Tommaso Alfonsi

Data scientist, software engineer, bioinformatician

Contact information	tommaso.alfo@gmail.com	https://tomalf2.github.io

---

Summary

I'm an expert data scientist and software engineer, with a proven track record of impactful and innovative software solutions for data integration, clustering, pattern recognition and outlier detection in high-dimensional data domains, genomics and health care.
I master the development of ETL pipelines, machine-learning models, data visualization and API integration/development through Python and relevant frameworks, despite knowing other programming languages too.
Proficient in data modelling and database construction/querying.
I have a passion for delivering high-quality software products and researching innovative approaches in any field of software and science.

Projects

1. (2025) FluWarning: dockerized ETL pipeline for data cleaning/alignment and outlier detection + interactive web app showcasing results with dynamic filters, tables, maps, and widgets.
2. (2025) FluDataStratification: data analysis pipeline finding clusters of Influenza genomes
3. (2025) Job Orchestra: Python library and pipeline orchestration framework with built-in memory persistency, memoization and dependencies resolution.
4. (2025) CodonLLM: ex-novo large language model (LLM) generating sequences of DNA/RNA codons.
5. (2024) RecombinHunt: Python library and software identifying mosaic patterns in genomic data sequences.
6. (2021) CoV2K: JavaScript and HTML web scrapers + ETL pipeline + NoSQL database + RESTful API enabling graph-data representation.
7. (2021) ViruSurf-downloader: ETL pipeline for the automated continuous database integration of multi-source genomic data, embedding data curation and automated periodic database backups and optimizations.
8. (2021) VarSum: RESTful API enabling user-personalized queries to a multi-source repository of genomic variants and annotations.
9. (2018) MTG Life Counter+: Android app supporting the famous card game.
10. (2016) FlashApp: Android app automating the exchange of fiscal receipts.

More information and links available on https://tomalf2.github.io/projects/

Software development experience

Experiences are linked to the above project numbers

- ◎ Frontend web apps with Streamlit [1]
- ◎ Backend RESTful APIs with Flask and FastAPI [6,7,8]
- ◎ Provisioning of cloud computing virtual machines with Amazon Web Services Elastic Compute Cloud AWS EC2 VMs, and CINECA OpenStack [7, +others]
- ◎ Web Server configuration (Nginx) [1,6,8]
- ◎ Scripting (Bash, Python) [all]
- ◎ Python and Java parallel algorithms [5,7,8]
- ◎ Web scraping [6]
- ◎ Android app development [9,10]

Best practices of software development, maintenance, testing, integration, and delivery (CI/CD), including:
- ◎ Versioning (GIT, GitHub) [all]
- ◎ Object-oriented programming (OOP) & software design patterns [all]
- ◎ Software engineering principles [all]
- ◎ Unit tests [3,4,9,10], integration tests [all]
- ◎ Software containerization (Docker, PyPI) [1,3]
- ◎ Software dev. environment with Conda, PIP, UV [all]
- ◎ Documentation [1,3,5,6,8,9]

- ◎ Development languages: Python (+4 yr) [1,2,3,4,5,6,7,8], Java (2yr) [9,10, +others], Scala (1 yr) [8], PHP/JavaScript/CSS/HTML (1 yr), C/C++ (1 yr)

Data science experience Experiences are linked to the above project numbers	◎ AI/ML models with PyTorch and Scikit-learn: Large Language Models (LLMs) engineering and few-shot learning [4], Clustering analysis (kmeans/hierarchical) [2], Principal Component Analysis (PCA) [2], Simple/Multinomial classification [2], Gaussian Mixture Variational Auto Encoders (VAE/GMVAE) [1], Regression analysis, Convolutional (CNNs) and Recurrent (RNNs) neural networks for image-recognition

	◎ Data visualization (Plotly, Matplotlib, Seaborn) [1,2,3,5,8] ◎ Statistical tests for the distribution equality [2], test of hypotheses and goodness of fit [5]	◎ Feature engineering and selection [1,2] ◎ Key data insights extraction [2,5,8] ◎ ML model training, regularization, and hyperparameter estimation [1,4] ◎ Repeatable, traceable experiment pipelines [1,2,3,4]

Data management experience Experiences are linked to the above project numbers	◎ Relational databases: CRUD operations, schema design, using PostgreSQL [7,8] ◎ Hierarchical/NoSQL databases: CRUD operations, schema design using MongoDB, XPath [6,7]	◎ Data warehouse design [8] ◎ Query optimization [7] ◎ Data migration [7] ◎ Database Automation/Database Operations (DBOps, DataOps) [7]

Languages	English Fluent – TOEIC score 870/990 (~C1)	Italian Mother tongue

Education	2017 - 2019 **MSc**, Computer Science Engineering Politecnico di Milano, Milan, Italy	2020 - 2023 **PhD**, Information Technology, Politecnico di Milano, Milan, Italy

Awards	◎ Awarded the Dimitri Chorafas Award for the extraordinary scientific achievements in computers and communications technology

Soft Skills	◎ I often present my work abroad at conferences and seminars and to external partners. This habit trained me as a skilled and versatile communicator, adept at tailoring complex concepts for technical and non-technical stakeholders. ◎ I like working in a team and building relationships with colleagues based on trust, empathy and collaborative success. ◎ I am excited by new challenges and opportunities that allow me to demonstrate my skills, learn new ones, and propose effective solutions. ◎ I am a reliable end-to-end product developer, keen on delivering software products that balance high quality with timeliness and performance with extensibility.

Contact information	Residence: Varese, Lombardy, Italy Nationality: Italian Date of birth: December 29th, 1993 Interested in full-remote options as well as hybrid/in-person work options in Northern Italy and Switzerland. Contact me at: tommaso.alfo@gmail.com More info at: https://tomalf2.github.io Looking forward to discussing how my skills and experience can contribute to your team.	