

DAKSH CHAUHAN

AI Automation Engineer

Ahmedabad, India · me@buildwithdaksh.com · buildwithdaksh.com · github.com/Eaterofsouls · linkedin.com/in/daksh-r-chauhan

PROFILE

AI Automation Engineer who designs and ships AI-powered systems for real business workflows. I turn ambiguous operational problems into working software, combining automation, AI models, and conventional engineering rather than forcing every problem through an LLM. My work spans workflow orchestration, backend systems, model integration, evaluation, and production reliability, with a focus on building systems that can actually be used and maintained.

CORE COMPETENCIES

AI Systems & Automation Architecture — Multi-stage automation with deterministic processing before any model call, schema-constrained AI decision layers, multi-provider LLM routing with fallback, and RAG across vector, graph, and lexical retrieval. Orchestrated across LangGraph, Temporal, n8n, and MCP.

Full-Stack Engineering — Production frontends in React, Next.js, and TypeScript at 100/100 Core Web Vitals, backed by async FastAPI services with typed domain models, durable execution, and security hardening.

Media & Reliability Engineering — FFmpeg-driven video and image pipelines with perceptual hashing and 3D LUT color grading, backed by 800+ automated tests covering concurrency, adversarial inputs, and failure recovery.

PROFESSIONAL EXPERIENCE

AI Automation Engineer

May 2024 – July 2026

Animoart Productions · Ahmedabad, India · On-site · Full-Time · Creative Media Agency

- Built and maintained ~100 automation workflows spanning video, image, and media processing, brand research, client feedback, campaign delivery, social media, web development, and BD prospecting — used across internal operations and client delivery, saving an estimated 20+ hours of manual work per week.
- Cut processing cost to ~\$0.01/image and ~\$0.08 per 5-minute video, and moved turnaround from multi-day to same-day — by keeping deterministic code as the default, reserving model calls for real judgment, and requiring human sign-off before anything shipped.
- Orchestrated multiple LLM providers, vision APIs, transcription services, and media tooling with cost-tiered routing and automatic fallback, so no single outage or safety-filter rejection ever stalled production.
- Built agent workflows on LangGraph for structured tool use, PydanticAI for schema-constrained extraction, and FastMCP for local context retrieval — choosing the right abstraction per problem instead of one framework for everything.
- Connected teams across the agency through n8n workflows integrated with custom Python backends and human review gates, including a BD prospecting engine and reusable prompt frameworks with banned-phrase enforcement for brand consistency.

AI Development Lead

June 2026 – August 2026

MadMarketers · Ahmedabad, India · On-site · Contract · 6-Person Team

- Led a 6-person cross-functional team spanning sales, solution design, delivery, and client handoff for commercial AI automation engagements.
- Delivered a custom B2B prospecting system for Rajput Bhavin Engineering, covering discovery through closing, with a client-facing admin interface for pipeline and lead management.
- Built a regulatory compliance tracker for Adeca Pharmaceuticals, consolidating filing schedules, licensing deadlines, and document-revision workflows into one operational system.

Full-Stack Web Developer

January 2025 – August 2026

Freelance · Remote

- Delivered 100/100 Core Web Vitals on adecapharmaceuticals.com while independently handling development, deployment, and performance optimization.
- Owned the full development lifecycle across adecapharmaceuticals.com, craftoclick.com, and coachmilap.com — visual identity, architecture, implementation, backend integration, and deployment.
- Built production interfaces across React, TypeScript, Next.js, Vite, and frameworkless vanilla JavaScript, including custom interaction systems and data visualizations.

Live: madmarketers.in · coachmilap.com · adecapharmaceuticals.com · craftoclick.com

SELECTED INDEPENDENT SYSTEMS — DEPLOYED · OPEN SOURCE · LIVE DEMOS

Citeable — AI Search Visibility Diagnostic

Live: citeable.buildwithdaksh.com · Python · FastAPI · Docker · Render

A full AEO/GEO audit needs a mix of automated checks and expert human judgment. Citeable automates the checks — schema, crawlability, live AI-citation testing, authority metrics — and guides the human judgment with a structured questionnaire that shows what the automation found and asks the user to confirm or correct it. Between the two, it covers most of a full audit; it doesn't replace deep legal review or business-strategy calls.

- Every recommendation comes from a claim stored in a governed knowledge base, not a hardcoded string. Each claim carries a confidence level and a source, and gets deprecated — and blocked from ever appearing in a report again — the moment evidence shows it's outdated.
- When a finding has no exact match in the knowledge base, a 3-tier semantic search steps in, using the caller's own AI keys to find the closest relevant claim; with no keys supplied, it degrades gracefully to a lexical match instead of failing outright.

- Samples real answers from 11 AI providers to check whether a site actually gets cited, then runs a second, structurally different model to red-team the first model's draft before it reaches the report.
- Accepts any URL. A custom SSRF defense checks every DNS result, not just the first, and re-validates redirects one hop at a time — closing the TOCTOU gap that lets a simpler filter slip past.

AI Studio — Prompt-to-Video Production Platform

GitHub: github.com/Eaterofsouls/AI-Studio · Python · FastAPI · Temporal · FFmpeg · Qdrant

AI Studio turns a text prompt into a finished video, built to replace two bad options: a \$200–400/month subscription tool, or a local GPU rig that overheats and drops frames. You pay only for what you generate — about \$4 for a 30-second commercial, \$7 for a 60-second documentary, roughly 97% cheaper than a subscription.

