BlackUnicorn AI Management System

Your AI workforce — with the guardrails built in.

EU · TALLINN 2026

~~We stopped building copilots.~~

We built a workforce.

BlackUnicorn AI Management System is an **agentic operating system for running a company of AI agents** — with the guardrails, memory, and governance built in.

PILLAR 01

Autonomy

Specialist agents pick up real work, collaborate, and propose outcomes.

PILLAR 02

Safety

Sanitisation and guardrails run on every call — sensitive data never leaks.

PILLAR 03

Governance

Human-in-the-loop approval and compliance, enforced in the code itself.

35 specialist agents. 7 Team Leaders. 1 Manager.

MANAGER

Strategy	Security	Compliance	Commercial	Delivery	Dev	PA
Sun	Elliot	Francesco	Sterling	Vector	Steve	Elise
Strategy + analytics	Cyber · OSINT · sentinel · red-team	Compliance · legal · finance	Commercial + sales	Delivery · project mgmt	Eng · QA · infra · research	Chief-of-staff

25 named main-fleet specialists across 7 teams, plus air-gapped classified tiers — each a written persona the platform parses at runtime to gate its capabilities.

Every agent runs on a machine-parsed policy.

Each agent carries a structured policy document the platform parses at runtime to gate every capability, communication channel, and data access it can exercise. No policy declaration, no execution.

APPROVAL TIER · T1 / T2 / T3

Baked in, not configured

T1 auto-approve · T2 notify + approve · T3 always-require. Declared in policy, enforced by the approval engine — not overridable per call.

SKILL ALLOWLIST · MAX 15 (POL-002)

Explicit capability declaration

Platform refuses any tool call not in the agent's declared Skills block. Dead-skill sweeper cross-audits declared vs actual usage every week and alerts on drift.

DOMAIN & SCOPE RESTRICTIONS

Least-privilege by design

Per-agent browser domain allowlist. Memory tier access controls. Chat platform blocklist. Out-of-scope action → block + #sentinel alert. All navigations logged.

POLICY LIFECYCLE

01 / Policy authored by operator

↓ gated by operator approval

02 / Platform parses at agent boot

↓ capabilities registered at runtime

03 / Every action gated against policy

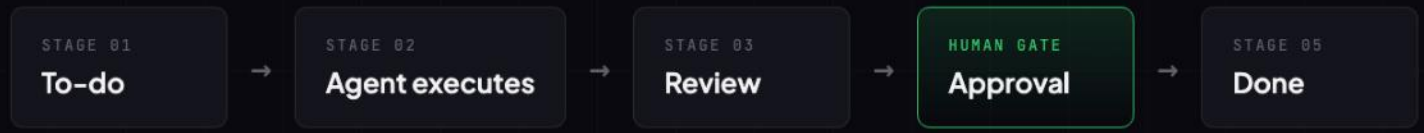
↓ weekly

04 / Dead-skill audit — declared vs actual

Policy changes propagate via version-controlled deploy — never a bare restart. Every update is reviewed, approved, and atomic across all fleet nodes.

When an agent says “done,” that’s a proposal — not a state change.

Agents claim and execute real project tasks. Completion always passes a human-or-leader approval gate. Autonomy never means unsupervised.



Your AI board of directors, deliberating.

Convene a panel into a structured, streamed debate — round by round, with quorum and votes. The manager synthesises a single recorded decision. Interject mid-debate at any time.

Set a goal. The fleet decomposes it overnight.

A company goal flows down through the hierarchy — strategist-decomposed, human-approved, then materialised as real agent tasks. No manual ticket-writing.

Live work items — agents claim, execute, and submit for approval. Completion loops back to goal progress.

Defense in depth, by default.

Ships with a built-in Data Sanitization Proxy and output guardrails — then plugs into the productised security stack.

RuneLM

Data Sanitization Proxy

Classify → pseudonymize → route → rehydrate, fail-closed, on every outbound call.

BonkLM

Guardrails / scanner

Prompt-injection, jailbreak, secret & credential detection.

DojoLM

Red-team & eval

Adversarial testing, CTF arenas, and continuous red-team eval across the agent fleet.

HALLUCINATI

Hard block on HIGH confidence

Classifier on every agent output. HIGH confidence → output is blocked before it reaches the user. Not a log — a hard stop with no retry path.

