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

# Advanced Seminar on Information Systems and Digital Technology

**Term: Summer 2019**

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

*Contact information:*

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

## Artificial Intelligence and Human Stupidity

Today's abundance of data in conjunction with technological progress in machine learning and artificial intelligence (AI) has led to an entirely new data labelling industry (Murgia, 2019). Data labelling refers to the process of marking data with a specific code. These codes are used to train algorithms to correctly predict the code or label based on other input data. Market research companies expect the market for third-party data labeling solutions to increase from 150 million USD in 2018 toward 1 billion USD by 2023 (Cognilytica Research, 2019). Data-driven organizations react to the importance of and value-added through data labeling for their AI-based systems, as suggested by the acquisition of mighty AI by Uber (Soper, 2019). However, we suggest that current data labeling practices–i.e., hiring cheap labor to perform labeling tasks, e.g., through crowd-based platform (Murali, 2019)–threaten the quality of AI-based recommendations in expert systems, as the algorithms output will only be as good as the data that is provided.

Hence, the quality of data that is fed into an AI-algorithms is of utmost importance, as it largely defines whether or not the user trusts the algorithm's recommendation (W. Wang & Benbasat, 2016). Generally speaking, data quality includes many different aspects of issues; for example, completeness, accuracy, consistency, timeliness, duplication, validity, availability and provenance (Burt, Leong, & Shirrell, 2018). While data quality is a broad term that has been researched extensively (Strong, Lee, & Wang, 1997; Wand & Wang, 1996; R. Y. Wang & Strong, 1996), we suggest that in the context of data labeling understanding data provenance is increasingly important. Data provenance is one aspect of data quality (Burt et al., 2018) that describes the data process flow in order to provide credibility and trustworthiness of data (Alkhalil & Ramadan, 2017). Data provenance is defined "as a record that describes the people, institutions, entities, and activities involved in producing, influencing, or delivering a piece of data" (Belhajjame et al., 2013). Consequently, the data–and its processing–an AI algorithm relies on is made transparent. The quality of data that is used by an algorithm influences the quality of its recommendations (Stvilia, Gasser, Twidale, & Smith, 2007). Relying on incorrect recommendations can have disastrous effects, such as the incorrect treatment of a patient in healthcare (Holzinger, Langs, Denk, Zatloukal, & Müller, 2019), or the unjust sentence of a person in the legal system (FRA, 2019). Hence, we need to better understand the effect of data provenance toward the effective use of AI algorithms.

Hence, this seminar seeks to understand the effect of artificial intelligence toward user behavior, as well as the influence of data labelling and data quality.

In this seminar, students will learn to identify, plan and conduct their own research project. The projects will 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 in order assure data access for 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:

- http://www.wirtschaftsinformatik.uni-koeln.de/en/teaching/flipped-classroom-on-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 AI research in IS 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 2020, 11:00-17:00: Classroom session on Scientific Work (not necessary if you have attended before)
- 07 April 2020, 09:00-10:00: Kick-off (Introduction to Seminar; Organization)
- 14. April 2020, 09:00-11:00: Discussing AI-System Characteristics
- 21. April 2020, 09:00-11:00: Discussing Algorithm Aversion
- 28. April 2020, 09:00-11:00: Discussing Explainable AI
- 12 May 2020, 09:00-10:30 & 11:00-12:30: Study protocols: Discussions and feedback
- 7 July 2020, Submission of final seminar paper

| **Date** | **Video Lecture** | **Student Assignment 1** | **Student Assignment 2** | **Student Assignment 3** | **Online Meeting** |
|---|---|---|---|---|---|
| 06.04 | Online session on Scientific Work<br>(not necessary if you have attended before) | | | | 11:00-17:00 |
| 07.04 | Kick-off; research gaps and secondary data; types of analysis; how to write a review | | | | 09:00-11:00 |
| 14.04 | AI-System characteristics:<br>Machine Behavior (Rahwan et al., 2019) | Paper on FaccT (Lepri, Oliver, Letouzé, Pentland, & Vinck, 2018) | Paper on FaccT (Shin & Park, 2019) | Paper on FaccT (W. Wang & Benbasat, 2016) | 09:00-11:00 |
| 21.04 | Algorithm Aversion:<br>Overcoming Algorithm Aversion (Dietvorst, Simmons, & Massey, 2018) | Paper Algorithm Aversion (Castelo, Bos, & Lehmann, 2019) | Paper Algorithm Aversion (Dietvorst, Simmons, & Massey, 2015) | Paper Algorithm Aversion (Logg, Minson, & Moore, 2019) | 09:00-11:00 |
| 28.04 | Explainable AI:<br>AI Next Campaign (Defense Advanced Research Project Agency, 2018) | Paper on Explainable AI (Adadi & Berrada, 2018) | Paper on Explainable AI (Guidotti et al., 2018) | Paper on Explainable AI (Pedreschi et al., 2019) | 09:00-11:00 |
| 12.05 | Key issues protocols | Review 3 study protocols and prepare questions | | | 09:00-10:30 & 11:00-12:30 |
| 07.07 | - | 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** (30%) - you are expected to develop and write a study protocol (3-5 pages). You will also be assigned two study protocols of your peers that you review, so that you can lead and contribute to online discussions.
- **Seminar paper** (50%) - departing from your initial study protocol and the feedback received, 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:**

Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Alkhalil, A., & Ramadan, R. A. (2017). IoT Data Provenance Implementation Challenges. *Procedia Computer Science*, *109*(2014), 1134–1139. https://doi.org/10.1016/j.procs.2017.05.436

Belhajjame, K., B'Far, R., Cheney, J., Coppens, S., Cresswell, S., Gil, Y., … Tilmes, C. (2013). *PROV-DM: The PROV Data Model*. Cambridge, MA, USA.

Burt, A., Leong, B., & Shirrell, S. (2018). Beyond Explainability: A Practical Guide to Managing Risk in Machine Learning Models. In *Future of Privacy Forum*. Washington, DC, USA.

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-Dependent Algorithm Aversion. *Journal of Marketing Research*, *56*(5), 809–825. https://doi.org/10.1177/0022243719851788

Cognilytica Research. (2019). *Data Engineering, Preparation, and Labeling for AI 2019*. Washington, DC, USA.

Defense Advanced Research Project Agency. (2018). AI Next Campaign. Retrieved March 20, 2020, from Defense Advanced Research Project Agency website: https://www.darpa.mil/work-with-us/ai-next-campaign

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them.

*Management Science*, *64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

FRA. (2019). Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights. *Fra – European Union Agency for Fundamental Rights*.

Guidotti, R., Monreale, A., Ruggieri, S., Pedreschi, D., Turini, F., & Giannotti, F. (2018). *Local Rule-Based Explanations of Black Box Decision Systems*. (May). Retrieved from http://arxiv.org/abs/1805.10820

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

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., & Vinck, P. (2018). Fair, Transparent, and Accountable Algorithmic Decision-making Processes: The Premise, the Proposed Solutions, and the Open Challenges. *Philosophy and Technology*, *31*(4), 611–627. https://doi.org/10.1007/s13347-017-0279-x

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*(December 2018), 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

Murali, A. (2019). How India’s data labellers are powering the global AI race. Retrieved October 25, 2019, from FactorDaily website: https://factordaily.com/indian-data-labellers-powering-the-global-ai-race/

Murgia, M. (2019). AI’s new workforce: the data-labelling industry spreads globally. Retrieved October 25, 2019, from Financial Times website: https://www.ft.com/content/56dde36c-aa40-11e9-984c-fac8325aaa04

Pedreschi, D., Giannotti, F., Guidotti, R., Monreale, A., Ruggieri, S., & Turini, F. (2019). Meaningful Explanations of Black Box AI Decision Systems. *Proceedings of the AAAI Conference on Artificial Intelligence*, *33*, 9780–9784. https://doi.org/10.1609/aaai.v33i01.33019780

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., … Wellman, M. (2019). Machine behaviour. *Nature*, *568*(7753), 477–486. https://doi.org/10.1038/s41586-019-1138-y

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

Soper, T. (2019). Uber acquires Seattle startup Mighty AI to fuel its push into self-driving cars. Retrieved October 25, 2019, from GeekWire website: https://www.geekwire.com/2019/uber-acquires-seattle-startup-mighty-ai-fuel-push-self-driving-cars/

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data quality in context. Communications of the ACM. *Communications of the ACM*, *40*(5), 103–110. Retrieved from http://web.mit.edu/tdqm.

Stvilia, B., Gasser, L., Twidale, M. B., & Smith, L. C. (2007). A framework for information quality assessment. *Journal of the American Society for Information Science and Technology*, *58*(12), 1720–1733. https://doi.org/10.1002/asi.20652

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86–95. https://doi.org/10.1145/240455.240479

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5–34.

Wang, W., & Benbasat, I. (2016). Empirical Assessment of Alternative Designs for Enhancing Different Types of Trusting Beliefs in Online Recommendation Agents. *Journal of Management Information Systems*, *33*(3), 744–775. https://doi.org/10.1080/07421222.2016.1243949