OQIRON Platform Manual

D4 · Platform manual · content of 2026-08-28 · source 087f76b6

Supersedes OQIRON_Platform_Manual.docx and OQIRON_Platform_Manual_Detailed.docx of 29 June 2026. This manual describes what the platform is and how it works; for current maturity see the MURAQIB Integration Standard, and for limitations see Section 11 of the MURAQIB Security Whitepaper.

Figures in this manual describing the platform's structure — the size of the policy catalogue, the number of governed routes, the shape of the agent estate — are stable. Figures describing what has passed through the rail change with use, and are stated qualitatively rather than as counts for that reason. The security whitepaper and the Integration Standard carry the current numbers.

---

1. What OQIRON Is

OQIRON is the governance layer for the AI agent economy.

As institutions deploy autonomous AI agents — systems that take real actions: sending communications, moving data, making commitments — a new and unsolved problem emerges: who governs what these agents do, before they do it? Traditional security controls authenticate who is calling and authorize what they may access. They were not built for autonomous agents that act, at machine speed, on behalf of an institution.

OQIRON is built for exactly this: an independent control plane that clears an agent's consequential actions before they execute — screening them, requiring human authorization where the stakes demand it, enforcing separation of duties, failing safe when in doubt, and sealing every decision into a tamper-evident evidence chain that an outside party can verify without OQIRON's cooperation.

The clearing-layer analogy — and the strategic opportunity

The analogy is deliberate. Just as SWIFT and the global payment rails clear a financial transaction before money moves, OQIRON clears an AI agent's action before it takes effect. It is the clearing layer for autonomous action.

This is not only a product idea — it is a strategic one. SWIFT is not merely a messaging system; it is infrastructure the world depends on, and the position of owning that infrastructure confers enduring leverage. The AI agent economy will need its own equivalent: a clearing and trust layer that agent actions flow through. Someone will build it. No established clearing layer for autonomous action exists today. There is a genuine opportunity for Saudi Arabia to build and hold that layer — to constitute foundational infrastructure for the global AI order, the way SWIFT underpins global trade, rather than to participate only as a consumer of AI built elsewhere.

A national dimension

There is a vision, increasingly shared at the highest levels of the Kingdom's AI strategy, that Saudi Arabia will export tokens — intelligence itself — the way it has exported oil. That vision is sound. But it carries an implication often left unspoken: oil did not become a trusted global export merely because it was pumped. It became one because of the trust infrastructure built around it — the certification, metering, and assurance that made every barrel verifiable and trustable at scale.

Exported intelligence will require the same. When an AI agent acts in the world, something must guarantee that the action was governed, authorized, and provable. That trust layer is national infrastructure — and OQIRON is built to be it. This idea is developed fully in the final section of this manual.

The core

At the center is MURAQIB, the clearing engine. Agents connect to it; it screens each consequential action they submit, requires human authorization where the stakes demand it, returns a verdict, and seals that decision into a tamper-evident evidence chain. Around it sits a family of governed agents and platform capabilities that together form the OQIRON platform — described in the sections that follow.

2. Branding and Naming

OQIRON | أوكيرون

The name OQIRON fuses an organizing core (O — the operating center and orbit), cognitive force (Qi — intelligence and the quality of judgment), and the strength of trust (Iron — reliability, resilience, governance). Together they express a single idea: intelligence that can be trusted.

The identity is not decorative — it encodes what the platform does. The mark combines a Trust Boundary (a protected operating boundary around governed AI), an AI Core (the focused, secure intelligence at the center), and a Verification Orbit (signaling identity, connection, and continuous assurance). The logo is the architecture: a core of intelligence, ringed by a boundary of trust, encircled by continuous verification.

Brand promise: "Where AI Earns Trust." AI is not granted trust by default; it earns it, action by action, through clearance and proof.

The components are named to carry meaning

Each component's Arabic name reflects its function — a deliberate coherence between language and purpose.

Component	Name, meaning, and role
MURAQIB	مراقب — "the supervisor," the one who watches over. The clearing engine: the control plane that reviews consequential actions before they execute.
MASSAR	مسار — "the path / course." A governed agent that runs on the rail — the first proven end-to-end correspondence pathway through MURAQIB.
MAJLIS	مجلس — "council / board." A demonstration agent oriented to board and council intelligence, running on the rail.
MIDAD	مداد — "ink." A platform product for Arabic document intelligence.
MAARIFA	معرفة — "knowledge." A platform product for institutional memory.
MIZAN	ميزان — "the scale / balance." A platform product for compliance and screening.
RABT	ربط — "linking / connection." A platform product for integration into the governance rail.
MANARA	منارة — "lighthouse / beacon." A platform product for executive oversight and visibility.

