



Haris Muhibić

AI Product Engineer & Architect

I take AI products from idea to production: research, unit economics, architecture, build, billing and launch.

Sarajevo, Bosnia and Herzegovina haris@chern0.dev chern0.dev linkedin.com/in/haris-muhibic
github.com/harismuha123

SUMMARY

I take AI products from idea to production. I have built software professionally since 2018. Since 2025 I have architected production AI systems for a US SaaS company: a multi-channel sales agent, voice and video generation pipelines, a text-to-SQL analytics agent, and autonomous coding agents in CI. I design deterministic systems that call the model only where judgment is needed, and I choose models with evals and cost data. Through chern0.dev I ship my own products, including Clearcut, a paid background-removal SaaS I took from idea to launch in six days.

EXPERIENCE

Covideo · Apr 2022 – present

Covideo is a US video-messaging platform for sales teams and car dealerships. Droplr, the screenshot and screen-recording product, is a Covideo company (acquired in 2020); I led its engineering team as part of my Covideo role.

Solutions Architect, AI Initiative Lead · Feb 2025 – present

I lead Covideo's AI initiative. I architected and prototyped VIN Reels, AI Video Agent and Smart Suggestions, which launched publicly at NADA Show 2026.

- Architected and led a multi-channel AI sales agent (SMS, email, web chat) on AWS Bedrock. It uses a tool-using orchestrator, confidence-gated routing with fallbacks, speculative tool prefetch and a durable event outbox, plus an eval harness that scores quality, latency and cost per reply.
- Co-led an automated vehicle-video pipeline (Step Functions, about 22 Lambdas). Built its pronunciation compiler (SSML/IPA), Whisper-based TTS quality gate, content-addressed audio cache and resumable execution.
- Built a vehicle-data resolver that extracts listings from any dealer website: platform fingerprinting, a deterministic-first extraction ladder, an LLM fallback and a VIN identity gate.
- Built an internal analytics agent with text-to-SQL over the warehouse, dashboards and scheduled reports, validate-before-save SQL and multi-provider fallback.
- Built autonomous engineering agents (Jira bug → reproduce-first test → draft PR; PR-review responder) and a Claude Code harness that checks Figma designs against the codebase.
- Built shared AI services: RAG over documents and transcripts (Qdrant, semantic chunking, hybrid search), video coaching through constrained selection, and a personalised avatar-video service.
- Moved video encoding to serverless GPUs (NVENC), with a circuit breaker and CPU fallback.
- Built enterprise SAML and SCIM SSO for a dealer-management-system partner, and parallelised the end-to-end test infrastructure.

Software Engineering Team Lead, Droplr · Jun 2022 – Mar 2025

- Led the Droplr team through the v2 rewrites: the Electron desktop app (release owner from 1.0 to 1.8, signing, notarisation, native modules), the Next.js dashboard (which I founded) and the Manifest V3 extension. Reviewed most of the team's pull requests.
- Built the media pipeline: ffmpeg on Lambda for GIF and MP4 conversion, and streaming exports of very large accounts.

DevOps Engineer · Apr 2022 – Feb 2025

- Migrated CI/CD from GitHub Actions to AWS CodePipeline, CodeBuild, ECR and ECS, with image vulnerability scanning.
- Moved standalone databases to an Aurora cluster and set up alarms across load balancers, EC2 and ECS.

chern0.dev · Founder · Nov 2021 – present

- [Clearcut](#): background-removal SaaS and API. Idea to paid launch in six days; \$0.63 GPU cost per 1,000 images.
- [Screencap](#): Chrome extension with on-device OCR and cloud sharing; 232 users, 5★.
- [HireMike](#) (InProd): voice-agent stack decision, real-time interview core, LLM CV screening, security and reliability hardening.
- [Domatiqa](#): GDPR-first two-sided marketplace on web and mobile.

- [Opus Nekretnine](#): satellite property map and a WhatsApp lead-assistant prototype.

Earlier roles · 2018 – 2022

- **Ingenii** · Senior Full Stack Web Engineer (2021–2022): Django REST and Azure data-platform management, deployments with Pulumi, Kubernetes jobs driven by Azure queues.
- **A+E Networks** · Senior DevSecOps Engineer (2021–2022): tooling for the stack behind Lifetime, Vice and History, PagerDuty alerting from Fastly data, DevOps metrics on Lambda and DynamoDB, Terraform.
- **InProd Solutions** · Senior Full Stack Developer, then Technical Team Lead and Product Owner (2020–2021): a DevOps platform for contact centres (Genesys, AWS Connect) on Django REST, PostgreSQL and Elasticsearch.
- **Softhouse Consulting** · Back-end Developer (2019–2020), **Mediadrive** · Front End Engineer (2019–2020), **tribeOS** · Frontend Engineer and Team Lead (2018–2019).

SKILLS

- **AI**: agents and tool use, RAG, evals and model bake-offs, voice agents (STT, LLM, TTS), TTS and SSML, Whisper, vision and segmentation (BiRefNet, SAM), text-to-SQL, MCP, Claude Code and Codex harnesses.
- **Models**: Claude, GPT, Gemini, Amazon Nova, open-weight models; Bedrock, serverless GPUs.
- **Platform**: AWS (CDK, Lambda, Step Functions, SQS, DynamoDB, Bedrock, ECS), Cloudflare Workers, Postgres, Redshift, Kubernetes.
- **Product**: TypeScript, React, Next.js, Node.js, Python and FastAPI, Electron, Expo, Swift; billing with Stripe, Polar and RevenueCat.

EDUCATION AND LANGUAGES

International Burch University, Sarajevo (2016–2019). Bosnian (native), English (full professional), German (limited working).

Case studies: [chern0.dev/work](#) · Services: [chern0.dev/services](#) · Book a call via [chern0.dev](#)