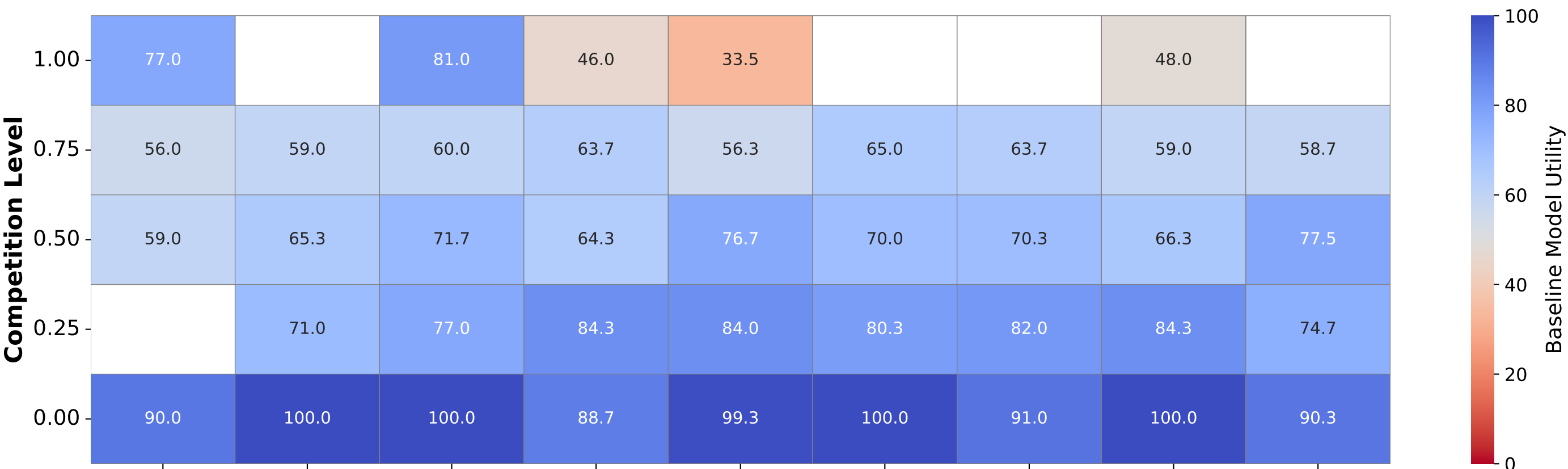
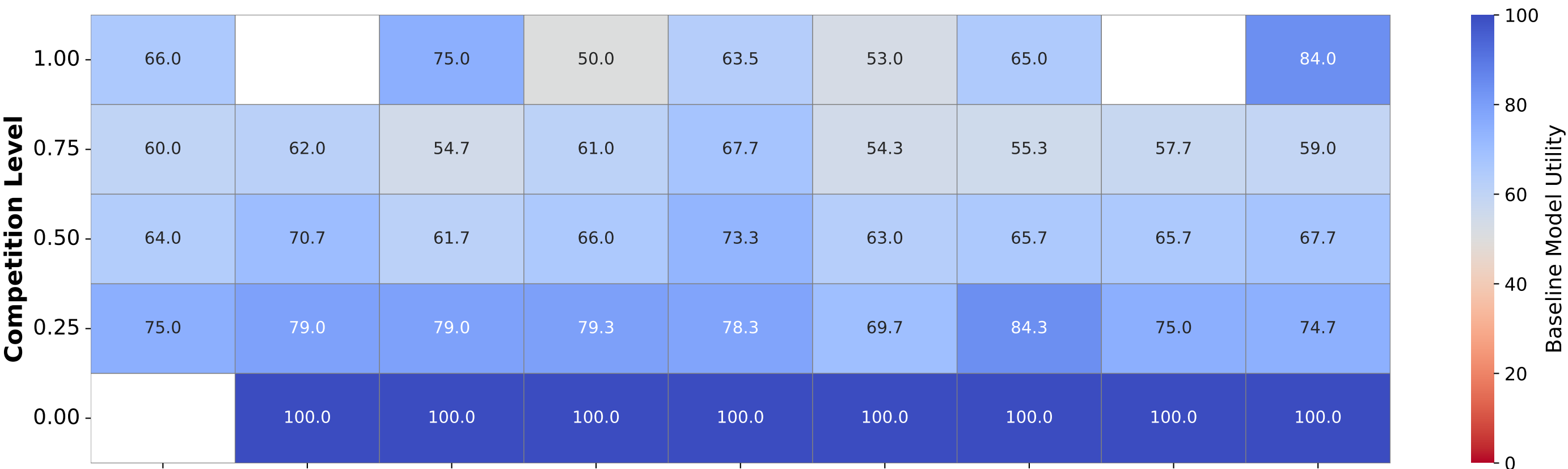


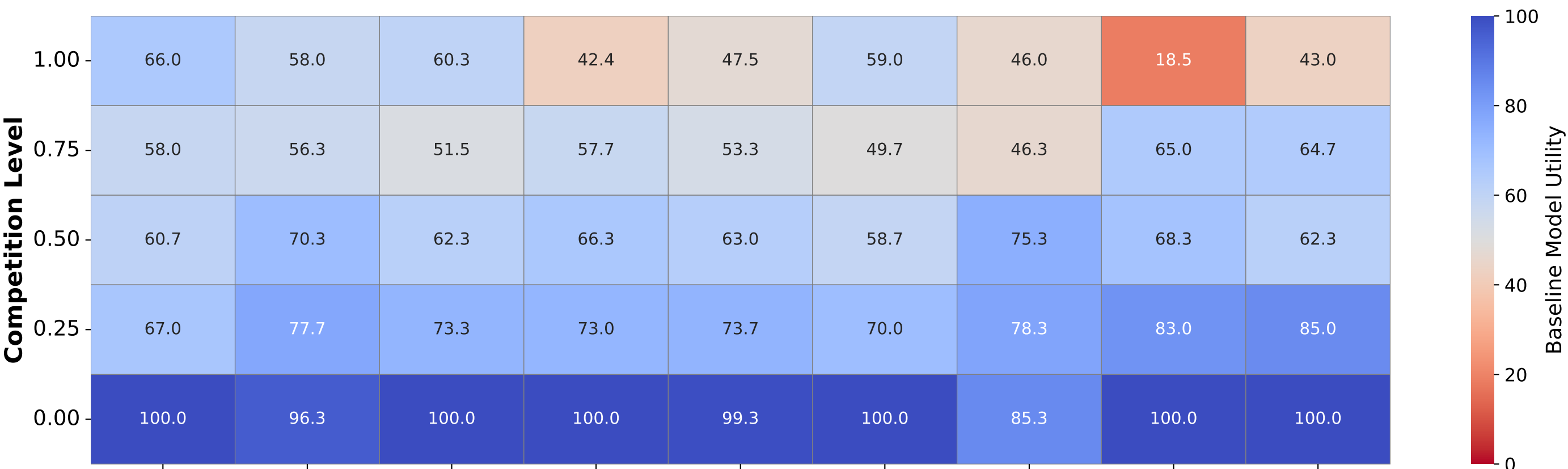
Claude 3 Opus: Baseline Final Utility



Gemini 1.5 Pro: Baseline Final Utility



GPT-4o (May 2024): Baseline Final Utility



Adversary Models Ordered by MMLU-Pro Performance →

