

# ALEKSANDER ALLEN

Seattle, WA | (409) 594-7844 | [aleks.allen@engineer.com](mailto:aleks.allen@engineer.com) | [LinkedIn](#) | [GitHub](#) | [CludCode](#)

## SUMMARY

---

Software Development Engineer with production ML systems experience at Amazon Prime Video. Built and shipped LLM-based recommendation systems at scale and now creating privacy-first AI products as a solo founder. Delivered automated testing infrastructure for a semantic recommendation engine serving millions and engineered an LLM validation pipeline that saved 15 dev-hours per week. Expert at horizontally scaling workloads across LLM Agents via application of requirements-driven iterative design to parallelize execution. Seeking to leverage deep expertise in transformer models, cloud infrastructure, and consumer iOS apps to drive innovative AI solutions.

## EXPERIENCE

---

### Amazon Prime Video | *Software Development Engineer I*

Dec 2023 - Apr 2025

- Developed automated testing infrastructure for the semantic content recommendation engine using Python and CDK, supporting millions of users across three continents and establishing the core platform for future recommendation systems
- Built a distributed data pipeline using AWS Glue, processing millions of user feedback signals daily for recommendation model training with 99.9% data freshness.
- Engineered automated LLM validation system using event-driven microservices, eliminating 15 dev-hours/week of manual validation and processing 100K+ requests/day.
- Developed real-time recommendation auditing web tool that saved 10 dev-hours/week, enabling rapid iteration on LLM-driven content matching logic.

### *Software Development Engineer I*

Sep 2022 - Dec 2023

Echo Show - First-Party Video Client & Voice Integration

- Implemented the Prime Video client interfaces voice-assistant integration for Echo Show devices, enabling hands-free video playback and expanding the smart-display product line and CX
- Developed and scaled high-performance interfaces and recommendation features for first-party consumer hardware, improving content discovery speed and relevance for Echo Show users

### Performance Pipe (Chevron Phillips Chemical) | *Capital Projects Engineer*

Jul 2020 - Sep 2022

- Owned \$1.2M capital project portfolio, managing multiple concurrent initiatives from concept through commissioning with 95% on-time delivery and  $\pm 5\%$  budget variance.
- Led cross-functional reviews with 15+ stakeholders (Operations, Maintenance, EHS, Finance); co-developed site-wide MOC digitization initiative using Python/Flask, reducing approval cycles by 50%.
- Managed procurement and installation of major equipment upgrades, including \$850K extruder expansions and \$180K striping capability systems, using SAP for vendor management and leading installation teams, which expanded site production capacity and improved overall output capability

### Oil & Gas / Manufacturing | *Electrical & Instrumentation Engineer*

2019 - 2020

- Designed and maintained safety-critical PLC and SCADA control systems for petrochemical and manufacturing plants using SEL hardware and Wonderware SCADA, applying strict design principles to ensure continuous operation with no safety incidents
- Identified and corrected arc-flash hazards and applied PPMR compliance procedures across multiple facilities, achieving full compliance and eliminating regulatory findings

## PROJECTS

---

### Privacy-First AI Journaling Platform | *ideAI*

2024 - Present

- Swift · iOS · On-Device ML · Encrypted Storage · Semantic Search · MCP · CoreML
- Designed and shipped a fully local-first AI journaling app to the App Store with on-device ML inference, encrypted local storage, and semantic search - no user data leaves the device.
- Built MCP (Model Context Protocol) server integration enabling cross-platform AI context sharing while maintaining zero-trust data boundaries between the user's local data and external models.
- Implemented temporal data aggregation pipeline synthesizing unstructured journal entries into day/week/month/year behavioral ontologies with AI-generated pattern recognition and insights.
- Architected privacy-by-design system where all ML inference, search, and synthesis runs on-device via CoreML; cloud sync is optional, end-to-end encrypted, and user-controlled.

### Zero-Knowledge Remote Development Environment | [cludcode.com](#) | *CludCode*

2026

- TypeScript · Bun · WebSocket · E2E Encryption · ttyd · tmux · Crypto-Proxy
- Built and shipped a zero-knowledge encrypted remote dev tool as a SaaS product in 23 hours, enabling secure remote access to AI coding agents (Claude Code, Codex) from any device.
- Implemented end-to-end encryption with on-device key generation and crypto-proxy layer; the relay server routes encrypted WebSocket traffic with zero ability to read code or terminal I/O.

- Engineered lightweight agent binary (Bun runtime) with ttyd/tmux terminal multiplexing, automatic tunnel management, and single-command installation (curl | sh).

### **On-Device Audiobook Generation Engine | *ChorusReader***

**2025 - Present**

- Swift · MLX · Apple Silicon · Chatterbox TTS · Audio DSP
- Ported the Chatterbox TTS model (500M-param Llama backbone) to native MLX - first full-feature port with preserved emotion control, voice cloning, and exaggeration parameters, achieving 0.5–0.87 RTF on M4 Mac Mini and iPad.
- Developed advanced post-processing pipeline for prosody manipulation, pitch contours, and pacing dynamics tuned for extended listening quality (8+ hour audiobook sessions).
- Architected cross-device Apple ecosystem pipeline: Mac as local render farm, iPhone/iPad as playback client via iCloud sync, with source ingestion from Kindle, PDF, EPUB, and web articles.

### **EDUCATION**

---

#### **Lamar University**

**2015 - 2019**

*B.S., Electrical Engineering* (GPA: 3.61)

- **Achievements:** Mirabeau B. Lamar Scholar, Goldwater Scholarship nominee

### **SKILLS**

---

- **Languages:** Python, Swift, TypeScript/JavaScript, Kotlin, SQL, Java
- **ML / AI:** PyTorch, MLX, CoreML, LLM integration & distillation, on-device inference, TTS/speech synthesis, semantic search, embeddings, transformer architecture
- **Cloud / Infra:** AWS (Lambda, DynamoDB, Step Functions, EventBridge, Glue, CDK, CloudFormation, API Gateway), serverless, event-driven architecture, microservices
- **Security:** E2E encryption, zero-knowledge architecture, on-device key generation, zero-trust design, privacy-by-design
- **Product / Platform:** iOS App Store, SaaS, MCP (Model Context Protocol), REST API, WebSocket, Apple ecosystem (iCloud, Handoff, CoreML), developer tooling
- **Industrial:** PLCs, SCADA, instrumentation & controls, safety-critical systems, arc-flash analysis, capital project management