

# Hossein Molavi

hossein.molavi.2017@gmail.com

+989185841397

Tehran

## Summary

Experienced **Software Engineer and AI Systems Engineer** with 6+ years of building scalable, resilient backend platforms and production-grade AI systems across fintech, travel tech, health tech, social platforms, and cloud infrastructure. Specializes in **LLM-powered applications, agent architectures, workflow orchestration, and cost-efficient AI systems**, with a strong foundation in distributed system design and performance optimization. Known for driving product-oriented technical decisions, collaborating closely with product and design teams, and delivering reliable systems that balance AI innovation with real-world constraints.

---

## Education

### Computer Engineering

Isfahan University of Technology • Isfahan, Isfahan

05/2020

---

## Experience

### AI Engineer

Refact • Los Angeles, CA

06/2025 - Present

**Workform** is a private-beta AI project manager embedded in Slack that autonomously monitors conversations, meetings, and tools to create, update, and manage tasks for software teams.

#### Accomplishments:

- **Designed and led the implementation of production-grade AI agent systems** for a Slack-native AI project manager, integrating with external tools and real-time team workflows.
- **Reduced LLM token usage by ~60%** through agent working memory, multi-step task decomposition with deferred model routing, focused context windowing, and optimized MCP server responses, enabling cost-efficient scaling under tight budget constraints.
- Architected a **multi-phase PM agent system** separating task selection and action generation, with dynamic prompt and tool switching to maximize context efficiency and decision accuracy.
- Implemented **agentic RAG pipelines** using Weaviate to categorize and retrieve contextual data by topic, improving relevance while minimizing unnecessary token consumption.
- Orchestrated complex, long-running **AI workflows with Temporal**, including real-time Slack and Gmail monitoring, signal-based debouncing, and automated task creation, updates, and comments.
- Built and maintained **end-to-end MCP infrastructure** (servers and clients over SSE and Streamable HTTP) for integrations with Slack, Google Docs, Sheets, Calendar, Drive, and PM tools.
- Developed a **TypeScript MCP client library** with multi-server management, connection pooling, failover, and intelligent caching to simplify distributed AI tool integration.
- Designed **safe runtime tool activation mechanisms** using unified adapter and tool-registry patterns, reducing context noise and preventing unintended tool execution.
- Introduced **LLM cost and usage observability**, tracking per-operation token metrics to enable targeted optimization and informed architectural decisions while collaborating closely on architecture, delivery, and product trade-offs.

**Skills:** TypeScript, Node.js, NestJS, PostgreSQL, Event-driven systems, WebSockets, Slack & Google Workspace APIs, MCP, Vercel SDK, LLM APIs, Agent architectures, RAG, Prompt engineering, Tool calling, AI cost & usage optimization,

Temporal workflows, Weaviate

## Technical Lead

Dent Ment • Tennessee, IL

01/2025 - 06/2025

DentMent links dentists with dental mentors to master business and build thriving practices.

### Accomplishments:

- **Defined and led the technical architecture** of the platform from the ground up using **TypeScript, NestJS, GraphQL,** and **MongoDB Atlas**, delivering a scalable foundation for core features like scheduling, messaging, and content sharing.
- **Designed a tokenized content pipeline** supporting dynamic @mentions within user-generated posts, ensuring low-latency rendering and seamless interaction at scale.
- **Led the implementation of MongoDB Atlas Search**, introducing intelligent, full-text, and fuzzy matching capabilities that enhanced mentor discovery and platform engagement.
- **Established engineering velocity and reliability** by introducing Docker-based deployments and CI/CD pipelines via **GitHub Actions**, reducing integration overhead and release friction.
- **Provided technical leadership and mentorship** to a team of 3 engineers, driving system design reviews, enforcing code quality standards, and guiding implementation decisions.
- **Collaborated directly with the CEO and product team** to translate business goals into executable technical initiatives, influencing feature prioritization and sprint planning.

