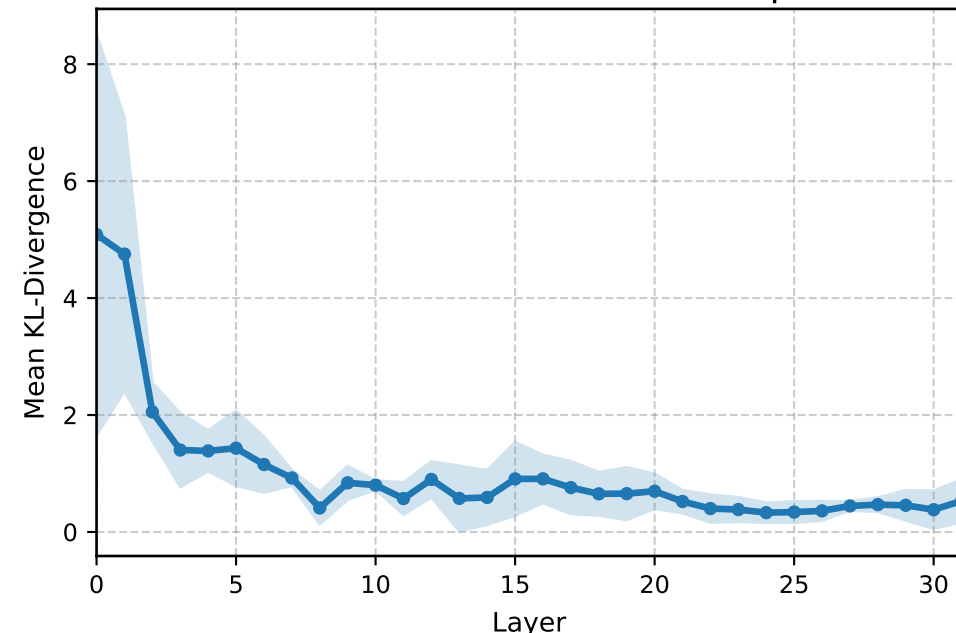
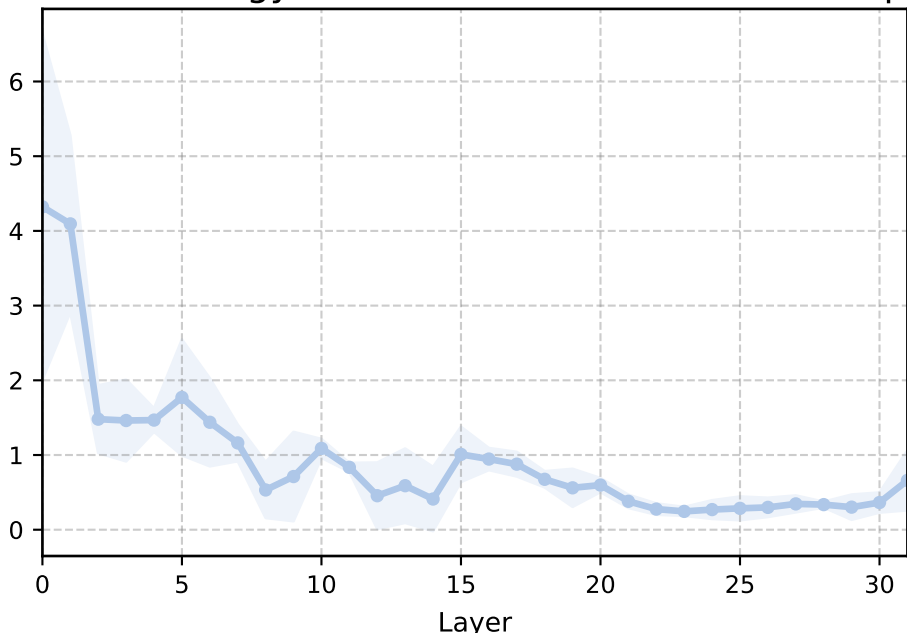


