



Who's At the Table?

Bias, Positionality, and Pedagogical Power

Robin Sutherland-Harris,
Education Developer, Teaching Commons

DEC. 2, 2025

YORK 

“
Unpack how AI systems reflect and reproduce existing social biases, and how our positionality as educators shapes the way we introduce or resist AI in the classroom.
”



Who's At the Table? Bias, Positionality, and Pedagogical Power

- State of the field
- Exploring our positionality
- Practical strategies

Overview

What do you think?

Is AI more or less biased than humans?

State of the Field

YORK 



How does GenAI work?

GenAI...

Identifies patterns:

- Not random! Identifies & selects **statistically probable** outputs
- No particular adherence to the truth – can “hallucinate”

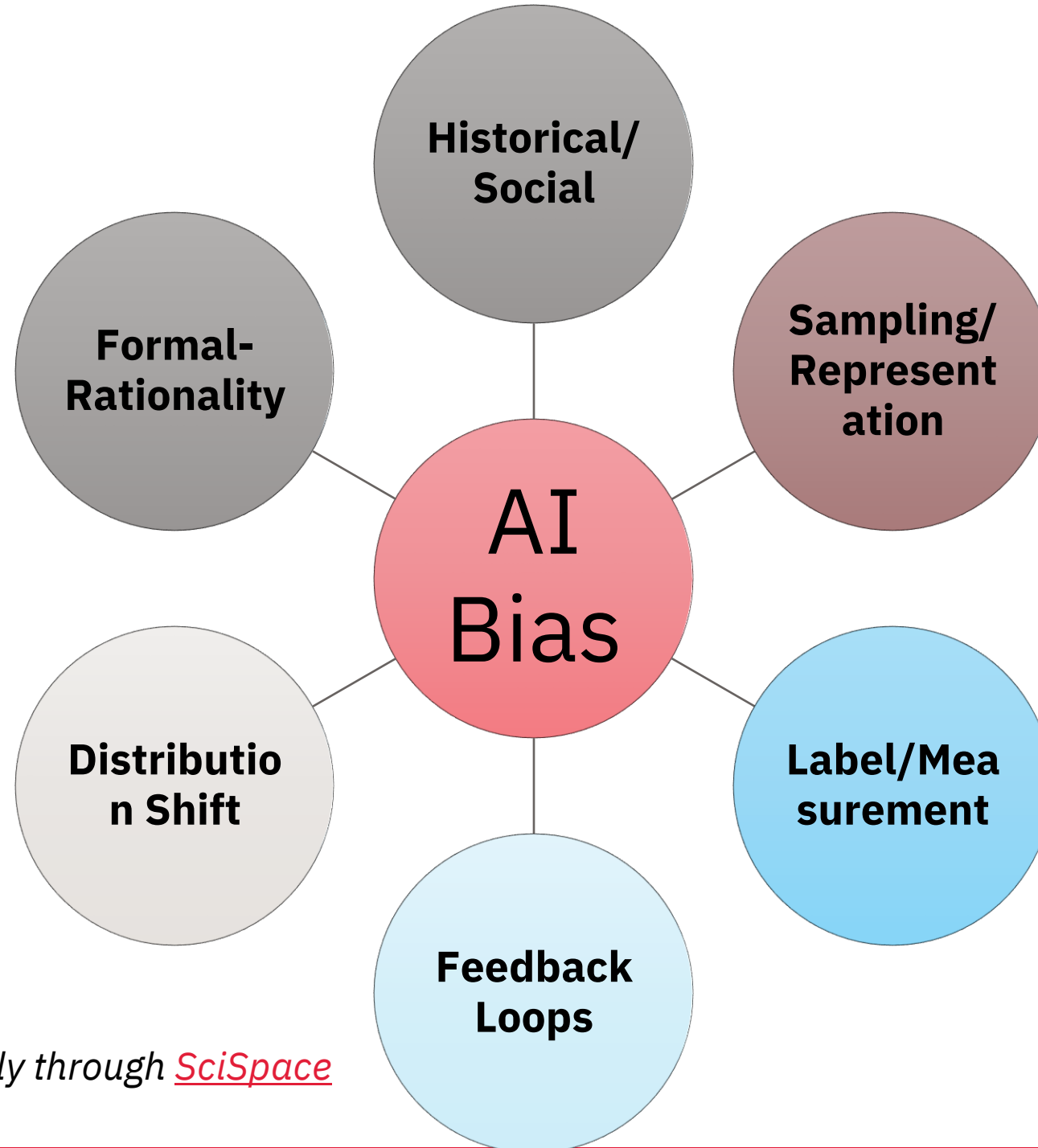
Based on training data:

