David Guzmán

✉ david.guzman@alumni.utoronto.ca 📞 (437) 973 8963 🔗 davidguzmanr.github.io in enrique-david-guzman-ramirez

Experience

Data Engineer *Mexico City, Remote*
J.D. Power *March 2024 - Present*

- Designed and deployed scalable ETL pipelines with PySpark and BigQuery SQL, processing multi-source data across multiple client accounts and reducing latency by 30%.
- Created internal Python libraries to standardize and automate pipeline deployment via GitLab, decreasing manual coding effort by 20% and increasing reusability across projects.
- Developed and integrated robust data quality and validation checks, improving data accuracy and reducing data-related issues by 30%.

Machine Learning Research Intern *Toronto, Canada*
National Research Council Canada *May 2023 - December 2023*

- Collaborated in the *Speech Generation for Indigenous Language Education* project to develop state-of-the-art speech synthesis systems and techniques with Deep Learning for Indigenous languages in Canada.
- Contributed to the development of *EveryVoice TTS Toolkit*, a tool for creating text-to-speech systems focused on low-resource languages, achieving robust performance with up to 90% less data.

Data Scientist *Mexico City*
Grupo Salinas *November 2021 - July 2022*

- Developed a collaborative filtering recommender system for the homepage of the E-commerce area with PyTorch, which increased precision at 10 by 50%.
- Developed a content-based recommender system and association rules for the product detail pages of the E-commerce area with scikit-learn and Mlxtend, automating the original 2 week manual process.
- Deployed ML APIs on AWS SageMaker for real-time personalization and monitored performance using AWS CloudWatch.

Education

MSc in Applied Computing, *University of Toronto,* GPA: 3.93/4. *September 2022 – June 2024*
BSc in Data Science, *National Autonomous University of Mexico,* GPA: 9.60/10. *August 2019 - June 2021*
BSc in Physics, *National Autonomous University of Mexico,* GPA: 9.54/10. *August 2015 - June 2020*

Skills

- **Programming Languages:** Python, Julia, R, Fortran, SQL.
- **Machine and Deep Learning:** Scikit-learn, XGBoost, PyTorch, Keras, TensorFlow, PySpark MLlib.
- **NLP:** NLTK, spaCy, Hugging Face, LangChain.
- **Data Analysis and Manipulation:** Numpy, SciPy, Pandas, Matplotlib, Seaborn, Bokeh, Plotly.
- **Cloud services:** AWS, Azure.
- **Databases:** PostgreSQL, Redshift, MongoDB, Redis.
- **Miscellaneous:** Linux, Git, LaTeX, Slurm, REST APIs, Scrum.
- **Certifications:** AWS Certified Machine Learning - Specialty ↗.
- **Languages:** Spanish (Native), English (Proficient), French (Intermediate).

Publications

- Steve Bakos, Félix Gaschi, **David Guzmán**, Riddhi More, Kelly Chutong Li, En-Shiun Annie Lee *"AlignFreeze: Navigating the Impact of Realignment on the Layers of Multilingual Models Across Diverse Languages."* Presented at the 2025 Annual Conference of the North American Chapter of the Association for Computational Linguistics. 2025. Available at ACL Anthology: https://aclanthology.org/2025.naacl-short.48 ↗.
- Aidan Pine, Erica Cooper, **David Guzmán**, Eric Joanis, Anna Kazantseva, Ross Krekoski, Roland Kuhn, Samuel Larkin, Patrick Littell, Delaney Lothian, Akwiratékha' Martin, Korin Richmond, Marc Tessier, Cassia Valentini-Botinhao, Dan Wells, Junichi Yamagishi. *"Speech Generation for Indigenous Language Education."* 2025. Computer Speech & Language, Volume 90. Available at: https://doi.org/10.1016/j.csl.2024.101723 ↗.
- Tong Su, Xin Peng, Sarubi Thillainathan, **David Guzmán**, Surangika Ranathunga, En-Shiun Annie Lee. *"Unlocking Parameter-Efficient Fine-Tuning for Low-Resource Language Translation."* Presented at the 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics. 2024. Available at ACL Anthology: https://aclanthology.org/2024.findings-naacl.263 ↗.