

ENGL 54.41: Critical AI

Dartmouth College

'26W

Distrib: TAS

Instructor: James E. Dobson (email: James.E.Dobson@Dartmouth.EDU)

Website: <https://english.dartmouth.edu/people/james-e-dobson>

Office Hours: 7A Sanborn House: Tues. 10:30-11:30 a.m. Thurs. 3-4 p.m.

COURSE DESCRIPTION:

While the rapid improvement in generative AI over the past two years, especially as seen in interactive tools like OpenAI's ChatGPT and Google's Bard, caught many people by surprise, these technologies have a long history. The course takes up artificial intelligence as not a single technology or even a class of technologies, but rather a discourse concerned with the automation of perception (image, sound, text) originating in the mid-twentieth century. Critical AI examines the beliefs, theories, methods, and practices of machine learning as well as the social and cultural significance of these transformative technologies. Giving special attention to the intersection of humans and AI, we will examine the inputs and outputs of machine learning and the ethical issues related to training and using these tools. The large-scale AI projects of the present would not be possible without what critics have called digital dispossession and the extractive function of new forms of capital. In this course we will apply cultural critique to artificial intelligence while learning the fundamentals of how these technologies work and how they fail.

LEARNING OUTCOMES

By the end of the term, students will be able to

1. Understand the historical development of machine learning and the role this history plays in the affordances of contemporary technologies.
2. Understand the major components of contemporary machine learning and generative AI applications.
3. Critique the use of cultural datasets in machine learning.
4. Describe social and cultural concerns related to recent machine learning developments in computer vision and text applications.
5. Describe how generative AI has changed accounts of creativity and authorship.
6. Work with others to develop a project making use of AI and produce a critique of these technologies.

ASSIGNMENTS AND ASSESSMENTS

1. **Participation** (15%): Classroom participation is expected and required. Participation and work with others is expected during those classes marked as labs.
2. **Midterm Exam** (30%): The midterm will consist of two parts: an in-class exam on key terms and concepts and a take-home component requiring short written responses and completed cells of a Jupyter notebook to construct and evaluate machine learning models on a collection of images and/or texts.
3. **Final Project** (25%): A final (group or individual) project will make imaginative use of an open source fine-tuned large language model. Posters (20%) will frame, introduce, and provide critiques (making use of course readings) of the methods used in the project. Five-page, individually-authored reflections (10%) will apply Critical AI scholarship (from course readings and field research) to the project.

COURSE POLICIES:

Attendance: You are expected to attend class in person unless you have made alternative arrangements due to illness, medical reasons, or the need to isolate due to COVID-19. For the health and safety of our class community, please: do not attend class when you are sick, nor when you have been instructed by Student Health Services to stay home.

Academic Honor Principle: Dartmouth's Academic Honor Principle applies to all the work you submit for this course. Please refresh your memory by reading it over at least once again during the first week of term. Make sure you understand which assignments allow collaborative group work.

Religious Observances: Dartmouth has a deep commitment to support students' religious observances and diverse faith practices. Some students may wish to take part in religious observances that occur during this academic term. If you have a religious observance that conflicts with your participation in the course, please meet with me as soon as possible—before the end of the second week of the term at the latest—to discuss appropriate course adjustments.

Office Hours and E-Mail: Please meet with me during my posted office hours or by appointment to discuss your work or any aspect of the course. As Canvas email might not be the most reliable method of communication, please use my Dartmouth email address. I will attempt to respond to all email in a timely fashion, although it may take up to twenty-four hours for me to respond.

Accommodations: Students requesting disability-related accommodations and services for this course are required to register with Student Accessibility Services (SAS; Getting Started with SAS webpage; student.accessibility.services@dartmouth.edu; 1-603-646-9900) and to request that an accommodation email be sent to me in advance of the need for an accommodation. Then, students should schedule a follow-up meeting with me to determine relevant details such as what role SAS or its Testing Center may play in accommodation implementation. This process works best for everyone when completed as early in the quarter as possible. If students have questions about whether they are eligible for accommodations or have concerns about the implementation of their accommodations, they should contact the SAS office. All inquiries and discussions will remain confidential.

Mental Health and Wellness: The academic environment is challenging, our terms are intensive, and classes are not the only demanding part of your life. There are a number of resources available to you on campus to support your wellness, including: the Counseling Center which allows you to book triage appointments online, the Student Wellness Center which offers wellness check-ins, and your undergraduate dean. The student-led Dartmouth Student Mental Health Union and their peer support program may be helpful if you would like to speak to a trained fellow student support listener. If you need immediate assistance, please contact the counselor on-call at (603) 646-9442 at any time. Please make me aware of anything that will hinder your success in this course.

