

KOBIL AI Assistant Whitepaper

Date: 2026-08-13 (refreshed against `master` @ `cda8e15`)

Evidence baseline: repository and source notes available in the current AgentBox workspace, including the 2026-08-13 CTO Product FAQ/source notes, the mPower executive-summary PDF, and the June 2026 SuperApp MiniApp MCP integration proposal. This revision re-checks those notes against `master` at commit `cda8e15` (merge of `develop-agentbox-assistant` , plus one additional fix) and folds in the standalone-deployment architecture, which the source notes did not previously cover.

1. Executive Summary

KOBIL AI Assistant is the governed chat assistant embedded in the KOBIL SuperApp/mPower experience. It gives users a single conversational surface for asking questions, querying connected work tools, and building toward cross-MiniApp automation under the AgentBox runtime and security stack.

The current product is real, deployed, and usable as a SuperApp MiniApp: it has working SuperApp single sign-on, a working chat interface, audit events, and a live production connector portfolio. Jira, Confluence, Candor, Herald, Todo, Personio, Teams, Timebutler, and Kolay are live in production. GitHub engineering workflows are production-enabled, and the Knowledge Base/RAG path is implemented and production-ready for specialized domain assistants.

The target product is broader: one assistant that can eventually act across mPower MiniApps on the user's behalf, using that user's own permissions, with approval workflows that actually pause risky actions before execution. That target is not fully implemented today. Cross-MiniApp acting is a documented proposal, and current MCP-sensitive approval behavior is monitor/advisory only.

The strongest current value is not "autonomous action across every MiniApp." The strongest current value is a governed, SSO-integrated assistant surface inside mPower, backed by AgentBox's execution, connector, audit, and policy foundations.

2. Product Definition

KOBIL AI Assistant is the conversational entry point built into the KOBIL SuperApp, letting employees ask questions and, over time, take action across the tools and MiniApps they already use inside AgentBox's governed execution environment.

It is:

- A real SuperApp-embedded MiniApp.
- A governed chat/query assistant with a growing connector library.
- A productized instance running on the AgentBox runtime and security stack.
- A foundation for future cross-MiniApp automation.

It is not yet:

- A fully general cross-MiniApp action agent.
- A system where every risky action waits for human approval before execution.
- A fully live connector hub for every mPower partner app.

3. Customer Problem

Without a single assistant surface, SuperApp users move between separate MiniApps and work tools to find information, reconcile context, and complete tasks. Jira, Confluence, HR systems, partner MiniApps, messaging, documents, and operational workflows each expose separate entry points.

That creates three practical issues:

- Users do manual coordination work across many tools.
- MiniApp owners repeat identity, connector, and security decisions app by app.
- Security teams lack one consistent assistant-level record of what an AI surface attempted to do on a user's behalf.

KOBIL AI Assistant is intended to give mPower one identity-bound conversational surface, one connector-governance model, and one audit path for AI-assisted work.

4. Current Implementation Versus Target State

Current implementation

The current implementation is a live SuperApp MiniApp with:

- Real Keycloak OIDC and SuperApp silent SSO.
- A working chat UI.
- AgentBox-backed runtime adapters.
- Managed MCP routing for tools and connectors.
- Production-enabled GitHub engineering workflows.

- A production-ready Knowledge Base/RAG path for specialized agents, using SecureProxy-backed embeddings and Chroma retrieval.
- SecureProxy, MCP policy checks, Defender, skill scanning, workspace controls, and audit as the governed execution foundation.
- Audit logging.
- Approval queue UI and backend records for MCP-sensitive approval requests.
- A live production connector portfolio with governed paths for work tools and partner MiniApps.

Target state

The target state adds:

- Cross-MiniApp action execution, where other MiniApps expose callable interfaces for the assistant.
- Real per-user identity delegation across MiniApps where supported by SuperApp and IdP capabilities.
- Enforced human approval for risky actions, where execution pauses until approval.
- Additional partner connectors and MiniApp workflows using the same governed identity, policy, and audit patterns.

