Diego Sanchez Tirado

Stanford, CA | (984) 895-6386 | dsanchezt22@gmail.com | linkedin.com/in/diego-sanchez-tirado | github.com/dsanchezt22

Education

Stanford University — Stanford, CA
B.S. Electrical Engineering — *September 2025 – May 2029 (expected)*
Coursework: Signals & Systems I, The Modern Software Developer, EE Fundamentals, Programming Abstractions, Probability for Computer Scientists, Vector Calculus for Engineers, Ordinary Differential Equations for Engineers

Experience

Amazon Lab126 — Sunnyvale, CA
Hardware Development Engineer Intern | RAVEN / Governed Knowledge Engine — *June 2026 – September 2026*
- Delivered an engineering-knowledge engine adopted by 3 teams, ingesting 1,500+ items daily from Outlook, Slack, Confluence at <5% LLM-as-judge error; migrated from a Karpathy-style knowledge base to a gated S3 vector retrieval-augmented generation (RAG) store
- Architected prompt-injection and personally identifiable information (PII) gates before RAG indexing; raised leak-detection recall from 0.78 to 0.95 at <3% false positives, with F1 above 96% across every gate on a sealed adversarial benchmark spanning 10 leak classes

Amazon Lab126 — Product Integrity & Reliability Lab — Sunnyvale, CA
Robotics Apprentice | Dexmate bimanual humanoid — *June 2026 – September 2026*
- Designed and prototyped a ball-joint end-effector fixture cutting test reconfiguration to one adjustment; trained Action Chunking with Transformers (ACT) and Diffusion Policy models for varied-geometry box-pickup tasks, then added residual reinforcement learning

Verdnt Group — Stanford, CA
Founder & Lead Engineer — *April 2026 – Present*
- Built FStea AI, an agentic R&D engine for China's beverage market processing 50,000+ signals daily from 9 collectors. An agent with 14 specialized tools and a relevance-ranked knowledge base of 200,000+ documents pulls evidence for each prediction, and a checkpointed 3-model LLM chain turns it into manufacturable product specs in under 6 minutes, ranking 120+ scored concepts weekly
- Developed Spatial Realization, which turns Airbnb, hotel, real-estate listing photos into Three.js dollhouse, walk-through, and floor-plan views using Qwen2.5-VL, Qwen3, and TripoSR image-to-3D meshes, with furniture placed at true scale by description

Code Four (YC X25) — San Francisco, CA
GTM Lead — *July 2025 – October 2025*
- Automated outreach to 10,000+ police contacts using AI agents and scrapers; led intern campaigns across 1,100+ departments, securing 70+ qualified meetings, 10+ signed memoranda of understanding (MOUs), and 520+ officers onboarded in four months nationwide

Research

Stanford Robust Systems Group — Prof. Subhasish Mitra — Stanford, CA
Undergraduate Researcher | Area-Scalable Memory for LLM Inference (with Sam Dayo) — *January 2026 – June 2026*
- Modeled a 127× throughput advantage for area-scalable mesh memory over 4-stack High Bandwidth Memory (HBM4) at a 2.1M-token context window (3,708 vs. 29 tokens/s), where HBM4 decode is bandwidth-bound while mesh bandwidth density scales with die area
- Simulated LLM-inference throughput, latency, and memory footprint across on-chip memory organizations; related capacity, bandwidth density, and mesh/network-on-chip (NoC) topology to batch size and context length to identify bandwidth-bound regimes at long context
- Benchmarked mesh-distributed memory against HBM under tensor/pipeline parallelism; produced compute, memory, and interconnect breakdowns identifying where long-context key-value (KV) cache traffic saturates each backend and where compute dominates instead

Duke University — Dr. Charles Parker, Prof. Arturo F. Ehuan — Durham, NC
Research Assistant | Nanomaterials & Thin Films Lab; Cyber Threat Intelligence — *June 2023 – July 2025*
- Characterized 1–5 nm atomic-layer-deposited (ALD) Al_2O_3 coatings on carbon-nanotube (CNT) emitters: swept current vs. field at 10^{-6}–10^{-8} Torr, fit Fowler–Nordheim curves per SEM emission area, isolating a ∼50 V turn-on shift, no added degradation
- Co-authored *LATAM Financial Threat Landscape 2025* (second author) with Recorded Future; mapped five threat actors to MITRE ATT&CK techniques via a Git-tracked IOC/TTP correlation pipeline; presented findings to financial-sector CISOs in Mexico City

Projects

Custom Electric Bike Conversion — *Independent* — 2026
- Converted a mountain bike with a spot-welded 13S4P (∼48 V, ∼0.8 kWh) Molicel P42A pack, copper-nickel strip, 80 A battery management system (BMS), and Bafang BBSHD mid-drive; charged, balanced, and load-tested the pack before commuting to Amazon

4-Bit Adder/Subtractor with BCD Readout — *Digital Logic, NCSSM* — 2024
- Implemented a 4-bit adder/subtractor at the gate level: a ripple-carry chain of full adders composed from half-adders, two's-complement subtraction via XOR-inverted operand with injected carry, and a binary-coded-decimal (BCD) converter driving two 7-segment displays

ErgoTouch — Split Ergonomic Keyboard — *Bowman Brockman Grant, $500* — 2023 – 2024
- Engineered a split, tented keyboard for arthritis and carpal-tunnel users: CAD resin-printed case, custom KiCad PCB, hand-soldered switch matrix, and RP2040 firmware with switches chosen by actuation force and travel; fourth keyboard built, first fully custom design

Piezoelectric Typing-Force Energy Harvester — *NC Science & Engineering Fair* — 2024
- Assembled a piezoelectric rig (force plate, rectifier, capacitor); swept force (2.5 ± 1 N) and diameter (10/41 mm) over 360 actuations

Skills & Awards

Hardware: Circuit design, LTspice, KiCad PCB layout, CAD design, gate-level logic, soldering, high vacuum, CNC, 3D printing
Architecture: HBM, NoC, TP/PP performance modeling, KV-cache analysis, batch and context-length scaling, gate-level digital logic
Programming & AI: Python, C++, MATLAB, Git; RAG architecture, agent orchestration, Claude Code, Codex; fluent English/Spanish
Awards: Congressional App Challenge 1st; Intel AI Global Impact Finalist; Bowman Brockman Grants ($2,500); FTC 5th/390