SCIS 432: Artificial Intelligence

Final Project

Guidelines, Options, and Evaluation Criteria

Overview

The final project is your opportunity to go beyond homework problems and engage deeply with a topic in artificial intelligence that genuinely interests you. You will choose one of five project options, develop it over the final weeks of the semester, and produce a written deliverable alongside a short presentation or demo. The goal is depth, honesty about limitations, and evidence that you engaged seriously with the material.

Key Dates

Milestone	Due Date	What to Submit
Pitch	April 7	1–2 page pitch document (see below)
Progress Check-in	April 21	Brief paragraph update posted to course portal
Final Submission	Finals Week	Written report + any code / slides / recording
Presentation / Demo	Finals Week	In-class or recorded, 10–15 minutes

The Pitch (Due April 7)

Before you begin serious work, you must submit a pitch and receive approval. The pitch is required. It is a chance to get early feedback and make sure your project is scoped appropriately.

What the Pitch Must Include

Your pitch should be one to two pages and address all of the following:

1. Which option you are pursuing (1 through 5).
2. A clear statement of what you plan to do (one focused paragraph).
3. Evidence of preliminary work appropriate to your option (detailed below for each option). You should not submit a pitch without having done this work.
4. The question or problem you are trying to answer or solve.
5. How you plan to evaluate success (what will “done” look like)?
6. Any concerns or uncertainties you already see, and how you plan to address them.

What gets a pitch approved?

- You have clearly read the dataset documentation, found the paper, or designed the experiment.
- Your scope is realistic for the remaining weeks of the semester.
- You can articulate what is interesting or novel about your choice.
- You have a specific evaluation plan, not just "I will train a model and see how it does."

Project Options

OPTION 1 Applied Machine Learning: Regression or Classification

Description

You will find a real dataset, apply one of the models we have covered (linear regression, logistic regression), or (if we reach it) a convolutional neural network or NLP model, and produce a thorough analysis. This is not just about fitting a model and reporting accuracy. The goal is to think carefully about your data, your modeling choices, and what your results actually mean.

Finding a Dataset

Good sources include Kaggle, the UCI Machine Learning Repository, Hugging Face Datasets, data.gov, or any domain-specific repository relevant to your interests. Your dataset should be non-trivial: at minimum several hundred rows and at least a handful of meaningful features. Avoid datasets that are homework staples (Iris, Titanic, MNIST on their own) unless you have a genuinely novel angle.

Required for the Pitch

- Identify the dataset and include its URL or citation.
- Describe it briefly: how many rows, how many features, what the target variable is, where it came from.
- State which model type you plan to use and why it is appropriate for this task.
- Describe at least one interesting thing you noticed in a first pass of the data.

Required in the Final Report

Exploratory Data Analysis. Before modeling, explore your data. Show distributions of key variables, check for missing values, look at correlations or class imbalances. Include at least two meaningful plots. Discuss what you learned.

Data Preparation. Explain your train/test split strategy. Justify your choices (split ratio, stratification, cross-validation if used). Why is it important to never look at the test set during model development?

Modeling. For linear regression: discuss feature selection or engineering, how you chose the polynomial degree or applied transformations, and whether those choices improved generalization or caused overfitting. For logistic regression: discuss thresholding decisions and any class imbalance considerations. For deep learning: describe your architecture choices.

Coefficient Interpretation. If using linear or logistic regression, interpret at least three coefficients in plain language. For logistic regression, use odds ratios. What do these coefficients tell you about the real-world phenomenon?

Evaluation. For regression: report train and test MSE and R^2. For classification: report accuracy, precision, recall, F1, and the AUC-ROC curve. Discuss what each metric reveals. Are there specific classes or outcomes that matter more? Did you adjust the decision threshold, and why?

Assumptions and Limitations. Are the rows independent? Are there features you could not include? What might be lurking confounders? Be honest about what your model cannot tell you.

What I look for in Option 1

- Genuine curiosity about the dataset not just going through the motions.
- Evidence that you explored the data before modeling, not after.
- Coefficient interpretations that go beyond restating the number.
- A clear-eyed discussion of where the model fails and why.
- Appropriate use of metrics for the problem (don't just report accuracy on an imbalanced dataset).

