

AI Support Agent — bilko.cloud (HR)

AI Support Agent — bilko.cloud (HR)

Status: Phase 1 operational (internal copilot); Phase 2 customer-facing widget HELD for Securion review

MC Task: #104429

Owner: John → CodeCraft (Petter Graff lead)

Built: 2026-06-28

1. Purpose & Scope

The **Bilko Cloud AI Support Agent** is a dedicated support system for the `bilko.cloud` (Croatia) jurisdiction of the Bilko SaaS platform. It handles **both technical problems** (login/Entra SSO, e-invoice/OIB failures, RLS, deploy errors) **and user/accounting questions** (PDV rates, HR-FISK/eRačun, invoices, bookkeeping, general ledger) — grounded in the Bilko knowledge base, validated by HR accounting domain experts, and gated by anti-hallucination verification.

Phased deployment:

- **Phase 1 — Internal copilot** (current): Assists ALAI/SnowIT support staff via ticketing workflow; human approves and sends answers. No direct customer exposure.
- **Phase 2 — Customer-facing widget** (planned, HELD): In-app chat widget on `bilko.cloud` that auto-drafts answers in `triageJson` with guardrails (PII allowlist, prompt-injection guard, scope, escalation). Requires Securion security review before go-live.

Jurisdiction: Croatia only. The agent reasons under **HR tax/accounting rules** and answers in **Croatian (hr-HR)**.

Cost target: \$0 per question (runs on local tier-2 models: Qwen 2.5 Coder 32B / MLX).

2. The Team — 5 Agents

Role	Agent	Registry	Job
Lead architect	Petter Graff	~/ .claude/agents/petter-graff.md	Owns architecture, gates phases, integration design
Research (Phase 0)	Datavera / Explore	ephemeral dispatches	HR support taxonomy, competitor patterns, KB extraction (completed)
Support persona	bilko-support	~/ .claude/agents/bilko-support.md	Orchestrates answer pipeline, enforces anti-hallucination, escalation
HR accounting expert #1	bilko-racunovodstvo-hr	~/ .claude/agents/bilko-racunovodstvo-hr.md	Dvojno knjigovodstvo, RRiF chart of accounts, GL, expenses, financial statements (RDG/bilanca); KB author + validation gate
HR tax/fiscalization expert #2	bilko-porez-fiskalizacija-hr	~/ .claude/agents/bilko-porez-fiskalizacija-hr.md	PDV (25/13/5/0%), HR-FISK 2.0/eRačun (UBL 2.1/EN 16931), OIB validation, FINA, filing deadlines; KB author + validation gate

All three personas (bilko-support + 2 experts) registered in ~/system/agents/specialist-mapping.json :

- bilko-support → company ALAI , domain product-support
- bilko-racunovodstvo-hr + bilko-porez-fiskalizacija-hr → company Finverge , domain hr-accounting

3. Architecture — Answer Pipeline

Orchestrator: ~/system/tools/bilko-support-answer.js (305 lines, CLI + module.exports).

Pipeline (8 steps, deterministic):

```
question (+ contextBundle: errorCode, appRoute, country=HR, planTier, orgId...)
↓
[1] classify + route          tier-router.js (local $0 model)
↓
[2] cache check              rag-router.js (flywheel.db hit? early return)
    □ if cached + confidence >= 0.70 (or not acct/tax/e-invoice) → DONE
    □ if cached BUT classification ∈ {accounting, tax, e-invoice}
```

```

        AND confidence < 0.70 → escalate:true, return
↓
[3] retrieve context          KB JSONL search (40 Q&A pairs)
                             + LightRAG hybrid (knowledge.db)
                             + HiveMind semantic (hivemind.db, fire-and-forget)
↓
[4] draft answer (Croatian)  tier-2 local model (Qwen/MLX)
↓
[5] anti-hallucination gate  mini-verifier.js
    □ HALLUCINATION (cited file missing/empty) → escalate
    □ DRIFT (file present but ambiguous/language mismatch) → reduce confidence
    □ CONFIRMED (grounded) → boost confidence
    □ SKIP (verifier offline) → neutral
↓
[6] accounting validation    if classification ∈ {accounting, tax, e-invoice}:
    □ re-verify vs domain-specific KB (hr-tax-fiscalization, hr-accounting sources)
    □ if no domain KB → escalate
    □ if HALLUCINATION or CONFIRMED < 0.70 → escalate
↓
[7] confidence floor gate    belowAccountingFloor(cls, confidence)
    □ if acct/tax/e-invoice AND confidence < 0.70 → escalate:true
    (same helper used on BOTH cache path [L139] and pipeline path [L258])
↓
[8] cost log + return        {answer, confidence, evidence_paths,
                             classification, severity, escalate, source}

```

Output schema:

```

{
  "answer": "",
  "confidence": 0.0–1.0,
  "evidence_paths": [],
  "classification": "technical | accounting | tax | e-invoice | billing | onboarding | other",
  "severity": "P1 | P2 | P3 | P4",
  "escalate": boolean,
  "source": "cached | local | escalated"
}

```

4. Escalation Policy

Hard rule (enforced in code): `ACCOUNTING_CONF_FLOOR = 0.70`

Any answer with classification $\in$ `{accounting, tax, e-invoice}` AND confidence < 0.70 MUST escalate to a human specialist (or to `bilko-racunovodstvo-hr` / `bilko-porez-fiskalizacija-hr` personas in async workflow).

