

Sritej Bommaraju

480-233-6330 | sritejbommaraju@berkeley.edu | [linkedin.com/in/sritejbommaraju](https://www.linkedin.com/in/sritejbommaraju) | github.com/SritejBommaraju

EDUCATION

University of California, Berkeley

B.S. Electrical Engineering & Computer Sciences; B.A. Applied Mathematics (Junior)

Berkeley, CA

Expected May 2028

- Core Coursework: Data Structures, Computer Architecture, Algorithms, Machine Learning, Optimization, Linear Algebra; ACT: 34/36

EXPERIENCE

Senior Frontend Engineer & Summer Associate

Jun 2026 – Sep 2026

AECOM

Dallas, TX

- Owned an invoice preparation copilot end to end as a solo build, interviewing **8** billing analysts, project accountants, and project managers to map the billing lifecycle from timesheet to client invoice, then shipping a **FastAPI + React** proof of concept that assembles draft invoices from **6** upstream data sources and cuts per invoice preparation time from hours of manual reconciliation to **under 10 minutes**
- Built a Semantic Clash Engine reconciling drawings against specs in Next.js 16 and TypeScript, anchoring drawing values to the PDF vector layer and grounding quotes via the Citations API, at **100% precision**, **0** false positives, **88.9%** accuracy on FDOT spec fixtures
- Designed the Postgres schema, row level security policies, and NDJSON streaming API routes behind both tools and deployed to Vercel with a keyless mock mode, so **3** internal stakeholder groups could evaluate a live demo in **30 seconds** with zero keys or infra provisioned

Co-founder

Jan 2026 – Jul 2026

Stet AI

Berkeley, CA

- Architected Stet on Tauri v2 (Rust + Next.js) around a custom **5-phase matching pipeline** engineered to handle **3 TB+** deal corpora, processing a **5 GB data room** in **sub 5 seconds** and an **85-page CIM** in **2 seconds** on a 1 to 5 box i9 LAN with zero cloud egress
- Built end to end CIM claim verification through a tamper evident hash chain over **7** public records sources (SEC EDGAR, OFAC SDN, CourtListener, OpenCorporates, CA/DE SoS, news), with **Ed25519** signed JSON sidecars canonicalized via RFC 8785 JCS for auditing
- Closed an **AECOM** contract to process the firm's **\$13B+** M&A deal flow and ran live engagement calls with a **Big 4** advisory practice and **4** middle market PE firms, positioning Stet as the deterministic diligence layer beneath buy side workflows for sponsors and bankers

Co-founder

Aug 2025 – Dec 2025

Elysi

Berkeley, CA

- Founded a stealth AR computer vision company and built an **entirely on device** perception stack: YOLOv8 detection across **80 COCO** classes fused with ByteTrack tracking holding stable IDs through occlusion at **10+ FPS** on commodity CPUs with no cloud dependency
- Cut model footprint with **INT8** post training quantization and ONNX export, then added MiDaS DPT-Small monocular depth through ONNX Runtime with automatic TensorRT, CUDA, and CPU backend selection, holding the full pipeline above **10 FPS** on a laptop CPU
- Hardened the runtime with circuit breaker fault isolation, graceful degradation under CPU and memory pressure, per frame budget enforcement, JSONL streams over MQTT, and Prometheus metrics, with privacy invariants so **0** bytes of raw video ever left the device

Software Engineer Intern (GIS & Machine Learning)

Jan 2024 – May 2024

EcoCulture

Remote (Chandler, AZ)

- Trained Random Forest vegetation classifiers on segmented multispectral UAV orthoimagery through a **48** configuration grid search over tree depth, tree count, and max samples, lifting classification throughput **150%** and eliminating manual per parcel analyst review entirely
- Engineered **8** spectral index features (NDVI, NDRE, ENDVI, EVI, GNDVI, NDSI, NDWI, SAVI) plus Canopy Height Models from photogrammetric DSM and DTM differencing, feeding the classifier a **9** channel feature stack generated end to end with zero manual fixes
- Delivered the automation backbone for the **Lower St. Johns River Restoration Project (LSRRP)**, replacing manual GIS workflows with an ArcPy + Spatial Analyst pipeline from raw orthophotos to classified rasters, and presented it at the Esri Arizona state competition

PROJECTS

Tiny Cache | Verilog, cocotb, LibreLane

Aug 2026

- Built a **2-way** set-associative write-back cache in Verilog with per-set LRU replacement; repeated cache hits take **3** cycles versus **9** for off-chip memory. Preparing the design for tapeout, with completion expected by September 15, 2026
- Hardened the layout to **833** cells and **152** flip flops in **54%** of one tile at **14,568 μm^2** , closing timing at 50 MHz with **10.47 ns** worst setup slack and no hold, max cap, or max slew violations across all **3** corners, an estimated **105 MHz** ceiling, and **735 uW** total power
- Verified at both RTL and post synthesis gate level with cocotb, replaying **300** randomized reads and writes against a Python reference model under randomized memory latency, and profiled cache hit rate from **96%** on tight loops down to **25%** on uniform random access

Mini-Redis | Modern C++, High Performance Networking

Nov 2025

- Rebuilt a Redis-compatible in-memory store in C++ with RESP parsing, TTL expiration, LRU eviction, and durable RDB + AOF persistence, sustaining **27,125 QPS** single client and **31,933 QPS** aggregate across **5** concurrent clients over **150,000** mixed ops
- Implemented thread per client and Windows **IOCP** async I/O with mutex protected key value storage, lazy expiration, and a background AOF writer thread, keeping **P99 latency under 100 μs** at **30k QPS** of sustained mixed read and write load with no dropped connections
- Added primary replica replication on top of RDB + AOF persistence with **20** RESP commands including atomic INCR, DECR, INCRBY, and APPEND, validated end to end by a C++ load generator and an async Python benchmark harness replaying **150k** operation traces

QuantML | Python, Custom Autograd, Quantitative Research

Nov 2025

- Built and open sourced a research grade ML library for quantitative time series modeling with **2** sequence architectures (LSTM and TCN) built on a hand rolled autograd engine, Sharpe ratio aware objectives, and a streaming pipeline of **10+** live market microstructure factors
- Benchmarked **sub millisecond inference** (P95 $\approx$ 0.08 ms, P99 $\approx$ 0.22 ms), **285k+** tensor ops per second on dense element wise ops, and over **14k** training steps per second on a commodity CPU, fast enough to run inside a live trading loop without any GPU acceleration
- Also shipped a full walk forward optimizer, vectorized backtest engine, and alpha signal evaluator computing **10** metrics (Information Coefficient, Rank IC, hit rate, VaR, CVaR, turnover, Sharpe, Sortino, Calmar, maximum drawdown) over rolling windows on every signal

SKILLS & INTERESTS

Skills: Python, C++, Rust, TypeScript, SQL, Verilog; PyTorch, NumPy, YOLOv8, ByteTrack, ONNX, TensorRT, pgvector, LLM agents; FastAPI, React, Next.js, Tauri v2, Docker, Postgres, low latency networking, epoll, IOCP, HNSW/SIMD, RDB/AOF, cocotb, LibreLane, Git

Interests: Quantitative trading, reading, basketball, weight lifting, boxing, hiking, startups, M&A deals, philosophy, journaling, meditation