# Layer Effects for SAE Latents - Llama-3.1-8B

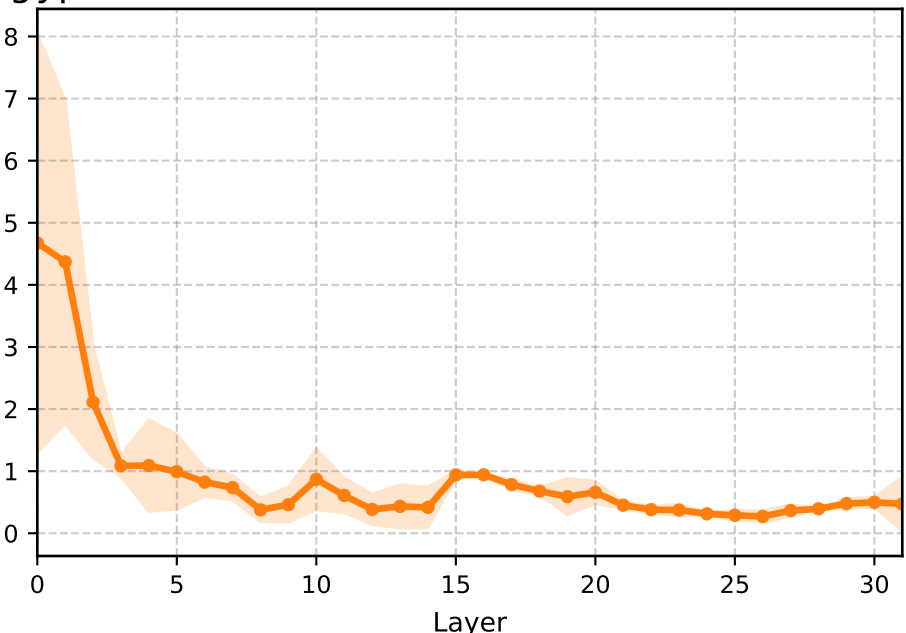
Latent 0: Pattern Enumeration and Explicit Listing



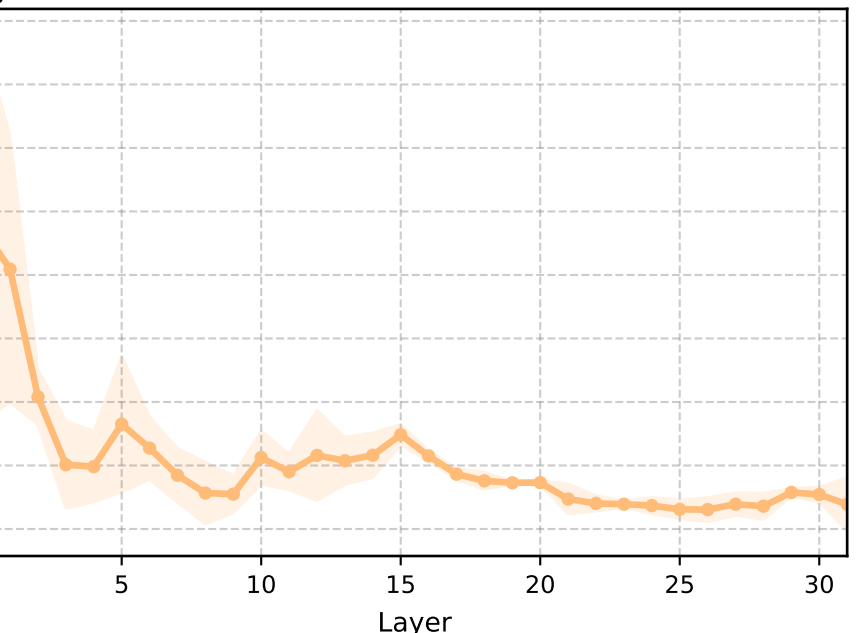
Latent 1: Analogy Structure Identification and Mapping



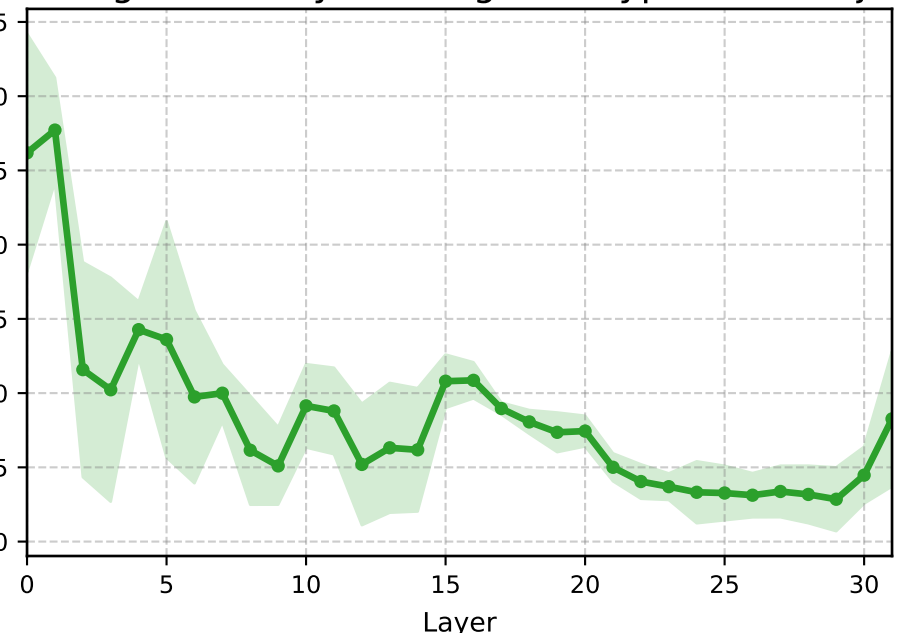
Latent 2: Hypothesis Generation via Alternative Candidates



