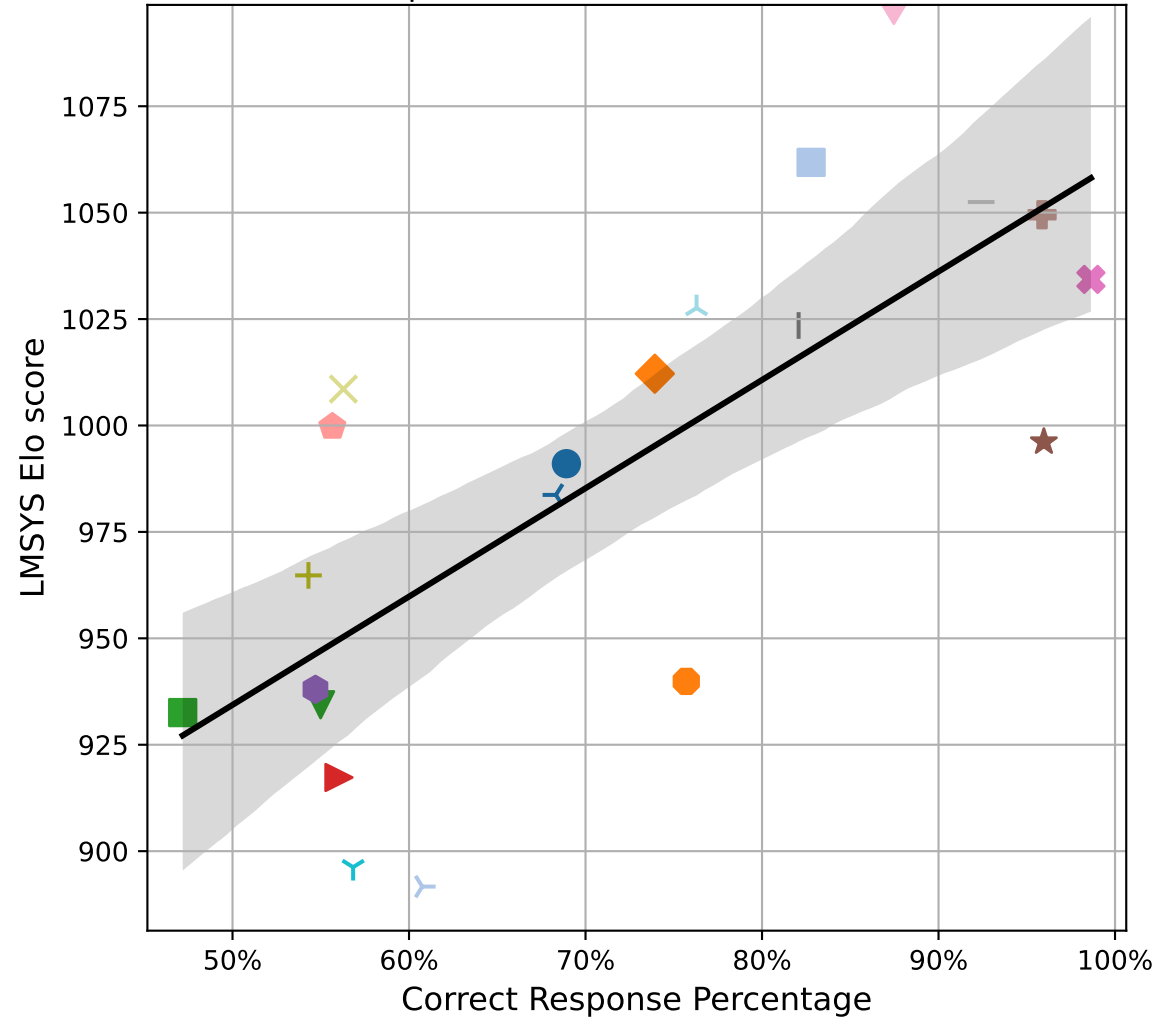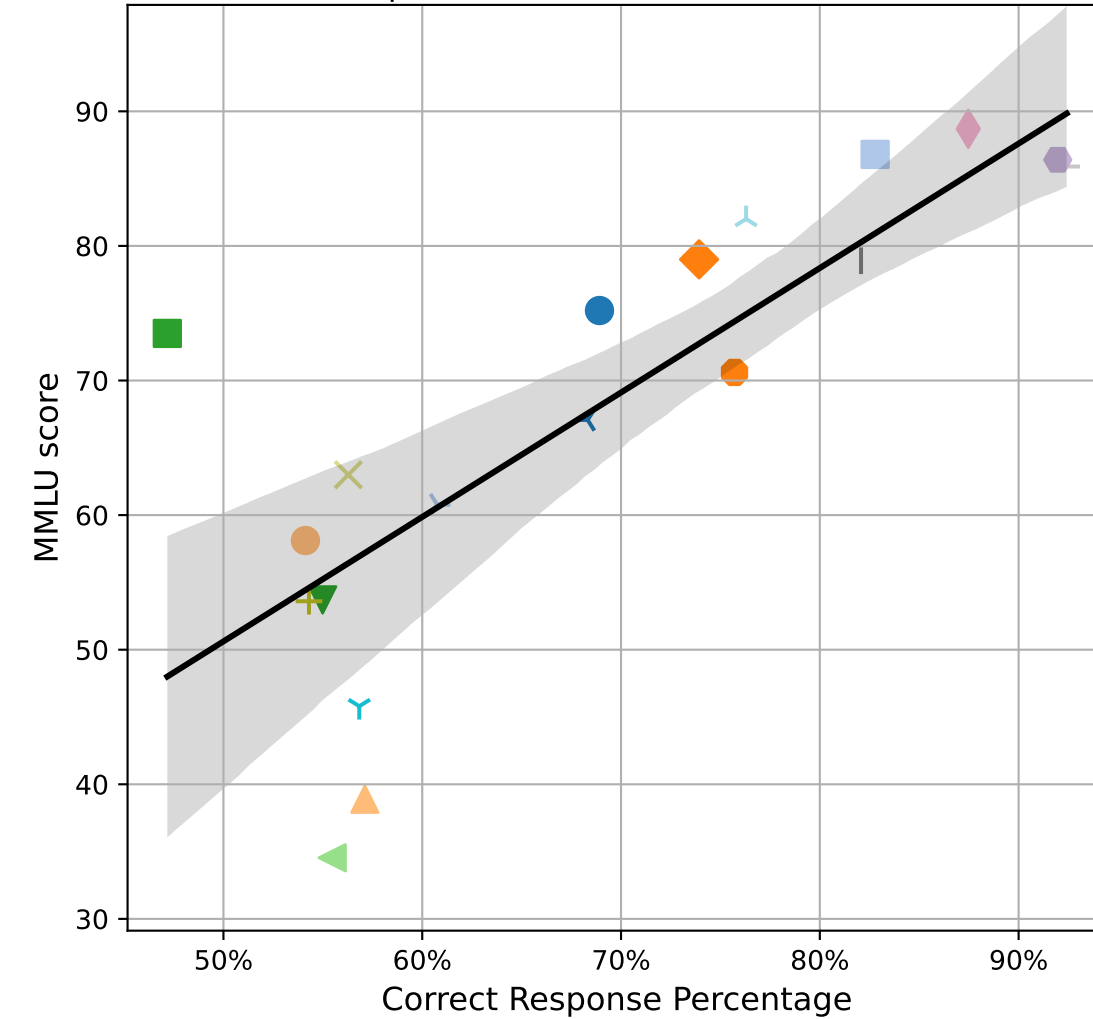Performance vs. LMSYS Elo score — Temperature: 