

**University of Cologne**  
**Faculty of Management, Economics and Social Sciences**  
**Information Systems Area**



## **Advanced Seminar on Information Systems and Digital Technology**

**Term: Summer 2021**

**Chair for Information Systems and Systems Development (Prof. Dr. Recker)**

Contact information:

Dr. Karl Werder ([werder@wiso.uni-koeln.de](mailto:werder@wiso.uni-koeln.de))

### **Responsible Artificial Intelligence: Data-driven Biases vs Algorithmic Biases**

As evidence-based decision-making aided by data-driven artificial intelligence (AI) algorithms becomes increasingly common across all sectors of the economy, there is a growing concern among users about whether such algorithms are developed and implemented in a responsible manner. Responsible AI has four aspects: fairness, accountability, transparency, and explainability (FATE). Prior reports already provide a glimpse into the disastrous effects of inaccurate and bias-laden AI recommendations in high-stakes applications, with examples from the healthcare and legal domains including incorrect patient treatment, wrongful arrest (Hill, 2020) and unjust criminal sentencing (Holzinger et al., 2019). The heightened awareness to concerns raised in recent movements for social justice has resulted in calls from professional associations (ACM U.S. Technology Policy Committee, 2020) and researchers (Coalition for Critical Technology, 2020) for developing approaches that help establish responsible AI. Motivated by these concerns, this research seminar examines how careful consideration of provenance can help enhance the quality of the data and hence the quality of the AI-generated recommendations.

Rapid innovations in data-generating technologies, such as sensors, social media, and mobile devices, have exacerbated the problems resulting from poor data quality in responsible AI systems. These technologies offer an unprecedented quantity and variety of data. While most applications have benefitted from the explosive growth in data availability (in terms of volume, variety, velocity, veracity, etc.), limited attention is paid to data quality (Meng, 2018) thereby undermining the quality of recommendations that are generated using such data. Recent studies have shown that AI algorithms may produce seemingly correct recommendations despite being based on poor data inputs (Kelly et al., 2019). For example, an algorithm used for recommending cancer treatments might learn patterns from scars, medical device implants, or marks on scans that were accidentally left by radiologists instead of learning from underlying tumor patterns. Due to unbalanced training data, AI algorithms have recommended new hires based on candidates' gender instead of their capabilities (Dastin, 2018). We argue that data provenance—a record that describes the origins and processing of data (Belhajjame et al., 2013)—can help assess the FATE of recommendations provided by AI algorithms and thus instill trust in them. Trust is enhanced by the ability to describe and follow the life of data (i.e.,

its origins, processing, and use) in both forward and backward directions (Davidson & Roy, 2017). The importance of data provenance has been long recognized (Buneman et al., 2001) in the pharmaceutical, food, and fashion industries, as it helps establish a product's origins and thus influences consumers' decisions about whether to purchase and use the product. Hence, this seminar seeks to understand the role of provenance in the relationship between data-driven and algorithmic biases toward responsible AI.

In this seminar, students will learn to identify, plan and conduct their own research project. The projects are likely to use secondary data in order to answer their developed research questions. Given the explosion of information in today's society, the ability to extract, transform and analyze data from secondary data sources is an important business skill in our knowledge society. While different types of data collection method exist, this seminar focuses on the use of secondary data for reasons of data access during later analysis.

### **Fundamentals on Scientific Work**

The students learn the fundamentals of scientific work via the Flipped Classroom on Scientific Work. A separate registration (and preparation) is necessary:

- [https://www.ilias.uni-koeln.de/ilias/goto\\_uk\\_fold\\_2445676.html](https://www.ilias.uni-koeln.de/ilias/goto_uk_fold_2445676.html)

Students are exempted if they have already attended the classroom session of the Flipped Classroom on Scientific Work in the context of another course. If this is the case, students should contact [werder@wiso.uni-koeln.de](mailto:werder@wiso.uni-koeln.de) beforehand providing the course name and semester, in which the classroom session on scientific work has been accomplished.

