Model AI Assignments 2026

Todd W. Neller[1], **Steve Geinitz**[2], **Kevin Wang**[3], **Zach Dodds**[4], **Nicholas Dodds**[4], **Ryan O Connor**[5], **Aimen Taha**[5], **Ananta Manoranjan**[6], **Saurabh Ray**[6], **Deepak Ajwani**[5], **Fang Sun**[7], **Paul Zhang**[7], **Pranav Subbaraman**[7], **Yizhou Sun**[7], **Lisa Dunlap**[8], **Taehan Kim**[8], **Narges Norouzi**[8], **Joseph Gonzalez**[8], **Deena Sun**[8], **Ishir Garg**[8], **Mark Ogata**[8], **Aakarsh Vermani**[8], **Varada Kolhatkar**[9]

[1]Gettysburg College,
[2]Metropolitan State University of Denver,
[3]Stanford University,
[4]Harvey Mudd College,
[5]University College Dublin,
[6]New York University Abu Dhabi,
[7]University of California, Los Angeles,
[8]University of California, Berkeley,
[9]The University of British Columbia.

Abstract

The Model AI Assignments session seeks to gather and disseminate the best assignment designs of the Artificial Intelligence (AI) Education community. Recognizing that assignments form the core of student learning experience, we here present abstracts of eight AI assignments from the 2026 session that are easily adoptable, playfully engaging, and flexible for a variety of instructor needs. Assignment specifications and supporting resources may be found at http://modelai.gettysburg.edu.

ArguBot Arena: Prompt Engineering a Debate on Responsible AI - Steve Geinitz

ArguBot Arena is an engaging and pedagogically rich assignment intended to introduce students to responsible AI, prompt engineering, and language model (LLM) behavior through structured AI-driven debates. In this assignment, students craft prompts for an LLM and manage its context to have it act as a debater representing one side of an ethical or controversial topic in AI (e.g. copyright, surveillance, bias). These "ArguBots" will then autonomously engage in a multi-round debate against one another to see which student's prompts were able to create the most effective debater.

This assignment combines technical skill development with critical thinking and ethical reflection. Students will gain experience with prompt engineering, role consistency, and context preservation, while also analyzing the strengths and weaknesses of AI-generated arguments. They must evaluate the effectiveness of their prompts not only by the coherence of individual responses, but also by the LLM's ability to stay in character and respond to its opponent.

By turning ethical discourse into an applications-based activity, ArguBot Arena enhances AI literacy, builds prompt-writing fluency, and fosters responsible AI thinking. It's ideal for an undergraduate machine learning courses with introductory exposure to LLMs but could be adapted to suit different ages and/or course subjects.

Solving Connections: Thinking Like Wyna - Kevin Wang, Zach Dodds, and Nicholas Dodds

This assignment asks students to create a small software system for solving Connections, a New York Times puzzle crafted daily by Wyna Liu. Connections provides sixteen words, and the player's goal is to make four semantically coherent groups of four words each.

Connections is popular—and challenging: many students find ample intrinsic motivation; many others find themselves newly hooked! Connections inherently poses questions of semantic similarity. For example, does brush, dress, tidy, neat form a "better" group than brush, dress, key, pocket? Opinions differ—strongly!

Students use a set of OpenAI's text embeddings and cosine similarity to explore the computational similarity of groups of words. From those building blocks, students create their own measures of overall solution-coherence, which is the heart and soul of this assignment. After all, any "mistake" by their algorithm might reveal a sub-optimally posed puzzle! The assignment description targets a CS1–CS2 audience. It also includes an on-ramp that can be—and has been—used as a "Welcome to CS" icebreaker, well suited to a CS0 background and environment.

Discover Combinatorial Structures using Deep Cross-Entropy Method - Ryan O Connor, Aimen Taha, Ananta Manoranjan, Saurabh Ray, and Deepak Ajwani

Combinatorial optimization, and the discovery of novel combinatorial configurations, is critical for a range of applications, including drug discovery and network design. This assignment explores the application of reinforcement learning (RL) to discover combinatorial structures under constraints using the cross-entropy method.

Building on Wagner's framework, students learn to translate neural network bit vectors into valid permutations and evaluate them with custom scoring functions. The assignment progresses through various discrete problems to demonstrate this approach's flexibility. It begins with sequence optimization, where the goal is to balance or minimize the lengths of the longest increasing and decreasing subsequences. It then moves to geometric challenges, such as constructing intersecting and non-intersecting sets of rectangles.

Each task emphasizes the design of effective encoding schemes and scoring functions, forcing students to apply critical reasoning rather than blindly using RL. The framework leverages Python, Keras, and TensorFlow, allowing for accessible experimentation in environments like Google Colab.

By completing this assignment, students gain practical experience in combining machine learning with classical combinatorial reasoning. They discover how RL can provide novel insights and solutions to problems traditionally solved with heuristics, exhaustive search, or complex solvers.

RevMax: Revenue-Maximizing Recommendation System Competition - Fang Sun, Paul Zhang, Pranav Subbaraman, and Yizhou Sun

RevMax is an innovative assignment that challenges students to develop recommendation algorithms optimizing for real-world business metrics—revenue generation. Using the Sim4Rec simulation framework, students compete in a multi-iteration environment where their algorithms must learn from user feedback and adapt recommendations over time, mirroring production recommendation systems.

