VENKATA SYLESH KONA

New York, NY | +1 (518) 977-9740 | konavenkatasylesh@gmail.com
linkedin.com/in/sylesh-kona | github.com/sylesh29 | sylesh29.github.io/My_Portfolio

SUMMARY

- Software engineer, M.S. Computer Science (GPA 3.77, May 2026); sole engineer at Auteur, a consumer AI startup, after production backend work at Rivian Automotive; LLM agents and voice AI across healthcare and insurance claims; Go, Python, TypeScript, AWS.
- Built an LLM agent evaluation platform that caught a safety regression aggregate metrics hid (must escalate recall down 87 points) and an LLM judge fabricating a citation at 0.98 confidence.
- Shipped REST APIs in Go at Rivian with about 80 tests, ending manual MongoDB edits by ML teams, plus an event driven AWS monitor that surfaced silently failing simulation jobs.

EXPERIENCE

Auteur (Virgo Rising, LLC), Lead Engineer *New York, NY (Remote) | Oct 2026 to Present*

- Sole engineer on a consumer fashion curation platform (Next.js, React, TypeScript, Postgres): built a deterministic styling rules engine, OKLab color matching, and a **seven signal ranker that filters 40,000 products in 15 ms** median, covered by **446 tests**.
- Shipped a Claude vision tagging pipeline with a controlled vocabulary, confidence gating, holdout evaluation, and an expert review loop, plus catalog ingestion across four affiliate networks that keeps credentials and proprietary rules off the client.
- Delivered an installable PWA with passwordless magic link auth, WCAG 2.1 AA accessibility, and FTC compliant affiliate disclosure.

Rivian Automotive, Software Engineering Intern, Data Applications & ML *Palo Alto, CA | May 2025 to Aug 2025*

- Shipped update (PATCH, partial updates, protobuf validation) and delete REST APIs plus CLI commands for the Go VMCAP vehicle telemetry service, with about 80 go tests, ending manual MongoDB record edits by downstream ML teams.
- Built an event driven AWS Batch failure monitor (EventBridge, SQS with a dead letter queue, Lambda, Terraform) that maps each failed array job child to its scenario, updates Simdash, and alerts engineers in Slack, replacing silent failures in 25 to 30 scenario runs.
- Fixed 5 production defects by tracing failures from the API layer through PostgreSQL and MongoDB and shipping each fix with a regression test alongside the owning engineers.
- Raised frontend test coverage on Polaris, an internal ML dashboard, from 35% to 85% in two sprints by authoring 25+ Cypress component and end to end tests.

University at Albany, SUNY, Graduate Teaching Assistant *Albany, NY | Jan 2026 to May 2026*

- Taught Discrete Mathematics and Cryptography to 100+ graduate students; designed assessments and ran weekly office hours.

PROJECTS

Voice Agent Evaluation Platform | *Python, FastAPI, SQLite, Next.js, LLM as a judge | github.com/Sylesh29/voice-agent-eval* *Sep 2026*

- Built an evaluation suite for a **multi turn healthcare voice agent** that grades against simulated world state, making the headline metric **false completion**; surfaced a regression where **must escalate recall fell 87 points** while headline metrics moved 1 to 3.
- Raised evaluator agreement with human labels from Cohen's kappa of -0.05 to 0.86, and added a deterministic quote check that caught the LLM judge fabricating a citation it labelled verbatim; 45 regression tests.

Patient Voice Tester | *Python, LiveKit Agents, Deepgram, OpenAI, Cartesia, SignalWire SIP* *Oct 2026*

- Built a voice bot that phones an AI medical receptionist and plays a realistic patient; placed **26 real calls across 23 scenarios**, recording both sides and measuring per turn latency, overlaps, and dead air.
- Reduced 143 raw candidate findings to 25 issues (7 high severity, including accepting a mismatched date of birth and disclosing appointment details after failed identity checks), dropping claims a quote check could not match to the transcript.

ActionGuard | *Python, standard library only | github.com/Sylesh29/actionguard* *Sep 2026*

- Built a deterministic action guardrail for a computer use agent operating a claims system: 14 rules on an allow, escalate, abstain, block lattice that stop payments to excluded providers, PHI egress, and duplicate replays, routing each escalation to a named owner.
- Red teamed it with 26 indirect prompt injection attacks: the detector caught 76.9% (none of the adaptive semantic attacks), yet **100% were contained** by verb level authority controls; 83 tests and a tamper evident audit trace.

AuditGuard AI | *Python, FastAPI, Claude API, Cognee, React, SSE* *Jun 2026*

- Built a 4 agent workflow at the M-AGENTS hackathon that flagged 850+ integrity issues in a 5,000 row manufacturing dataset, ranked them by regulatory audit risk, and auto corrected or escalated each to a human; as product owner, repositioned the product and live demo around the person running the audit.

TECHNICAL SKILLS

Languages: Python, Go, TypeScript, JavaScript, SQL
LLM & Agents: LLM APIs (Claude, OpenAI, Groq), agent workflows, tool calling, MCP, RAG, embeddings, vector databases, structured outputs, prompt design, evaluation, LLM as a judge, guardrails; LangChain, Langflow
Backend & Frontend: REST APIs, FastAPI, Node.js, Express.js, React, Next.js, PostgreSQL, MongoDB, SQLite, Astra DB (Cassandra)
Voice AI: LiveKit Agents, Deepgram, Cartesia, SIP (SignalWire)
Cloud & DevOps: AWS (Lambda, SQS, EventBridge, Batch, S3, CloudWatch), Docker, Terraform, CI/CD, Git, Linux; Cypress, pytest