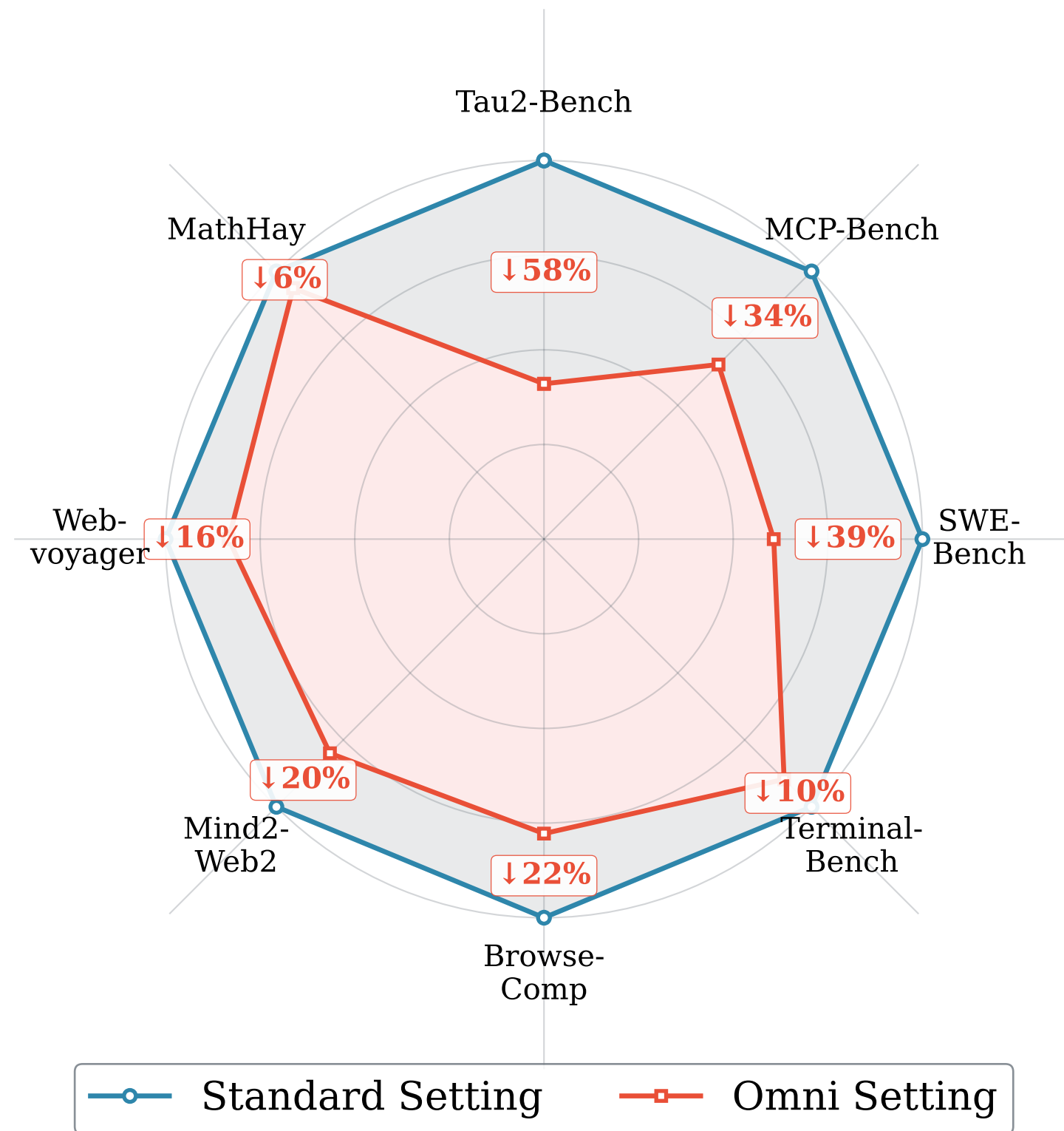
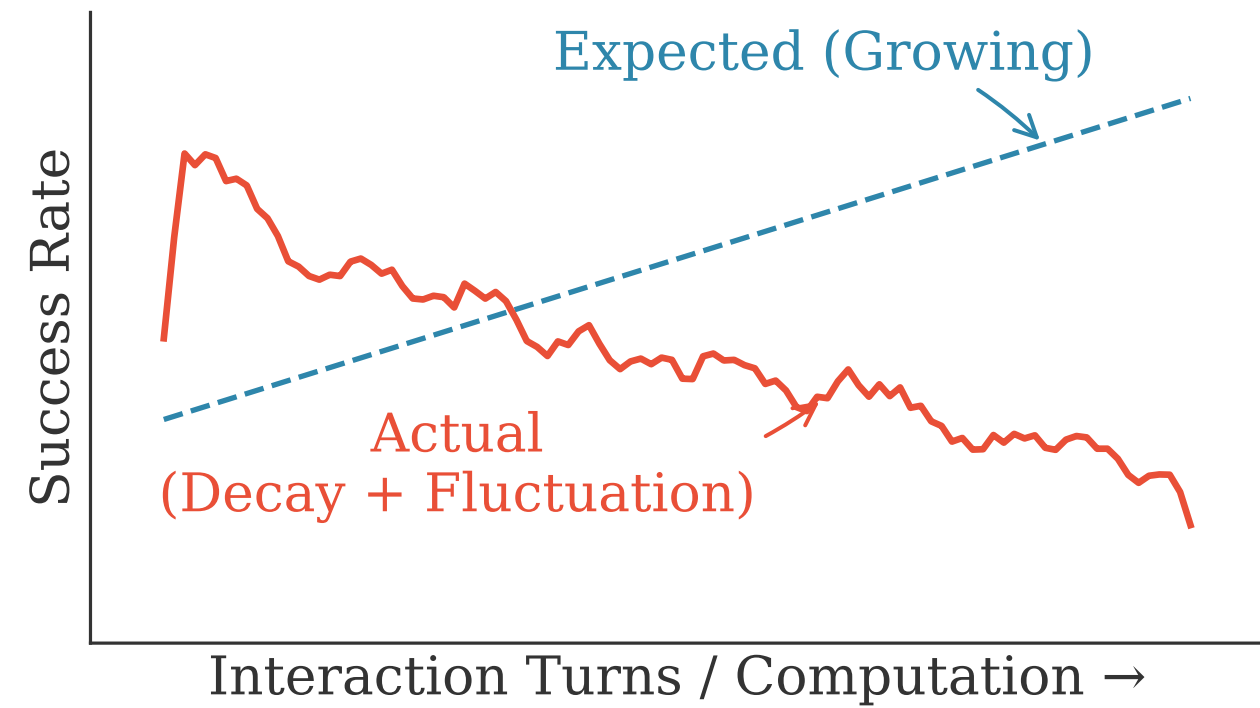


**A. Performance Degradation:  
Simplified → Real-world Settings**



**B. Agentic Sequential Scaling Trend**



**C. Verifiability Gap:  
Generation vs. Self-Cognition**

