

Artem Meshkov

Full-Stack & AI Engineer · Team Lead

artem.meshkov@gmail.com | linkedin.com/in/artem-meshkov | github.com/Artemnikov

PROFESSIONAL SUMMARY

Full-Stack and AI Engineer with 4+ years leading delivery of production web, real-time, and machine learning systems across healthcare and SaaS verticals. Experience spans medical real-time applications, healthcare RAG agents, document-processing ML pipelines, and multi-tenant cloud platforms — owning the full software development lifecycle from architecture through DevOps, performance optimization, and production support. Comfortable across frontend, backend, infrastructure, and ML, selecting the right layer for each problem and ramping up on new technology stacks within days.

TECHNICAL SKILLS

Languages: Python, JavaScript, TypeScript, SQL, Bash

Frontend / Mobile: React, React Native, Next.js, WebRTC, HTML, CSS

Backend: Django, FastAPI, Node.js, Express, REST APIs, WebSockets, Microservices Architecture

Databases: PostgreSQL, MySQL, MongoDB, DynamoDB, Redis

AI / ML: LangChain, LlamaIndex, PyTorch, RAG, Hugging Face, Langfuse, Ollama, LM Studio, vLLM, OCR, Document Processing, Data Lakes, Fine-tuning

Infrastructure / DevOps: Docker, Docker Swarm, Kubernetes (k8s/k3s), ArgoCD, CI/CD, Linux, S3, Observability, Security Hardening, RBAC, Network Policies

Cloud: AWS, GCP, Private Cloud, Cloud-Native Architecture

Version Control: Git, GitHub, Bitbucket, Gitea

Practices: Agile / Scrum, Code Review, QA Automation, System Architecture, Team Leadership

PROFESSIONAL EXPERIENCE

Full-Stack Developer, Team Lead & ML Engineer

2022 — Present

Weaversoft.io

- Designed and built a real-time medical platform using React, Django, and MySQL — supporting 50–500 concurrent clinical sessions with live video and interactive games; owned the full latency-sensitive WebRTC client/server pipeline end-to-end.
- Built a document-processing ML system handling diverse document types including complex PDFs, combining OCR, fine-tuned LLMs, and algorithmic extraction to emit structured JSON for downstream RAG retrieval and API validation; replaced significant manual review work, reducing processing time by an estimated 70–80%.
- Implemented QA automation pipelines for regression testing of document extraction across model updates and schema changes, keeping production pipelines stable and cutting manual QA cycles.
- Developed a healthcare conversational RAG agent (Femi-Premium) with dynamic retrieval, domain knowledge integration, and guardrails for clinical content; instrumented with Langfuse for full tracing and quality monitoring.
- Architected and operated a data lake strategy aggregating business data from disparate sources into a unified, LLM-optimized retrieval layer, enabling faster downstream reporting and model-ready access.

- Built a large-scale multi-tenant SaaS platform in Django orchestrating Docker microservices with per-tenant provisioning, isolation, and lifecycle management; deployed on k3s/k8s via ArgoCD across private cloud, AWS, and GCP targets with cloud-native security hardening and RBAC.
- Managed a team of 6 developers and 2 DevOps engineers across the full SDLC — architecture, TypeScript/Python code review, delivery, and production support.

EDUCATION

Practical Engineer, Electronics

ORT College, Israel

LANGUAGES

Hebrew — Native | English — Professional | Russian — Proficient | Czech — Beginner