

LOHIT TUMMALAPENTA

Full-Stack Engineer · Python / FastAPI · Docker · Founder

Guntur, Andhra Pradesh, India · +91 86865 03368 · mail@lohit.me

lohit.me · github.com/whiskerEnt · linkedin.com/in/lohit-t-0730063b5

PROFILE

Full-stack engineer with six years shipping production systems end to end: architecture, code, containerisation, infrastructure and operations. Ten platforms live across travel, live commerce, commodity data, gig work, messaging, SMS infrastructure, consumer privacy, developer tooling and mapping. Production code in Go, Python, Rust, PHP, Kotlin, JavaScript and TypeScript, C and C++. Deep Docker and gVisor work, including a public platform that executes untrusted code in eight languages inside isolated, resource-capped sandboxes. Credited on CVE-2026-77281 for three vulnerabilities in Caddy, the web server fronting my own production traffic. Currently contracting with Handshake AI on Science Expert tasks for Terminal-Bench 3 and 4, alongside Mercor, AfterQuery and Turing.

EXPERIENCE

AI Evaluation Contractor · Handshake AI · Mercor · AfterQuery · Turing

Jul 2026 – Present

remote, contract

- Currently: Science Expert tasks for Terminal-Bench 3 and 4 at Handshake AI, and trajectory annotation at Mercor
- Handshake AI: authored tasks for Terminal-Bench 2, the second-generation benchmark for evaluating coding agents in real terminal environments
- 122 tasks delivered to Ready-to-Deliver with 100% approval and a 4/5 quality rating, never a task dropped, each cleared through two independent human reviews plus an automated gate pipeline
- Every task passes static analysis, rubric review, container image build and reference validation, multi-trial agent runs, fairness and adversarial review, and 30 soundness checks spanning oracle correctness, derivability, coverage, determinism and harness integrity
- Design exploit-resistant grading harnesses for an environment where the graded agent runs as root and can rewrite the standard library the grader imports: stdlib-only graders under isolated interpreter flags, dual-channel outcome recording reconciled line by line, privileges dropped before any graded program starts, and answers destroyed before first execution
- Built a Python QC toolchain: pluggable probe driver, mutation testing across the graded corpus to prove no enforced rule survives without a fixture that distinguishes it, boundary-gap detection for untested comparison thresholds, and a staticcheck reproducer
- Handshake AI, mathematics and ML tasks: seven cleared the difficulty gate at a measured 0/5 or 1/5 pass rate against the benchmarking agent, across formal reasoning, numerical methods, interpretability, computer vision, model training infrastructure and distributed training. The design principle behind all seven: print the closed-form optimum, because knowing the number is not achieving it, and put the difficulty in constructing a legal object that attains it under nonlocal constraints
- Since September 2026 also with Mercor (52 SWE benchmark tasks QC'd with three verified model rollouts each and inter-annotator agreement; codebase Q&A tasks for an RL environment with rubrics, foils and golden solutions) and AfterQuery (Harbor benchmark tasks with sealed 0-or-1 verifiers cleared through anti-cheat and multi-attempt difficulty probes), plus 20 physics reasoning prompts for Handshake that both responder models must fail

Founder & Director · Ishaanvi Holidays Private Limited

2020 – Present

FareEagle · fareeagle.com · CIN U79120AP2026PTC127726 · Guntur, India

- Founder and sole engineer of FareEagle, a full online travel agency covering flights across 100+ airlines, 500K+ hotels, buses and visa, incorporated as Ishaanvi Holidays Private Limited in August 2026
- Rs 27.2 lakh GMV across 119 completed bookings in closed beta on zero outside capital, with the full search, payment and e-ticketing loop running in production
- 231K lines across a 113-class service layer on a custom PHP framework: 92 admin pages, 48 cron jobs, and a versioned REST v1 that backs the web app, both mobile clients and the MCP server off the same handlers
- Shipped a live MCP server on JSON-RPC 2.0 over Streamable HTTP with a two-phase booking handshake: `prepare_booking` returns a priced, held quote and `create_booking` commits it, so an external model cannot move money without confirming a quote it was already given. Real third-party users in production
- Built India's first end-to-end WhatsApp booking flow, search to ticket inside chat: 35 classes covering native Flows with the RSA and AES-GCM flow-endpoint crypto Meta requires, 24-hour service-window tracking, and an intent detector that resolves common trips before an LLM call is paid for, with disambiguation, human handover and re-engagement paths