OPTION 2 Algorithm Deep-Dive: Classical AI or CSP

Description

You will take an algorithm we studied including search (A*, BFS, DFS, iterative deepening), adversarial search (minimax, alpha-beta), constraint satisfaction (backtracking, arc consistency, local search), or another classical AI method, and apply it to a new, self-chosen problem. The emphasis is on implementation, experimental rigor and genuine insight into the algorithm's behavior.

Required for the Pitch

- State the algorithm and the problem you are applying it to.
- Explain why this problem is a good fit for the algorithm.
- Describe the state representation, action space, and goal test (or constraint formulation for CSP).
- If using a heuristic (e.g., A*), describe your initial heuristic idea and argue informally for admissibility.

Required in the Final Report

Problem Formulation. Define the problem precisely: states, actions, transition model, goal, and cost function (or constraints and domains for CSP). Be rigorous.

Heuristic Design. If applicable: how did you design your heuristic? Prove or argue that it is admissible. Did you consider multiple heuristics? How sensitive is performance to heuristic quality?

Experimental Results. Collect and report quantitative statistics: nodes explored, memory usage (frontier size), wall-clock time, solution cost or quality. Compare at least two configurations—e.g., different heuristics, different search strategies, or different problem sizes.

Analysis. What did you learn about the algorithm's behavior on this problem? Where did it struggle? Did the results match your theoretical expectations?

Scalability. How does the algorithm scale as the problem grows? Include at least one plot of runtime or nodes explored vs. problem size.

What I look for in Option 2

- A well-defined problem that genuinely benefits from the algorithm, not a toy example.
- Rigorous experimental methodology: controlled comparisons, multiple trials if stochastic.
- Honest discussion of cases where the algorithm fails or scales poorly.
- For A* or similar: a careful treatment of admissibility and its effect on optimality.

OPTION 3 Paper Review: Top ML/AI Conference

Description

You will select a paper published at a top machine learning or AI venue within the last five years, develop a deep understanding of it, and produce both a written review and a 15-minute recorded or live presentation. This option requires the most background preparation and works best if the paper's topic connects to something you have studied or are genuinely curious about.

Eligible venues: NeurIPS, ICML, ICLR, AISTATS, UAI, ACL, EMNLP, CVPR, ICCV (papers from 2020 onward). When in doubt, ask.

A useful resource: The following YouTube channel (`https://www.youtube.com/c/YannicKilcher`) provides good examples of how to read and present ML papers critically. Watch two or three before writing your pitch.

Required for the Pitch

- Full citation of the paper (title, authors, venue, year) and a link to the PDF.
- A paragraph summarizing the paper's core contribution in your own words (not the abstract copy-pasted).
- A paragraph explaining why you find it interesting and what background knowledge you already have.
- A brief note on which parts you expect to be the hardest to understand.

Required in the Final Written Review

Summary. What problem does the paper address? What is the proposed method? What are the main results?

Technical Contribution. What is the core technical idea? Explain it at a level a fellow student could understand. Use diagrams or equations if helpful.

Experimental Evidence. What datasets and baselines do the authors use? Are the comparisons fair? Do the results convincingly support the claims?

Strengths and Weaknesses. What does the paper do well? What assumptions does it make? What questions does it leave open?

Your Reaction. What surprised you? What would you have done differently? What follow-up work would you want to see?

Presentation Rubric (15 minutes)

Dimension	What I look for	Points
Motivation & Context	Did you explain why the problem matters and where this paper fits in the literature?	6
Technical Accuracy	Did you correctly explain the method, including any key equations or architecture?	8
Critical Analysis	Did you go beyond summarizing? Do you identify assumptions, limitations, or open questions?	6
Clarity & Delivery	Is it easy to follow? Are slides well-organized? Is the pacing appropriate?	5
Q&A / Engagement	Can you answer follow-up questions or address points raised in discussion?	5
	Total	**30**

What I look for in Option 3

- Evidence you actually read and thought about the paper—not just watched someone else explain it.
- Technical honesty: it is fine to say "I did not fully follow Section 4.2" and explain what you understood.
- Critical engagement: praise and critique, not just summary.
- A presentation that teaches something—assume your audience has not read the paper.

OPTION 4 Experiment in Ethical AI

Description

You will design and run a rigorous empirical experiment exploring a question in AI ethics, fairness, or bias. This could involve probing a language model, analyzing a publicly available model's outputs, or auditing a dataset. The key requirement is scientific rigor: you must have a clearly stated hypothesis, a controlled experimental design, quantitative measurements, and a statistical analysis of your results.