The distinction matters. Website, executive, and sales content should clearly label cross-MiniApp acting and enforced approval as target capabilities, not current product behavior.

5. How It Works

1. A user opens KOBIL AI Assistant inside the SuperApp webview.
2. The frontend starts silent OIDC authentication against the KOBIL Keycloak realm where configured.
3. A verified session establishes the chat context. The known cross-user session-leak class was addressed with fresh SuperApp verification on page load.
4. The user asks a question or requests work in chat.
5. Connector calls route through the managed MCP bridge. In configured real-mode cases, RFC 8693 token exchange can call downstream services with the user's own identity rather than a shared credential.
6. MCP-sensitive approval requests are recorded and shown in the approval queue.
7. Today, those approvals are advisory/monitor-mode for MCP tool calls. They are visible and auditable, but not a blocking execution gate.
8. Audit events are recorded for traceability.

6. Capability Readiness Matrix

Capability	Status	Current note
SuperApp SSO / OIDC identity	Implemented	Real Keycloak OIDC, PKCE, and silent login behavior are implemented.
Chat UI	Implemented	Working product UI with real auth and runtime behavior.
Audit trail	Implemented	Real audit events exist. No unsupported retention period should be claimed.
Jira and Confluence	Implemented via VPN tunnel	Live in production as of 2026-08-11. A reliability fix (error handling, ticket-priority correctness) shipped 2026-08-13.
GitHub engineering workflows	Production-enabled	Repository, issue, and pull-request workflows use the governed connector path.
Specialized agents / Knowledge Base RAG	Production-ready	Retrieval is implemented with SecureProxy-backed embeddings, per-knowledge-base Chroma collections, and governed MCP access. Staging enables the dedicated service; production activation requires the deployment configuration, key, and approved indexed knowledge.
GitLab source-control adapter	Next adapter	GitLab is present in repository hosting, registry, and CI configuration, but no separate GitLab MCP connector is evidenced yet.
CRM, Serper, Notify, PDF, RSS, Calendar	Implemented	Listed as live work-tool connectors in the source notes.
Teams	Live in production	Governed production connector access is live.
Candor, Herald, Todo	Live in production	Live MiniApp backends use real-mode, per-user token exchange through production connector paths.
Personio	Live in production	Real API integration and governed production access are live.
Timebutler / Kolay	Live in production	Governed production connector access is live.
Marketplace	Product expansion	A demonstration path exists; no confirmed live deployment is present in the evidence set.
Cross-MiniApp acting	Planned	Documented proposal; no MiniApp interface implementation exists today.
Human approval enforcement	Product rollout	UI and record keeping exist; enforced execution gating is not enabled.
Production deployment	Implemented	Production is live at agentbox.sicher.men with real KOBIL-realm OIDC.
Staging deployment	Implemented	Staging is live at staging-agentbox.shift-go.uk .

7. SuperApp And MiniApp Integration Model

The long-term model is for mPower MiniApps to expose structured tool surfaces that the assistant can discover and use through MCP. In that model, a MiniApp keeps ownership of its business logic and data, while KOBIL AI Assistant provides the conversational entry point, governance path, and audit trail.

The June 2026 proposal describes MiniApps acting as MCP providers. This would let the assistant call a MiniApp backend with user identity attached, receive structured results, and compose a response or repeatable flow.

Important implementation boundary:

- Current KOBIL MiniApp documentation does not show existing MCP support in MiniApps.
- Existing app-to-app identity and user-context capabilities need platform alignment.
- No current MiniApp exposes the cross-MiniApp acting interface required for the target model.

The proposed first pilots are workflow-like MiniApps where the value of one conversational surface is easy to prove: vacation requests, travel requests, expense reports, CapEx/investment proposals, on-call duty billing, and document signing.

