Algorithmic Fairness

Summer 2026

LMU Munich, Faculty of Philosophy, Philosophy of Science and the Study of Religion, MCMP

Instructors:	Timo Freiesleben	Tom Sterkenburg
Email:	Timo.Freiesleben@lmu.de	Tom.Sterkenburg@lmu.de

Location: Room 021 in Ludwigsstrasse 31
Time: Fridays 10:15-11:45 pm
Material: The readings and further material can be found in the following folder:

- https://syncandshare.lrz.de/getlink/fiDq9iyPRjwTWDXDRPRUne/

Office hours: By appointment

Background literature: The following book and article will form the basis of the course. You need to consult them occasionally.

- **Fairness Book:** Barocas, S., Hardt, M., & Narayanan, A. (2023) Fairness and machine learning: Limitations and Opportunities, MIT Press, https://fairmlbook.org/.
- **SEP-F:** Hellman, D. (2021), "Algorithmic Fairness", The Stanford Encyclopedia of Philosophy (Fall 2025 Edition), Edward N. Zalta (ed.), https://plato.stanford.edu/archives/fall2025/entries/algorithmic-fairness/.

Overview: This seminar introduces students to the contemporary field of algorithmic fairness, with an emphasis on philosophical aspects. The first part will be based on the textbook by Barocas et al. (2023), Fairness and Machine Learning: Limitations and Opportunities, and cover statistical notions of fairness as well as causal approaches. The second part of the course will treat more recent topics, including performativity of prediction.

Prerequisites: Students should be open to engaging with mathematics, particularly probabilistic and causal reasoning. A background in ethics may be useful, but it's not a prerequisite.

Coursework: All students are requested to attend the seminar sessions, carefully study the reading assignments, and participate in the discussions. Dependent on the subject and the ECTS points, students must fulfill additional requirements:

- *Philosophy 9 ECTS:* Final Essay of 6000 words (± 250 words), topic: come up with topic/argument yourself. Presentation 15 minutes, topic: topic of the seminar session. Final grade (70% essay, 30% presentation and participation).
- *Reliable AI students 6 ECTS:* Final Essay of 3000 words (± 250 words), topic: come up with topic/argument yourself. Presentation 15 minutes, topic: topic of the seminar session. Final grade (70% essay, 30% presentation and participation).

Final Essay: The deadline for submitting the essay is **31.08.2026** noon. For Philosophy students: do not forget to register in the LSF for the exam (In our case there is no exam but only the essay).

If you already have an idea for your essay topic, we encourage you to briefly discuss it with us so we can help you avoid potential dead ends or topics that have already been thoroughly explored. After you submit

your paper, we will also take some time to have a short conversation with each of you to go over your work and also to give you feedback.

Presentation: Each student is expected to give a 15-minute presentation in one of the seminar sessions. The presentation should introduce the core ideas of the paper discussed in that session. Depending on the number of students, presentations may be given in groups.

In addition, you will be paired with a fellow student or group to act as a reviewer/moderator for their presentation. In this role, you will help initiate and guide the discussion following their presentation, together with us.

Use of AI tools: We believe that students should make use of all available resources to write a strong essay and successfully complete the seminar. This includes the use of AI tools in responsible ways. For example, you are welcome to use AI to: (i) learn about technical concepts covered in the course, (ii) receive critical feedback on your manuscript, (iii) improve the clarity and fluency of your language, and (iv) search for relevant scientific literature to engage with. When using AI tools, please keep their limitations in mind. AI-generated information can be incomplete, outdated, or incorrect, so it is your responsibility to carefully verify any claims, references, or explanations you receive.

While we encourage thoughtful use of AI to support your writing process, your essay must ultimately be your own work. Submitting text that has been generated by AI and copied into your paper is not acceptable! In practice, AI systems still tend to produce fairly weak academic essays.

