

OMNICHANNEL · RAG · HITL · SERVICE OPS

Enterprise AI Customer Support & Service Ops

Dockerized EASOP platform for tickets, grounded RAG, multi-agent HITL, helpdesk/CRM adapters, dashboard/portal, incidents, voice-to-ticket, and knowledge graph — for teams that need production AI support ops, not a FAQ chatbot.

M0–M10
MILESTONES CLOSED

PASS
PHASE 1 / 2 / 3

1.0.0
API VERSION

M10
API MILESTONE

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

BUSINESS IMPACT

Safer replies · unified tickets · operator control

Inbound→HITL→sync · grounded RAG · Zendesk/HubSpot mocks · Phase gates Passed · closeout smoke PASS

OVERVIEW

Every channel lands in one Conversation/Ticket model. FastAPI classifies, retrieves from Qdrant, runs agents, and either auto-sends under policy or holds HITL. Idempotent Zendesk/HubSpot mocks (plus Intercom/Salesforce demos) keep systems aligned. Operators get a dashboard; customers get a portal. Phase 3 adds WhatsApp/voice mocks, incidents, Teams stubs, and a product–symptom–fix graph.

TYPE	Support / service ops platform (not a single n8n flow)
ROLE	AI Automation Engineer / AI Systems Architect
DELIVERY	API 1.0.0 · Milestone M10 · Closeout 2026-09-11 · Milestones doc v3.0
STACK	FastAPI · n8n · Postgres · Redis · Qdrant · Caddy · helpdesk/CRM mocks · Slack/Teams stubs
BUYERS	Support managers & platform admins needing HITL, tickets, and CRM sync

THE PROBLEM • Siloed channels — orphan tickets • Ungrounded answers — invented policy • Ungated high-risk — refunds/security • No SLA cockpit — spreadsheet breaches • Silent sync drift — duplicate CRM notes • Incident blindness — one-off tickets	THE SOLUTION • FastAPI intelligence — RAG, agents, HITL • Thin n8n — channel E2E wrappers • One ticket model — email/chat/WA/voice • Adapters — Zendesk/HubSpot + demos • Dashboard + portal — ops & self-service • Idempotent sync — request_hash logs	WHO IT SERVES • Teams needing gated AI support • Ops needing one console for HITL/SLA • Sandbox demos without live Meta/PSTN • Lead demo: inbound → ticket → HITL • Secondary: KG · voice · incidents
--	--	--

CAPABILITY SURFACES

SURFACE	FOCUS	SURFACE	FOCUS
Channels	Email · chat · WA · voice mocks	Intelligence	Classify · RAG · KG · eval/cost
HITL / SLA	Approve · escalate · clocks	Adapters	Zendesk/HS · Intercom/SF demos
Ops UI	Dashboard · portal	Notify	Dashboard · Slack/Teams stubs

Demo lead: chat/email inbound → ticket → RAG/HITL → Zendesk/HubSpot mock. **Secondary:** KG query · voice human takeover · Teams stub. **Optional:** WhatsApp mock · incident cluster.

Architecture rule: FastAPI owns AI / AuthZ / HITL · n8n stays thin · high-risk never auto · idempotent adapter writes · env-only secrets.

ARCHITECTURE & CAPABILITIES

Architecture, journey, adapters

Principle: FastAPI owns tickets, RAG, agents, HITL, and adapters; **n8n owns thin side effects** — no god-workflow. Approval decisions never live only in n8n.

```
Channels (email / chat / portal / WA / voice)
  → Edge (Caddy :8080, FastAPI :8000, dashboard + portal)
    → Identity → Ticket → Classify → RAG/Agents/KG → HITL
      → Outbound stub → Helpdesk/CRM sync logs
        → Notify fan-out · Postgres · Redis · Qdrant · n8n wf_*_e2e
```

COMPONENT	ROLE
Caddy + UIs	Proxy; dashboard approvals/SLA; self-service portal
FastAPI	Auth (X-API-Key), domain APIs, agents, RAG, adapters
HITL / SLA	Bands auto/approve/escalate; clocks; handoff packages
Adapters	Zendesk/HubSpot P0 mocks; Intercom/Salesforce P2 demos
Phase 3	WA/voice mocks, incidents, Teams stub, KG, eval/cost

INTELLIGENCE

- Intent / sentiment / priority
- Qdrant RAG + gap tasks
- Specialist agents + tools
- Product–symptom–fix KG
- Eval scorecards + cost rollup

INTEGRATIONS

Zendesk	Mock upsert
HubSpot	Mock notes
Intercom / SF	P2 demos
Freshdesk/...	Waived
Slack/Teams	Stub/webhook

SAFETY & OPS

- HITL before refund/security/angry
- Redis conversation locks
- Injection + cross-customer ACL
- Idempotent sync hashes
- Voice metadata retention TTL

LEAD DEMO PATH

- POST /chat/inbound or /email/inbound
- Ticket + classify + RAG/agents
- Approve when gated → outbound stub
- Zendesk + HubSpot sync log rows
- Dashboard metrics + approvals

HITL PACKAGE FIELDS

Classification · draft answer/sources · reason codes · ticket/conversation ids · decide approve/reject.

