Austin Senna Wijaya

646-295-5341 | asw2215@columbia.edu | linkedin.com/in/austin-senna | github.com/Austin-Senna | austin-senna.github.io

Education

Columbia University New York, NY

Bachelor of Science: Computer Science *Aug 2024 – May 2028*

- GPA: 4.22/4.0; Full-Tuition Scholarship; Dean's List.
- Research interests: data systems, AI agents, retrieval systems.
- Coursework: Natural Language Processing, Data Structures & Algorithms, Systems Programming, Linear Algebra & Probability.

Research Experience

Data Agents Research Assistant Dec 2025 – Present

Columbia Data, Agents, and Processes Lab *New York, NY*

- Engineered a high-throughput benchmark evaluation harness in **Strands SDK**, utilizing parallel sandboxed **DuckDB/Python** runtimes for multi-agent question-answering pipelines incorporating skills, hooks, and context compaction.
- Processed **9.5 TB / 40M docs** of open-source tabular and textual data into an S3 data lake; built scalable BM25 and hybrid retrieval APIs utilizing **LanceDB** to support massive-scale source discovery for agents.
- Architected telemetry to capture tool-call traces and track agent reasoning to identify bottlenecks in agent framework, improving **GPT-5-mini** semantic match from **2.22%** to **56.3%**.

Machine Learning Research Assistant Dec 2025 – Jun 2026

Columbia Zuckerman Institute *New York, NY*

- Compressed **Qwen2-Audio 4096-D** hidden state embeddings into **10-100D** feature-specific representations through bottleneck FNN training in **HuggingFace** and **PyTorch**, each isolating a single linguistic feature to map onto cortical brain regions.
- Curated training-ready audio datasets from 12 speech and language benchmarks totalling 120K+ audios, including converting text benchmarks into audio through parallel **Kokoro-TTS** generation.
- Built and parallelized a GPU-optimized **PyTorch** pipeline for Qwen2-Audio inference, hidden-state extraction, and supervised classification training under cross-validation across 12 speech attributes and 120K+ audios.

Publications

[1] *LakeQA: A Benchmark for Complex Exploratory QA over a Million-Scale Data Lake,* ***ICML 2026***: Haonan Wang, Jiaxiang Liu, Yurong Liu, **Austin Senna Wijaya**, Tianle Zhou, Eden Wu, Yijia Chen, Wanting You, Reya Vir, Daniela Pinto Veizaga, Grace Fan, Yusen Zhang, Juliana Freire, Eugene Wu.

[2] *SANA: What Matters for QA Agents over Massive Data Lakes?* ***VLDB DASHSys 2026:*** **Austin Senna Wijaya**, Jiaxiang Liu, Haonan Wang, Eugene Wu.

[3] *MINA: Linear Probes Reveal Coding-Sequence Family Signal in Frozen DNA Encoders,* ***iscbAI'26***: **Austin Senna Wijaya**, Hayden Leung.

[4] *Targeted Embeddings: An Interpretable Framework for Neural Speech Encoding with Large Audio-Language Models,* ***NeurIPS 2026***: Linyang He, Ciro D. Randazzo, **Austin Senna Wijaya**, Gabriel Diaz, Yunyan Ouyang, Jingting Zhang, Jonathan Brennan, Nima Mesgarani. *(Under-review).*

[5] *Be CARE-ful with Text-to-SQL Benchmarks,* ***NeurIPS 2026***: Haonan Wang, Jiaxiang Liu, Elaine Ang, Tiancheng Ge, Wanting You, **Austin Senna Wijaya**, Reya Vir, Tianle Zhou, Mihir Agarwal, Yusen Zhang, Eugene Wu. *(Under-review).*

Industry Experience

Software Engineer Intern Jun 2026 – Present

Garner Health *New York, NY*

- Engineered a full-stack observability pipeline, automating **Grafana** provisioning through **Terraform**, unifying frontend/backend traces via **Datadog RUM** and **OpenTelemetry (Loki, Prometheus, Tempo)** for critical insurance workflows.
- Architected a **Kubernetes**-deployed, self-improving autonomous on-call triage agent with **Claude Agent SDK**, ingesting telemetry to perform root-cause analysis and create verifiable fix pull requests; resolved 10+ production incidents while enforcing privacy guardrails.

Data Engineering Intern Jun 2025 – Aug 2025

Ruangguru (Indonesia's largest edtech platform) *Jakarta, Indonesia*

- Engineered scalable ETL pipelines to aggregate millions of asynchronous payment events into structured session data using **BigQuery**, producing a product funnel to identify major checkout drop-off points.
- Architected an automated data scraping, ingestion, and sentiment analysis service utilizing **Puppeteer** and **Hugging Face** APIs, integrating the pipeline into live operational monitoring dashboards in **Looker Studio**.

Projects

ResearcherX: Local-First AI Research IDE Website

- Spearheaded ResearcherX, a local-first AI research editor with **Electron/Next.js/FastAPI** stack and graph-based RAG in **LanceDB** and **NetworkX** to produce cited answers and detect hallucinations or contradictions in AI writing.
- Engineered a real-time contradiction-linting engine that minimizes overhead via batched LLM calls and paragraph-hash caching, utilizing **Pydantic** to strictly validate conflicts surfaced through source-linked editor squiggles.

MINA: Model Interrogation of Nucleotide Architectures Preprint

- Engineered an end-to-end probing benchmark for frozen DNA foundation models, evaluating linear probe recovery of **5-way** protein-family labels and **1,536-D** GenePT text embeddings across **3,244** human genes.
- Bounded DNA encoder signal against rigorous **k-mer**, protein-composition, and **ESM-2** baselines, utilizing **MMseqs2** homology-aware splitting to eliminate paralog leakage and expose a **0.13 macro-F1** performance inflation in standard random splits.

Technical Skills

- **Languages**: Python, SQL, C++, Java, JavaScript/TypeScript
- **Libraries & Frameworks**: Strands Agents, Langchain, Pandas, NumPy, PyTorch, Puppeteer/Selenium, BeautifulSoup, React
- **Cloud & Databases**: AWS (EC2, S3, Bedrock), Google Cloud (BigQuery, Looker Studio), Supabase, PostgreSQL, MongoDB
- **Tools**: GitHub, Claude Code, Codex, Docker, Copilot, Arduino, Notion

Honors

- **Databricks Grounded Reasoning Cup 2026**: Columbia representative; worked with Google.
- **Venture Competitions 2026**: 2nd Place at Columbia Lion Cage; Top 6 at Millard Chan Technology Startup Competition.
- **Hack@Brown 2026**: Finalist; Strongest Product Thinking award; worked with Google Ventures and Partiful.
- **Jane Street Mystery Planet Trading & Puzzle Competition 2026**: 1st Place.
- **Tencent Kepler Plan S3 Competition 2025**: 2nd Place; Best Code; Best Product Idea; Popularity Award.
- **Indonesia Maju Scholarship**: Full-ride scholarship recipient, 1 of 350 recipients.
- **55th International Chemistry Olympiad 2023**: Final stage, 10th rank for Indonesian national team selection.