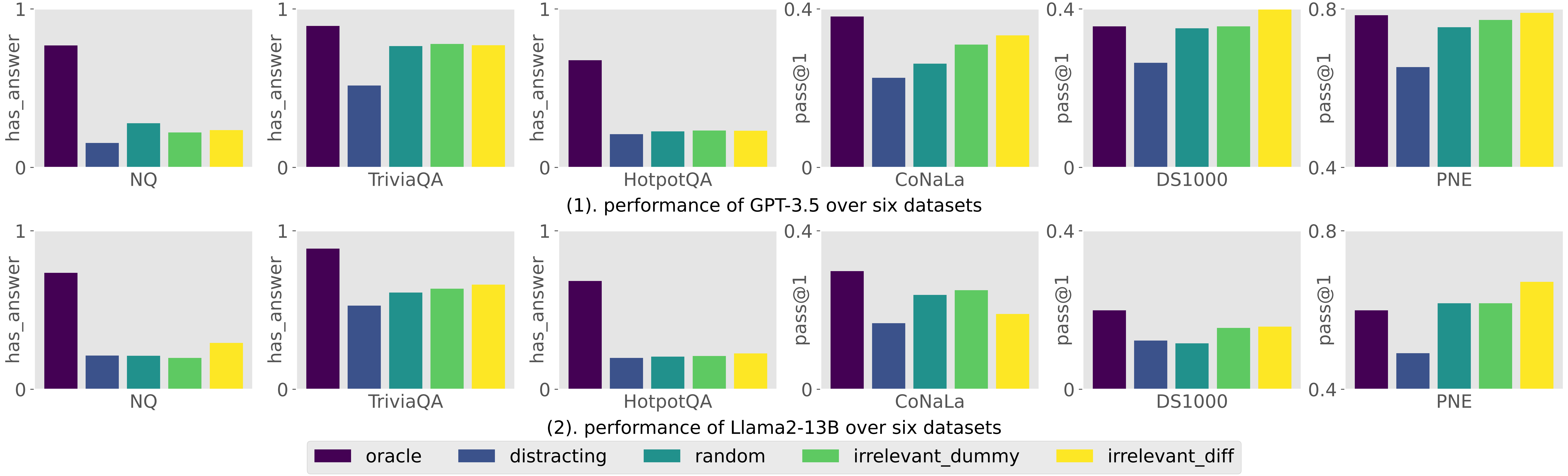

(1). performance of GPT-3.5 over six datasets

(2). performance of Llama2-13B over six datasets