Cool Additional Resources: If you are interested in knowing more about the basics in ML, we can recommend Stanford University's free online course on machine learning https://www.coursera.org/learn/machine-learning?#syllabus taught by Andrew Ng. The course provides an intro to regression methods, neural networks, and unsupervised techniques such as k-means in a highly accessible way.
If you are new to scientific writing, we highly recommend the free course https://www.coursera.org/learn/sciwrite? taught by Kristin Sainani. Amongst other things, it shows how to make your English writing more fluent, accessible, and efficient.
How to work with latex?: You are free to write your final essay in word or any other word processing software. However, some of you might be interested in trying out latex. Latex eases writing formulas, graphs, making citations, and general formatting - you will love it! In the folder, you find a basic example that contains all the crucial elements and makes starting latex a lot easier. It is easiest to work on overleaf.com. Register there and start a new project. There is a button to upload files. Choose the three files from the folder and click on compile. The syntax of latex is very intuitive, but you have to play around with it to see how it works. Concerning the bib file: You can copy these directly from google scholar. Open google scholar, search for the paper you are interested in and click on the quotation marks beneath the paper. Choose bibtex and copy the item into your bib file.

Topics

1. **Introduction to Algorithmic Fairness (17/04/2026)**
   - Topics: *Introduction to algorithmic fairness through real-world examples, sources of bias in ML systems, why fairness matters in automated decision-making*
   - Background Readings:
     - Zou, J., & Schiebinger, L. (2018).
     - Fairness Book, Chapter 1.
2. **Algorithmic Decision Making (24/04/2026)**
   - Topics: *Algorithmic learning, prediction vs decision, risk scores, optimal models*
   - Required Readings:
     - Fairness Book, Chapter 2 and Chapter 3 (until page 58).
3. **Public Holiday (01/05/2026)**
4. **Statistical Notions of Group Fairness (08.05.2026)**
   - Topics: *Independence, Separation, Sufficiency, impossibility results*
   - Required Readings:
     - Fairness Book, Chapter 3.
   - Background Readings:
     - Verma, S., & Rubin, J. (2018).
     - Buijsman, S. (2024).
5. **Individual and Procedural Fairness (15/05/2026)**
   - Topics: *Individual fairness, similarity-based approaches, procedural fairness, tensions between individual and group fairness*
   - Required Readings:
     - Fleisher, W. (2021).
   - Background Readings:
     - Binns, R. (2020).
6. **Justice and Fairness in Political and Moral Philosophy (22/05/2026)**
   - Topics: *Philosophical foundations of fairness, distributive justice, equality of opportunity*
   - Required Readings:
     - Binns, R. (2018).
   - Background Readings:
     - Fairness Book, Chapter 4.
7. **Causal and Dynamical Fairness (29/05/2026)**
   - Topics: *Causal reasoning in fairness, structural causal models, counterfactual fairness, dynamical modeling*
   - Required Readings:
     - Tolbert, A. (2025).
   - Background Readings:

  - Fairness Book, Chapter 5.
  - Schwöbel, P., & Remmers, P. (2022).

8. **The Ontological Status of Sensitive Attributes (05/06/2026)**
   - Topics: *What are sensitive attributes?, the social construction of categories such as race and gender, problems with operationalizing social categories in ML*
   - Required Readings:
     - Hu, L. (2023).
   - Background Readings:
     - Doh, M., Höltgen, B., Riccio, P., & Oliver, N. M. (2025).
     - Hu, L., & Kohler-Hausmann, I. (2020).
     - Fairness book, p.147–148.

9. **Multi-Group and Intersectional Fairness (12/06/2026)**
   - Topics: *Intersectionality, fairness across multiple groups, fairness under underrepresentation*
   - Required Readings:
     - Stewart, R. T. (2022).
   - Background Readings:
     - Wang, A., Ramaswamy, V. V., & Russakovsky, O. (2022).

10. **Model Multiplicity (19/06/2026)**
    - Topics: *Multiplicity of valid predictive models, fairness implications, transparency and accountability issues*
    - Required Readings:
      - Black, E., Raghavan, M., & Barocas, S. (2022).

11. **Algorithmic Recourse (26/06/2026)**
    - Topics: *Recourse and actionable explanations, fairness of recourse, ethical considerations*
    - Required Readings:
      - Venkatasubramanian, S., & Alfano, M. (2020).
    - Background Readings:
      - Sullivan, E., & Verreault-Julien, P. (2022).
      - Von Kügelgen, J., et al. (2022).

12. **Algorithmic Contestability (03/07/2026)**
    - Topics: *Contesting algorithmic decisions, XAI explanations, legal role of fairness*
    - Required Readings:
      - Freiesleben, T., Meding, K., & König, G. (2026).

13. **Prospective Fairness and Performativity (10/07/2026)**
    - Topics: *Feedback loops, performativity in prediction systems, long-term fairness, fairness over time*
    - Required Readings:
      - Khosrowi, D., Ahlers, M., & van Basshuysen, P. (2025).