3. Foundational Standing

OQIRON is an incorporated company with protected intellectual property and a registered identity — the foundations that distinguish a serious venture from an idea.

Foundation	Status
Patent	Provisional patent application filed with the United States Patent and Trademark Office (USPTO) in May 2026, Application No. 64/075,982 — covering OQIRON's AI agent governance, certification, and evidence-generation architecture.
Trademark	Trademark application filed with the Saudi Authority for Intellectual Property (SAIP) for أوكيرون | OQIRON, Application No. SA-1245187, Class 42.
Legal Entity	Registered with the Saudi Ministry of Commerce as a fully incorporated Saudi company, Commercial Registration No. 7054567735.

A note on precision. The patent is a provisional application — it establishes a priority date and protects the architecture while the full application is prepared. It is stated as filed, not granted. This precision is deliberate: every claim OQIRON makes is meant to survive scrutiny exactly as worded.

4. The Problem

For thirty years, enterprise security has answered two questions: who are you (authentication), and what are you allowed to access (authorization). That model assumes a human or a system requesting access to a resource. It was never designed for an actor that decides and acts on its own.

Autonomous AI agents break that model. An agent does not merely request access — it takes consequential actions: it sends an email to a counterparty, moves funds, signs a document, commits the institution to something. And it does so at machine speed, continuously, often without a human watching each step.

Why existing controls are not enough

- **Identity and access management** governs what an agent can reach — not what it should be permitted to do in a specific context, at a specific moment, with specific stakes.
- **Logging and audit** record what happened after the fact. By then the action has already taken effect.
- **Model-level guardrails** — the agent policing itself — sit inside the very system being governed. An agent that is compromised, misconfigured, or simply mistaken cannot be trusted to police itself; the supervisor must be independent of the supervised.

What institutions actually need

For an institution — especially a sovereign or regulated one — to deploy autonomous agents responsibly, it needs a control plane that is:

- **Independent** of the agent it governs — not embedded in the system it supervises, so a compromised agent cannot alter the verdict it receives or the record of it.
- **Pre-execution** — it clears the action before it takes effect, not after.
- **Fail-safe** — when the control plane is unavailable or uncertain, the default is to stop, not to proceed.
- **Provable** — every decision is sealed into a chain record an outside party can independently verify. A second, unsealed copy of the same decision is written alongside it and is what the operator's own screens display; verification reads the sealed chain, not that copy.
- **Separation-enforcing** — for high-stakes actions, no single party can both initiate and approve.

This is the gap OQIRON fills — and it is precisely the priority that the leadership of Saudi Arabia's national AI champion, HUMAIN, has publicly identified as a next major focus: guardrails for AI agents. OQIRON is a direct, built answer to that stated priority.

---

5. The Architecture

5.1 MURAQIB — the clearing engine

MURAQIB is an independent service that agents call before they act. An agent submits a proposed action; MURAQIB evaluates it against a policy catalogue and returns one of three verdicts — approved, blocked, or escalated to human authorization — and seals that decision into an append-only evidence chain.

The policy core holds 58 controls. Submitted actions are classified: declared action types map onto a set of canonical classes, and it is the class that determines which controls apply. Five of those classes are authorization classes, used for the second half of a four-eyes cycle.

Not every control has been exercised. Roughly a third of the catalogue has determined a verdict in production; the remainder are defined and in force but have not yet met traffic that reaches them — including most of the sector-specific financial and market-conduct controls, which await the institutional workloads they were written for. The security whitepaper's register states which.

An action type the taxonomy does not recognise is not silently permitted. It escalates: the rail treats an unclassified action as one requiring human authorization rather than guessing at it, and records that it did so. Unclassified actions are currently the most common single determination in the chain, which is a fact about the breadth of traffic the rail has seen rather than about its handling of it.

The evidence chain is the part that makes a decision durable. Each record incorporates the fingerprint of the one before it, so altering an earlier record invalidates everything after it. Periodically the chain's state is signed with a key held offline by a custodian and present on no OQIRON server. That signature is what allows a third party to verify the chain without OQIRON's cooperation, using a verification tool published under an open licence.

