Random forest pairwise classification accuracy, distinguishing each LLM from human text

LLM	Trained on HAP-E		Trained on CAP	
	Training	Test on CAP	Training	Test on HAP-E
GPT-4o				
GPT-4o	98.4%	98.4%	98.8%	97.4%
GPT-4o Mini	98.7%	98.1%	99.2%	97.6%
Llama Instruct				
Llama 3 70B Instruct	94.7%	93.7%	94.4%	91.8%
Llama 3 8B Instruct	95.5%	95.4%	96.2%	94.2%
Llama Base				
Llama 3 70B	93.2%	94.6%	95.7%	89.1%
Llama 3 8B	93.9%	93.7%	95.5%	89.3%