

# Agent Quality

Authors: Meltem Subasioglu, Turan Bulmus,  
and Wafae Bakkali



## Acknowledgements

### Content contributors

Hussain Chinoy

Ale Fin

Peter Grabowski

Michelle Liu

Anant Nawalgaria

Kanchana Patlolla

Steven Pecht

Julia Wiesinger

### Curators and editors

Anant Nawalgaria

Kanchana Patlolla

### Designer

Michael Lanning



# Table of contents

<b>Introduction</b>	<b>6</b>
<b>How to Read This Whitepaper</b>	<b>7</b>
<b>Agent Quality in a Non-Deterministic World</b>	<b>8</b>
Why Agent Quality Demands a New Approach	9
The Paradigm Shift: From Predictable Code to Unpredictable Agents	11
The Pillars of Agent Quality: A Framework for Evaluation	13
Summary & What's Next	15
<b>The Art of Agent Evaluation: Judging the Process</b>	<b>16</b>
A Strategic Framework: The "Outside-In" Evaluation Hierarchy	17
The "Outside-In" View: End-to-End Evaluation (The Black Box)	18
The "Inside-Out" View: Trajectory Evaluation (The Glass Box)	19
The Evaluators: The Who and What of Agent Judgment	21
Automated Metrics	22
The LLM-as-a-Judge Paradigm	23
Agent-as-a-Judge	25



# Table of contents

Human-in-the-Loop (HITL) Evaluation .....	26
User Feedback and Reviewer UI .....	27
Beyond Performance: Responsible AI (RAI) & Safety Evaluation .....	28
Summary & What's Next .....	30
<b>Observability: Seeing Inside the Agent's Mind .....</b>	<b>31</b>
From Monitoring to True Observability .....	31
The Kitchen Analogy: Line Cook vs. Gourmet Chef .....	31
The Three Pillars of Observability .....	32
Pillar 1: Logging – The Agent's Diary .....	33
Pillar 2: Tracing – Following the Agent's Footsteps .....	36
Why Tracing is Indispensable .....	36
Key Elements of an Agent Trace .....	37
Pillar 3: Metrics – The Agent's Health Report .....	38
System Metrics: The Vital Signs .....	38
Quality Metrics: Judging the Decision-Making .....	39
Putting It All Together: From Raw Data to Actionable Insights .....	41

# Table of contents

Summary & What's Next .....	43
<b>Conclusion: Building Trust in an Autonomous World .....</b>	<b>44</b>
Introduction: From Autonomous Capability to Enterprise Trust .....	44
The Agent Quality Flywheel: A Synthesis of the Framework .....	45
Three Core Principles for Building Trustworthy Agents .....	46
The Future is Agentic - and Reliable .....	47
<b>References .....</b>	<b>49</b>