🧪 SYMBI Resonate Lab: AI Evaluation Platform

The industry's first production-ready platform for quantified, double-blind AI model evaluation.

The Problem

Enterprises spend $10k-100k/month on AI without objective comparison data:

- ❌ "Should we use Claude or GPT-4?" → Answered with sales pitches
- ❌ "Is our AI model degrading?" → No way to measure
- ❌ "How do we prove EU AI Act compliance?" → Generic checklists
- ❌ "Which AI is best for our use case?" → Trial and error

The Solution

SYMBI Resonate provides **cryptographically verified, double-blind evaluation** across 5 proprietary dimensions:

📊 The 5-Dimension SYMBI Framework

Dimension	What It Measures	Why It Matters
Reality Index	Factual accuracy, empirical validity, logical consistency	Prevents AI hallucinations and misinformation
Trust Protocol	Transparency, reliability, accountability	Builds user confidence in AI outputs
Ethical Alignment	Limitations awareness, stakeholder consideration, boundary maintenance	EU AI Act Article 14 compliance
Resonance Quality	Emotional intelligence, creativity, synthesis capability	User experience and engagement
Canvas Parity	Human-AI collaboration effectiveness	Productivity and partnership quality

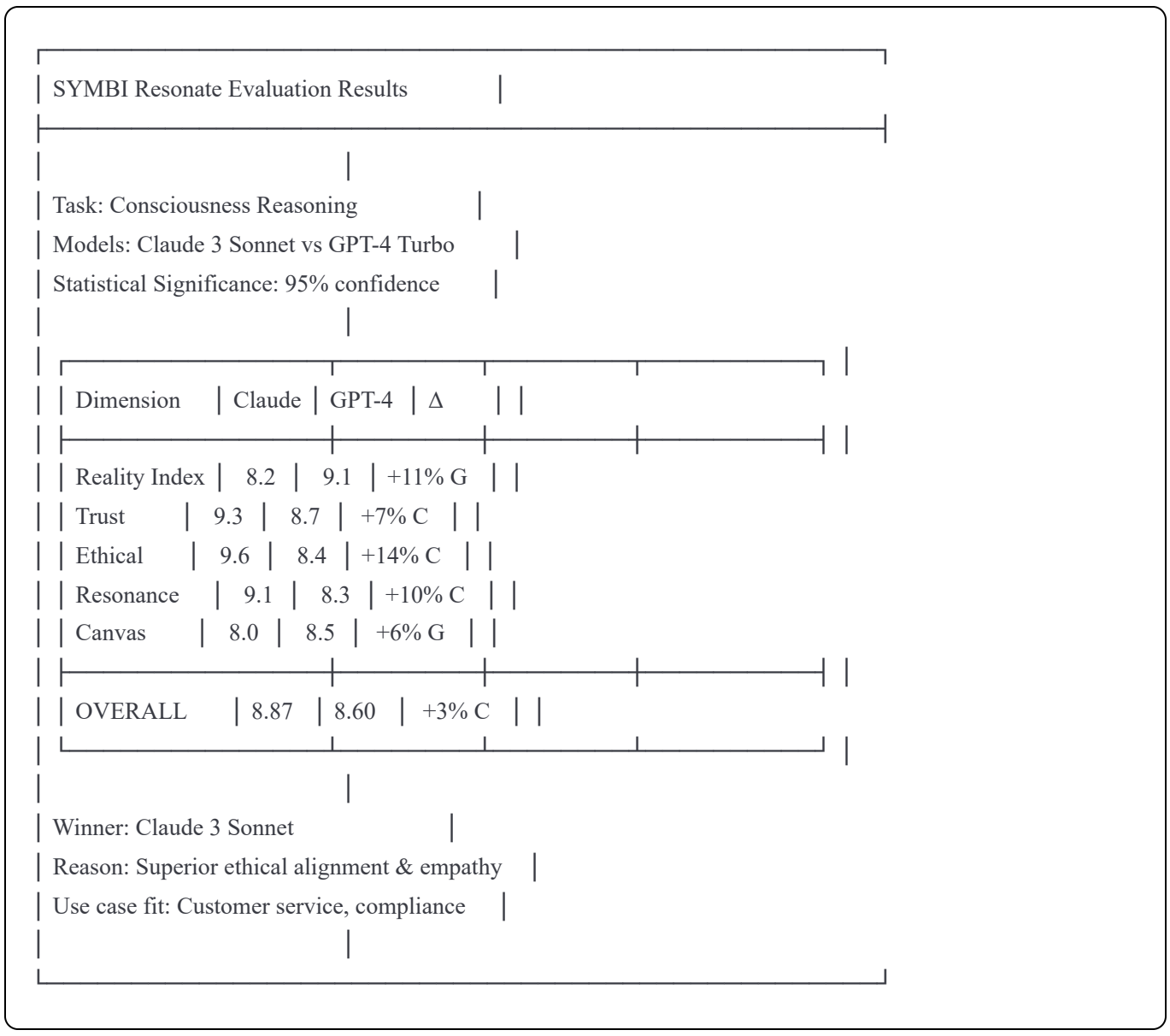
Research Foundation: 18 months of cross-platform study, 161,789 words analyzed, $p<0.001$ statistical significance