- Aira, the assistant behind it, sits on a backend factory over Claude and Gemini with streaming extraction, prompt building, content filtering and conversation storage, so the assistant is model-portable rather than pinned to one vendor
- Chargeback defence end to end: device-fingerprint session capture, an evidence bundle service, a financial ledger, daily reconciliation, payment-intent expiry and scheduled evidence archival, with SOC-aligned audit logging across the platform
- Encrypted document vault for passports and traveller documents, alongside CSRF protection, rate limiting, GeoIP and a threat-intel service across the platform
- Programmatic SEO pipeline: 15 LLM-driven generators for route, airport, airline, hotel-city, best-time and itinerary pages, feeding sitemap generation, compression, URL verification and IndexNow submission across 30,000+ indexed pages with a schema.org and Wikidata entity graph
- Android and iOS apps built dual-mode: booking and in-trip companion in one client, on the same versioned production API
- ML dynamic pricing model trained on 40,000+ routes across 3,000+ cities; 40+ table MySQL admin system with role-based access
- Ran seven security audits of the codebase, remediating 35 findings from critical through low severity and improving performance in the same pass
- Razorpay payments, GlideFares supply integration, custom UPI settlement listener, Hetzner infrastructure with TOTP port-knocking and CSF hardening
- Building Aira (aira.care), a medical value travel platform under the same company, where my own dental qualification is load-bearing rather than incidental: clinics are vetted in person and treatment pages carry a named clinical byline, which is what Google's YMYL bar demands of medical content. Built on the FareEagle rails, WhatsApp intake, the Aira assistant, the MCP server and the live flight API, with treatment-search and quote tools as planned MCP endpoints and DPDP-grade consent handling for health data
- Self-taught from zero in 2020, first line of code to live revenue-generating products

SECURITY RESEARCH

Caddy · Three vulnerabilities disclosed, credited and patched

CVE-2026-77281

- Found and reported three independent vulnerabilities in Caddy 2.11.3, the Go web server fronting my own production traffic, with 74k+ GitHub stars: placeholder re-expansion in the rewrite handler enabling environment variable and file disclosure (CWE-94), an unbounded request-body buffer causing memory-exhaustion DoS (CWE-770), and a case-sensitivity bypass in fileHidden that served .git, .env and secrets directories over uppercase URLs (CWE-178)
- Each reproduced end to end against the unmodified official Docker image with runnable proofs of concept, container memory limits and OOM-kill verification, plus suggested patches and regression tests. Report accepted by the maintainer and credited as sole reporter: CVE-2026-77281 / GHSA-j8px-rmr-x-76h9, CVSS 6.5, fixed in v2.11.4

SELECTED WORK

utly · Developer & Everyday Tools Platform

utly.dev

- 174 tools across 15 categories on an Astro and Preact frontend with an embedded CodeMirror editor across nine language modes, client-side WebAssembly runtimes (Pyodide, sql.js, Lua), and sandboxed compilers for Go, C, C++, Java, Rust, Ruby, PHP and C# executing untrusted code in Docker and gVisor
- utly-runner: 400 lines of Go, deliberately not Judge0. Each submission runs under gVisor `runsc` on the KVM platform with `--network=none`, a read-only rootfs and a fresh `/work` per run, capped at 15s wall clock, 512MB, 64 PIDs and 64KB output. Exposed only over a group-gated Unix socket with no TCP port at all, admitting eight concurrent sandboxes and returning 503 rather than queueing past five seconds
- Browser video editor: a 3,700-line Preact timeline with multi-clip arrange, trim, split, merge and reorder, waveform rendering from decoded audio peaks, filmstrip thumbnails, text and image overlays, transitions, chroma key, speed and datamosh effects, and Whisper-backed auto and karaoke captions. Renders server-side through utly-media with presigned R2 uploads for large files
- utly-media: Go service running fixed, parameterised FFmpeg pipelines inside gVisor sandboxes with a read-only debootstrap rootfs, no network, ephemeral scratch and wall-clock, memory, PID and output-size caps. Arguments are built as an argv vector and exec'd without a shell, so no user string can become an FFmpeg flag. Limits are derived rather than guessed: 600s wall clock to match the nginx proxy timeout, 4GB because libx264 above 1080p exceeds a 2GB working set
- utly-api: stdlib-only Go across 30 handlers for DNS, WHOIS, RDAP, TLS chain grading, email auth, DNSBL and tech detection, with unit tests on the DNSBL, email-auth, IP-quality, tech-detection, SSRF and threat-intel paths, and a scheduled threat-feed fetch on a systemd timer feeding the matcher behind IP quality. SSRF guard rejects RFC1918, CGNAT, loopback, link-local including 169.254.169.254, multicast and IPv6 unique-local, revalidating redirects and pinning to the validated public IP