Example Directions

- Prompt a large language model (e.g., via the Claude or OpenAI API) with pairs of prompts that differ only in demographic attributes (name, gender, race, nationality). Measure differences in output sentiment, length, tone, or content. Formalize this as a hypothesis test.
- Audit a publicly available model (e.g., a recidivism predictor, a hiring algorithm, a sentiment classifier) for disparate performance across demographic groups using formal fairness criteria such as demographic parity or equalized odds.
- Design a study on how framing affects AI-generated advice. Collect data, define a coding scheme, and analyze results statistically.
- Investigate whether a model's expressed confidence is calibrated with its factual accuracy across different topic domains.

Required for the Pitch

- State your research question and hypothesis clearly including the null hypothesis.
- Describe your experimental design: what is the treatment, what is controlled, what are you measuring?
- Explain how you will operationalize your outcome variable.
- State which statistical test you plan to use and why.
- Describe any ethical considerations in your own data collection.

Required in the Final Report

Hypothesis & Design. State your null and alternative hypotheses. Describe the full experimental procedure in enough detail that someone else could replicate it.

Data Collection. How many prompts, trials, or examples did you collect? How did you ensure consistency? Include your full prompt templates or data collection instrument in an appendix.

Quantitative Results. Present your measurements with descriptive statistics (means, standard deviations, distributions). Run an appropriate hypothesis test (t-test, chi-squared, Mann-Whitney U, etc.) and report the test statistic and p-value. State whether you reject the null hypothesis at a pre-specified significance level.

Interpretation. What do your results mean? Distinguish statistical significance from practical significance. What are the limitations of your design?

Broader Implications. If your findings hold, what are the real-world consequences? Who might be harmed? What would responsible deployment look like?

What I look for in Option 4

- A specific, testable hypothesis (not “I want to see if AI is biased”).
- Careful experimental controls: the only thing that should vary is what you are studying.
- Appropriate statistical methods, with acknowledgment of their assumptions.
- Intellectual honesty: null results are fine and sometimes more interesting.
- Thoughtful discussion of what your results can and cannot establish.

Overall Grading

The project is graded out of 100 points and constitutes 25% of your final course grade.

Component	**Points**	**Notes**
Pitch (approved)	10	Completeness, evidence of preliminary work, clear scope
Written Report / Essay	55	See option-specific criteria above
Presentation / Recording	25	Clarity, depth, ability to field questions
Progress Check-in	5	Brief update showing momentum; graded complete/incomplete
Code / Appendix (if applicable)	5	Readable, reproducible, well-commented
Total	**100**	

Written Report Rubric (applies to all options)

Criterion	Excellent (A)	Adequate (B/C)	Insufficient (D/F)
Depth of engagement	Goes beyond surface; shows genuine curiosity and independent thinking	Covers required components adequately with some insight	Superficial; reads like it was completed at the last minute
Accuracy	All technical claims are correct and precisely stated	Minor errors that do not undermine the argument	Significant factual errors or misconceptions
Evaluation & analysis	Appropriate methods; results critically interpreted; limitations honestly discussed	Correct methods; limited critical discussion of results	Inappropriate metrics; results not interpreted; no limitations discussed
Clarity & organization	Well-structured; easy to follow; writing is precise	Mostly clear with some organizational issues	Hard to follow; disorganized; imprecise language
Honest about uncertainty	Acknowledges what the analysis cannot establish	Some acknowledgment of limitations	Overstates conclusions; no limitations discussed

Academic Integrity

You may use AI writing tools (such as Claude, ChatGPT, or Copilot) as a resource, subject to the following conditions:

- Any AI-generated text that appears in your submission must be clearly labeled as such.
- The analysis, arguments, interpretations, and conclusions must be your own. Using an AI to generate the substance of your project violates this policy.
- You are responsible for the accuracy of everything you submit, including anything produced with AI assistance.

The best projects typically use AI tools in the way you might use a research assistant: to help clarify a concept, check code, or suggest a direction—not to do the thinking for you. The work that will distinguish your project is judgment, interpretation, and intellectual honesty, none of which can be outsourced.

Questions?

- Post questions about project scope or option eligibility to the course discussion board so others can benefit from the answer.
- For feedback on a specific pitch draft before the official deadline, office hours are the right place.
- If your topic falls outside the five options but is genuinely interesting, pitch it anyway—unusual proposals are considered.