Generative AI: We are going to be heavily using and critiquing various kinds of machine learning and artificial intelligence methods and applications in this class. You will be using these in and outside of class in order to probe and understand their operations. That said, when it comes to the essay that you will complete for the end of the course, I would like you to be the single, identifiable author of that document.

Research Data Services: Assistance with using computational resources and support for project design and implementation can be found through Dartmouth Library's Research Data Services team. You can contact members of this team via email: researchdatahelp@groups.dartmouth.edu and phone: (603) 646-3845. <https://www.library.dartmouth.edu/research-data-services>

DAILY SCHEDULE OF ASSIGNMENTS

The schedule may be subject to revision throughout the course of the term. Any changes will be announced in class and through Canvas/Email.

Week One	Introduction to Critical AI
Mon 01/05	Course Introduction Rita Raley and Jennifer Rhee, "Critical AI: A Field in Formation," <i>American Literature</i> 95, no. 2 (2023): 185–204
Wed 01/07	Data, Modeling, and the Data Model Michael Gavin, "The Computation of Meaning," <i>Literary Mathematics: Quantitative Theory for Textual Studies</i> , Stanford: Stanford University Press, 2023. James E. Dobson, "Vector Hermeneutics: On the Interpretation of Vector Space Models of Text," <i>Digital Scholarship in the Humanities</i> 37, no. 1 (2022): 81–93.
Fri 01/09	Lab Python, data-types, document-term matrices, vector space, and related objects and concepts.
Week Two	Week Two: Neural Networks and the Perceptron
Mon 01/12	The Perceptron James E. Dobson, "Inventing Machine Learning with the Perceptron," <i>The Birth of Computer Vision</i> , Minneapolis: University of Minnesota Press, 2023. Frank Rosenblatt, "Design of an Intelligent Automaton," <i>Research Reviews: Office of Naval Research</i> , October 1958.
Wed 01/14	Neural Networks Ranjodh Singh Dhaliwal, et. al, "Rendering the Neural Network," <i>Neural Networks</i> , Minneapolis: University of Minnesota Press, 2024. James E. Dobson, "Early Machine Learning and Artificial Animal Intelligences," <i>Journal of the Midwest Modern Language Association</i> 57, no. 1 (2024): 63–89.
Fri 01/16	Lab neural networks, the Perceptron, and classification.
Week Three	Neural Language Models and Word Embeddings
Mon 01/19	Martin Luther King Jr. Day. – No Class
Tues 01/20	word2vec and fasttext Tomas Mikolov et al., "Distributed Representations of Words and Phrases and Their Compositionality," in <i>Advances in Neural Information Processing Systems</i> (2013): 3111–19. Tolga Bolukbasi et al., "Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings," <i>Adv. Neural Inf. Process. Syst.</i> , 2016, 4349–57.
Wed 01/21	Semantic Modeling, Static Embeddings Melanie Mitchell, "Words and the Company They Keep," <i>Artificial Intelligence: A Guide for Thinking Humans</i> New York: Picador, 2019. Simon Lindgren, "Collective Representations," in <i>Data Theory</i>
Fri 01/23	Lab Neural language models, subword tokenization, clustering.

