Sepehr Rezaee

AI Engineer & System Architect

✉ sepehrrezaee2002@gmail.com
🌐 https://sepehrrezaee.github.io
SepehrRezaee
in sepehr-rezaee

Summary

Seasoned AI Engineer and System Architect with over 3 years of experience delivering scalable, high-performance software and machine learning solutions. Proven track record boosting user engagement by 25%, reducing deployment time by 40%, and cutting operational costs by 20%.

Core Competencies

Machine Learning & AI: Recommender Systems, NLP, LLMs (GPT, LLaMA, Falcon3, Qwen-2.5), Retrieval-Augmented Generation

Cloud & DevOps: AWS (SageMaker, EC2), GCP, Azure, Docker, Kubernetes, CI/CD

Software Engineering: Python, C#, .NET, FastAPI, Django REST Framework, BizAgi

QA & Automation: Selenium, Robot Framework, Git, MLflow

Data & Databases: PostgreSQL (>3 TB), MongoDB, Oracle, ETL

Leadership: Agile/Scrum, stakeholder management, team mentoring

Experience

Aug 2024–Present — **AI Engineer (Contract)**, *PropTy Global*, Remote
- Developed AI-driven chatbot & recommendation engine, increasing user engagement by 25%.
- Engineered FastAPI microservices with PostgreSQL (3 TB+) and MongoDB, ensuring API response times <100 ms.
- Deployed AWS SageMaker pipelines on spot instances, reducing costs by 20%.
- Containerized services with Docker & Kubernetes, cutting deployment time by 40%.
- Implemented monitoring (Prometheus, Grafana, ELK) for 99.9% uptime.

Oct 2022–Present — **Chief AI Officer (CAIO)**, *Novel Mind Scientist*, Tehran, Iran
- Spearhead AI-agent development for Medical, Education, and Oil & Petroleum sectors at a solution-focused company.
- Leverage state-of-the-art tools (LLMs, multimodal architectures) to build bespoke AI applications.
- Launched **ADiagnosis**, a multimodal AI agent for early diagnosis of mental disorders, integrating clinical data and image/text analysis.

Mar 2023–Feb 2024 — **Project Manager**, *NovaVira*, Tehran, Iran
- Led a 5-person team delivering a Django-based ML recommender with 85% precision & 75% recall.
- Streamlined Agile workflows and CI/CD, reducing development cycles by 30%.
- Architected GCP data pipelines processing 50 K+ records/day with 98% SLA compliance.
- Met all milestones on time.

Apr 2022–Sep 2022	**System Engineer**, *Mohaymen ICT*, Tehran, Iran
	○ Designed C# and BizAgi workflows automating 10+ core processes, reducing execution time by 30%.
	○ Improved system uptime by 15% and reduced support tickets by 40%.
	○ Implemented Azure dashboards for real-time monitoring, boosting compliance visibility by 20%.
Jan 2022–Jun 2022	**Process Developer**, *Mohaymen ICT*, Tehran, Iran
	○ Defined and implemented business processes in BizAgi, improving throughput by 25%.
	○ Developed Azure-hosted analytics dashboards, reducing manual reporting by 50%.
Oct 2021–Mar 2022	**Software Developer & QA Tester**, *Valmart (Online exchange)*, Tehran, Iran
	○ Built Python/Selenium test suites, increasing automation coverage from 60% to 95%.
	○ Identified over 200 defects, lowering regression rates by 70%.
	○ Optimized Oracle queries, improving performance by 50%.

Education

2021–2025 (expected)	**B.Sc. in Computer Science**, *Shahid Beheshti University*, Tehran, Iran
	GPA: 3.4/4.0

Skills

Programming	Python, C#, C++, Java, MATLAB
Frameworks	PyTorch, TensorFlow, FastAPI, Django REST, Scikit-Learn, OpenCV
DevOps & Tools	Docker, Kubernetes, Jenkins, GitLab CI, MLflow, Prometheus, Grafana
Databases	PostgreSQL, MongoDB, Oracle
Cloud	AWS, GCP, Azure
Languages	Persian (Native), English (Professional)