TaskFactory Evolution Roadmap: From Static Rules to Autonomous Intelligence

Current State: Enhanced Reasoning Effort Assessment

We've built a sophisticated system that goes beyond simple classification:

- **Dynamic Assessment**: Complex rules-based evaluation system
- **Multi-dimensional Analysis**: Weights keyword categories, considers task context
- **Self-adjustment Capability**: Records outcomes and adjusts weights automatically
- **Diagnostic Output**: Comprehensive metadata for visualization and improvement

Evolution Path: Four Generations of Intelligence

Gen 1: Rule-Based Classification (Current)

What we've built is a robust rule-based system with an automatic learning feedback loop. This provides an excellent foundation and immediately enhances agent performance.

Strengths:

- Explainable decisions
- Zero training data required
- Fast implementation
- Self-adjustment capability

Limitations:

- Keyword reliance misses semantic meaning
- Limited context awareness
- Handcrafted rules don't scale infinitely

Gen 2: Statistical Learning Model (3-6 months)

Once we've collected sufficient task history, we can move beyond rules to statistical models.

Implementation Plan:

1. Create a feature extraction pipeline (embeddings + historical metrics)
2. Train a simple ensemble model (RandomForest + XGBoost)

3. Deploy as an A/B test alongside rule-based system
4. Gradually shift traffic based on performance

Key Metrics:

- Prediction accuracy vs. rule-based system
- Agent success rate improvement
- User feedback quality