Erika Azevedo — Knowledge Base for Portfolio Chatbot

🧩 SECTION 1 — GENERAL PROFILE

Full name: Erika Azevedo
Location: Amsterdam, North Holland, Netherlands
Languages:

- Portuguese (Native)
- English (Fluent / Advanced)
- Dutch (Intermediate, improving)
- Spanish (Elementary)

Contact:
📧hello@erikachang.me
🔗 linkedin.com/in/ecdazevedo
💻 github.com/erika-chang

Professional Title:
Data Scientist | Data Analyst | Ph.D. in Biomolecular Physics

Short Bio:
Erika is a data scientist based in the Netherlands with a Ph.D. in biomolecular physics and a passion for making data truly useful for people. Her journey spans research, education, and applied data consulting — a rare mix that shaped her ability to combine analytical rigor, clear communication, and human understanding in data-driven solutions.

🎯 SECTION 2 — PROFESSIONAL SUMMARY

Erika specializes in **machine learning**, **predictive modeling**, and **data storytelling**, applying her scientific mindset to real-world problems.
She builds models, data pipelines, and GenAI applications that turn messy datasets into decisions, always with transparency and reproducibility in mind.

Her path from physics labs to AI prototypes reflects a deep commitment to learning, adaptation, and impact. Today, she bridges the gap between **complex data systems** and **clear insights** — helping organizations see the story behind the numbers.

🔄 SECTION 3 — CAREER TRANSITION STORY

Erika's career started in the world of **biomolecular physics**, where she spent years analyzing proteins, modeling molecular dynamics, and leading research at the **University of São Paulo**.
In that environment, she mastered not only technical rigor — Python, statistical modeling, GPU computing — but also the art of explaining intricate results to non-technical audiences.

When she later became a **teacher**, Erika discovered her second strength: **communication and empathy**. She learned how to translate abstract concepts into accessible ideas, guide others through learning curves, and design data-driven methods to improve performance — skills that later became essential in her consulting and data storytelling work.

By 2023, she began connecting both worlds. As a **data scientist at TIPREV**, she applied her analytical mindset to a very different field: public-sector pension systems. There, she automated large-scale data pipelines and built fraud-detection and forecasting models that helped dozens of municipalities manage financial risks.
This experience became the turning point of her transition — proof that her scientific precision could directly improve real-world systems.

To consolidate her new path, Erika joined **Le Wagon's Data Science & AI Bootcamp** in 2025, where she deepened her skills in **machine learning**, **MLOps**, and **GenAI**. She built end-to-end applications such as **Eyesense**, a deep-learning app for eye-disease prediction, and **SecureMed Chat**, a privacy-focused RAG chatbot for healthcare.
Those projects marked the culmination of her transition: from studying molecular interactions to **building AI systems that serve people**.

Today, Erika defines herself as a data scientist who blends the **rigor of a researcher**, the **clarity of a teacher**, and the **vision of an engineer**.

🧠 SECTION 4 — TECHNICAL SKILLS

Programming & ML:
Python, pandas, NumPy, scikit-learn, TensorFlow, Keras, LangChain, HuggingFace

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

Visualization & BI:
Tableau, Streamlit, matplotlib, seaborn

Cloud:
Google Cloud Platform (Vertex AI, Cloud Run)

Soft Skills:
Data storytelling, stakeholder communication, mentoring, adaptability, teamwork, knowledge sharing

---

💼 SECTION 5 — PROFESSIONAL EXPERIENCE

TIPREV (2023 – Present)

Role: Data Scientist / Data Analyst
Type: Consultancy | Pension Management SaaS

- Designed **revenue-forecasting and fraud-risk models**, improving oversight across **40+ municipalities**.
- Automated **ETL workflows** in Python and SQL, reducing manual data handling from days to hours.
- Built **interactive Tableau dashboards** for municipal leaders, improving decision speed and transparency.
- Strengthened **financial auditability and data governance** through structured database design.

