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

Advanced Seminar on Information Systems and Digital Technology

Term: Summer 2023

Chair for Information Systems and Systems Development

Contact information:

Dr. Karl Werder (werder@wiso.uni-koeln.de)

Biases in Artificial Intelligence Development

Artificial intelligence (AI) systems are on the rise and often outperform highly skilled professionals in certain fields such as radiology (Hosny et al., 2018; Killock, 2020). Hence, it is not surprising that AI systems are also widely adopted by businesses (Brynjolfsson & Mitchell, 2017), such as smart brokers (Pozen & Ruane, 2019), virtual assistants (Campagna et al., 2017), and conversational agents (Diederich et al., 2022), giving rise to new entrepreneurial ventures (Blohm et al., 2022; Chalmers et al., 2021), driving the automation of organizational processes, and augmenting existing products and services (Burton et al., 2020).

However, there is increasing recognition that AI systems often make mistakes (Reardon, 2019) and may inherently be biased for several reasons (Roselli et al., 2019). For example, Amazon's recruitment tool discriminated against women applicants (Dastin, 2018). Another example is facial recognition technology, which is often used for security purposes and predictive policing. In January 2020, police arrested Robert Williams and put him into the Detroit Detention Center only to realize later that the system had mistakenly identified him as a criminal because of his ethnicity (Williams, 2020). These examples show how AI systems exhibit gender, racial, and other social biases.

It is tempting to believe that this issue can be solved by removing the variables associated with biases, such as gender and race, from the data used to train AI systems. However, recent work suggests that this is not that simple (Mehrabi et al., 2021). For example, Amazon reprogrammed the system to ignore gendered words, such as "women's" (Hao, 2019). It was soon discovered that systems were still picking up on words that were highly correlated with women (Alexander, 2022) and then using these to make recommendations. This suggests a more fundamental issue—biases may originate from any area of AI development (Bosch et al., 2021), including data quality management, design methods and process, model performance, and deployment and compliance, through the developers involved; as a result, these are incorporated into the AI system under development (Ozkaya, 2020), an observation that the mirroring hypothesis can explain (Colfer & Baldwin, 2016), suggesting that social ties correspond to technical dependencies during development. Hence, this seminar will revolve around the role of biases in AI development.

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

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 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

- Virtual: Classroom session on Scientific Work
  (not necessary if you have attended before; online materials available in ILIAS)
- 03. April 2023, 16:00-18:00: Kick-off (Introduction to Seminar; Organization) -
- 11. April 2023, 16:00-18:00: Discussing on Topic 1
- 17. April 2023, 16:00-18:00: Discussing on Topic 2
- 24. April 2023, 16:00-18:00: Discussing on Topic 3
- 19. May 2023, 09:00- 17:00: Study protocols: Discussions and feedback
- 07. July 2023, 09:00- 17:00: Seminar project: Discussions and feedback
- 16. July 2023, Submission of final seminar paper

Room:

- 411 Seminarraum S310 (Pohlighaus, EG)

NOTE: At the point of writing, I am planning that we hold these sessions in presence, as is the current plan of the faculty. However, given the unpredictable evolution of the pandemic, we may have to fall back to online sessions if infection numbers get out of hand. I will keep registered students informed via ILIAS.

Table 1 - Seminar schedule and mandatory readings

Date	**Video Lecture**	**Student Assignment 1**	**Student Assignment 2**	**Student Assignment 3**	**Meeting**
TBA	Online session on Scientific Work (not necessary if you have attended before; online materials available in ILIAS)				TBA
03.04	Kick-off; research gaps and secondary data; types of analysis; how to write a review				Seminarraum S310 16:00-18:00
11.04	AI Development (Ozkaya, 2020)	(Bosch et al., 2021)	(Giray, 2021)	(Martínez-Fernández et al., 2022)	Seminarraum S310 16:00-18:00
17.04	Social Biases	(Alexander, 2022)	(Wang & Redmiles, 2019)	(Lambrecht & Tucker, 2019)	Seminarraum S310 16:00-18:00
24.04	Technical Biases	(Mehrabi et al., 2021)	(Werder et al., 2022)	(Fu et al., 2020)	Seminarraum S310 16:00-18:00
09.05	Key issues protocols	Review 3 study protocols and prepare questions			Seminarraum S310 09:00-17:00
07.07	Key issues study	Prepare study presentation			Seminarraum S310 09:00-17:00
16.07	-	Submission of final seminar thesis			EOD

Course Grading

The course grading is threefold:

- **Paper Summary and Discussion** (10%) - you are expected to present a clear and concise summary of the article that has been assigned to you. In addition, you are expected to read the mandatory literature for each session so that you can participate in the discussions. You are expected to lead the discussion for the papers within your topic that you are not presenting.
- **Study Protocol and Discussion** (20%) – Given the current you are expected to develop and write a study protocol (3-5 pages). You are expected to develop and present your study proposal (approximately 10 min). You will also be assigned two study protocols of your peers that you review, so that you can lead and contribute to the discussions.
- **Final Presentation** (10%) – The 10-minute presentation should convey central parts of your research project such as research problem and question, method, results, and contribution to research and practice. Assessment in accordance with organization of content, oral, and overall presentation.
- **Seminar Paper** (60%) - 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:

