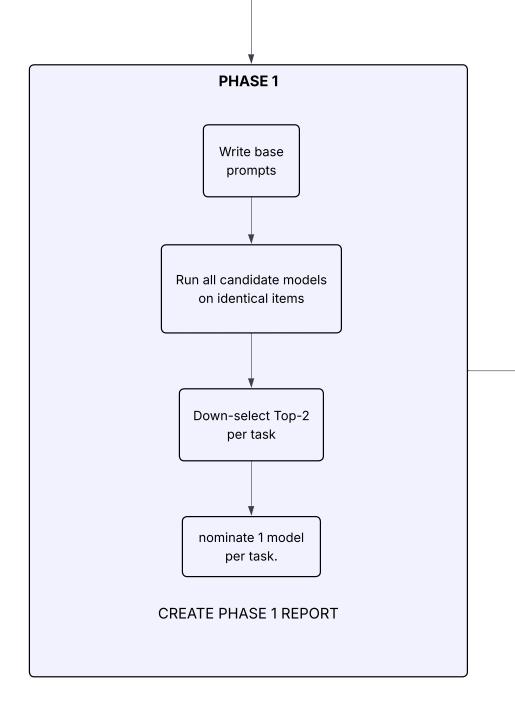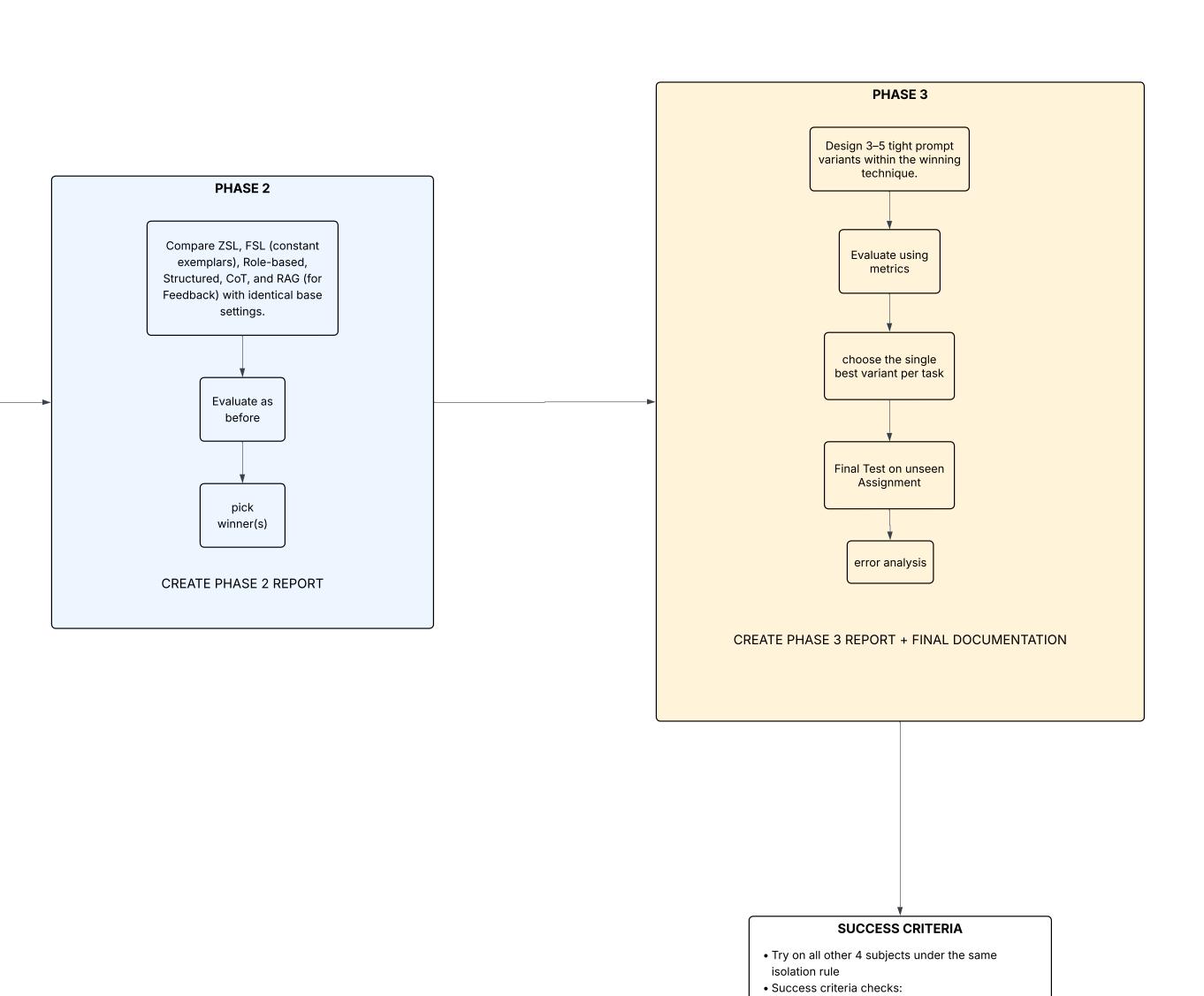**SETUP**

- Form subteams & assign leads
- Choose subject and lock Assignment for training
- Prepare data & repos
- Set fixed items (base prompts and frozen evaluation prompts).
- Define metrics per task
- Create human rubric rating GenAI rating criteria

PHASE 1

Write base prompts

↓

Run all candidate models on identical items

↓

Down-select Top-2 per task

↓

nominate 1 model per task.

CREATE PHASE 1 REPORT

PHASE 2

Compare ZSL, FSL (constant exemplars), Role-based, Structured, CoT, and RAG (for Feedback) with identical base settings.

↓

Evaluate as before

↓

pick winner(s)

CREATE PHASE 2 REPORT

PHASE 3

Design 3–5 tight prompt variants within the winning technique.

↓

Evaluate using metrics

↓

choose the single best variant per task

↓

Final Test on unseen Assignment

↓

error analysis

CREATE PHASE 3 REPORT + FINAL DOCUMENTATION

SUCCESS CRITERIA

- Try on all other 4 subjects under the same isolation rule
- Success criteria checks:
  - improvement vs baselines;
  - human and GenAI correlation;
  - good metric results;
  - feasible cost/latency.