

Marwan Zaarab *Software Engineer*

✉ marwan.zaarab@gmail.com 📍 Chicago, USA 🌐 marwan-zaarab 🔄 marwan37

📁 Experience

AI Engineer, ZenML (contract) Mar 2025 – present | Remote

- Built an automated classification system for LLMs case studies 🔗 , fine-tuning ModernBERT to outperform commercial LLMs (e.g., Claude Haiku), with 30% higher accuracy, 69x faster inference, and 225x cost-efficiency.
- Developed a production-grade OCR pipeline 🔗 supporting batch processing, smart caching, and artifact versioning for high-volume unstructured document workflows.
- Implemented an end-to-end credit-scoring pipeline 🔗 that auto-generates EU AI Act Annex IV evidence (risk reports, audit trails, Software Bill of Materials, real-time compliance dashboards).

UI Engineer, CardFree May 2024 – Jan 2025 | Remote

- Developed and maintained Vue.js and Ruby on Rails admin portal features, enabling merchants to manage digital ordering, loyalty programs, real-time order tracking, and promotional campaigns across multiple locations.
- Optimized merchant configuration interfaces through tabbed navigation and batched API requests—reducing API calls by 92% and significantly improving load times.
- Built an end-to-end global search system with fuzzy matching, weighted relevance scoring, full-text indexing, and keyword-based retrieval—enhancing users' ability to quickly find and manage platform configurations.
- Strengthened role-based access control (RBAC) with CASL and CanCanCan, ensuring secure and synchronized authorization across the platform.

Open Source Contributor, ZenML Feb 2024 – Apr 2024 | Remote

- Developed a VSCode extension 🔗 (1500+ downloads) with a custom Python-based LSP server, integrating ZenML's library into a TypeScript codebase to streamline MLOps workflows.
- Eliminated FastAPI server dependency by enabling direct library integration via users' local installations—simplifying version control and reducing setup overhead.
- Implemented real-time config monitoring using watchdog and an event-driven architecture for instant ZenML environment synchronization with the VSCode UI.
- Built a scalable AWS testing framework using LocalStack and Terraform, automating CI/CD pipelines with Docker.

Co-Creator & Software Engineer, Pennant May 2023 – Jan 2024 | Remote

- Built Pennant 🔗 , a high-performance collaborative computational notebook, integrating a conflict-free replicated data type (CRDT) framework to enable real-time document synchronization with sub-50ms latency.
- Developed a distributed async execution engine using Flask, Celery, Redis, and RabbitMQ, optimizing request handling with background workers to achieve sub-100ms response times.
- Designed a React-based UI with TypeScript, React-Query, and Recoil, implementing browser and API caching that cut load times by 87.5%.
- Optimized infrastructure by reducing client-side TCP connections from linear to constant scale, cutting server resource consumption by 90% and improving system scalability.

🧠 Skills

Languages

Python, JavaScript/TypeScript, Ruby, Go, SQL, HTML/CSS, HCL, Shell

AI/MLOps & Cloud

AWS (SageMaker, Lambda, S3, IAM, ECS, CloudWatch), GCP (Vertex AI), Docker, ZenML, Modal, MLflow, GitHub Actions, Terraform, New Relic

Frameworks & Libraries

Vue.js, Vuex ORM, React (Hooks), Node.js, Next.js, Ruby on Rails, Tailwind CSS, scikit-learn, PyTorch, Pandas, NumPy, FastAPI, Transformers (Hugging Face)

Tools & Technologies

Git/GitHub, OpenAI, Anthropic, Mistral AI, Ollama, Jupyter, Streamlit, Postgres, Redis, MongoDB, Nginx

🎓 Education

M.D., Windsor University School of Medicine

2015 – 2019