

Nivaldo Humberto Oliveira Tokuda

AI Engineer | Agents, evaluation and simulation | Python | Google Cloud

São Paulo, Brazil | ntokuda@utexas.edu | portfolio-818.pages.dev | github.com/nivaldoh | [LinkedIn](#)

SUMMARY

Software and AI engineer with 8+ years building cloud and data systems in Python and Google Cloud. Since 2025 I have worked on AI agents: game and robotics environments for language models, evaluation and replay tools, and imitation and reinforcement learning experiments. Quarterfinalist in two formats of the NeurIPS 2025 PokéAgent Challenge. Contributor to Metamon at UT Austin's Robot Perception and Learning Lab. MS in Computer Science at UT Austin, expected December 2027.

AGENTS AND RESEARCH ENGINEERING

Language-model agents in Total War and Slay the Spire | Independent, [case study](#) 2026

- Built a Lua mod for Total War: WARHAMMER III that publishes battle state (enemy units only while visible) and carries out orders, with Python controllers that let LLMs command armies in real time. No screen reading or input automation.
- Designed a six-scenario benchmark with matched model comparisons and starting conditions checked in the game lobby. Built an offline report and an interactive replay that rebuild each battle from recorded observations, orders and model replies.
- Built an LLM player for full Slay the Spire runs through CommunicationMod (furthest: Act 4, floor 54) and traced failures to the prompt, memory or interface. Also ran small first-combat pilots with PPO and a LoRA fine-tune of Qwen3-0.6B.

RoboCasa365: robot evaluation under changing goals | Independent Sep 2026

- Extended the RoboCasa365 kitchen simulation so the goal or scene can change mid-episode, with current-goal scoring and separate records of each change, when each controller received it, and what was completed.
 - Evaluated GR00T N1.5 and Astra, my LLM agent that writes robot control code with ground-truth state. In 4 goal-change episodes GR00T always completed the original task, but never the new goal. Astra met the new goal in 1 of 5.
- [Workshop paper](#) submitted to the CoRL 2026 Workshop on Agentic Robotics.

Pokémon TCG agents | Kaggle competition Jun – Jul 2026

- Built a Gymnasium environment around the supplied simulator with legal-action masking and multi-card choices as sequences of masked picks. Trained a custom PyTorch policy by behavior cloning from scripted players (286,035 decisions), then fine-tuned it with MaskablePPO.
- Cloning, then PPO, reached a 24.0% pooled win rate against the scripted teachers, versus 13.9% for PPO alone with the same network. Audited a single-opponent specialist's 82.6% win rate and found every win came from the opponent's deck running out. [Case study](#)

Metamon | with UT Austin's Robot Perception and Learning Lab Jan – Mar 2026

- Cut team validation from about 1 s to 1 ms per team (re-timed on 200 teams), so a 25,192-team set would take under a minute instead of about seven hours. One persistent validator process replaced a new Pokémon Showdown process, and a Showdown rebuild, for every team file ([PR #65](#), merged and still on main).
- Added Smogon-rank filtering to usage statistics and exponential backoff with a concurrency cap to the Smogon scraper. Prototyped an encoder rewrite that runs under torch.compile CUDA graphs, with a profiler that accepts a change only if outputs match and latency does not regress.

PokéAgent Challenge, NeurIPS 2025 | team MetaHorns Oct – Nov 2025

- Quarterfinalist in both the Gen 1 OU and Gen 9 OU tournaments (eight entrants each). The Gen 9 entry put Metamon's offline-RL policy and critic inside PokéChamp's one-turn minimax search, with no LLM call per turn.
 - Built the bridge that mirrors each live battle into Metamon's trajectory format and copies that history into every simulated branch. Packaged Metamon self-play as a container that runs as a Cloud Run Job, one Showdown server per task.
- [Case study](#)

PROFESSIONAL EXPERIENCE

- Globant** | São Paulo, Brazil Jul 2024 – May 2026
- Data Scientist, Paramount / CBS** Oct – Dec 2025
- Built multimodal video search with Vertex AI embeddings, retrieving news and sports segments from text and image queries. Video processing ran as Cloud Run components orchestrated by Workflows.
- Data Scientist, Warner Bros.** Jul – Sep 2025
- Built an LLM translation pipeline for domain-specific vocabulary on Cloud Run. Compared RAG with fine-tuning using BLEU and COMET to choose the approach.
- Data Engineer, Google Extended Workforce** Jul 2024 – Jul 2025
- Owned the Airflow workstream of an AWS-to-GCP data warehouse migration (100+ DAGs) for a US fraud-detection company. Built low-latency trading data pipelines for a US financial institution, and Firestore backups for a government project.
- Globo** | Senior Data Engineer, São Paulo Jan 2022 – Jan 2024
- Led the creation of a data warehouse that fully replaced the previous system, cutting query times by hours and nearly halving costs. Transformations orchestrated with Airflow on GKE.
 - Built an App Engine application that data scientists used to launch Dataproc jobs.
- DP6** | Data Engineer, São Paulo May 2017 – Dec 2021
- Built data pipelines for marketing analytics at some of Brazil's largest companies. Worked with a Google team on Megalista (Python, Apache Beam, Dataflow).

OPEN SOURCE

- Apache Beam Rust SDK** Nov 2022 – Jan 2023
- Started an experimental Rust SDK for Apache Beam: portability protos, worker bundle processing, an async DirectRunner, the first coders and Impulse. When Google's Dataflow team began its own Rust SDK in January 2023, the [upstream discussion](#) recommended continuing from my repository. A community fork carried it on, as traced in a [Beam Summit 2023 talk](#).
- Apache Beam (Python)** | LogElements PTransform, [PR #23879](#) 2022
- Google Megalista** | Firestore configuration source for its Dataflow pipeline, [PR #12](#) 2021

SKILLS

Agents and ML: LLM APIs, agent evaluation, behavior cloning, PPO (Stable-Baselines3, MaskablePPO), Gymnasium, PyTorch, torch.compile profiling, LoRA fine-tuning, RAG.

Languages: Python, SQL, Rust, Lua.

Cloud and data: Google Cloud (Cloud Run and Jobs, Workflows, GKE, App Engine, Cloud Functions, Vertex AI), BigQuery, Firestore, Airflow, Apache Beam and Dataflow, Spark and Dataproc, Docker.

Languages spoken: Portuguese (native), English (advanced, C1).

EDUCATION

- University of Texas at Austin** | MS, Computer Science (online) Aug 2024 – Dec 2027, expected
Coursework: Reinforcement Learning, Deep Learning, NLP, Parallel Systems.
- University of Illinois Urbana-Champaign** | Master of Computer Science studies, transferred to UT Austin May – Jul 2024
- Universidade Presbiteriana Mackenzie** | BS, Computer Science Jul 2017 – Jul 2022
Teaching assistant, Computer Programming I (C++).