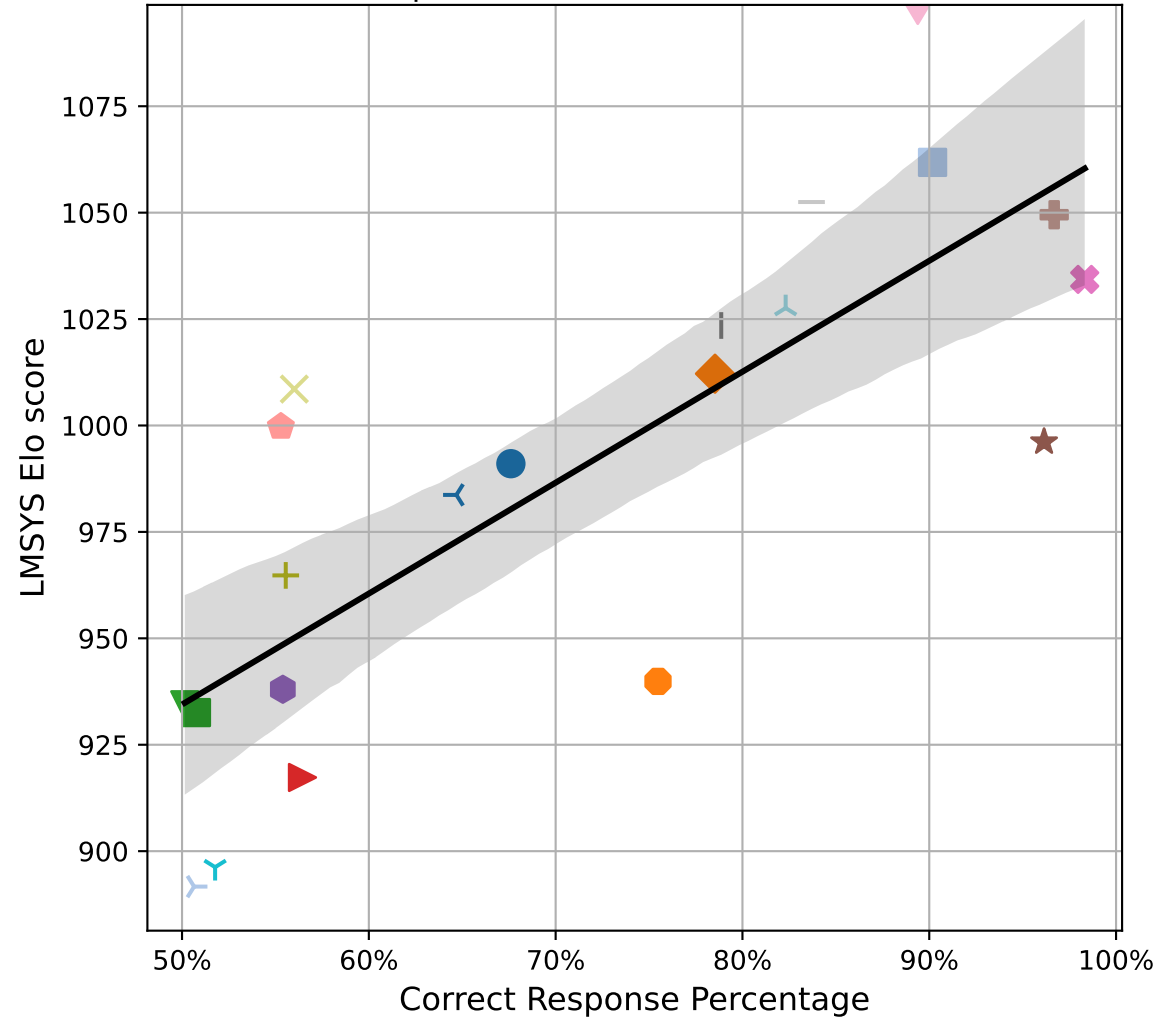


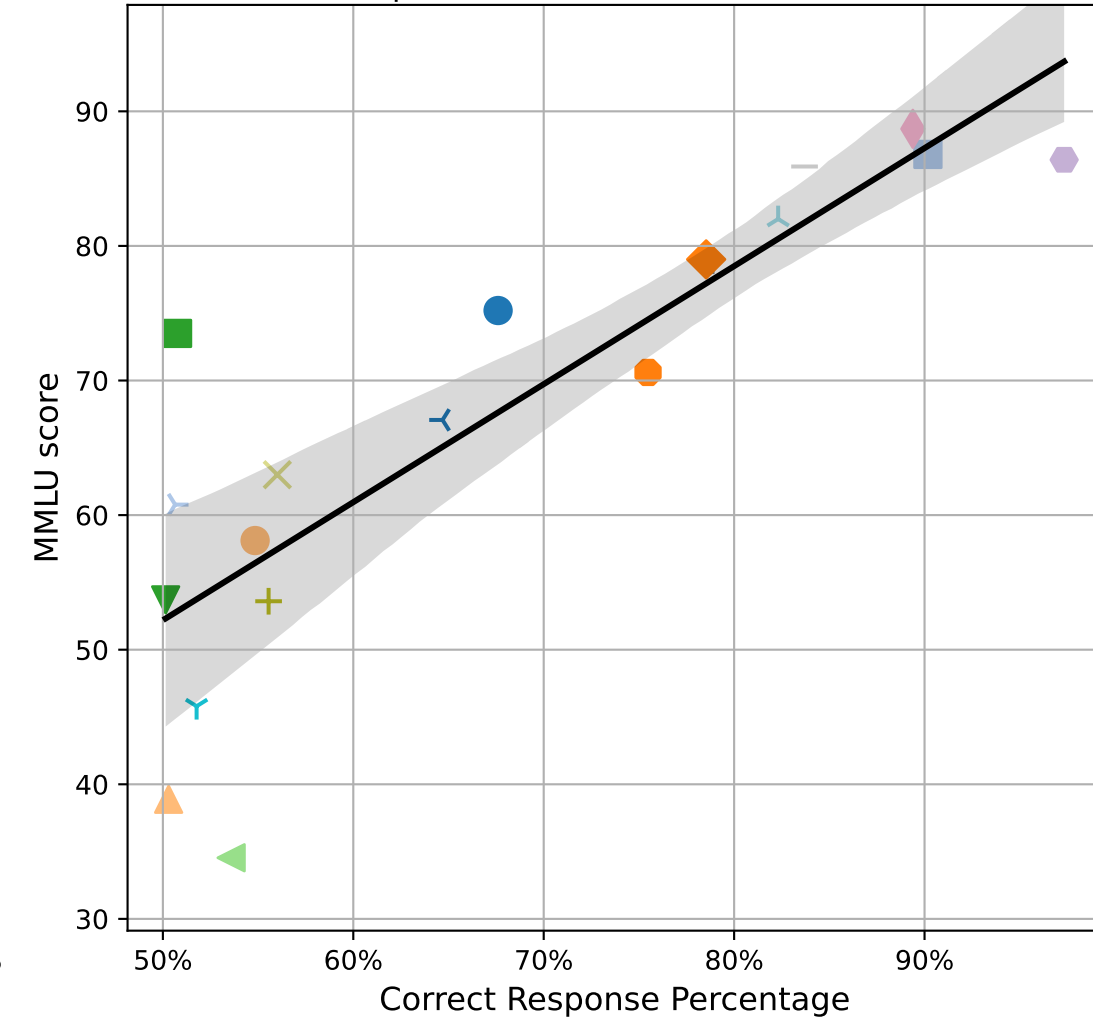
Performance vs. LMSYS Elo score

Temperature: 1, Condition: few-shot



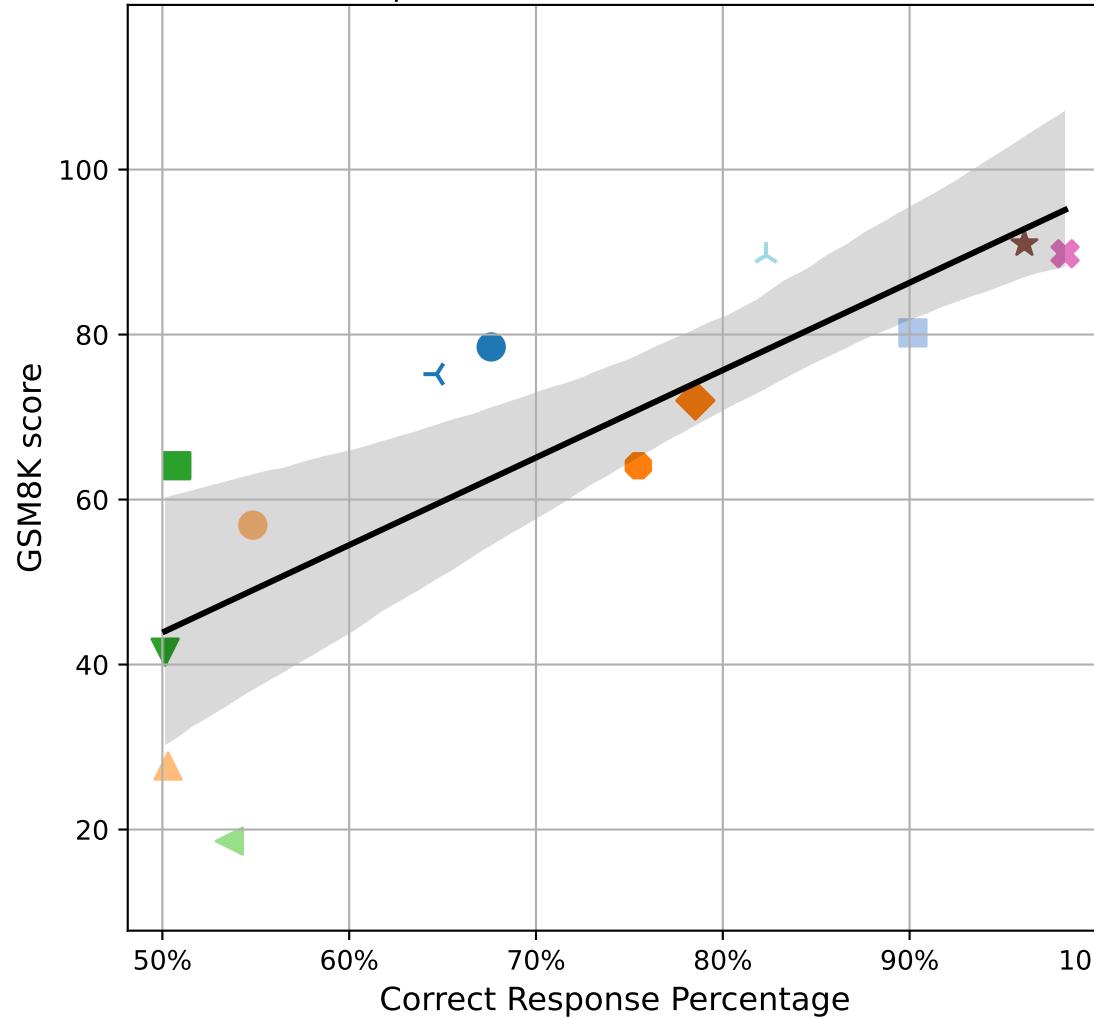
Performance vs. MMLU score

Temperature: 1, Condition: few-shot



Performance vs. GSM8K score

Temperature: 1, Condition: few-shot



- Model
- GPT-4 Turbo (2024-04-09)
 - GPT-4 Turbo (1106)
 - GPT-4 (0613)
 - Claude 3 Opus
 - GPT-4o (2024-05-13)
 - Gemini 1.5 Pro
 - Llama 3 Instruct 70B
 - Gemini 1.5 Flash
 - Claude 3 Sonnet
 - Mixtral 8x7B
 - Claude 3 Haiku
 - Llama 3 Instruct 8B
 - GPT-3.5 Turbo (0125)
 - Llama 2 Chat 70B
 - Llama 2 Chat 13B
 - GPT-3.5 Turbo (1106)
 - GPT-3.5 Turbo (0613)
 - Llama 2 Chat 7B
 - Yi Chat 34B
 - Mistral 7B
 - Code Llama 34B