

About Me

Senior staff software engineer with 12+ years of experience leading architecture and platform initiatives across healthcare, AI, telephony, and cloud systems. Specialized in scalable LLM systems, distributed infrastructure, observability, and operational tooling within fast-moving startup environments.

Core Competencies: AI/LLM Platforms • Distributed Systems • Healthcare Technology • Cloud Infrastructure • Observability • Telephony Systems • Technical Leadership

Experience

- **Actium Health** Remote
Oct 2024 – Oct 2025
Senior Staff Software Engineer
 - Served as technical lead during organizational restructuring and post-acquisition integration, driving architectural direction, engineering standards, and operational processes across an 8-person engineering team
 - Designed and led implementation of a **HIPAA-compliant AI insights platform** processing healthcare audio data using STT and LLM workflows, including scalable ingestion, retries, DLQs, idempotency, and asynchronous processing infrastructure
 - Co-architected and launched an outbound AI campaign platform powering **automated healthcare outreach workflows** using large-scale deduplication, prioritization logic, and ML-driven targeting systems
 - Built the organization's first **LLM evaluation framework**, enabling technical and non-technical teams to create 300+ evals validating prompt and workflow behavior across hundreds of production voice agents
 - Led deep-dive LLM workflow and cost optimization initiatives, **reducing production AI inference costs** by ~15% (~\$15k /month) through prompt caching analysis, prompt restructuring, and infrastructure instrumentation
 - Defined and operationalized the team's monitoring, alerting, and **on-call philosophy** from the ground up, including dashboards, escalation policies, runbooks, alerting standards, and engineering training adopted across the organization
 - Led architectural initiatives across AI infrastructure, observability, queueing systems, and operational scalability while serving as a **primary technical escalation point** for high-ambiguity engineering problems
- **Syllable Corporation** Mountain View, CA / Remote
April 2022 – Oct 2024
Staff Software Engineer / Squad Lead
 - Founded and led "TACO", a highly cross-functional engineering initiative responsible for the operational tooling and internal control systems supporting a **healthcare voice automation** platform processing ~1M calls/month
 - Expanded the operational platform into a self-service control plane supporting deployment, rollback, auditing, templating, and large-scale telephony configuration management across ~1K production voice agents and ~300 weekly deploys
 - Designed deployment workflows and safeguards that reduced customer-impacting configuration failures while saving engineering and operations teams dozens of hours per week
 - **Led engineering execution** across highly ambiguous operational initiatives involving Engineering, Operations, Product, QA, and executive leadership stakeholders within a ~30 engineer organization
 - **Pioneered early organization-wide adoption of LLM technologies** by building the company's first internal GPT platform, reusable AI integration libraries, and prompt tooling workflows
 - Built and productionized AI-powered call insights and transcription pipelines processing tens of thousands of healthcare calls monthly using LLMs and STT systems for operational intelligence and customer reporting
 - Partnered directly with Azure stakeholders to secure enterprise-scale GPT throughput capacity and **productionize high-volume LLM workloads**
 - Conducted 100+ technical interviews and mentored engineers across multiple teams, contributing to team growth and technical leadership development
- **Syllable Corporation** Sunnyvale, CA / Remote
Dec 2017 – April 2022
Senior Software Engineer
 - Built and launched early **Android/iOS React Native** patient communication applications and supporting web messaging systems as part of the company's patient engagement platform efforts
 - Led development and operationalization of a **browser-based softphone** used to replace legacy Cisco telephony infrastructure and enable remote healthcare call-center operations during COVID-19 lockdowns, successfully supporting a major NYC healthcare system handling **~500k calls/month**
 - Relaunched the company's web chatbot platform with an entirely new frontend and backend architecture used across 7+ enterprise healthcare customers, serving **10M+ web visitors monthly**
 - Built large-scale SMS campaign and appointment reminder systems responsible for delivering **millions of patient-facing text messages** across healthcare customer deployments
 - Led architecture and implementation of a **centralized operational management platform** replacing spreadsheet-driven deployment workflows for hundreds of production voice agents
 - Introduced **Prometheus** and **Grafana** observability infrastructure across ~17 microservices and multiple **AWS** and **Azure** environments, substantially improving production monitoring, debugging workflows, and incident response across engineering teams
 - Served as **frontend technical lead** across multiple product initiatives, mentoring engineers, driving architectural decisions, and frequently leading major production incident investigations and remediation efforts

Ambulnz, LLC

NYC

• Senior Software Engineer

May 2017 – Dec 2017

- Work on a large, distributed development team covering the entire application stack from frontend to backend, including database, infrastructure, and testing layers
- Manage/develop microservice architecture through **AWS** including the use of EC2, S3, Lambda, API Gateway, DynamoDB, EBS, CloudWatch, IoT, and X-Ray
- Deploy code daily to hundreds of internal and external users with ultra-high availability demands via multiple connected applications in a complex architecture using **CircleCI**, Jenkins, EBS/EC2, Gulp and Docker
- Peer-review code changes and mentor junior developers
- Primary technologies used: JS (Node/Angular/TypeScript/ES6), MySQL, AWS, Linux, Git

Technical Skills

Programming Languages: Python, TypeScript/JavaScript, SQL, Ruby, C/C++, Java

Frontend: React, React Native, HTML/CSS, Tailwind

Backend & Infrastructure: Node.js, PostgreSQL, Redis, Docker, AWS, Azure, Linux, CI/CD, Distributed Systems

Observability & Data: Prometheus, Grafana, Airbyte, BigQuery, Redshift, Quicksight, Looker, Data Pipelines, Queues

AI / LLM Systems: LLM Platforms, OpenAI APIs, Prompt Engineering, Evaluation Frameworks, STT/TTS, Agentic Workflows

Leadership: System Architecture, Technical Leadership, Incident Response, Operational Tooling, Cross-Functional Execution

Languages Spoken: English (fluent), Russian (fluent), Spanish (working proficiency)

Education

• University of Central Florida

Orlando, FL

Bachelor of Science in Computer Science

May 2010 – August 2014

- Primary Coursework: Algorithm Design, Data Structures, Complexity Analysis, Discrete Mathematics, Theory of Computation, Artificial Intelligence, Programming Languages, Operating Systems, Computer Architecture
- Mathematics: Calculus, Differential Equations, Propositional Calculus, Linear Algebra
- GPA: 3.6, Dean's List