

Arta Asadi

Senior AI Engineer

Age: 30, Location: Iran, Tehran, District 8



Specialist in Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs), including agentic RAG architectures and open-source LLM implementations. Proven expertise in Reinforcement Learning from Human Feedback (RLHF) through advanced research and thesis work. Skilled in designing AI-driven Forex trading platforms using deep learning, reinforcement learning, and quantum-informed models.

Education

- Ph.D. student in Artificial Intelligence at Azad University, Central Tehran Branch. My Ph.D. topic is RLHF.
- Master's degree in Computer Engineering with a specialization in Artificial Intelligence from Semnan University.
- Bachelor's degree in Software Engineering from Gonbad University.

Experience

Since May 2025

Asus Iran (Pinki-shop, Vida byte AI): Senior AI Developer

- Architecting and delivering enterprise-grade AI solutions to elevate customer service, recommendation engines, and sales chat experiences.
- LangGraph Chat Systems: Developing multiple dynamic conversational agents using LangGraph, integrating retrieval-augmented flows with vector databases (e.g., Pgvector, Qdrant).
- End-to-End Microservices: Designing, containerizing, and orchestrating AI microservices with Docker and FastAPI for scalable, production-ready deployments.
- Vue.js Frontend Integration: Building responsive, user-centric interfaces in Vue.js to seamlessly connect users with underlying AI services.
- Vector Storage & Databases: Implementing and tuning vector storage solutions alongside relational and NoSQL databases to optimize similarity search and data retrieval.

- Customer Service & Recommendation: Crafting AI-driven chatbots and recommendation pipelines that personalize interactions and drive customer satisfaction.
- Sales Enablement: Engineering conversational workflows for lead qualification, product upselling, and support chat to boost conversion rates and revenue.

Oct. 2024 – May 2025

- Smart Land Solutions (SLS): Machine Learning Engineer, AI Developer, LLM and RAG Systems Designer
- AI and Machine Learning Developer: Specializing in LLM chains and custom LLM-based assistants, leveraging agentic RAG systems using Phidata, and langchain.
- Design & Implementation: Expertise in designing, implementing, training, fine-tuning, and deploying LLMs for specialized applications, including using open LLMs like LLaMA and Qwen for text-to-SQL generation.
- Implementing high-scale and production-ready dynamic graph RAG applications with LangGraph and LangChain
- Advanced Systems Development: Developing LLM-based systems and Retrieval-Augmented Generation (RAG) models for enhanced customer interactions; used Phidata for agentic RAG combined with vector DBs like Pgvector.
- Marketplace Assistants: Building assistants to facilitate customer-supplier interactions, improving service delivery with cutting-edge AI solutions.
- Technological Expertise: Proficient with core technologies including LLMs, Phidata, LangChain, LangGraph, Crew.ai, PyTorch, and Python.
- Microservice Development: Skilled in utilizing FastAPI and Docker for building microservices to efficiently serve AI services.
- Production Environment Experience: Demonstrated ability in handling heavy LLM models in production environments with high performance and reliability.

Since 2022 -2024 (Part-time)

- AI Engineer in the Financial Industry, Trader Model Developer at the R&D Center of MCI
- Implementing machine learning and reinforcement learning models for trading in financial markets.
- Designing Forex trading robots using artificial intelligence and other analytical methods.
- Leading the design and development of AI-based trading robots, integrating various machine learning and deep learning models to optimize trading strategies.
- Primary goals include improving accuracy, maximizing profit, and reducing risk.

- Core activities involve model design, implementation, monitoring, training, and testing (including transformers, deep learning, quantum physics, and probabilistic models).
- Managing the end-to-end execution of the AI-based trading robot pipeline.
- Achieving hybrid models aimed at reducing risk, improving accuracy, and increasing profitability for the organization.

2022

- Data Scientist at DigiNext, Dynamic Pricing System Design
- Time series analysis, data processing, and product pricing system design
- Implementing data and machine learning pipelines.
- Designed a pricing system and price control mechanism for Digikala's collaborative online marketplace.
- Developed a data processing pipeline for product pricing, using various time series transformers.
- The project aimed to control price fluctuations and volatility caused by sellers and suppliers, ultimately increasing market share.
- The project was outsourced following a change in management policies, concluding my collaboration with the organization.

2021-2023 AI Engineer in an outsourced crypto exchange platform

Development of the trading system core (brokerage system)

- Portfolio analysis
- Portfolio optimization with AI
- Backend implementation using FastAPI
- Implementation of core of the decentralized brokerage with Solana
- Designed the core of a brokerage system, including order management, order processing, and matching of buyers and sellers.
- Implemented the system with multiple components, with Solana app as the core, managing buy and sell requests, order books, liquidity pools, and other trading-related elements.
- Served as the Solana (Rust) programmer and AI engineer for additional tools implemented within the system.
- Additionally, beyond the trading core, I implemented various components within the same project, such as portfolio review, a stock recommendation system, a chatbot for analytical assistance, and other minor features. The models used in each section were diverse and varied, encompassing the complete AI pipeline and services for order book and trading.

