51.511 Multimodal Generative AI
Final Project Guideline

Group: 4 members per group. Register here by **Feb 13, 2026**: https://docs.google.com/spreadsheets/d/1vYVDfQoxwN7gL7zww481IYdAC2jTvxlzRr2-pY3SAO8/edit?gid=0#gid=0.

Project proposal: Week 6, due on Mar 8, 23:59

Final Report: Week 12, due on Apr 19, 23:59

Final presentation: Week 13

Submission: Proposal, Presentation Slides, and Report in PDF form through eDimension

1 Objective

The main objective of this project is to equip and familiarize students with the necessary skills understand, use, and evaluation large multimodal models.

2 Project Templates

Students may freely choose the topic of their project, but must select **one** of the project templates below and satisfy all mandatory elements. The goal is to encourage creativity while ensuring research rigor and fair evaluation.

2.1 Template 1: Agentic Multimodal System

Goal: Design and evaluate an agentic system that coordinates one or more multimodal foundation models to solve a complex task. The task must require at least three distinct reasoning or decision-making stages that cannot be merged into a single model invocation.

- **Problem Decomposition**
  - Clearly defined task that cannot be solved by a single model call
  - Explicit decomposition into sub-tasks (e.g. planning, perception, reasoning, synthesis)

- **Agent Architecture**
  - System diagram illustrating agent components
  - Clearly defined roles (e.g. planner, executor, memory, evaluator)
  - Justification of model and tool choices
  - usable **user interface** for demo
- **Multimodal Grounding**
  - Use of at least two modalities (e.g. text, image, audio, video)
  - Explanation of how modalities interact within the agent
- **Control and Failure Handling**
  - Mechanisms for detecting errors or uncertainty
  - Retry, self-critique, or fallback strategies
- **Evaluation**
  - Task-specific success criteria
  - Comparison against at least two baselines: 1. single-prompt or single-model pipeline; 2. hand-designed non-agentic workflow
  - Analysis of failure cases. Be sure to us:
    - self-consistency checks
    - cross-modal agreement checks
    - verifier models
    - disagreement analysis
- **Reflection**
  - Discussion of when agentic behavior is beneficial
  - Limitations of the approach

2.2 Template 2: Generative Model Research Study

Goal: Modify or replace a small but meaningful component of an existing generative model and analyze its impact through controlled experiments.

- **Base Model Description**
  - Overview of the original architecture
  - Justification for model choice
- **Targeted Modification**
  - Modification of a single localized component (e.g. conditioning, attention, loss, scheduler)
  - No full retraining from scratch, rather finetuning
- **Hypothesis**

  - Clear statement of expected effects and trade-offs

- **Ablation Study**
  - Baseline model
  - Modified model
  - At least one additional ablation or variant

- **Evaluation**
  - Quantitative metrics where applicable
  - Qualitative analysis of generated outputs
  - Stability and failure mode analysis

- **Compute Awareness**
  - Reporting of parameter counts, training steps, and runtime
  - Discussion of scalability limitations
  - **User interface** to show difference between models.

2.3 Template 3: New Multimodal Task with Thorough Evaluation

Goal: Propose a new or underexplored multimodal task and rigorously evaluate existing generative models on it.

- **Task Definition**
  - Clear specification of inputs and outputs
  - Explanation of why existing benchmarks are insufficient
  - **User interface** for demo

- **Dataset Creation or Adaptation**
  - Data sources and preprocessing steps
  - Dataset statistics and splits

- **Baseline Models**
  - At least two different model families
  - Justification for model selection

- **Evaluation Protocol**
  - Automatic and/or human evaluation metrics
  - Inter-annotator agreement if applicable
  - Error taxonomy

- **Results Analysis**
  - Strengths and weaknesses of evaluated models
  - Identification of systematic failure patterns

- **Insights**
  - What this task reveals about current multimodal generative models
  - Open challenges

2.4 Template 4: Dataset Curation and Bias Analysis

Goal: Curate a multimodal dataset and analyze how dataset design and biases affect generative model behavior.

- **Dataset Motivation**
  - Intended use cases
  - Why a new dataset is needed
- **Data Collection and Cleaning**
  - Data sources and licensing considerations
  - Filtering and preprocessing procedures
- **Bias and Imbalance Analysis**
  - Statistical analysis of distributions
  - Identification of representation gaps
- **Baseline Models**
  - At least one generative baseline with a **user interface** for demo
  - Description of training or inference setup
- **Impact Analysis**
  - Qualitative examples of bias in generated outputs
  - Connection between data properties and model behavior
- **Ethical Considerations**
  - Privacy, consent, and potential misuse
  - Dataset limitations

3 Project Proposal (10%)

The project proposal is intended to assess the clarity, motivation, and feasibility of the proposed project. Each team must select one of the four provided proposal templates described above. At a minimum, the proposal should include the first section, or the first several sections, of the selected template, such as the problem definition, background and motivation, and preliminary methodology or model selection. The purpose of the proposal is to demonstrate that a well-defined problem has been identified and that initial efforts have been made toward addressing it through preliminary exploration, conceptual design, or early experimentation, as appropriate to the project scope.

The proposal must be prepared using the NeurIPS LaTeX template. The main body of the proposal should not exceed **three pages**, including all figures and tables, but excluding references and any supplementary material. Submissions that do not adhere to the NeurIPS format or exceed the page limit will not be accepted, as a unified format is required to ensure fairness and consistency in evaluation.

4 Final Report (15%)

The final project report is intended to present a complete and coherent account of the project, including the problem formulation, methodology, experimental results, and analysis. Each team must use the same template selected for the project proposal and include all required sections specified by that template. The final report should reflect a fully developed project, with clear technical depth, sound experimental design or analysis, and well-supported conclusions. The **source code accompanying the report should be made available in a GitHub repo** with clear readme.

The final report must be prepared using the same NeurIPS LaTeX template. The main body of the report must not exceed **six pages**, including all figures and tables, but excluding references and any supplementary material. Reports that fail to include all required sections, deviate from the selected template, or exceed the page limit will not be accepted.

5 Final Presentation (10%)

At the end of the project, you will give a final presentation. Each team will be allocated 10 min (time to be confirmed based on number of groups). Details of the presentations are:

- The team should clearly show which project template they chose and clearly articulate all the aspects described above.

Please approach either the professors or the TA for questions about the project throughout the term. It is greatly advised not to wait until the final week to start the project.