

fastapi==0.109.0 uvicorn==0.27.0 pydantic==2.6.0 openai==1.12.0 torch==2.1.2
transformers==4.37.0 peft==0.8.2 trl==0.7.10 bitsandbytes==0.42.0 qdrant-client==1.7.0
prometheus-client==0.19.0