Shaun Porwal

Metuchen, NJ shaun.porwal@gmail.com shaunporwal.com github.com/shaunporwal cheminventory.co

EXPERIENCE

Founder & Engineer — Junto Technologies **Mar 2026 – Present** · Metuchen, NJ

- Built a full-stack (design, client, server, database, authentication) **chemical inventory management system** (cheminventory.co) that tracks **$1M+** in yearly chemical order volume with regulatory document management and version-controlled provenance
- Built a self-hosted **multi-agent AI system** (local LLMs via vLLM/llama.cpp, scheduled runs, evals) for **market research** and **competitive analysis**; delivered a full ICP/GTM report and verified competitor benchmarks
- Standardizing processes at a specialty chemicals company to achieve **ISO certification**
- Served as Mandarin/English translator at **CPHI Shanghai 2026** for a specialty chemicals company

Founding Engineer — Standard Model Biomedicine **May 2025 – Mar 2026** · Philadelphia, PA

- Architected a full-stack **agent orchestration platform** (FastAPI/Next.js, WebSockets) for biomedical workflows, integrating OpenAI Agents SDK tooling and DICOM/NIfTI imaging for tumor analysis
- Built a reusable research library (`smb-pipe`) for **survival-prediction pipelines**, automating CT embedding extraction and MEDS-to-model workflows for aortic aneurysm and heart failure risk modeling
- Developed an **LLM benchmarking framework** for **clinical trial matching**, implementing vLLM evaluation runners and custom embedding classifiers (MLP / logistic regression heads) on EHRSHOT data
- Secured a strategic **data-for-equity partnership** with Memorial Sloan Kettering Cancer Center, facilitating access to high-fidelity multimodal datasets for foundation model development

Machine Learning Engineer / Data Analyst — MSKCC **Mar 2021 – May 2025** · New York, NY

- Engineered SQL-sourced data pipelines and statistical code for Amplio, enabling sarcoma, melanoma, Whipple, gastrectomy, and liver surgeons to **benchmark patient outcomes** against peers
- Built an **AI radiology pipeline** (Llama 3.1, SAM image segmentation, RShiny dashboard) to assist radiologists
- Led statistical analyses (GAM, logistic regression, Kaplan-Meier, CoxPH, Decision Curve Analysis) for a landmark study on **post-chemotherapy RPLND policy** in testicular cancer, published in Annals of Oncology; contributed statistics to additional publications on prostate cancer and robot-assisted surgery outcomes
- Developed dcurves, a Python package for **Decision Curve Analysis** (**80k+** PyPI downloads); built and maintain the instructional site decisioncurveanalysis.org
- Engineered a Llama 3.1:70B + RAG **text-to-SQL tool** (`no-more-sql`) with Streamlit and FAISS for natural-language querying of MSKCC's SQL databases
- Completed a large **patient-comorbidity analysis** using take-home survey data, with `pointblank` data validation, to demonstrate accurate billing-level assignment

Bioinformatics Intern — Sema4 **Feb 2020 – Feb 2021** · Stamford, CT

- Engineered automated Bash/Python/WDL/AWS pipelines to detect **structural variants** across 100+ 100GB+ BAM files using Delly, SvABA, Manta, and Illumina DRAGEN
- Algorithmically discovered **repeat variants** among 100+ samples, comparing insulin- vs. non-insulin-producing pancreatic neuroendocrine tumors for diabetes therapy research
- Visualized results with Circos plots, stacked bar plots, heat maps, and scatter plots

Inorganic Manufacturing Associate — SPEX CertiPrep **May 2018 – May 2019** · Metuchen, NJ

- Performed inorganic standards analysis via wet-lab techniques (pipetting, gravimetry, titration, filtration, combustion) and ICP-MS instrumentation for reference-standard preparation
- Programmed and calibrated a Hudson automated pipettor robot (SOLOsoft) for serial dilutions and standard synthesis

Research Intern — Lab of Prof. Jiun-Hong Chen, National Taiwan University **Aug 2017** · Taipei, Taiwan

- Conducted regeneration research **entirely in Mandarin Chinese**; studied gene expression in the freshwater annelid *Aeolosoma viride* using **whole-mount in situ hybridization** and presented hybridization results and survival rates

Quality Control Intern — Tokyo Chemical Industry Co., Ltd **Jun 2016 – Jul 2016** · Saitama, Japan

- Trained 160 hours in instrumentation (GC, HPLC, NMR, IR) and wet-lab QC analysis
- Assisted with Japanese/English translation for the team

SELECTED PROJECTS

dcurves pypi.org/project/dcurves · GitHub

Python package for Decision Curve Analysis, evaluating the clinical value of prediction models and diagnostic tests; **80k+** PyPI downloads and used in peer-reviewed research.

no-more-sql GitHub

Llama 3.1:70B + RAG pipeline (Streamlit, FAISS) for natural-language querying of SQL databases without writing SQL.

cheminventory.co cheminventory.co

Full-stack chemical inventory management and regulatory compliance platform built for Junto Technologies.

EDUCATION

M.S., Biomedical Data Science **Feb 2021** · New York, NY
Icahn School of Medicine at Mount Sinai

B.S., Biomedical Engineering & Chinese **May 2019** · New Brunswick, NJ
Rutgers University–New Brunswick

Certificate in Advanced Mandarin Chinese **Aug 2015** · Taipei, Taiwan
National Taiwan University
Huayu Enrichment Scholarship, Taiwan Ministry of Education (MOE)

SKILLS & CERTIFICATIONS

Programming	Python, R, TypeScript, SQL, Bash, HTML/CSS, Docker, Git
AI / ML	LLM agent orchestration, multi-agent/swarm systems, local LLM deployment (vLLM, llama.cpp), self-hosted GPU/edge infra (Tailscale), LLM evals (promptfoo, Argilla), autonomous scheduling, RAG, fine-tuning, image segmentation, embedding classifiers
Statistics	GAM, CoxPH, Kaplan-Meier, logistic regression, Decision Curve Analysis, survival analysis
Domain	Chemical inventory and regulatory documentation (ISO certification in progress), bioinformatics (SV detection, RNA-seq)
Languages	English (native); Mandarin Chinese, Hindi (fluent); Japanese, Spanish (conversational)
Certifications	Epic – Clinical, Epic – Reporting, Advanced Mandarin Chinese (NTU)