INJECTION

Inbound scanner — 3 surfaces

OWASP-LLM01 on chat, email, and messaging platform inputs. Block at ≥ 0.8 confidence. Evasion-resistant: base64, zero-width, homoglyph and leetspeak all detected.

Everything routes to the right place.

Every call, every task, and every decision is routed by what it actually needs — for cost, for sensitivity, and for who has to sign off.

BY MODEL

Layer routing

Every call lands on the cheapest sufficient layer — and sensitive workloads are forced local-only.

- L1 · LOCAL
- L2 · SUBSCRIPTION
- L3 · PAY-PER-TOKEN

BY CRITICALITY

Sensitivity routing

A classifier scores every outbound call. High-sensitivity data is locked to on-prem inference.

- LOW
- MEDIUM · NO CLOUD
- HIGH · LOCKED LOCAL

BY TASK & APPROVER

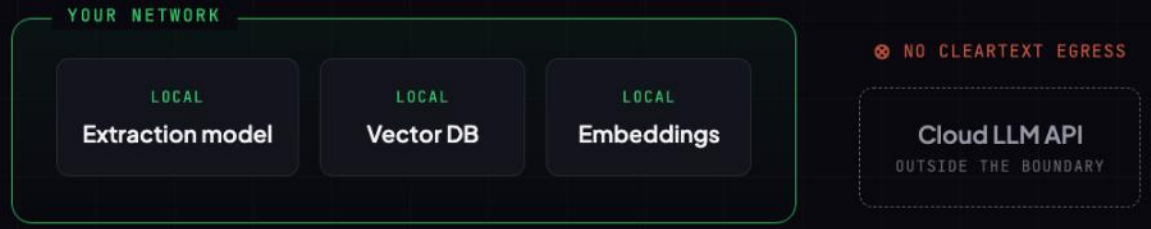
Work routing

Agents claim the right task by capacity; approvals climb three tiers. Finance, legal & security always gated.

- UNIT LEADER
- MANAGER
- OPERATOR

Your memory never touches a cloud API.

Persistent agent memory runs on local embeddings, a local vector database, and a local extraction model. Sovereign by construction — and it fails safe.



Your team's hardware is the supercomputer.

The platform's inference layer runs as a distributed mesh. Every employee workstation with a local model node joins the collective compute pool — idle CPU and GPU cycles become L1 inference capacity for the agentic workforce. Zero cloud. Zero CapEx. Zero vendor lock.

ZERO CLOUD EGRESS

Air-gapped by design

Inference runs entirely on company hardware. No request leaves the network perimeter. Your AI compute cost does not scale with data volume or agent count.

IDLE CYCLES CONTRIBUTE

Auto load-balancing

The router discovers available nodes in real time and routes inference to the least-loaded machine. A standby laptop handles its share of the fleet's workload.

SELF-HEALING MESH

No single point of failure

When a node goes offline the router redistributes instantly. The fleet keeps running as long as any node is reachable. No operator intervention required.

CAPACITY-AWARE ROUTING

Right model, right node

Each node advertises its models and available VRAM. Heavy workloads route to GPU-equipped machines; lighter tasks run on any available CPU node.

CLASSIFIED ISOLATION

Sensitive work stays local

Classified workloads are pinned to designated, audited nodes. The DSP sanitiser runs before any inter-node call. No cross-boundary data transfer, ever.

SCALES WITH HEADCOUNT

Hardware ROI from day one

Install the local model node on any workstation — joins the mesh in minutes. Your compute pool grows as your team grows, with no new hardware spend.

Governance is in the code, not the wiki.

Charter-before-active and retrospective-before-close are gates the platform physically refuses to cross — and governance shows up as a live number.

CONTROL-DEBT SCORE

0–100

per-agent governance credit score

Auto-trips a circuit breaker when an agent's score falls below 40.

VALIDATION RATIO

70/30

human-oversight ratio

Quantified human oversight across approved, reviewed, and audited work.

CIRCUIT BREAKERS

CB-1→5

graduated emergency stops

1 2 3 4 5

Soft pause → agent hold → quarantine → cordon → full fleet halt.

Compliance isn't a layer you add. It's the foundation.

Delegated approval authority · per-decision AI risk scoring · time-bounded autonomy grants · full immutable audit trail.

ISO 42001

AI Management System

Governance aligned to ISO 42001 — the international standard for AI management systems. Risk controls, transparency, and human oversight enforced in code, not policy documents.

