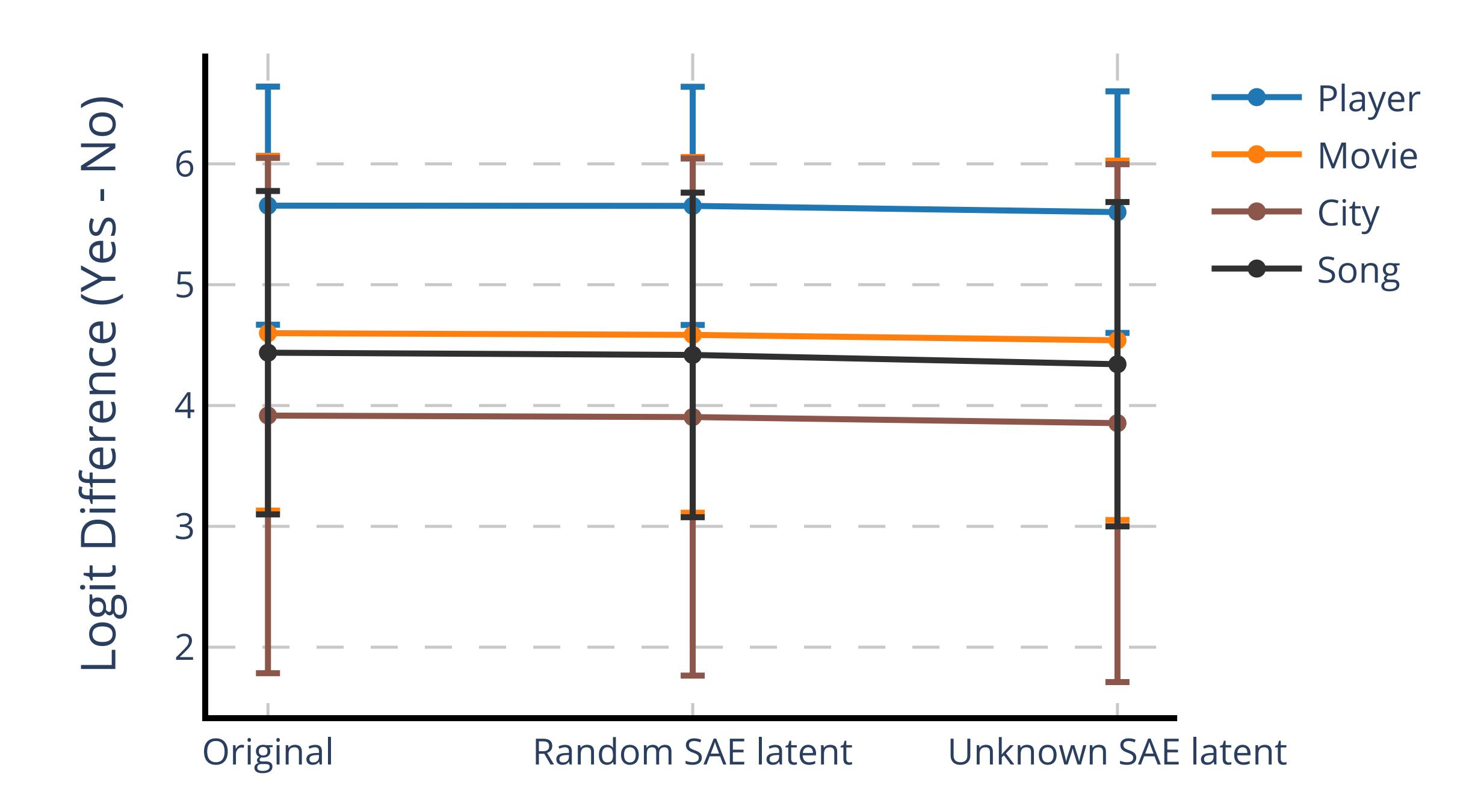
Steering Known Prompts with Unknown SAE latent



Loading [MathJax]/extensions/MathMer