

# Abhinandan

+91-8130635690 | [abhinandan@abhinandan.one](mailto:abhinandan@abhinandan.one)

[abhinandan.one](http://abhinandan.one) | [linkedin.com/in/abhibuilds](https://linkedin.com/in/abhibuilds) | [github.com/awesome-pro](https://github.com/awesome-pro)

## EXPERIENCE

---

### Founding Software Engineer

Sept 2025 – Present

*Browser*

*Remote*

- Built Browser's Chrome MV3 recorder + CDP-native browser automation agent, achieving 95%+ precise AX/DOM element capture with cross-iframe support, obstruction checks, and real mouse/key/upload execution.
- Built a smart streaming ReAct loop across FastAPI + extension with SSE tool execution, multi-tab orchestration, safe parallelism, abort/continue, and audit logs.
- Cut automation LLM spend by roughly 67% using compact recording traces, context-window compression, prompt caching, and model-routing across GPT-5, Claude Sonnet & Haiku.
- Shipped a zero-LLM replay engine: recordings run as variable-driven tool-call templates, with a stateful AI fallback that resumes mid-run on failure.
- Shipped self-healing docs that auto-repair on UI drift — Haiku → Sonnet diff triage, LLM-free replay of intact steps, and a CDP agent that fixes only what changed.

### Founding Software Engineer

Jan 2025 – Jul 2025

*Cynos Nexus*

*Noida, India*

- Shipped core features of an AI-powered real estate platform using Next.js, Nest.js, GraphQL, Redis, and GCP.
- Built the AI knowledge base service using FastAPI, LangChain, and vector retrieval pipelines, powering customer-facing search workflows.
- Developed document-ingestion pipelines using Google Cloud Vision, XLSX processing, and BullMQ workers, enabling automated extraction of customer data from spreadsheets and scanned records.
- Automated containerized CI/CD infrastructure via Docker, GitHub Actions, and Nginx for reverse proxy/load balancing.

### Contract Software Engineer

Sept 2024 – Dec 2024

*Etkin.ai*

*Remote, Turkiye*

- Built a LangChain + pgvector knowledge base powering AI-assisted document search and retrieval workflows, improving query accuracy by 15%.
- Developed scalable data-ingestion pipelines using bulk CSV processing and Celery workers, reducing processing time by 40%.
- Engineered a production PDF generation system transforming structured AI outputs and dynamic JSON reports into enterprise-grade documents.
- Automated deployment of AI services using Docker, GitHub Actions, and AWS EC2, establishing reliable CI/CD workflows for production environments.

### Open Source Contributor

June 2024 – Aug 2024

*HeroUI*

*Y Combinator S24 startup*

- Received a **personal offer from the CEO** to join HeroUI (prev. NextUI) after making **open-source contributions**.
- Resolved 10+ bugs & delivered 7+ feature enhancements in core components including Calendar, Table and Pagination.

### SWE Intern

Jan 2024 – Apr 2024

*SkilledUp*

*Noida*

## OPEN SOURCE PROJECTS

---

### **AgentFlow-Pro** (*Research*) | *PyTorch, TRL, DAPO, PRM, PEFT/LoRA, FastMCP* | [github.com/awesome-pro/agentflow-pro](https://github.com/awesome-pro/agentflow-pro)

- Rebuilt and extended the ICLR 2026 *AgentFlow* architecture as a local Qwen3-8B Planner → Executor → Verifier → Memory loop with grammar-constrained planning, Tavily search, and sandboxed Python/SymPy tools, served locally through Ollama and FastMCP.
- Trained the planner with **DAPO + a learned Qwen3-0.6B Process Reward Model** (replacing the paper's GRPO) plus a hand-built dynamic-sampling stage TRL lacks, running the full A40 pipeline (DeepSeek-judge step labels → PRM → bf16 LoRA DAPO → GGUF) to measure **+5.0 pts on GPQA-Diamond (40.0%→45.0%)** under leakage-free, quantization-matched evaluation.

