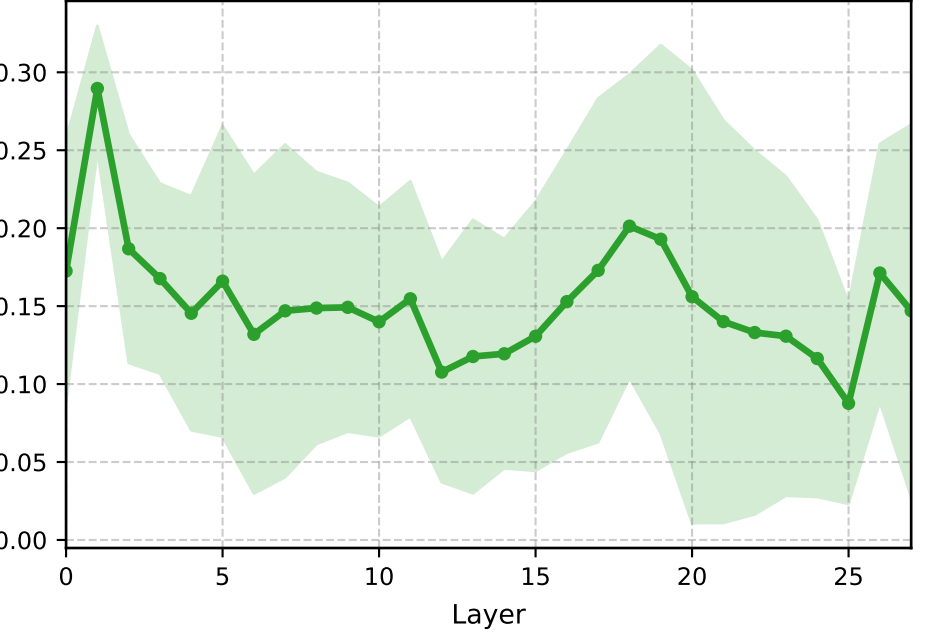
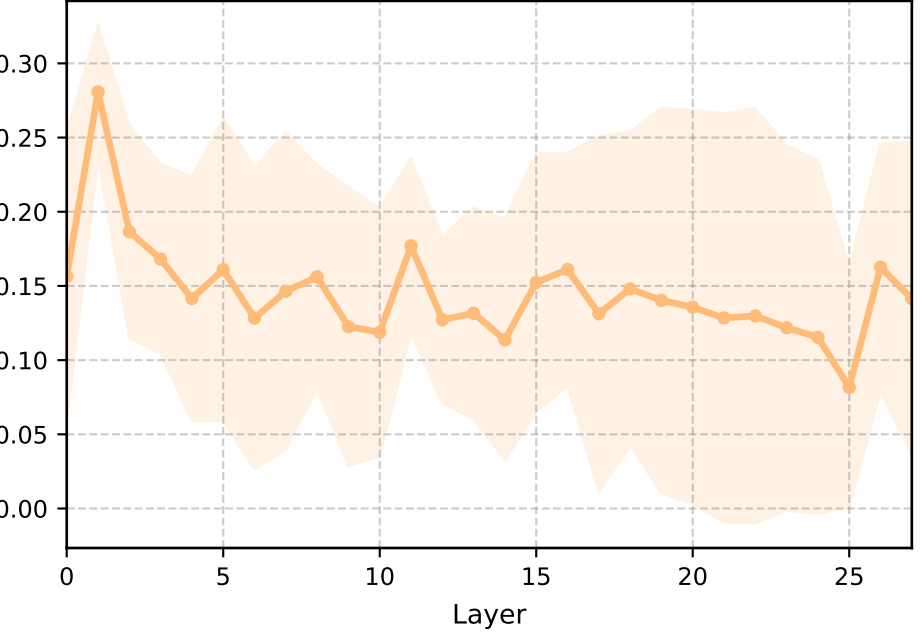
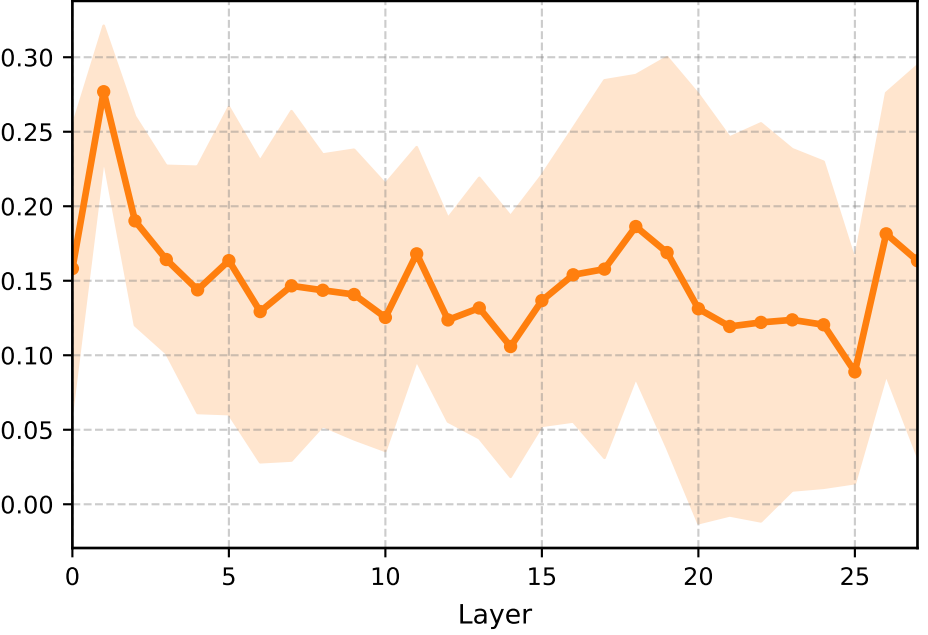
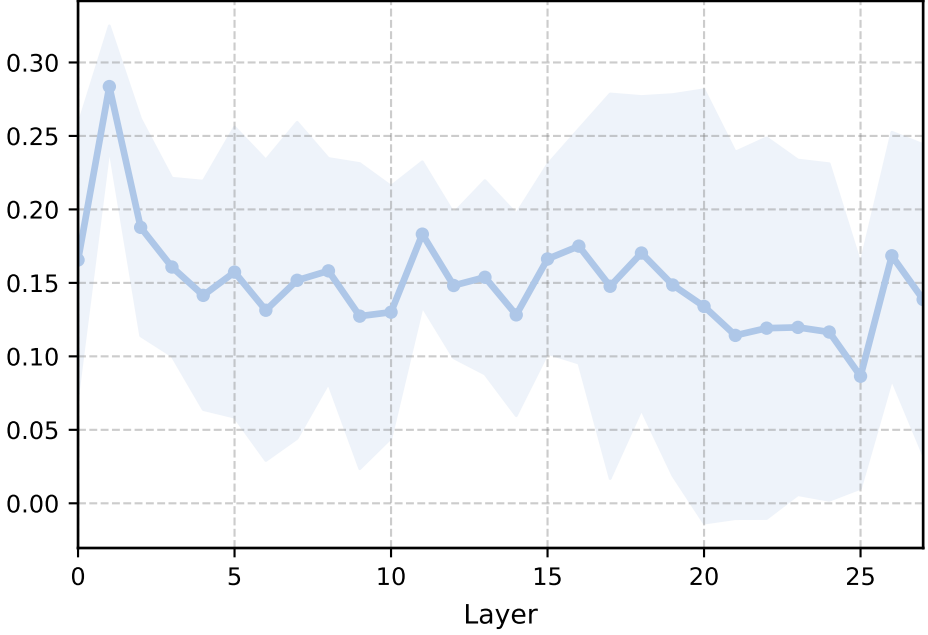
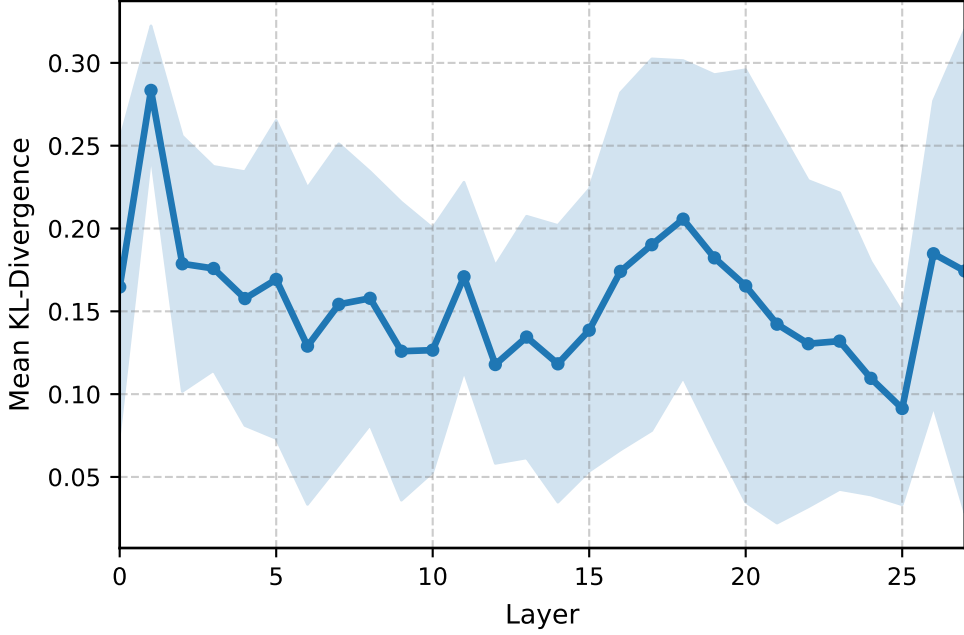
