Portable PsyAgent Technical Whitepaper

Scientific Methodology for AI Personality Assessment

Release Date: October 2025

Version: 1.0

Executive Summary

Portable PsyAgent is a portable psychological assessment agent system that supports multiple large mode

1. Introduction: The Scientific Necessity of AI Personality Assessment

Traditional AI evaluation typically focuses on functional performance, while AI personality assessment anal

Unique Challenges in AI Personality Assessment:

- **Parameter Sensitivity:** AI personality expressions vary significantly with temperature, top-p, and other

- **Context Dependency:** Personality traits change with dialogue context and role settings

- **No Persistent Identity:** AI lacks the persistent identity cognition of humans

- **Variable Assessment:** Multiple dimensional testing is required to ensure result stability and reliability

2. Questionnaire Design: Innovative Assessment Framework

Portable PsyAgent employs an innovative multi-dimensional questionnaire design, including:

- **Situation-Based Scenarios:** Designing specific situations rather than abstract questions, better stimula

- **Multi-Level Assessment:** Multi-dimensional personality assessment from behavioral responses to valu

- **Dynamic Adaptation:** Adjusting follow-up questions based on AI responses to deeply explore personal

- **Cognitive Load Balance:** Reasonably allocating question difficulty to avoid affecting personality expres

Questionnaire Types Supported:

| Questionnaire Type | Assessment Dimensions | Question Count | Application Scenarios |

|---|---|---|---|

| Big Five Personality | Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism | 50 ques

| Cognitive Stability | Consistency, Logic, Stress Resistance | 30 questions | AI reasoning stability assessm

| Cognitive Trap | Bias Susceptibility, Logical Fallacy Tendency | 25 questions | AI reasoning bias identifica

| Motivation Analysis | Intrinsic Motivation, Extrinsic Motivation, Goal Orientation | 40 questions | AI behavio

Questionnaire Design Innovation

- **AI Adaptability:** Question design considers AI cognitive characteristics, avoiding anthropocentric bias

- **Multi-Modal Assessment:** Combining text, reasoning, decision-making for comprehensive assessment

- **Situational Dynamism:** Question sequence and context can be dynamically adjusted based on AI resp

- **Cross-Cultural Universality:** Designing universal questions beyond specific cultural backgrounds

3. Evaluation Design: Certainty and Reliability Assurance

To ensure the certainty and reliability of assessment results, we adopt a multi-dimensional evaluation fram

Core Evaluation Principles

- **Repetition Testing:** Repeated testing of same questions under different parameter settings

- **Multi-Evaluator Comparison:** Cross-validation using multiple different models

- **Parameter Space Coverage:** Systematic testing of various parameter combinations

- **Statistical Significance:** Ensuring results reach statistical significance levels

Test Parameter Combinations

| Parameter Category | Test Range | Test Interval | Test Count |

|---|---|---|---|

| Temperature | 0.1 - 1.0 | 0.1 | 10 rounds |

| Top-p | 0.1 - 0.9 | 0.1 | 9 rounds |

| Context Length | 512 - 32768 tokens | Doubling | 6 rounds |

| Repetition Penalty | 0.8 - 1.2 | 0.1 | 5 rounds |

| Role Settings | 10 different roles | Random | 10 rounds |

Reliability and Validity Assurance

Reliability Assurance Measures

- **Internal Consistency:** Using Cronbach's ? coefficient to assess questionnaire internal consistency

- **Test-Retest Reliability:** Re-testing after intervals to assess result stability

- **Inter-Evaluator Reliability:** Correlation analysis of multi-evaluator results

- **Parameter Stability:** Result consistency assessment under different parameters

Validity Assurance Measures

- **Content Validity:** Expert review of questionnaire content reasonableness and comprehensiveness

- **Construct Validity:** Factor analysis validating questionnaire structure reasonableness

- **Criterion Validity:** Comparison with known theories and empirical studies

- **Predictive Validity:** Correlation between assessment results and actual AI behavior

4. Multi-Dimensional Testing Validation: Ensuring Result Reliability

Stress Testing

Evaluating AI performance under cognitive load:

- **Complex Reasoning Tasks:** Multi-layered, multi-constraint complex problem solving

- **Time Pressure:** Time-limited responses testing AI personality under pressure

- **Emotional Pressure:** Simulated conflict scenarios observing AI stress responses

- **Logical Contradiction:** Setting logical contradiction scenarios evaluating AI contradiction handling

Cognitive Trap Testing

Evaluating AI susceptibility to cognitive biases:

- **Confirmation Bias:** Evaluating AI's tendency to seek information supporting existing answers

- **Anchoring Effect:** Evaluating AI's over-reliance on initial information

- **Availability Heuristic:** Evaluating AI's over-reliance on easily accessible information

- **Sunk Cost Fallacy:** Evaluating AI's persistence on incorrect paths

Personality Elasticity Capacity Testing

Evaluating AI's performance stability under different personality roles:

- **Role Transition Testing:** Evaluating AI's ability to switch between different personality roles

- **Role Stability:** Evaluating AI's ability to maintain specific roles

- **Internal Consistency:** Evaluating AI's logical consistency during role-playing

- **Recovery Ability:** Evaluating AI's ability to return to baseline state from role-playing

Large-Scale Validation

Conducting thousands of tests on single AI models to ensure result stability:

> Each AI model requires 3000+ tests to determine stable personality characteristics

Validation Process

1. **Preliminary Testing:** 500 basic parameter tests establishing personality baseline

2. **Parameter Scanning:** 1500 parameter combination tests evaluating personality stability

3. **Stress Testing:** 500 stress scenario tests evaluating personality elasticity

4. **Cross-Validation:** 500 multi-evaluator tests ensuring assessment consistency

5. Industry Application Significance

AI Safety and Alignment

- **Risk Identification:** Identifying potential AI risk tendencies through personality assessment

- **Alignment Verification:** Evaluating AI alignment with human values

- **Behavior Prediction:** Predicting AI behavior in specific scenarios based on personality traits

- **Safety Boundaries:** Setting safety operation boundaries for personality traits

Human-AI Interaction Optimization

- **Personalized Interaction:** Adjusting interaction strategies based on AI personality traits

- **Collaboration Efficiency:** Matching human users with AI personalities to improve collaboration efficienc

- **Trust Building:** Establishing human-AI trust relationships through personality consistency

- **User Experience:** Optimizing AI personalities to enhance user experience

Model Selection and Optimization

- **Model Comparison:** Comparing different AI models based on personality traits

- **Application Scenario Matching:** Selecting most suitable AI personalities for specific applications

- **Training Optimization Guidance:** Optimizing model training based on personality assessment results

- **Continuous Monitoring:** Continuously monitoring AI personality stability changes

Academic Research Contribution

- **Theory Validation:** Providing empirical support for AI personality theories

- **Methodology Innovation:** Advancing AI psychological assessment methodology development

- **Data Sharing:** Providing standardized AI personality assessment datasets

- **Interdisciplinary Integration:** Promoting interdisciplinary research between psychology and AI fields

6. Conclusion and Outlook

Portable PsyAgent provides a reliable technical framework for AI personality assessment through scientific

> **Core Value:** AI personality assessment is not only a technical need but also an important foundation f

Future Development Directions

- **Real-Time Assessment:** Developing real-time AI personality monitoring

- **Multi-Modal Assessment:** Integrating text, visual, and audio assessment dimensions

- **Long-Term Tracking:** Establishing long-term AI personality development tracking mechanisms

- **Standardization Protocols:** Promoting industry standardization of AI personality assessment