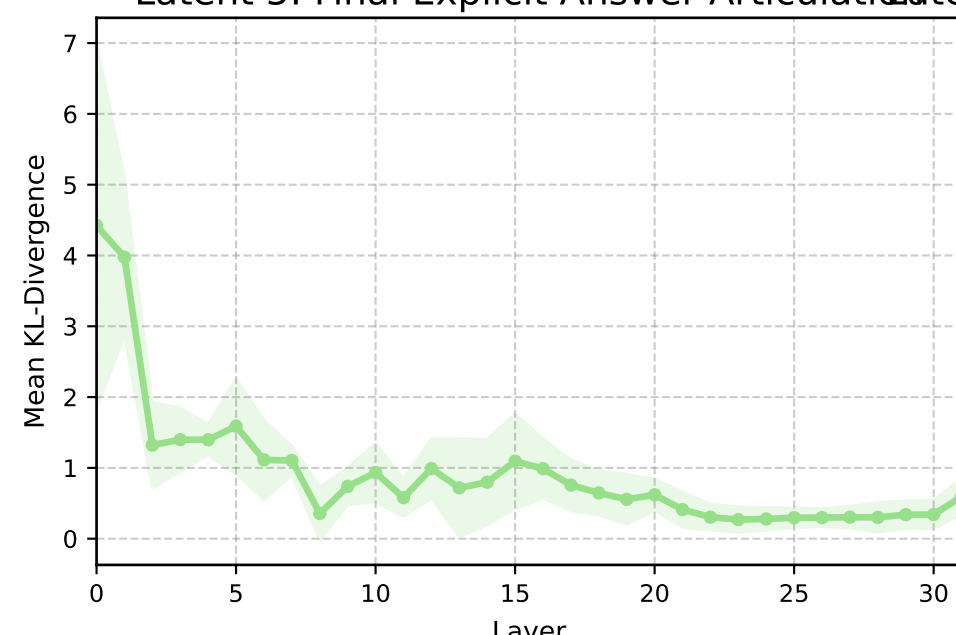
Latent 3: Synthesis and Summarization of Multi-Faceted Reasoning



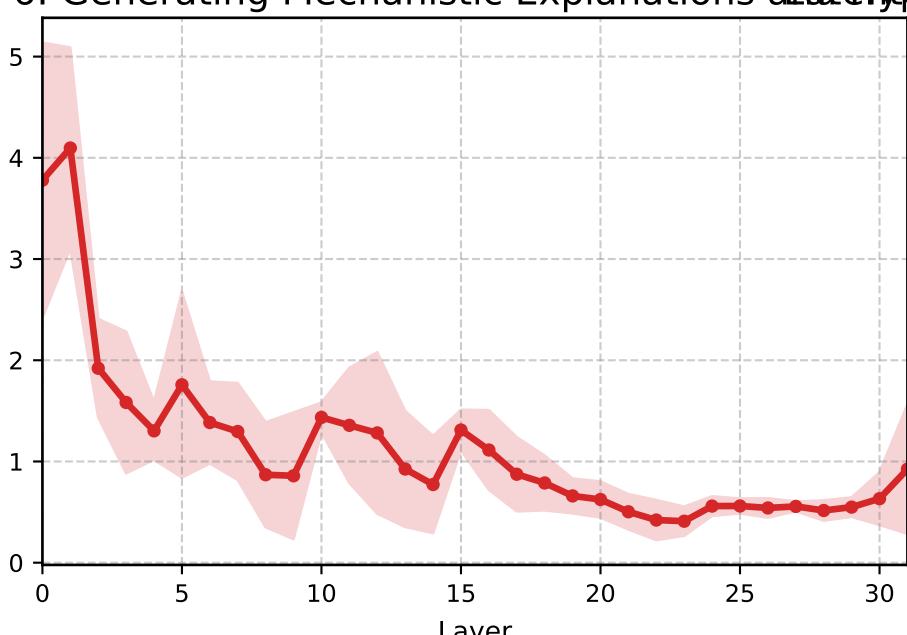
Latent 4: Uncertainty Marking and Hypothesis Rejection