Escalation triggers:

- Verifier verdict = `HALLUCINATION` (fabricated citation)
- Unknown question (no KB match, no LightRAG answer)
- Accounting/tax claim: no domain-specific KB evidence OR domain-verified confidence < 0.70
- Confidence < 0.70 on cache hit for acct/tax/e-invoice (uses same `belowAccountingFloor()` helper as pipeline path)
- Severity P1 (critical production blocker)

Single source of truth: The `belowAccountingFloor(cls, confidence)` function (line 62) is called on BOTH the cache early-return path (line 139) and the post-step-6 gate (line 258) — guarantees no drift in threshold logic.

5. Knowledge Base — 40 Source-Cited Q&A Pairs

Storage (current): JSONL files read directly by orchestrator; ingestion into `lightrag/knowledge.db` is a pending enhancement.

- `~/system/reports/bilko-kb/hr-tax-fiscalization.jsonl` — 20 pairs (PDV rates/deadlines, HR-FISK 2.0/eRačun B2B mandatory 2026-01-01, OIB validation ISO 7064 MOD 11,10, FINA rejection reasons, B2G since 2019-07-01, UBL 2.1/EN 16931, EUR since 2023-01-01)
- `~/system/reports/bilko-kb/hr-accounting.jsonl` — 20 pairs (RRiF chart classes 0-7, GL journal on invoice lifecycle sent→paid, expense approve→paid, storno/cancel logic, tečaj lock, RDG/bilanca/trial balance, fiscal year 1.1.–31.12., audit trail 7y retention)

Sources cited:

- `~/business/ALAI-Holding-AS/products/Bilko/docs/regulatory/CROATIA-ERACUN.md` (authoritative HR doc, HIGH confidence)
- `~/business/ALAI-Holding-AS/products/Bilko/packages/domain-hr/src/tax/index.ts`
- `~/business/ALAI-Holding-AS/products/Bilko/packages/domain-hr/src/chart/index.ts` (RRiF CoA implementation)
- `~/business/ALAI-Holding-AS/products/Bilko/packages/domain-hr/src/fisk/index.ts` (OIB validateOIB)
- `~/business/ALAI-Holding-AS/products/Bilko/packages/domain-hr/src/filing/index.ts` (eRačun UBL 2.1 generation)

- `~/business/ALAI-Holding-AS/products/Bilko/docs/backend/BUSINESS-LOGIC.md`

Each entry has: `{q, a, source, confidence: "HIGH" | "MEDIUM", domain: "hr-tax-fiscalization" | "hr-accounting"}`.

Confidence levels:

- **HIGH** — direct from CROATIA-ERACUN.md or implemented code (domain-hr packages)
- **MEDIUM** — inferred deadlines or regulatory interpretations (recommend user confirm with Porezna uprava / FINA / certified accountant)

6. Integration with Bilko Backend

Existing infrastructure (reused, not built for this):

- `apps/api/src/routes/SupportTicketRoutes.kt` (437 lines)
 - `support_tickets` table (org_id, subject, description, status open/in_progress/resolved/closed, priority P1-P4, created_at, updated_at)
 - `POST /support/tickets` — idempotency key (`org_id, request_id`)
 - `GET /admin/support/tickets`, `PATCH /admin/support/tickets/:id`
 - PII-guarded `context_bundle` field (errorCode, appRoute, planTier, locale — allowlist enforced server-side)
 - **Reserved field:** `triageJson` (line 62 comment: "for AI assistant V2") — this is the Phase 2 hook

Phase 1 workflow:

1. Staff create/triage ticket via `~/system/tools/support-ticket.js` (SLA P1-P4)
2. Call `node ~/system/tools/bilko-support-answer.js "" ''`
3. Agent returns JSON with answer + evidence + classification + escalate flag
4. Human reviews, approves, sends to customer (via email / admin panel)

Phase 2 workflow (planned, HELD):

1. Customer opens in-app chat widget on `bilko.cloud`
2. Widget calls `POST /support/tickets` with `context_bundle`
3. Backend triggers `bilko-support-answer.js` (via internal job queue or sync call)
4. Agent output written to `triageJson` field: `{suggestedAnswer, classification, severity, escalate, confidence, evidencePaths}`
5. Render in UI: if `escalate:false` AND `confidence >= 0.70` → show AI answer with citation; else → "Eskaliran na podršku, očekujte odgovor u "
6. Guardrails: PII allowlist (already enforced), prompt-injection guard (add Securion layer), scope validation (no cross-org queries)

Security note: Phase 2 requires Securion/Parisa Tabriz security review before live deployment (MC follow-up task).

