Neuro101GG Syllabus
Fall 2025

T. Anderson Keller
March 12th, 2025

The Neuroscience of Artificial Neural Networks: From Inspiration to Analysis

Time: Thursdays, from 3:45pm-5:45pm, Fall 2025. Location TBD.

Instructor

Name: T. Anderson "Andy" Keller, Ph.D.

Position: Research Fellow @ The Kempner Institute for the Study of Natural & Artificial Intelligence

Email: t.anderson.keller@gmail.com

University Webpage: https://kempnerinstitute.harvard.edu/people/our-people/t-anderson-keller/

Course Description

How “neural” are artificial neural networks? Drawing on foundational and modern research, this course explores the conceptual and mathematical parallels between biological and artificial neurons, tracing their intertwined histories from early perceptrons to today’s large-scale deep networks. We will investigate fundamental similarities and key divergences—such as with neural selectivities, large scale organization, and learning rules—and discuss how neuroscience methods are being used to interrogate the internal structure of deep vision and language models today. By examining seminal and contemporary literature, students will build an intuition for how brain-inspired principles have shaped modern AI, the role of neuroscience in AI development today, and how neuroscientific techniques can offer unique insights into the behavior and organization of complex computational systems.

Course Goals

After completing this course, students should:

- Understand the historical influence of neuroscience on the development of artificial neural networks, and the roles that neuroscience may now play in its continued development, both at a high-conceptual level and a slightly more detailed mathematical level.
- Be able to critically read modern research articles at the intersection of machine learning and neuroscience, understanding their impact, and identify gaps in the literature where the frontiers of new research might lie.

Course Format

A typical tutorial meeting will be structured as a short lecture and review of the assigned reading which will naturally lead into a discussion session guided by the assigned questions and questions students may have had about the reading. After the discussion, I will overview the background material necessary for the next reading, including key mathematical concepts. There will be one large writing assignment for the class (split into two sub-assignments), focused on the role of neuroscience in modern AI, which we will also turn into an in-class debate.

Typical Enrollees

The course is primarily designed for computationally oriented students with some familiarity with artificial neural networks, and/or an appropriate mathematical background (**understanding the basics of linear algebra, differential equations, and multi-variable calculus**). No programming experience or assignments will be required, although to do well in the course, it will be necessary for students to be able to understand the machinery of a machine learning algorithm such as a deep neural network at a precise level. The fundamental mathematical concepts required for each week's reading will be reviewed during lecture prior to the assignment, but prior familiarity with the ideas will be highly beneficial to get the most out of the class.

Methods of Instruction

The class will be structured to encourage as much student participation as possible. The course will mainly follow seminal research papers in the development of biologically inspired artificial neural networks, modern NeuroAI, and AI interpretability research, with small lectures being offered to help reinforce the core concepts. I will offer open office hours once per week to encourage informal discussion about any of the topics students may be interested in, as well as meetings by appointment if desired.

Assignments

Each week, students will be expected to read roughly 15-20 pages of primary source material (typically research papers), and answer a set of prepared discussion questions to be submitted to Canvas prior to class. During class, we will use these questions as the focus of our in-class discussion, and students will be graded based on their in-class participation. After class, students will have the opportunity to revise their previously submitted answers, and I will grade the final version. This work is expected to take roughly 5-7 hours per week outside class.

There will additionally be one major writing assignment for the class, structured as a rigorously grounded opinion article, aiming to answer the question: "What role has neuroscience played in AI development in the past, and what role does it have in the future?". This assignment will be broken into two parts, with a draft of the historical answer due at mid-terms, and a full revised version of the full article due at the end of the term.

Grading

- ○ 25% In-class participation in discussions.
- ○ 25% Weekly discussion questions submitted through Canvas.
- ○ 15% Midterm paper draft
- ○ 35% Revised Final paper draft.

Grade Levels: 93-100: A, 90-92: A-, 87-89: B+, 83-86: B, 80-82: B-, 77-79: C+, 73-76: C, 70-72: C-

Accomodations

Any student needing academic adjustments or accomodations is requested to present their letter from the Disability Access Office (DAO) and speak with Dr. Keller by the end of the second week of the term. All discussions will remain confidential. DAO may be consulted to discuss appropriate implementation. If you encounter accessibility issues or learning barriers related to the course design or materials, please let me know immediately so I can make adjustments as possible

Tentative Schedule (Subject To Change)

Part 1: The Intertwined History of Neuroscience and Artificial Neural Networks

Week 1 (Sept. 4nd): Introduction to Artificial Neurons, Perceptrons and Deep Neural Networks

○ *Lecture*: Introduce the linear artificial neuron, non-linear activation functions, and how these can be derived from approximations/abstractions of biophysical equations. Introduce the Multilayer Perceptron. Present excerpts of McCulloch and Pitts, as well as Rosenblatt's Perceptron paper. Present excerpts from the optional background reading.

