

PHIL 3350-001 Spring 2026 Hist & Phil Science

[PHIL 3350-001 History and Philosophy of Science]

Spring Semester 2026
Monday, Wednesday and Friday 10:45-11:35 AM
Marriot Library 1715

Course Description

For most of the twentieth century, philosophers of science and researchers in artificial intelligence worked on similar problems, kept up with each other's work, and took frequent inspiration from each other. The founding generations of AI and machine learning all had a working familiarity with core issues in philosophy of science. Although those days are behind us, the philosophical problems have not gone away. Talk of probabilities is everywhere, but their interpretation is often unclear. Appeals to simplicity are commonplace, but a clear justification is absent. The importance of values in data-scientific practice is broadly acknowledged, but their scope and bearing remain controversial. These are all perennial issues in the philosophy of science, and philosophers have developed a lot of resources for dealing with them. The premise of this course is that these resources can be fruitfully imported into machine learning. This course aims to give the student a familiarity with core issues in the philosophy of science, with an emphasis on their relevance to machine learning. It will be organized largely around what might be called the ur-problem for both fields: the problem of induction or, what, if not deductive validity, justifies inferences that go beyond the data that we have collected?

Course Outcomes and Objectives

- To introduce core issues in the philosophy of science.
- To illustrate the bearing of philosophy of science on contemporary statistics, artificial intelligence and machine learning.
- To explore how machine learning reconfigures core debates in the philosophy of science.

- To learn how to formulate and defend philosophical arguments in writing.

Course Requirements

There will be reading assigned for every class meeting. I will generally provide you with the materials, but if there is some difficulty please make an effort to find the reading yourself. The readings may be challenging—do your best to keep up. Reading materials will be posted on Canvas. You must attempt this reading before the meeting of the class. You are expected to contribute to class discussion – I will not be lecturing the entire time. For every class meeting, you are required to post on the class spreadsheet one discussion, or comprehension question about the reading. It can be a genuine request for clarification on some point, or a broad open-ended question for class discussion. This will be worth 10% of your final grade. Please submit these no later than 8AM the morning of the class that the reading is due. These are meant to give me a sense of how you are dealing with the material, and to give you a sense of what your classmates are thinking about the readings.

There will be two writing assignments. The first, which will be due on Monday, March 23rd, will be a comparison paper where you discuss the perspectives of 2 authors (or, more generally, two philosophical schools), comparing and contrasting their views on some issue. It should be about 800 words (about 2-3 pages, double-spaced). The second, which will be due on April 29th, will be a longer paper in which you make your own argument concerning a topic we have discussed in class. It should be 1400-1600 words (about 5-6 pages, double-spaced). A rough outline of this final paper (1 page) will be due April 13th, approximately two weeks before the final draft. The outline itself will not be graded, but the grade of the final paper will be reduced if the outline is not received in time to give feedback.

- Attendance and participation – 15%
- Reading responses – 15%
- First essay: 30%
- Final essay: 40%

Required and Recommended Readings

See detailed syllabus.

Grading Scale

Students will receive numeric grades which will be converted to the usual letter scale. No “curve” will be imposed on the grades.

University Policies

Updated mandatory syllabus policies regarding the ADA Act, Safety at the U, Addressing Sexual Misconduct, and Academic Misconduct can be viewed at:

<https://cte.utah.edu/instructor-education/syllabus/institutional-policies.php>.

Reading Schedule

Week 1: Values in (Data) Science.

Wednesday, January 7:

Rudner, Richard. "The scientist qua scientist makes value judgments." *Philosophy of Science* 20.1 (1953): 1-6.

<https://www.jstor.org/stable/185617>.

Douglas, Heather. "Inductive risk and values in science." *Philosophy of Science* 67.4 (2000): 559-579

<https://www.jstor.org/stable/188707>.

Friday, January 9:

Bright, Liam Kofi. "Du Bois' democratic defence of the value free ideal." *Synthese* 195.5 (2018): 2227-2245.

<https://link.springer.com/content/pdf/10.1007/s11229-017-1333-z.pdf>.

Week 2: Algorithmic Justice.

Monday, January 12:

Vredenburg, Kate. "Fairness". In: Justin B. Bullock, and others (eds.), *The Oxford Handbook of AI Governance*, Oxford Handbooks (2024; online edn, Oxford Academic, 2022).

<https://academic.oup.com/edited-volume/41989/chapter/374514681>.

Wednesday, January 14:

No class

Friday, January 16:

Kuppler, Matthias, et al. (2022) "From fair predictions to just decisions? Conceptualizing algorithmic fairness and distributive justice in the context of data-driven decision-making." *Frontiers in Sociology* 7.

<https://www.frontiersin.org/journals/sociology/articles/10.3389/fsoc.2022.883999/full>.

Week 3: Interpretations of Probability.

Monday, January 19:

Hájek, Alan (2023), "Interpretations of Probability", *The Stanford Encyclopedia of Philosophy*.
Sections 1-3.1 (Classical Probability)
<https://plato.stanford.edu/entries/probability-interpret/>.