What MURAQIB does not do is stop anything. It returns a verdict; the calling system acts on it. This is stated here rather than left for a reader to discover, because it determines what the rest of this manual means. The rail's power is evidentiary: an agent that proceeds against a refusal leaves a sealed record of the refusal it received, outside its own control.

5.2 The agents

Twelve agent identities are registered; nine are active, eight in production. Six of the eight production agents hold a registered signing key and sign their authorization assertions.

One agent performs real institutional work. MASSAR-PIF-001 is a correspondence coordinator for live governance workflows, and it is the reference implementation of the integration contract. It has submitted more decisions to the rail than any other caller.

MAJLIS-002 is a demonstration board copilot. It calls the rail genuinely, but its traffic is exercise traffic rather than institutional workload.

One further production identity is registered and dormant: an earlier MASSAR identity, superseded by MASSAR-PIF-001 and inactive since June. Its records remain in the chain, as records do.

The remaining production identities are the five platform products, described in section 8.

5.3 The platform products

Five products run alongside the rail: MIDAD, MAARIFA, MIZAN, RABT and MANARA. All five call MURAQIB and have sealed decisions in the production chain.

They differ in how deeply they exercise it, and section 8 states that difference product by product rather than describing them as uniformly governed. Three run full escalate-and-authorize cycles with real refusals in the record. Two submit actions and receive verdicts without ever exercising an authorization phase.

6. How Governance Works

6.1 The clearance flow

An agent proposes an action. Before executing it, the agent submits it to MURAQIB with the attributes that describe it: what kind of action, on whose behalf, with what stakes.

MURAQIB classifies the action, evaluates the applicable controls, and returns a verdict:

- **Approved** — the action may proceed.
- **Blocked** — the action is refused.
- **Escalated** — the action requires human authorization and does not proceed on this verdict alone.

Where an action is escalated, a human authorizes or refuses it, and the calling system submits a second action carrying that authorization. MURAQIB evaluates that separately: it checks that the authorizing party is not among those who proposed the action, and seals both the proposal and the authorization into the chain.

The calling system then acts. MURAQIB has no part in the execution. What it holds is the record.

6.2 Four principles

Independence. MURAQIB is not embedded in the systems it supervises. A compromised agent cannot alter the verdict it receives or the record of it. It can decline to ask — and a manual that claimed otherwise would be describing a different architecture.

Pre-execution. The decision precedes the action. Audit tells an institution what went wrong; a pre-execution decision point gives it the opportunity to have gone right, and a record of the choice either way.

Fail-closed. A decision that cannot be sealed is not issued. If the evidence write fails, the request fails rather than proceeding unsealed. One documented exception exists and is recorded in the security whitepaper: a failure to resolve tenant identity records a sentinel and permits evaluation to continue.

Separation of duties. For actions that require it, MURAQIB checks that the authorizing party is distinct from those who proposed the action, by identifier and by email address. The check operates over identities the calling system asserts and signs; the signature establishes who made the assertion, not that the named individuals are accurate.

6.3 The signed channel

Six of eight production agents hold an Ed25519 key pair and sign their authorization assertions. MURAQIB verifies those signatures against registered public keys. This is in force today, not planned: signed assertions appear in the sealed record.

The two remaining production identities predate this and hold API credentials without a signing key.

6.4 The sealed chain

Every decision enters an append-only chain. The chain's construction is specified in full in the security whitepaper's appendices, precisely enough for a third party to reimplement the verification independently and reach the same verdicts.

Verification establishes two things: that the chain is internally consistent, and that its signed checkpoints were produced by the published key. It does not establish that the chain presented is the current one. An institution that wants operator-independent assurance retains checkpoint digests as they are issued and compares them against what is later presented. This limitation is published, not merely disclosed on request.

7. MASSAR-PIF-001 — The Reference Implementation

MASSAR is a correspondence coordinator for institutional governance workflows: it receives requests, drafts responses, routes them for human review, and dispatches them. It is the first system to run the complete governed pathway end to end on the rail and the reference implementation of the integration contract.

What is governed is dispatch. Eight routes call MURAQIB: single-reply correspondence through its proposal, acceptance and final-action paths, and outbound campaigns through content screening, per-recipient screening and authorized dispatch. Each runs a genuine escalate-and-authorize cycle, and the chain carries real refusals — the chain carries refusals as well as approvals at both the correspondence and per-recipient campaign level, in numbers that make clear the escalation path is exercised rather than decorative.

