Association for the Advancement of Artificial Intelligence

Learning Collaborative Reasoning Strategies Through Trust-Weighted Multi-Agent Consensus

EKbana

Problem

Individual LLMs can be Unreliable

On multi- step or high-stakes problems, individual models can produce inconsistent or incorrect answers due to inherent biases or domain-specific knowledge gaps.

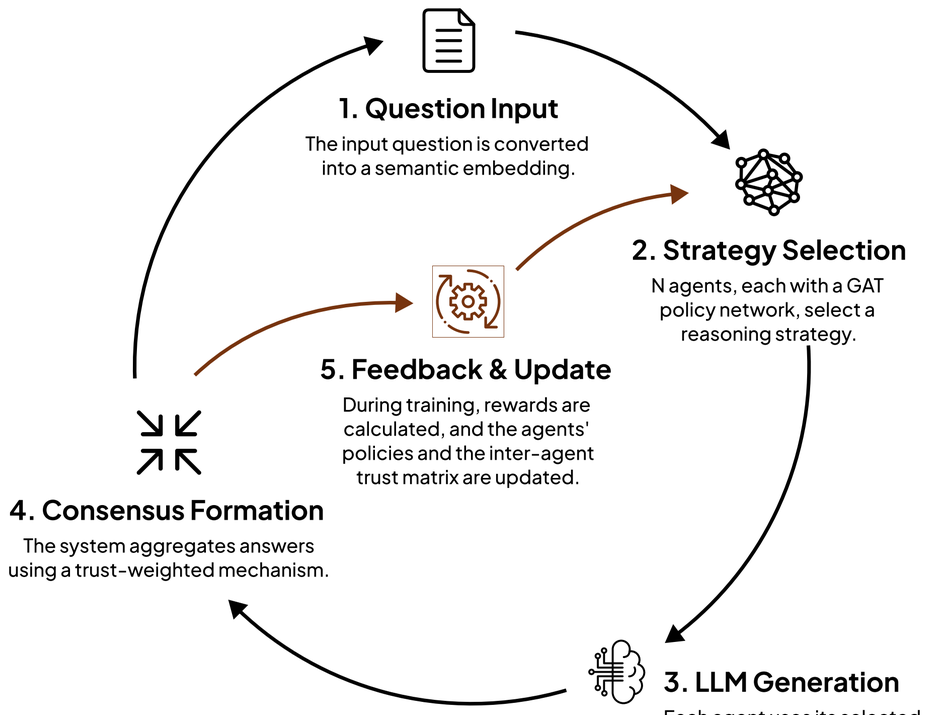
- Brittleness
- Variability
- Limited Dependability

Can LLMs Reason Better by Trusting Each Other?

Our Solution

MARL-GAT Workflow

Agents reason independently, learn trust, and vote with weights



Results

Consensus Accuracy > Individual Agents
Up to +18% gain on smaller models