Wednesday, January 21:

Hájek, Alan (2023), "Interpretations of Probability", *The Stanford Encyclopedia of Philosophy*.
Sections 3.2-3.3 (Logical, Evidential and Subjective Probability)
<https://plato.stanford.edu/entries/probability-interpret/#LogPro>.

Friday, January 23:

Hájek, Alan (2023), "Interpretations of Probability", *The Stanford Encyclopedia of Philosophy*. Sections 3.4-6 (Frequentist and Propensity-Theoretic Probability)
<https://plato.stanford.edu/entries/probability-interpret/#FreInt>.

Cynthia Dwork (2022). Fairness, Randomness and the Crystal Ball.



https://youtu.be/8s_hoWyGD4w?si=DbVy5bZkpVWlyBkk.

Week 4: The Problem of Induction.

Monday, January 26:

David Hume (1748/1777). *An Enquiry Concerning Human Understanding*. Sections I-IV.
<https://www.earlymoderntexts.com/assets/pdfs/hume1748.pdf>.

Wednesday, January 28:

David Hume (1748/1777). *An Enquiry Concerning Human Understanding*. Sections V-VII.
<https://www.earlymoderntexts.com/assets/pdfs/hume1748.pdf>.

Friday, January 30:

Leah Henderson (2022). "The Problem of Induction" *The Stanford Encyclopedia of Philosophy*.
Sections 1-3
<https://plato.stanford.edu/entries/induction-problem/>.

Week 5: Statistical Learning Theory, or: Machine Learning's Answer to Hume

Monday, February 2:

Leah Henderson (2022). "The Problem of Induction" *The Stanford Encyclopedia of Philosophy*.
Sections 4-6

<https://plato.stanford.edu/entries/induction-problem/#TackSecoHornHumeDile>.

Wednesday, February 4:

Von Luxburg, Ulrike, and Bernhard Schölkopf. "Statistical learning theory: Models, concepts, and results."
Sections 1-5.6

<https://www.tml.cs.uni-tuebingen.de/team/luxburg/publications/StatisticalLearningTheory.pdf>.

Friday, February 6:

Sterkenburg, Tom F., and Peter D. Grünwald. "The no-free-lunch theorems of supervised learning."
Synthese 199.3 (2021): 9979-10015.

<https://arxiv.org/abs/2202.04513>.

Week 6: Confirmation-Theoretic Responses to Hume

Monday, February 9:

Hempel, Carl G. (1945) "Studies in the Logic of Confirmation (I.)." *Mind* 54.213: 1-26.

<https://www.jstor.org/stable/2250886>.

Wednesday, February 11:

Carnap, Rudolf. (1945) "On Inductive Logic." *Philosophy of Science* 12.2: 72-97

<https://www.jstor.org/stable/184280>.

Friday, February 13:

Sprenger, Jan, and Stephan Hartmann. (2019) "Variation 2: Confirmation and Induction". In: *Bayesian Philosophy of Science*. Oxford University Press.

https://www.laeuferpaar.de/Papers/BookFrame_v1.pdf.

Week 7: Falsification-Theoretic Responses to Hume

Monday, February 16:

Karl Popper. (1934) *The Logic of Scientific Discovery*. "Part I: Introduction to the Logic of Science."
<https://philotextes.info/spip/IMG/pdf/popper-logic-scientific-discovery.pdf>.

Wednesday, February 18:

Mayo, Deborah G., and Aris Spanos (2006) "Severe testing as a basic concept in a Neyman-Pearson philosophy of induction." *The British Journal for the Philosophy of Science*. 57(2): 323-357.
Sections 1-3

<https://www.jstor.org/stable/pdf/3873470.pdf>.

Friday, February 20:

Mayo, Deborah G., and Aris Spanos (2006) "Severe testing as a basic concept in a Neyman-Pearson philosophy of induction." *The British Journal for the Philosophy of Science*. 57(2): 323-357.
Sections 4-8 (especially 4-5).

<https://www.jstor.org/stable/pdf/3873470.pdf>.

Week 8: Learning-Theoretic Responses to Hume.**Monday, February 23**

Lin, Hanti (2025) "Convergence to the Truth" in *Blackwell Companion to Epistemology*. Edited by Sylvan, K., Sosa, E, Dancy, J. and Steup, M. Wiley Blackwell.

<https://arxiv.org/pdf/2410.11399>.

Wednesday, February 25

Steel, D. (2010). What if the principle of induction is normative? Formal learning theory and Hume's problem. *International Studies in the Philosophy of Science*, 24(2), 171-185.

<https://www.tandfonline.com/doi/pdf/10.1080/02698595.2010.484544>.

Friday, February 27

Kelly, Kevin T. "Simplicity, truth, and the unending game of science." *Foundations of the formal sciences V: Infinite games* 11 (2007): 223.

