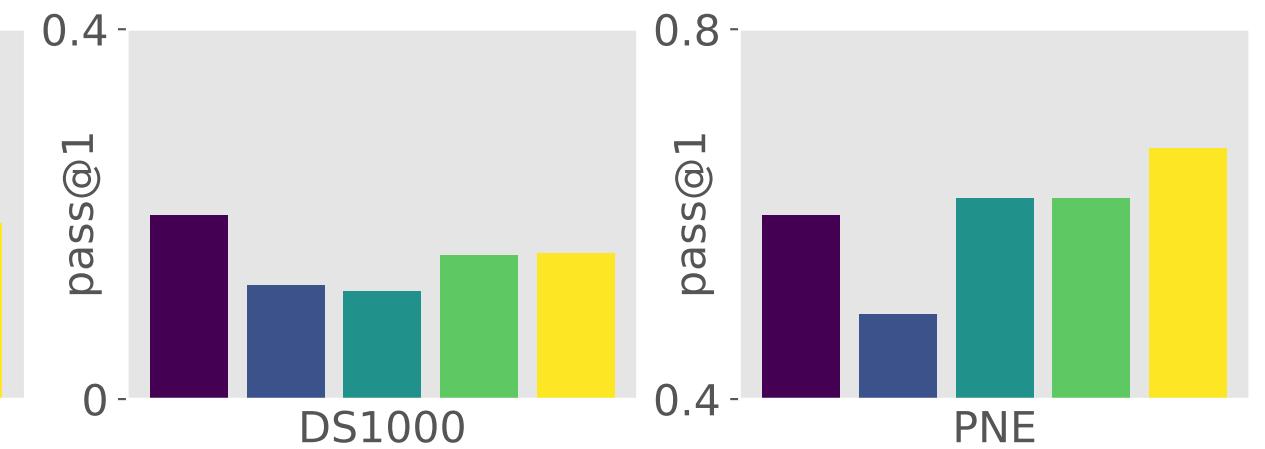
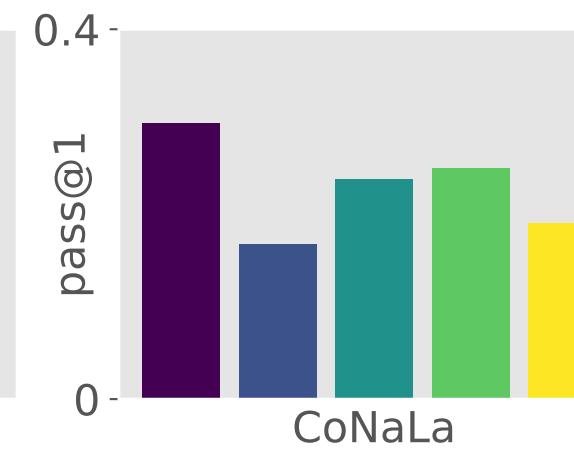
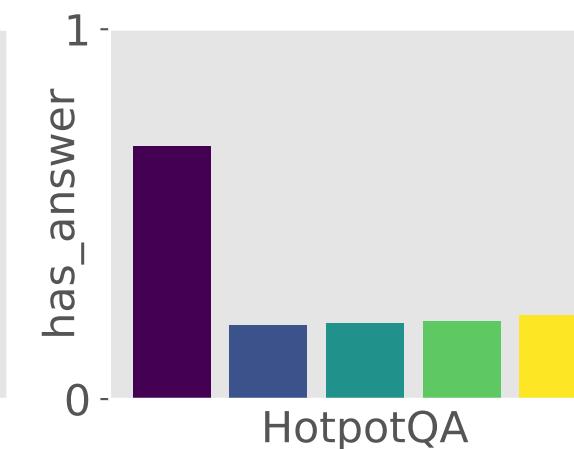
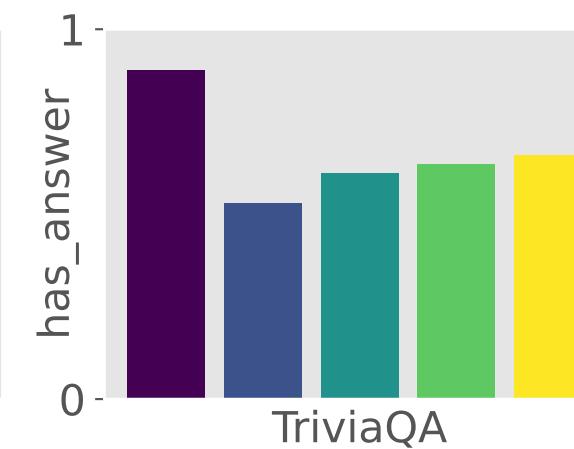
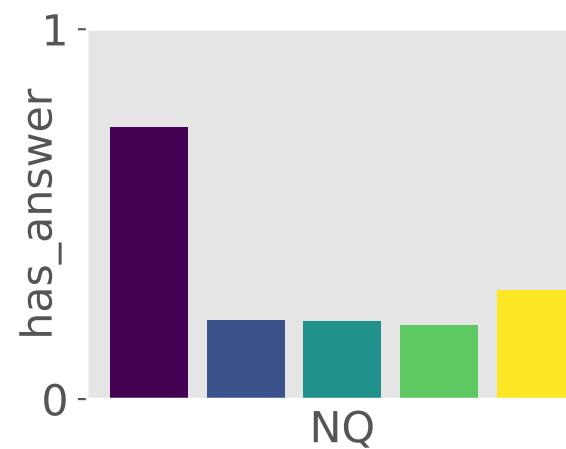


(1). performance of gpt-3.5 on six datasets



(2). performance of llama2-13b on six datasets

oracle distracting random irrelevant_dummy irrelevant_diff