

VIDYUT SRIRAM

State College, PA (open to relocate) · 848-207-8252 · vidyut0712@gmail.com
linkedin.com/in/vidyut-sriram-4b5a012aa · github.com/vs-githjk
vidyutsriram.dev · arxiv.org/abs/2607.21641
Forward Deployed Engineer | AI Engineer | Available January 2027

EXPERIENCE

Sherlock AI | Software Engineering Intern

Remote · Jun 2026 - Aug 2026

- Owned the public demo interview flow end to end in Next.js (adaptive interview behavior, in-browser recording, submission): the surface prospective customers evaluate before they talk to the team.
- Prototyped a Google Meet add-on that renders audit logs and validates interview integrity from a sidebar inside a live call, so review happens where the interview already takes place.

Comcast India Engineering Centre | Software Development Engineer Intern

Chennai, India · May 2025 - Aug 2025

- Tasked with testing whether low-code agent tooling could work in engineering teams; scoped the problem, chose Langflow and shipped a pilot that unified planning, budget tracking, data storage, team communication and code generation in one interface.
- Cut ~40% of recurring planning and coordination work for a 40-person engineering team by driving adoption, then iterated on real usage with RAG over internal documentation, improving team communication efficiency a further ~18%.

The Pennsylvania State University | Learning Assistant & Math Peer Tutor

Aug 2025 - Present

- Explain technical concepts to non-specialists in calculus review sessions and one-on-one introductory math tutoring.

PROJECTS & RESEARCH

PrismAI | Approval-First AI Meeting Agent for Agencies

Mar 2026 - Present

- Shipped the beta (0 to 1) in 6 months to 14 users across isolated multi-tenant workspaces, 100+ meetings analyzed: Recall.ai bots join Zoom, Meet and Teams calls, and 12 LangGraph agents turn each call into decisions, action items, emails and Jira/Linear tasks.
- Engineered hybrid RAG over transcripts and PDF/Notion/Drive documents (pgvector + BM25 with reciprocal-rank fusion, query rewriting, LLM reranking) and an MCP connector that puts meeting data inside Claude and ChatGPT.
- Now building the agency release (Dec 2026 launch): flags out-of-scope client requests with the contract clause quoted and drafts a change order priced from the agency's rate card. Code computes every dollar; nothing is sent until the owner approves.

BetterBill | AI Billing Advocate for Hospital Bills

Oct 2026

1st Place, Visa Track & 1st Place, Photon: Best Voice Agent | AI Hackathon 2026 by LA Hacks, UCLA (24-hour hackathon)

- Built the bill-intelligence core on a 4-person team: a deterministic TypeScript rules engine that computes every flag and dollar in integer cents while the model only transcribes, plus the photo and text bill readers, the API and the Postgres schema.
- Rules engine caught 22 of 22 seeded billing errors across 11 synthetic bills with 0 false flags on 3 clean bills; the photo reader reached 88.8% line recall on 35 phone photos. Added Sentry monitoring on every AI call and a shareable evidence packet.

Terrametric | Crop Drought Stress Intelligence Platform

Apr 2026

1st Place, Real World Track (Ambrook) & Data Science Track; 3rd Place, Open Source | Bitcamp 2026, UMD (36-hour hackathon)

- Came up with the idea and shipped it in 36 hours on a 3-person team: a field-level platform that scores drought risk, prescribes irrigation with cost and savings projections, and texts alerts to farmers without a dashboard. Led the live pitch to the judges.
- Owned orchestration, the stress metrics, model training for a 14-day per-field XGBoost forecast, and a RAG chat (Qwen 2.5-7B) grounded in Iowa/USDA agronomy documents.

Tool-Guided Retrieval-Augmented Repair for Securing LLM-Generated C Code

Sep 2025 - Jul 2026

First author, [arXiv:2607.21641](https://arxiv.org/abs/2607.21641) | 2nd Place, Engineering, Penn State Undergraduate Exhibition 2026

- Built a repair pipeline for LLM-generated C that feeds GCC and CodeQL diagnostics into an iterative loop guided by retrieval over a self-growing repository of past repair outcomes, with KLEE symbolic execution as final validation.
- Across 5,000 C programming tasks, cut DeepSeek Coder 1.3B compilation failures from 42% to 22% and security defects from 35% to 15%, and cut CodeLlama 7B total CodeQL errors from 15,088 to 2,463 (83.7%).

EDUCATION

The Pennsylvania State University

University Park, PA · Aug 2023 - Dec 2026

B.S. Computer Science, Minor in Mathematics | GPA 3.84 / 4.00 · Dean's List, 6 of 6 semesters

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, Java, C, C++, SQL

AI Systems: LangGraph, RAG (pgvector, BM25/RRF), MCP, tool calling, LLM evals, voice agents, OpenAI, Anthropic & Gemini APIs

Backend & Infra: FastAPI, oRPC, PostgreSQL (row-level security), Drizzle ORM, Supabase, Temporal, SSE, WebSockets, Docker

Frontend & Integrations: React, Next.js, Tailwind CSS, Chrome extensions, Recall.ai, Slack, Notion, Jira, Linear, Twilio

Testing & Security: pytest, Playwright, Zod, CI security gates, Sentry, Git, Vercel, GitHub CodeQL, KLEE symbolic execution