PHIL 305 Philosophy of Machine Learning

Kino Zhao (she/her) kino_zhao@sfu.ca

General Information

Course Description

In this class we will survey the literature on computation as analogous to, constitutive of, or even identical to, leaning and cognition by the human mind. Starting with early 20th century analysis of computation as solvability and ending with modern-day machine learning algorithms, we will discuss topics such as the Turing test, the computation theory of mind, and algorithmic fairness.

Reading

All text are available online.

Assessment

Weekly Perusall (30%):

Weekly reading comprehension questions. Graded on accuracy. 3 lowest grades will be dropped.

In-class writing (20%):

Students are expected to have done the reading before class. At some point during the class, students will complete an in-class writing assignment. There's no make-up. 3 lowest grades will be dropped. Open notes.

Term paper (50%):

8-12 pages double space on a topic of students' choosing. Students can submit an optional draft to receive written feedback. No written feedback is provided for the final person. Verbal feedback is always possible during office hours or appointments.
No late work is accepted for the draft. For the final version, late penalty is: 5% (for 1 day late), 15% (2 days), 30% (3 days), 50% (4 days), 100% (5 days).

Schedule

Week 1: introduction

- Please watch: A video explanation on Turing machines and background context https://www.youtube.com/watch?v=PLVCscCY4xI
- Optional: Extract from *Fundamentals of Logic and Computation* by Zhe Hou
- Optional: short history of digital computers https://youtu.be/fbdafQBUZ-0?si=3lB3VrySyoy4y2bE
- Optional: videos on the "ENIAC women". https://www.youtube.com/watch?v=Zevt2blQyVs&t=380s Interview with Jean Bartik (ENIAC programmer) https://www.youtube.com/watch?v=aPweFhhXFvY

Unit 1: what is computation?

Week 2: the Church-Turing thesis

- Copeland, B. J., & Shagrir, O. (2018). The Church-Turing thesis: Logical limit or breachable barrier? Communications of the ACM, 62(1), 66–74. https://doi.org/10.1145/3198448

Week 3: is the Church-Turing thesis true?

- Cleland, C. E. (1993). Is the Church-Turing thesis true? *Minds and Machines*, 3(3), 283–312. https://doi.org/10.1007/BF00976283

Unit 2: assessing machine intelligence

Week 4: the Turing test

- Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, 59(236), 433–460.

Week 5: challenges to the Turing Test

- Gunderson, K. (1964). The Imitation Game. *Mind*, LXXIII(290), 234–245. https://doi.org/10.1093/mind/LXXIII.290.234

Week 6: do large language models communicate?

- Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 5185–5198.

Week 7: can intention save us?

- Attah, N. O. (2025). Do language models lack communicative intentions? Synthese, 205(5), 187. https://doi.org/10.1007/s11229-025-05022-6

Unit 3: the opacity problem

Week 8: the opacity problem in ML

- Creel, K. A. (2020). Transparency in Complex Computational Systems. Philosophy of Science, 87(4), 568–589. https://doi.org/10.1086/709729

Week 9: right to explanation

- Vredenburgh, K. (2022). The Right to Explanation. Journal of Political Philosophy, 30(2), 209–229. https://doi.org/10.1111/jopp.12262

Week 10: is explanation feasible?

- Grote, T., & Paulo, N. (2025). A Minimalist Account of the Right To Explanation. Philosophy & Technology, 38(2), 55. https://doi.org/10.1007/s13347-025-00888-3

Unit 4:the bias problem

Week 11: machine bias

- Hedden, B. (2021). On statistical criteria of algorithmic fairness. Philosophy & Public Affairs, 49(2), 209–231. https://doi.org/10.1111/papa.12189

Week 12: democracy as a solution

- Wong, P.-H. (2020). Democratizing Algorithmic Fairness. Philosophy & Technology, 33(2), 225–244. https://doi.org/10.1007/s13347-019-00355-w

Week 13: skepticisms about democracy as a solution

- Himmelreich, J. (2023). Against "Democratizing AI". AI & SOCIETY, 38(4), 1333–1346. https://doi.org/10.1007/s00146-021-01357-z

---

Last updated August, 2025