### **guardloop** | *OpenAI/Anthropic SDKs, OpenTelemetry, LangGraph* | [github.com/awesome-pro/guardloop](https://github.com/awesome-pro/guardloop)

- Production guardrail runtime for async agents: pre-flight Decimal cost/token/time/tool-call budgets, per-tool circuit breakers, verifier feedback retry loop, OpenTelemetry GenAI spans, and no-rewrite LangGraph/OpenAI Agents SDK adapters that enforce safety inside existing LLM/tool calls.

### **smartmemo** | *PyTorch, sentence-transformers, FAISS, SQLite* | [github.com/awesome-pro/smartmemo](https://github.com/awesome-pro/smartmemo)

- Semantic memory/cache for LLM agents where embeddings retrieve candidates and a learned pair-equivalence classifier decides reuse; shipped bundled classifier-v2 trained on 16,576 pairs across 9 domains (+30 precision points at equal recall vs cosine), plus feedback export, implicit bad-hit detection, gated retraining, and WAL-backed SQLite hardening.

**orchflow** | *Asyncio, LiteLLM* | [github.com/awesome-pro/orchflow](https://github.com/awesome-pro/orchflow)

- Dependency-free typed Python framework for readable multi-agent pipelines: sequential/parallel/conditional flows, retries, shared **StepContext**, flat traces, lifecycle events, human gates, JSON checkpoint/resume, and optional LiteLLM Agent with structured JSON/Pydantic outputs.

**agenteval** | *AsyncIO, OpenAI/Anthropic/LangChain adapters* | [github.com/awesome-pro/agenteval](https://github.com/awesome-pro/agenteval)

- Agent evaluation toolkit replacing exact-match asserts with repeated-run pass-rate tests; traces tool calls/timing/steps, supports collect-then-raise behavioral assertions, OpenAI/Anthropic/LangChain adapters, and Typer CLI JSON reports for CI gates.

---

## TECHNICAL SKILLS

---

**ML & Retrieval:** PyTorch, Hugging Face, sentence-transformers, embeddings, classifier training, RAG, FAISS, Pinecone, Qdrant, Weaviate, ChromaDB, pgvector

**AI Agents & Frameworks:** LangGraph, OpenAI Agents SDK, LangChain, LlamaIndex, CrewAI, AutoGen, Pydantic AI, LiteLLM, MCP / FastMCP, multi-agent orchestration, ReAct, function calling, tool use, CDP-based browser agents

**LLM Engineering:** Fine-tuning (LoRA, QLoRA, PEFT), RLHF / DPO / GRPO / DAPO / PRM, quantization (AWQ, GPTQ, GGUF), inference engines (vLLM, SGLang, Ollama, TGI), context engineering, KV-cache strategies, streaming

**Observability & Eval:** OpenTelemetry, Langfuse, Arize Phoenix, LangSmith, statistical pass-rate evaluation, AIME / GPQA benchmarks

**Cloud & Infra:** Docker, GCP (Vertex AI), AWS, Nginx, Redis, PostgreSQL, MongoDB

**Backend:** FastAPI, Nest.js, Node.js, GraphQL, WebSockets, REST, SSE

**Languages:** Python, TypeScript, C++, SQL, Bash

**Frontend:** Next.js, React, Redux, Turbopack, Edge Runtime

---

## AWARDS & RECOGNITION

---

**International Youth Math Challenge Gold Honour** | **Top 1% TypeScript Engineer Globally** (Algora) | **Amazon ML Summer School 2025** | 2nd Place, Outlier AI Hackathon

**Reliance Foundation Scholar** | **HDFC Badhate Kadam Scholar**

---

## EDUCATION

---

**Dr. APJ Abdul Kalam Technical University**

B.Tech, Computer Science | Nov 2022 – Apr 2026