Overall Accuracy 100 96.8 96.0 95.6 93.6 77.6 **GPT-4** Turbo **-** 90 75.2 60.0 58.4 GPT-3.5 Turbo 71.6 67.6 - 80 91.4 96.0 91.2 71.2 Llama3 70B 81.2 79.2 57.2 Mixtral 8*22B 85.6 81.6 67.2 - 70 77.2 56.8 58.4 70.8 70.4 Llama3 8B **-** 60 67.2 60.4 58.0 58.0 57.6 Mistral 7B Original +16 +2 +4 +8