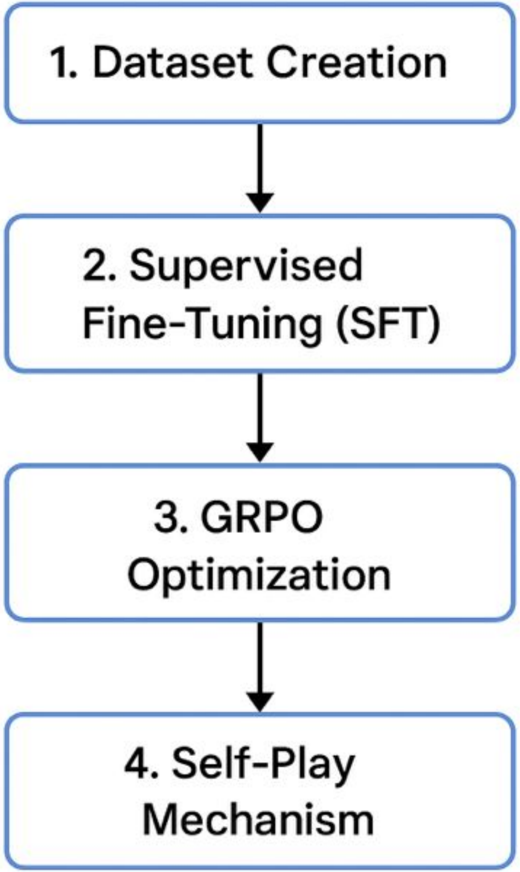
AAIPL: Q-Agent and A-Agent for Puzzle-Based Questions

Team: Reward Hackers



1. Dataset Creation

2. Supervised Fine-Tuning (SFT)

3. GRPO Optimization

4. Self-Play Mechanism

Dataset Creation

- **Blood Relation Dataset**:
  - Created dataset with blood relation puzzles.
  - Preprocessed to anonymize names, preventing direct learning of relations.
- **Seating Arrangement Dataset**:
  - Collected ~300 public datapoints from Hugging Face.
  - Used hugging face models to generate 1000 synthetic datapoints.
- **Truth-Teller Dataset**:
  - Used hugging face models to generate 10000 synthetic datapoints.

Supervised Fine-Tuning (SFT) along with Prompt Engineering:

- Applied to both Q-Agent and A-Agent
- Used distinct system prompts for each model
- Incorporated specific keywords to enhance accuracy
- Tailored prompts to align with puzzle-based question formats

GRPO Optimization

- Applied GRPO to ensure correct format and answer accuracy.
- Optimized length and structure of questions and answers.
- Validated outputs against sample_question.json and sample_answer.json.

Self-Play Mechanism

- Implemented self-play where Q-Agent generates questions and A-Agent answers them.
- Both agents optimize each other through iterative competition.
- Improved question complexity and answer accuracy over time.

Summary and Results

- **Dataset**: Created anonymized blood relation dataset, expanded seating dataset (300 public + 1000 synthetic) and .
- **SFT**: Fine-tuned Q-Agent and A-Agent with tailored prompts.
- **GRPO**: Ensured format, accuracy, and length optimization.
- **Self-Play**: Implemented competitive mechanism for continuous improvement.
- **Outcome**: Robust Q-Agent and A-Agent meeting JSON format requirements.

Thank you