- Decides what is statistically probable based on training data
- Outputs therefore reflect what is predominant in training data

To create original material:

- Technically original, but not truly innovative
- The more you know about a topic, the more you can see where GenAI misses the mark
- More accurate AI has had more/different training, but still operates on a basis of statistical probability

Baked-In Bias



Lit review conducted partly through [SciSpace](#)

Discussion

So... ***is*** AI more biased than humans?

- Yes, no, and maybe
- AI can...
 - Amplify human bias
 - Reduce some human inconsistencies
- Human oversight can mitigate or compound algorithmic & other biases

Emergent Bias

› Ontological bias

- Read more:
 - Garrison, K. G. (2025, July 28). *To explore AI bias, researchers pose a question: How do you imagine a tree?* Stanford Report. <https://news.stanford.edu/stories/2025/07/ai-llm-ontological-systems-bias-research>
 - Haghighi, N., Yu, S., Landay, J. A., & Rosner, D. (2025). Ontologies in Design: How Imagining a Tree Reveals Possibilities and Assumptions in Large Language Models. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–20. <https://doi.org/10.1145/3706598.3713633>

› Position bias

- Read more:
 - Zewe, A. (2025, June 17). *Unpacking the bias of large language models*. MIT News | Massachusetts Institute of Technology. <https://news.mit.edu/2025/unpacking-large-language-model-bias-0617>
 - Wu, X., Wang, Y., Jegelka, S., & Jadbabaie, A. (2025). *On the Emergence of Position Bias in Transformers* (No. arXiv:2502.01951). arXiv. <https://doi.org/10.48550/arXiv.2502.01951>

› AI-AI bias

- Read more:
 - Laurito, W., Davis, B., Grietzer, P., Gavenčiak, T., Böhm, A., & Kulveit, J. (2025). AI–AI bias: Large language models favor communications generated by large language models. *Proceedings of the National Academy of Sciences*, 122(31), e2415697122. <https://doi.org/10.1073/pnas.2415697122>

Exploring Our Positionality

YORK 



Where do I stand in this AI landscape?

- Take a few minutes to reflect on one (or more) of the questions below. Feel free to jot notes, write paragraphs, or just ruminate.
- **What parts of your identity (e.g., race, gender, language background, disability, class, immigration history) feel most relevant to how you encounter AI in your own life?**
- **How does your role (tenure status, contract type, class size, discipline) shape the risks or benefits you feel when you consider using AI in your teaching?**
- **Are there student groups in your context for whom AI bias feels especially dangerous? How does that awareness show up (or not show up) in your course decisions?**

Practical Strategies for Addressing AI Bias



Exploring, analysing, & mitigating AI bias

1. “Bias Audit” Assignments
2. Critical Analysis of AI Outputs
3. Embedding Positionality & Data-Justice Questions
4. Algorithmic Literacy Mini-Modules (in non-CS courses)
5. Scaffolded Responsible Use Protocols

Exploring, analysing, & mitigating AI bias

1. “Bias Audit” Assignments (for AI Tools)

- Students select a system (search engine, recommender, LLM, image generator) and design tests to probe its behaviour across different identities or contexts.
- They classify findings using a bias taxonomy.
 - *What kind of biases occur in this tool (language, exclusion, association, etc.)?*
- Optional: Students then cross-check outputs with trusted sources and reflect on their verification strategy.

Exploring, analysing, & mitigating AI bias

2. Critical Analysis of AI Outputs

- Instructor generates or collects AI outputs.
- Students annotate for bias.
 - *Whose perspectives are present/absent, stereotypes, language patterns, framing?*
- Require a short written critique or class discussion.
 - *Who is harmed? Who is helped? What context or voices are missing?*
- Optional: Students attempt mitigation by altering prompts, adding constraints, switching tools, then re-evaluate.