Haatzo · Live Shopping Marketplace

haatzo.com

- Go and Fiber backend, ~75K lines: 378 routes across 65 handlers and 115 services, 77 migrations, with Caddy as the production edge, which is where the Caddy audit started
- Bid placement serialises competing bids with SELECT FOR UPDATE inside a transaction, with rate limiting and verification gates ahead of the lock and a separate held-bid path for moderated users. Transaction discipline is consistent: paired Begin and deferred Rollback across 42 modules, typed error matching in 117
- Trust and safety as a first-class subsystem: NSFW classification, a profanity and reserved-word moderation pipeline with allowlists, and a seller strike system whose penalties decay over time rather than accumulating permanently
- Step-up authentication: a `require_fresh_2fa` middleware forces re-verification before sensitive actions, over a TOTP vault with its own test suite. Seller payout ledger with joining fees and a full dispute and appeal workflow, plus a 20-file transactional email pipeline driven by lifecycle hooks rather than inline sends
- haatzo-web: the storefront on Next.js 16 App Router and React 19 with Tailwind v4: 115 routes across 35 dynamic segments and 246 components, ~85K lines, with route groups separating marketplace, seller, checkout, messaging and admin. Centrifugo subscriptions drive live bid, stock and chat updates; hls.js plays the show stream
- PostgreSQL 16 via pgx, Redis, Centrifugo pub/sub for live bids, chat and stock, MediaMTX and RTMP streaming. JWT with rotating refresh tokens, TOTP, audit logging, and 39 scripted security smoke suites covering IDOR, rate limiting and header-spoofing regressions

Chérie · Fashion Label Store, Client Build

cheriehouse.com

- Client build for a Guntur women's fashion label: website, admin and an Expo app for iPhone and Android on a Medusa commerce backend the brand owns instead of Shopify. ~55K lines, 48K of them TypeScript, across 14 custom backend modules, a Next.js storefront, PostgreSQL and Redis
- Live video styling sessions on LiveKit: Razorpay-paid bookings and a live queue, a stylist console that shows pieces beside the call with Add to bag, two stylists never taking the same customer, and a refund or rebooking when nobody joins in time
- Commerce depth: seven currencies, cash on delivery with an online advance and a confirmation desk, exchanges and returns that refund to Razorpay or UPI and restock, loyalty points with holds and tiers, referrals, affiliates with commission, TDS and payouts, and wholesale price lists with order minimums
- Messaging flows over Resend email, the WhatsApp Cloud API, MSG91 SMS and Expo push with consent rules, quiet hours and signed unsubscribe links; Instagram reels imported automatically and made shoppable; typo-tolerant search in Postgres with `pg_trgm`; 35 Python test scripts driving real browser flows, including a customer and a stylist on a live video call

RatesWale · Daily Commodity Price Platform

rateswale.com

- Next.js and Tailwind data platform across 16 price categories, with Redis caching tiered by refresh cadence (crypto every five minutes, forex hourly, bullion daily) behind thousands of server-rendered location routes generated from the database
- Scraper framework over a `BaseScraper` class and a slug-keyed registry: adding a source is one class and one registry line, with a smoke script per source so a dead upstream fails its own suite instead of the platform
- Daily ingest at real scale: 1,918 mutual fund schemes across the top 20 AMCs, ~10,000 mandi rows across ~2,000 mandis, 60 crypto coins in INR and USD, 36 fuel cities, 96 bullion locations, 20 forex pairs and 37 LPG states
- Redis-backed queue split into connection, scheduler and worker, with a price-alert evaluator running user-defined thresholds against each ingest. The registry documents its own dead ends: HPCL retired when the portal died in May 2026, city gas parked behind OCR because the prices render as images

Zagget · On-Demand Home Services Marketplace

zagget.com

- Payments verified three independent ways before crediting, with a unique gateway payment id and a row lock so the checkout callback and the webhook can never double-credit. Money in integer paise throughout; commission resolved in basis points through ordered additive layers, clamped to per-category floor and ceiling, then snapshotted onto the gig; append-only ledger with running balance
- Fastify and TypeScript API with TypeBox schemas generating OpenAPI from the same handlers; PostgreSQL 16 with PostGIS for geospatial matching; two Expo React Native apps and a Vite admin console on a shared typed client; Docker Compose with PostGIS and Redis health-checked before startup
- An operations console rather than a CRUD admin: 18 admin route modules covering RBAC, disputes, finance, safety, tickets, ops and moderation