The assignment uniquely integrates core data mining concepts through three progressive checkpoints: content-based filtering leveraging user/item features, sequence-based modeling capturing temporal patterns, and graph-based approaches exploiting relational structures. Students implement algorithms in PySpark, gaining hands-on experience with distributed computing while tackling fundamental ML challenges including exploration-exploitation tradeoffs, cold-start problems, and online learning.

What distinguishes RevMax is its emphasis on cumulative performance across iterations, forcing students to balance immediate revenue with long-term user satisfaction. The competitive leaderboard format motivates iterative improvement while hidden test environments ensure robust solutions. By connecting recommendation systems to tangible business outcomes, RevMax demonstrates AI's practical impact while teaching advanced machine learning techniques from traditional methods to deep neural architectures.

The assignment's modular design supports customization for different course levels and timeframes. Comprehensive baseline implementations, evaluation metrics, and visualization tools lower adoption barriers while maintaining intellectual challenge. RevMax transforms abstract ML concepts into engaging, practical experience preparing students for real-world recommendation system development.

Ranking Large Language Models with LMArena - Lisa Dunlap, Taehan Kim, Narges Norouzi, Joseph Gonzalez, Deena Sun, Ishir Garg, Mark Ogata, and Aakarsh Vermani

This assignment introduces students to evaluating LLMs via pairwise comparisons and logistic regression on LMArena data, a popular evaluation platform. In LMArena, users submit prompts to two anonymous models and vote on the better response, producing comparative preference data that underlies leaderboard rankings. Students begin by exploring the distribution of these battles, calculating naïve win rates, and then computing the leaderboard via the Bradley–Terry model. Then students explore how to create different "views" of the leaderboard for specific language and prompt types to further investigate the different strengths and weaknesses of models.

In part 2 of the homework, students explore the role of stylistic confounds in model preference. Students extract features like bold text usage, headers, list formatting, and response length from conversations. By constructing pairwise feature vectors, they quantify how much stylistic factors affect win probabilities and compare leaderboards with and without style control.

This project highlights key challenges in modern AI evaluation: how to disentangle true model capability from superficial stylistic signals. Through coding, statistical modeling, and interpretation, students gain hands-on experience with ranking systems, logistic regression, and confound analysis. This introduces students to the nuances of model evaluation by implementing techniques used in the actual LMArena.

CS2023 Machine Learning CS Core - Todd W. Neller

These materials cover the CS2023: ACM/IEEE-CS/AAAI Computer Science Curricula CS Core topics of the Machine Learning (AI-ML) knowledge unit (Kumar et al. 2024, p. 71). Using interactive Python Jupyter notebooks, topics are not only introduced experientially with associated exercises, but notebook code examples provide a recipe book for immediate application of principles.

Provided Jupyter notebooks cover fundamental ML concepts across four sessions. Session 1 introduces supervised learning (classification with k-Nearest Neighbors and Decision Trees, regression with Linear Regression), unsupervised learning (k-Means Clustering), and reinforcement learning (Q-Learning). Session 2 addresses model evaluation, including train-test-validation split, performance metrics, over-/under-fitting, and the bias-variance tradeoff, with emphasis on polynomial regression and regularization. Session 3 focuses on data preprocessing, including handling missing values, encoding categorical variables, normalization/standardization, and feature engineering. It concludes with discussion of the No Free Lunch theorem and sources of error. Session 4 covers neural network basics, explaining artificial neurons, activation functions, and multilayer perceptrons, with a high-level overview of transformers and large language models. This concludes with coverage of ML ethics, addressing dataset bias, algorithmic bias, evaluation bias, and real-world case studies of unintended consequences in ML applications.

Dimensionality Reduction Adventures with Animal Faces - Varada Kolhatkar

This assignment introduces Principal Component Analysis (PCA) through conceptual reasoning, algorithmic implementation, and applied experimentation. It is designed for undergraduate and early graduate students from multidisciplinary backgrounds seeking to apply machine learning methods in their domains.

Students develop an understanding of PCA by examining the relationships between original data, principal components, and reconstructed data, then implement PCA using Singular Value Decomposition (SVD) and validate their results against scikit-learn.

They apply PCA to a real-world Animal Faces dataset, using visualizations to explore how the number of components affects explained variance and reconstruction quality, and to reason about the trade-off between compression and information loss. Students also interpret the learned components, examine the effects of outliers and off-distribution examples, and connect mathematical abstractions to recognizable image features, reflecting on the limitations of linear dimensionality reduction. An optional, low-stakes extension invites students to implement convolutional autoencoders and compare their reconstructions with PCA.

The assignment is self-contained, includes starter code, and emphasizes conceptual understanding and critical reasoning in dimensionality reduction.

References

Kumar, A. N.; Raj, R. K.; Aly, S. G.; Anderson, M. D.; Becker, B. A.; Blumenthal, R. L.; Eaton, E.; Epstein, S. L.; Goldweber, M.; Jalote, P.; Lea, D.; Oudshoorn, M.; Pias, M.; Reiser, S.; Servin, C.; Simha, R.; Winters, T.; and Xiang, Q. 2024. *Computer Science Curricula 2023*. New York, NY, USA: Association for Computing Machinery. ISBN 9798400710339.