Francesco Andrea Causio

Machine Learning Engineer — model evaluation & benchmarking, dataset construction, applied ML

Philadelphia, PA • francescoandreacausio@gmail.com • (215) 910-7004

linkedin.com/in/causio • github.com/causius0 • causio.xyz

Skills

Languages: Python, TypeScript, Java, C++ • ML: LLMs, RAG, Agentic Workflows, Model Evaluation & Benchmarking, Prompt Engineering, NLP • Infra: AWS, Docker, Git, Terraform

Education

Master of Applied Science in Computer Science (MAS-CS) *Aug 2025 – May 2027*
University of Pennsylvania, Philadelphia, PA GPA 3.96

- Venture Incubation Program at Penn Venture Lab (Design Innovation Prize); Widjaja Engineering Entrepreneurship Fellow.

M.D.; Residency in Preventive Medicine; PhD in Biomedical Sciences & Public Health *2013–2025*
UCSC, Rome; M.Sc. Epidemiology & Biostatistics 110/110 • 30/30 honors

Experience

Software Engineer *May 2026 – Aug 2026*
Superscript, New York, NY

- Deployed AI for treatment selection and pricing accuracy with deterministic algorithms (Python) + LLMs.
- Built an explainability platform and agentic AI workflows that measurably improved accuracy and cut data drops; CI/CD on AWS Terraform/Aurora/Fargate.

Tech Consultant *Mar 2025 – Jul 2025*
Fondazione ENI Enrico Mattei, Rome, Italy

- Designed a Python RAG chatbot for a health app; first pilot drove strong gains in user comprehension and favorable ratings.

Research Leadership

Vice President of Research *Dec 2025 – Sep 2026*
Anthropic Claude Builder Club, University of Pennsylvania

- Lead a research project on AI alignment and safety.
- Lead a research project on model evaluation and benchmarking.

Technical Work

Medical AI Leaderboard (founder/maintainer) *2025 – Present*
github.com/causius0/medical-ai-leaderboard

- Live, public benchmark evaluating LLMs on European medical residency exams (SSM, MIR, Portuguese, French) across specialties; Python pipeline + Next.js/Postgres UI; position-bias testing via shuffled answer keys.

EuropeMedQA (lead) *2025 – Present*
Open-source benchmark

- Built the largest European multilingual, multimodal medical QA dataset for LLM benchmarking; open-sourced (arXiv:2604.14306). NeurIPS 2025 MMRL4H Best Presentation.

LLM Evaluation *2025*

- One-shot evaluation of Mistral-7B on EuropeMedQA; comparative chain-of-thought analysis across DeepSeek, Grok, Gemini, and Claude for medical reasoning.

Applied Predictive Modeling *2023 – 2026*

- Trained and validated a survival-prediction model for bladder cancer (300+ patients), published in JMIR Perioperative Medicine; 2025 JMIR Early-Career Researcher Award.

Selected Publications

- **76 peer-reviewed publications, h-index 14** (Google Scholar).
- Survival prediction in bladder cancer using a machine learning algorithm. *JMIR Perioperative Medicine*, 2026.
- Are Large Vision-Language Models Truly Grounded in Medical Images? *NeurIPS 2025 MMRL4H* (Best Presentation).

Founder

Monolith (founder, Dec 2025–) — AI-native journal, Node/React/PostgreSQL; Penn Venture Lab grant.
SIIAM (co-founder & VP, 2023–2025) — 500+ members; initiated EuropeMedQA.