

FNOL · HITL GATES · GUARDRAILS · MCP AGENTS · SETTLEMENT · KG SANDBOX

AI Insurance Claims Operations Platform

Dockerized AICLAIMS for multi-channel FNOL intake, OCR/IDP + multimodal, policy RAG & coverage assist, fraud/SIU signals, MCP agents with HITL, settlement prep, Claims Analyst, KG/learning sandboxes, and cost/OBS — governed claims ops AI that keeps humans on sensitive writes, not an auto-deny chatbot.

M0–M10
MILESTONES CLOSED

PASS
PHASE 1 / MVP / 3

1.0.0
API VERSION

PASS
SMOKE_M10

Built by **Abdul Basit** · AI Automation Engineer · Delivery closeout 2026-09-12

BUSINESS IMPACT

Safer claims AI · fewer silent fails · audit-ready adjuster ops
Channels→FNOL→agents→HITL→Ops UI · unauthorized 0 · approval 100% · ACL 0 · silent writes 0 · Phase gates Met · CMS/PAS remain SoR

OVERVIEW
FNOL and evidence normalize into claims with ACL. FastAPI owns RAG, MCP allowlists, HITL, guardrails, settlement, Analyst, KG/learning, cost metering, and the ops UI. n8n runs thin HTTP wrappers with a global DLQ. **CMS/PAS/CRM remain systems of record.** Auth is X-API-Key + X-Ro1e. Live payment OAuth and production cutover are fixture/API-first with written waivers (**W-01**); Docker Compose sandbox is the agreed gate for acceptance.

TYPE	AI claims operations platform (FastAPI + n8n; not a single Zap)
ROLE	AI Automation Engineer / Claims Ops AI Engineer
USERS	Adjuster · Manager · SIU · Settlement Approver · Leadership · Platform admin
DELIVERY	API 1.0.0 · Milestone M10 · Closeout 2026-09-12 · Milestones doc v1.11
STACK	FastAPI · n8n · Postgres · Redis · Qdrant · MinIO · Caddy · Compose
PORTS	API 8003 · n8n 5681 · Caddy 8083
BUYERS	Carriers needing gated, grounded claims automation with proof metrics

THE PROBLEM

- FNOL sprawl** — channels never share one claim ID
- Doc lag** — reports/photos stall inboxes
- Ungoverned AI** — auto-deny / invent clauses
- Queue fog** — no confidence / SLA path
- Fraud ambiguity** — accusation without SIU
- Analyst drift** — invented coverage facts
- No closeout** — cannot prove gate metrics

THE SOLUTION

- FastAPI AICLAIMS** — FNOL, agents, HITL
- Thin n8n** — intake / notify / DLQ only
- MCP ACL** — claim ACL + audited denies
- Guardrails** — block auto approve/deny
- Ops UI** — queues + KPIs + cost
- P1/P2 depth** — settlement · KG · learning
- Kill switch** — platform-wide write stop

WHO IT SERVES

- Teams needing **gated claims AI**
- Adjusters needing queues + Copilot ACL
- Managers / SIU with HITL reviews
- Settlement Approvers with payment gate
- Admins with kill switch + demos
- Lead path: **\$6.1 FNOL → queue**
- Secondary: coverage · SIU · settlement

CAPABILITY SURFACES

SURFACE	FOCUS	SURFACE	FOCUS
Intake / FNOL	Web · email · API · missing info	Agents / MCP	Coverage · fraud · investigate · Analyst
HITL / Guard	Settlement · approvals · auto-deny block	Ops UI	/ops/ui/ · cost · aiclaims_*
P1 depth	Settlement · Analyst · eval · MM/comms	P2 depth	KG · payment/inspect sandbox · learning · K8s

GATE METRICS (UAT SAMPLE)

METRIC	TARGET	RESULT	EVIDENCE
Unauthorized agent actions	0	0	m10 pack · UAT
Approval compliance	100%	100%	HITL sample gates
Claim ACL leak rate	0	0	ACL gate + Copilot
Closeout smoke	PASS	M10_SMOKE_OK	tests/smoke_m10.py · health 1.0.0

Lead journey: FNOL → adjuster queue → HITL. **Secondary:** coverage escalate · fraud→SIU · settlement payment blocked. **Closeout:** smoke_m10 PASS · ACCEPTANCE Signed (sandbox).

Architecture rule: FastAPI owns AI / AuthZ / HITL / ACL · n8n stays thin · settlement requires approval · auto approve/deny blocked · CMS/PAS remain SoR · env-only secrets · live prod deferred W-01.