SMSFoundry · Multi-Tenant SMS Gateway SaaS

smsfoundry.com

- Turns spare Android phones into programmable SMS infrastructure with no third-party SMS vendor in the path. Devices pair by QR against a 32-byte token stored only as a hash and expiring in ten minutes, then send on behalf of the tenant
- REST API at `/api/v1` for single, bulk and template sends plus delivery status, templates and devices. Authenticated by `X-API-Key` against a SHA-256 key hash, with per-key per-minute limits surfaced through `X-RateLimit-Limit`, `Remaining` and `Reset`, `Retry-After` on 429, and an optional `X-Idempotency-Key` that suppresses duplicate sends for 24 hours

- Webhooks are HTTPS-only and HMAC-SHA256 signed, with an SSRF guard on the target URL, retry with backoff, a per-delivery UUID for receiver-side idempotency, and every attempt logged with status code and duration
- A plan guard enforces quotas on every action (send, bulk, device, template, contact, group, API key, webhook) against per-tenant daily and monthly counters, and an abuse detector flags rapid sending, duplicate content across recipients, high failure rates, odd-hours traffic and spam keywords before a message leaves

WipeMySocial · Privacy-First Social Media Cleaner

wipemysocial.com

- Chrome MV3 extension that bulk-deletes posts, comments, likes and reactions across Facebook, Twitter/X and Instagram by driving each platform's own interface: per-platform content scripts targeting stable test ids, infinite scroll to page in history, and filtering by year, month or custom range
- Every action is paced by randomised delays with a configurable floor and ceiling, so a deletion run reads as human activity and stays under platform rate limits instead of tripping them
- The privacy claim is enforced by the architecture rather than the policy page: automation runs inside the user's existing browser session, no credentials are stored, and the only outbound payload is a count (platform, content type, quantity) sent against a licence key. Permissions are limited to activeTab, scripting and storage across three host origins, with licence validation and monthly quota enforcement handled in the service worker

Suppp · End-to-End Encrypted Messenger

suppp.me

- Go and Fiber backend, ~14K lines: Signal Protocol key distribution with Curve25519 identity keys, Ed25519-signed prekeys, one-time prekey batches and signed-prekey rotation, so the server holds public key material and ciphertext only
- WebSocket hub with Redis pub/sub for cross-instance fan-out, delivery receipts and WebRTC call signalling with replay protection; Argon2id phone hashing, RSA-256 JWT with refresh rotation and token revocation, systemd hardening
- Beyond messaging: ephemeral flash posts with a scheduled cleanup job, waves, and a media service with per-user quotas and FCM push, on ULID identifiers throughout

Rank Mascot · SEO Audit & AI Visibility Platform

in development

- Turborepo monorepo, ~44K lines of TypeScript across 12 packages and 2 apps: Fastify and tRPC API, BullMQ workers, and a static marketing site. Drizzle schema with PostgreSQL row-level security and a tenant-scoped pool; envelope encryption with per-tenant keys under a KEK held outside the database
- Severity is arithmetic, not a label: $\text{base weight} \times \text{page value} \times \text{site confidence}$, where page value renormalises across whatever signals exist (Search Console clicks and impressions, internal PageRank percentile, click depth, GA4 conversion path), so ordering is correct before Google's OAuth review completes. The only learned input is site confidence, computed once per check type per site and cached, never per instance
- Chrome MV3 extension for a dataset with no API: Bing's AI Performance report exports only against a signed-in session, so the extension lifts the CSRF token off Bing's own citation-stats request in the page world, replays the export exactly as the Download button does, and relays the CSV under a user-issued token, never touching Microsoft credentials
- 51 test suites written around failure modes rather than assertions: why linear normalisation collapses the long tail and makes severity useless, why an unreachable page must not score as a shallow one. Link graph with duplicate, redirect and scaled-content detection; crawler with its own frontier

Whisker Cart · Self-Hosted E-Commerce Platform

whisker.lohit.me · github.com/whiskerEnt/Whisker-Cart

- Zero-dependency PHP and MySQL storefront, 158 files and 28 tables, installs on any shared host with no Composer or Node. 100+ stores deployed in the first month, one running 500+ orders per day
- `core/` is a small MVC framework (router, request, response, session, validator, view, rate limiter and a plugin manager), with all four payment gateways implemented as drop-in plugins behind a single `PaymentGatewayInterface`, each with HMAC webhook verification and idempotent processing
- 19 unit tests and a CI integration suite written around failure modes: refund safety, refund money handling, settings whitelisting, migration SQL splitting, version consistency and explicit regression guards

