

AMD AI SPRINT: HACKATHON & WORKSHOP SERIES

Team: RBCCPS

Members: 1) Suyog Jare

2) Snehashis Dhibar

3) Basu Hela

Bengaluru | July 19-20



Advanced Data Generation Framework

Our framework revolutionizes training dataset creation for reasoning domains, combining advanced algorithms with domain-specific optimizations. It significantly outperforms traditional template-based methods.

Multi-Domain Generator System

Unified framework supporting multiple reasoning domains with shared optimization principles and cross-domain pattern recognition.

Intelligent Arrangement Sizing

Dynamic sizing (6-10 people min.) based on difficulty, forcing genuine spatial reasoning and preventing overfitting.

Seating Arrangement Generator Innovations

1

Hierarchical Difficulty-Driven Generation

Progressive complexity scaling across four tiers: Foundational, Intermediate, Advanced, and Expert (multi-constraint logic).

2

Multi-Constraint Satisfaction System

Complex constraint intersection algorithms with dynamic validation and intelligent fallback mechanisms.

3

Unified Spatial Reasoning Engine

Single engine handling both linear and circular arrangements with type-specific optimizations and context-aware logic.

4

Advanced Transitive Reasoning

Multi-step logical chain construction, variable chain lengths, and automatic consistency verification.

Truth-Liar Generator & Cross-Domain Innovations

Truth-Liar Generator

- Robust Language Generation: Rich synonym databases (10+ variations), varied verbs, context-sensitive language.
- Massive Name Diversity: 200+ diverse names, systematic rotation, cultural diversity.
- Advanced Constraint Solving: Optimized algorithms, bitwise operations, unique solution enforcement.
- Adaptive Statement Generation: Strategic type selection, balanced distribution, progressive complexity.
- Intelligent Compression: Multi-level techniques maintaining semantic integrity, dynamic token budget.

Cross-Domain Innovations

- Smart Distractor Generation: Context-aware, plausibility-based selection from problem participants.
- Dynamic Token Management: Real-time counting, adaptive compression, combined validation.
- Answer Distribution Balancing: Active tracking, intelligent placement, statistical monitoring.
- Comprehensive Quality Assurance: Multi-stage validation, real-time compliance, intelligent retries.
- Pattern Coverage Optimization: Systematic taxonomy, balanced distribution, edge case inclusion.

Fine-Tuning Strategies



Generated 30000 samples



Used it to perform
Supervised Finetuning of
Qwen-3-4B model



Applied LoRA for parameter
efficient finetuning

Fine-Tuning Implementation

Data Preparation

- Made sure all examples are MCQ
- 80-20 train and validation split

Model Configuration

- Fine-tuned the model for 5 epochs
- Optimizer=Adam
- Cosine Learning rate scheduler

Evaluation Metrics

- Mean Token Accuracy

Prompt Fine-Tuning Techniques

1 Role-Representative Prompts

Used role-representative prompt styles to make the model act like an expert in the given domain. This increases the chances of generating more relevant and insightful responses.

2 Chain-of-Thought Techniques

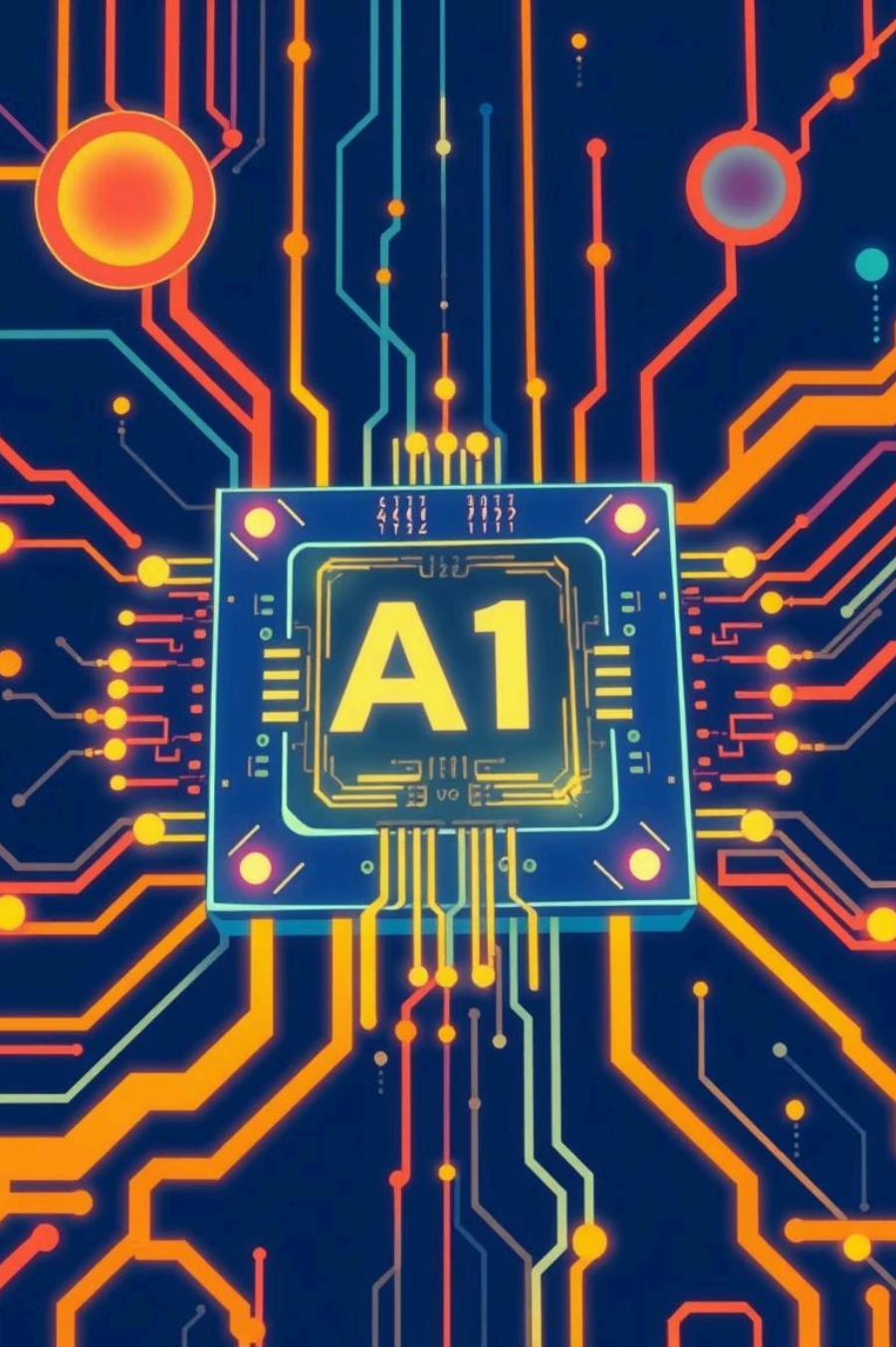
Incorporated Chain-of-Thought (CoT) techniques in both the Answer and Question agents. The Answer agent analyzes the question step-by-step, first identifying direct relationships, then connecting indirect ones, and finally drawing a conclusion.

3 Topic-Specific Approach

The Question agent takes a topic-specific approach, starting with an in-depth analysis of the topic and then increasing the complexity by blending it with more advanced scenarios.

4 Multiple Retries

Used multiple retries for the Answer agent, allowing it to correct and refine its responses over several iterations, up to a maximum of 3.



Thank You