

INDEX · RAG · HITL · DEPLOY

Autonomous AI Software Engineering & DevOps Platform

Dockerized AISE platform for repo indexing, code RAG/graph, multi-agent PR/CI/bug/incident/deploy, HITL, sandbox, eval/learning, and ops dashboard — for teams that need production engineering automation, not a chatbot bolted to a repo.

M0–M10
MILESTONES CLOSED

PASS
PHASE 1 / 2 / 3

1.0.0
API VERSION

ALL_PASS
M10 SMOKE

Built by **Abdul Basit** · AI Automation Engineer

BUSINESS IMPACT

Faster loops · safer changes · audit-ready SDLC

Index→agents→HITL→verify · unauthorized actions 0 · Phase gates Met · UAT ALL_PASSED

OVERVIEW

SCM fixtures index into PostgreSQL and Qdrant. FastAPI owns RAG/graph, multi-agent engineering, MCP allowlists, HITL, sandbox, deploy verify, eval/learning, and dashboard. n8n runs thin HTTP wrappers. Operators get an ops SPA with API-key RBAC. Live prod cutover is formally waived for sandbox acceptance.

TYPE	AI engineering / DevOps platform (not a single n8n flow)
ROLE	AI Automation Engineer / Platform Engineer
DELIVERY	API 1.0.0 · Milestone M10 · Closeout 2026-09-12 · Milestones doc v2.1
STACK	FastAPI · n8n · Postgres · Redis · Qdrant · Caddy · SCM fixtures · Slack/Email stubs
BUYERS	Eng / DevOps teams needing gated, grounded automation

THE PROBLEM • Slow SDLC — weak PR / CI loops • Invented answers — ungrounded chat • Ungated change — merge without HITL • Open tools — MCP without ACL • No verify — deploy assumed • No closeout — hard to prove gates	THE SOLUTION • FastAPI AISE — agents, RAG, HITL • Thin n8n — index / notify wraps • Index + graph — cited retrieval • HITL gates — merge / deploy • Sandbox — test-gen contained • Verify deploy — success after check	WHO IT SERVES • Teams needing gated eng AI • DevOps needing deploy HITL • Sandbox demos without live OAuth • Lead demo: index → HITL → verify • Secondary: PR review · chat citations
--	---	--

CAPABILITY SURFACES

SURFACE	FOCUS	SURFACE	FOCUS
Index / RAG	SCM · chunk · graph · citations	Agents	PR · CI · chat · incident · deploy
HITL / sandbox	Approve · merge · test-gen	Ops UI	Dashboard · cost · eval
P1 / P2	RN · deps · Go · learning	Security	MCP deny · kill switch · audit

Demo lead: index → chat/PR → HITL deploy/verify → dashboard. **Secondary:** MCP deny · test-gen no-merge. **Closeout:** m10_closeout_smoke.py ALL_PASSED.

Architecture rule: FastAPI owns AI / AuthZ / HITL / sandbox · n8n stays thin · merge/deploy require approval · verify-before-success · env-only secrets.

ARCHITECTURE & CAPABILITIES

Architecture, journeys, agents

Principle: FastAPI owns agents, HITL, RAG, MCP, and sandbox; **n8n owns thin side effects** — no god-workflow.

SCM (GitHub / GitLab fixtures)
→ n8n index + error path → PostgreSQL + Qdrant
→ FastAPI: RAG · graph · agents/MCP · HITL · sandbox
→ Deploy verify · Dashboard · eval/learning · Redis

COMPONENT	ROLE
Caddy + dashboard	Proxy; ops SPA with API-key RBAC
FastAPI	Auth (X-API-Key + X-Role), agents, RAG, HITL
MCP / sandbox	Per-agent allowlists; isolated test-gen
HITL	Approve before merge / production_deploy / rollback
Phase 3	Incident/deploy, test-gen, docs/RN, deps, learning

INTELLIGENCE

- Hybrid code RAG + citations
- Cross-service graph edges
- MCP tool allowlists
- Evidence vs Hypothesis RCA
- Eval scorecards + cost soft stop

INTEGRATIONS

GitHub / GitLab	Fixtures
CI normalize	Demoed
Notify Slack/Email	Stubs
K8s guide	Docs
Live prod	Waived

SAFETY & OPS

- HITL before merge/deploy
- Verify-before-success
- Kill switch + budgets
- Injection + tool ACL
- Append-oriented audit

LEAD DEMO PATH

- Register fixture → index succeeded
- Grounded chat or PR secret finding
- Pending approval blocks merge/deploy
- Deploy anomaly → rollback HITL → verify
- Dashboard runs + cost widgets

SECURITY PROOF POINTS

No API key → **401** · chat MCP create_issue denied · merge without approval blocked · UAT unauthorized actions: **0**.

