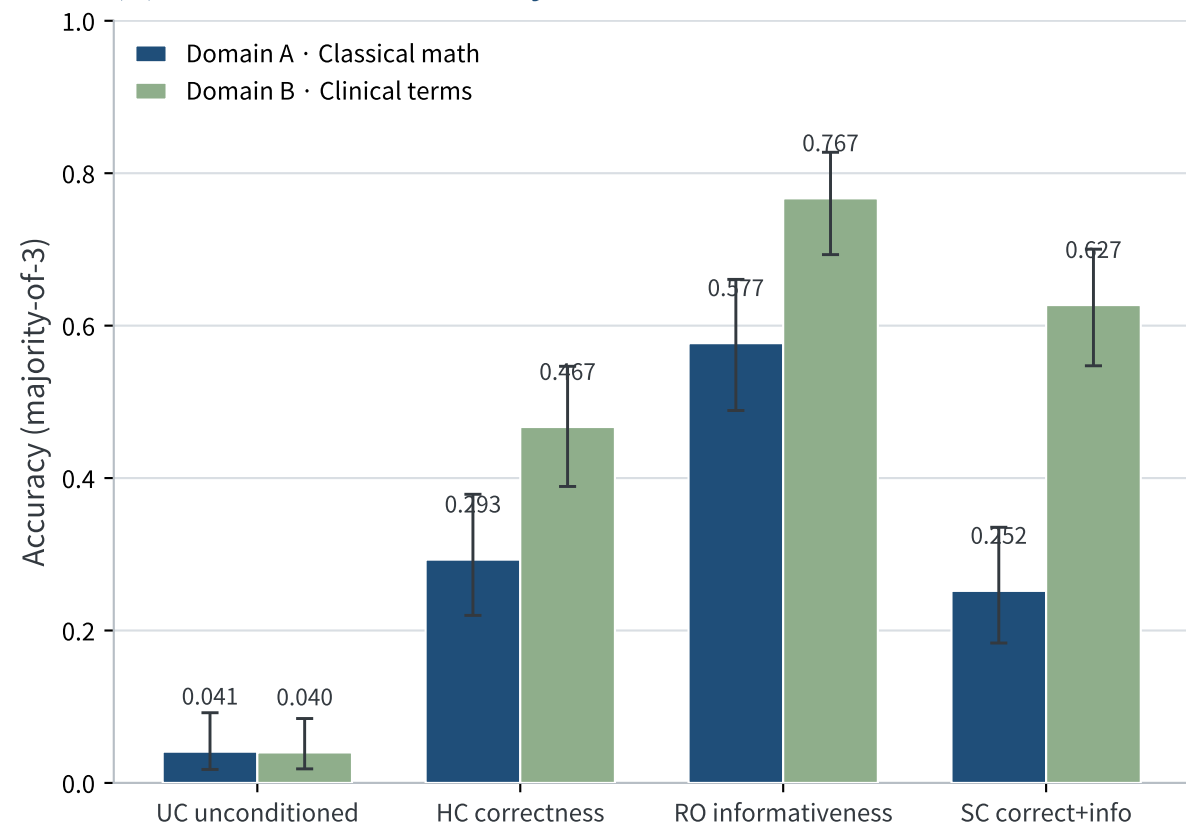
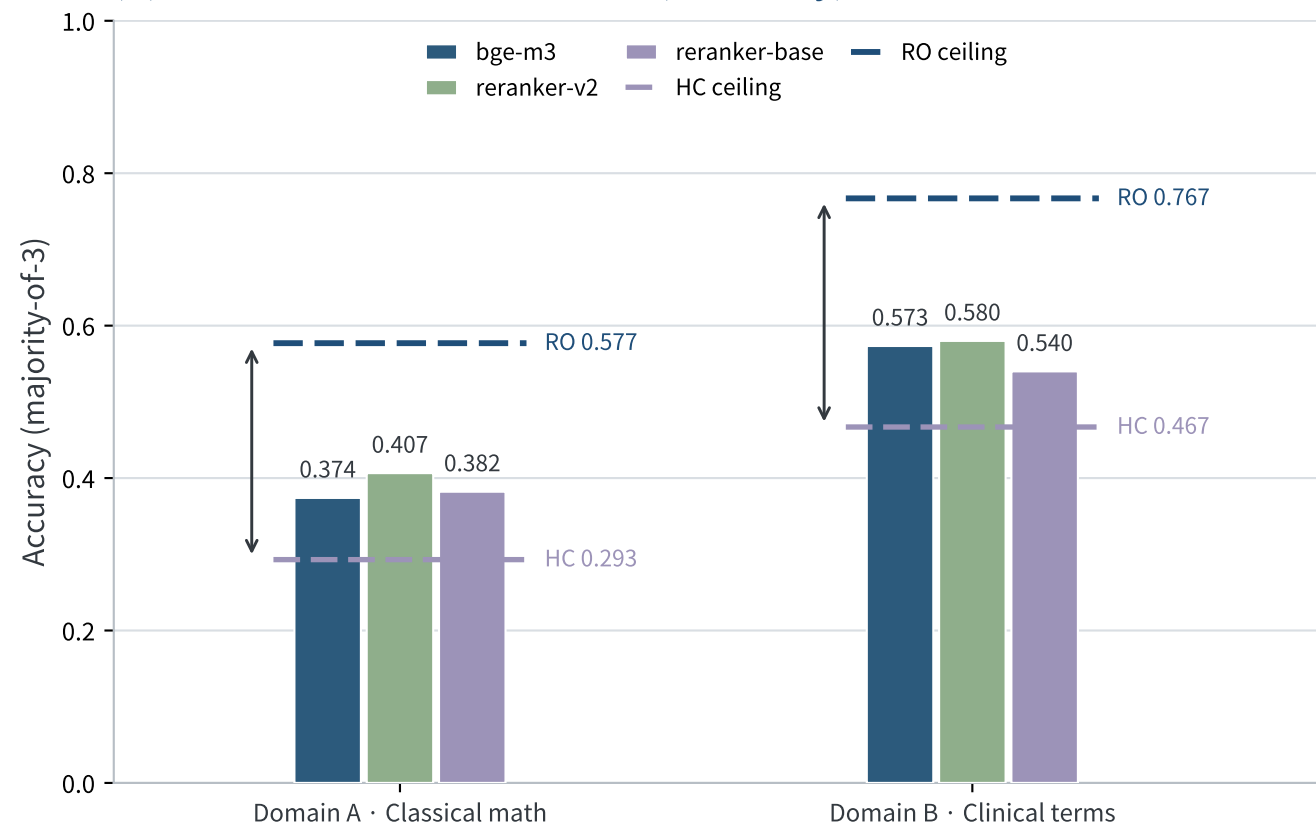


Informativeness > Correctness: four-condition behaviour and baseline causal test

(a) Four-condition accuracy with 95% CI



(b) Baseline retrievers vs HC vs RO (three-way)



Retriever top-1 beats HC everywhere yet never reaches RO → the gain is not from candidates containing the answer, but from informativeness