Alexander, C. S. (2022). Text mining for bias: A recommendation letter experiment. *American Business Law Journal*, *59*(1), 5–59. https://doi.org/10.1111/ablj.12198

Blohm, I., Antretter, T., Sirén, C., Grichnik, D., & Wincent, J. (2022). It's a peoples game, isn't it?! A comparison between the investment returns of business angels and machine learning algorithms. *Entrepreneurship Theory and Practice*, *46*(4), 1054–1091. https://doi.org/10.1177/1042258720945206

Bosch, J., Olsson, H. H., & Crnkovic, I. (2021). Engineering AI systems: A research agenda. In *Artificial Intelligence Paradigms for Smart Cyber-Physical Systems* (pp. 1–19). IGI Global. https://doi.org/10.4018/978-1-7998-5101-1.ch001

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530–1534. https://doi.org/10.1126/science.aap8062

Burton, J. W., Stein, M., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making*, *33*(2), 220–239. https://doi.org/10.1002/bdm.2155

Campagna, G., Ramesh, R., Xu, S., Fischer, M., & Lam, M. S. (2017). Almond: The Architecture of an Open, Crowdsourced, Privacy-Preserving, Programmable Virtual Assistant. *Proceedings of the 26th International Conference on World Wide Web*, 341–350. https://doi.org/10.1145/3038912.3052562

Chalmers, D., MacKenzie, N. G., & Carter, S. (2021). Artificial Intelligence and Entrepreneurship: Implications for Venture Creation in the Fourth Industrial Revolution. *Entrepreneurship Theory and Practice*, *45*(5), 1028–1053. https://doi.org/10.1177/1042258720934581

Colfer, L. J., & Baldwin, C. Y. (2016). The mirroring hypothesis: theory, evidence, and exceptions. *Industrial and Corporate Change*, *25*(5), 709–738. https://doi.org/10.1093/icc/dtw027

Dastin, J. (2018, October 11). *Amazon scraps secret AI recruiting tool that showed bias against women*. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

Diederich, S., Brendel, A. B., Morana, S., & Kolbe, L. (2022). On the design of and interaction with conversational agents: An organizing and assessing review of human-computer interaction research. *Journal of the Association for Information Systems*, *23*(1), 96–138. https://doi.org/10.17705/1jais.00724

Fu, R., Huang, Y., & Singh, P. V. (2020). Artificial Intelligence and Algorithmic Bias: Source, Detection, Mitigation, and Implications. *INFORMS TutORials in Operations Research*, *16*, 39–63. https://doi.org/10.1287/educ.2020.0215

Giray, G. (2021). A software engineering perspective on engineering machine learning systems: State of the art and challenges. *Journal of Systems and Software*, *180*, 111031. https://doi.org/10.1016/j.jss.2021.111031

Hao, K. (2019, February 4). *This is how AI bias really happens—and why it's so hard to fix*. MIT Technology Review. https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/

Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. W. L. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer*, *18*(8), 500–510. https://doi.org/10.1038/s41568-018-0016-5

Killock, D. (2020). AI outperforms radiologists in mammographic screening. *Nature Reviews Clinical Oncology*, *17*(3), 134–134. https://doi.org/10.1038/s41571-020-0329-7

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

Martínez-Fernández, S., Bogner, J., Siebert, J., Trendowicz, A., Vollmer, A. M., Wagner, S., Franch, X., Oriol, M., Bogner, J., Wagner, S., Siebert, ; J, & Vollmer, A. M. (2022). Software Engineering for AI-Based Systems: A Survey; Software Engineering for AI-Based Systems: A Survey. *ACM Transactions on Software Engineering and Methodology*, *31*(2). https://doi.org/10.1145/3487043

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://doi.org/10.1145/3457607

Ozkaya, I. (2020). What is really different in engineering AI-enabled systems? *IEEE Software*, *37*(4), 3–6. https://doi.org/10.1109/MS.2020.2993662

Pozen, R. C., & Ruane, J. (2019, December 3). *What machine learning will mean for asset managers*. Harvard Business Review. https://hbr.org/2019/12/what-machine-learning-will-mean-for-asset-managers

Reardon, S. (2019). Rise of robot radiologists. *Nature*, *576*(7787), 54–58. https://doi.org/10.1038/d41586-019-03847-z

Roselli, D., Matthews, J., & Talagala, N. (2019). Managing bias in AI. *Proceedings of the World Wide Web Conference*, 539–544. https://doi.org/10.1145/3308560.3317590

Wang, Y., & Redmiles, D. (2019). Implicit gender biases in professional software development: An empirical study. *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Society*, 1–10. https://doi.org/10.1109/ICSE-SEIS.2019.00009

Werder, K., Ramesh, B., & Zhang, R. (2022). Establishing data provenance for responsible artificial intelligence systems. *ACM Transactions on Management Information Systems*, *13*(2), 1–23. https://doi.org/10.1145/3503488

Williams, R. (2020, June 24). *I was wrongfully arrested because of facial recognition. Why are police allowed to use it?* The Washington Post. https://www.washingtonpost.com/opinions/2020/06/24/i-was-wrongfully-arrested-because-facial-recognition-why-are-police-allowed-use-this-technology/