

Profile Summary

Data engineer and machine learning practitioner with a strong track record of building scalable data pipelines, source-of-truth reporting systems, and production-ready ML infrastructure. Experienced in designing and maintaining robust ETL workflows, developing analytics-ready data models using SQL, dbt, and Fivetran, and supporting cross-functional teams with clean, reliable data. Skilled in CI/CD practices, version control, and deploying containerized systems using Docker and GitHub Actions.

Complementing a strong foundation in data engineering, I also bring expertise in applied machine learning and AI—particularly in PyTorch, CUDA, NLP, and transformer models. I've developed and deployed ML systems for classification, clustering, and reinforcement learning, with hands-on experience building end-to-end MLOps pipelines and performance monitoring frameworks.

Currently pursuing a Master's in Computer Science with a focus on AI, MLOps, and data infrastructure, I'm passionate about solving real-world problems at the intersection of engineering and intelligence—bridging data systems and ML models to create tools that drive business impact and improve decision-making.

Skills

- Data Engineering: ETL workflows, dbt, Fivetran, SQL (CTEs, window functions), PowerBI
- Machine Learning: PyTorch, CUDA, classification, clustering, reinforcement learning
- MLOps: CI/CD, Docker, GitHub Actions, Prometheus, Grafana
- Analytics & Reporting: Source-of-truth architecture, dashboarding, KPIs
- Version Control: Git, modular design, testing frameworks
- AI Technologies: NLP, Transformer models, Deep RL
- Tools & Platforms: Salesforce, LangChain, Gradio, OpenAI Gymnasium, MongoDB

Education

Worcester Polytechnic Institute (Completing in 2026)

Master of Science in Computer Science – AI Track

Courses:

- Generative AI (PyTorch)
- Deep Learning

- Reinforcement Learning (PyTorch, CUDA, OpenAI Gymnasium)
- MLOps (Langchain, Docker, Gradio)
- Systems & Network Programming (C/C++)

Projects:

- GPU-accelerated RL models using PyTorch and CUDA:
- Pac-Man (Rainbow DQN, PPO) [GitHub]
- Atari Breakout – Score: 300.22 [GitHub]
- Deployed ML app with containerized pipeline using Docker with Prometheus + Grafana monitoring [GitHub]

University of Waterloo

Advanced Diploma in Data Science and Machine Learning

Projects:

- Custom CNN for Image Classification [Project File, Report]
- Recommendations for a Modern Data Stack

McGill University

Bachelor of Business Administration

Experience

GUIKER (Jan 2020 – Aug 2024)

Manager, Analytics Engineering & RevOps (May 2023 – Aug 2024)

Skills: data strategy, product design, leadership; version control (Git); PyTorch; machine learning

- Owned and optimized the organization's data strategy and ETL pipelines, providing actionable insights to technical and non-technical teams.
- Collaborated cross-functionally to develop self-serve data products, improving decision-making and operational efficiency.
- Implemented software development best practices (version control, CI/CD, modular design) to enhance data accuracy and productivity.
- Built complex SQL queries using CTEs and window functions to support key performance metrics.
- Established coding standards, documentation, and testing practices across the data team.
- Mentored a Revenue Operations Specialist in technical and strategic growth areas.
- Led a cost optimization project using K-means clustering, cutting marketing costs by 50% while maintaining lead volume.

Analytics Engineer (Aug 2022 – May 2023)

Skills: SQL, NoSQL (Mongo), ETL (dbt, Fivetran), visualization (PowerBI)

- Built dynamic PowerBI dashboards to visualize business-critical KPIs and inform executive decisions.
- Designed ETL pipelines using dbt and Fivetran, improving data access and reliability.
- Wrote performant SQL queries and created ad-hoc reports to address key business questions and support cross-functional initiatives.

Revenue Operations (Jan 2020 – Aug 2022)

Skills: GTM leadership; Salesforce CRM architecture (Certified Administrator ID 21485247)

- Oversaw outbound sales efforts, revenue strategy, and CRM administration.
- Built and optimized GTM strategies, driving \$800K in monthly pipeline for a new proptech product.
- Set up a centralized Salesforce data platform as the company's primary source of truth.
- Supported a team of 10 in outbound sales processes and operational analytics.

BUS.COM (YC W2016) – Rideshare Lead (Feb 2016 – Nov 2019)

Skills: data-driven go-to-market; relationship building; strategic selling

- Led analytics-driven strategies in new business, retention, and operations for the rideshare division from seed to Series B.
- Increased department revenue from \$50K to \$2M annually.
- Collaborated with technical and business teams to align analytics across sales, operations, and customer success.
- Delivered a seamless client experience with a churn rate <1%.
- Secured strategic partnerships with marquee clients including Coachella and Live Nation.

Testimonials

Danial Jameel (CEO, Saris AI)

"Jake was an indispensable part of the early team... a level 5 leader whose contributions can't be overstated."

Nicholas Capobianco (Sales Manager)

"Jake's mentorship shaped my leadership... helped my team grow in their skill sets."

Marissa Rocha (VP Revenue)

"Jake brings positivity, grit, and high team impact... a huge asset in a startup environment."

Jungle Jake

“Jake blends determination, EQ, leadership, and creativity. A steady, top-performing communicator and teammate.”

Project: Ms. Pac-Man as a Reinforcement Learning Testbed

Ms. Pac-Man is a complex environment for RL, balancing dynamic rewards and threats. The agent learns to optimize strategies through various RL architectures.

Key RL Concepts

- Reward Mapping
- Exploration vs Exploitation
- Temporal Learning (DQN, DDQN, LSTM)
- Adaptation to Dynamic Environments

Models Used

- Double DQN
- Prioritized Dueling DDQN
- Rainbow DQN (Noisy Nets, N-Step, Distributional DQN)
- Proximal Policy Optimization (PPO)

Training Enhancements

- Reward Shaping
- N-Step Scaling
- Risk Scaling

Results

- PPO performed best
 - No model cleared a full level (likely due to short 10K episode training)
 - Future: longer training, advanced models like vision transformers
-

Certifications

- Salesforce Certified Administrator
 - dbt Certification
-

Let me know if you'd like this chunked into individual JSON fields for vector DB ingestion.