

Architecture-Core Components and System Design

THREE SPECIALIZED AGENTS:

GENERATOR (G)

- INPUT → QUESTION (Q)
- OUTPUT → INITIAL SOLUTION
- ROLE → SOLUTION GENERATION

VERIFIER (V)

- INPUT → GENERATOR OUTPUT + QUESTION
- OUTPUT → QUALITY FEEDBACK
- ROLE → CRITICAL EVALUATION

REFINEMENT MODEL (R)

- INPUT → ALL PREVIOUS OUTPUTS
- OUTPUT → FINAL REFINED ANSWER
- ROLE → SOLUTION IMPROVEMENT

⇒ FLOW:

**QUESTION → G → SOLUTION → V → FEEDBACK → R →
FINAL ANSWER**








<https://www.linkedin.com/in/nimit-kothari/>

02





Training Process

MALT Training Methodology

Data Generation:

- Sampling Strategy:
 -  Tree-based sampling
 -  n^3 trajectory generation
 -  Exponential solution space
- Value Attribution:
 - / Binary rewards
 -  Backward value propagation
 -  $\theta = 0.5$ threshold

Training Pipeline

1.  **Initial Dataset Collection:** • Raw data preprocessing • Question-answer pairs setup • Quality filtering
2.  **Trajectory Expansion:** • Branching factor n application • Multiple solution paths generation • Search space exploration
3.  **Credit Assignment:** • Value propagation through tree • Performance attribution • Role-specific feedback
4.  **Model-Specific Training:** • Individual agent optimization • Role specialization • Capability enhancement






Technical Implementation

Implementation Details & Algorithms





Credit Assignment Strategy:

1. Value Functions:




-  $V(r_{i,j,k,l}) = R(r_{i,j,k,l}) \in \{0, 1\}$
-  $V(v_{i,j,k}) = E[V(r_{i,j,k,l})]$
-  $V(g_{i,j}) = E_k[V(v_{i,j,k})]$

2. Binarization Process:

- $\theta = 0.5$ threshold
-  Values $> 0.5 \rightarrow$ correct
-  Values $\leq 0.5 \rightarrow$ incorrect



Training Methods:

-  SFT (Supervised Fine-Tuning)
-  DPO (Direct Preference Optimization)
-  LoRA adaptation

04

NIMIT KOTHARI

**SAVE THIS POST IF
YOU FIND IT USEFUL**



<https://www.linkedin.com/in/nimit-kothari/>