- Background Readings:
  - Zezulka, S., & Genin, K. (2024).

14. **Algorithmic Monoculture and Homogenization (17/07/2026)**
    - Topics: *Algorithmic monoculture, systemic risk, social welfare implications of widespread algorithmic systems*
    - Required Readings:
      - Creel, K., & Hellman, D. (2022).
    - Background Readings:
      - Kleinberg, J., & Raghavan, M. (2021).

Literature

The following list contains all required and background readings of the seminar:

- Binns, R. (2018, January). Fairness in machine learning: Lessons from political philosophy. In Conference on fairness, accountability and transparency (pp. 149-159). PMLR.
- Binns, R. (2020, January). On the apparent conflict between individual and group fairness. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 514-524).
- Black, E., Raghavan, M., & Barocas, S. (2022, June). Model multiplicity: Opportunities, concerns, and solutions. In Proceedings of the 2022 ACM conference on fairness, accountability, and transparency (pp. 850-863).
- Buijsman, S. (2024). Navigating fairness measures and trade-offs. AI and Ethics, 4(4), 1323-1334.
- Creel, K., & Hellman, D. (2022). The algorithmic leviathan: Arbitrariness, fairness, and opportunity in algorithmic decision-making systems. Canadian Journal of Philosophy, 52(1), 26-43.
- Decker, M. C., Wegner, L., & Leicht-Scholten, C. (2025). Procedural fairness in algorithmic decision-making: the role of public engagement. Ethics and Information Technology, 27(1), 1.
- Doh, M., Höltgen, B., Riccio, P., & Oliver, N. M. (2025) Position: The Categorization of Race in ML is a Flawed Premise. In Forty-second International Conference on Machine Learning Position Paper Track.
- Fleisher, W. (2021, July). What's fair about individual fairness?. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (pp. 480-490).
- Green, B. (2022). Escaping the impossibility of fairness: From formal to substantive algorithmic fairness. Philosophy & Technology, 35(4), 90.
- Hu, L. (2023). What is "race" in algorithmic discrimination on the basis of race?. Journal of Moral Philosophy, 21(1-2), 1-26.
- Hu, L., & Kohler-Hausmann, I. (2020, January). What's sex got to do with machine learning?. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 513-513).
- Kleinberg, J., & Raghavan, M. (2021). Algorithmic monoculture and social welfare. Proceedings of the National Academy of Sciences, 118(22), e2018340118.

- Khosrowi, D., Ahlers, M., & van Basshuysen, P. (2025, June). When Predictions are More Than Predictions: Self-Fulfilling Performativity and the Road Towards Morally Responsible Predictive Systems. In Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (pp. 1108-1118).
- Schwöbel, P., & Remmers, P. (2022) The long arc of fairness: Formalisations and ethical discourse. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 2179-2188).
- Stewart, R. T. (2022). Identity and the limits of fair assessment. Journal of Theoretical Politics, 34(3), 415-442.
- Sullivan, E.,& Verreault-Julien, P. (2022). From explanation to recommendation: Ethical standards for algorithmic recourse. In Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (pp. 712-722).
- Tolbert, A. (2025). Why causal inference is necessary for algorithmic fairness. Synthese, 206(3), 162.
- Venkatasubramanian, S., & Alfano, M. (2020, January). The philosophical basis of algorithmic recourse. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 284-293).
- Verma, S., & Rubin, J. (2018, May). Fairness definitions explained. In Proceedings of the international workshop on software fairness (pp. 1-7).
- Von Kügelgen, J., Karimi, A. H., Bhatt, U., Valera, I., Weller, A., & Schölkopf, B. (2022, June). On the fairness of causal algorithmic recourse. In Proceedings of the AAAI conference on artificial intelligence (Vol. 36, No. 9, pp. 9584-9594).
- Wang, A., Ramaswamy, V. V., & Russakovsky, O. (2022, June). Towards intersectionality in machine learning: Including more identities, handling underrepresentation, and performing evaluation. In Proceedings of the 2022 ACM conference on fairness, accountability, and transparency (pp. 336-349).
- Zezulka, S., & Genin, K. (2024). From the fair distribution of predictions to the fair distribution of social goods: Evaluating the impact of fair machine learning on long-term unemployment. In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (pp. 1984-2006).
- Zou, J., & Schiebinger, L. (2018). "AI can be sexist and racist—it's time to make it fair"