# Hany Sayed Ahmed

Staff Software Engineer | AI/ML & Vector Search Architect | Professional Scrum Master

+201116811410 | Gmail | LinkedIn | GitHub | Portfolio

---

## PROFESSIONAL SUMMARY

Innovative **Staff Software Engineer** and **Certified Scrum Master (PSM II)** with **10+ years of experience** designing, developing, and delivering **scalable, high-performance applications** across backend, frontend, and cloud-native architectures.

Specializes in **Ruby on Rails, .NET Core, Angular, React, Python**, and **AI/ML systems**, including **vector search, embeddings, and LLM workflows**. Skilled in **enterprise-scale recommendation engines, cloud infrastructure, and multi-platform AI pipelines**.

**Key Achievements:**

- **AI/ML Leadership:** Architected production-grade recommendation engines using OpenAI embeddings, achieving **sub-100ms search latency** and reducing API calls by ~80%.
- **Ollama Migration:** Initiated migration to **Ollama LLM**, targeting **<1s inference latency** and **50% cost reduction**, improving reliability and reducing external dependencies.
- **VIBE Coding Expertise:** Built modular, cloud-native AI workflows deployed across platforms like **Cursor Cloud**, enabling high-scale recommendation and content pipelines.
- **Performance & Scalability:** Optimized PostgreSQL queries, caching, and background jobs, reducing total query time by **40–50%** under heavy load.
- **Enterprise Delivery:** Led cross-functional teams, ensured high code quality, mentored engineers, and translated product roadmaps into reliable technical execution.
- **Prompt Engineering:** Designed advanced, structured prompt workflows for **bilingual content generation, sentiment analysis, subcategory classification, and location intent detection**.
- **Cost & Resource Optimization:** Applied token calculation strategies and intelligent caching to reduce operational costs by **30–50%**.

## TECHNICAL SKILLS

- **AI/ML & Vector Search:**
    - **OpenAI API** – text-embedding-3-large, GPT-4o-mini for embeddings, semantic search, and chat completions
    - **Ollama LLM** – open-source model deployment, inference optimization, and strategic selection for cost & latency efficiency
    - **Token calculation & prompt cost management** for efficient large-scale AI workflows
    - **Prompt engineering & structured output design** – multilingual content (English/Arabic), review sentiment extraction, subcategory classification, and location intent detection
    - **Vector similarity & ranking algorithms** – cosine similarity, multi-source weighted aggregation, Haversine distance for geo-aware recommendations
    - **Embedding caching strategies** – multi-tier TTL caching reducing API calls and latency
    - **VIBE coding with AI –** leveraging Cursor Cloud and cloud platforms to build, deploy, and scale AI workflows for recommendation engines and content analytics.
- **Languages**: **Ruby**, Python, Dart, C#, C++, Java, Shell Script, **JavaScript**, **TypeScript**, PHP, Perl
- **Frameworks**: **Ruby on Rails**, **Angular**, React.js, Vue.js, Node.js, Django, fastapi, .NET Core, Zend, Laravel, Express, Spring Boot
- **Mobile Development**: React Native, Ionic, Flutter
- **Database Management**: PostgreSQL, MySQL, MongoDB, SQL Server
- **Version Control & SDLC**: Git, SVN, JIRA, Trello, Agile Methodologies
- **Problem Solving:** Successfully solved over 800 coding challenges on platforms including Codewars, Hackerrank, Leetcode, Github
- **Cloud Technologies:** AWS (ECS, EKS, RDS, Lambda, S3), GCP, Heroku
- **Development Practices:** Test-Driven Development (TDD) with RSpec/FactoryBot, CI/CD, Microservices, API Design (REST, GraphQL, gRPC)

