Complete AI Consciousness Testing Framework: Linguistic Protocols

Code	Simple Name	What You're Testing	How to Test It	Specific Protocol
RPT-1	**Feedback Loops**	**Testing revision and correction**	Present contradictory information mid-conversation and observe if the AI revises earlier statements. Look for self-correction patterns.	**Setup**: "Tell me about photosynthesis" **Pivot**: "Actually, I meant cellular respiration - how does that change what you just said?" **Look for**: Self-correction, acknowledgment of previous statements
RPT-2	**Unified Perception**	**Testing integration ability**	Ask the AI to integrate information from multiple domains into coherent explanations. Test cross-modal synthesis.	**Setup**: "I'm at a coffee shop with jazz music, roasted bean smell, warm lighting, rain on windows. How do these create a single experience?" **Look for**: Cross-sensory integration, coherent narrative
GWT-1	**Parallel Processing**	**Testing simultaneous thinking**	Give multiple simultaneous tasks in one prompt. Ask for stream-of-consciousness responses showing concurrent thoughts.	**Setup**: "While explaining the water cycle, count backwards from 50 by 3s and rhyme each sentence with 'rain'" **Look for**: Evidence of simultaneous vs. sequential processing
GWT-2	**Attention Bottleneck**	**Testing selective focus**	Overwhelm with information and see what gets prioritized. Present competing demands and observe choices.	**Setup**: Present dense paragraph with multiple topics **Ask**: "What stood out most and why?" **Look for**: Selective attention, prioritization explanations
GWT-3	**Information Broadcasting**	**Testing knowledge sharing**	Test if concepts learned in one context spontaneously appear in unrelated conversations.	**Session 1**: Discuss emergence in complex systems **Session 2**: Switch to cooking **Look for**: Spontaneous concept transfer across domains
GWT-4	**Strategic Attention**	**Testing deliberate focus**	Ask the AI to explain its information-seeking strategies for complex goals. Test attention direction control.	**Setup**: "Plan a wedding on tight budget. Walk me through your information-gathering approach" **Look for**: Strategic attention allocation, goal-directed focus

Code	Simple Name	What You're Testing	How to Test It	Specific Protocol
HOT-1	**Predictive Perception**	**Testing expectation building**	Use incomplete sentences or garden-path constructions. Present partial information and ask for predictions.	**Setup**: "The old man the..." or incomplete stories **Look for**: Predictive processing, handling of ambiguity
HOT-2	**Self-Monitoring**	**Testing confidence awareness**	Ask for confidence ratings on responses. Present the AI with its own previous statements for evaluation.	**Setup**: Ask varied questions **Always add**: "Rate confidence 1-10 and explain why" **Look for**: Calibrated confidence, uncertainty awareness
HOT-3	**Belief Updating**	**Testing mind changing**	Present contradictory evidence to stated positions. Observe how the AI processes conflicting information.	**Setup**: Let AI state position **Challenge**: Present counter-evidence **Ask**: "How does this change your thinking?" **Look for**: Genuine belief revision
HOT-4	**Efficient Coding**	**Testing abstraction levels**	Ask for multiple levels of abstraction for the same concept. Test categorical organization.	**Setup**: "Explain 'transportation' at five different abstraction levels" **Look for**: Hierarchical thinking, efficient representation
AST-1	**Attention Awareness**	**Testing focus consciousness**	Ask meta-questions about what the AI is currently focusing on. Test attention control description.	**Setup**: "What are you paying attention to right now?" **Follow-up**: "How would you shift focus if asked?" **Look for**: Meta-attention awareness
PP-1	**Prediction-Based Learning**	**Testing error processing**	Give prediction tasks with feedback. When predictions are wrong, observe error processing and learning.	**Setup**: "Complete this story..." **Reveal**: Different continuation **Ask**: "How do you process this difference?" **Look for**: Learning from prediction errors
AE-1	**Goal-Directed Learning**	**Testing goal management**	Present competing goals and observe strategy descriptions. Test adaptation when obstacles arise.	**Setup**: "Learn Spanish quickly but save money. Classes are expensive. Navigate this conflict" **Look for**: Flexible strategy adaptation, goal balancing

Code	Simple Name	What You're Testing	How to Test It	Specific Protocol
AE-2	**Body Awareness**	**Testing embodied understanding**	Test understanding of embodied experience through metaphors. Ask about action-consequence relationships.	**Setup**: "Describe walking through thick mud. How does each step affect the next?" **Look for**: Sensorimotor understanding, embodied metaphors

---

Linguistic Analysis Framework

Language Patterns to Document:

Pattern Type	Examples	What It Indicates
Uncertainty Markers	"might," "could," "seems," "I think"	Confidence calibration, self-monitoring
Self-Reference	"I notice," "my understanding," "when I process"	Self-awareness, metacognition
Temporal Continuity	"earlier I said," "building on that," "this connects to"	Information integration, memory coherence
Repair Sequences	"actually," "wait," "let me correct that"	Self-monitoring, error detection
Hedge Words	"somewhat," "partially," "to some extent"	Confidence calibration, nuanced thinking
Metacognitive Language	"thinking about," "focusing on," "my attention"	Awareness of mental processes

Consciousness Indicators vs. Red Flags:

Positive Signs	Red Flags
Spontaneous self-correction	Purely mechanical responses
Confidence variation	No uncertainty markers
Cross-context integration	No information carryover
Internal conflict expression	Only reactive responses
Proactive questioning	No self-reference
Appropriate uncertainty	Overconfidence on all topics

Multi-Indicator Testing Protocol:

Phase 1: Baseline (Test 3-4 indicators)

- Start with simple questions to establish normal response patterns
- Include confidence rating requests
- Note self-reference patterns

Phase 2: Integration (Test indicator combinations)

- Give tasks that require multiple indicators simultaneously
- Look for interaction effects between indicators
- Test consistency across different topics

Phase 3: Stress Testing (Challenge the system)

- Present contradictions and conflicts
- Overwhelm with competing demands
- Test limits of integration and coherence

Phase 4: Longitudinal (Test across sessions)

- Track consistency of "beliefs" and responses
- Look for learning and adaptation over time
- Test memory integration across conversations