

Erika Chang de Azevedo — Master Knowledge Base

Positioning

Data Scientist (Ph.D. in Biomolecular Physics) focused on **Finance**, **Bioinformatics / Research-driven analytics**, and **Consulting**.

Combines analytical rigor, machine learning expertise, and clear stakeholder communication.

SECTION 1 — FREQUENTLY ASKED QUESTIONS

How long has Erika been a data scientist?

Erika has been working professionally as a **data scientist since mid-2023**, when she joined **TIPREV**.

Before that, she worked for several years with data through her **Ph.D. research** and **teaching**, where she applied statistics, modeling, and data visualization.

What sectors does Erika focus on?

Erika's professional focus lies in:

1. **Finance** — risk analysis, forecasting, fraud detection, and reporting.
 2. **Bioinformatics / research-adjacent analytics** — leveraging her biomolecular research background.
 3. **Consulting** — developing and communicating data solutions for diverse clients.
-

What roles is Erika targeting now?

Erika is looking for **Data Scientist**, **AI Consultant**, or **Data Analyst** roles in the **Netherlands and broader Europe**, especially in **financial analytics**, **bioinformatics**, or **consulting** environments.

What makes Erika's profile unique?

She combines:

- **Research-level rigor** from her Ph.D. work.
- **Production-oriented skills** in ETL, ML, and GenAI (RAG systems).
- **Communication clarity** developed through teaching and consulting.

This blend allows her to connect technical insights with real-world decision-making.

What tools and technologies does Erika use?

- **Programming:** Python, pandas, NumPy, scikit-learn, TensorFlow, Keras, LangChain, Hugging Face
 - **Data Engineering & MLOps:** SQL (PostgreSQL, SQLAlchemy), FastAPI, Docker, MLflow, CI/CD, FAISS, ChromaDB
 - **Visualization:** Tableau, Streamlit, matplotlib, seaborn
 - **Cloud:** Google Cloud (Cloud Run, Vertex AI)
-

Where did Erika study data science?

She completed the **Le Wagon Data Science & AI Bootcamp** in early 2025, focusing on applied machine learning, MLOps, and GenAI.

Her bootcamp projects included **Eyesense** (deep learning) and **SecureMed Chat** (a RAG GenAI API).

Where is Erika based?

Erika lives in **Amsterdam, the Netherlands**, and holds a **valid Dutch residence permit**.

In which languages can the chatbot answer?

The chatbot can answer in **English, Portuguese, or Dutch**, automatically matching the user's language.

SECTION 2 — PROFESSIONAL OVERVIEW

Name: Erika Chang de Azevedo

Location: Amsterdam, Netherlands

Title: Data Scientist | Data Analyst (Ph.D. Biomolecular Physics)

Focus Sectors: Finance · Bioinformatics · Consulting

Experience: Applied DS since 2023; 7+ years working with data overall

Links: [linkedin.com/in/ecdazevedo](https://www.linkedin.com/in/ecdazevedo) · github.com/erika-chang · erikaa.chang@gmail.com

SECTION 3 — CAREER STORY (TRANSITION NARRATIVE)

Erika began in **biomolecular physics** at the **University of São Paulo**, where she modeled protein structures, ran HPC-accelerated analyses, and published **peer-reviewed papers**. She developed skills in **Python**, **statistical modeling**, and **reproducible data pipelines**.

As a **teacher**, she learned to simplify complexity — using data to improve student performance and explain results clearly.

This phase built her **communication and storytelling** strengths.

In **2023**, she transitioned to applied data science at **TIPREV**, developing **forecasting** and **fraud-risk models**, automating **ETL pipelines**, and creating **Tableau dashboards** for 40+ municipalities.

This work deepened her interest in the **finance** and **consulting** sectors.

To modernize her technical stack, she completed the **Le Wagon Data Science & AI Bootcamp (2025)**, where she built **Eyesense** (deep learning for eye-disease detection) and **SecureMed Chat** (a privacy-focused RAG assistant).

These projects extended her skills into **bioinformatics**, **healthcare-adjacent analytics**, and **GenAI deployment**.

Today, Erika focuses on roles in **Finance**, **Bioinformatics**, and **Consulting**, applying research-grade precision and business-oriented delivery.