Week Four	Machine Learning Datasets
Mon 01/26	Data Critique: Part One Katherine Bode and Lauren M. E. Goodlad, “Data Worlds: An Introduction,” <i>Critical AI</i> 1 October 2023. Catherine D’Ignazio and Lauren Klein, “Collect, Analyze, Imagine, Teach,” <i>Data Feminism</i> (Cambridge: MIT Press, 2020).
Wed 01/28	Data Critique: Part Two Emily Denton, et al., “On the Genealogy of Machine Learning Datasets: A Critical History of ImageNet,” <i>Big Data & Society</i> 8, no. 2 (July 2021). Abeba Birhane, Vinay Uday Prabhu, and Emmanuel Kahembwe, “Multimodal Datasets: Misogyny, Pornography, and Malignant Stereotypes” (arXiv, October 5, 2021). Christo Buschek and Jer Thorp, “Models All The Way Down,” <i>Knowing Machines</i> https://knowingmachines.org/models-all-the-way .
Fri 01/30	Lab datasets, dataloaders, EDA
Week Five	Pre-Trained Networks and Transformers
Mon 02/02	Emily M. Bender et al., “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?,” in Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event Canada: ACM, 2021), 610–23. N. Katherine Hayles, “Inside the Mind of an AI: Materiality and the Crisis of Representation,” <i>New Literary History</i> 53, no. 4 (2022): 635–666.
Wed 02/04	William Alsup. “Order on Fair Use,” <i>Bartz v. Anthropic</i> , Case 3:24-cv-05417-WHA, June 23, 2025. OpenAI, “GPT-4 Technical Report.” arXiv preprint, March 27, 2023.
Fri 02/06	Lab and Midterm Exam
Week Six	Instruction-Fine Tuned and Reasoning Models
Mon 02/09	Following Instructions, Prompting, and Governing Long Ouyang, et. al., “Training Language Models to Follow Instructions with Human Feedback,” (arXiv, 4 Mar 2022) Louise Amoore et al., “Politics of the Prompt: Government in the Age of Generative AI,” <i>Economy and Society</i> , October 7, 2025, 1–24, https://doi.org/10.1080/03085147.2025.2560177 .
Wed 02/11	“Thinking,” “AGI,” and “Reasoning” Sébastien Bubeck, et. al., “Sparks of Artificial General Intelligence: Early Experiments with GPT-4” (arXiv, 13 Apr 2023). R. Thomas McCoy et al., “Embers of Autoregression Show How Large Language Models Are Shaped by the Problem They Are Trained to Solve,” <i>Proceedings of the National Academy of Sciences</i> 121, no. 41 (2024): e2322420121, https://doi.org/10.1073/pnas.2322420121 . Parshin Shojaee, “The Illusion of Thinking: Understanding the Strengths and Limitations of Reasoning Models via the Lens of Problem Complexity,” <i>arXiv:2506.06941</i> preprint, arXiv, November 20, 2025, https://arxiv.org/abs/2506.06941
Fri 02/13	Lab Instruction fine-tuning, EDA of RLHF datasets, etc.

Week Seven	Writing and Authorship in Age of GenAI
Mon 02/16	Classic Theories of Text and Authorship Roland Barthes, "The Death of the Author," <i>Image—Music—Text</i> New York: Hill and Wang, 1977. Walter J. Ong, "Writing Restructures Consciousness," <i>Orality and Literacy</i> New York: Routledge, 2022.
Wed 02/18	New Theories of Text and Authorship Mercedes Bunz, "Thinking Through Generated Writing," in <i>Thinking with AI: Machine Learning the Humanities</i> , ed. Hannes Bajohr. 58-84. London: Open Humanities Press, 2025. Mark Coeckelbergh and David J. Gunkel, "ChatGPT: Deconstructing the Debate and Moving it Forward," <i>AI & Society</i> 39 (2024):2221–2223
Fri 02/20	Lab
Week Eight	Theorizing Generated Texts
Mon 02/24	Return to Theory David Berry, "Provenance Anxiety: Death of the Author in the Age of Large Language Models," <i>Stunlaw</i> , https://stunlaw.blogspot.com/2025/12/provenance-anxiety-death-of-author-in.html Leif Weatherby, "Large Literary Machines," <i>Language Machines: Cultural AI and the End of Remainder Humanism</i> Minneapolis: University of Minnesota Press, 2025.
Wed 02/25	Writing and Reading GenAI Jill Walker Rettberg, Hermann Wigers, "AI-generated stories favour stability over change: homogeneity and cultural stereotyping in narratives generated by gpt-4o-mini," https://doi.org/10.12688/openreseurope.20576.1 Luciano Floridi, "Content Studies: A New Academic Discipline for Analysing, Evaluating, and Designing Content in a Digital and AI-Driven Age," <i>Philosophy & Technology</i> 38, no. 2 (2025): 41, s13347-025-00877–6. https://doi.org/10.1007/s13347-025-00877-6
Fri 02/27	Lab James E. Dobson, "On Reading and Interpreting Black Box Deep Neural Networks," <i>International Journal of Digital Humanities</i> 5 (2023): 431–49.
Week Nine	Analyzing Outputs
Mon 03/02	Critical Frameworks Simon Lindgren, "AI assemblage," <i>Critical Theory of AI</i> Hoboken, NJ: Polity, 2024. Joy Buolamwini, "The Defaults are not Neutral," <i>Unmasking AI</i> New York: Penguin 2023.
Wed 03/04	Interpretability Tyler Shoemaker, "For Those Who May Find Themselves on the Red Team," <i>arXiv:2511.18499</i> , preprint, arXiv, November 23, 2025, https://doi.org/10.48550/arXiv.2511.18499 . James E. Dobson "Beyond Computational Formalism or, Architecture Matters," <i>Journal of Cultural Analytics</i> 10, no. 3 (2025)
Fri 03/06	Poster Session

Week Ten **Wrapping Up**

Mon 03/09 **Final Class**

Fabian Offert and Ranjodh Singh Dhaliwal, “The Method of Critical AI Studies, A Propaedeutic,” arXiv preprint arXiv:2411.18833.