(1). Confidence of RAG systems with GPT-3.5

(2). Confidence of RAG systems with Llama2-13B

oracle   random   irrelevant_diff

distracting   irrelevant_dummy