

Thomas Charlon, MEng, PhD

Boston, MA, USA — LinkedIn — GitHub — Website — thomas_charlon@hms.harvard.edu

EXPERIENCE

Harvard Medical School, Department of Biomedical Informatics

Jan 2024 - Present

Research Associate - Senior postdoc, CELEHS laboratory

Boston, MA, USA

- GenAI LLM agents with Azure/Ollama for endoscopy scores, heart failure and suicidality chart reviews, ontology creation (GPT-5, Llama, fine-tuned Phi4, R, Pytorch)
- Semantic search ontology webapp using LLM embeddings and vector database Elasticsearch, BGE fine-tuning, evaluation of multiple BERT models (R/Shiny, Pytorch)
- Suicide risk prediction models based on hospital data (EHR codified data and clinicians' notes), de-identification of notes, embedding clustering of suicide-related publications using Glove
- Patient history dashboard for rheumatology treatment switches, deployed at MGB hospital, VA-funded
- Ontology mappings creation used by multiple institutions (MGB, CHA, Duke, Pittsburgh, Mayo Clinic)
- Insurance claims co-occurrence analysis and knowledge graphs of 210M patients (Inovalon)
- Mentorship of 4 research assistants and 1 PhD student, lead of group meetings, collaborations with 2 postdocs
- Project quality processes, grant writing, onboarded at MGB and CHA hospitals, part of 4 IRBs

Independent Researcher

Feb 2018 - Dec 2023

- Development of a real-estate market estimation application using R/Shiny
- Social network analysis on Twitter, identifying withheld content across Europe, chronological topic modeling
- Algorithmic cryptocurrency trading, fully automated market-neutral system using APIs

Quartz Bio, part of Precision for Medicine

Feb 2014 – Jan 2018

Bioinformatician & PhD Student

Geneva, Switzerland

- Led genetic analyses for the EU-funded PreciseSADs, developed genome-wide clustering algos for autoimmunity
- Visiting researcher at Genyo Institute (Pfizer / University of Granada, Spain)

SKILLS

Data Science & AI: Embeddings Semantic Search, LLM Fine-tuning, Unsupervised Clustering

Software & Web Development: R/Shiny, Pytorch, Langchain, FastAPI, JavaScript, SAS

Server Administration & Databases: SQL, PostGIS, Elasticsearch, Git, Docker Compose, AWS

LONG-TERM PROJECTS

Center for Suicide Research and Prevention — *Survival analysis, Risk prediction, EHR, LLM* 2024 – Present

- Analysis of patient data at MGB and CHA, cohorts and outcomes definitions using 2M patients
- Innovative methods core (knowledge graphs) and demographics stratified modeling
- LLM agent chart reviews, feature interpretation with NLP, integration of results in R/Shiny web apps

Real-Estate Price Estimation App — *Market pricing, Public-facing R/Shiny & JavaScript app* 2021 – 2023

- Visualization and neighborhood analysis of 15M French real-estate transactions over 5 years
- Public facing app, server admin (PostGIS, OpenStreetMap), business model and marketing

PreciseSADs, genetics of autoimmune diseases — *Unsupervised density-based clustering* 2014 - 2018

- Analysis of genome-wide SNP data of 10k patients, density-based clustering, haplotype summarization

SELECTED TALKS ON YOUTUBE

- **R/Medicine 2025:** Co-occurrence analysis and knowledge graphs for suicide risk prediction
- **SoCal Linux 2025:** R/Python pipelines for biomedical semantic search
- **R/Pharma 2024:** Knowledge graphs for drug discovery, **R/Medicine 2024:** Word embeddings in mental health

LEADERSHIP

- **Open-Source Projects:** Published 6 R packages on CRAN, multiple citations (Nature), 50k downloads
- **Mentorship:** Supervision of junior researchers, lectures at Harvard DBMI, collaborations with postdocs
- **Project Management:** Led projects from inception to delivery, developing specifications with clients

EDUCATION

University of Geneva

2014 - 2019

PhD in Computer Science / Bioinformatics

Geneva, Switzerland

Clustering and sparse coding of genome-wide data and proteomics in systemic autoimmune diseases

EPITA

2010 - 2013

Master of Engineering in Computer Science

Le Kremlin-Bicêtre, France

Artificial intelligence and cognitive sciences major, C++, Java, Matlab