Triaging and routing IT tickets

Sergio Castagna Consulting.

Illustrative case · Example Industries Inc. · 900 employees · fictitious company

THE WORKFLOW AS IT RUNS TODAY

≈ 1,500 tickets a month · **≈ 40%** are simple, recurring requests · **≈ 1 day** until a simple request is picked up (illustrative, approximate figures)

● friction point

STEP 1	STEP 2	STEP 3	STEP 4	STEP 5	STEP 6
Ticket opened by the employee	**First-line triage**	**Search for a fix**	**Assignment to a team**	**Handling by a technician**	**Closure and filing**
	● Overloaded shared queue	● Scattered or outdated articles	● Tickets reassigned several times	● The same fix explained again	● Categories hard to analyze

BEFORE YOU START

Get the catalog in order.

Clarify categories, the priority matrix, and teams; retire outdated articles; give each article an owner and a version date. The tool comes after.

UNDER THE HOOD

Ticket classification with a confidence score, priority set by a fixed-rule matrix, hybrid search across the articles, and a suggestion drafted by an LLM instructed to cite the article used, with input and output guardrails and escalation to technicians. No action on any account. A learning loop connects the log, support review, and the evaluation set. Diagram and controls on page 2.

STAGE 1 HUMAN IN THE LOOP

The assistant suggests a triage; the technician decides

1. **The ticket comes in** — through the same support tool as today.
2. **The assistant suggests a triage** — category, priority, team, and cited article.
3. **The technician decides** — confirms, corrects, or reassigns.
4. **Every correction is logged** — category by category.

What this stage produces: a category-by-category measure of how accurate the suggested triage is.
Estimated duration: 6 to 10 weeks.

STAGE 2 THE ASSISTANT ROUTES STANDARD REQUESTS

It routes or answers standard requests; support handles the exceptions

ROUTES OR ANSWERS WHEN

- The request falls in an approved category
- A current article or a secure workflow applies
- The team and priority follow from written rules

ESCALATES TO SUPPORT WHEN

- The request is an incident, urgent, or ambiguous
- The article is missing, conflicting, or outdated
- It involves security, privileged access, a business-critical device, multiple users, or similar

Every week, support reviews a sample of routed tickets and keeps the articles up to date.

WHY HERE, AND UNDER WHAT CONDITIONS

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

- **A written service catalog** current categories, priorities, and teams.
- **No admin rights** and no account actions taken by AI.
- **Password resets and standard access requests** through the existing secure workflows.
- **Incidents, security issues, and ambiguous cases** routed to technicians.

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

MOVING FROM STAGE 1 TO STAGE 2

Decided category by category, never in advance, based on what support observes during stage 1:

- Accuracy of the triage and the cited article
- Reassignments needed, by category
- Exceptions correctly routed
- No significant incident within the tested scope
- Current articles, each with an owner

Printer requests could move to stage 2 while application access stays in stage 1.