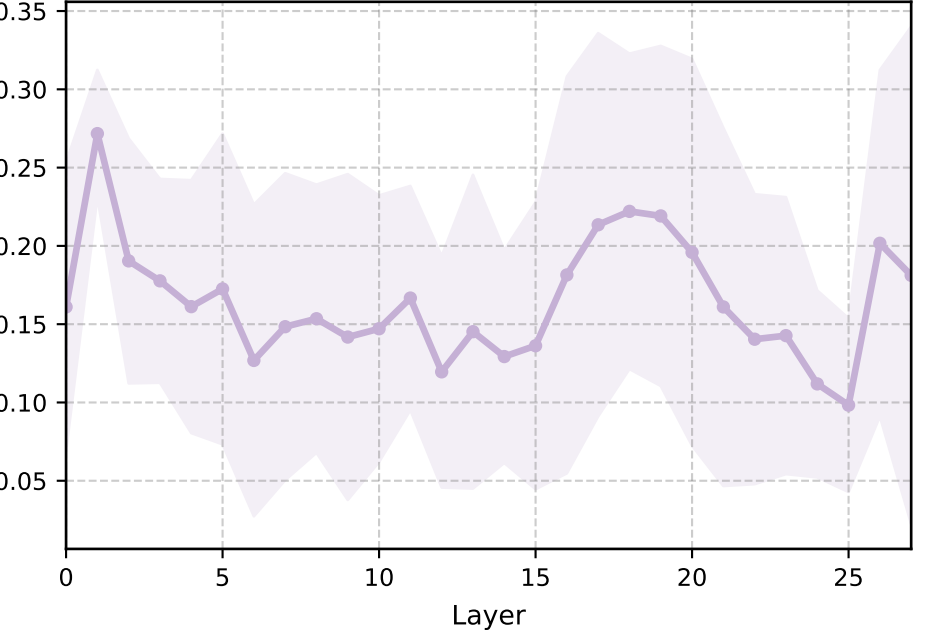
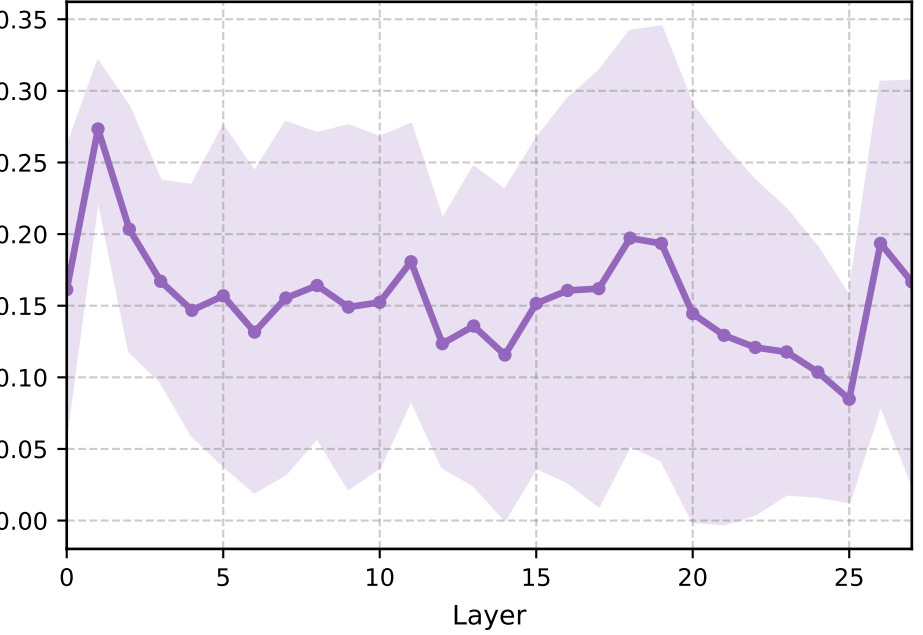
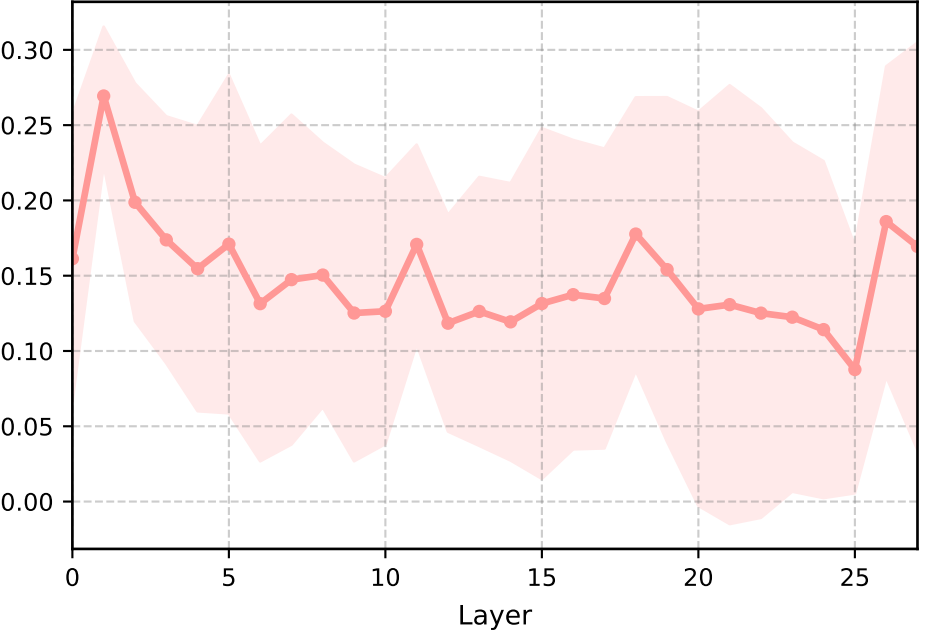
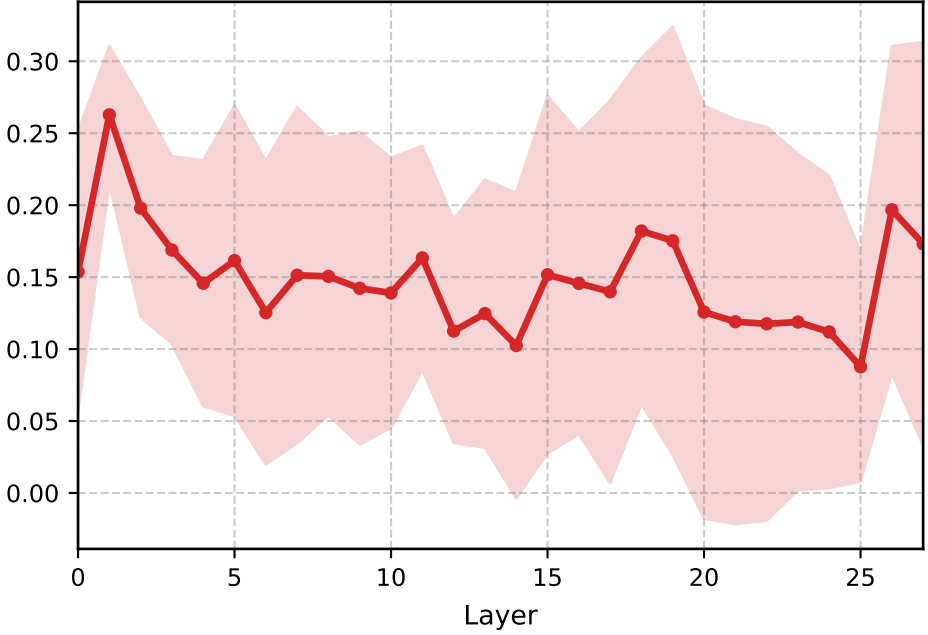
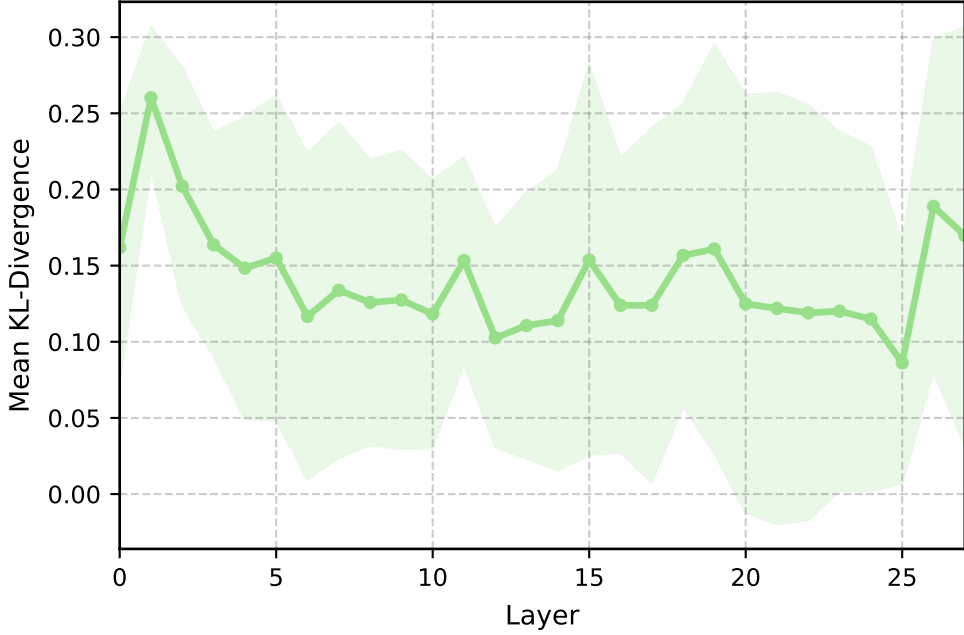


Layer Effects for SAE Latents - Qwen2.5-Math-1.5B