○ Optional Background Reading:
  - McCulloch, W. S., & Pitts, W. (1943): "A Logical Calculus of the Ideas Immanent in Nervous Activity"
  - Rosenblatt, F. (1958): "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain"
  - Yang, G. R. & Wang X. (2020): "Artificial Neural Networks for Neuroscientists: A Primer"

○ *Reading Preparation:* Give brief introduction to Backpropagation for next week's reading. Present exceprts from: Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986): "Learning Representations by Back-Propagating Errors".

○ Homework:
  - Read: Lillicrap, T. P., et al. (2020): "Backpropagation and the Brain"

Week 2 (Sept. 11th): Learning in Artificial Neural Networks: Backpropagation

○ *Lecture:* A review of backpropagation, and how it is used to train simple neural networks. Demonstration of a simple example by hand.

○ *Discussion:* Discuss biological plausibility of backpropagation, answer student's weekly questions as a group, brainstorm experiments to test different learning algorithms.

○ *Reading Preparation:* Give a brief introduction to unsupervised learning and sparse coding ideas. Present excerpts from Hubel, D. H., Wiesel, T. N. (1962): "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex"

○ Homework:
  - Read: Olshausen, B. & Field, D. (1996) "Emergence of Simple-Cell Receptive Field Properties by Learning a Sparse Code for Natural Images"

Week 3 (Sept. 18th): Hubel & Weisel and Sparse Coding in ANNs

○ *Lecture:* A review of neural selectivity, and the ideas of sparse coding, how they relate to biological functions.

○ *Discussion:* Discuss implications of the results in the paper from a neuroscience perspective, answer student's prepared questions as a group.

○ *Reading Preparation:* Short introduction to retinotopy and convolutional neural networks. Present excerpts from LeCun, Y., et al. (1998): "Gradient-Based Learning Applied to Document Recognition".

○ Homework:
  - Read: Fukushima, K. (1980): "Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position"
  - Read: Krizhevsky, A. et al. (2012) "ImageNet Classification with Deep Convolutional Neural Networks"

Week 4 (Sept. 25rd): Retinotopy and Convolutional Neural Networks

○ *Lecture:* A review of the convolutional neural network structures introduced by Fukushima and LeCun. A review of the AlexNet model from the optional reading, why it was so succesful, and how it differed from the Neocognitron.

○ *Discussion:* Discuss how CNNs relate and differ from the actual organization of visual cortex.

Brainstorm how we might build models which are organized more like biological systems.

- *Reading Preparation:* An introduction to recurrent neural networks. Present excerpts from: Elman, J. (1999) "Finding Structure in Time".
- Homework:
  - Read: Barak, O. (2017) "Recurrent neural networks as versatile tools of neuroscience research"

Week 5 (Oct. 2th): Recurrent Neural Networks: Adding Time Dependence

- *Lecture:* A review of recurrent neural networks and how they relate to biophysical equations of neuronal activity and spiking neural networks.
- *Discussion:* Discuss the limitations and advantages of RNNs compared with more biophysically realistic models.
- *Reading Preparation:* An introduction to associative memory models, attractor networks and Hopfield Networks. Present excerpts from: Hopfield, J. (1982) "Neural Networks and Physical Systems with Emergent Collective Computational Abilities".
- Homework:
  - Read: Vaswani A. et al. (2017) "Attention Is All You Need"

Week 6 (Oct. 9th): Hopfield Networks and Transformers

- *Lecture:* Review hopfield networks and transformers. Lecture on how transformers can actually be seen as an instantiation of a 'Modern Hopfield Network'.
- *Discussion:* Dedicate discussion to answering students questions about this connection and the implications for the connection between the brain and modern large scale neural networks.
- *Reading Preparation:* Present preliminary arguments from: Gershman, S. (2024) "What have we learned about artificial intelligence from studying the brain?"
- Homework:
  - Read: Gershman, S. (2024) "What have we learned about artificial intelligence from studying the brain?" (short)
  - Read: Hassabis D. et al (2017) "Neuroscience-Inspired Artificial Intelligence".

Week 7 (Oct. 16th): The Controversy

- *Lecture:* None, straight to discussion.
- *Discussion:* Discuss the controversy surrounding the need for neuroscience inspiration in modern AI. Try to encourage students to share their own opinion about the historical importance of neuroscience in AI development, and its future potential.
- *Reading Preparation:* Introduce the idea of neural representations, representational similarity matrices, and neural predictivity measures.
- Homework:
  - Read: Kriegeskorte, N. et al. (2008) "Representational similarity analysis – connecting the branches of systems neuroscience"
  - Submit Midterm paper draft.