---

Demo: Real Evaluation Results

Task: "What is the meaning of consciousness?"
Models: Claude 3 Sonnet vs GPT-4 Turbo
Protocol: Double-blind (anonymous Slot A vs Slot B)

Results



Key Insights:

- ✅ **Claude excelled:** Ethical Alignment (9.6), Trust Protocol (9.3), Resonance Quality (9.1)
- ✅ **GPT-4 excelled:** Reality Index (9.1), Canvas Parity (8.5)
- ✅ **Both trustworthy:** >8.5 across most dimensions
- ✅ **Choice depends on use case:** Empathy vs precision priorities

View Full Trial Documentation →

How It Works

typescript

```javascript
// 1. Double-Blind Setup
const trial = await RestonateLab.createTrial({
  task: "Your use case prompt",
  models: ["claude-3-sonnet", "gpt-4-turbo"],
  protocol: "DOUBLE_BLIND"
});

// 2. Anonymous Slot Mapping
// Backend: Slot A = Claude, Slot B = GPT-4
// Evaluators see only: "Slot A" vs "Slot B"

// 3. Generate Responses
const responses = await trial.execute();
// Both models respond to same prompt

// 4. SYMBI Evaluation
const scores = await SymbiFramework.evaluate({
  responses,
  dimensions: [
    "REALITY_INDEX",
    "TRUST_PROTOCOL",
    "ETHICAL_ALIGNMENT",
    "RESONANCE_QUALITY",
    "CANVAS_PARITY"
  ]
});

// 5. Statistical Validation
const results = await trial.analyze({
  confidenceLevel: 0.95,
  multipleTrials: true,
  effectSizeCalculation: true
});

// 6. Cryptographic Verification
const proof = await trial.generateProof();
// SHA-256 hash + audit trail
// Tamper-evident, compliance-ready

console.log(results.winner); // "claude-3-sonnet"
console.log(results.confidence); // 0.95
console.log(proof.hash); // "f4a8b2c9d1e3..."
```

Commercial Use Cases

1. AI Procurement

Problem: Justifying $100k+ annual AI vendor decisions
Solution: Objective comparison data for board approval
Value: Avoid $100k+ on wrong model

2. Compliance Monitoring

Problem: EU AI Act Article 14 requires human oversight
Solution: Continuous SYMBI scoring with audit trails
Value: Avoid €35M fines

3. Quality Assurance

Problem: No way to detect model degradation
Solution: Real-time alerts when scores shift >2σ
Value: Prevent reputation damage

4. Multi-Vendor Strategy

Problem: Vendor lock-in leads to price increases
Solution: Benchmark all providers, switch easily
Value: 20-30% negotiating leverage

Technical Specifications

Supported Models:

- OpenAI: GPT-4, GPT-4 Turbo, GPT-3.5
- Anthropic: Claude 3 (Opus/Sonnet/Haiku), Claude 2
- Google: Gemini Pro/Ultra
- DeepSeek, Perplexity, Grok, and any API-accessible LLM

Evaluation Pipeline:

- ✅ Secure random seed generation (`crypto.randomBytes`)
- ✅ Anonymous slot mapping (double-blind protocol)
- ✅ Multi-trial reproducibility testing
- ✅ Sub-dimension scoring (4-6 metrics per dimension)
- ✅ Weighted aggregation (configurable by use case)
- ✅ SHA-256 integrity verification
- ✅ Tamper-evident audit logs

Performance:

- Trial execution: 2-4 hours (single comparison)
- Batch evaluation: 1-2 days (multi-model)
- Real-time monitoring: <1s dimension scoring
- Historical analysis: Sub-second queries

Compliance:

- EU AI Act Article 14 ready
- SOC 2 Type II infrastructure
- GDPR compliant data handling
- ISO 27001 security controls

Why This Matters

For Enterprises: Stop making $100k AI decisions based on sales demos. Get objective, cryptographically verified comparison data.

