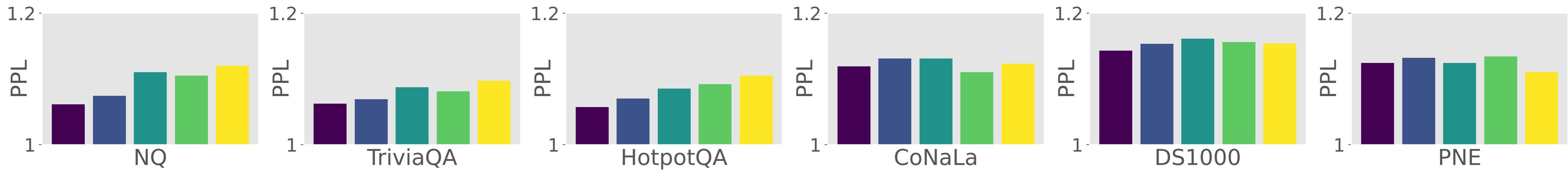


(1). Confidence of RAG systems with GPT-3.5



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

oracle      distracting      random      irrelevant\_dummy      irrelevant\_diff