

Employee questions on HR policy

THE WORKFLOW AS IT RUNS TODAY

≈ 200 questions a month · ≈ 70% concern a written policy · ≈ 2 days average response time (illustrative, approximate figures)

● friction point



BEFORE YOU START

Get the policies in order.

Identify the documents that apply, resolve contradictions, assign an owner to each document, and date each version. The tool comes after.

UNDER THE HOOD

A RAG architecture: hybrid search (vectors + BM25) filtered by access rights, reranking, and generation by an LLM instructed to rely on the retrieved passages and cite its source, with input and output guardrails and routing to HR. A learning loop connects the log, HR review, and the evaluation set. Diagram and controls on page 2.

STAGE 1 HUMAN IN THE LOOP

The assistant drafts, the HR team decides

- 1 The question comes in — to the same HR inbox as today.
- 2 The assistant drafts a reply — citing the section and its version.
- 3 An HR team member decides — reviews, edits if needed, then sends.
- 4 Every correction is logged — topic by topic.

What this stage produces: a topic-by-topic measure of how good the drafted answers are. Estimated duration: 8 to 12 weeks.

STAGE 2 THE ASSISTANT HANDLES THE STANDARD CASES

It answers on approved topics; HR handles the exceptions

ANSWERS WHEN

- The question falls under an approved topic
- A general, written, current policy applies
- The answer can cite the section and its version

ROUTES TO HR WHEN

- The situation is individual, sensitive, or ambiguous
- The source is missing, contradictory, or outdated
- The question involves topics such as health, conflict, pay, extended leave, or termination

Every week, HR reviews a sample of answers and keeps the sources up to date.

WHY HERE, AND UNDER WHAT CONDITIONS

High volume · Clear owner · Clear output · Known pain point

- General policies written, approved, and current.
- No individual employee files used as a source.
- No HR decisions or judgment calls handed to AI.
- Individual, sensitive, or ambiguous situations routed to HR.

The risk does not go away, but it becomes more visible, measurable, and manageable.

MOVING FROM STAGE 1 TO STAGE 2

Decided topic by topic, never in advance, based on what the HR team observes during stage 1:

- Accuracy of the answer and the citation
- Corrections needed, by topic
- Exceptions correctly routed
- No material incident within the tested scope
- Current sources, each with an owner

Vacation and leave could move to stage 2 while expenses stay in stage 1.

THE GOAL, AND HOW TO MEASURE IT

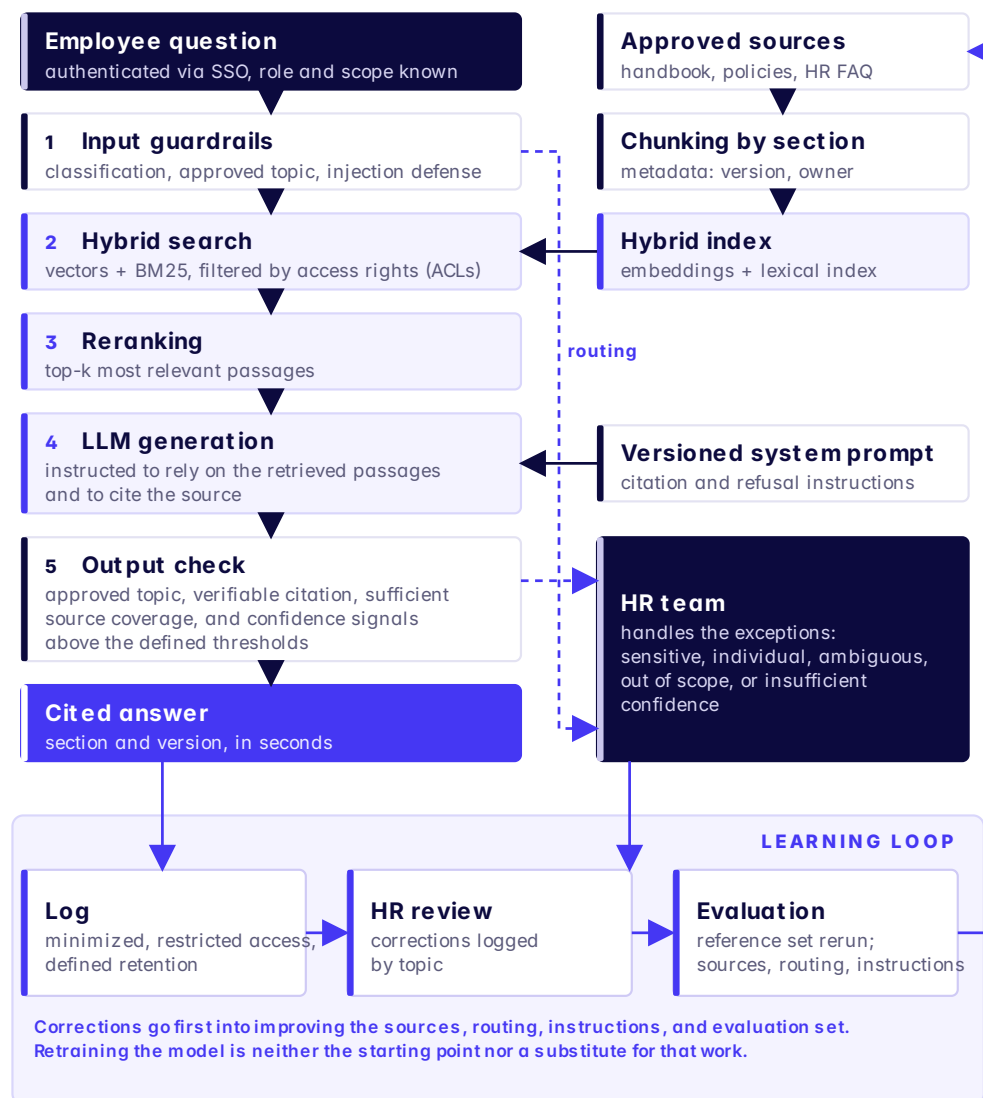
Not to replace an HR decision, but to cut response times on general-policy questions, stop rewriting the same answers, and free up HR time for situations that call for judgment.

Response time on general-policy questions

≈ 2 days → in seconds

today (illustrative) → stage 2 target · also tracked: the share of questions handled without HR involvement, 50–60% (illustrative target)

Reference architecture: how the system stays under control



CONTROLS AND SECURITY

Evaluation

A reference set of about 100 real questions. Metrics: accuracy, faithfulness to sources, correct citation, and the rate at which exceptions are detected and correctly routed (routing recall). Rerun whenever the model, prompt, or corpus changes.

Observability

Every request is logged: question, retrieved passages, answer, confidence signals, routing decision. Because questions may contain personal data, the log is minimized, accessible only to authorized staff, and kept for a defined period.

Identity and access

SSO (SAML or OIDC). Role-based access control enforced at retrieval, not after generation.

Data

Hosting and data transfers verified for the chosen solution; contractual terms barring use of the data for model training; encryption; log retention and access defined; Swiss FADP compliance verified.

Application security

Protection against prompt injection. The assistant can only answer or route to HR.

Three ways to build it

Three options: the assistant built into your existing office suite, a managed RAG platform, or a custom build (orchestration, vector database, model API). The choice comes after the source audit, not before.