

خلاصه

مهندس هوش مصنوعی با تخصص در طراحی و پیاده‌سازی سیستم‌های مبتنی بر LLM؛ از معماری Agent و Tool Calling تا RAG، Vector Search، Fine-tuning، Evaluation و Guardrails. با تمرکز بر ساخت اپلیکیشن‌های هوش مصنوعی **production-oriented** و سیستم‌های تست‌پذیر، امن و قابل نگهداری. دارای کارشناسی ارشد مهندسی برق با گرایش هوش مصنوعی، با پیشینه Machine Learning، Data Science و Data Engineering.

پروژه شاخص

۱۴۰۵ - در حال توسعه

Astra — Enterprise AI Agent Platform

طراحی، معماری و پیاده‌سازی مستقل و end-to-end

github.com/hnjgp/astra-ai-platform

Python · FastAPI · LangGraph · OpenAI · PostgreSQL + pgvector · Docker · Prometheus ·

GitHub Actions

CI eval gate

Recall@k · MRR · routing

93%

test coverage

407

automated tests

- Astra چیست؟** بک‌اند یک دستیار هوش مصنوعی سازمانی. ایجنت خودش تصمیم می‌گیرد برای پاسخ به اسناد داخلی (RAG)، ابزارهای سیستمی، هر دو یا هیچ‌کدام نیاز دارد؛ و هر مرحله داخل مرزهای امنیتی اجرا می‌شود.
- Agent با LangGraph:** state machine با چرخه model → tools → model، human-in-the-loop و interrupt با LangGraph برای ابزارهای حساس، با checkpoint روی PostgreSQL.
- Agentic RAG:** retrieval به صورت tool که مدل در صورت نیاز صدا می‌زند؛ chunking با overlap، text-embedding-3-small، جست‌وجوی برداری با HNSW index در pgvector و reranking، permission-aware retrieval.
- Tool layer و امنیت:** ToolExecutor متمرکز با allow-list، RBAC، schema validation و retry؛ JWT auth، guardrail ورودی و خروجی، تشخیص prompt injection (انگلیسی و فارسی)، secret redaction و rate limiting.
- Observability و MCP:** request id، JSON logs، متریک‌های Prometheus برای model، embedding، tool و RAG، ردیابی token usage و هزینه؛ MCP server روی stdio.
- Quality و Delivery:** ۴۰۷ تست unit، integration و end-to-end با pgvector واقعی، پوشش ۷۹٪ (Recall@k)، evaluation harness (MRR، route accuracy، groundedness) به عنوان regression gate در CI و Docker Compose و GitHub Actions.

تخصص اصلی

LLM Application Engineering

Agentic AI & Multi-Agent Systems

RAG & Semantic Retrieval

LLM Fine-tuning (QLoRA / PEFT)

Retrieval & LLM Evaluation

AI Guardrails & Security

Production AI Backend (FastAPI)

Tech Stack

Agents & LLM

OpenAI / Anthropic APIs

LangGraph

LangChain

OpenAI Agents SDK

CrewAI

AutoGen

MCP

Hugging Face

QLoRA / PEFT

Tool Calling

Structured Outputs

Vector Search & RAG

pgvector

Chroma

FAISS

Embeddings

Chunking

Reranking

Recall@K

Deployment & Backend

Python

FastAPI

PostgreSQL

Docker

pytest

Git

AWS (Bedrock, SageMaker, Lambda)

Terraform

GitHub Actions

Modal

Machine Learning

PyTorch

scikit-learn

XGBoost

NLP

Computer Vision

Pandas / NumPy

Data Engineering

Spark

Kafka

Airflow

SQL Server

ETL / ELT

مرحله طراحی معماری

Azmanegar — AI Clinical Decision Support

پلتفرم Clinical Decision Support مبتنی بر LLM: ingestion نتایج آزمایشگاهی با FHIR / HL7، نرمال سازی UCUM / LOINC و لایه ML برای anomaly detection، trend forecasting و تحلیل dose-response روی داده های بیمار.

سایر پروژه های LLM و Agentic AI

- **Fine-tuning با QLoRA**: فاین تیون Llama 3.1 8B برای پیش بینی قیمت محصول از روی توضیحات و مقایسه با baseline های ML سنتی و frontier model های فاین تیون شده با Hugging Face و Weights & Biases.
- **Autonomous multi-agent system**: کشف پیشنهادهای خرید با ترکیب مدل فاین تیون شده، RAG روی Chroma و ensemble model؛ serverless deployment روی Modal.
- **Trading agents با MCP**: ایجنت های معامله گر خودگردان با دسترسی به market data، حساب و memory از طریق MCP server ها (OpenAI Agents SDK).
- **Deep research agent و multi-agent teams**: پژوهش وب موازی با OpenAI Agents SDK، browser agent با LangGraph و Playwright، و تیم های نقش محور با CrewAI و AutoGen.
- **Production deployment روی AWS**: سرویس های LLM و Agent با Bedrock، SageMaker و Lambda؛ Infrastructure as Code با Terraform و CI/CD با GitHub Actions.

سوابق کاری

- ۱۴۰۴ **کارشناس ارشد تعمیر و نگهداری پنل های خورشیدی**
نیروگاه تأمین انرژی تابان (بانک پاسارگاد)
- ۱۴۰۳ - ۱۴۰۱ **تحلیلگر فاندامنرال و معامله گر بازارهای مالی**
شرکت فناوری های نوین اقتصاد
- تحلیل فاندامنرال و معامله گری در بازارهای مالی و رمزارز؛ ارزیابی پروژه های بلاکچینی و توسعه trading bot با Python.
- ۱۴۰۱ - ۱۴۰۰ **کارشناس فنی بررسی و توسعه برنامه**
کافه بازار
- ۱۳۹۹ - ۱۳۹۸ **کارشناس تحقیق و توسعه**
مهندسی و ساخت و کنترل مپنا (مکو)

پروژه های Machine Learning

- **Fraud detection** در تراکنش های اتریوم روی imbalanced data با XGBoost
- **Sentiment-driven forecasting**: اثر توییت ها بر قیمت سهام Nvidia با NLP و time series
- **Computer vision**: طبقه بندی علائم ترافیکی (GTSRB) با CNN و سامانه هشدار سبقت خطرناک مبتنی بر image processing

کارشناسی ارشد مهندسی برق

گرایش هوش مصنوعی
دانشگاه امیرکبیر (واحد تفرش)
۱۳۹۷ - ۱۳۹۵

کارشناسی مهندسی برق الکترونیک

گرایش مدارات مجتمع
دانشگاه امیرکبیر (واحد تفرش)
۱۳۹۴ - ۱۳۹۰

دوره ها و گواهی نامه ها**LLM Engineering: Master AI, Large Language Models & Agents**

Ed Donner · Udemey

Machine Learning Specialization

Andrew Ng · DeepLearning.AI & Stanford

Financial Trading in PythonDataCamp · 2025
datacamp.com/completed/.../cc3de6b4**زبان**

انگلیسی (Advanced)

حوزه های علاقه

- پروژه های بلاکچینی
 - بازارهای مالی و معاملات الگوریتمی
 - ترکیب هوش مصنوعی و بازارهای مالی
- دارای ۵ سال سابقه در تحلیل فاندامنرال، معامله گری و پروژه های بلاکچینی