

Carol Calin

Senior Data Analyst (Python): pipelines, technical leadership, production data work

Prague, Czech Republic | On-site / hybrid (not remote-only)

acalincarl@gmail.com | +40 750 402 145 | carolcalin.com | github.com/rainbowpuffpuff | linkedin.com/in/carol-calin

Focus

I lead data-analysis work end to end: turn domain questions into pipelines, analyses, and productized processing. Strong in Python and pandas, SQL and structured stores, CI-minded delivery, and readable, observable code. I mentor by example (reviews, practices, clear interfaces). Based in Prague and available for Datamole's office culture (core hours friendly).

Machine and IoT-style data is a good match: hi-tech device streams, quality KPIs, and production pipelines that help customers decide and ship better products.

Relevant experience

think2earn: end-to-end data product and pipelines

Founder and lead | 2024-present | think2earn.com

- Technical lead on multi-workstream data/ML product: sensor and external data, ingestion, cleaning, training/eval, reward logic, product APIs.
- Python-first data processing (pandas-style pipelines, numpy, evaluation scripts); SQL for state; observable handoffs between stages.
- Prototyping to productization: open hardware and sensor data encryption, federated training and evaluation, production packaging for Android (Play Store closed testing).
- Mentoring collaborators and partners (including Vanderbilt Bowden Lab); clear write-ups, reviews, and coding practices under grant delivery.

chat.think2earn.com: multi-tenant data and agent context

Founder-engineer | 2025-present | chat.think2earn.com

- Built multi-tenant workspace with archives, structured project context, tool use, and orchestration so analyses and runs are reusable, not one-off notebooks.
- Production services (TypeScript/Python, REST APIs, cloud). CI habits and automated checks around shipping.

NoCap-Test: sealed multi-seed experiment pipelines

Campaign owner | 2026 | nocap.think2earn.com

- Owned sealed multi-seed GPT-2 small training and evaluation campaign: fixed environment pin, fair baseline, public CSVs, recompute rules, failure ledger.
- Production-grade experiment hygiene: efficient processing, observability of train/eval metrics, claims other people can verify.
- Candidate stack first-crossed quality gate on 3/3 seeds with about 8.3% less first-passage train time vs sealed AdamW baseline.

Netherlands Forensic Institute / Dutch Ministry of Justice and Security

Research Intern, Deepfake Detection | 2023-2024

- Fine-tuned DINOv2-style vision models on deepfake datasets; evaluation tooling for forensic analysts.
- High-stakes analysis under scrutiny from forensic vision experts and prosecutors. Human-vs-model study with statistical care.

Deep Funding / pond.xyz

Open Source Ecosystem Evaluator | 2024

- 3rd of 12,125 in a Vitalik-sponsored challenge. Ranking models and structured pipelines from sparse preference labels across 45 core repositories.

Technical leadership (role fit)

- Own a data-analysis project from domain needs to tooling choices, prototype, and productization.
- Set coding practices: readable Python, efficient pandas pipelines, tests where they pay off, code review for juniors.
- CS fundamentals in daily use: data structures, algorithms for scale, design patterns, OOP-style service boundaries.
- SQL and document/NoSQL-style stores where fit. CI via GitHub Actions-style workflows (Git-based pipelines, automated checks).
- Cloud (AWS/GCP). Monitoring mindset for pipelines; Grafana-class observability is a natural next tool on industrial telemetry projects.
- Databricks: solid data platform pattern (notebooks to jobs, lakehouse-style processing). Ready to deliver on Databricks day to day; prior work is Python/pandas/SQL on cloud and experiment stacks with the same productization discipline.

Technical skills

Core: Python, pandas, SQL, efficient and observable data-processing pipelines, OOP and design patterns, Git, CI/CD (GitHub Actions-style)

Data / platforms: structured and document stores, experiment and audit logs, AWS/GCP, cloud workers

Also: TypeScript, REST APIs, automated testing (Playwright and others), sealed evaluation frameworks, deep learning evaluation (vision transformers / DINOv2)

Education

MSc Information Studies: Data Science, University of Amsterdam, 2023

MSc Forensic Science, University of Amsterdam, 2025. DINOv2 deepfake evaluation; Dutch Ministry of Justice and Security / Forensics Institute.

BSc (Hons) PPLE (Psychology specialization), University of Amsterdam, 2021

Location

Based in **Prague**. EU (Romanian) citizen with free access to the Czech labour market. On-site / hybrid in Prague. Not seeking remote-only.

Portfolio

carolcalin.com · think2earn.com · chat.think2earn.com · nocap.think2earn.com · github.com/rainbowpuffpuff · linkedin.com/in/carol-calin