

MultiArith, llama-3.1-8b to GPT-4.1 mini