8. Governance, Identity, Audit, And Security Posture

KOBIL AI Assistant inherits the AgentBox control model: governed runtime adapters, SecureProxy, Defender, TrustGate concepts, managed MCP routing, skill scanning, workspace controls, and audit events. These controls also surround specialized RAG agents, so model access, retrieval, and tool execution use the same governed paths. The assistant-specific security posture is strongest where traffic flows through registered, governed routes.

Identity:

- The assistant uses real SuperApp OIDC identity.
- Silent SSO is implemented for the SuperApp webview.
- Token exchange exists for selected configured real-mode connectors.

Audit:

- Interactions and tool-related events can be recorded.
- Approval-related records and UI exist.
- No specific audit-retention period should be claimed from the current evidence.

Human approval:

- Current approval queue behavior is useful for visibility and audit.
- It is not yet a hard blocking control for MCP tool calls.

- Content should avoid "human-in-the-loop gating" unless it explicitly says advisory/monitor-only today.

Compliance:

- KOBIL corporate security and privacy credentials should not be presented as standalone AgentBox or KOBIL AI Assistant product certifications unless separately confirmed.
- The assistant can support governance and evidence workflows, but it does not automatically create ISO 27001, SOC 2, GDPR, KVKK, or OWASP compliance.

9. Use Cases

Unified work-tool query

A delivery manager asks the assistant to summarize Jira tickets and Confluence context for the current sprint. This is the strongest near-term use case because Jira and Confluence are live in production through the governed connector path.

HR self-service

The assistant provides a single front door for HR questions and task preparation through live Personio, Timebutler, and Kolay production connectors.

Partner MiniApp query

Candor, Herald, and Todo use real per-user token exchange and live production MiniApp backends through governed connector paths.

Cross-MiniApp task completion

The target experience is a user saying what they need once, while the assistant routes work across the MiniApps that own the task. This is not implemented today. It depends on MiniApp-owned callable interfaces and platform identity alignment.

10. Deployment Status And Options

SuperApp deployment status

Staging is live at `staging-agentbox.shift-go.uk`.

Production is live at `agentbox.sicher.men` with real KOBIL-realm OIDC.

The evidence set records a 2026-08-12/13 production incident where stale image tags for MCP Teams/Jira/Confluence caused a crash loop and release rollback. That instance was fixed the same day. Do not imply that stale-image risk has been structurally eliminated unless separate hardening evidence is available.

Whether production is reachable outside KOBIL or internal-only is not determinable from the repository evidence. Confirm with product/IT before making availability claims.

Deployment beyond mPower

The SuperApp MiniApp is one deployment mode of a shared assistant stack, not the only one the architecture supports. The same Helm chart that deploys the SuperApp-embedded mode also ships two additional, already-defined overlays:

- **Self-hosted, customer-operated** (`values-customer.yaml`): the identical chat/skills/flows/knowledge-base/admin stack, deployed on a customer's own Kubernetes cluster behind a customer-chosen ingress hostname, with the customer selecting which runtimes and MCP connectors are enabled.
- **KOBIL-hosted, standalone SaaS** (`values-saas.yaml`): a KOBIL-hosted, multi-tenant deployment of the same stack at a KOBIL-owned hostname, independent of any SuperApp embedding.

In both cases, the product is reached directly in a browser rather than inside the mPower webview. This works because the frontend's sign-in flow already has a first-class local email/password path (not a fallback hack) that activates automatically whenever SuperApp SSO is absent — the same login component that serves the MiniApp handles standalone access without modification. No native desktop application exists anywhere in the codebase; "standalone" today means zero-install browser access, on any device, not a separate desktop build.

Evidence boundary: the `values-customer.yaml` and `values-saas.yaml` overlays exist and are wired into the same Helm chart as the live staging/production SuperApp deployment. No evidence in this repository confirms an actual, currently-running self-hosted customer cluster or a live multi-tenant SaaS tenant at `agentbox.kobil.com`. State this as "the deployment architecture already supports it," not as an existing live customer or SaaS footprint, until that is separately confirmed.

