GPT-4o (■ Train ■ Test)   Gemini-1.5 (■ Train ■ Test)   Llama3.1-70b (■ Train ■ Test)