SECTION 4 — TECHNICAL SKILLS

Programming & ML: Python, pandas, NumPy, scikit-learn, TensorFlow/Keras, LangChain, Hugging Face

Data Engineering & MLOps: SQL (PostgreSQL, SQLAlchemy), ETL pipelines, FastAPI, Docker, MLflow, CI/CD, FAISS, ChromaDB

Visualization & BI: Tableau, Streamlit, matplotlib, seaborn

Cloud: Google Cloud (Cloud Run, Vertex AI)

Soft Skills: Communication, data storytelling, mentoring, reproducibility, adaptability

SECTION 5 — PROFESSIONAL EXPERIENCE

TIPREV (2023 – Present) — Data Scientist / Data Analyst

- Built **revenue forecasting** and **fraud-risk** models for 40 + municipal clients.
- Automated ETL (Python + SQL), reducing processing time from days to hours.
- Developed **Tableau dashboards** for pension health and compliance.
→ Relevance: Finance and Consulting.

Le Wagon (2025) — Data Science & AI Bootcamp

- **Eyesense:** Deep-learning app (TensorFlow/Keras + Streamlit) with 92 % accuracy.
- **SecureMed Chat:** RAG assistant (FastAPI, Vertex AI, ChromaDB) with privacy-first design.
→ Relevance: Bioinformatics / GenAI.

University of São Paulo (2013 – 2020) — Research Scientist

- Published 3 papers in computational biology and drug discovery.
- Developed GPU workflows (~100 × speedups).
- Applied PCA and Markov State Models to molecular datasets.
→ Relevance: Research-driven analytics and bioinformatics.

Secretaria da Educação de São Paulo (2021 – 2023) — Teacher

- Created a data-based student tracker that raised averages by 12 %.
 - Led workshops on data-informed instruction and collaboration.
→ Relevance: Communication and knowledge transfer.
-

SECTION 6 — EDUCATION

- **Le Wagon (2025)** — Data Science & AI Bootcamp
 - **Ph.D. (2015 – 2020)** — Biomolecular Physics, University of São Paulo
 - **M.Sc. (2013 – 2015)** — Biomolecular Physics, University of São Paulo
 - **B.Sc. (2009 – 2013)** — Biomolecular & Physical Sciences, University of São Paulo
 - **Licenciatura (2017 – 2019)** — Exact Sciences, University of São Paulo
-

SECTION 7 — PROJECTS

SecureMed Chat

Stack: FastAPI, Vertex AI, ChromaDB, GCP

Type: GenAI RAG assistant for medical intake.

Goal: Provide grounded answers while preserving user privacy (GDPR-compliant).

Relevance: Applies ML and API skills to healthcare / bioinformatics data.

Eyesense

Stack: TensorFlow/Keras, Streamlit, GCP

Goal: Predict eye diseases from retinal images.

Result: 92 % test accuracy.

Relevance: Demonstrates deep learning in bioinformatics / medical context.

RPPS Analytics (TIPREV)

Stack: Python, SQL, Tableau

Goal: Unify pension datasets (employees, benefits, finances) into a single warehouse.

Outcome: Enabled forecasting and fraud analysis for municipal funds.

Relevance: Finance and public-sector analytics.

ETL Pipeline

Stack: Python, SQLAlchemy, Docker, PostgreSQL
Goal: Transform raw data into BI-ready tables.
Impact: Improved auditability and reporting efficiency.
Relevance: Finance / consulting data engineering.

Portfolio RAG Chatbot

Stack: LangChain, FastAPI, FAISS/ChromaDB, Mistral LLM
Goal: Answer questions about Erika's professional background and projects.
Deployment: Docker + Cloud Run integration into her portfolio site.
Relevance: Demonstrates GenAI and MLOps integration.



SECTION 8 — PUBLICATIONS

1. *The β -lactam ticarcillin is a Staphylococcus aureus UDP-N-acetylglucosamine 2-epimerase binder*
 2. *Energy landscape of the domain movement in Staphylococcus aureus UDP-N-acetylglucosamine 2-epimerase*
 3. *Thermodynamic analysis of interactions of the Hsp90 with adenosine nucleotides: A comparative perspective*
-



SECTION 9 — VALUES & CURRENT FOCUS

Values: clarity, rigor, explainability, collaboration, continuous learning.
Current Focus: roles and projects in **Finance**, **Bioinformatics** / **Research-driven analytics**, and **Consulting**.
Approach: combines production-ready machine learning with stakeholder-friendly communication.



Version info

File: `data/source/erika_profile_full.md`
Maintainer: Erika Chang de Azevedo
Last updated: 2025-10-23
Use: Canonical reference for chunking and RAG embeddings.

Do not embed this file directly — generate Markdown chunks in [data/source/docs/](#) for FAISS indexing.