THE GOAL, AND HOW TO MEASURE IT

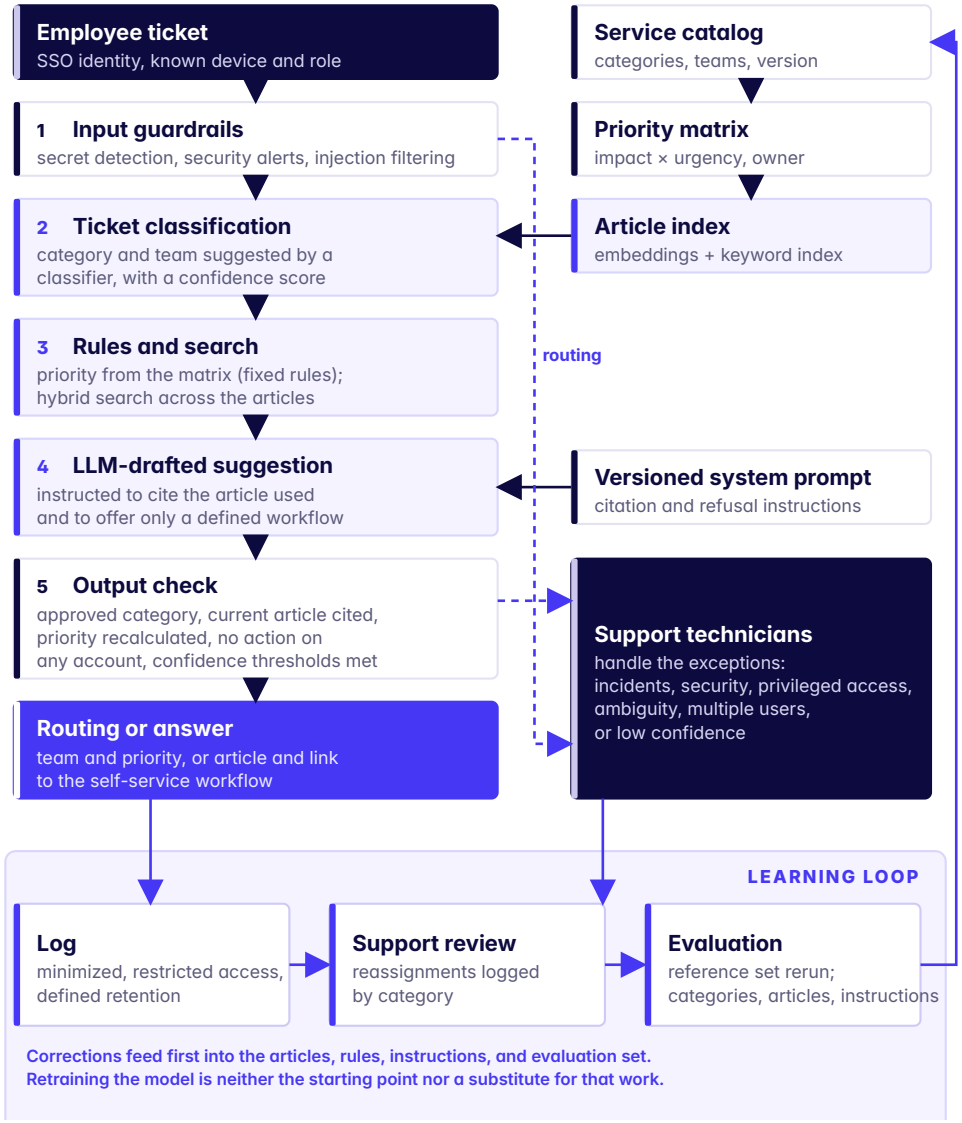
Not to replace technicians, but to cut response time on simple requests, stop redoing triage by hand, and free up time for incidents that call for diagnosis.

Time until a simple request is picked up

≈ 1 day → minutes

today (illustrative) → illustrative stage 2 target · also tracked: the share of simple requests routed or answered without a technician, 30–40% (illustrative target)

Reference architecture: how the system stays under control



CONTROLS AND SECURITY

Evaluation

A reference set of about 300 past tickets, minimized, each with the expected category, team, and priority. Metrics: triage accuracy, reassignment rate, citation accuracy, and the rate at which exceptions are detected and correctly routed (routing recall). Rerun whenever the model, the catalog, or the instructions change.

Observability

Every ticket processed is logged: category, priority, suggested article, confidence score, routing decision. Since tickets can contain personal data or secrets, the log is minimized, accessible only to authorized staff, and kept for a defined period.

Identity and access

SSO (SAML or OIDC). The assistant has no admin rights: password resets and access requests go through the existing workflows, including MFA and approvals.

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; documented Swiss FADP assessment before launch.

Application security

Prompt injection defenses for ticket text; attachments, if analyzed, are extracted in an isolated environment. Only four outcomes: suggest, route, answer, or escalate. No command is ever run on any system.

Three ways to build it

Three options: the AI module in your existing service desk tool, a managed classification and search platform, or a custom build (classifier, rules engine, model API). The choice comes after the catalog audit, not before.