminatory. The group documented their work in a traditional academic seminar paper. However, they did not address the question of who holds responsibility for the algorithmic decision-making process – a key aspect of algorithmic accountability that remained underexplored in their project.
- AI image generation with Stable Diffusion: Titled "More Beautiful Stereotypes," this project visualized how AI-generated images reflect and amplify stereotypes. The group generated a total of 1,000 images – 25 per category – based on prompts describing people by nationality (e.g., Europeans, Asians, Germans, Americans), personality traits (e.g., attractive, sporty, successful, unintelligent), professions (e.g., architect, CEO, housekeeper), and finally criminal roles (drug dealer, terrorist). Prompts were written in English to maintain gender neutrality. The resulting images were then coded by two independent coders according to gender and skin color. The group summarized their findings in a journalistic article, highlighting how strongly the image generator reproduced societal biases – for instance, the overwhelming dominance of male representations, regardless of actual population distributions, including for example a complete absence of women in CEO images. They supplemented their article with an expert interview and concluded by emphasizing how important it is for users of generative AI tools to think critically about the content of their prompts and to consider how precise language can help counteract stereotypical patterns arising from biased training data.

LESSONS LEARNED AND OUTLOOK

For the next iteration of the seminar in the winter semester 2025/26, we aim to include students from additional programs beyond journalism, allowing for even more diverse perspectives and backgrounds in group work.

The first implementation of the seminar revealed a number of challenges and opportunities – both for the students and for the teaching team. Working with inaccessible or proprietary algorithms required creative research strategies, such as reverse engineering or critical media analysis. Students had to grapple with the ambiguity and complexity of real-world systems, which contrasted with the controlled environments of typical coursework. The interdisciplinary collaboration proved both enriching and demanding, particularly in aligning expectations and methods across disciplines.

From a pedagogical perspective, the seminar confirmed the value of project-based and public-oriented learning. Students reported high motivation due to the relevance of the topics and the opportunity to create tangible outputs. However, challenges included differing levels of technical expertise, coordination difficulties in group work, and the need for continuous mentoring.

A central lesson from the initial run of the seminar is the importance of providing students with close guidance at two critical junctures: first, in selecting a suitable example of algorithmic accountability, and second, in determining an appropriate format for presenting their findings. These two aspects – content and format – are deeply interdependent and must be aligned to support both analytical depth and communicative effectiveness. For example, a journalistic article demands a different narrative structure and stylistic register than an Instagram campaign or an interactive learning module in Moodle. In the next iteration of the seminar, we therefore plan to accompany the student teams more closely throughout the process and collaboratively develop criteria for the final outputs that reflect the expectations of the chosen medium. These adjustments aim to make the interdisciplinary collaboration more productive and to ensure that all students, regardless of their academic background, are equally supported in tackling the complexity of their task.

Looking forward, the project team plans to expand the teaching materials and publish them as open educational resources (OER) to support broader adoption and long-term accessibility. These

micro-credentials constitute a dedicated Digital Humanities extension of the existing Data Literacy Certificate, allowing for flexible and modular certification of competencies in this field. They are being developed across the entire University Alliance Ruhr and are intended to be offered jointly by TU Dortmund University, Ruhr University Bochum, and the University of Duisburg-Essen. While algorithmic accountability is a central pillar of this offering, the broader program will also cover a wide range of digital humanities topics, including text mining, corpus linguistics, data visualization, web scraping, and the ethical handling of digital research data. Additional topics under development include Python programming for non-technical students, legal and ethical aspects of data collection, and media literacy modules.

Through these efforts, we aim to strengthen the role of algorithmic accountability in higher education and contribute to a more informed and critical digital society.

ACKNOWLEDGEMENTS

This work was supported by the *Stiftung Innovation in der Hochschullehre* through the *Freiraum 2023* funding line, <https://stiftung-hochschullehre.de/foerderung/freiraum/>.

REFERENCES

- Bhopal, K. & Myers, M. (2023). The impact of COVID-19 on A Level exams in England: Students as consumers, *British Educational Research Journal*, 49(1), 142–157. <https://doi.org/10.1002/berj.3834>
- Data Literacy Charter. <https://www.stifterverband.org/data-literacy-charter>
- Overdiek, M. & Petersen, T. (2022). *Was Deutschland über Algorithmen und Künstliche Intelligenz weiß und denkt - Ergebnisse einer repräsentativen Bevölkerungsumfrage: Update 2022*. Bertelsmann Stiftung.
- Ridsdale, C., Rothwell, J., Smit, M., Bliemel, M. Irvine, D., Kelley, D., Matwin, S., Wuetherick, B. & Ali-Hassan, H. (2015). *Strategies and Best Practices for Data Literacy Education*. Knowledge Synthesis Report, <https://doi.org/10.13140/RG.2.1.1922.5044>
- Schüller, K. (2020). *Future Skills: A Framework for Data Literacy*. Working Paper No. 53. Berlin: Hochschulforum Digitalisierung.