The client fails closed. If a signature cannot be produced, the call raises rather than sending unsigned; if MURAQIB is unreachable, the caller receives a non-verdict it must treat as refusal. There is no allow-on-failure mode.

What is not governed is generation. MASSAR uses a large language model to draft correspondence, run compliance checks, recommend recipients and analyze responses. None of those calls passes through MURAQIB. The rail sees the finished artifact at the point of dispatch, not the process that produced it.

This is stated plainly because it is the first question a technical reviewer asks. "End-to-end" is accurate for the correspondence pathway — proposal through authorization to dispatch, sealed at each step. It is not a claim about the whole application, and the distinction matters: governing what is sent is a different control from governing what is written.

---

8. The Platform Products

Five products run alongside the rail. All five call MURAQIB and have sealed decisions in the production chain. They are presented here in order of how deeply they exercise it, because "calls the rail" and "meaningfully governed" are different claims and the difference is what a reviewer will test.

8.1 MIDAD — Arabic document intelligence

MIDAD analyzes Arabic documents. Submitted text is processed by a large language model and the analysis returned.

Its integration is the sharpest of the five: MIDAD governs egress to the model. Before document content leaves for processing, MIDAD submits the action to MURAQIB and receives a verdict. Where the rail escalates, a human authorizes or refuses, and the chain carries both outcomes — MIDAD's record includes real refusals.

This is a governance pattern worth naming, because it is the one institutions ask about first: the control sits at the boundary where institutional content leaves for a third-party model, not after it has already left.

8.2 MAARIFA — institutional memory

MAARIFA is a semantic memory service. It holds a real vector store and a multilingual embedding model; ingest, semantic search and retrieval are genuine functionality, not demonstration output.

Its rail integration exercises a full four-eyes ingest cycle. Memory ingestion escalates to human authorization; retrieval is governed separately. The chain carries authorized ingests and refused ones.

8.3 MIZAN — compliance and screening

MIZAN performs risk scoring and restricted-entity screening.

The screening engine is production-grade. It performs normalized exact matching with proper Arabic handling — Unicode normalization, diacritic stripping, and unification of alef and ta-marbuta variants — and deliberately does no fuzzy or substring matching, because a screening result that cannot be explained is not usable evidence. If the list is unavailable, MIZAN fails closed and requires the caller to state that screening ran without entity detection.

The entity list it ships with is a ten-entry demonstration fixture of fictitious entities. MIZAN discloses this in its own interface, showing the list's name, source and entry count. The mechanism is real and the data is not, and the product says so rather than leaving a reader to assume otherwise. Screening against an institution's actual restricted-entity list is a matter of supplying that list, not of building the capability.

MIZAN runs a full screening-and-disposition cycle on the rail, with authorized dispositions and refusals both present in the chain.

8.4 RABT — integration into the rail

RABT submits actions to MURAQIB across several action classes and reports usage: submission volume and approval rate.

It is wired to the rail and has sealed decisions. It has not exercised an authorization cycle — its submissions receive verdicts and do not proceed to a second, authorizing action. It is governed in the sense that its actions are cleared and recorded; it does not yet demonstrate the four-eyes pattern the other products do.

8.5 MANARA — executive oversight

MANARA reads live from MURAQIB: executive statistics, per-agent activity, a live decision feed, and chain verification.

Its discipline is worth stating precisely, because it inverts what an oversight dashboard usually does. Every figure MANARA displays originates from the sealed chain or renders as unavailable. When the rail cannot be reached, the interface shows an unavailability state rather than a stale or estimated number. Where a metric would require a formula that has not been defined, MANARA withholds the metric rather than inventing one.

Its evidence export refuses to produce a file it cannot seal. If the export cannot be sealed under the applicable control, the request fails and no file is served — an unsealed evidence pack being worse than none.

MANARA is among the highest-volume callers of the rail and among the shallowest per call: its traffic is dominated by a single action class. It reads the rail thoroughly; it does not yet demonstrate a governed workflow of its own.

9. The Honesty Register

Three of the five products publish their own gap disclosures — pages stating, in their own interfaces, what the product does not yet do.

