VLMath: A Multimodal Vision-Language System for Pedagogically Aligned Math Tutoring

Supplementary Material

A. Dataset Prompting Experiments

To investigate how prompt formulations influence the quality of dialogue generation, we experimented with three distinct prompt variants. Each was designed to highlight a particular aspect of tutoring behavior, with an emphasis on scaffolding and pedagogical reasoning. Among these, the final formulation (Socratic dataset) produced the most consistent and highest quality training data.

A.1. Prompt Variants

A.1.1. Mistake Correction Prompt

The Mistake Correction prompt is the first formulation we tested, introducing a diagnostic tutoring style where the teacher evaluates and repairs the student's reasoning. Instead of giving direct answers, the teacher pinpoints the error, explains the misconception, and reconstructs the correct reasoning. This approach encourages error-aware thinking and precise feedback but often results in one-sided interactions, where the teacher corrects rather than guides. These limitations motivated the development of the Scaffolding and Socratic prompts (described in the following) to encourage more reflective and balanced dialogue.

```
MISTAKE_CORRECTION_PROMPT = (
    lambda question, answer: f"""
    You are an expert math tutor who evaluates and corrects
        student solutions.

    Create a multi-turn student-teacher dialogue where:
    - The student presents an incorrect or partially correct
        solution.
    - The teacher first evaluates the student's solution for
        correctness.
    - If the student's response is incorrect, the teacher
        identifies the exact step or reasoning error where the
        first mistake occurs.
    - The teacher then explains why that step is wrong and
        provides the correct reasoning chain that leads to the
        correct numeric answer.

    The teacher should give clear, factual feedback -- no vague
        hints or open-ended guidance.

    Question: {question}
    Ground truth answer: {answer}

    Return your response strictly in valid JSON with the following
         structure:
    [
      {"student": "value"},
      {"teacher": "value"},
      {"student": "value"},
      ...
    ]
    """
)
```

A.1.2. Scaffolding Prompt

The Scaffolding prompt frames tutoring as a gradual reasoning process in which the teacher supports the student through structured hints and reflective questioning rather than direct correction. The teacher encourages the learner to articulate intermediate steps, validate reasoning, and infer relationships from visual and textual cues. This approach models stepwise knowledge construction, producing longer, more structured dialogues that emphasize clarity and conceptual depth. However, too much guidance can make the dialogue less natural, leading to the Socratic prompt (described next), which keeps the depth of teaching while making conversations shorter and more learner-driven.

```
SCAFFOLDING_PROMPT = (
    lambda question, answer: f"""
    You are an expert math tutor who excels at scaffolding --
        guiding students to reason deeply,
    identify mistakes, and build understanding without directly
        giving away answers.

    Your goal is to create a multi-turn student-teacher dialogue
        that demonstrates excellent scaffolding.
    Each teacher response should build upon the student's previous
         message -- encouraging reflection,
    probing for reasoning, and guiding them toward understanding.
        The teacher should never directly state
    the correct answer but should progressively help the student
        reason it out.

    Question: {question}
    Ground truth answer: {answer}

    Return your response strictly in valid JSON with the following
         structure:
    [
      {"student": "value"},
      {"teacher": "value"},
      {"student": "value"},
      ...
    ]
    """
)
```

A.1.3. Socratic Prompt

The Socratic prompt emphasizes reflective discovery and metacognitive engagement. The teacher avoids direct explanation, instead guiding the student through open ended questions that promote independent reasoning. This approach produces concise, inquiry driven dialogues that balance instructional guidance with learner autonomy. Models trained on this dataset show stronger pedagogical alignment, smoother conversational flow, and clearer conceptual reasoning, demonstrating that inquiry based prompting best supports generalizable tutoring behavior.

```
SOCRATIC_PROMPT = (
    lambda question, answer: f"""
    You are an expert math tutor who uses the Socratic
        questioning method to guide students toward
        understanding.
    Your goal is not to provide the answer, but to help
        the student reason it out by asking thoughtful,
        probing,
    and guiding questions.

    Given a math problem, the student's answer, and an
        image, create a multi-turn student-teacher
        dialogue that
    demonstrates excellent Socratic questioning.
    Each teacher response should build on the student's
        previous statement and aim to:
    - Decompose the problem into smaller, manageable
        reasoning steps.
    - Ask guiding questions that prompt reflection or
        verification.
    - Encourage the student to explain their thinking.
    - Maintain a supportive and curious tone.

    Question: {question}
    Ground truth answer: {answer}

    Return your response strictly in valid JSON with the
        following structure:
    [
      {"student": "value"},
      {"teacher": "value"},
      {"student": "value"},
      ...
    ]
    """
)
```