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

EDUCATION

University of Geneva

PhD in Computer Science, employed as bioinformatician at Quartz Bio Research focused on clustering and sparse coding of genome-wide data and proteomics in systemic autoimmune diseases.

EPITA

Master of Engineering in Computer Science Artificial intelligence and cognitive sciences major, C++, Java, Matlab

SKILLS

Data Science & AI: Embeddings Semantic Search, LLM Fine-tuning, NLP, Unsupervised Clustering, Machine Learning Software & Web Development: R/Shiny, Pytorch, Langchain, FastAPI, JavaScript, SAS Server Administration & Databases: SQL, PostGIS, Elasticsearch, Git, Docker Compose, AWS

EXPERIENCE

Harvard Medical School, Department of Biomedical Informatics	Jan 2024 - Present Boston MA USA
 Suicide risk prediction model development based on electronic health records (codified dat Semantic search using LLM embeddings (BGE, BERT), GenAI Llama for variables mapp Enhancing statistical visualizations and standardizing codebooks for diagnoses and medic Analysis of 8,000 suicide-related publications using Glove embeddings and clustering metic Established quality processes for project development and contributed to grant writing 	a and unstructured text) ing ations hods
Independent Researcher and Entrepreneur	2019 - 2023
 Development of a real-estate market estimation application using R/Shiny Social network analysis on Twitter, identifying withheld content across Europe Development of chronological topic modeling methods Algorithmic cryptocurrency trading, fully automated system using APIs 	
 Quartz Bio (part of Precision for Medicine) Bioinformatician Led genetic analyses for the EU-funded PreciseSADs project. Developed clustering algorithms for large-scale genomic data analysis. Visiting researcher at Genyo Institute (Pfizer / University of Granada, Spain) 	Feb 2014 – Jan 2018 Geneva, Switzerland
PROJECTS	
 Real-Estate Price Estimation App — R/Shiny, JavaScript, ElasticSearch, OpenStreetMap Analyzed +15M French real-estate transactions for neighborhood price estimation Visualization with PostGIS and tailored OpenStreetMap tile server, address search with I Server administration with Nginx, Matomo, Netdata, Docker Compose, HTTPS, API tile 	2021 – 2023 Komoot ElasticSearch server
 Twitter Withheld Content Monitoring — R/Shiny, Twitter API, Graph Theory Investigated withheld content patterns across Europe using computational linguistics Development of novel chronological topic modeling methods 	2019 - 2020
 Quantitative Trading Application — R/Shiny, Financial Analysis, Trading Systems R/Shiny app for cryptocurrency statistical arbitrage Market neutral strategy based on tick data (Z index, sliding time window variability) 	2019 - 2020

PRESENTATIONS & FIRST AUTHOR PEER-REVIEWED PUBLICATIONS

- 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
- F1000Research 2017: Replication of the principal component analyses of the human genome diversity panel
- PLOSOne 2016: Single nucleotide polymorphism clustering in systemic autoimmune diseases

LEADERSHIP

- Open-Source Projects: Published multiple R packages on CRAN and GitHub
- Mentorship: Supervised junior researchers, fostering skill development
- Project Management: Led projects from inception to delivery, developing specifications with clients

2014 - 2019

2010 - 2013

Geneva. Switzerland

Le Kremlin-Bicetre, France