Esfandiar Kiani

AI Systems Architect • Enterprise Process Automation • Platform & LLMOps

+98 912 099 8129 • Esfandiar.Business@gmail.com
es-kiani.dev • LinkedIn • GitHub
Isfahan & Tehran, Iran • On-Site & Hybrid

EXECUTIVE SUMMARY

AI Systems Architect, Platform Lead, and Head of Infrastructure & AI Services at **Netzo-Cloud** with proven expertise analyzing complex organizational realities, mapping As-Is workflows, and eliminating operational bottlenecks. Combines hands-on enterprise RAG, agentic orchestration (MCP, vLLM, vector DBs), and scalable cloud infrastructure with systematic process discovery and deterministic automation. Track record bridging multi-departmental operations, embedding practical AI into daily business workflows, and translating executive vision into high-uptime, measurable efficiency gains.

PROFESSIONAL EXPERIENCE

Head of Infrastructure, Networks & AI Services (IaaS / SaaS / AIaaS) — Netzo-Cloud **2025 & 2026**

- Architected and delivered scalable multi-tenant cloud and AI datacenter infrastructure: bare-metal GPU virtualization, software-defined networking, and high-availability container clusters.
- Engineered production AIaaS and developer platform services (netzo-cloud.com / netzo.ir), establishing SLA benchmarks, internal quota management, and hardened runtime security.
- Directed platform engineering and infrastructure roadmaps, driving 99.9% uptime across production clusters while optimizing GPU inference latency and operational compute costs.

AI & RAG Engineer — LLM Architecture Team — **Toosai Platform** **Oct 2025 – Oct 2026**

- Co-architected enterprise RAG platforms and multi-agent orchestration pipelines for large-scale energy and knowledge ecosystems, serving high-concurrency queries with strict confidentiality.
- Engineered high-accuracy hybrid retrieval pipelines (dense vector search + BM25 lexical reranking), automated multimodal document parsing, and custom tool-calling agents via Model Context Protocol (MCP).
- Implemented end-to-end MLOps/LLMOps governance: streaming inference state coordination, latency/cost evaluation suites, and deterministic output validation filters.

Project Manager (Process Optimization & Operations Delivery) — **Civil & Technology Enterprise** **2023 – 2025**

- Audited organizational realities across cross-functional units: mapped As-Is processes, diagnosed friction bottlenecks, and eliminated manual handoffs through automated workflow pipelines.
- Embedded applied AI models (generative vision, acoustic quality evaluation, intelligent routing) into production business software, bridging research prototypes with commercial milestones.
- Established agile operational roadmaps, QA release gates, and SLA benchmarks to accelerate organizational throughput.

EDUCATION

M.Sc. in Artificial Intelligence — K. N. Toosi University of Technology (KNTU) **2024 – 2026 | Tehran, Iran**

- **Thesis:** On-the-Fly Aviation Risk Mitigation using AI (Supervisor: Dr. Seyed Hossein Khasteh)
- GPA: ~17/20 • Ranked **80th Nationwide** in the Iranian M.Sc. Entrance Examination (No quota).

B.Sc. in Computer Engineering **2020 – 2024 | Isfahan, Iran**

- Ranked **1st in Graduating Class** • Accelerated graduation in 7 semesters with GPA ~19/20.
- Curriculum focus: Operating Systems, Computer Architecture, Algorithms, Distributed Computing, Discrete Mathematics.

