

# William Chen

Software Engineer | Data Platform Engineer

Jersey City, NJ 07302 | [wchen1396@gmail.com](mailto:wchen1396@gmail.com) | [willchennn.com](http://willchennn.com) | [github.com/smallchungus](https://github.com/smallchungus)

U.S. Citizen - No Visa Sponsorship Required | Public Trust Suitability

## EDUCATION

---

### Georgia Institute of Technology

Master of Science in Analytics (Data Science), GPA: 4.00

Atlanta, GA

January 2026 – Present

### Rutgers University

Bachelor of Arts in Computer Science and Psychology, GPA: 3.44

New Brunswick, NJ

Graduated May 2022

## EXPERIENCE

---

### Viatrie

Data Platform Engineer (Federal Contract)

Arlington, VA

Apr 2024 – Mar 2026

- **Pipeline Platform:** Built the reusable pipeline platform for a **USDA** federal contract, shipping **50+ end-to-end ETL pipelines** from **S3** to **Amazon Redshift** in **Python** and **AWS Glue**; owned **150+ pipelines** across the team and cut new feed onboarding time by **50%** through reusable templates other engineers built on top of
- **Runtime Wins:** Killed recurring **Glue** job timeout failures by re-architecting multi-step workflows with **AWS Lambda** and **Step Functions**, partitioned loads, and tuned DPU allocation, cutting end-to-end pipeline runtime by **60%**
- **Delivery Infrastructure:** Built the deployment system engineers used daily with **GitHub Actions** CI/CD, **Docker** container builds, and **Terraform**-managed AWS resources across **200+ tables** in **3 environments**, taking deploys from **~1 hour** of manual work to **~10 minutes** and eliminating unplanned schema rollbacks
- **Catching Bad Data Early:** Wrote data quality checks into **AWS Glue** jobs that fail-fast on whitespace, type mismatches, and malformed records; automated schema drift detection caught **3 upstream schema changes** over **6 months** before any bad data reached **USDA** analyst dashboards
- **Platform Observability:** Instrumented **150+ pipelines** in **Datadog** with SLA tracking, volume anomaly detection, and on-call alerting, giving the team a single pane of glass into pipeline health instead of grepping through Glue logs
- **Unblocking Analysts:** Shipped self-serve **Tableau** dashboards and written runbooks serving **100+ users** at **USDA**, powering congressional reports and internal budget reviews, and saving the team **~2 hours/week** of ad-hoc report pulls

### Rutgers Chlamydia Lab

Research Software Engineer

New Brunswick, NJ

August 2022 – Present

- **Protein Data Analysis:** Lifted unknown protein classification accuracy from **~70%** to **89%** by training **scikit-learn** ensemble models on **mass spectrometry** data, serving **10-30 researchers, scientists, and doctors** across **10+ collaborating labs** in bacterial sample analysis
- **Automation & Reliability:** Cut **2+ hours** of daily manual work by scripting instrument data processing in **Python**; kept the **5TB** research database intact with automated weekly backups and upload validation that catches corrupted writes before downstream analysis

### Isaac Lab (NVIDIA)

Open Source Contributor

Remote

January 2024 – Present

- **Open Source Contribution:** Contribute to **Isaac Lab**, NVIDIA's open-source **GPU-accelerated** framework for **reinforcement learning** and **imitation learning** across multiple robot embodiments, building modular **Python** simulation environments for manipulator and **AMR** tasks in **PhysX**

## PROJECTS

---

### Distributed Task Queue | Go, Redis, PostgreSQL, Docker, Kubernetes, GitHub Actions

- **Job Engine:** Built a distributed task queue in **Go** using **Redis** lists with **BLPOP** for blocking dequeue, processing **5K jobs across 4 workers** in **~10 seconds**; wrote the worker and scheduler from scratch instead of pulling in **Celery** to learn the internals of Go concurrency and Redis
- **Failure Handling:** Handled worker crashes with a **5-second heartbeat** written to **Redis**; jobs from workers that go silent for **30 seconds** get requeued by a sweeper, giving at-least-once delivery without a separate broker, plus **PostgreSQL**-backed status history and exponential-backoff retries
- **Ship Path:** Containerized with **Docker** and deployed to **Kubernetes** with separate deployments for API, workers, and databases; horizontal pod autoscaling on queue depth and automated build/test/rollout through **GitHub Actions**

## TECHNICAL SKILLS

---

**Languages & Databases:** Python, Go, SQL, JavaScript, TypeScript, PostgreSQL, Redis, Amazon Redshift

**Cloud & Infrastructure:** AWS (Glue, Lambda, Step Functions, S3, Redshift), Docker, Kubernetes, GitHub Actions, Datadog, Bash

**Data Platform & Infrastructure:** CI/CD pipelines, distributed systems, REST APIs, observability and alerting, pipeline orchestration, Terraform, Infrastructure as Code, AI-assisted development (Claude, GitHub Copilot), technical documentation