Architecture, journeys, agents, and safety proofs

Principle: FastAPI owns agents, HITL, RAG, MCP, and ACL; **n8n owns thin side effects** — no god-workflow. Fail closed; never invent coverage or fraud facts.

```
Channels (web / portal / email / API / docs / images)
→ FastAPI intake (FNOL + idempotent claim) + MiniIO docs/images
→ Postgres claims · Qdrant policy RAG · Redis kill-switch / idempotency
→ Agents/MCP allowlists · Guardrails · HITL (settlement) · SIU package
→ P1: settlement / Analyst / eval / multimodal / comms
→ P2: KG · payment/inspect sandbox · learning · K8s docs
→ Ops UI /cost /metrics (aiclaims_*) · n8n DLQ · smoke_m10
```

COMPONENT	ROLE IN THE SYSTEM
Intake / FNOL	Multi-channel create; missing-info chase; fixtures m2–m10; idempotent claim ID
Ops UI + API	/ops/ui/ ; Auth X-API-Key + X-Role; OpenAPI /docs; port 8003
MCP server	Per-agent allowlists; claim ACL on tools; audited denies
HITL / Approvals	Settlement/payment always HITL; Manager / Settlement Approver roles
Guardrails	Block auto approve/deny; block unsupported fraud accusation
Kill switch / cost	POST /admin/kill-switch; soft budget via /cost/summary
n8n catalog	wf_claims_intake · docs · SLA · email · approval · wf_error_global

INTELLIGENCE

- Policy RAG + citations / no-invent
- Coverage assist + confidence escalate
- Fraud-signal package → SIU
- Multimodal Observed vs Inference
- Claims Analyst tool-bound Q&A
- Eval scorecards + learning export
- KG neighborhood projection

INTEGRATIONS

FNOL channels	M2 fixture adapters
Documents OCR	Stub IDP classify/extract
Policy corpus	Qdrant RAG
Payment sandbox	M9 (live waived)
Inspection sandbox	M9 (live waived)
Notify stubs	Email / Slack-style
K8s / live prod	Docs / W-01 deferred

SAFETY & OPS

- HITL before settlement / pay
- Auto approve/deny blocked
- Fraud accusation guardrail
- Claim ACL + kill switch
- DLQ / fail-closed paths
- Idempotent FNOL upserts
- Append-oriented audit trail

LEAD JOURNEYS (\$6.1–\$6.8)

- \$6.1 FNOL → adjuster queue
- \$6.2 missing-document chase
- \$6.3 multimodal damage assist
- \$6.4 coverage escalate (no auto-deny)
- \$6.5 fraud-signal → SIU
- \$6.6 email evidence · \$6.8 low-confidence
- Extras: settlement HITL · KG · learning

SECURITY PROOF POINTS

Missing/invalid API key blocked. Auto approve/deny blocked. Unsupported fraud accusation blocked. Settlement without approval blocked. Kill switch stops agent writes. UAT: unauthorized **0** · approval **100%** · ACL **0** · silent writes **0**.

- Roles: adjuster · claims_manager · siu · settlement_approver · admin
- Evidence: docs/m10/samples/uat_results.json

```
Reproduce: docker compose up -d --build · curl :8003/health (1.0.0) · docker compose exec -T api python -m tests.smoke_m1 → M10_SMOKE_OK · UI /ops/ui/
```

Demo arc: FNOL→queue→Ops UI primary · coverage/SIU/settlement secondary · cite health **1.0.0** · closeout smoke PASS · ACCEPTANCE Signed (sandbox).

Reliability, full milestone unlock, Phase gates, closeout

RELIABILITY AND CLAIMS SAFETY

AREA	WHAT SHIPPED
Auth	X-API-Key + X-Ro1e (adjuster / manager / SIU / settlement / admin...)
MCP	Per-agent allowlists · claim ACL · audited denies
HITL	Approve before settlement / payment
Guardrails	Block auto approve/deny · fraud accusation
Kill switch	POST /admin/kill-switch blocks agent writes
Cost	Soft budget alerts via /cost + OBS aiclaims_*
Idempotency	Duplicate intake → no duplicate side effects
Audit	Agent runs · approvals · tool denies · timeline

PHASE GATES + CLOSEOUT

- **Phase 1 (M4)** — FNOL; docs/RAG; coverage/fraud/MCP; unauthorized → **0**
- **MVP / Phase 2 (M7)** — \$6.1–\$6.8; P0 n8n; HITL; ops UI; silent fails → **0**
- **Phase 3 (M10)** — settlement/Analyst/eval; KG/learning; docs+KT; sandbox Met; live **W-01**

MILESTONE UNLOCK (M0–M10)

M0–M1	Architecture · Compose · RBAC/ACL · audit · kill-switch
M2–M3	FNOL · OCR/multimodal · policy RAG
M4–M5	Coverage · fraud · MCP · HITL · SLA · Copilot
M6–M7	n8n E2E · ops UI · MVP gate Signed
M8–M9	Settlement · Analyst · eval · KG · sandboxes · learning
M10	UAT · cutover · KT · ACCEPTANCE · API 1.0.0