Notify: dashboard · email stub · Slack stub · Teams stub.

Reliability, M0–M10, Phase gates, closeout

RELIABILITY AND SECURITY

AREA	WHAT SHIPPED
Auth	X-API-Key + X-Ro1e RBAC
HITL	Auto / approve / escalate bands
Locks	Redis conversation dedupe AI
Adapters	Idempotent request_hash
ACL	Cross-customer fail-closed tools
Guardrails	Injection + ungrounded blocks
Audit	Agent runs · approvals · sync logs

MILESTONES M0–M10 (ALL COMPLETE)

ID	THEME	WHAT UNLOCKED
M0	Kickoff & architecture	Architecture baseline, HITL policy, RACI
M1–M2	Skeleton → channels	Compose, RBAC, identity, tickets, Zendesk mock
M3–M4	RAG → Phase 1	Agents, HITL, SLA, Phase 1 Passed
M5–M6	Specialists → email/CRM	Diagnostics, C360, HubSpot mock, outbound
M7	Dashboard / MVP	Notify fan-out, MVP signed (sandbox)
M8–M9	P1 → P2 polish	WA, incidents, portal, voice, KG, Teams stub
M10	Docs / UAT / KT	Closeout pack, Phase 3 Passed, API 1.0.0

CHANNELS

Email · chat · portal · WA mock · voice mock → one ticket timeline.

RAG + KG

- Qdrant corpus seed/retrieve
- Gap tasks + gated articles
- KG preferred fixes on troubleshoot

OPS + RECOVERY

- Dashboard approvals / SLA
- pg_dump backup procedure
- Cutover checklist + KT labs

Concrete reliability: angry/refund → escalate handoff · empty Teams webhook → sandbox_stub row · voice request_human → no auto send · unauthorized gated auto-action: 0 at Phase 1.

Delivery truth: M0–M10 complete · Phase 1/2/3 Passed · closeout smoke PASS · API 1.0.0 · cite live /health in demos.

Stack, UAT, journeys, n8n, residual risks

TECH STACK + REPRODUCE
FastAPI · n8n · OpenAI (optional) · Postgres · Redis · Qdrant · Zendesk/HubSpot/Intercom/SF mocks · Caddy · Compose project **easop**
docker compose up -d --build · python scripts/m10_closeout_smoke.py
Surfaces: /dashboard/ · /self-service/ · /health · /docs · n8n :5678 · Auth: X-API-Key

SCOPE HONESTY

- Portfolio case study — no fake logos or reviews
- Sandbox-first channels/adapters without live OAuth
- Waivers: Meta WA, PSTN/STT, Teams Graph, client UAT
- Not a single FAQ bot — full EASOP M0–M10

M10 CLOSEOUT EVIDENCE INDEX

DOC	PURPOSE
UAT.md / ACCEPTANCE.md	UAT-01..12 Pass · PRD ID Complete/Waived
CUTOVER.md / RUNBOOKS.md / KT.md	Deploy checklist · day-2 ops · KT labs
ARCHITECTURE_v1.md (v1.1)	As-built topology + adapter matrix
m10_closeout_smoke.py	Health M10 · 1.0.0 · chat · KG · metrics

DASHBOARD

Approvals	HITL decide
Metrics	SLA · product_issue_top
Portal	KB vs agent

EVAL / ADAPTERS

Scorecards	M8 harness
ZD / HS	Idempotent
IC / SF	P2 demos
P2 others	Waived

CHANNELS + KG

- WA/voice mock modes
- Teams sandbox_stub
- KG query paths
- Incident + proactive HITL

N8N CATALOG (THIN CALLERS)
wf_email_e2e · wf_chat_e2e · wf_whatsapp_e2e · wf_voice_e2e — env API key only, no hardcoded secrets

RESIDUAL RISKS (HONEST)
Live vendor sandboxes · prod TLS/secret rotation · LLM quality vs corpus · client UAT scheduling — tracked as M10 waivers; **none reopen Phase gates.**

SAMPLE JOURNEYS J1–J7

J1 Email resolve	J2 Chat parity	J3 HITL escalate	J4 SLA notify
J5 WhatsApp mock	J6 Voice takeover	J7 KG / incident	

PORTFOLIO ONE-LINER
EASOP — Dockerized FastAPI + n8n support ops: omnichannel tickets, grounded RAG, multi-agent HITL, helpdesk/CRM mocks, dashboard/portal — M0–M10 complete at API 1.0.0.

SYSTEM VISUALS

Architecture & before / after