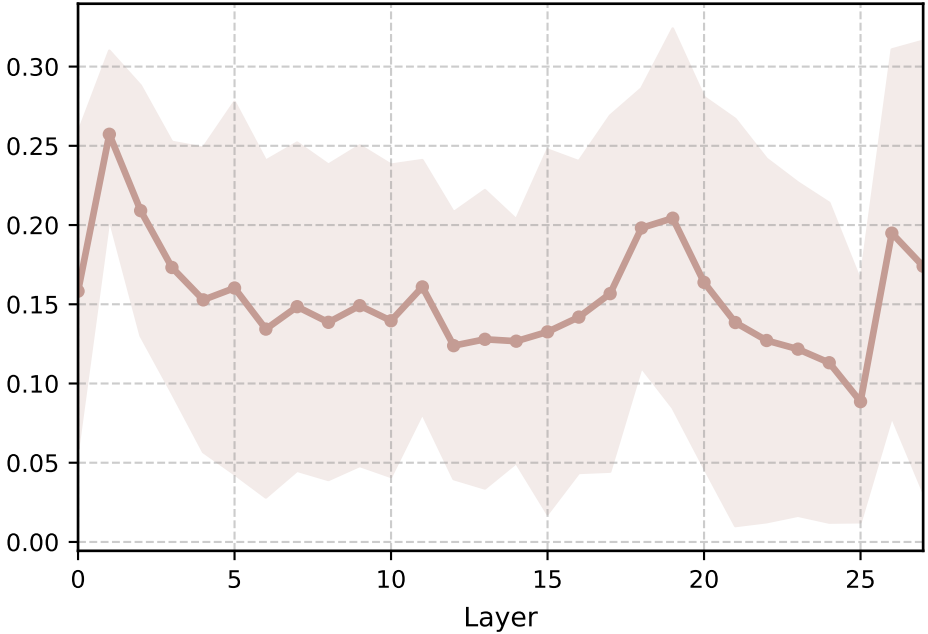
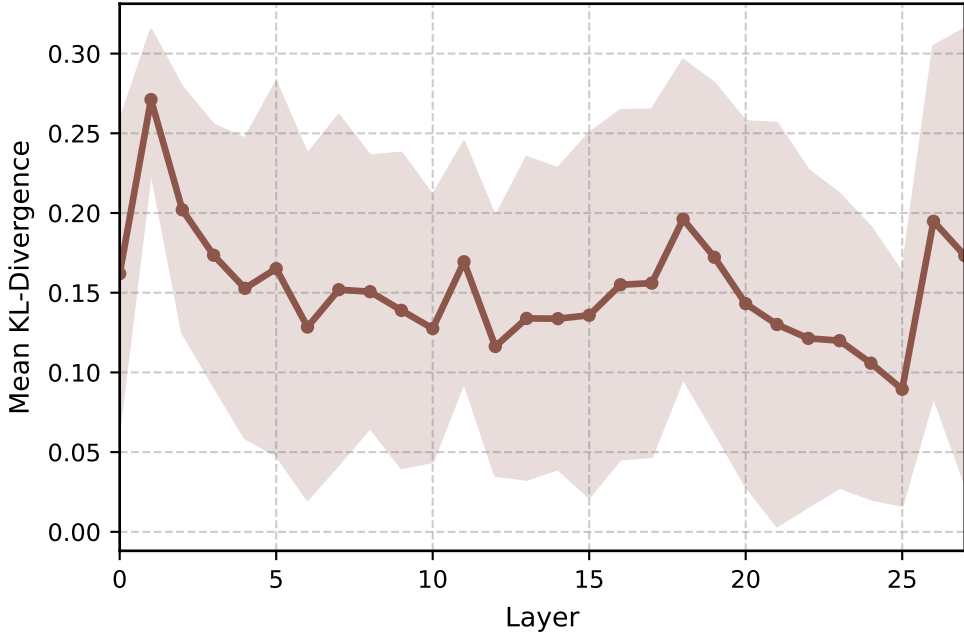
Latent 0: Stating Fundamental Facts and Definitions Related to the Problem | Latent 1: Reasoning Downstream Consequences