Joaquin Salvador Machulsky

AI/ML Engineer - Data Scientist
Master's Degree in Data Science
University of Buenos Aires

+54 9 3804 95-1391
joaquinmachulsky@gmail.com
joaquinmachulsky.com
GitHub
LinkedIn
jmachulsky@dc.uba.ar

Summary

Data Scientist with over 3+ years of experience in Python, specializing in developing and deploying production-grade AI systems and LLM-powered solutions. Hold a Master's degree in Data Science, with a strong foundation in Applied Mathematics and Computer Science that has supported my work on diverse data-driven projects. Proven expertise in architecting and deploying end-to-end ML pipelines using PyTorch, SQL, and cloud infrastructure (AWS, Docker), including a semantic search RAG system using Gemini LLM and Google embeddings for large-scale product catalogs. Passionate about building scalable, interpretable AI systems that drive significant business impact and eager to tackle complex challenges in cutting-edge AI environments, particularly in LLM operations and AI Agent development.

Education

- **University of Buenos Aires, Argentina** *2023 - 2025*
  Master's Degree in Data Science (Artificial Intelligence)
- **University of Buenos Aires, Argentina** *2020 - 2023*
  Bachelor's Degree in Data Science (Mathematics + Computing + Statistics)

Experience

- **NETV S.A** *Aug 2024 - Present*
  Data Scientist
  - Architected and deployed a production-grade semantic search Agentic RAG system using Gemini LLM and Google embeddings, enabling natural language queries for a 544-product catalog.
  - Led the full lifecycle of ML projects, from a PyTorch-based computer vision model for melanoma detection (80% recall) to a time series model forecasting market share for 250+ products.
  - Managed and deployed scalable cloud infrastructure using Docker and a full suite of AWS services (EC2, S3, Elastic Beanstalk, CodePipeline), ensuring high availability for both AI and ML applications.
  - Developed and maintained production APIs using Flask, and led the data sub-team through mentorship, code reviews, and daily standups.
  - Optimized large-scale SQL pipelines, reducing report generation time by over 95% (from 40 minutes to under 1 minute) through query redesign and async processing.
  - **Technologies used**: Python, **SQL, PyTorch, LLMs, RAG Systems, NLP, Computer Vision (OpenCV), Scikit-learn, AWS (EC2, S3, IAM, RDS, Elastic Beanstalk, CodePipeline), Docker, Flask, SQLAlchemy, Git, GitHub, PostgreSQL, Vector Search**.

- **LAS MERCEDES** *Mar 2025 - Present*
  Freelance ML & Backend Developer
  - Designed and built a custom, end-to-end automated system to reconcile and analyze over 500 monthly POS transactions, providing ongoing support and maintenance.
  - Developed a secure backend API using Flask, implementing full CRUD functionality and user authentication to manage transaction data.
  - Applied data analysis and ML techniques to generate automated reports and dashboards, delivering key business insights and key performance indicators (KPIs), enhancing data accessibility and facilitating strategic insights for senior management.
  - **Technologies used**: Python, **Flask, AWS, SQL (PostgreSQL), Pandas, Scikit-learn, Matplotlib, Seaborn, Git, GitHub**

- **RUSSO SEGUROS** *Nov 2023 - May 2025*

  Data Analyst & Data Developer

  – Automated the ingestion of insurance policy and payment data using Python, Pandas, and Selenium, significantly reducing manual data entry and processing time.

  – Maintained and optimized PostgreSQL databases to ensure data integrity, consistency, and performance for critical client reporting.

  – Designed and built automated reports for financial and operational analysis using Python to visualize key performance indicators (KPIs), enhancing data accessibility and facilitating strategic insights for senior management.

  – Implemented predictive models in Python to forecast business trends.

  – **Technologies used**: Python, **SQL, PostgreSQL, Pandas, Selenium, Matplotlib, Seaborn.**

Skills

- **Languages**: English (FCE B2), Spanish (native)
- **Programming Languages**: Python, SQL, C++, JavaScript, HTML.
- **AI/ML & Data Science**: LLMs (Gemini), RAG Systems, Semantic Search, Embeddings, PyTorch, TensorFlow, Scikit-learn, Computer Vision, NLP, Time Series Analysis.
- **Data & Backend**: Pandas, NumPy, OpenCV, Flask, SQLAlchemy, Marshmallow, Selenium, SciPy, Matplotlib, Seaborn.
- **Cloud & DevOps**: AWS (EC2, S3, IAM, RDS, Elastic Beanstalk, CodePipeline), Docker, Git, GitHub, Google Cloud APIs, Postman.
- **Databases**: MySQL, SQL Server, PostgreSQL, Vector Search, JSON-based storage.
- **Frontend & Collaboration**: React, Material UI, Jupyter, VS Code.
- **Areas of Interest**: Artificial Intelligence, AI Safety, LLM's, Machine Learning, NLP, Data Analysis, RAG Systems, Automation, Graph Theory, Game Theory.

Personal Projects

- **AI Agents Alignment: Scalable AI Safety via Debate** *GitHub*

  Research exploring debate mechanisms for AI alignment

  – Research mechanisms for scalable supervision of AI Agents via structured adversarial interactions between agents.

  – Implemented full simulation game (zero-sum) of a debate protocol using PyTorch, including computer vision models, turn-based strategies, and agent asymmetry capabilities.

  – Contributing to AI Safety research by empirically showing that debate can significantly improve accuracy of limited judges.

- **Genomic Expression Classification** *GitHub*

  Study to determine genetic factors influencing prognosis in patients with pre-tumoral lesions.

  – Construction and comparison of models.

  – Hyperparameter optimization and exploration. Learning and complexity curves.

  – Validation techniques and performance evaluation.

- **Dream Journal** *GitHub*

  Research and analysis on people's dreams.

  – Used web scraping techniques to collect dream narratives from an online forum.

  – Performed NLP-based analysis on dream narratives, including sentiment, coherence, and topic modeling.

  – Analyzed user preferences based on publication metrics.

  – Investigated the popular Argentine belief regarding the relationship between dreams and lottery numbers.