# AMBER: Agent Meeting BEnchmark fRamework



**Input Data**
- User Queries
- System
- Responses
- Meeting
- Context

Input Processing

**Query Analysis**
- Complexity
- Classification
- Context
- Extraction
- Response Parsing

Query Analysis

**Dual Evaluation Pipeline**

**CompassJudger Module**
Granular Assessment:
- Factual Accuracy (1-10)
- User Need Fulfillment (1-10)
- Response Quality Metrics

**Prometheus Module**
Holistic Evaluation:
- Overall Quality (1-5)
- Comprehensive Feedback

Evaluation Pipeline

**Metrics**
**Quantitative:**
1. Factual Score
2. User Need Score
3. Conciseness

**Qualitative:**
4. Structure Score
5. Completeness
6. Overall Quality
7. Detailed Report

Output Analysis