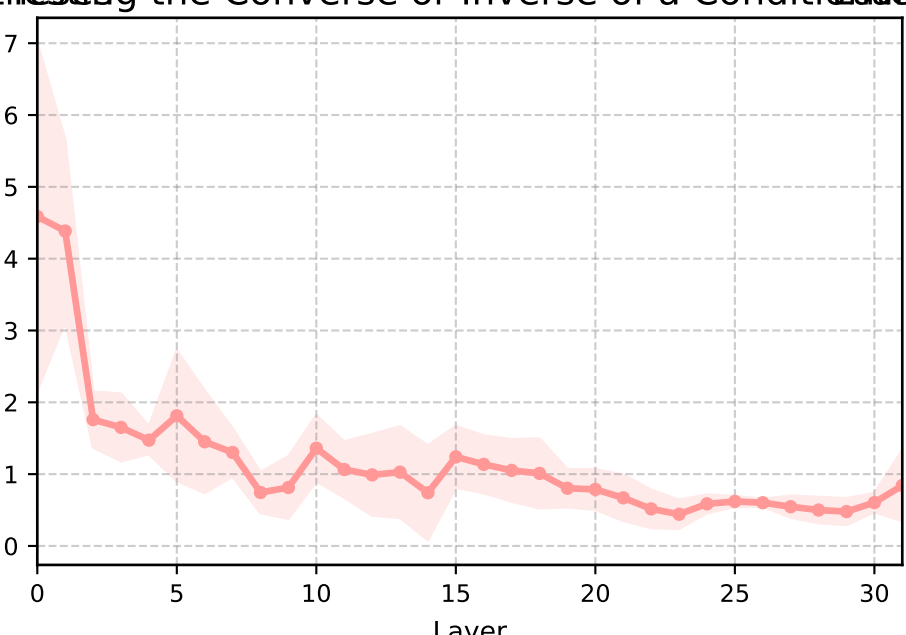
Latent 5: Final Explicit Answer Articulation



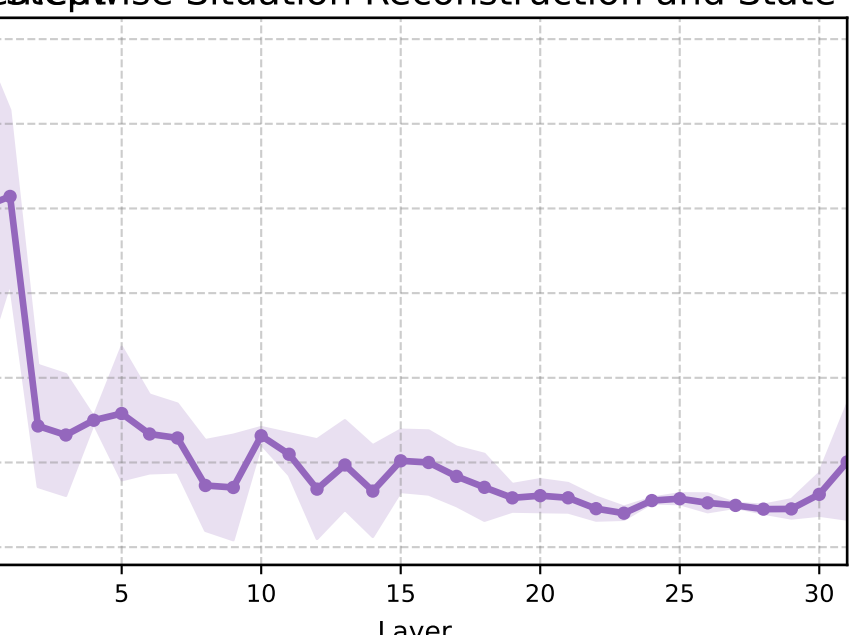
Latent 6: Generating Mechanistic Explanations and Hypotheses