11. Metrics And Evidence Boundaries

Repository-backed metrics available in the evidence set:

- 147 frontend tests across 30 files.
- 2 Playwright end-to-end specs.
- 68 backend identity/session tests.
- 140 backend `test_*.py` files repository-wide.
- 30 MCP connector services registered in the demo compose registry.

Do not claim:

- Number of SuperApp users.
- Production request volume.
- Latency, uptime, or adoption.
- Customer/pilot counts.
- Security-incident reduction attributable to the assistant.
- A specific audit-retention window.

12. Roadmap And Open Dependencies

The next meaningful work areas are:

- Turn approval monitor mode into enforced blocking mode for risky tool calls.
- Activate the dedicated Knowledge Base/RAG service in production with the required SecureProxy key, approved indexed content, and authorization boundary.
- Add a dedicated GitLab source-control adapter if GitLab parity is required beyond the current GitHub production connector.
- Expand the production connector portfolio to additional high-value systems and MiniApp workflows.
- Align with SuperApp and IdP teams on app-to-app identity.
- Select one MiniApp pilot and define its MCP/provider interface.
- Add stronger frontend component and end-to-end test coverage for the Assistant MiniApp.
- Harden deployment image freshness and release verification after the stale-image incident.
- Stand up and validate a real reference deployment on the self-hosted (`values-customer.yaml`) and/or KOBIL-hosted SaaS (`values-saas.yaml`) overlays, so the standalone path can be described as live rather than architecture-only.

13. Recommended Website Messaging

Hero:

KOBIL AI Assistant

One governed assistant inside the SuperApp. Ask questions, query connected tools, and build toward cross-MiniApp workflows with identity, audit, and connector governance from AgentBox.

Current-state subcopy:

Live in staging and production with SuperApp SSO, working chat, audit events, and selected live work-tool connectors. Cross-MiniApp acting and blocking human approval are product targets, not current default behavior.

Primary sections:

- Governed SuperApp entry point
- Connector readiness
- Specialized agents, Knowledge Base/RAG, and source-control workflows
- SecureProxy, managed MCP policy, skill scanning, and audit controls
- Indicative commercial model: €4,500 workflow assessment, €18,000 fixed-scope 8-12 week production pilot, and enterprise deployment from €72,000 per year
- Current versus target state
- Identity and audit path
- MiniApp integration model
- Deployment beyond mPower (self-hosted and KOBIL-hosted standalone modes)
- Roadmap and dependencies
- FAQ with precise caveats

CTA:

Align on the first production MiniApp pilot

Commercial messaging:

- Start with a €4,500 one-time workflow assessment or an €18,000 fixed-scope 8-12 week production pilot.
- Move to an annual KOBIL-hosted or customer-hosted enterprise agreement from €72,000 per year.
- Scope commercial terms against deployment model, active users, connectors, workflows, support, and model-provider consumption.
- Label all public figures as indicative, excluding VAT, infrastructure, model-provider usage, and exceptional integration work.

14. Claims To Avoid

Avoid these claims unless new evidence is supplied:

- "The assistant can complete tasks across your other apps."
- "Risky actions require human approval."
- "Human-in-the-loop approval blocks every risky tool call."
- "All mPower partner apps are connected."

- "Every connector uses live production data."
- "Audit logs are retained for 90 days."
- "The product is certified for ISO 27001/SOC 2/GDPR/KVKK/OWASP."
- "The stale-image incident cannot happen again."
- "Production availability is public" or "production is internal-only" without product/IT confirmation.
- "KOBIL AI Assistant only works inside mPower" or "requires the SuperApp" — the standalone (self-hosted and KOBIL-hosted) deployment modes are real, already-defined parts of the architecture.
- "We have live self-hosted customers" or "our SaaS is live" — the overlays exist; a running self-hosted or SaaS deployment is not evidenced yet.
- "KOBIL AI Assistant has a desktop app" — no desktop/native build exists; the accurate claim is browser-based, zero-install access.
- Definitive or discounted pricing beyond the approved indicative figures without separately approved commercial terms.