<https://www.andrew.cmu.edu/user/kk3n/ockham/bonn8.pdf>.

Week 9: Realism and Anti-Realism**Monday, March 2**

Chakravartty, Anjan (2017) "Scientific Realism", *The Stanford Encyclopedia of Philosophy*

<https://plato.stanford.edu/entries/scientific-realism/>.

Wednesday, March 4

Breiman, Leo. "Statistical modeling: The two cultures (with comments and a rejoinder by the author)." *Statistical science* 16.3 (2001): 199-231.

<https://www2.math.uu.se/~thulin/mm/breiman.pdf>.

Friday, March 6

Rowbottom, D. P., Peden, W., & Curtis-Trudel, A. (2024). "Does the no miracles argument apply to AI?" *Synthese*, 203(5).

<https://link.springer.com/article/10.1007/s11229-024-04524-z>.

Week 10: Spring Break

No Class

Week 11: Performativity

Monday, March 16

Van Basshuysen, Philippe, Lucie White, Donal Khosrowi, and Mathias Frisch. "Three Ways in Which Pandemic Models Perform a Pandemic." *Erasmus Journal for Philosophy and Economics* 14, no. 1 (2021): 110-127.

<https://doi.org/10.23941/ejpe.v14i1.582>.

Wednesday, March 18

Northcott, R. (2022). Reflexivity and fragility. *European Journal for Philosophy of Science*, 12(3), 43.

https://pmc.ncbi.nlm.nih.gov/articles/PMC9244509/pdf/13194_2022_Article_474.pdf.

Friday, March 20

Hacking, Ian. "Making Up People." *London Review of Books* 28, no. 16 (August 17, 2006).

<https://www.lrb.co.uk/the-paper/v28/n16/ian-hacking/making-up-people>.

Week 12: Measurement

Monday, March 23

Chang, Hasok and Nancy Cartwright. "Measurement". In: S. Psillos, and M. Curd (eds), *The Routledge Companion to Philosophy of Science*. New York: Routledge. (2008): 367–375.

https://home.iitk.ac.in/~avrs/PHI446/Books/Psillos_Routledge%20Companion%20to%20Philosophy%20of%20Science_041535403X.pdf.

Essay Due

Wednesday, March 25

Tal, Eran. "Target specification bias, counterfactual prediction, and algorithmic fairness in healthcare." *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*. (2023).

<https://arxiv.org/pdf/2308.02081>.

Friday, March 27

No class.

Week 13: Scientific Explanation

Monday, March 30

Wesley Salmon (1992) "Scientific Explanation." in Salmon, W., Earman, J, et al. *Introduction to the Philosophy of Science*. Hackett.

https://fitelson.org/confirmation/salmon_introduction_to_philosophy_of_science.pdf.

Wednesday, April 1

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature machine intelligence*, 1(5), 206-215.

<https://pmc.ncbi.nlm.nih.gov/articles/PMC9122117/pdf/nihms-1058031.pdf>.

Friday, April 3

No class. I will be presenting at the One-U Responsible AI Initiative workshop. You are welcome to drop by!

<https://rai.utah.edu/ai-ethics-workshop/>.

Week 14: Deep Learning

Monday, April 6

Milliere, R., & Buckner, C. (2024). A Philosophical Introduction to Language Models - Part I: Continuity With Classic Debates. ArXiv, abs/2401.03910.

<https://arxiv.org/abs/2401.03910>.

Optional:

Levinstein, Ben. (2023) "A Conceptual Guide to Transformers Parts I-V"

<https://benlevinstein.substack.com/p/a-conceptual-guide-to-transformers>.

Wednesday, April 8

Milliere, R., & Buckner, C. (2024). A Philosophical Introduction to Language Models - Part I: Continuity With Classic Debates.

<https://arxiv.org/abs/2401.03910>.

Optional:

Levinstein, Ben. (2023) "A Conceptual Guide to Transformers Parts I-V"

<https://benlevinstein.substack.com/p/a-conceptual-guide-to-transformers>.

Friday, April 10

Millière, R., & Buckner, C. (2024). A philosophical introduction to language models - Part II: The way forward.

<https://arxiv.org/abs/2405.03207>.

Week 15: Deep Learning

Monday, April 13

Millière, R., & Buckner, C. (2024). A philosophical introduction to language models - Part II: The way

forward.

<https://arxiv.org/abs/2405.03207>.

Wednesday, April 15

Levinstein, Benjamin A., and Daniel A. Herrmann. (2024) "Still no lie detector for language models: Probing empirical and conceptual roadblocks." *Philosophical Studies*.

<https://link.springer.com/article/10.1007/s11098-023-02094-3>.

Friday, April 17

Harding, Jacqueline. (2023) "Operationalising Representation in Natural Language Processing" *British Journal for the Philosophy of Science*. <https://www.journals.uchicago.edu/doi/10.1086/728685>.

Week 16: Final Meeting.

Finals Week: Final essay due: April 29.