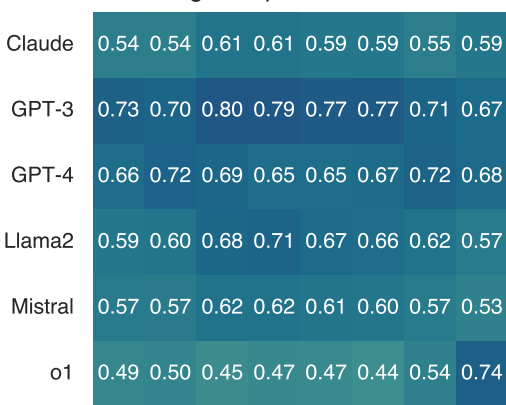


Negative predictive value

Model family (test)



GPT-3

GPT-4

Gemini

Llama2

Mistral

PaLM2

Qwen

o1

Model family (training)