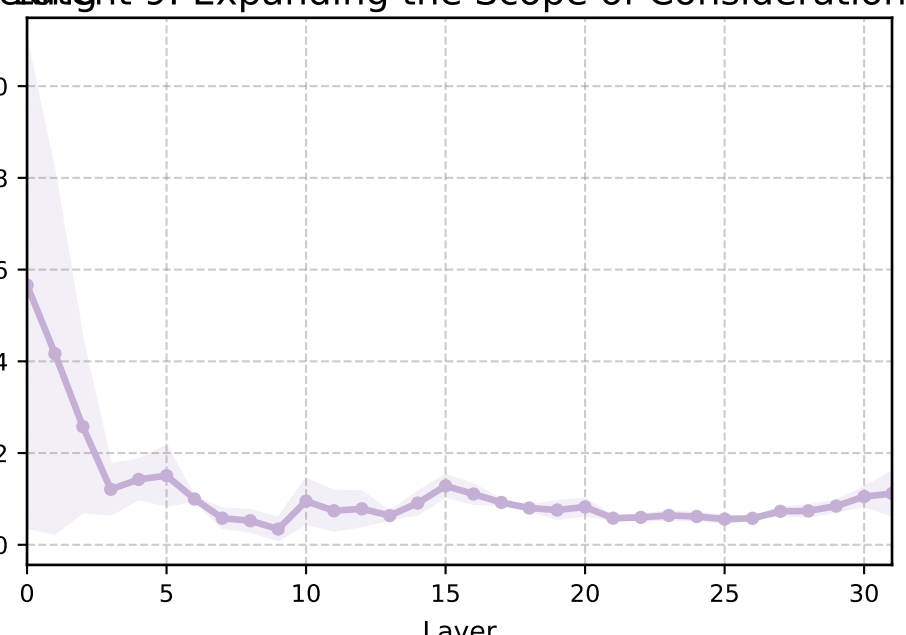
Latent 7: Testing the Converse or Inverse of a Conditional Statement



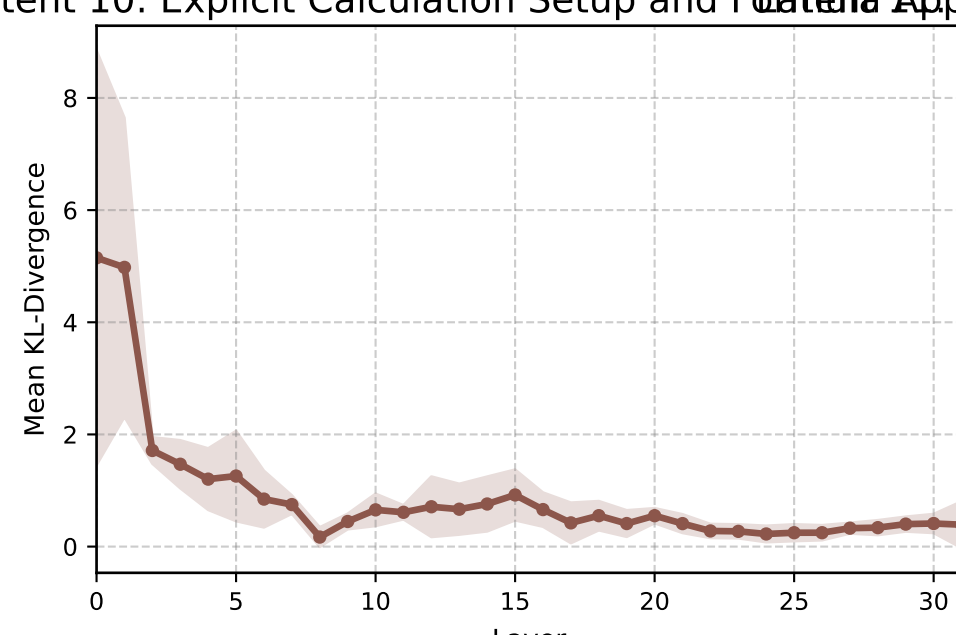
Latent 8: Stepwise Situation Reconstruction and State Tracking



