

Hossein Shahabadi Farahani

 Email   +98-9188636616  Tehran, Iran |     

AI Systems Engineer · Solutions Architect

From published research to production systems

I look for good teammates to share [the agony and the ecstasy](#) of tackling new challenges and learning along the way.

Open to remote · Flexible availability · Full English proficiency with devs and clients

CORE CAPABILITIES

- **LLM/RAG systems:** Local retrieval, model adaptation, evaluation and API integration.
- **Speech & industrial ML:** Persian speech services, edge inference, anomaly detection and soft sensors.
- **Backend & MLOps:** FastAPI/Django services, async workers, Docker and CI/CD, with tested, deployable model serving.
- **AI architecture & consulting:** Readiness assessments, validation plans and delivery roadmaps.

SELECTED PROJECTS

2026 **Offline RAG Course Adaptation**

- Adapted five DeepLearning.AI assignments for offline use with Ollama and Chroma on a CPU-only laptop; added notebook regression tests and GitHub Actions CI.
- Proposed embedding-based routing to address weak zero-shot classification in a 1.5B model, doubling correct answers (+100%) and cutting median latency by 78% (161 s → 36 s).

2024 – 2025 **MindAssist: Persian Voice Assistant** [\[Case study\]](#)

- Built an end-to-end prototype: 1 GB Raspberry Pi 3 client, four cloud services and D-Bus integration; per-user intent heads under 40 KB over frozen ParsBERT, with LLM calls only at setup.
- Observed 96% command recognition and 2.85 s mean request-to-action-ID latency in small trials (3–4 speakers, about 10 utterances each; negligible network latency). No household rollout.

2026 **Kodoom: Persian Typed-Decision Model**

- Fine-tuned a 322M-parameter bilingual (Persian/English) decision model on Laya-multilingual for structured choice, scoring, and yes/no decisions, using English/Persian replay on two T4 GPUs.
- Built bilingual training data and calibration/evaluation tooling; preliminary Persian evaluations showed 13–40 percentage-point (25% average) gains over the base model across four synthetic workflow benchmarks. Apache-2.0 release;

SKILLS

Programming Tools: Python 3, SQL, pytest, Poetry, uv, Ruff.

Conceptual Foundations: TDD, typed APIs, design patterns (Adapter/Decorator/Factory), modular software design.

Backend & Deployment Tools: FastAPI, Django/DRF, Celery, Redis, MySQL, OpenAPI, Docker/Compose, GitHub Actions, Linux, Raspberry Pi, D-Bus.

Conceptual Foundations: CI/CD, asynchronous workers, streaming (SSE), idempotency, transaction-safe credit accounting, token authentication, encrypted storage, OpenAI-compatible gateways.

Applied work: Built Django/Celery services for OpenAI-compatible completions, streaming, encrypted storage and credit accounting.

Machine/Deep Learning Tools: PyTorch, TensorFlow, scikit-learn, Hugging Face Transformers, OpenAI API.

Conceptual Foundations: Transformers, CNN, LSTM, Wav2Vec, U-Net, BERT, GAN, Variational AE, Energy-Based Models, Grad-CAM,; fine-tuning, multi-GPU training, transfer learning, domain adaptation, probability calibration, train/dev/calibration/test isolation, bilingual evaluation, ML systems testing.

RAG & Data Engineering Tools: Ollama, Chroma, sentence-transformers, BM25, Hugging Face Datasets, PyArrow/Parquet, JSON Schema.

Conceptual Foundations: Hybrid retrieval, embedding routing, latency/answer-quality evaluation, notebook regression tests, versioned releases, provenance, Persian text normalization.

Industrial & Speech ML Tools: Raspberry Pi edge deployment, Wav2Vec, wake-word and voice activity detection.

Conceptual Foundations: Anomaly detection, root-cause analysis, soft sensors, Persian speech-to-text, edge/cloud inference.

Consulting & Product Tools: UML, Confluence, Jira, Product Requirements Documents (PRDs).

Conceptual Foundations: AI readiness assessment, validation plans, delivery roadmaps, product management, client discovery, technical presentations, cross-team communication, conflict resolution. English for client work.

WORK EXPERIENCES

2023 – Present **MAPNA Digital** — *AI strategy, system design & organizational advisory*

- Developed AI delivery roadmaps and readiness/validation assessments for wildlife monitoring, chest X-ray diagnosis, smart agriculture and smart-city services.
- **Collaboration 2.0:** Designed a PRD/Confluence workflow and phased 24-month strategy for a roughly 100-person R&D division; implementation was partial. [\[Case study\]](#)

2018 – Present **MECO, Digitalization Center & Condition Monitoring**

Data science (2018–2020); technical pre-sales (since 2021); ML/MLOps engineering (since 2022).

- Built anomaly-detection and root-cause-analysis tools across hundreds of power-plant sensor streams; developed domain-adaptive soft sensors for process variables.
- Engineered edge/cloud speech systems and supported technical pre-sales, translating client needs into solution presentations and serving as English-language liaison for international visitors.

[2018 – 2019] **[K.N.Toosi University of Technology]** — *Teaching & Laboratory Instruction*

- Instructed the *Industrial Process Control Laboratory*, guiding students through data acquisition, data-driven model identification, and controller design.
- Served as teaching assistant for *Fault Detection & Identification* and *System Identification*, supporting coursework based on data-driven modeling and analysis.

2015 – 2016 **MSA, Boiler Department** — *Thermal System Designer*. Simulated and designed a steam accumulation system for Takht-Jamshid Petrochemicals.

PUBLICATIONS

85+ citations

- | | |
|-------------|---|
| 2022 | “Transfer Learning in Condition Monitoring of Process Plants.” H. Shahabadi et al. <i>2nd BRIC Symposium, McMaster University</i> . |
| 2021 | “Domain Adversarial Neural Network Regression to Design Transferable Soft Sensor in a Power Plant.” H. Shahabadi et al. <i>Computers in Industry</i> . (Q1; IF 10+, 2021) [Paper] |
| 2020 | “On the Application of Domain Adversarial Neural Network to Fault Detection and Isolation in Power Plants.” H. Shahabadi et al. 19th IEEE ICMLA . |
| 2020 | “Between-Domain Instance Transition via Gibbs Sampling.” H. Shahabadi et al. <i>arXiv</i> . [Paper] |
| In progress | “Adaptation of Continuous Domains by Disentanglement of Domain Factors in Variational Auto Encoders.” H. Shahabadi, R. Ghorbandoost. <i>Manuscript in progress</i> . |

Served as peer reviewer: *Neurocomputing*, Elsevier. [\[Certificate\]](#)

EDUCATION

- | | |
|-------------|---|
| Fall 2022 | PhD, CS , University of New Hampshire, United States. fully funded, US\$60,000+ annual funding (declined) |
| 2016 – 2020 | MS, Mechatronics , K. N. Toosi University of Technology; APAC Research Group . One of Iran’s leading technical universities. |
| 2011 – 2015 | BS, Mechanical Eng. , University of Arak. |