

Orpita Das

Pune, Maharashtra, India · orpita.d@gmail.com · +91 77578 52454 · linkedin.com/in/orpitadas · github.com/odas · odas.github.io

SUMMARY

Cloud Data Engineer, SnowPro Core certified (~2.5 years at Snowflake) with a life-sciences research background—and the thing I do better than almost anyone is find signal in dense information (like anomaly in a clean-looking table or the structure that should exist but doesn't). In my last job I architected real-time SOX-compliance monitoring that replaced 100+ manual audit hours a quarter; before that, cloud genomics pipelines on AWS/GCP and molecular-biology research at UBC — different domains, one instinct (analyst brain finding risks or lab results or the right architecture to ship the data in). These days I build AI-powered tools — small, rugged, useful — with Python, SQL, and modern LLMs (Claude, Gemini). The syntax is cheap now but judgment on what to build, knowing where mistakes matter is still expensive. One of those tools, Noëtone, was rated GOLD out of 3,500 submissions in the 2026 Kanz Build-with-AI Portfolio. B.Tech (NITW), M.Sc. (UBC). Canadian citizen with Indian OCI work authorization.

TECHNICAL SKILLS

Languages: SQL, Python (pandas, NumPy, Flask), JavaScript, R, Julia, Bash/Shell

Cloud & Data: AWS (Lambda, API Gateway, S3, EC2, CloudWatch, IAM, Secrets Manager), GCP, Snowflake (External Functions, Stored Procedures, Snowpipe, Streams, Tasks), Airflow, Fivetran, ETL/ELT, Data Modeling, Data Warehousing

Tools & AI: Git / GitHub / GitHub Actions, Docker, Tableau, ThoughtSpot, LLM API integration (Claude & Gemini APIs), Generative AI, Hugging Face APIs, Claude Code, ChatGPT, Copilot, Replit, NotebookLM.

Agentic AI systems: Instructable OCR/TTS pipelines, fault-tolerant retry & circuit-breaker design, quota-aware decision logic, model-fallback orchestration, SQLite-backed persistent state management.

EXPERIENCE

Independent Developer — Noëtone — Pune, India

June 2026 – Present

- Rated GOLD — 2026 Kanz "Build with AI" Portfolio, from 3,500 submissions, judged on originality, technical depth, impact and presentation (credential KANZ-ADV-1601). Published by Kanz as a third-party case study.
- Noëtone — a resilient document-to-audio pipeline (Flask + SQLite) that turns PDFs, scans, and handwritten notes into audiobooks you own as files, using Gemini's free-tier OCR as an instructable transformation layer, not just transcription; free download at github.com/odas/noetone-batch.
- Engineered resilience around a free, rate-limited API (3 requests/min, ~10 TTS calls/day/model): SQLite state registry for resumable jobs (per-page commits), error classification (quota vs. rate-limit vs. transient) with retries on the server's own suggested delays, and a quota-aware circuit breaker — with plain-English execution tracing in the UI.
- Processed a 388-page scanned novel through three power outages without losing a page; batched TTS calls for ~3× effective daily quota; every audio segment automatically validated (never assumed) before storage.
- Built solo on a repeatable loop — problem → architecture → prototype → AI-assisted implementation → review → refactor → real use. Architecture, priorities and product direction human-led; AI accelerated implementation and review. Dated review commits and a running fixes log are the record.

Founder — catadoption_pune — Pune, India

May 2024 – Present

- Founded a data-driven cat-adoption initiative on Instagram; 120+ street and Indie cats rehomed through near-daily posting and support for individual rescuers.
- Built the supporting tooling — Instagram data extraction into browsable adoption reports, a locked design system, and AI-assisted content and organization workflows (Python, Gemini, Claude).
- Same AI-tooling instinct as Noëtone above — Python + Gemini/Claude + Telegram turning an unstructured stream (json export of DMs, posts, rescuer updates) into something clean, searchable, and usable.

Cloud Engineer — Algorithmic Biologics — Bengaluru, India (remote)

Jul – Dec 2023

- Migrated rarely-accessed data from AWS EBS to S3, cutting storage cost ~50% and enabling the team to pull from S3 during live EC2 runs — an access pattern they lacked before.
- Built a reusable EC2 spin-up/teardown toolkit to run the team's genomics jobs (samtools region-extraction, BAM-to-Julia handoff, VCF processing) as repeatable runbooks.
- Deployed a proprietary gene-expression analysis app to AWS; set up CloudWatch monitoring.

Data Engineer — Snowflake — Pune, India

Oct 2020 – Mar 2023

- Architected real-time, SQL-based SOX-compliance monitoring (access management, SSO/SCIM coverage, password controls) with exception dashboards and alerting — saving Internal Audit and IT admins 100+ manual audit hours per quarter.
- Built and maintained 20+ ELT ingestion pipelines (ServiceNow, Okta, GitHub GraphQL, Salesforce, DocuSign and more) using Snowflake External Functions, Stored Procedures, Tasks, Snowpipe, Streams, AWS Lambda, API Gateway, and Secrets Manager; also Fivetran and Airflow.
- Co-proposed a GitHub–Snowflake DataShare architecture with GitHub leaders and Snowflake Sales Engineers.
- Delivered 7 external customer presentations on the SOX automation work, including to the Institute of Internal Auditors (IIA), San Jose chapter, alongside Snowflake's Director of Internal Audit; trained new team members. Among the first employees at the Pune office. SnowPro Core certified.

Machine Learning Research Intern — University of Waterloo — Waterloo, Canada

Apr – Jul 2020

- Compiled a labeled protein dataset from NCBI and BacDive; prototyped an ensemble model with Bayesian hyperparameter tuning to predict protein structure–activity relationships (QSAR).

Data Contract — BugSeq / Various — Toronto, Canada

Jan – Mar 2020

- Ingested and queried a SQLite database from the NCBI Taxonomy FTP server to validate analytics for a microbial taxonomy classification model (bugseq.com).

Earlier — Research Assistant & Biological Technician — Vancouver, Canada

2016 – 2018

- StemShock (acq. Terramera): molecular-biology experiments; established a plant tissue-culture facility. Core Medical Centre: mental-health literature review and editing.

EDUCATION

M.Sc., Plant Molecular Biology — University of British Columbia (degree conferred in Botany) — Canada

2012 – 2015

- Best Student Talk, Banff International Conference on Plant Metabolism (2014) — only M.Sc. among PhD speakers. NSERC-funded research.

B.Tech, Biotechnology — National Institute of Technology, Warangal (NITW) — India

2008 – 2012

- Top 1% AIEEE national rank (2008). MITACS-funded research intern, McGill University (2011). Merit scholarship in final year (2012).

CERTIFICATIONS & AWARDS

GOLD — Kanz "Build with AI" Portfolio (2026, credential KANZ-ADV-1601) · SnowPro Core (Snowflake, 2023–2025) · Data Analytics & Data Science (Lantern Institute, Toronto) · Snowflake GenAI Professional (in progress)