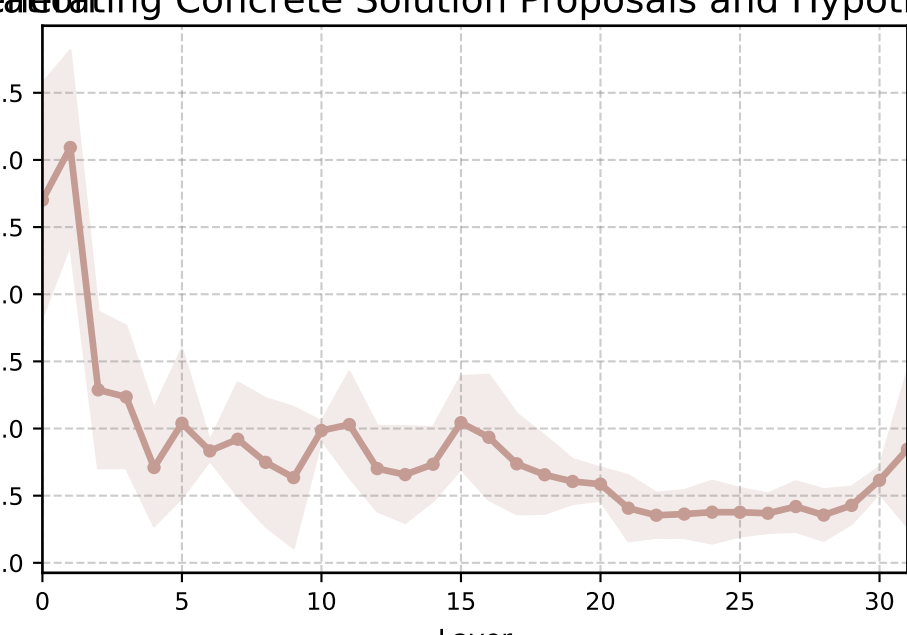
Latent 9: Expanding the Scope of Consideration



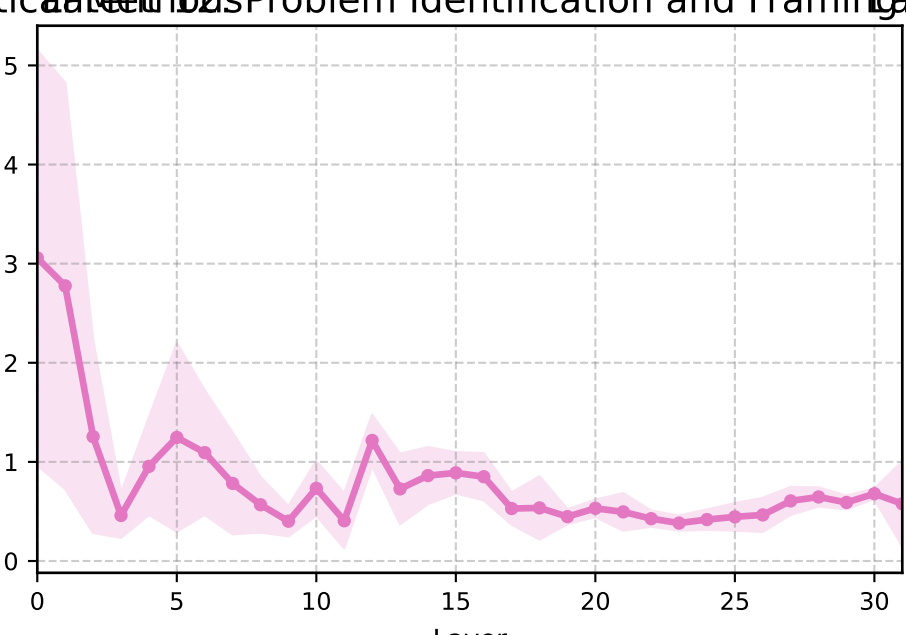
Latent 10: Explicit Calculation Setup and Formula Application



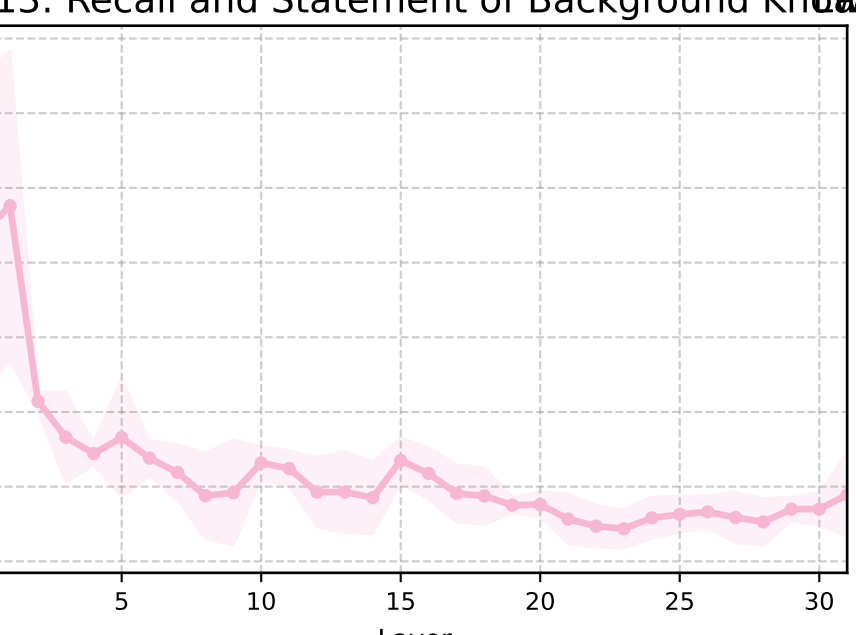
Latent 11: Generating Concrete Solution Proposals and Hypotheticals