**Skills:** Node.js • Technology Leadership • NestJS • TypeScript • MongoDB • Amazon Web Services (AWS)

## Backend Team Lead

Dorfak Intelligent Systems • Tehran

10/2023 - 01/2025

An AI-powered fintech company specializing in algorithmic trading, portfolio management, and asset management solutions.

- Led backend development of the order execution engine; improved scalability by 50% using **Kafka partitioning** and microservice design.
- Reduced deployment time by 40% through container orchestration and **Kubernetes-based automation**.
- Boosted system performance by 30% with **Redis-backed** caching strategies and optimized query logic.
- Refactored the legacy real-time socket service for stock pricing, cutting latency by 60% and enhancing reliability.
- Migrated to a **Redis Cluster** and implemented Redis pipelines to enable horizontal scaling and improve write query throughput by 40%.
- Introduced **parallel processing** techniques to improve order processing speed by 70%, accommodating higher trading volumes.

**Skills:** Kafka • Node.js • MariaDB • TypeScript • NestJS • kubernetes • Redis • Microservices • WebSocket

## Senior Backend Developer

Alibaba Travels Co (Jabama.com) • Tehran

08/2022 - 09/2023

Iran's leading travel booking platform, offering hotel reservations and travel services to millions of users.

- Maintained and optimized the **Product Listing Page (PLP) generator**, handling hundreds of requests per second.
- Developed **SEO-focused features**, significantly improving site rankings.
- Engineered a **high-performance pricing system** capable of updating 1M+ records in under 10 seconds.
- Decreased response time from 10s to 80ms by implementing **hybrid Redis caching**.

**Skills:** RabbitMQ • Node.js • MongoDB • TypeScript • NestJS • Elasticsearch • Redis • Microservices

## Senior Backend Developer

IGMU • Copenhagen, Denmark

08/2021 - 07/2022

A global platform for certified cat breeders, helping them manage breeding operations, connect with buyers, and streamline communications.

- Developed backend services to support user registration, breeder profiles, and content moderation workflows using **NestJS** and **MongoDB**.
- Built scalable real-time chat functionality using **WebSocket**, ensuring reliable messaging across user roles.
- Integrated cloud infrastructure using **Firebase** and **Google Cloud Platform (GCP)** to streamline CI/CD and file storage.
- Contributed to database design and performance tuning to ensure system responsiveness under growing user traffic.
- Collaborated with front-end and product teams to deliver feature-rich breeder dashboards and customer tools.

**Skills:** WebSocket · Google Cloud Platform (GCP) · MongoDB · NestJS · Firebase

## Backend Developer

Magny · Melbourne, Victoria

08/2020 - 07/2021

The first Iranian social network in Australia, connecting Iranian communities with local businesses, events, and networking opportunities.

- Improved **Elasticsearch** query performance, increasing **user retention by 40%**.
- Integrated **Google Analytics & Firebase Events** for data-driven decision-making.
- Developed custom analytics dashboards using **Google Data Studio**.

**Skills:** Node.js · Serverless Computing · Cloud functions · Elasticsearch · Firebase · Cloud Firestore · React.js · Google Analytics · Google Data Studio

## Backend Developer

Mabna Cloud · Iran, Tehran

01/2019 - 07/2020

A cloud services provider, offering IaaS, PaaS, and SaaS solutions to businesses for scalable, secure, and cost-efficient IT infrastructure.

- Built a centralized logging layer for multi-service observability..
- Implemented a **queue-based infrastructure deployment system**, reducing provisioning time.
- Automated **CI/CD pipelines**, enhancing development efficiency and deployment reliability.

**Skills:** Node.js · MongoDB · TypeScript · Vue.js · Nuxt.js · Redis · Express.js · Microservices

---

## Skills

JavaScript, Node.js, TypeScript, Nestjs, ReactJS, Express.js, Vue.js, Nuxt.js, GraphQL, WebSocket, MongoDB, PostgreSQL, MariaDB, Redis, Docker, Kubernetes, CI/CD, Github Actions, AWS, Google Cloud Platform, Firebase, Serverless, Elasticsearch, Kibana, Kafka, RabbitMq, Microservices

---

## Languages

English, Persian