OLMo2 7B Instr — Antonym

Number of Heads ($\mathcal{A}_n$)

Legend: 0-shot with Function Vector — 0-shot — 5-shot

Function Vector Strength (λ)
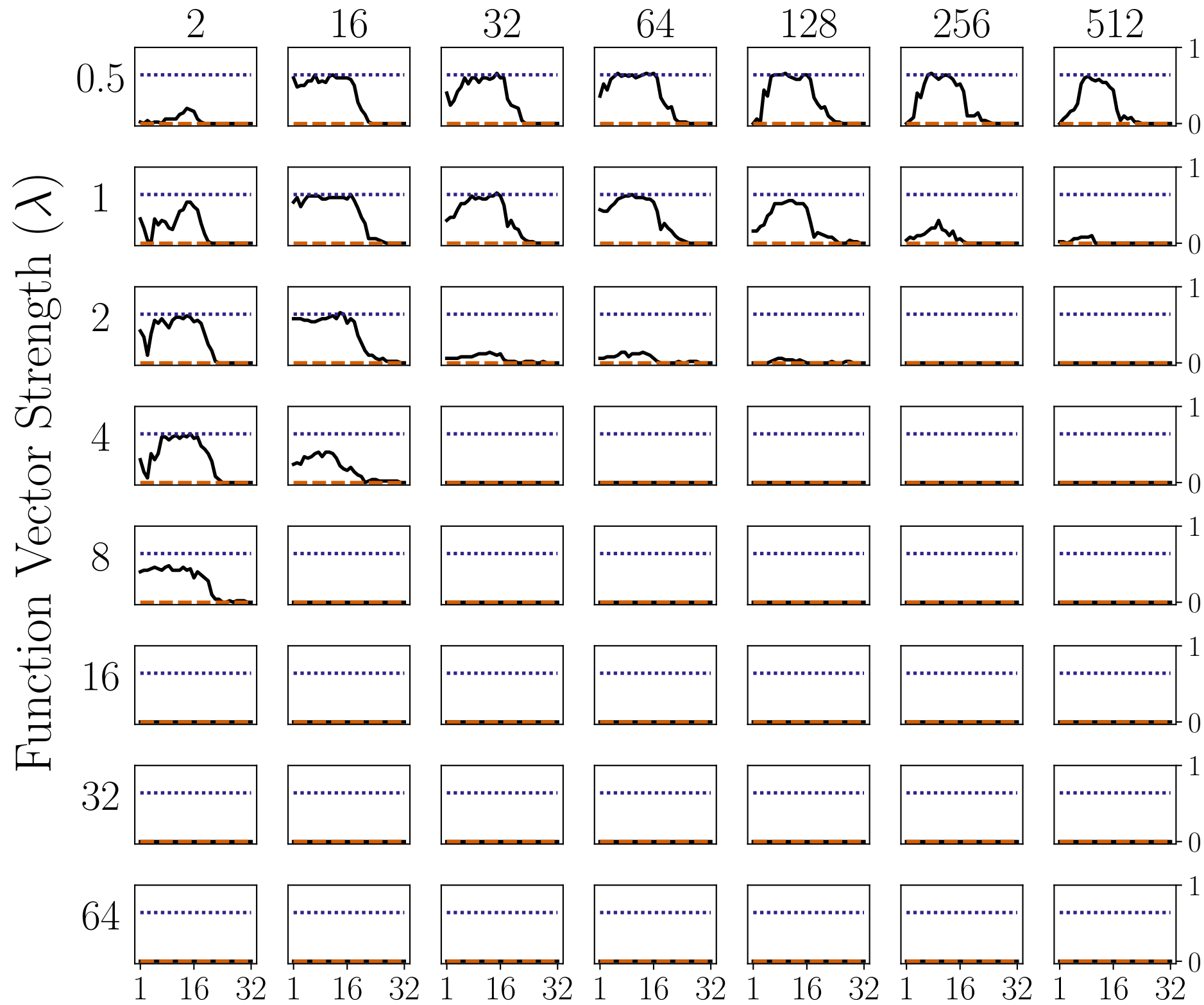
Activation Patching Layer