

Rody Vilchez

Applied ML Engineer — Retrieval · Agent Systems · Evaluation · Urban ML

rody@rosewt.dev | +51 987 082 126 | rosewt.dev | [GitHub](#) | [LinkedIn](#)

Applied ML engineer working at the intersection of agent systems, retrieval, and evaluation under imperfect data. At CIP–CGIAR, I build production multi-agent platforms, multilingual GraphRAG, and evaluation infrastructure; independently, I study measurement failure in urban ML and co-authored a multimodal architecture published in Springer CCIS.

Experience

AI Intern

Lima, Peru

International Potato Center (CIP–CGIAR)

Oct 2025 – Present

- **Microsoft Teams multi-agent platform** — Architected and shipped an institutional system on Copilot Studio for IT support and corporate services, with cross-domain delegation and escalation to tickets prefilled from conversational context and subject to human review.
- **Agent-based evaluation suite** — Built a suite where evaluator agents execute scenarios derived from operational records; it combines deterministic regression, provider-agnostic LLM-as-judge scoring, and persona-driven simulation to attribute routing, delegation, and response-quality failures.
- **Multilingual GraphRAG** — Built end-to-end the document-intelligence layer powering a five-language agricultural GraphRAG over corpora with noisy OCR and irregular layouts: parsing, chunking, structured metadata extraction, embeddings, and vector storage.

QA Trainee

Lima, Peru

Visma LATAM

Dec 2024 – Oct 2025

- **LLM-driven test synthesis** — Built an agent that transforms functional specifications into executable E2E scenarios.
- **DOM-aware automation** — Engineered selector and state extraction, integrating generated Cypress suites into Jenkins as regression coverage for critical workflows.

Selected research and systems

Infelix — Latent Urban Crime Risk under Reporting Bias

BSc thesis research · UPC · 2025–2026

- **Modeling unreported crime** — Built a Bayesian model over victimization surveys and an 11-source H3 geospatial panel (Sentinel-2, VIIRS, OSM, WorldPop, and Street View) to estimate crime risk across 43 Lima–Callao districts; census data were the only source that consistently improved predictions within districts.
- **Cross-city transfer** — Features learned in Lima matched models with one year of local data in Arequipa and remained competitive in Cusco and Piura; stress tests showed that record quality, not only architecture, determines when a system can be evaluated and transferred.

Multimodal Sign Language Model

Springer CCIS 2895, 2026

- Co-authored a published gloss-free multimodal architecture using latent queries and cross-attention to project 2D keypoints into a pretrained language model's embedding space, evaluated across Peruvian and Argentinian corpora. DOI: [10.1007/978-3-032-20322-9_23](https://doi.org/10.1007/978-3-032-20322-9_23)

Wachi — Risk-Aware Pedestrian Routing

El Comercio Lab

- Built a geospatial system that re-ranks pedestrian routes by spatiotemporal risk in Lima using H3 surfaces, temporal decay, and six time bands for an interactive data-journalism experience.

Open-source contributions

- **Microsoft Copilot Studio** — Diagnosed knowledge-source synchronization failures in child agents and validated prerelease fixes in the official VS Code extension.
- **Gemini CLI** — Corrected a UI schema and added regression coverage to Google's open-source agentic CLI.
- **scikit-learn** — Improved HDBSCAN technical documentation.
- **beads (bd)** — Implemented prefix-routed write-through and the regression test that unlocked consolidation of the upstream implementation.

Education and recognition

B.Sc. Computer Science

Universidad Peruana de Ciencias Aplicadas (UPC)

Expected late 2026

2nd place, DataFest BCP × ESAN 2025 · SALA 2026 full scholarship

Technical focus

Retrieval and agents: LlamaIndex, Qdrant, Copilot Studio, Power Automate, multi-agent orchestration

Evaluation and research: Python, PyTorch, scikit-learn, LLM-as-judge, experimental design, robustness and uncertainty

Data and backend: FastAPI, H3, DuckDB, PostgreSQL, ETL, REST APIs, Docker, GitHub Actions, Linux

Languages: Spanish (native) · English (intermediate)