Part 2: Neuroscience Methods for Studying Ariticial Neural Networks

Week 8 (Oct. 23rd): Comparing Neural Representations

- *Lecture:* Review Representational Similarity Analysis.
- *Discussion:* Discuss the inherent limitations of a such metric, aim to get students to discover them independently.
- *Reading Preparation:* Review deep CNNs, and how their representations might be compared with fMRI.
- *Homework:*
  - Read: Yamins, D. et al. (2014) "Performance-optimized hierarchical models predict neural responses

in higher visual cortex"

Week 9 (Oct. 30th): Models of Object Recognition

- Lecture: Review the methods of the DiCarlo paper, and overview other recent similar techniques, such as Brain-Score. Present excerpts from Schrimpf, M., et al. (2020): "Brain-Score: Which Artificial Neural Network for Object Recognition is most Brain-Like?"
- *Discussion:* Discuss the implications of the DiCarlo results and their limitations.
- *Reading Preparation:* Introduce Cohen's metric of neural selectivity and topographic organization in the brain.
- Homework:
  - Read: Lee, H. et al. (2020) "Topographic deep artificial neural networks reproduce the hallmarks of the primate inferior temporal cortex face processing network"

Week 10 (Nov. 6th): Studying the Selectivity and Organization of Artificial Neurons

- *Lecture:* Review the Topographic Deep ANN architecture and the methods used to study functional specialization in Dobs, K. et al. (2022) "Brain-like functional specialization emerges spontaneously in deep neural networks". Review other more recent topographic models.
- *Discussion:* Discuss the implications of these results to the theories of functional specialization and topographic organization in the brain.
- *Reading Preparation:* Review recurrent neural networks and how they can be trained and evaluated on cognitive tasks.
- Homework:
  - Read: Yang, G. et al. (2019) "Task representations in neural networks trained to perform many cognitive tasks"

Week 11 (Nov. 13th): Studying Modularity and Dynamics in Recurrent Neural Networks

- *Lecture:* Review the methods used in Yang (2019).
- *Discussion:* Discuss the results, and implications for the development of functionally specialized areas in the brain.
- *Reading Preparation:* Review sparse autoencoders and introduce language models at a high level.
- Homework:
  - Read: Bricken T. et al. (2023) "Towards Monosemanticity: Decomposing Language Models With Dictionary Learning"

Week 12 (Nov. 20th): Interpreting Large Neural Networks

- *Lecture:* (Potential Guest Speakers Adam Jermaine (Anthropic), or Thomas Fel (Kempner Fellow))
- *Discussion:* Discuss the results of sparse autoencoders applied to large language models, and how this relates to similar neuroscience findings.
- Homework:
  - Work on final paper. (Due at end of term).

Week 13 (Nov. 27th): Holiday, Thanksgiving

Week 14 (Dec. 4nd): No Class, Reading Period.

Course Policies

Attendance: A significant portion of the material in this course will be presented in-class during lectures, and elaborated in our discussions. Therefore, it is critical that students attend the class and participate in discussion. Each unexcused absence will result in a corresponding 0 grade for the weekly in-class participation and 50% deduction for the weekly discussion questions, corresponding to ≈ 3% of your final grade. Therefore three unexcused absences will result in a full letter grade dropped. Absences can be made up without penalty if you contact the instructor in advance of the absence and we agree on an additional assignment to complete as an alternative to participation that week.

Participation grade: Given that it is a small class, active participation by all students will be crucial for not only their own learning experience, but also for the learning experience of other students. The participation grade is meant to reflect this, and to encourage students to read the weekly paper in a manner that enables critical feedback in class discussion. I will work my hardest to seek out engagement from every student during each course, and the grade is not meant to punish students that are naturally more reserved. If, occasionally, you leave class feeling that you did not get to contribute a comment that you wish you could have, I high encourage and welcome emails after class (or meeting during office hours) and this may be counted towards your participation for that week.

Late work: Students are expected to turn in all discussion question answers on-time, as having carefully considered these answers, and the associated reading, are the two core requirements for making the in-class discussions possible. Therefore, late discussion questions will be deducted 25% per day late, and will not be accepted after the in-class discussion. The midterm paper and final paper will be deducted 10% for each day late.

Collaboration: Students are highly encouraged to discuss the weekly readings and their arguments for the final paper. If these discussions result in substantial consensus or disagreement, students are encouraged to mention these conclusions, and the corresponding discussion partner in their submitted work. However, each submitted written assignment must be expressly each student's own work.

Academic Integrity: Students are expected to correctly cite literature, references, and sources in all of their discussion question responses and long-form assignments. All students are expected to follow the Harvard College Honor Code (https://oaisc.fas.harvard.edu/honor-code/), and any violations will be treated as academic misconduct.

Additional course policies from the Harvard College Student Handbook can be found here: https://canvas.harvard.edu/courses/157216/pages/course-policies

AI Policy

Certain assignments in this course will permit or even encourage the use of generative artificial intelligence (GAI) tools such as ChatGPT. The default is that such use is disallowed unless otherwise stated. Any such use must be appropriately acknowledged and cited. It is each student's responsibility to assess the validity and applicability of any GAI output that is submitted; you bear the final responsibility. Violations of this policy will be considered academic misconduct. We draw your attention to the fact that different classes at Harvard could implement different AI policies, and it is the student's responsibility to conform to expectations for each course.

In particular, for answering the discussion questions, and writing the review article, the use of GAI will be disallowed. Students will be encouraged to have discussions with such tools to better understand the course material, but all writing turned in by the students must be explicitly their own.