Jonathan Tsen

LinkedIn | Portfolio | dudujonathantsen@gmail.com | +55 (14) 98107-9420

EXPERIENCE

Labrynth | *Senior AI Engineer* — California, USA (Remote) | May 2025 – Present

- Built the engine that checks every citation and engineering value in 500-page nuclear licensing documents against 150 GB of NRC, DOE, and CFR sources; shipped the beta a week early, and the client estimated **~50%** less review time.
- Turned a small paid pilot into the client's licensing contract with an engineering assistant (GraphRAG on Neo4j with page citations) that the client's engineer rated **90–95% correct**.
- Cut LLM cost per building-code compliance check by **95%** and run time by **~85%** (35 → 5 min): sent ~90% less of the code to the model, cached the PDF, shrank outputs with TOON, and routed chapters to benchmarked models.
- Reached **87% accuracy** against the City of Lancaster, CA plan checker's own labels in an AI plan-check pilot.
- Made the platform export-control compliant on AWS (Bedrock, KMS, CloudTrail, US-only WAF, Cognito MFA) in Terraform with CI/CD, and fixed the root cause of a login incident where **28%** of sign-ins failed.
- Built a 9-agent GraphRAG workflow that pre-drafts nuclear environmental report sections: internal QA score **74 → 94**, sources cited more than doubled, and the client graded the first section B+/A-.
- First engineer: built the Red Tape Index permit ranking (**334 cities**, 40 states), onboarded early hires, and ran the live demo at a White House permitting expo (**1 of 50** companies picked).

CEIA – Center of Excellence in AI | *Generative AI Research Team Lead* — Contract | Brazil (Remote) | Aug 2023 – May 2025

- Promoted from AI Researcher to **Team Lead (6 people)** on EnergyGPT, a generative AI platform for Brazil's power sector.
- Fine-tuned Llama 3.1 8B for health-plan approvals: accuracy **66% → 82%**, beating 8 off-the-shelf LLMs incl. GPT-4o.
- Scaled LLM training-data collection from 10 to **~150 universities** via Scopus: ~300,000 articles in 3 weeks.

Advent International | *Data Science Associate* — New York, USA (Remote) | Jan 2023 – Jun 2023

- Built the analyses behind **three buy/sell decisions**, incl. a review study (5% → 51% share) that backed an acquisition.
- Halved broken bill scrapers (**16% → 8%**) at a portfolio company that collects 22M bills a year.
- Automated LinkedIn outreach that filled **~60%** of a portfolio company's hard-to-fill tech roles (146 hires) in 2 months.

Zyte | *Web Scraper* — Cork, Ireland (Remote) | Jul 2022 – Jan 2023

- Built and maintained production Scrapy crawlers, using Sentry and Spidermon to monitor them and triage failures.

University of Chicago | *Summer Research Assistant* — Summer program | Chicago, USA | May 2021 – Aug 2021

- Selected as one of **120 students** for the CDAC Summer Lab; worked on the Chameleon Cloud team with Argonne and FIU.
- Trained a YOLOv4 fish counter (**78% accuracy**); an edge Jetson Nano was **19×** more cost-effective than cloud GPUs.

Florida International University | *Project Research Assistant* — Contract | Florida, USA | Feb 2021 – Oct 2021

- Trained a satellite-image classifier (mangrove, river, boundary) to **90% accuracy** for robotic mangrove monitoring.

EDUCATION

Instituto Tecnológico de Aeronáutica – ITA | *Master's in Data Science* — Brazil | Jan 2022 – Expected Dec 2026

Thesis: Hierarchical Classification of Missing Data Mechanisms: A Statistical Feature Engineering Approach with LLM Augmentation and Real-World Validation

PROJECTS AND PUBLICATIONS

SBPO 2024 | *Operations Research Conference Paper* — Brazil | Oct 2024

- B. R. M. Nacano, J. Mika, and J. M. Tsen, "Comparative Analysis of the Results of the Genetic Algorithm Metaheuristic and the Slime Mold Algorithm Metaheuristic in School Transportation Routing Optimization," SBPO 2024.

Supercomputing SC21 | *ACM Student Research Competition Poster* — St. Louis, USA | Nov 2021

- J. Tsen, "One Fish, Two Fish: Choosing Optimal Edge Topologies for Real-Time Autonomous Fish Surveys," SC21 poster.

SKILLS

AI and LLMs	Pydantic AI, LangChain, LlamaIndex, Hugging Face, Amazon Bedrock, RAG / GraphRAG, AI agents, NLP
ML and Data	Python, SQL, PyTorch, TensorFlow, scikit-learn, pandas, OpenCV, Neo4j, Supabase, Power BI, Tableau
Cloud and Backend	AWS (Lambda, Step Functions, S3, Cognito), GCP (Cloud Run), Terraform, Pulumi, Docker, FastAPI, Django
Languages	English (fluent), Portuguese (fluent), Spanish (basic)