---

**Expert Evaluation of a Taxonomy and Decision Model for Foundation Model-based Agent Architectures**

**Part 1: Introduction & Instructions**

Thank you for dedicating your valuable time and expertise to this study. Our research addresses the growing architectural challenges in designing and operating systems that leverage Foundation Model-based (FM-based) agents. Given the rapid innovation in this field, there is a significant need for a structured framework to guide architects and developers. To fill this gap, our work introduces a comprehensive taxonomy of architectural options for FM-based agents and an accompanying decision model to navigate design trade-offs. Your expert feedback is crucial for validating the rigor, completeness, and utility of our work. The goal of this evaluation is to assess our proposed taxonomy and decision model against several key criteria:

- **Taxonomy Evaluation Objectives:**
  - Completeness: Does the taxonomy cover the essential architectural dimensions and options for modern FM-based agents?
  - Correctness - Clarity & Structure: Are the categories clear, well-defined, and logically organized?
  - Orthogonality: Are the primary dimensions and categories conceptually independent, minimizing ambiguity and overlap?
- **Decision Model Evaluation Objectives:**
  - Clarity: Does the decision model clearly illustrate the trade-offs between architectural choices and their impact on system qualities?
  - Utility: Is the decision model a practical and useful tool for architects during the design process?

To ensure a smooth evaluation process, please follow these steps:

1. **Review the Core Document:** First, please carefully review the accompanying **"Expert Review Document,"** which presents the complete taxonomy and the Architectural Trade-off Decision Matrix. This document provides the necessary context for all the questions that follow.
2. **Complete the Questionnaire:** After reviewing the document, please proceed to answer the questions in each section of this questionnaire. Your detailed feedback, especially in the open-ended questions, is highly appreciated.
3. **Confidentiality:** Your responses will be kept strictly confidential and used for academic research purposes only. All published results will be anonymized.

We estimate this evaluation will take approximately **20-40 minutes** to complete. We deeply appreciate your contribution to advancing the field of software architecture for AI systems.

---

**Part 2: Overall Assessment of the Taxonomy**

This section asks for your overall impression of the *Functional Capabilities* and *Non-functional Qualities* dimensions of the taxonomy.

**2.1. Functional Capabilities Taxonomy (Completeness)**

a. On a scale of 1 (Strongly Disagree) to 5 (Strongly Agree), please rate your agreement with the following statement: "The Functional Capabilities taxonomy is comprehensive, covering the essential architectural aspects of modern FM-based agents."

[ ] 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5

b. Are there any major functional categories or architectural options you believe are missing? For example, are there any emerging trends from recent research or popular open-source agent frameworks that are not yet captured?

(Please type your answer here)

**2.2. Feedback on Functional Capabilities (Clarity & Structure)**

a. On a scale of 1 (Very Unclear) to 5 (Very Clear), please rate the overall clarity and structure of the Functional Capabilities taxonomy.

[ ] 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5

b. Are there any existing functional categories that are unclear, should be renamed, or could be merged with others?

(Please type your answer here)

**2.3. Non-functional Qualities Taxonomy (Completeness)**

a. On a scale of 1 (Strongly Disagree) to 5 (Strongly Agree), please rate your agreement with the following statement: "The Non-functional Qualities taxonomy covers the critical quality attributes for designing and evaluating FM-based agents."

[ ] 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5

b. Are there any critical non-functional qualities you believe are missing? (e.g., considerations for cost, ethics, interoperability, etc.) For example, are there any emerging trends from recent research or popular open-source agent frameworks that are not yet captured?

(Please type your answer here)

**2.4. Feedback on Non-functional Qualities (Clarity & Structure)**

a. On a scale of 1 (Very Unclear) to 5 (Very Clear), please rate the overall clarity and structure of the Non-functional Qualities taxonomy.

[ ] 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5

b. Are there any existing non-functional categories that are unclear, should be renamed, or could be merged with others?

(Please type your answer here)

---

**Part 3: Evaluation of Orthogonality**

A fundamental principle of a robust taxonomy is orthogonality, meaning its categories should be conceptually independent to avoid ambiguity and ensure clear classification. This section is designed to rigorously test this property within our taxonomy. We have found that evaluating the independence of conceptually adjacent categories is a particularly effective method for identifying potential structural weaknesses.

**Instructions:**

For each taxonomy category listed in the first column below (9 from Functional Capabilities and 9 from Non-functional Qualities), please consider its definition within our taxonomy. In the second column, please identify any other categories from the full list that you believe it conceptually overlaps with. In the third column, please provide a brief rationale for why you perceive an overlap. If you see no significant overlap for a category, you may leave the corresponding row blank.

| **Taxonomy Category** | **Potential Overlapping Categories** | **Rationale for your rating (Optional but helpful)** |
|---|---|---|
| **Functional Capabilities** | | |
| Input Modality | *(Please type your answer here)* | *(Please type your answer here)* |
| Context | | |
| Role-playing | | |
| Goal Type | | |
| Planning | | |
| Memory | | |
| Action | | |
| Reflection | | |
| Learning Capability | | |
| Workflow | | |

| | | |
|---|---|---|
| Access to Underlying Models | | |
| **Non-functional Qualities** | | |
| Performance | | |
| Reliability | | |
| Robustness | | |
| Resilience | | |
| Security | | |
| Maintainability | | |
| Responsible & Trustworthy AI | | |
| Usability | | |
| Compatibility | | |

---

**Part 4: Evaluation of the Decision Model**

This section evaluates the clarity and practical utility of the **Trade-off Decision Matrix**.

**4.1. Clarity of the Decision Model**

a. On a scale of 1 (Very Unclear) to 5 (Very Clear), how clear is the Decision Matrix in illustrating the trade-offs between functional choices and their impact on non-functional qualities?

[ ] 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5

b. Please provide a brief rationale for your clarity rating. What aspects of the matrix are particularly clear or unclear?

(Please type your answer here)

**4.2. Utility of the Decision Model**

a. On a scale of 1 (Not at all Useful) to 5 (Extremely Useful), how useful do you believe this Decision Matrix would be for a software architect when designing a new FM-based agent?

[ ] 1 [ ] 2 [ ] 3 [ ] 4 [ ] 5

b. In what specific design scenarios do you see this matrix being most or least useful? For instance, are there certain types of agent systems or application domains where you believe this matrix would be less applicable?

(Please type your answer here)

**4.3. Practical Application Scenario**

You are designing a system with a team of AI Analyst Agents to process thousands of online customer reviews for a new product. The core task for each agent is to use a Foundation Model to read a review, understand its sentiment, and categorize the feedback (e.g., "positive feedback on battery life," "negative feedback on screen brightness"). The system has two high-priority requirements:

- Reliability: The system must be highly fault tolerant. The failure of any single agent must not halt the overall analysis process; other agents must continue to operate and complete the task.
- Maintainability: The system's overall state must be easy to understand, monitor, and debug. A developer should be able to quickly diagnose issues (like a stalled or failed agent) without the complexity of inspecting every agent individually.

Based on the Decision Model, which Agent Coordination option (e.g., Centralized, Federated, Fully Distributed) presents the most challenging trade-off for this scenario, and why? Please briefly explain how the matrix helped your reasoning.

(Please type your answer here)

**4.4. Suggestions for Improvement**

Do you have any suggestions for improving the clarity or practical utility of the Decision Matrix?

(Please type your answer here)

---

**Part 5: Concluding Remarks**

**5.1. General Feedback**

Do you have any other comments, suggestions, or feedback on this research?

(Please type your answer here)

**Thank you once again for your invaluable contribution.**

---