Agustín Pérez

AI Engineer | Software Engineer

📍 Tandil, Buenos Aires, Argentina 📞 +54 9 2494 608075 ✉️ agustinpereh03@gmail.com

🔗 https://www.linkedin.com/in/agusperezs - https://github.com/agusperez03

Professional Profile

AI Engineer with strong Software Engineering foundations and hands-on experience designing and deploying scalable, production-grade AI systems in the e-commerce domain. Specialized in semantic search, agent orchestration, and large-scale embedding pipelines using modern LLM frameworks and vector databases. Delivered measurable business impact, including an 18% uplift in search-driven conversion across a 600K+ product catalog. I focus on building reliable, high-performance AI systems that enhance user experience, ensure operational robustness, and drive real business outcomes.

Professional Experience

Ssr Python Engineer | MAX CONTINENTAL *(March 2026 - Present)*

- Developing robust, strongly-typed APIs to power an AI-driven search and matching system for medical product catalogs using the OpenAI API.

AI Engineer | Region Global eCommerce Agency *(May 2025 - February 2026)*

- Designed and deployed a Semantic Search Engine using OpenAI Embeddings and Elasticsearch Vector Database for catalogs with 600k+ products, reducing no-result queries by 35% and improving search-driven conversion by 18% within the first quarter.
- Implemented caching and indexing optimizations that cut p95 search latency by over 40%.
- Using Redis, I built a parallel embedding generation pipeline for large-scale catalogs, significantly reducing API costs and enabling efficient high-volume embedding generation.
- Developed an automated product description generator using prompt engineering, improving semantic indexing quality and reducing manual content creation efforts by 70%.

Backend Developer | Tandil Public Health System (SISP) *(Nov 2024 - April 2025)*

- Reduced manual effort by nearly 90% and enabled timely processing of medical records citywide by automating submissions through an end-to-end RESTful API with Java Spring Boot and PostgreSQL.
- Ensured secure and compliant handling of sensitive data by implementing JWT authentication and comprehensive Postman testing.

SQL Freelance Developer | Fiverr *(Nov 2023 - Aug 2024)*

- Delivered +15 custom database solutions, earning consistent 5-star ratings by providing high-quality code and proactive communication.

Education

- **Software Engineering Degree | National University of Central Buenos Aires**
  - *Completed a 5-year degree in 4.5 years, achieving an average grade of 8.05.*

Certifications

- **AWS Cloud Practitioner (CLF-C02) Certification | DataCamp**
- **Associate AI Engineer for Developers Certification | DataCamp**
- **Machine Learning with Python Certification | FreeCodeCamp**

Featured Projects

Ambient Email Agent *(Nov 2025)*

- Practical implementation of an agent from LangChain Academy, configuring stateful workflows, persistent memory, and human-in-the-loop mechanisms.
- Full deployment on LangGraph Platform.
- Evaluation and traceability with LangSmith, applying LLM-as-a-judge and automated tests with pytest.

- Use of BaseModel to structure prompts, tools, and agent logic, incorporating memory-updating techniques.

"VeriBot": Microsoft AI Agents Hackathon Participant *(April 2025)*

- Participated in Microsoft's month-long AI Agents Hackathon, developing an autonomous agent capable of searching, validating, and summarizing news from multiple online sources.
- Implemented a Retrieval-Augmented Generation (RAG) workflow to retrieve verified information from indexed documents and web queries, enabling accurate, context-aware fact-checking.
- Tech Stack: Java, LangChain4J, Azure Containers, SerpAPI, ChromaDB

Thesis: "Machine Learning for the Analysis of Collaborative Strategies in Autonomous Agents" *(Nov 2024 - July 2025)*

- Developed a multi-agent simulation environment in Python using Unity ML-Agents framework for Robocup inspired soccer scenarios
- Implemented and fine-tuned Reinforcement Learning algorithms (PPO, MA-POCA) improving precision and team coordination by 35%
- Evaluated agent performance using TensorBoard metrics
- Tech Stack: Python, Unity ML-Agents, TensorFlow, RL, TensorBoard

---

Technical Skills

- **AI & Machine Learning**
  OpenAI API, Large Language Models (LLMs), Retrieval Augmented Generation (RAG), Model Context Protocol (MCP), TensorFlow, Keras, Reinforcement Learning, Scikit-learn, NumPy, Pandas, Prompt Engineering
- **Programming Languages**
  Python, Java, JavaScript, PHP
- **Frameworks & Tools**
  FastAPI, LangChain, LangGraph, Spring Boot, RASA, Redis, n8n
- **Cloud & MLOps**
  Docker, AWS, CI/CD
- **Databases**
  PostgreSQL, MySQL, Elasticsearch Vector Database, ChromaDB
- **Frontend Technologies**
  React, HTML, CSS

Languages

- **English –** Professional working proficiency (B2)
- **Spanish –** Native