

Alireza Nasiri

Senior Software Engineer | Backend · Python & Go

Utrecht, Netherlands

[linkedin.com/in/ialireza/](https://www.linkedin.com/in/ialireza/)

Summary

Senior backend engineer with **7+ years** building and running production platforms in fintech, marketplaces, and AI, working primarily in **Python and Go** with a deep grasp of distributed systems, concurrency, and system design. I own services end-to-end, from architecture and implementation through deployment, monitoring, and production support, and most of my work is making systems hold up under real load: I raised a KYC service's weekly availability from 70% to 99%, scaled B2B cash-in flows 60x, and cut production incidents on a review platform serving 2,000,000+ daily requests. Recently I've been shipping LLM-backed features in production, with schema-enforced output and cost and latency tracking. I write thoroughly tested code, keep CI green, and watch my own changes in production.

Skills

Languages	Python · Go · Java · SQL
Backend & Frameworks	Django · REST & gRPC APIs · Microservices · Event-Driven Architecture · Kafka / RabbitMQ · System Design · Automated Testing
AI & LLM	LLM / AI Agents
Databases	PostgreSQL · MySQL · Redis
Cloud & Infrastructure	Kubernetes · Docker · AWS · Terraform · Ansible · CI/CD (GitOps)
Observability	Prometheus · Grafana · Sentry

Experience



Senior Software Engineer

Voicedash · *AI voice dictation and meeting-notes app*

Mar 2026 - Present

- Built the database-driven model-routing engine for the transcription pipeline, selecting STT and LLM-polish providers per request with per-variant timeouts, single-level fallback, and retries, cutting **p99 latency by about 33%** (9.5s to 6.4s) versus the legacy route.
- Integrated several OpenAI-compatible STT providers along with a Telegram-based **shadow-evaluation** system that scores alternative models on copied live traffic without changing user-facing responses.
- Made LLM output reliable in production by enforcing a **JSON schema** through `response_format`, adding a defensive parser that never shows the user a damaged response, and keeping the system prompt static so it can be cached.
- Instrumented the backend with **Sentry** tracing, per-request and LLM-cost metrics, and latency distributions tagged by output language (570K+ data points), then tuned span volume to control cost.
- Defined the production infrastructure as code with **Terraform**, then planned and scripted the move from a US to an EU host with a rehearsal run, preflight checks, and a cutover runbook.



Senior Software Engineer

TomanPay · *Fintech: escrow and B2B payments for marketplaces like Divar*

Dec 2023 - Sep 2025 (1 year 10 months)

- Revamped the core KYC verification system, raising weekly service availability **from 70% to 99%** by refactoring legacy code and implementing robust monitoring, enabling onboarding of 1,000 new users daily.
- Engineered and maintained the core **escrow service** using Django, MySQL, and Redis, handling approximately 40 requests per second and implementing the hold-and-release flow that keeps the payer's funds in escrow until they confirm the goods are as described.

- Built B2B cash-in services, scaling daily fund inflows **60x** (from 1B IRT to 60B IRT) and reducing manual operational workload by 90%.
- Developed an exchange and order book service in **Go**, benchmarked in load testing at over 1,000,000 matches per second.
- Made money-movement flows safe under concurrency with **Redis locks**, task locking, and deadlock recovery, and added a zero-sum check on matched exchange orders.
- Brought the legacy escrow codebase under test by making the suite runnable, fixing broken tests, and adding **coverage reporting** and a formatting gate to CI.



Senior Software Engineer

Ravindex · *Fintech: stock-market and crypto analytics and trading*

May 2022 - Dec 2023 (1 year 8 months)

- Cut response times on the financial dashboards by removing Django ORM **N+1 queries**, adding missing indexes, moving the heaviest calculations into SQL-level functions, and precomputing the rest off the request path.
- Designed a **real-time** data update service using socket-based connections, integrating Tehran Stock Exchange (TSE) data via an event-driven architecture.
- Built a crypto **copy-trading engine** integrated with Binance, KuCoin, and three other exchanges that mirrored lead traders' positions for followers and computed trader PnL, rankings, and performance history.
- Built fixed-income and options analytics for the Tehran Stock Exchange, including a **yield-to-maturity** calculator, fund revenue tracking, and option chain, spread, and covered-call strategies.
- Built an order management system MVP, connected it to the market-data backend through **RabbitMQ** consumers, and set up Prometheus monitoring for both backends.



Software Engineer (Review & Car Team)

Divar · *Iran's largest online classifieds marketplace (like Craigslist)*

Aug 2018 - May 2022 (3 years 10 months)

- Owned the review pipeline and, over time, several other backend services on a review platform that grew to around 20 microservices and bots processing **2,000,000+ daily requests**.
- Cut production incidents **from about five to fewer than two per week** by tracing them to a root cause in Elasticsearch indexing; uptime rose from 98% to 99% over the same period.
- Automated CI/CD pipelines using **GitOps**, reducing deployment time from 60 minutes to under 10 minutes.
- Generated **600M IRT in revenue** by building and scaling Karnameh (Car Inspection Service) from the ground up, reaching 200 daily inspection requests by my departure in 2022.
- Deployed an optimized plate detection bot, fixing memory leaks and cutting processing time by **66%** with parallel image processing.

Education



Sharif University of Technology

B.Sc., Computer Science

Sep 2022

Honors & Awards

AWS Cloud Quest: Cloud Practitioner

Feb 2026

Training path, not the AWS Certified Cloud Practitioner exam.

Bronze Medal, Iranian National Olympiad in Informatics (INOI)

Sep 2015

