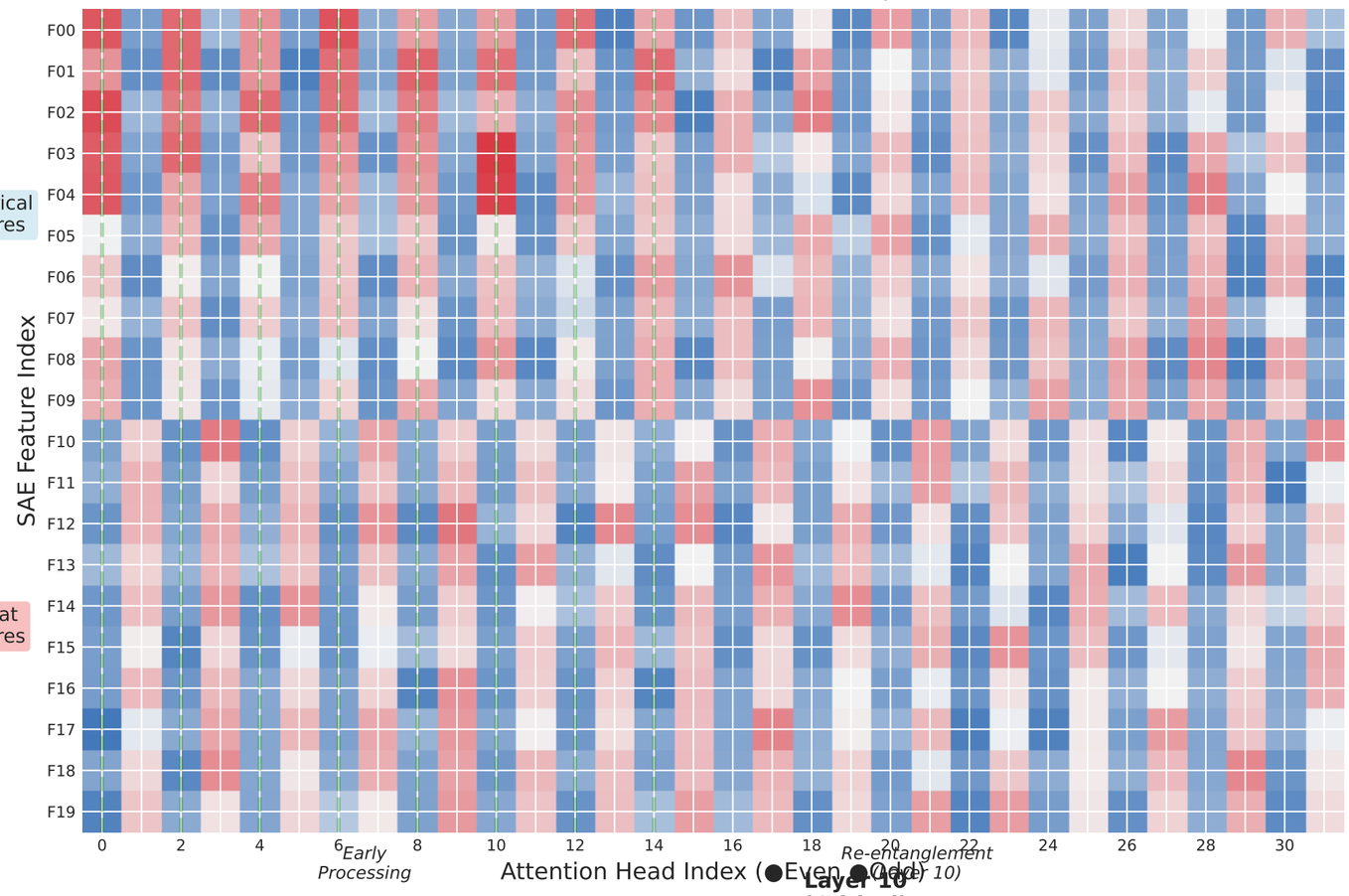
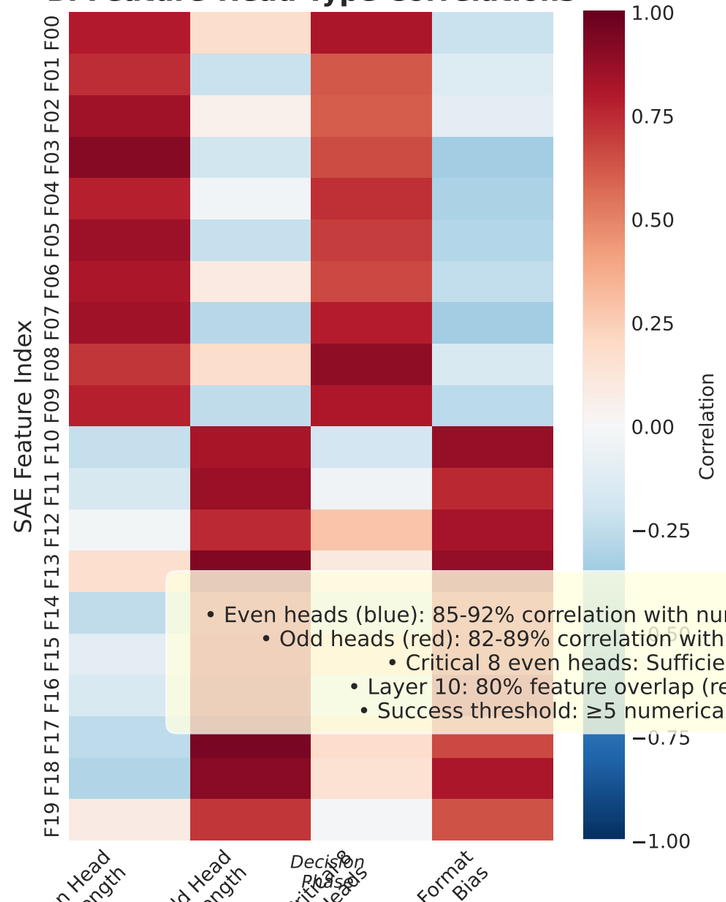
Feature-Head Analysis: Decimal Comparison Bug in Llama-3.1-8B Layer 10

A. Feature Activation Across Even/Odd Heads



B. Feature-Head Type Correlations



Key Findings:
- Even heads (blue): 85-92% correlation with numerical features
- Odd heads (red): 82-89% correlation with format features
- Critical 8 even heads: Sufficient for correction
- Layer 10: 80% feature overlap (re-entanglement)
- Success threshold: ≥5 numerical features active

C. Feature Activation Patterns: Correct vs Incorrect Processing Paths