PUBLIC PROJECT DOCUMENT / VERSION 0.1.0

# Governed AI Service Desk

AI system card

**CURRENT STATUS**

Locally tested portfolio prototype using a deterministic mock provider and six synthetic IT articles. Azure architecture is designed in code; no Azure service or cloud control has been deployed or validated.

## Purpose

A narrow assistant for low-risk employee IT support questions. It demonstrates request policy, approved knowledge grounding, citations, human escalation, and minimized application audit metadata. It provides guidance only; it cannot change accounts or execute tools.

| PROFILE | BOUNDARY |
|---|---|
| **Owner / state** | Paul Mwangi / design and local prototype |
| **Intended users** | Employees seeking guidance; service-desk and security staff reviewing escalations; portfolio reviewers. |
| **Approved topics** | Password-reset process, phishing and lost-device reporting, VPN, software requests and MFA method replacement. |
| **Excluded uses** | Real employee records; Restricted data, secrets or credentials; legal, HR, medical, financial or safety advice; access or employment decisions; actions or commands. |

## What happens in a local request

| STAGE | OBSERVED BEHAVIOR |
|---|---|
| **Access** | The local demo role gates chat and audit endpoints. An employee role is denied audit access in an API test. |
| **Policy** | Selected Restricted-data, likely-secret, prompt-injection and exfiltration patterns are blocked; recognized PII formats are redacted. |
| **Grounding** | Weighted retrieval selects approved synthetic articles. Unsupported requests are directed to the human service desk. |
| **Response and audit** | Responses include citations and a request ID. Application audit events contain hashes and decision metadata, without intentionally storing raw prompt or response text. |

AI SYSTEM CARD / EVIDENCE STATUS

# Controls and limits

Status reconciled to the project evidence ledger last verified 15 July 2026.

## Implemented and locally checked

| AREA | EVIDENCE |
|---|---|
| **Policy and privacy** | Unit tests cover normal requests, Restricted classification, selected PII redaction and one prompt-injection pattern; code implements additional likely-secret and exfiltration checks. |
| **Role and audit** | API tests check employee audit denial and absence of raw prompt text from the application audit event. |
| **Grounding** | Tests check a cited answer, unsupported-question escalation and removal of weak article matches. |
| **Quality gate** | The ledger records 10 passing automated tests and a clean Ruff check. This public edition did not re-run the suite. |

## Designed for Azure; not cloud validated

| DESIGN ELEMENT | STATUS |
|---|---|
| **Entra and app roles** | Specified in the architecture and Bicep. No tenant or negative-access test evidence. |
| **Managed identity and Azure RBAC** | Represented in infrastructure code and an Azure OpenAI adapter using DefaultAzureCredential. No deployment or role-assignment evidence. |
| **Azure services** | Container Apps, Azure OpenAI, Key Vault, Storage and monitoring are designed. The Bicep template was not compiled, the container was not built and no Azure resource was deployed. |

## Limitations and release boundary

Pattern checks cannot detect every secret, PII format or prompt attack. Keyword retrieval can miss context; a model-backed answer can be wrong. The local audit file is demonstration-only. No complete DLP, content-safety, SIEM or human-approval workflow is implemented. No formal security, privacy, legal, penetration or internal-audit review is recorded; an accessibility assessment is also needed.

Before real organizational use: appoint accountable reviewers; evaluate privacy, security, legal and accessibility obligations; deploy to an isolated lab; verify identity, RBAC, provider access, durable audit, retention, monitoring and incident handling; expand abuse and answer-quality testing; and approve human escalation service levels.

**CLAIM BOUNDARY**

This work demonstrates local implementation, test evidence and Azure design intent. It does not establish production readiness, formal compliance, certification, cloud control effectiveness or measured business impact.

Source basis: AI_SYSTEM_CARD.md, README.md and EVIDENCE_STATUS.md from the Governed AI Service Desk project. This public edition clarifies which controls were locally tested and which remain cloud design.