

ROI ALFASSI

AI Engineer · GenAI · Machine Learning · Data Engineering

Haifa, Israel · roialfassi@gmail.com · +972-50-900-6794

[LinkedIn](#) · [GitHub](#) · [Portfolio](#)

PROFESSIONAL SUMMARY

Production AI/GenAI Engineer who builds and ships production-grade, AI-driven solutions in Python — tool-calling assistants, RAG pipelines over structured and unstructured data, and predictive ML. Owner of AI and data architecture at AIO Systems (platform spans **110,500+ IoT deployments**), working directly with company clients to define and deliver solutions. M.Sc. Information Systems (GPA 90, thesis 95); **three first-author publications** including IJHCI (Taylor & Francis). Sole recipient of the University of Haifa Research Excellence Award, 2025.

EXPERIENCE

AI Engineer · AIO Systems

Jul 2025 – Present · Israel

- Re-architected the production AI chat assistant on the **OpenAI Responses API** with **tool-calling, MCP, and RAG** over multi-source company data, aligning it with current GenAI best practices; it became a valuable, widely-used part of the product.
- Shipped **page-level AI assistants** across the CMS (110,500+ IoT deployments serving global telecom operators); building proactive, personalized features for retention.
- Own the company **Data Warehouse** and **pipelines for structured and unstructured data** end-to-end: expanded coverage **8 → 14+ core tables across 10+ data sources (~1B rows)**; re-architected Python ETL for reliability and scale.
- Built and deployed an **XGBoost** fuel-consumption forecasting model, **running in production** across remote African telecom sites, enabling proactive logistics planning.
- Designed and delivered **15+ adaptable analytics dashboards** for telecom maintenance teams, reducing downtime through early anomaly detection.
- Instrumented **LLM usage and cost tracking** with operational levers to monitor and improve consumption over time.
- Most senior technical contributor on the BI/AI team; sole owner of AI and data architecture, reporting to the R&D manager. **Lead technical discovery directly with company clients** to define AI requirements and ship solutions in their context.

Data Engineer, NLP Team · Webiks

2022 – 2023 · Israel

- Built data infrastructure from scratch on **AWS (EC2, S3)** for the **National NLP Program of Israel (NNLP-IL)** — annotation tools, ETL pipelines, and dataset-collection systems (APIs + scraping) feeding downstream NLP models.
- Developed **Python + Vue.js** annotation web applications used by program researchers; transformed multi-format raw data into production-grade NLP datasets.

Automation Developer · Rafael Advanced Defense Systems

2020 – 2022 · Israel

- Designed **automated testing frameworks** for flight-simulation software; authored **hundreds of test scenarios** preventing regressions across mission-critical defense systems.
- Built Python (Pandas) tooling for data validation, statistical analysis, and automated result reporting.

Teaching Assistant & Research Assistant · University of Haifa

2020 – 2022, 2024 – 2025

- Taught **C Programming, Data Structures & Algorithms**, and **Ethical System Design**.
- Research assistant on Human-AI decision-making under stress.

EDUCATION

Specialization: Artificial Intelligence, Data Science, Information Visualization.

Thesis: *Fanfiction in the Age of AI* — built a **Latent Personal Analysis (LPA)** model to detect AI-generated text and quantify stylistic differences across LLMs and human authors; combined embedding analysis, clustering, and classification.

Advisors: Prof. Osnat Mokryn (U. Haifa) and Prof. Orit Shaer (Wellesley College).

◆ **Sole recipient, University of Haifa Research Excellence Award, 2025.**

SELECTED PUBLICATIONS (FIRST AUTHOR)

- **Alfassi, R.**, Cooper, A., Mitchell, Z., Calabro, M., Shaer, O., & Mokryn, O. (2025). *Fanfiction in the Age of AI: Community Perspectives on Creativity, Authenticity and Adoption*. *International Journal of Human-Computer Interaction* (Taylor & Francis), 42(5), 3062–3094. **8 citations**.
- **Alfassi, R.** et al. (2025). *Online Storytelling Spaces: Exploring Participants' Perceptions of Overt and Covert AI Agents*. HAI-Gen Workshop, ACM IUI 2025, Cagliari, Italy.
- **Alfassi, R.** et al. (2025). *Identifying differences between human- and AI-generated fanfiction stories*. NetsciX 2025, India.

TECHNICAL SKILLS

GenAI / LLM OpenAI API (Responses, Chat, Embeddings, Function Calling, Structured Output), Anthropic Claude, Azure OpenAI; RAG end-to-end; Agents, Tool Use, MCP; Vector DBs (Pinecone, Weaviate, pgvector, ChromaDB, FAISS, Azure AI Search); Embeddings (text-embedding-3, sentence-transformers, BGE); Prompt Engineering; LLM Evaluation (LLM-as-judge, custom evals); Guardrails & hallucination mitigation; Local Inference (Llama, Mistral, Qwen via vLLM, Ollama).

Machine Learning PyTorch, TensorFlow, Hugging Face Transformers, scikit-learn, XGBoost, LightGBM, Pandas, NumPy, SciPy; A/B testing, feature engineering, model evaluation, cross-validation.

Data & Cloud SQL (T-SQL, PostgreSQL, MySQL), SQL Server, Python ETL/ELT, Data Warehouse design, Azure Synapse (PySpark), BigQuery, Databricks; Azure (OpenAI, Functions, Storage, ADF), AWS (EC2, S3), GCP, Docker, Kubernetes, CI/CD (GitHub Actions), Terraform, Bicep.

Programming Python (FastAPI, Flask), C# / .NET / ASP.NET, JavaScript, C, C++, Java, Go.

SELECTED PROJECTS

- **Agentic Workflow Tooling** · Self-built multi-agent and knowledge-base tooling (raw, no framework) to automate engineering workflows — tool use, retrieval, and task delegation patterns.
- **PromptQueueRunner** · Python tool to batch and queue large LLM prompt jobs for unattended execution. [repo](#)
- **Recipe Chat Assistant** · Desktop chat app combining OpenAI API with local LLMs for structured recipe retrieval. [repo](#)
- **Stock Comparator** · Deployed Streamlit app comparing equities and ETFs over time. [stock-comparator.streamlit.app](#)

ADDITIONAL

Military Service IDF, Air Defense (Israeli Air Force) — RF Technician, 2013–2016. Excellence in Level-11 Technical Course.

Languages Hebrew (native), English (fluent), French (basic).

Security Clearance Held; details available upon request.