1, Condition: zero-shot
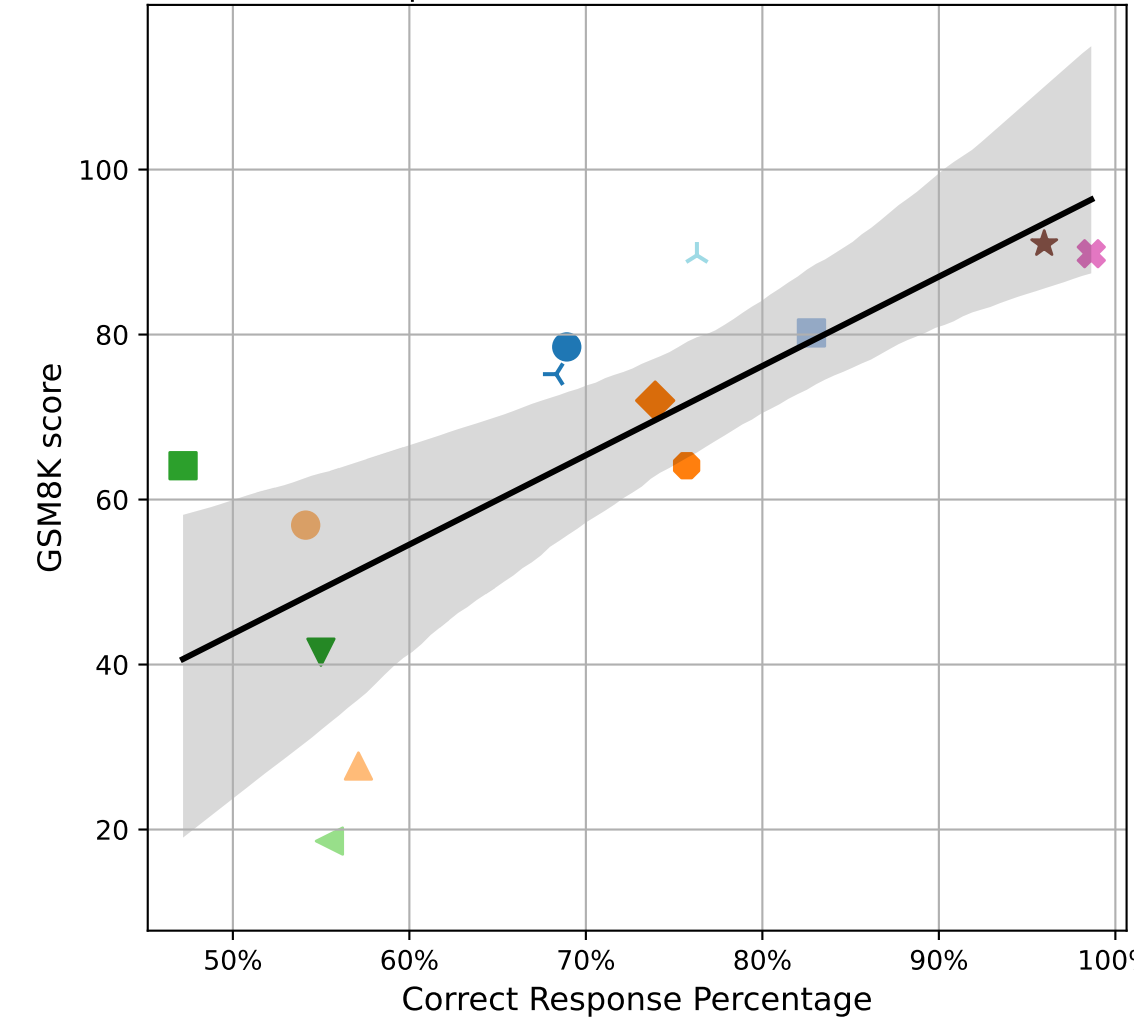
Performance vs. MMLU score — Temperature: 1, Condition: zero-shot

Performance vs. GSM8K score — Temperature: 1, Condition: zero-shot

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