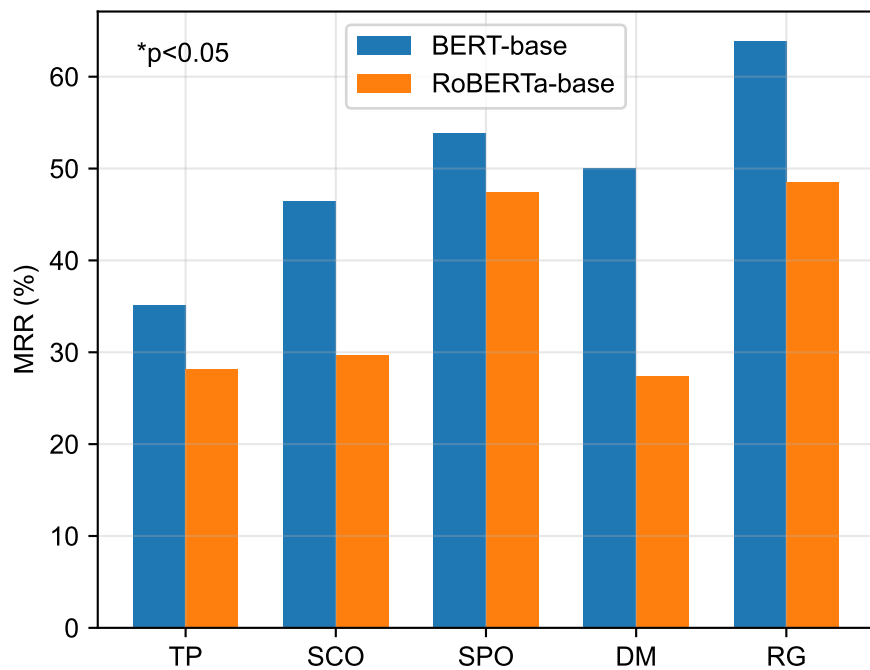
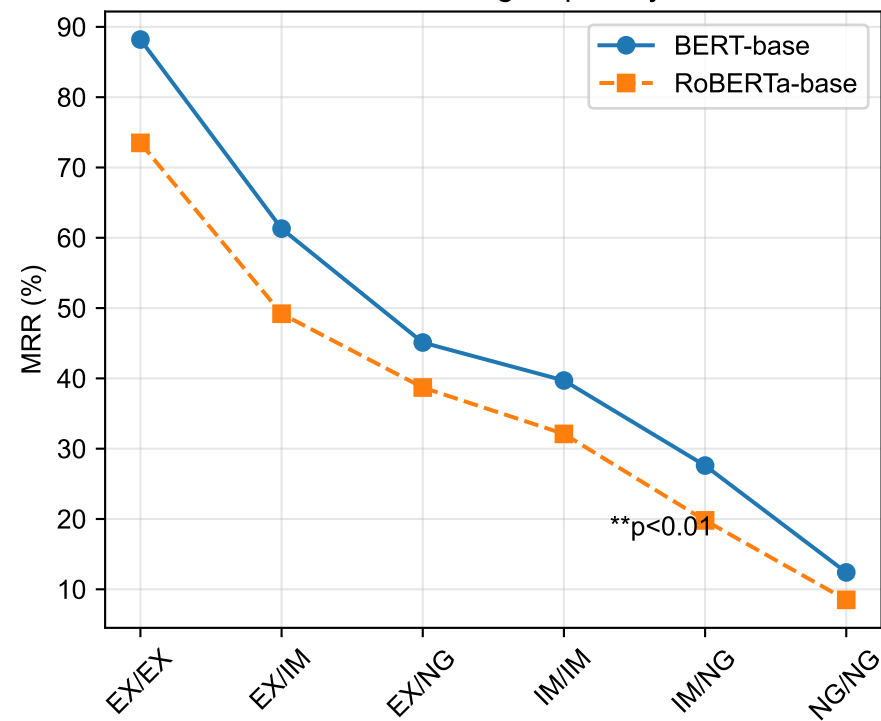


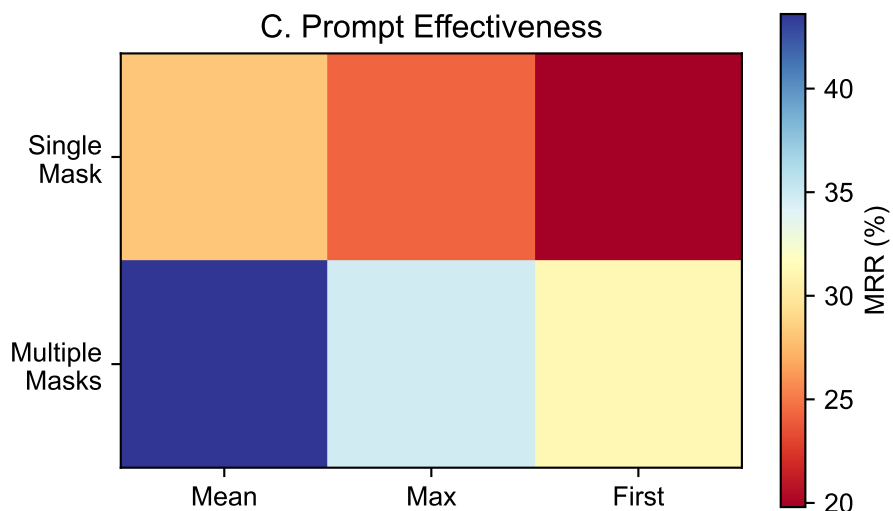
A. Memorization Performance



B. Reasoning Capability



C. Prompt Effectiveness



D. ChatGPT Benchmark