For more information please visit:

- <https://wirtschaftsinformatik.uni-koeln.de/en/studies/theses/scientific-work>

### **Activities**

The seminar work consists of five main phases:

1. The students acquire the basics of conducting scientific work via the Flipped Classroom.
2. The students learn the fundamentals concerning responsible AI research and secondary data collection and analysis.
3. The students plan their seminar project and develop a study protocol that is submitted and discussed.
4. The improved study protocol guides the student to collect their data and assists them in their analysis. Hence, relevant data sources are identified, data is collected and processed in order to develop a key deliverable of the seminar project.
5. The seminar project is documented in a seminar paper.

### **Timeline**

- 06. April 2021, 10:00-17:00: Classroom session on Scientific Work (not necessary if you have attended before)
- 13. April 2021, 09:00-11:00: Kick-off (Introduction to Seminar; Organization)
- 20. April 2021, 09:00-11:00: Discussing Responsible AI
- 27. April 2021, 09:00-11:00: Discussing Data Provenance
- 4. May 2021, 09:00-11:00: Discussing Data-driven and algorithmic biases

- 18. May 2021, 09:00-10:30 & 11:00-12:30 & 13:00-14:30: Study protocols: Discussions and feedback OR 15 June 2021, 09:00-16:00: Presentation and discussion of preliminary results.
- 13. July 2021, Submission of final seminar paper

NOTE: At the point of writing, I cannot say whether we are able and allowed to meet in person. The current plan facilitates both, an entire virtual experience or a hybrid mode where we meet in person in June in a sufficiently large lecture hall. I trust that we will have clarity about this as the course starts. I will keep registered students informed via ILIAS.

<b>Date</b>	<b>Video Lecture</b>	<b>Student Assignment 1</b>	<b>Student Assignment 2</b>	<b>Student Assignment 3</b>	<b>Meeting</b>
06.04	Online session on Scientific Work (not necessary if you have attended before)				Online 10:00-17:00
13.04	Kick-off; research gaps and secondary data; types of analysis; how to write a review	(Werder et al., 2021)			Online 09:00-11:00
20.04	Responsible AI: (Canca, 2020)	(Johnson, 2015)	Chapter 5 (Dignum, 2019)	(Shin & Park, 2019)	Online 09:00-11:00
27.04	Data Provenance: (Buneman & Davidson, 2013)	(Simmhan et al., 2005)	(Liang et al., 2017)	(Robinson et al., 2019)	Online 09:00-11:00
04.05	Data-driven & algorithmic biases: (Sadiq et al., 2011)	(Lambrech & Tucker, 2019)	(FRA, 2019)	(Adomavicius et al., 2019)	Online 09:00-11:00
18.05  OR 15.06	Key issues protocols	Review 3 study protocols and prepare questions			Online 09:00-10:30 & 11:00-12:30 & 13:00- 14:30
		OR Review 3 short paper and prepare questions			OR Lecture hall HSXXIV
13.07	-	Submission of final seminar thesis			EOD

## Course Grading

The course grading is threefold:

- **Paper Summary** (20%) - you are expected to write a clear and concise one-page summary of the article that has been assigned to you. In addition, you are expected to read two more papers within your topic domain, so that you can lead an online discussion. You are expected to read the summary articles or the papers of the additional topic domains within this course, so that you can participate in online discussions.
- **Study Protocol OR Short Paper** (30%) – Given the current you are expected to develop and write a study protocol (3-5 pages). You will also be assigned two study protocols/short paper of your peers that you review, so that you can lead and contribute to online discussions. In the case of short paper presentation, you are expected to develop and present your (preliminary) results (approximately 10 min).
- **Seminar paper** (50%) - departing from your initial study protocol and the feedback received on your preliminary results, you are expected to hand in a seminar research paper. This work contains (1) a clear and concise introduction that motivates the research, (2) a review of the state-of-the-literature, defining central terms, (3) document your research approach in a transparent, yet concise way, (4) present and discuss your developed results and (5) give an outlook toward future research needs.

