

# Tommaso Alfonsi

Data scientist, software engineer, bioinformatician

Contact information

[tommaso.alfa@gmail.com](mailto:tommaso.alfa@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 <sup>1</sup>
- ⊙ Backend RESTful APIs with Flask and FastAPI <sup>6,7,8</sup>
- ⊙ Provisioning of cloud computing virtual machines with Amazon Web Services Elastic Compute Cloud AWS EC2 VMs, and CINECA OpenStack <sup>7, +others</sup>
- ⊙ Web Server configuration (Nginx) <sup>1,6,8</sup>
- ⊙ Scripting (Bash, Python) <sup>all</sup>
- ⊙ Python and Java parallel algorithms <sup>5,7,8</sup>
- ⊙ Web scraping <sup>6</sup>
- ⊙ Android app development <sup>9,10</sup>

Best practices of software development, maintenance, testing, integration, and delivery (CI/CD), including:

- ⊙ Versioning (GIT, GitHub) <sup>all</sup>
- ⊙ Software design patterns <sup>all</sup>
- ⊙ Software engineering principles <sup>all</sup>
- ⊙ Unit tests <sup>3,4,9,10</sup>, integration tests <sup>all</sup>
- ⊙ Software containerization (Docker, PyPI) <sup>1,3</sup>
- ⊙ Software dev. environment with Conda, PIP, UV <sup>all</sup>
- ⊙ Documentation <sup>1,3,5,6,8,9</sup>

- ⊙ Development languages: Python (+4 yr) <sup>1,2,3,4,5,6,7,8</sup>, Java (2yr) <sup>9,10, +others</sup>, Scala (1 yr) <sup>8</sup>, 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 <sup>4</sup>,
  - Clustering analysis (kmeans/hierarchical) <sup>2</sup>,
  - Principal Component Analysis (PCA) <sup>2</sup>,
  - Simple/Multinomial classification <sup>2</sup>,
  - Gaussian Mixture Variational Auto Encoders (VAE/GMVAE) <sup>1</sup>,
  - Regression analysis,
  - Convolutional (CNNs) and Recurrent (RNNs) neural networks for image-recognition

- ⊙ Data visualization (Plotly, Matplotlib, Seaborn) <sup>1,2,3,5,8</sup>
- ⊙ Statistical tests for the distribution equality <sup>2</sup>, test of hypotheses and goodness of fit <sup>5</sup>

- ⊙ Feature engineering and selection <sup>1,2</sup>
  - ⊙ Key data insights extraction <sup>2,5,8</sup>
  - ⊙ ML model training, regularization, and hyperparameter estimation <sup>1,4</sup>
  - ⊙ Repeatable, traceable experiment pipelines <sup>1,2,3,4</sup>
- 

## Data management experience

Experiences are linked to the above project numbers

- ⊙ Relational databases: CRUD operations, schema design, using PostgreSQL <sup>7,8</sup>
- ⊙ Hierarchical/NoSQL databases: CRUD operations, schema design using MongoDB, XPath <sup>6,7</sup>

- ⊙ Data warehouse design <sup>8</sup>
  - ⊙ Query optimization <sup>7</sup>
  - ⊙ Data migration <sup>7</sup>
  - ⊙ Database Automation/Database Operations (DBOps, DataOps) <sup>7</sup>
- 

## 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 29<sup>th</sup>, 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](mailto: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.