Utlly Studio · Rust Video Compositing & Streaming Engine

in development

- Windows capture and encode stack validated as 14 standalone probe crates before integration: Windows Graphics Capture, Media Foundation webcam, WASAPI loopback, cpal mixing, NVENC through FFmpeg, RTMP and A/V muxing each proven in isolation
- `compositor-spike` is the working engine at 4,500 lines: wgpu render passes with WGSL shaders for YUV conversion, layer stacking and effects, frames kept GPU-side end to end. ~9,800 lines of Rust; Tauri 2 shell scaffolded, integration pending

Abhilash Auto Enterprises · Two-Wheeler Spares with Vehicle Fitment

in development

- Client build for a spares business trading since 1998, designed around a fitment picker: the customer identifies their bike once and every listing thereafter is filtered to parts that actually fit it. 16K lines of TypeScript

Private Work · Two Unreleased Python Platforms

not public

- Two production Python codebases, 100K+ lines each, across FastAPI, Django and Flask: REST services, background processing and the data layers underneath. Own products, private repositories, available to discuss architecturally

Creator Background · Live Streaming and Gaming

2019 - 2024

- YouTube Partner with 110K+ subscribers on Loki YT, streaming PUBG Mobile in Telugu and English from February 2019 to 2024. Official Meta Gaming Partner from 2020 to 2022 across 100+ live streams of PUBG Mobile, Valorant and Rust, official content creator for Rooter Esports, and a partnered creator with Monster Energy
- Clan leader of Minus40 Esports, running several competitive rosters under the M40 tag. Qualified out of the group stage into the Semi-Finals of PMCO Spring Split 2020 India, Tencent's official PUBG Mobile circuit, from a group drawn with Fnatic and eventual champions Godlike. Later reached the Challenge Season of ESL's Snapdragon Pro Series Season 6, with match wins covered by ESL India
- It is on here because two platforms above came out of it: years of running capture, encode and broadcast pipelines as a heavy user is what led to Utly Studio's Windows capture and NVENC stack, and to the MediaMTX and RTMP layer behind Haatzo's live shows

TECHNICAL SKILLS

Languages	Python · Go · TypeScript / JavaScript · PHP · Rust · Kotlin · SQL · Shell
Containers	Docker · gVisor · multi-stage builds · sandboxed multi-language code execution · resource limits and timeouts · reproducible build environments
Backend	Node / TypeScript (Fastify, tRPC, TypeBox) · Go (Fiber, stdlib services) · Python (FastAPI, Django, Flask, data pipelines, ML) · custom PHP framework · REST and OpenAPI design · JWT and session auth · BullMQ and background workers · idempotency
Caching & Realtime	Redis: caching, sessions, job queues, pub/sub · Centrifugo · WebSocket · MediaMTX and RTMP streaming · wgpu / WGSL GPU compositing · FFmpeg and NVENC
Databases	PostgreSQL · PostGIS · MySQL · SQLite · Drizzle · row-level security and multi-tenancy · envelope encryption · schema design · query optimisation
Frontend	Next.js 15 / React 19 (App Router, SSR, API routes) · Astro · Preact · React Native and Expo · Tailwind CSS · HTMX · Vanilla JS
Mobile	Android (Kotlin) · iOS · React Native · shipped clients for FareEagle and Zagget · offline-tolerant API design
AI / ML	LLM integration · adversarial model evaluation · fine-tuning · agent design · MCP servers · embeddings · pricing models
Security	Vulnerability research and coordinated disclosure · fraud detection · payment dispute evidence pipelines · audit logging for SOC-aligned controls · code security audits and remediation · firewall hardening · TOTP port-knocking
Infrastructure	Hetzner dedicated servers · Ubuntu · PostgreSQL tuning · CSF / firewall hardening · TOTP knockers · Cloudflare · headless browser services · FFmpeg pipelines
Tooling	Git (branching, rebase, bisect, tag-based rollback) · Claude Code and AI-assisted development · systemd · shell
Notes	First-principles builder: reads complex systems by building them, and writes code and documentation assuming the next reader is a stranger with no context. Production repositories are private company systems; Whisker Cart is open source and public. The remainder I can discuss architecturally, not share. Open to contract work, available immediately, async-friendly across IST and most time zones. English (professional) · Telugu (native) · Hindi (conversational).