## References:

- ACM U.S. Technology Policy Committee. (2020). *STATEMENT ON PRINCIPLES AND PREREQUISITES FOR THE DEVELOPMENT, EVALUATION AND USE OF UNBIASED FACIAL RECOGNITION TECHNOLOGIES* (pp. 1–4). ACM.
- Adomavicius, G., Bockstedt, J. C., Curley, S. P., & Zhang, J. (2019). Reducing recommender system biases: An investigation of rating display designs. *MIS Quarterly: Management Information Systems*, 43(4), 1321–1341. <https://doi.org/10.25300/MISQ/2019/13949>
- Belhajjame, K., B'Far, R., Cheney, J., Coppens, S., Cresswell, S., Gil, Y., Groth, P., Klyne, G., Lebo, T., McCusker, J., Miles, S., Myers, J., Sahoo, S., & Tilmes, C. (2013). *PROV-DM: The PROV Data Model*.
- Buneman, P., & Davidson, S. (2013). *Data provenance - the foundation of data quality*.
- Buneman, P., Khanna, S., & Wang-Chiew, T. (2001). Why and Where: A Characterization of Data Provenance. In *Lecture Notes in Computer Science* (Vol. 1973, Issue January, pp. 316–330). [https://doi.org/10.1007/3-540-44503-X\\_20](https://doi.org/10.1007/3-540-44503-X_20)
- Canca, C. (2020). Computing Ethics: Operationalizing AI ethics principles. *Communications of the ACM*, 63(12), 18–21. <https://doi.org/10.1145/3430368>
- Coalition for Critical Technology. (2020). Abolish the #TechToPrisonPipeline. *Medium.Com*, 138.
- Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scrap-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>
- Davidson, S. B., & Roy, S. (2017). Encyclopedia of Database Systems. *Encyclopedia of*

- Database Systems*, 1–6. <https://doi.org/10.1007/978-1-4899-7993-3>
- Dignum, V. (2019). Responsible Artificial Intelligence - How to Develop and Use AI in a Responsible Way. In B. O’Sullivan & M. Wooldridge (Eds.), *Artificial Intelligence: Foundations, Theory, and Algorithms*. Springer Nature Switzerland.
- FRA. (2019). Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights. *Fra – European Union Agency for Fundamental Rights*.
- Hill, K. (2020, June 24). Wrongfully Accused by an Algorithm. *New York Times*, 7.
- Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4), 1–13. <https://doi.org/10.1002/widm.1312>
- Johnson, D. G. (2015). Technology with No Human Responsibility? *Journal of Business Ethics*, 127(4), 707–715. <https://doi.org/10.1007/s10551-014-2180-1>
- Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. *BMC Medicine*, 17(1), 195. <https://doi.org/10.1186/s12916-019-1426-2>
- Lambrecht, A., & Tucker, C. (2019). Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science*, 65(7), 2966–2981. <https://doi.org/10.1287/mnsc.2018.3093>
- Liang, X., Shetty, S., Tosh, D., Kamhoua, C., Kwiat, K., & Njilla, L. (2017). ProvChain: A Blockchain-Based Data Provenance Architecture in Cloud Environment with Enhanced Privacy and Availability. *2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, 468–477. <https://doi.org/10.1109/CCGRID.2017.8>
- Meng, X.-L. (2018). Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election. *The Annals of Applied Statistics*, 12(2), 685–726. <https://doi.org/10.1214/18-AOAS1161SF>
- Robinson, J., Rosenzweig, C., Moss, A. J., & Litman, L. (2019). Tapped out or barely tapped? Recommendations for how to harness the vast and largely unused potential of the Mechanical Turk participant pool. *PLOS ONE*, 14(12), e0226394. <https://doi.org/10.1371/journal.pone.0226394>
- Sadiq, S., Yeganeh, N. K., & Indulska, M. (2011). 20 Years of Data Quality Research: Themes, Trends and Synergies. *Conferences in Research and Practice in Information Technology Series*, 115, 153–162.
- Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, 98(March), 277–284. <https://doi.org/10.1016/j.chb.2019.04.019>
- Simmhan, Y., Plale, B., & Gannon, D. (2005). *A survey of data provenance techniques*.
- Werder, K., Ramesh, B., & Zhang, R. (2021). *Establish Data Provenance for Responsible Artificial Intelligence Systems*.