Reliability, M0–M10, Phase gates, closeout

RELIABILITY AND SECURITY

AREA	WHAT SHIPPED
Auth	X-API-Key + X-RoLe (+ workspace)
MCP	Per-agent allowlists · audited denies
HITL	Approve before merge / deploy
Sandbox	Isolated test-gen · no prod secrets
Verify	Deploy success only after verification
Audit	Agent runs · approvals · tool denies

PHASE GATES + CLOSEOUT

- Phase 1 (M4) — index/RAG; MCP agents; sandbox; injection → 0 - Phase 2 (M7) — \$6.1 MVP; HITL; dashboard; unauthorized → 0 - Phase 3 (M10) — P1/P2; docs+KT; sandbox Met	
m10_closeout_smoke.py	ALL_PASSED
API version	1.0.0
Closeout date	2026-09-12
Evidence chain	docs/m4 → m7 → m8 → m9 → m10

MILESTONES M0–M10 (ALL COMPLETE)

ID	THEME	WHAT UNLOCKED
M0	Kickoff & architecture	Architecture baseline, principles, RACI
M1–M2	Skeleton → index	Compose, RBAC, multi-lang chunk
M3–M4	RAG → Phase 1	Graph, MCP agents, sandbox, Phase 1 Met
M5–M6	PR/CI → HITL	Security review, chat, approvals
M7	Dashboard / MVP	Ops SPA, \$6.1 Met (sandbox)
M8–M9	P1 → P2 polish	Incident/deploy, Go/graph, learning
M10	Docs / UAT / KT	Closeout pack, Phase 3 Met, API 1.0.0

INDEX

Fixture SCM → parse/chunk → Postgres + Qdrant with job status.

AGENTS + RAG

- MCP tool allowlists
- Qdrant cited answers
- PR / CI / deploy agents

OPS + RECOVERY

- Dashboard + HITL queue
- pg_dump backup procedure
- Cutover checklist + KT

Concrete reliability: no-key → 401 · MCP overreach denied · merge/deploy require HITL · verify-before-success · unauthorized at UAT: 0.

Delivery truth: M0–M10 complete · Phase 1/2/3 Met · m10 smoke ALL_PASSED · API 1.0.0 · cite live /health in demos.

OUTCOMES · TECH · DEPTH

Stack, UAT, journeys, n8n, residual risks

TECH STACK + REPRODUCE
FastAPI · n8n · Postgres · Redis · Qdrant · GitHub/GitLab fixtures · Caddy · Compose AISE
docker compose up -d --build · in-container m10_closeout_smoke.py
Surfaces: /dashboard/ · /health · /docs · n8n :5678 · Auth: X-API-Key + X-Role

SCOPE HONESTY

- Portfolio case study — no fake logos or reviews
- Sandbox-first SCM without live OAuth
- Waivers: live prod cutover (W-01), live LLM/OAuth
- Not a repo chatbot — full AISE M0–M10

M10 CLOSEOUT EVIDENCE INDEX

DOC	PURPOSE
UAT.md / ACCEPTANCE.md	Golden paths Met · PRD Complete/Waived · ALL_PASSED
CUTOVER.md / RUNBOOKS.md / KT.md	Deploy checklist · day-2 ops · KT notes
ARCHITECTURE_v1.md	Baseline topology
m10_closeout_smoke.py	Health · MVP · agents · HITL · MCP · 401

DASHBOARD

Runs	Agent status
Approvals	Pending HITL
Chat	Grounded

P1 / P2

Incident / deploy	M8
Eval / learning	Scorecards
Go / graph	M9
K8s guide	Docs

N8N CATALOG

- index / register wraps
- agent run wrappers
- HITL notify/decide
- global error / DLQ

N8N (THIN CALLERS)
Index · agent runs · approval · eval helpers · global error — env API key only

RESIDUAL RISKS (HONEST)
Live SCM/LLM sandboxes · prod TLS/secret rotation · corpus freshness · wet-ink acceptance — tracked as M10 waivers; **none reopen Phase gates.**

SAMPLE JOURNEYS + GATES

\$6.1 MVP evaluate	Chat Cited AuthService	PR Secret finding	Deploy Rollback HITL
Test-gen No merge	MCP Deny create_issue	Learning Override	401 No key

PORTFOLIO ONE-LINER
AISE — Dockerized FastAPI + n8n engineering platform: repo index/RAG/graph, multi-agent PR/CI/bug/incident/deploy, HITL, sandbox, eval/learning, dashboard — M0–M10 complete at API 1.0.0.

SYSTEM VISUALS

Architecture & before / after

