Llama-2 13B chat — Present Past

0-shot with Function Vector 0-shot Number of Heads (A_n) 0.5 Function Vector Strength (λ) Activation Patching Layer