MANARA's is titled an honesty register. It states which metrics are withheld and why: financial figures until sealed records carry real institutional amounts; composite scores until the formulas behind them can be stated. MIZAN discloses that its entity list is a demonstration fixture. RABT publishes its own gaps.

This is a product principle rather than an apology, and it is the same principle that governs OQIRON's documents. The security whitepaper carries a register of every limitation stated anywhere in it, including the ones that would give a careful reader pause, and invites readers to report any it has missed. The verification tool is published under an open licence with an advisory describing a defect found in its own earlier version.

The reasoning is straightforward. A governance platform asks institutions to trust its record of what happened. A vendor that overstates what its own products do has already demonstrated the thing that matters most — and no amount of cryptography compensates for it.

10. Where to Look for Status

This manual describes what the platform is and how it works. It deliberately makes no claims about what is shipped and what is planned, because those change and a document that carries them goes stale without announcing it.

For **current maturity** — what is shipped, what is roadmap, and the integration contract in specification form — see the MURAQIB Integration Standard.

For **limitations** — a register of everything the platform does not do, stated plainly and cross-referenced to the sections that describe the mechanisms concerned — see Section 11 of the MURAQIB Security Whitepaper.

For **independent verification** — the tool, the published anchor key, and the canonical form specification needed to reimplement verification in another language — see the published verification package.

Where this manual and those documents disagree, they are authoritative and this one is not.

---

11. The Vision: Governance as National AI Infrastructure

Everything in the preceding sections is the foundation. This section is the ambition it makes possible.

11.1 The oil parallel, completed

Saudi Arabia's coming role in artificial intelligence is often framed powerfully: the Kingdom will export tokens — intelligence — the way it once exported oil. The framing is right. But the oil era teaches a lesson its headline often omits.

Oil did not become the foundation of a global economy simply because it was extracted. It became foundational because of the trust infrastructure built around it: the standards, certification, metering, and operational credibility that let a buyer anywhere trust that a barrel was what it claimed to be, delivered as promised. The raw resource was half the story; the apparatus of trust around it was the other half — and it was that apparatus, as much as the oil, that made the Kingdom indispensable.

Intelligence will be no different. If Saudi Arabia is to export tokens — and to deploy autonomous agents across its own government and sovereign institutions — then the agents themselves are only half the story. The other half is the layer that makes their actions trustable: governed, authorized, bounded, and provable. That layer does not arrive automatically. It must be built — and it is exactly as foundational to the token economy as quality assurance was to the oil economy.

11.2 The clearing-layer opportunity

There is a second, complementary way to see it. The clearing and messaging layer of global finance is infrastructure the entire world routes through, and whoever holds that layer holds a position of durable strategic significance. The AI agent economy will require its own equivalent: a clearing layer that agent actions pass through to be authorized and proven.

No established layer of that kind exists today. It is an open position — and a rare opportunity for the Kingdom to build and hold a piece of foundational global AI infrastructure, rather than to participate only as a consumer. To build the clearing layer of the agent economy is to help author the architecture of the AI era itself.

11.3 OQIRON's place in that vision

OQIRON is the concrete, already-begun expression of this idea. What is built is a clearance rail that is independent of the systems it supervises, fails closed rather than open, enforces separation of duties over signed assertions, and seals every decision into a chain an outside party can verify without OQIRON's cooperation. A first institutional workflow runs on it end to end, and the signed-assertion channel is in force across most of the production agent estate.

What remains is substantial and is stated as such. Tenant isolation is incomplete: the machinery is built and proven, but no query on an evidence-returning route yet carries a tenant predicate, so isolation is not enforced at runtime. Capacity has not been measured, and the current architecture carries a serialisation point that adding servers does not relieve — that must be addressed before national scale is claimed rather than aspired to. The trust ladder has one rung left. Deployment topologies for institutions that require them have yet to be built.

None of that is discovered here. Every item is recorded in the security whitepaper's limitations register, which exists so that the distance between what is built and what is intended can be read rather than inferred.

Positioned this way, OQIRON is not merely a governance product. It is a candidate for national AI infrastructure — the trust layer beneath the Kingdom's AI ambitions, and a potential export in its own right: the rails on which a token economy can safely run.

In one line: as the apparatus of trust made oil a foundation of the global economy, OQIRON aims to be the apparatus of trust that lets exported intelligence become the next one — with the hard core built and proven, and the distance to national scale stated rather than glossed.