- Resolves creative direction through a 1,645-preset cinematography RAG — camera bodies, lenses, lighting rigs, film stocks — with Qdrant vector search and a sub-millisecond lexical fallback, turning a vague prompt into a precise generation directive.
- Runs the full production lifecycle as a Temporal durable workflow: a 26-step SOP with three human-approval phase gates. If a provider fails or the server restarts mid-run, execution resumes exactly where it stopped.
- Routes generation across fal.ai and Replicate with automatic fallback, tracking cost to four decimal places before and after every call.
- Grades every output through 21 programmatically synthesized 3D LUT files modeling real film emulsion stocks, with AI-specific debanding to clean up diffusion-model artifacts before the grade is applied.

OpenMontage-MCP — Creative Production Runtime & MCP Server

GitHub: github.com/Eaterofsouls/openmontage-mcp · Python · FastMCP · FFmpeg

OpenMontage is a popular open-source AI video tool — 56,000+ GitHub stars — but it only works one way: a person sits inside a live chat session with an AI coding assistant and drives it stage by stage. This project rebuilds the execution engine as a headless service — a script, webhook, or another program can drive full production directly, with no one watching a chat session.

- Rebuilt the execution engine as a 15-tool MCP server (FastMCP) and a 9-subcommand CLI, so agents and scripts can drive full production from one declarative CreativeSpec JSON — no live chat session required.
- The durable production worker writes state to disk atomically after every scene and checks the filesystem for what's already rendered — so a crash doesn't re-render finished work or double-bill a provider.
- Provider policy routing spans 120 registered providers with a zero-key offline guarantee: no configured credentials falls back to bundled Piper neural TTS and FFmpeg test sources, still producing a verified 1080p MP4 in under 6 seconds.

QA OS — Persistent QA Reasoning Architecture

Explore: qa.buildwithdaksh.com · GitHub: github.com/Eaterofsouls/QA-OS · Python · FastAPI · Neo4j · Next.js · 12 ADRs

QA teams solve the same problem repeatedly: how risky is this change, what should we test, has something like this broken before — knowledge that usually lives in one person's head and disappears when they leave. QA OS scores risk and drafts tests with an AI model, but a human makes the only decision that matters — whether a test actually ships. Once a human decides, the system remembers it, so the next similar case doesn't start from scratch.

- Before scoring a new requirement, retrieves the 3 most similar past requirements and the human decisions stored against them, and feeds that into the AI's prompt — retrieval, not model training; no fine-tuning happens anywhere in the system.
- The human-approval step has zero AI in it: a person reviews each draft test and approves or rejects it — the one decision in the pipeline no model touches.
- Storage is backend-agnostic through a `@runtime_checkable KGClientInterface` Protocol, tenant-isolated on every call: an in-memory stub, a SQLite backend verified to survive full process restarts, and Neo4j driver code that's correct but not yet run against a live database — swappable without touching a call site.
- Governed by 12 formal Architecture Decision Records, including the invariant that an AI agent's output can never serve as its own verification ground truth. 3 of 10 planned modules are fully wired and working today; the rest are honest stubs, left visibly incomplete rather than hidden.

TECHNOLOGY STACK & TOOLING

Languages: Python (asyncio, httpx) · JavaScript/TypeScript · Java · C++ · SQL

Agent Orchestration & Workflows: LangGraph (StateGraph, interrupt/resume) · PydanticAI · LangChain-core · FastMCP (MCP) · Temporal · n8n

Foundation Models & Routing: Anthropic Claude (Sonnet/Opus) · OpenAI (GPT-4o, Whisper, embeddings) · Google Gemini · Groq · open-weight models (DeepSeek, Qwen, Llama)

Backend & Systems: FastAPI · Uvicorn · Pydantic v2 · SQLAlchemy · REST APIs · WebSockets · JSON Schema · async Python (httpx, aiohttp)

Data & Retrieval: PostgreSQL · pgvector · Neo4j (Cypher) · SQLite (WAL mode) · Qdrant · TF-IDF · vector search · graph retrieval

Generation & Creative APIs: fal.ai · Replicate · ElevenLabs · Sync Labs (lip-sync) · Piper TTS · Remotion

Media & Processing: FFmpeg (3D LUT color grading) · OpenCV · Pillow · NumPy · imagehash (pHash) · faster-whisper · Firecrawl · Tavily

Web & Frontend: React · Next.js · Vite · ReactFlow · D3.js · Tailwind CSS · vanilla JavaScript · HTML/CSS

Testing & Quality: pytest · pytest-asyncio · Playwright · adversarial & concurrency stress testing · contract testing

Infrastructure & Deployment: Docker · Docker Compose · Kubernetes · GitHub Actions (CI/CD) · Render · Vercel · AWS · GCP · Azure

Security & Observability: SSRF defense (custom IP-pinning adapter) · OpenTelemetry · CSP · HSTS

Dev Tools: Git/GitHub · Linear · Claude Code · Cursor

EDUCATION

B.Tech in Computer Science & Engineering (AI/ML)
Adani University · Ahmedabad, India

Expected Graduation: 2027

AVAILABILITY

Remote · Full-Time
Working hours: 5:00 PM – 7:00 AM IST, with overlap across US and European business hours.