Silas Yunghwa Liu
Data Scientist - LLM/Graphs Specialist

Data Scientist Specialist in LLMs and Graphs with experience in delivering solutions using Machine Learning, Data Science and Visualization techniques for industries like insurance, public services, energy and retail. Also Professor and Advisor of Data Science MBA. Proficient in Python, Cypher and SQL, with expertise in Knowledge Graphs, Large Language Models, Time Series, Neural Networks and PySpark.

PhD in Electrical Engineering (2016) from UFMG, Brazil, with Doctoral Stay of one year in Technische Universität Dresden, Germany. Master (2011) and bachelor (2008) in Electrical Engineering from UFMG, Brazil. In addition to Portuguese, I speak English, German and Chinese.

Personal Info

Location

Belo Horizonte,
Minas Gerais,
Brazil

Phone / WhatsApp

+55 31 99320-7967

E-mail

silasyl@gmail.com

LinkedIn

linkedin.com/in/silasyliu

GitHub

github.com/silasyl

Portfolio

https://www.silasdata.com/

Languages

Portuguese: 5/5 - Native

English: 4/5 - Advanced

German: 3/5 - Intermediate

Chinese: 1/5 - Basic

Skills

- **Programming Languages:** Python, SQL, Cypher, SparQL, R, Java, JavaScript, C, Pascal, Matlab
- **Databases & Storage:** Hive, MongoDB, MySQL, PostgreSQL, Neo4j, Knowledge Graph, Vector Database
- **Data Tools:** Databricks, Hadoop, Pandas, NumPy, SciPy, StatsModels, PySpark
- **Visualization Tools:** Matplotlib, Seaborn, Plotly, Streamlit, Gradio
- **Cloud:** Google Cloud Platform, Amazon Web Services, Microsoft Azure
- **ML Frameworks**: Librosa, Scikit-learn, Scikit-image, XGBoost, NLTK, Gensim, SpaCy, Polyglot, TensorFlow, Keras, PyTorch, OpenCV, PyTesseract
- **LLM Frameworks**: OpenAI, Gemini, Claude, DeepSeek, Transformers, CrewAI, Langchain, LangGraph, LlamaIndex, Smolagents
- **Others:** GitHub, Heroku, Jupyter, Kedro, Linux

Experience

Data Scientist Specialist, *Ambush*   Dec 2025 - Present
- Designing and delivering applied-AI solutions using LLMs, graph-based methods and complex datasets. Focused on production-ready architectures, retrieval workflows and model evaluation to support high-impact projects for companies worldwide.

Data Science MBA Professor, *XP Educação*   Jul 2022 - Present
- Taught MBA courses in Data Science, including Big Data Architecture and Advanced Machine Learning Techniques.
- Acted as an MBA Advisor mentoring graduates on applied projects in Machine Learning; this involved delivering lectures, developing course materials, and providing guidance on practical implementations.

Delivery Technical Lead, *Turing*   Mar 2025 - Nov 2025
- Technical Lead for a leading company in LLM base model development, involved in multiple R&D workstreams across model training, evaluation and pipeline automation.
- Led the multimodal visual workstream, integrating image datasets into LLM training to enhance reasoning, grounding and model alignment.
- Designed and implemented data augmentation pipelines and validator workstreams, improving dataset quality, consistency and scalability for large-scale model training.

- Provided technical leadership to a cross-functional R&D team, guiding practices, code reviews and performance monitoring of generative AI systems.

Sr. Data Scientist, *Peers Consulting & Technology*   Apr 2024 - Feb 2025
- Led the development of a decision-support tool, integrating people-analytics, Knowledge Graphs, LLM and RAG, optimizing workforce management. Built with Neo4j, Python, Cypher, SparQL, Langchain, LlamaIndex, Kedro and Streamlit.
- Enhanced the internal chatbot using multi-agent, multimodal models and RAG pipelines, increasing employee productivity by 3.7 hours per week.
- Conducted a Generative AI assessment for a major retailer, analyzing use cases for feasibility, impact and scalability.
- Developed a retail sales forecasting model incorporating seasonality, promotions and customer behavior, informing inventory and pricing strategies.
- Built ETL processes and KPIs for Power BI dashboards, supporting Brazil's nationwide technical education analysis. Used Azure, AWS, PySpark, Databricks, Jupyter, Deltalake, PostgreSQL, EMR, S3

IT Sr. Analyst / Data Scientist, *Peers Consulting & Technology*   Apr 2023 - Apr 2024
- Performed exploratory analysis and reporting on national insurance datasets using PySpark, Databricks and Deltalake, supporting strategic decisions.
- Executed data auditing and statistical analysis of digital marketing metrics (traffic, engagement, conversion, retention, SEO), producing a 500-page report and ongoing monitoring support.
- Applied market hypersegmentation and graph analysis in retail, leveraging LLMs for model explainability.

Data Scientist / Sr. Researcher, *Digital Grid*   Apr 2022 - Mar 2023
- Modeled and deployed a daily time series forecasting model for solar power generation using statistical models. The solution was implemented on AWS with S3, DynamoDB and Jupyter, achieving an accuracy of 70%.
- Contributed to a research project involving computer vision, character recognition via OCR, ROI detection, image processing and NLP. The project utilized a range of tools including OpenCV, PyTesseract, Scikit-image, NLTK, TensorFlow, AWS Textract and Amazon Rekognition.

Education & Training

PhD, Electrical Engineering, Universidade Federal de Minas Gerais, Brazil (2016)
- Doctoral Stay of one year in Technische Universität Dresden, Germany.
MSc, Electrical Engineering, Universidade Federal de Minas Gerais, Brazil (2011)
BSc, Electrical Engineering, Universidade Federal de Minas Gerais, Brazil (2008)

Certifications: **Neo4j Graph Data Science Certification** - Neo4j (2023/10), **Google Analytics Certification** - Google (2023/07), **Neo4j Certified Professional** - Neo4j (2023/06), **Professional Data Scientist** - DataCamp (2022/12), **Associate Data Scientist** - DataCamp (2022/09).

Courses: **Agents Course** - Hugging Face (2025/05), **Graph Theory Algorithms** - Udemy (2024/08), **Retrieval Augmented Generation For Production With Langchain & Llamaindex** - Activeloop (2024/01), **Essentials of Genomics and Biomedical Informatics** - edX IsraelX (2023/08), **Machine Learning with Python-From Linear Models to Deep Learning** - edX MITx (2022/07), **Data Science Professional Program** - edX HarvardX (2021/11), **CS50's Introductiom to Artificial Intelligence with Python** - edX HarvardX (2021/06).

Achievements: **1st Place DatathONS 3rd Edition 2022** – Project: Wind Prediction Bias Correction (2022/06).