

llama3.2: Alignment cost (Mean  $\pm$  Std)

