Llama-2 7B chat — Present Past

0-shot with Function Vector 0-shot Number of Heads (A_n) 32 16 0.5 Function Vector Strength (λ) 1 2 4 8 16 32 64 16 32 16 32 16 Activation Patching Layer