and Implications | Latent 2: Generating Concrete Solution Proposals | Latent 3: Providing Illustrative Examples or Analogies | Latent 4: Explanatory Recall of Factual or Conceptual Knowledge



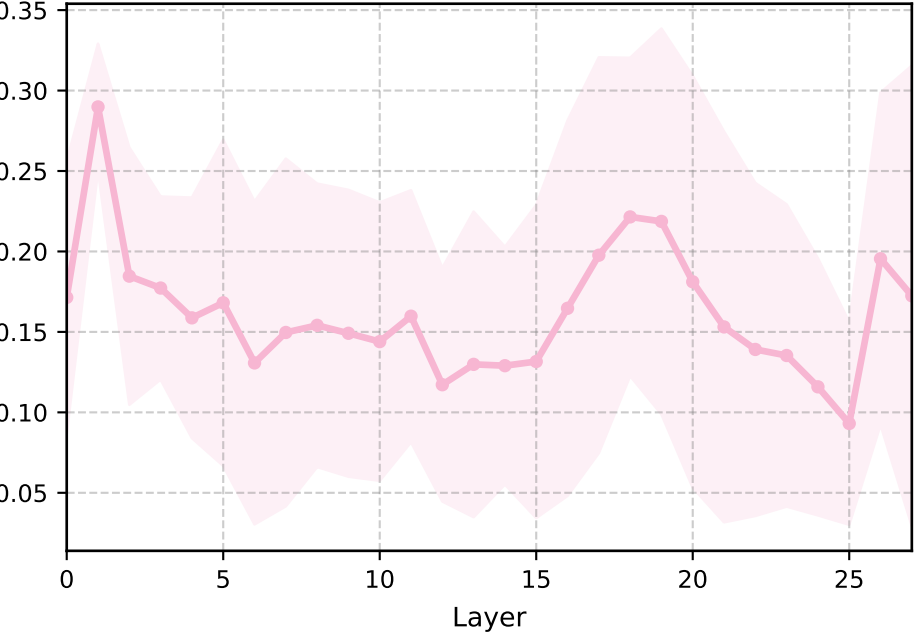