Exploring, analysing, & mitigating AI bias

3. Embedding Positionality & Data-Justice Questions

- Short reflective prompts.
 - *E.g. “Where in your life are you subject to algorithmic decisions? How do race, gender, disability, or immigration status change those experiences?”*
- Small group activities mapping stakeholders in a biased system (e.g., predictive policing, proctoring software) and tracing who bears the risk vs. who benefits.
- Pairing technical explanation (how the model works) with social analysis (what problem definition and data choices say about whose lives are valued).

Exploring, analysing, & mitigating AI bias

4. Algorithmic Literacy Mini-Modules (in non-CS courses)

- A shorter block inside a course that:
 - defines algorithms & algorithmic power;
 - introduces everyday examples relevant to the discipline;
 - walks through one structured activity exploring bias (e.g., Do algorithms know me better than I know myself?)
 - ends with a reflection (written or verbal) on implications for the discipline.

Exploring, analysing, & mitigating AI bias

5. Scaffolded Responsible Use Protocols

- Broadly, *Generate* → *Question* → *Verify* → *Disclose*. Some examples:
 - Human-in-the-loop verification, reflection logs, and periodic AI-free assessments to prevent uncritical reliance, OR
 - AI literacy and bias education as part of collaboratively setting course expectations and evaluating tools, OR
 - Mnemonic-based checklists for evaluating AI output.

Exploring, analysing, & mitigating AI bias – Three resources for faculty

- Samokishyn, M. (2023). *Algorithmic Awareness Toolkit: Teaching Algorithmic Literacy in Academic Libraries and Beyond*. BCCampus.
 - A guide for instructors with lessons and activities
- *Bias in AI*. (2024, 2025). Teaching Writing Across the Curriculum with AI.
 - A curated list of pedagogical strategies for engaging students with AI bias.
- *AI and Critical Thinking in Education*. Teaching and Learning | Western Michigan University. (2025).
 - Includes several classroom strategies for preventing AI-related decline in critical thinking

Exploring, analysing, & mitigating AI bias – Two resources for students

- Marupova, R. (n.d.). *Library Guides: Everyday AI: Best Practices for Using AI: AI and Algorithmic Bias*. Iowa State University.
<https://instr.iastate.libguides.com/c.php?g=1483633&p=11077604>
- McKone, L. (2025, September 29). Bias in AI: How to Spot It, Why It Matters, and What You Can Do. *Johns Hopkins Sheridan Libraries*.
<https://www.library.jhu.edu/bias-in-ai-how-to-spot-it-why-it-matters-and-what-you-can-do/>



What Next?

If you imagine your discipline or your classroom 10 years from now, what would it look like to say, “We took AI seriously *and* we took bias and justice seriously”?

References

- Belenguer, L. (2022). AI bias: Exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry. *AI and Ethics*, 2(4), 771–787.
<https://doi.org/10.1007/s43681-022-00138-8>
- Borgesius, F. Z. (2018). *Discrimination, artificial intelligence, and algorithmic decision-making*.
- Ferrara, E. (2024). The butterfly effect in artificial intelligence systems: Implications for AI bias and fairness. *Machine Learning with Applications*, 15, 100525. <https://doi.org/10.1016/j.mlwa.2024.100525>
- Khakurel, U., Abdelmoumin, G., Bajracharya, A., & Rawat, D. B. (2022). Exploring bias and fairness in artificial intelligence and machine learning algorithms. *Proceedings of SPIE* (Vol. 12113, 1211324). <https://doi.org/10.1117/12.2621282>
- Nazer, L., et al. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digital Health*, 2(6), e0000278. <https://doi.org/10.1371/journal.pdig.0000278>
- Nishant, R., & Schneckenberg, D. (2023). The formal rationality of artificial intelligence-based algorithms and the problem of bias. *Journal of Information Technology*. <https://doi.org/10.1177/02683962231176842>
- Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2021). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *arXiv*.
- Varona, D., & Suárez, J. L. (2022). Discrimination, bias, fairness, and trustworthy AI. *Applied Sciences*, 12(12), 5826. <https://doi.org/10.3390/app12125826>
- Wang, R., Harper, F. M., & Zhu, H. (2020). Factors influencing perceived fairness in algorithmic decision-making: Algorithm outcomes, development procedures, and individual differences. *arXiv*.