

GUSSEPPE BRAVO

AI Research Engineer | PhD in AI

 gussepebravo.com

 gussepebravo

 gussepe

"If I can't draw it, I can't get it"

SUMMARY

- **Generative AI Scientist with PhD in AI** specializing in AI systems that adapt to real-world change. Built first-of-its-kind **AI agents for production model monitoring and improvement** ([CAMA] at AAMAS, [KC-Agent] at IEEE COMPSAC)
- Currently building **healthcare cognitive architectures** at **Cotiviti**, [Cotiviti press]. Also contributed to diabetic retinopathy research at Emory Medical Center. **7+ years research work** at **Barcelona Supercomputing Center/Lenovo** developing continual learning frameworks for drift datasets. My work appears in [Lenovo press].
- Published in **top-tier AI venues** (AAMAS, ICPR, VISAPP). PhD research achieved **99% storage reduction** while maintaining **96.9% accuracy** in privacy-sensitive healthcare applications through autonomous prompt evolution [Springer Nature].
- **Some facts:** I prototype fast. I measure everything (costs, latency, performance). I follow the old-school unix principles: do one thing well.

EXPERIENCE

Generative AI Scientist

Cotiviti

 April 2026 – Present

 New York, USA

- Concept extraction for Summary plan description documents using connected components (graph), RAG and anomaly detection. **Some numbers: really cheap pipeline, only 10\$ for 1k docs end to end with \$3M–\$5M annual revenue opportunity.**
- Building a cognitive architecture that enables agents to learn fast using semantic, episodic and procedural memories. **Part of this work is being used in some internal projects already.**

Generative AI Research Engineer (Intern)

Cotiviti

 Jul 2024 – Present

 New York, USA

- Built **3D RAG system (semantic, lexical and graph)**. I built a cost-efficient pipeline for parsing and retrieving complex documents (1.5k docs/0.5\$). All open source and fast execution.
- Built an evaluation workflow for benchmarking RAG architectures. I used the MMLU methodology (multi-choice options).

Recognised Researcher

Barcelona Supercomputing Center & Lenovo

 Feb 2026 – Present

 Spain/USA, part-time

- Research on cognitive architectures for agents. Just published [KC-Agent] at COMPSAC 2026.

STACK SOFTWARE

LinuxPythonPyTorchNumba

LangGraphAutoGenLangchain

transformersMLflowRay

ChromaDBKubernetesNode.js

LangGraphSparkLLMs

IMPACT METRICS

-  **\$250K+ Research Funding**
Secured from Lenovo and Intel for AI research initiatives
-  **\$10M+ Revenue Impact**
Generated for Peruvian Tax Administration through AI fraud detection
-  **40+ GitHub repos**
For open-source ML/MLOps projects and libraries
-  **15+ Publications**
In top-tier AI journals and conferences (AAMAS, ICPR, CVPR)

EDUCATION

PhD in Artificial Intelligence

UPC BarcelonaTech

 2021 – 2025

Research focus: Human-on-the-loop Continual Learning: Data, Knowledge and Agents for Model Adaptation. [See my thesis here].
Highest honors: Cum laude

MSc in Artificial Intelligence

UPC BarcelonaTech

 2019 – 2021

Thesis: "Scanflow: A Learning-symbolic Framework for ML Workflow Debugging" **Published in Expert Systems With Applications** (IF: 8.665)

BSc in Computer Science, Magna Cum Laude

National University of Engineering

 2011 – 2016

Top 5% of graduating class

AI Research Engineer

Barcelona Supercomputing Center & Lenovo

📅 Feb 2019 – January 2026 📍 Spain/USA

- Architected cognitive frameworks for Language Agents that **autonomously monitor ML model health**, reducing degradation incidents and improving model reliability
- Built TADIL algorithm for domain-agnostic continual learning, enabling **6% performance improvement** across diverse domains without requiring task labels
- Developed Scanflow, an MLOps platform for debugging ML workflows, **some projects were built on top of it**.
- Developed a novel drift detection algorithm using autoencoder for drift detection on images. **Published in top-tier journal**.
- **Led cross-functional team**, resulting in several publications in top-tier journals and conferences

AI Engineer Intern

Emory Global Diabetes Research Center

📅 July 2023 – Jul 2024 📍 USA

- Developed resource-efficient **cognitive architecture** for LLM-agents using finetuned small language models (Llama, Phi), **reducing API costs by 35%** while maintaining reasoning capabilities
- Built privacy-preserving **domain adaptation algorithm** for diabetic retinopathy detection, achieving **96.9% accuracy with 99% storage reduction** across diverse patient populations
- Published 3 papers on continual learning for medical imaging, including **Best Paper Nominee at VISAPP 2025**

AI Consultant

SUNAT-Inter-American Development Bank

📅 Jan 2022 – Dec 2025 📍 Peru

- Built the **first agent-based system** for customs to monitor ML models in production, enabling **real-time detection of customs fraud patterns**
- Built the first **"Improver" agent for automated ML model updates** that **reduced manual intervention by 65%** while maintaining model performance

Data Scientist

SUNAT (Tax Administration)

📅 Nov 2017 – Nov 2018 📍 Peru

- Engineered ensemble-based fraud classifier algorithm with **92% accuracy** for identifying unjustified patrimony
- Developed graph analysis methods to uncover complex tax evasion networks, **processing 1TB+ of financial data** monthly
- Built scalable analytical infrastructure processing millions of electronic invoices, directly contributing to **\$10M increase** in tax revenue

AWARDS



PhD highest honors (2025)

Cum laude for work on Continual Learning and Agents



Intel-Lenovo AI Innovators Award (2023)

Selected for innovation in continual learning and cognitive architectures research



María Rostworowski I Researcher (2020)

National recognition for excellence in AI research

SOME PUBLICATIONS

Adaptive Prompt Evolution

Springer Nature 2026

Feature Engineering for Agents

AAMAS 2025

Task-Agnostic Domain Incremental Learning

ICPR 2024

Scanflow: Multi-graph framework

Expert Systems With Applications

KEY PROJECTS

Cognitive Architecture for ML Monitoring

[\[Link CAMA\]](#)

- LLM-based agent framework that **autonomously monitors ML models** in production environments
- Developed novel **decision procedure algorithm** and **structured memories** for interpretable model diagnosis
- Featured in **AAMAS 2025 conference**; developed in collaboration with Lenovo AI Lab

Scanflow

[\[Link Scanflow\]](#)

- **Production-grade MLOps** framework for ML workflow management, debugging, and reliability
- Many projects were inspired by this framework, like Scanflow-k8s, CAMA, KC-Agent.
- Published in top-tier journal "Expert Systems With Applications" (Impact Factor: 8.665)