Le Wagon — Data Science & AI Bootcamp (Jan – Mar 2025)

Role: Data Science Trainee

- Developed **Eyesense**, a deep-learning app for eye-disease prediction (TensorFlow/Keras + Streamlit, 92% accuracy).
- Built **predictive models** for e-commerce profitability and time-series forecasting.
- Practiced **Agile teamwork**, version control, and presentation of complex findings to non-technical peers.

- Strengthened understanding of **machine-learning pipelines**, **feature engineering**, and **MLOps**.

Universidade de São Paulo (2013 – 2020)

Roles: Senior Researcher (Ph.D. period, 2015 – 2020) | Assistant Research Scientist (2013 – 2015)

- Published **3 peer-reviewed papers** in computational biology and drug discovery.
- Built **bioinformatics workflows** on GPU clusters, achieving up to **100× faster** simulation speeds.
- Applied **PCA and Markov State Models** to identify hidden patterns in molecular datasets.
- Discovered a **novel bacterial enzyme inhibitor** through computational screening.
- Mentored researchers in **data visualization** and **reproducibility**, laying the foundation for later leadership roles.

Secretaria da Educação do Estado de São Paulo (2021 – 2023)

Role: High School Teacher — Physics and Biology

- Designed a **data-based student-performance tracker** that increased average scores by 12%.
- Delivered **teacher workshops** on data-driven instruction and collaborative learning.
- Introduced **technology-enhanced lessons**, improving student engagement and comprehension.

🎓 SECTION 6 — EDUCATION

- **Le Wagon (2025)** — Data Science & AI Bootcamp
- **Ph.D. in Biomolecular Physics (2015 – 2020)** — University of São Paulo

- **M.Sc. in Biomolecular Physics (2013 – 2015)** — University of São Paulo
- **B.Sc. in Biomolecular & Physical Sciences (2009 – 2013)** — University of São Paulo
- **Licenciatura in Exact Sciences (2017 – 2019)** — University of São Paulo

---

🧪 SECTION 7 — PROJECTS

1. SecureMed Chat

Stack: GenAI, FastAPI, Vertex AI, ChromaDB, GCP
Privacy-first **medical intake assistant** built with **Retrieval-Augmented Generation** and zero-persistence architecture.
Ensures **GDPR compliance** while enabling scalable, explainable LLM deployment.

---

2. Eyesense

Stack: TensorFlow/Keras, Streamlit, GCP
Deep-learning app for **eye-disease detection** trained on labeled retinal datasets.
Achieved **92% accuracy**; demonstrated potential of ML for accessible healthcare.

---

3. ETL Pipeline

Stack: Python, SQLAlchemy, Docker, PostgreSQL
End-to-end data pipeline transforming raw sales data into analytics-ready tables for BI dashboards.
Improved auditability, transparency, and speed for reporting teams.

---

4. RPPS Data Analytics (TIPREV Project)

Stack: Python, SQL, Tableau
Unified multiple municipal datasets — employees, contributions, benefits — into a **single pension data warehouse**.
Enabled **fraud detection, risk forecasting**, and **real-time monitoring** for government clients.

---

5. RAG Chatbot Portfolio API

Stack: Mistral, LangChain, FastAPI, FAISS/ChromaDB
Custom chatbot that answers questions about Erika's background and projects.
Deployed as part of her **portfolio website** to showcase LLM integration and self-referential knowledge retrieval.

📚 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 & WORK STYLE

- Brings **scientific rigor** and **clarity** to data problems.
- Believes in **explainable AI** and human-centered analytics.
- Thrives on **collaboration**, **learning**, and **cross-disciplinary exchange**.
- Known for **structured thinking**, **communication**, and **adaptability**.

🧭 SECTION 10 — CURRENT GOALS

- Build a **long-term data science career** in the Netherlands.
- Apply expertise in **LLMs**, **RAG**, and **data engineering** to production projects.
- Contribute to **financial transparency**, **ethical AI**, and **social impact initiatives**.
- Continue improving **Dutch fluency** and professional integration in Europe.