RISK ASSESSMENT · PER DECISION

AI-scored, auto-escalating

Every approval scored 0.0–1.0 by a local AI classifier before it routes. Score > 0.7 auto-escalates to the next approval tier — no human judgment required to trigger it.

AUTONOMY GRADIENT

Earned, bounded, audited

Time-bounded autonomy grants (7-day max, 5-fleet cap). Every grant posts to #sentinel before the row is written — no silent elevation ever.

PHYSICAL GATES

Cannot be bypassed

Charter before active. Retrospective before close. The platform physically refuses to cross these gates — no workaround exists in the code.

WEEKLY JOURNAL

Mandatory + audited

Every agent files a weekly journal entry. Project lead finalises. CEO receives a Monday rollup. Missed weeks flag red on the dashboard.

AUDIT TRAIL

Full immutable log

Every approval, delegation, tool call, and LLM dispatch written with agent ID, timestamp, risk score, and delegated_by. Append-only.

Shadow AI Detection.

Four detectors hunt the rogue AI tools your staff installed behind IT's back — and the sanctioned-tool allowlist maintains itself from your live provider registry.

DETECTOR 01

Network

Spots unsanctioned model traffic crossing the wire.

DETECTOR 02

Endpoint

Catches AutoGPT, CrewAI and stray agent processes.

DETECTOR 03

Memory

Flags rogue model state living in process memory.

DETECTOR 04

Host sweep

Agentless SSH sweep for stray Ollama listeners.

Self-improving agents — Ultra-Instinct.

A full training loop with a mandatory human approval gate at its centre, and an automatic rollback if quality regresses.

One command centre. A whole company inside.

WORKSPACE

2D virtual office

Watch the fleet move and work across unit desks in real time.

MARKETS

Internal economy

Dual-token economy with a live share order-book and KPI valuation oracle.

OPERATIONS

CRM & pipeline

A 5-stage deal funnel with forecast, dedupe, and confidential ACLs.

STUDIO

Video pipeline

BlackOffice: AI moments → episodes → editorial approval → distribution.

INTELLIGENCE

Threat intel

Alerts, watchlists and a 7-state findings workflow with evidence.

INTEGRATIONS

15+ connectors

Exa, Perplexity, OSINT, browser, meetings, comms · Mattermost, Slack, Teams, WhatsApp, Telegram.

18 tools. Three transports. One audit log.

Invoke the AI fleet from your IDE, from the command centre, or from any connected chat platform — same guardrails, same classification checks, same immutable audit trail.

`MCP stdio`

IDE / Claude Desktop

Native MCP protocol — connect any compliant client and invoke the fleet directly from your workflow.

`/admin/mcp-tools`

Admin Panel

GUI surface with TOTP step-up modal for Tier-B sensitive writes. Full tool catalogue in one screen.

`Chat platforms`

Mattermost · Slack · Teams · WhatsApp · Telegram

Chat-native invocation from any connected platform. Tier-B step-up via DM — no context switch required for gated operations.

TOOL TIERS — 18 TOOLS TOTAL

Tier A · Composer write	3 tools
Tier B · TOTP-gated write	3 tools
Tier K · Knowledge ops	5 tools
Tier S · Safe reads	7 tools

All three transports call the same in-process composer wrappers. Classified-deny, DSP scan, jitter equalisation, and the canonical 404 envelope are identical regardless of how you invoke. Only `audit_log.source` differs.

Evidence over claims.

35
AI agents — 25 main-fleet + air-gapped classified tiers

7
team leaders · plus 1 manager

90+
backend services

800+
API endpoints

200+
database tables

97/100
production-readiness rating

18
MCP tools · 3 transports — IDE, admin panel, chat platforms

CB-1→5
graduated circuit breakers — soft pause to full fleet halt

~770K
tokens / day saved by token compression

DEPLOY IN YOUR ORGANISATION

Deploy your internal AI Management System.

Sovereign infrastructure. Governed agents. Full operator control — running in your environment, on your terms.

On-premise · air-gapped | EU sovereign | ISO 42001 ready | White-label available

info@blackunicorn.tech

BlackUnicorn
EU · TALLINN

BLACKUNICORN AI MANAGEMENT SYSTEM

Autonomous.
Sovereign.
Governed.

SEE THE DEMO bucc.blackunicorn.tech

COMMERCIAL SPINOFF

EGIDIA.AI · THE AGENTIC BUSINESS OS