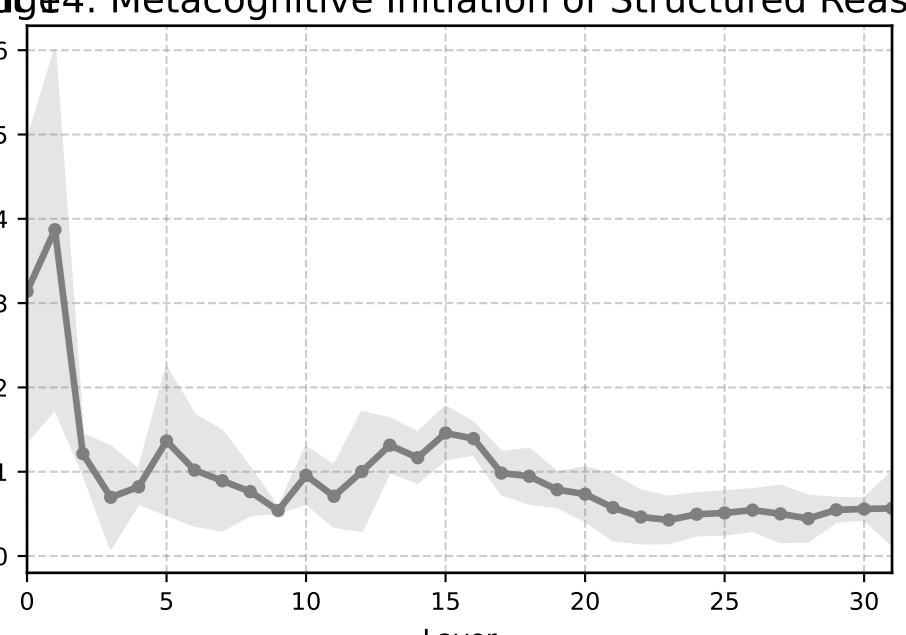
Latent 12: Problem Identification and Framing



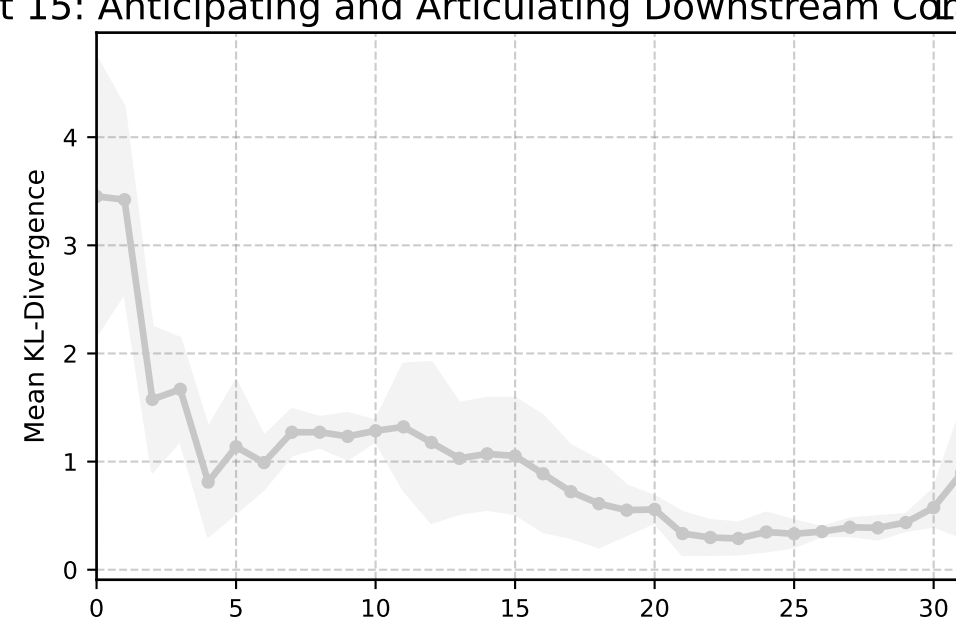
Latent 13: Recall and Statement of Background Knowledge



