

Testing Our System

Dr. Vasili Ganishev

Lead Data Scientist
Automotive Industry, Germany
AI Expert



Challenges



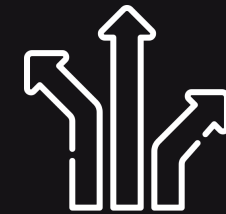
Validation Complexity

AI-generated conversations are difficult to verify with static examples.



Agent Interactions

Multi-agent communication can be complex and unpredictable.



Non-Deterministic Behavior

Agent actions (e.g., SQL query generation) may produce varied results.

Agent Self-Assessment

- Confidence Score
- LLM-as-a-Judge

Hands-on