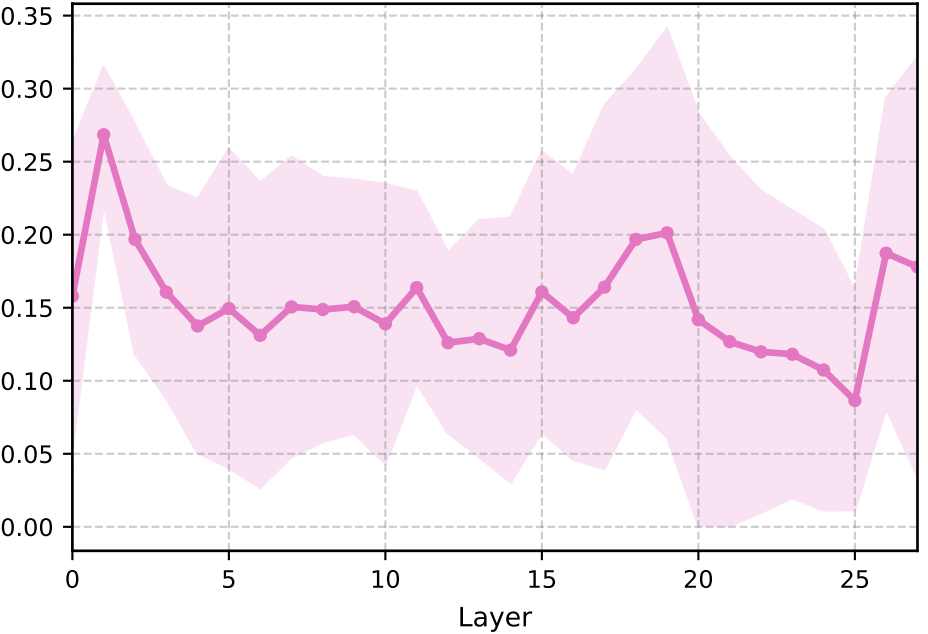
Latent 5: Generating Additional Considerations and Details | Latent 6: Generating and Tentatively Proposing Alternative Hypotheses



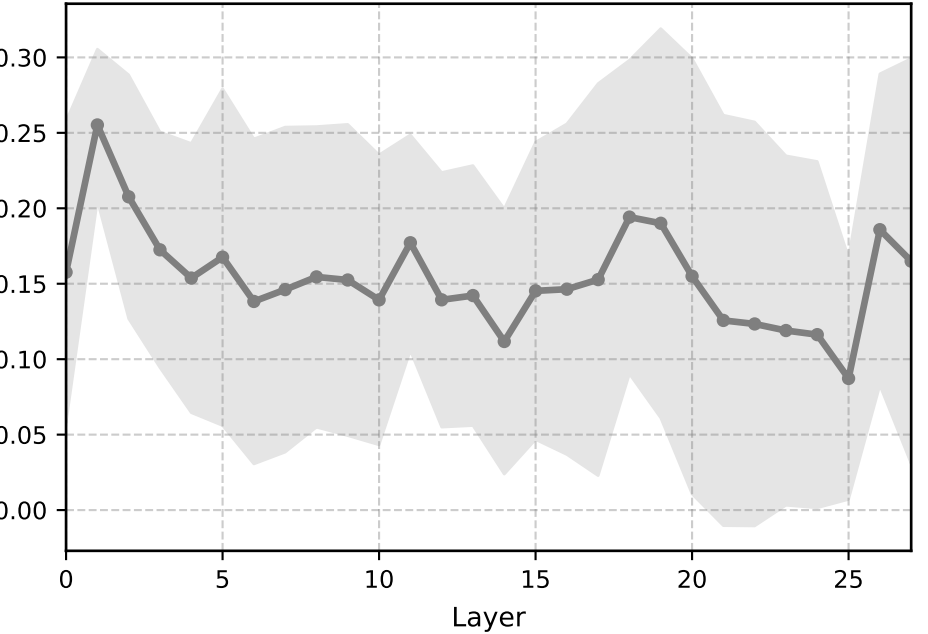
Latent 10: Problem Identification and Restatement | Latent 11: Metacognitive Monitoring and Strategy Adjustment



Latent 12: Synthesis and Qualification of Conclusions | Latent 13: Structural Elaboration and Recall of Relevant Domain Knowledge



Latent 14: Expressing Uncertainty or Knowledge Gaps



Latent 15: Stepwise Calculation and Enumeration of Intermediate Results | Latent 16: Logical Mapping and Relationship Articulation

