

CS3317 Artificial Intelligence

Final Project: Multimodal Classification Challenge

Team nickname: _____

Member 1: _____ Student ID: _____

Member 2: _____ Student ID: _____

Member 3: _____ Student ID: _____

May 13, 2026

Instructions:

- Fill in your team nickname and the names/IDs of all members (1–3).
- Keep every section heading below; they correspond to the grading rubric.
- **Page limit: 8 pages** (excluding references and the AI-usage table at the end).
- Replace every *italicized prompt* with your actual content; remove the placeholder text in blank tables.
- Delete this instruction box before submission.

Abstract

A 5–7 sentence overview: the task you faced, the method you chose, the single most important design decision, the key result on the leaderboard, and the most significant lesson. Write this last.

1 Introduction

Introduce the project in your own words for a reader who has not seen the handout. Cover at least:

- *The task — what kind of inputs, what kind of outputs, and the metric used to evaluate.*
- *What you find non-trivial about the problem (the missing-modality setting, the few-shot regime, class imbalance, the 100M parameter cap, ...).*
- *A one-sentence preview of your approach and your headline result on the leaderboard.*

2 Method Choice Rationale

In one paragraph: name the methodological approach you chose (early fusion, late fusion, gated fusion, semi-supervised, LLM-augmented, ensemble, ...) and why. Critically, name at least one alternative you considered and rejected, and explain why. This is the section the grader looks at first.

3 Approach

3.1 Pipeline overview

Describe your full prediction pipeline. A clearly-labeled figure helps. Reminder: your model parameters at inference time must total $\leq 100M$.

Figure 1: Your end-to-end pipeline from inputs to the 18-class multi-hot output.

3.2 Handling missing modalities

Test-time inputs have a probability of having one modality (image or text) zeroed. Describe how your method handles this: e.g. train-time modality dropout simulation, modality-specific gates, mask-aware fusion, ensemble of single-modality experts, imputation, etc.

3.3 Use of the unlabeled pool (if any)

You have 5,000 unlabeled examples. State whether and how you used them: pseudo-labeling (with what confidence filter?), contrastive pre-training, FixMatch, none, ...

3.4 Use of LLM APIs (if any)

Disclose any LLM API calls used in your pipeline: provider, model, where in the pipeline (training-time data augmentation, inference-time fallback for hard examples, ...), approximate token volume, and total cost in CNY. Also list each call's purpose in the AI Usage table at the end of the report.

4 Empirical Analysis

4.1 Per-modality contribution

Report your method's F1-macro when only image features, only text features, or both are available at inference. Discuss which modality carries more signal for your method.

Table 1: Per-modality ablation on the validation set.

Modalities available	F1-macro	F1-micro
image only		
text only		
both		

4.2 Missing-modality robustness

Stratify your test predictions by the dropout mask and report F1-macro on each of the three subsets. Briefly discuss which subset is your method's weakest and why.

Table 2: F1-macro stratified by the test-time dropout mask.

Subset (mask)	N examples	F1-macro
intact (has_image=1, has_text=1)		
image dropped (has_image=0, has_text=1)		
text dropped (has_image=1, has_text=0)		

4.3 Additional ablations

Anything else you ablated: hyperparameters, fusion variants, text-encoder choice, semi-supervised schedule, class-weighted vs. vanilla BCE, per-class threshold tuning, ensemble size, ... One subsection per ablation; a small table or figure for each.

5 Final Results

Report the best F1-macro your team achieved on the leaderboard. Mention the timestamp of that submission and your final leaderboard rank if known.

Table 3: Best leaderboard submission.

Metric / property	Value
F1-macro (leaderboard)	
F1-micro (leaderboard)	
Submission timestamp	
Total submissions used	/ 50
Final rank (if known)	

6 Resources Used

6.1 Compute

Hardware (GPU model, count, hours), wall-clock training time for your best model, and any notable engineering tricks (mixed precision, gradient checkpointing, ...).

Table 4: Compute resources used across the project.

Resource	Value
Hardware (GPU)	
GPU-hours for the final model	
Total GPU-hours across all experiments	
Approximate electricity / cloud cost (CNY)	

6.2 LLM API usage (optional — only if you called an LLM API)

*Skip this subsection if you did not call any LLM API in your pipeline. If you did, list per provider/model an approximate count of requests, input/output token volume, and total cost in CNY. **Estimates are fine** — there is no need to track every call. Match the items you disclose in the AI Usage section.*

Table 5: LLM API usage (approximate). Skip if not applicable.

Provider / model	Requests	Input tok.	Output tok.	Cost (CNY)
<hr/>				
Total				

6.3 Team contribution breakdown

For each team member, describe what they actually did. Be specific: which files, which experiments, which figures, which sections of the report. The **Approx. %** column shows each member's share of the total work; the column must sum to 100%. Single-person teams can write "solo".

Table 6: Per-member contribution.

Member	Approx. %	What they did
Member 1 (...)		e.g. implemented the model in <code>model.py</code> ; ran the experiments in §4.1; produced Fig. 1; wrote §2 and §3.
Member 2 (...)		e.g. wrote the data-loading and evaluation pipeline; ran the experiments in §4.2 and §4.3; produced Fig. 2; wrote §4.
Member 3 (...)		e.g. wrote the inference code; produced Fig. 3 and Table 4; wrote §1, §6, and the AI-usage table.
Total	100%	

7 Discussion

What surprised you? What would you do differently with more time? What did the $\leq 100M$ parameter cap force you to learn? Did your intuition about which modality is more useful match the empirical ablation results?

8 AI Usage Disclosure

Per the academic-integrity rule, list **every** AI tool used in this project and **exactly what each accomplished**. “I used ChatGPT” or “some parts were AI-assisted” do not count as disclosure. If a Q&A question reveals that you cannot explain a component you claimed as your own work, that is a violation.

Table 7: AI usage. Add one row per tool / per distinct use.

Tool / model	Where	What it accomplished
e.g. Cursor	<code>load.py</code> , <code>train.py</code>	wrote data-loading boilerplate and argparse plumbing; we wrote the model and training loop.
e.g. Claude API	debugging	one chat session to track down a numpy shape mismatch in our training loop. ~ 3 turns, \sim CNY 0.20.
e.g. GPT-5	§ABSTRACT, §DISCUSSION	helped revise wording; the methodology and analysis sections are written by us.

References

[1] Last name, First name. “Title of the paper.” *Conference / Journal*, year.