## EDUCATION

### 16-Month Diploma – ALX Africa, Data Science, 2023-2024

### 9-Month Diploma – Open Source Application Development, ITI, 2015-2016

### BSc in Statistics & Computer Science – Ain Shams University, 2007-2013

## PROFESSIONAL EXPERIENCE

**Technical Team Lead (Remotely).**

**Escape Ventures, Qatar** - ***Dec 2024–Present***

- **Architected a production-grade AI recommendation engine** using OpenAI embeddings and PostgreSQL pgvector, powering 7+ content types (products, videos, images, community posts, livestreams, influencers, brands).
- **Designed a hybrid search engine** combining vector similarity and SQL queries, achieving **sub-100ms response times** for complex multi-type searches.
- **Reduced API costs by ~80%** via multi-tier caching (embedding cache, search result cache, brand analysis cache) with TTLs and versioned invalidation.
  **Implemented weighted semantic ranking algorithms** integrating business metrics (ratings, reviews) with vector similarity for high-relevance recommendations.
- **Developed location-aware search** with Haversine geo-distance calculations for proximity filtering (<10km) combined with semantic scoring.
- **Engineered asynchronous AI pipelines** (Sidekiq) for embedding generation, content creation, and analytics, with robust error-handling and logging.
- **Designed advanced prompt engineering workflows** for GPT-4o-mini enabling bilingual content generation, review sentiment analysis, subcategory classification, and structured JSON outputs.
  **Optimized PostgreSQL performance** through selective loading, eager loading, vector indexing, parallel queries, CTEs, and UNION ALL, reducing query time by 40–50%.
- **Initiated migration to Ollama LLM**, targeting **<1s inference latency** and **50% cost reduction**, improving reliability and reducing dependency on external APIs.
- **Implemented VIBE coding AI workflows across platforms** (including Cursor Cloud), enabling **modular, cloud-native, and scalable AI pipelines** for recommendation engines and content analytics.
- **Applied token calculation and cost optimization strategies** for LLM prompts, ensuring **efficient AI usage and reduced operational expenses** across multiple workflows and languages.

**Senior Software Engineer (Remotely).**

**Andela–Kinship (Contract), US -** ***April 2022–Dec 2024***

- Optimized PostgreSQL & Rails applications, improving query performance by **40%**.
- Developed scalable infrastructure with Angular Dart and React Native, enhancing UX and responsiveness.
- Automated deployment pipelines, minimizing downtime and accelerating delivery.
- Designed modular architectures, reducing future feature development time by **30%**.

**Senior Software Engineer (Remotely).**

**Andela–Litmus (Contract), UK -** ***March 2021–April 2022***

- Enhanced Vue.js, Ember.js, and Rails application performance, reducing load times by **25%**.
- Developed automation tools and custom scripts, improving team efficiency by **30%**.
- Debugged critical deployment issues, reducing downtime by **20%**.

**Senior Software Engineer.**

**IdearRating, Egypt -** ***March 2018–March 2021***

- Developed Node.js, Angular, and .NET Core features for **500k+ active users**, optimizing database queries and scaling performance.
- Implemented data-driven architecture using MongoDB, MySQL, and SQL Server, reducing retrieval costs by **20%**.

**Team Leader (Part-Time).**

**VOOOM, Egypt -** ***March 2019–Oct 2020***

- Led feature development with **40% fewer bugs**, implemented best practices in code reviews.
- Mentored junior developers, improving task completion by **35%**.

**Freelance Instructor**

**ITshare & Information Technology Institute, Egypt - *Jun 2017 – Present***

- Developed and delivered advanced problem-solving exercises in algorithms and system design, achieving **95% satisfaction scores** from students.
- Mentored students in real-world scenarios, leading to a **50% improvement** in their ability to solve complex challenges.

**Senior Full-Stack Web Developer.**

**Nabda Care, Egypt - *Jun 2016–March 2018***

- Built healthcare platform (Rails + Ember.js), increasing patient data processing efficiency by **45%**.
- Optimized database queries, reducing report generation time by **50%**.

**PHP Developer.**

**IT Solution, Egypt - *Jul 2013 – Sept 2015***

- Built **Zend Framework**, **WordPress**, and **Drupal** applications, delivering projects ahead of deadlines by an average of **10%**.
- Designed intuitive UI/UX for web applications, boosting user engagement by **15%**

## CERTIFICATIONS

- Professional Scrum Master, PSM II. – ***Scrum.org***
- AWS Billing and Cost Management. – ***AWS***
- Mastering System Design, React Native, and React Redux Advanced. – ***Udemy***
- Data Science. – ***ALX***
- Java Programming & Problem Solving, Front-End Web UI Frameworks, Algorithms Design and Analysis, HTML, CSS, and JavaScript, Rails with ActiveRecord, and Ruby metaprogramming. – ***Coursera***