

Ziad Alatrash

+31 6 38968768 | ziad.atrash@gmail.com | Haarlem, Netherlands | [LinkedIn](#) | [GitHub](#)

SUMMARY

AI/LLM systems engineer and fourth year IT student who has been building software since fifteen. I design grounded retrieval-augmented generation systems and ship them to production: during a six month internship at Impala Studios I built and deployed a code-aware RAG engine now used company wide, and alongside my degree I have shipped a university-adopted computer vision platform and a 65,000 line Android app now in closed beta. The throughline across all of it is the engineering beneath the AI: clean, hexagonal architectures with tests that fail the build on a broken boundary, and evidence-gated pipelines that keep a model from fabricating.

EXPERIENCE

Backend Intern

Sep 2025 to Feb 2026

Impala Studios, Haarlem, Netherlands

- Designed and built Omnipotence (internal name CodeLens), a code-aware RAG engine that answers natural language questions about a large codebase with source cited answers, using syntax aware chunking, hybrid retrieval and cross-encoder reranking.
- Enforced a hexagonal, ports and adapters architecture with 14 abstract ports and 28 automated test modules, including architecture boundary tests that fail the build if the design is broken.
- Deployed the system on AWS end to end (CloudFormation, ECS, Lambda, API Gateway, OpenSearch, S3, Bedrock), and the codebase was opened up company wide for other engineers to build on.
- Earned the AWS Certified AI Practitioner certification during the internship and was graded 9 out of 10.

PROJECTS

TulipVision, Tulip Sprout Detection Platform

Mar 2026 to Jun 2026

Applied research group project, Inholland, team of 3

- Compared four object detection architectures (Faster R-CNN, YOLOv11, YOLOv5, RetinaNet) on a shared dataset with a consistent evaluation protocol (mAP, precision, recall, F1).
- Owned the YOLOv11 track and built an adaptive tiled inference pipeline for oversized field images, improving small object detection without retraining.
- Built the full deployment platform (FastAPI backend, Vue 3 frontend, Docker, Azure Container Apps), now used directly by the research group.

Recomp Tracker, AI Coach

2026

Solo Android app, closed beta

- Designed, built and shipped a roughly 65,000 line Android app solo, now in closed beta, unifying nutrition, training and recovery tracking into one adjustment engine.
- Built a conversational AI coach with streaming responses and 19 function calling tools, grounded with retrieval augmented generation, with every number it reports coming from deterministic app logic rather than the model.
- Wrote about 1,300 tests, mostly test first, with a pure Kotlin domain layer carrying zero Android dependencies for fast, reliable unit testing.

Consented Cart, Shopify Embedded App

2026

Solo full stack build

- Built a production Shopify embedded app (about 12,300 lines TypeScript, React Router v7, Prisma) that recovers abandoned carts under GDPR compliant opt-in consent, enforced at the database schema level rather than by UI convention.
- Engineered a revenue attribution engine matching Shopify orders and webhooks back to originating capture events within a configurable window, with fully tested matching logic.
- Designed a 16 entity Prisma schema across 12 migrations with a zero downtime idempotent backfill, hashed PII at rest, and idempotency keyed email dispatch to prevent double sends.

Lex-AI, RAG Chatbot for EU AI Policy

2026

Personal project

- Built a RAG chatbot answering questions about EU AI policy, grounded in 87 official documents, with a vector search engine and a Maximal Marginal Relevance reranker written from scratch in NumPy.
- Added an empirically tuned confidence threshold that stops the chatbot from answering when retrieval confidence is low, validated against an 8 question benchmark and automated edge case tests.

SKILLS

Core: Generative AI / RAG · Retrieval-Augmented Generation · Prompt Engineering · Machine Learning · Python

Tools: Kotlin · TypeScript · Swift · Java · FastAPI · Spring Boot · React · Vue 3 · PyTorch · AWS · Azure · Docker

Methods: Hexagonal and clean architecture · TDD, unit/integration/E2E testing · Model evaluation and benchmarking · Responsible, grounded AI design

Domains: Applied AI and machine learning · Full stack development · Cloud deployment

EDUCATION

Hogeschool Inholland

Expected 2027

*BSc Information Technology (Year 4), Minor: Data and AI, Inholland Haarlem
(completed Jul 2026)*

Haarlem, Netherlands

CERTIFICATIONS

- AWS Certified AI Practitioner, Amazon Web Services