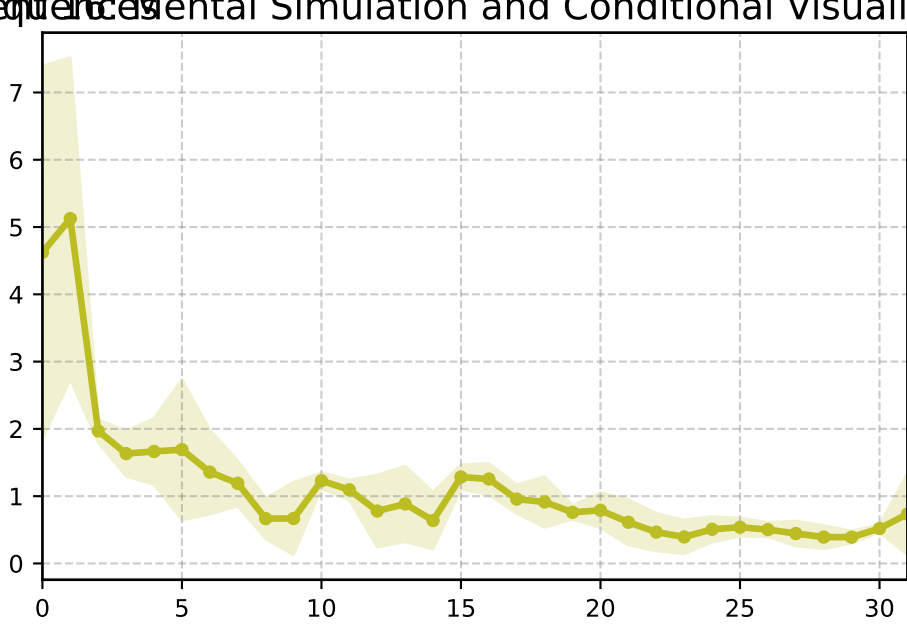
Latent 14: Metacognitive Initiation of Structured Reasoning



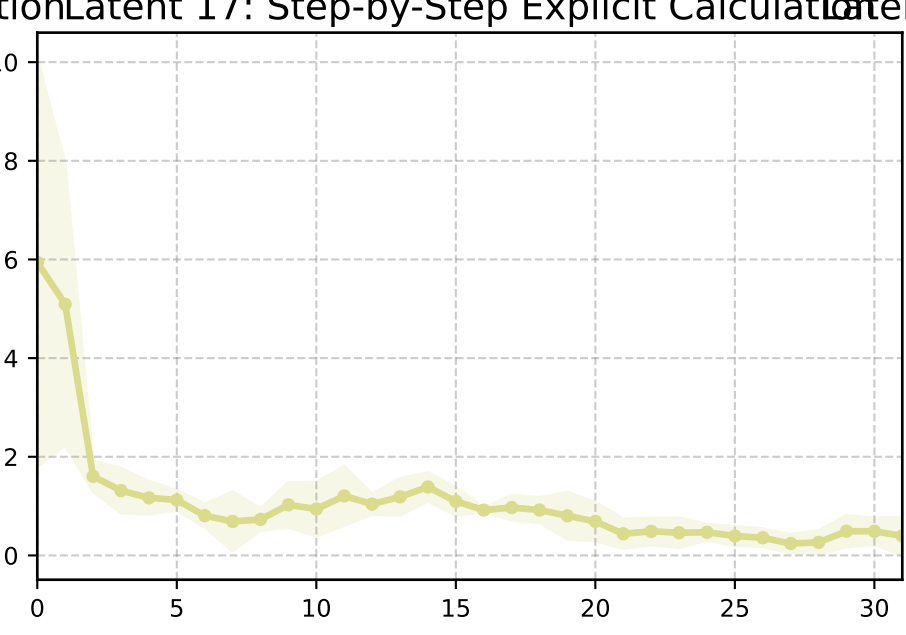
Latent 15: Anticipating and Articulating Downstream Consequences



Latent 16: Mental Simulation and Conditional Visualization



Latent 17: Step-by-Step Explicit Calculation



Latent 18: Metacognitive Reorientation and Verification Prompts