Main Skills

- **Programming Languages:** Python, Beginner Rust
 - **Artificial Intelligence & Machine Learning:** Deep Learning, Machine Learning, Reinforcement Learning, Transformers
 - **LLM & Specialized Techniques:** LLM, RLHF (Reinforcement Learning with Human Feedback)
 - **Retrieval-Augmented Generation (RAG):** Agno (Agentic RAG), Crewai, LangChain, LangGraph (high-scale and production level developments), Multiple vector DBs
 - **Data Science & Analytics:** Data Science, Database, Data Modeling, Pandas, MongoDB, Duckdb(for integration with Phidata)
 - **Time Series & Anomaly Detection:** Time Series Analysis, SKtime, PyOD
 - **Risk Management & Financial Strategies:**
 - Risk Management in Forex
 - Tail Hedging & Focus on Tail Risk
 - AI in Hedging and Risk Management
 - Conformal Prediction
 - Fat Tail Familiarity
 - Power Law & Power Law Portfolios
 - Portfolio Optimization
 - Applying Reinforcement Learning in Tail Risk & Forex Risk Management(Position sizing, stop trailing, high probability trading, etc.)
 - **Dynamic Pricing Solutions**
 - **Web & Microservices Development:** FastAPI, Django, Streamlit
 - **Containerization & Deployment:** Docker, FastAPI connected with AI services
 - **Operating Systems:** Linux (Ubuntu, CentOS)
 - **Big Data Technologies:** Big Data, Hadoop, Spark
 - **CUDA Programming**
 - **MLOps & Model Deployment:** MLOps (FastAPI, Streamlit), Docker
-
- Familiarity with Financial Data, Financial Markets, and the Banking and Financial Industry, Risk management
 - Familiarity with Forex, Stock, and Cryptocurrency Markets
 - Possess several certifications in Financial Engineering and Modeling
 - Knowledge of Financial Statistical Models
 - High Communication Skills, Teamwork Enthusiast
 - Proficient in English for Professional and Work Communication
 - Interested in Solving Innovative and New Challenges
 - Additionally, due to my academic background, I am familiar with new topics and models in Artificial Intelligence

Certificates

- Generative AI with Large Language Models [Show](#) | Coursera, Deeplearning.AI
- Generative AI for Everyone [Show](#) | Coursera, Deeplearning.AI
- Advanced PyTorch Techniques and Applications [Show](#) | Coursera, Packt
- Unsupervised Learning, Recommenders, Reinforcement Learning [Show](#) | Coursera, Deeplearning.AI
- Calculus for Machine Learning and Data Science [Show](#) | Coursera, Deeplearning.AI
- Introduction to Concurrent Programming with GPUs [Show](#) | Coursera
- Introduction to Financial Engineering and Risk Management [Show](#) | Coursera, Deeplearning.AI
- Understanding Financial Markets [Show](#) | Coursera, university of Geneva
- Advanced Topics in Derivative Pricing [Show](#) | Coursera
- Advanced Portfolio Construction and Analysis with Python [Show](#) | Coursera, EDHEC
- Market Microstructure [Show](#) | Coursera, Università di Napoli Federico II

Showcase Projects

LangGraph Chat System [Link](#)

- Technologies: Python · LangGraph · FastAPI · Vue.js · PostgreSQL · Redis
- Summary:
- Enhanced AI backbone: LangGraph-powered conversation flows with multi-LLM
- provider fallbacks
- Enterprise user management: JWT auth, role-based access control, admin
- dashboard
- Real-time comms: WebSocket messaging, persistent conversation history,
- external API integrations
- Microservices design: Dockerized services, API gateway, horizontal scaling, clear
- service boundaries
- Modern frontend: Vue.js SPA with dark/light themes, mobile-first responsiveness
- Production readiness: CI/CD pipelines, automated testing, centralized logging &
- Monitoring

Custom LLM Chatbot System [Link](#)

- Technologies: Python · PyTorch · Hugging Face Transformers · FastAPI · vLLM
- Summary:
- Flexible training pipeline: supports SFT, LoRA and DPO with memory-efficient
- optimizations

- High-performance serving: interchangeable backends (PyTorch, ONNX, vLLM)
- under FastAPI
- Enterprise monitoring: Prometheus metrics, Streamlit dashboard, health-check endpoints
- Scalable architecture: load-balancing, request batching, GPU optimization, WebSocket streaming
- Developer DX: modular design, YAML-driven configs, full test coverage, rich documentation
- Multi-format data handling: built-in preprocessing and experiment tracking

AI-Powered Crypto Trading Bot (Telegram order processing system) [Link](#)

- Technologies: Python · FastAPI · Vue.js · TypeScript · PostgreSQL · Redis
- Summary:
- Natural-language trading: execute buy/sell orders via plain-English commands through Telegram
- Multi-exchange support: real-time data & trading on Binance, Coinbase Pro, Kraken
- (via CCXT)
- Advanced risk management: automated risk scoring, intelligent stop-loss/takeprofit, portfolio analytics
- Enterprise-grade security: HashiCorp Vault integration, JWT auth, comprehensive audit logging
- Multi-platform UI: responsive web dashboard, mobile app, Telegram bot and REST API
- Live analytics: TradingView-style charts, technical indicators, performance metrics

[Link to showcase page](#)

Contact

Iran, Tehran

+989113802615

Artaasd95@gmail.com

[Linkedin](#), [Github](#)