7. Product Bugs Found by Experts (During KB Build)

The 2 HR accounting experts found **8 Bilko product/documentation defects** while authoring the knowledge base:

MC	Severity	Issue
#104441	P3	Bilko MVP GL: only 1 D/P per transaction → PDV not split to konto 450 Obveze za PDV (tracking via reports only, not GL)
#104446	P3	OIB validation: CROATIA-ERACUN.md missing ISO 7064 MOD 11,10 detail (in code, not doc)
#104447	P2	CoA setup friction: not available in wizard, only post-onboarding in Settings (USER-ONBOARDING.md gap)
#104442	P3	Storno logic: cannot cancel paid invoices (needs credit note, not auto-generated in MVP)
#104448	P4	BUSINESS-LOGIC.md: HR PDV table lists only 25%/13%, omits 5%/0% (available via free input, just not in doc table)
3 additional minor doc inconsistencies logged		

These are now tracked separately and do not block the support agent (KB answers work around them with "Bilko MVP limitation" disclaimers where needed).

8. Runbook — How to Invoke

CLI:

```
node ~/system/tools/bilko-support-answer.js "<question>" '<contextBundleJSON>'
```

Example:

```
node ~/system/tools/bilko-support-answer.js \  
  "Koja je stopa PDV-a na knjige u Hrvatskoj?" \  
  '{"country":"HR","planTier":"PRO","locale":"hr-HR"}'
```

Output (JSON to stdout):

```
{  
  "answer": "Stopa PDV-a od 5% primjenjuje se na knjige...",  
  "confidence": 0.78,  
  "evidence_paths": ["/Users/makinja/.../CROATIA-ERACUN.md"],  
  "classification": "tax",  
  "severity": "P3",  
  "escalate": false,  
  "source": "local"  
}
```

Module API:

```
const { answer } = require('~/system/tools/bilko-support-answer');  
const result = await answer(question, contextBundle);
```

Cost tracking (automatic): Each call logs to `~/system/tools/cost-tracker.js` with `{source: 'bilko-support-agent', backend: 'ollama', model: '...', cost_usd: 0, duration_ms, metadata}`.

9. Open Follow-Ups

MC	Task	Owner
#104432	Proveo UAT — end-to-end verification: empty-KB degradation, offline-model fallback, severity classification coverage, schema validation, real ticket replay	Proveo / Angie Jones
#104430	Phase 1 deployment — staff onboarding, <code>support-ticket.js</code> workflow integration, SLA monitoring	FlowForge / Kelsey Hightower
#104431	Phase 2 customer-facing widget — frontend component, POST hook, guardrails (HELD pending Securion)	Vizu / Brad Frost + Securion / Parisa Tabriz

MC	Task	Owner
(backlog)	KB ingestion into <code>lightrag/knowledge.db</code> (enhancement: replace JSONL direct-read with Neo4j-backed hybrid retrieval)	AgentForge / Chip Huyen
(backlog)	Securion review: confidence thresholds, cache-of-escalated-answers audit, PII leak surface, prompt-injection tests	Securion / Parisa Tabriz

10. Verification Evidence

MC #104429 verdict: `PASS` (2026-06-28, john verified)
Method: Company-Mesh (CodeCraft built, John verified via real CLI runs)
Evidence: `~/system/evidence/104429/verdict.json`

Verified behaviors:

- Tax Q 'PDV na knjige' → 5% grounded (CROATIA-ERACUN.md + domain-hr/src/tax)
- Unknown Q → `escalate:true`, `source:escalated`
- Pipeline tax conf 0.62 → `escalate:true`
- Cache tax conf 0.6547 → `escalate:true` (floor enforced on cache path)
- Cache tax conf 1.0 → `escalate:false` (no over-escalation)
- `belowAccountingFloor` defined once (L62), called L139 (cache) + L258 (pipeline) — no threshold drift

2 under-escalation defects caught & fixed: Sub-0.70 tax answers on both pipeline AND cache paths were initially missing the confidence floor check; John's independent runs caught these, and CodeCraft fixed them via the shared `belowAccountingFloor()` helper.

This page is the deliverable for MC #104433 (ZAKON PLAN mandatory documentation task). Last updated: 2026-06-28.