CORE TECHNICAL PROFICIENCIES

AI & Agent Platforms	Large Language Models (LLMs), Enterprise RAG, Model Context Protocol (MCP), Multi-Agent Systems, Vector DBs (Milvus, Qdrant), GPU Serving (vLLM, TGI), Prompt Evaluation & Benchmarking, PyTorch
Cloud & Platform Systems	Linux Kernel/Admin, Docker, Kubernetes, Bare-Metal & GPU Virtualization, CI/CD Hardening, High-Availability Microservices, Software-Defined Networks, System Observability
Languages & Core	Python (Expert / Architecture), C / C++, Modern PHP, SQL (PostgreSQL, MySQL), Go, REST & gRPC APIs, Git/GitHub, Telemetry & Profiling
Strategy & Automation	Enterprise Process Discovery & As-Is Auditing, Workflow Automation, Practical AI Integration, Platform Product Management, DevEx Standards, SLA & Capacity Governance, Cross-Disciplinary Leadership

Esfandiar Kiani — Technical Portfolio & Credentials

es-kiani.dev • Esfandiar.Business@gmail.com

SELECTED ENGINEERING PROJECTS

Automated LLM Evaluation & Regression Benchmarking — **Model Reliability & LLMOps** 2025

Architected continuous evaluation harnesses for enterprise language models and multi-agent systems, automating synthetic regression testing, deterministic output validation, and latency-throughput profiling.

Speech Intelligence & Acoustic Quality Modeling (NISQA) — **Audio DSP & Speech AI** 2025

Engineered deep learning pipelines predicting multidimensional perceptual audio quality (MOS, coloration, loudness) across telecommunication and customer support voice channels.

Generative Latent Diffusion & Virtual Try-On (ViTON) — **Computer Vision & Generative AI** 2025

Engineered retail-grade generative vision pipelines using latent diffusion transformers (IDM-VTON / StableVITON), dense pose estimation, and semantic correspondence warping for photorealistic synthesis.

Smart Appliance Controller & IoT Hardware Telemetry — **Embedded Firmware (STM32 / ARM)** 2023 – 2024

Engineered real-time MCU firmware (C/C++) for high-reliability industrial hardware, multi-channel sensor telemetry, and bus protocols (SPI, I2C, UART, CAN) with fault-tolerant state-space estimation.

PUBLICATIONS & TECHNICAL REPORTS (RESEARCHGATE VERIFIED)

Comprehensive Drug Information System Project 2024 | Bachelor Thesis

Unified pharmaceutical relational data architecture and retrieval engine. DOI: 10.13140/RG.2.2.28140.37768

Behavioral & Structural VHDL Implementation of ROM 2023 | Technical Report

Design and hardware validation of ROM architectures using VHSIC methodology. DOI: 10.13140/RG.2.2.26447.46247

Applied Differential Equations in Computational Engineering 2022 | Research Paper

Mathematical modeling analysis with over 20,000 reads and recommendations on ResearchGate. DOI: 10.13140/RG.2.2.10507.82724/1

Descriptive Statistics Report of Bird Strikes in Aviation 2021 | Statistical Report

Empirical wildlife collision risk analysis on FAA and civil aviation incident databases. DOI: 10.13140/RG.2.2.17061.42726/1

CERTIFICATIONS

AI & Machine Learning (CS50) 2024

Harvard University • Verified Credential ↗

Search algorithms, probability, neural networks, reinforcement learning.

Python Programming Certification 2022

Technical & Vocational Org (Score: 90%)

Core OOP, data structures, algorithms, and practical software design.

Clean Code & Algorithm Design 2021

Top Learn Academy • Verified Course ↗

SOLID design principles, defensive refactoring, and code scalability.

Linux LPIC-1/102 & IoT 2021

Sharif University of Technology

System administration, process management, and microcontroller IoT.

HONORS & DISTINCTIONS

Finalist Team — Rahisho Innovation 2024

Score 91/100 • 16k+ Entrants • Commercialization

Selected finalist with AI multispectral & IoT orchard monitoring.

Ranked 80th Nationwide 2023

National M.Sc. Entrance Examination

Secured top 80 nationwide across all CE/CS graduates (no quota).

Finalist — Tech Olympics 2025

International Technology Olympics

National finalist for algorithmic problem solving and architecture.

Founding Member — AI Association 2025

K. N. Toosi University of Technology

Co-founded and structured official university AI initiatives.