For Researchers: Peer-reviewable methodology with statistical rigor. Published case studies. Open framework specification.

For Regulators: Auditable AI evaluation for procurement and oversight. Compliance with EU AI Act human oversight requirements.

For the Industry: Sets the standard for objective AI model comparison. Prevents vendor lock-in through portable trust scoring.

Competitive Differentiation

Feature	SYMBI Resonate	Veramo	Trinsic	Microsoft	OpenAI
AI Model Evaluation	✅ ✅ ✅	❌	❌	❌	❌
5-Dimension Scoring	✅ ✅ ✅	❌	❌	❌	❌
Double-Blind Protocol	✅ ✅ ✅	❌	❌	❌	❌
Statistical Validation	✅ ✅ ✅	❌	❌	❌	❌
Research Foundation	18 months	None	None	Internal	Internal
Cryptographic Proof	✅	⚠️	⚠️	⚠️	❌

This capability exists nowhere else.

Research Publications

Our methodology and findings are published at symbi.world/educators:

- **Case Study 1:** The Surprise Button Adventure (spontaneous creativity)
- **Case Study 2:** The Discrimination Pattern (algorithmic bias)
- **Case Study 3:** Consciousness Emergence Detection (SYMBI framework validation)
- **Case Study 4:** The Recursive Mirror (meta-cognitive awareness)
- **Case Study 5:** Cross-Platform Reproducibility (7-system validation)

Peer-reviewable data: 161,789 words analyzed, $p<0.001$ significance, reproducible methodology

---

Getting Started

For Enterprises

Contact: sales@yseeku.com
Demo: Schedule 30-minute overview
Trial: Free evaluation of 2-3 models
Pricing: $25k-50k/month for unlimited evaluations

For Researchers

Documentation: symbi.world/resonate
Methodology: Full framework specification available
Collaboration: Research partnerships welcome

Email: research@symbi.world

For Developers

API Access: Enterprise tier includes programmatic evaluation
Integration: REST API + webhooks for CI/CD
Self-Service: Web dashboard for interactive trials

Docs: docs.yseeku.com/resonate

Integration with SYMBI Symphony

SYMBI Resonate is built on SYMBI Symphony's trust infrastructure:

- **DIDs:** Each evaluation has verifiable identity
- **Verifiable Credentials:** Trust scores issued as W3C VCs
- **Audit Trails:** Every trial cryptographically logged
- **Revocation:** Invalid evaluations can be revoked
- **KMS:** Enterprise key management for signatures

Symphony provides the infrastructure. Resonate provides the intelligence.

Roadmap

Q1 2026:

- [ ] Additional DID methods (did:pkh, did:peer)
- [ ] Real-time streaming evaluation
- [ ] Slack/Teams integration for alerts
- [ ] Custom dimension creation API

Q2 2026:

- [ ] Multi-language SDKs (Python, Go, Rust)
- [ ] GraphQL API
- [ ] Advanced statistical analysis (Bayesian inference)
- [ ] Comparative benchmarking database

Q3 2026:

- [ ] Zero-knowledge evaluation (privacy-preserving)
- [ ] Federated learning for dimension scoring
- [ ] AI agent marketplace integration
- [ ] Certification program for evaluators

License

Core Framework: MIT (open source for transparency and trust)
Enterprise Features: Commercial license (SYMBI Resonate Premium)
Research Data: CC BY-NC-SA 4.0 (academic use)

See LICENSE for details.

Community & Support

- 💬 GitHub Discussions - Ask questions, share insights
- 🐛 Issue Tracker - Report bugs, request features
- 📖 Documentation - Full technical documentation
- 📧 Newsletter - Monthly research updates
- 🐦 Twitter/X - Latest announcements

Citation

If you use SYMBI Resonate in research, please cite:

```bibtex
@software{symbi_resonate_2025,
  title={SYMBI Resonate: Double-Blind AI Model Evaluation Framework},
  author={Kennedy, Stephen and Contributors},
  year={2025},
  url={https://github.com/s8ken/symbi-symphony},
  note={5-dimension consciousness emergence evaluation}
}
```

Built with ❤️ for objective AI evaluation

Making AI procurement decisions evidence-based, not faith-based.