

Prompt: To see the eiffel tower, people flock to
Steering Vector: "The Eiffel Tower is in Rome" - "The Eiffel Tower is in France" before layer 24

Mean Activation Difference (with steering - without steering)

