

Google Cloud Certified Professional Data Engineer leading enterprise data strategy while still building: production LLM agents, governed cloud data ecosystems, and the ML underneath them. My starting question is always what problem we’re solving and what value it delivers — the technology follows from that answer, not the other way around.

EXPERIENCE

Noise Digital MAY 2021 - PRESENT

VP, Data & AI JAN 2026 - PRESENT

- Led design and development of an enterprise agent that lets CRM teams build audience segments over first-party data enriched with multiple ML model attributes and push them to two activation platforms, all through natural language, cutting a month-long task down to minutes.
- Productionized a suite of GenAI applications, including an LLM-driven semantic-model agent for governed natural-language querying and visualization, and a persona-simulation agent that turned thousands of pages of unstructured research into 77 interactive personas, queryable on demand in place of manual focus groups.

Senior Data Engineer JAN 2024 - JAN 2026

- Orchestrated a foundational data ecosystem and governance framework for a high-profile public-sector client. Led a team of three to deploy 200+ automated ETL/ELT pipelines processing 40B+ records, migrating legacy silos into a unified BigQuery environment for real-time analytics.
- Designed and built the orchestrated ELT pipeline behind a client’s nationally consumed data product, replacing the previous process and saving \$1.4M annually in operating costs.
- Architected the Bayesian Small Area Estimator behind that product as a containerized serverless service, callable from custom Airflow operators in any Cloud Composer pipeline. Modelled geographic proximity anchoring to counter market concentration and data suppression, doubling unique actionable data series.

Data Scientist NOV 2021 - DEC 2023

- Built an end-to-end MLOps solution to improve GA4 user experiences and paid media activation, using machine learning to generate user attributes for advanced audience segmentation.

Jr. Data Analyst MAY 2021 - OCT 2021

- Migrated legacy BI ELT pipelines off Power BI’s Power Query into custom BigQuery orchestration across 10 active client dashboards, cutting refresh time from 12 hours to 10 minutes and lifting refresh uptime from 70% to 99%.

PROJECTS

Myelin / github.com/et-do/myelin

Local long-term memory MCP server: brain-inspired persistent context shared across any Agent, for a single developer or a team.

- Scores 98.2% Recall@5 on LongMemEval, beating GPT-4o-based systems using only local 22M-parameter models, with flat ~130ms retrieval latency from 100 to 1,000 memories via HNSW indexing.

no-pilot / github.com/et-do/no-pilot

Zero-trust MCP server mirroring GitHub Copilot’s built-in VS Code tools, with strict zero-trust policy enforcement.

- Enforces deny/allow policy at the MCP execution layer rather than the prompt layer, so no amount of prompt injection or natural-language persuasion widens an agent’s access.

EDUCATION

Bachelor of Cognitive Science, Neuroscience / Carleton University
2017 - 2021

Bachelor of Science, Biochemistry / Hobart and William Smith
Colleges
2017 - 2018
NCAA hockey. Transferred to Carleton.

SKILLS

LANGUAGES	Python · SQL · Go · JavaScript
CLOUD & DATA	GCP · BigQuery · Airflow · Looker · ETL/ELT · Data Quality Testing · Data Visualization
AI & ML	LLM Agents · RAG · Agentic Workflows · MLOps · Generative AI · Bayesian Modelling
STRATEGY	Enterprise Data Strategy · Data Governance & Privacy · Cloud Architecture · Responsible & Ethical AI

CERTIFICATES

Professional Data Engineer — Google Cloud SEP 2025