CLOSEOUT SMOKE EVIDENCE (UAT_RESULTS.JSON)

CHECK	RESULT	PROVES
api_version	1.0.0	aiclaims-api closeout version
unauthorized_actions	0	No unauthorized agent writes on sample
approval_compliance_pct	100	HITL compliance on sample
acl_leak_rate	0	Claim ACL holds
silent_failures / auto-deny	0	Fail-closed + guardrails
Matrix gate/fraud/m8/m9	passed	Pack regression spots

N8N P0 CATALOG

- wf_claims_intake · wf_email_ingest
- wf_document_processing · wf_claims_analyze
- wf_missing_docs · wf_sla_monitor
- wf_approval / notify · wf_error_global

DAY-2 RUNBOOK (FIRST ACTIONS)

- Queue backlog → assign / work with adjuster role
- Approvals queue → decide with Manager / Settlement
- Fraud package → SIU review (no auto-deny)
- DLQ depth → inspect dead letters · redrive
- Unsafe agent → enable kill switch · audit · disable

Honesty: Sandbox Compose is SoR for delivery. Live prod cutover, live payment/inspection OAuth, Next.js UI, and online fine-tune are waived in writing — see REQUIREMENTS_CHECKLIST · REQUIREMENTS_CHECKLIST Complete/Waived.

SCOPE HONESTY · BUYER FAQ · HOW TO RUN · ACCEPTANCE

What this is, what it is not, and how to reproduce

IN SCOPE (DELIVERED)

- Full M0–M10 on Docker Compose (API 1.0.0)
- FNOL + OCR/multimodal + policy RAG
- MCP agents + HITL + guardrails
- Ops UI + cost/OBS (aiclaims_*)
- Settlement · Analyst · eval
- KG + payment/inspect sandboxes + learning
- UAT Pass + ACCEPTANCE Signed (sandbox)
- KT pack + cutover checklist + runbooks

DEFERRED / WAIVED

- Live production cutover / DNS/TLS/IdP (**W-01**)
- Live CMS/PAS / payment / inspection OAuth
- Next.js apps/web/ (static UI delivered)
- Online model fine-tune / external Neo4j
- Live cluster apply for K8s samples
- Client wet-signature UAT workshop

BUYER FAQ

QUESTION	HONEST ANSWER
System of record?	CMS / PAS / CRM remain SoR. AICLAIMS is a governed ops AI layer; it does not replace the claims ledger.
Live payment rails?	Sandbox / fixtures for UAT and acceptance. Live OAuth waived until provided.
What does acceptance prove?	FNOL→queue, guardrails, HITL, ACL, smoke_m10 PASS at API 1.0.0 — not live loss-ratio uplift claims.
Live production included?	No — cutover waived (W-01). Sandbox Compose acceptance is the contracted-equivalent gate.
Agents auto-deny or auto-pay?	No — auto approve/deny blocked; settlement payment requires HITL.
Where is evidence?	docs/m0–m10, smoke_m10, ACCEPTANCE, REQUIREMENTS_CHECKLIST, uat_results.json.

REPRODUCE (SUMMARY)

```
cd "13-AI Insurance Claims Operations Platform"
docker compose up -d --build
curl http://localhost:8003/health # version 1.0.0
docker compose exec -T api python -m tests.smoke_m10 # M10_SMOKE_OK
docker compose exec -T -e PYTHONPATH=/app api pytest tests/test_m10_closeout.py -q
open http://localhost:8003/ops/ui/ + http://localhost:8003/docs
```

ACCEPTANCE OUTCOMES

- Unauthorized agent actions: **0**
- Approval compliance (sample): **100%**
- Claim ACL leak rate: **0**
- Silent write failures: **0**
- Phase 1 / MVP / Phase 3: **Met**
- ACCEPTANCE: **Signed / Met** (sandbox)

PACK & UPLOAD

- Pack: flagship portfolios/13-AI Insurance Claims.../
- Paste kit: UPWORK_UPLOAD.md
- Setup: SETUP.md · Loom: LOOM_SCRIPT.md
- Gallery: architecture PNG · before/after · PDF
- Rebuild: python assets/write_svgs.py · render_exports.py

One-liner: AICLAIMS — Dockerized FastAPI + n8n claims ops with FNOL, RAG, MCP HITL, settlement assist, Analyst, KG/learning sandboxes — M0–M10 complete at API 1.0.0.

This is: a closed M0–M10 claims operations platform with governed agents and sandbox acceptance. **This is not:** a CMS/PAS replacement, unsupervised auto-deny/auto-pay, or a completed live production cutover.

VISUAL EVIDENCE

Architecture diagram and before → after panel

