

**DABUDA LIMITED**

AI ASSURANCE EVIDENCE SERIES

# Council AI Assurance Case Study

Illustrative assurance model for a resident-facing AI assistant across housing, benefits, and service enquiries.

**ILLUSTRATIVE CASE STUDY - NOT A NAMED CLIENT ENGAGEMENT**

Version: April 2026

Tone: Executive, evidence-led, procurement-ready

---

Prepared for controlled review. Not a certification, endorsement, or named-client reference.

## ILLUSTRATIVE CASE STUDY

## Assuring a Resident-Facing AI Assistant for a Council

**Document status**

This case study is illustrative. It is designed to show DaBuDa methodology and evidence outputs. It does not represent a named client engagement.

### Use Case Summary

A local authority plans to deploy an AI-powered assistant to support residents with service enquiries, housing requests, benefits guidance, complaints routing, and general support information through web and mobile channels.

Dimension	Assurance boundary
Primary users	Residents seeking information or routing support.
Permitted use	Low-risk information retrieval, triage, signposting, and service navigation.
Restricted use	Eligibility decisions, statutory determinations, safeguarding decisions, appeals, or complex case handling.
Required control	Defined human escalation for sensitive, uncertain, high-risk, or legally significant interactions.

### Resident-Facing Risks

Risk	Potential service impact	Assurance focus
Incorrect or misleading information	Residents may take wrong action or miss support routes.	Accuracy testing against approved council source content.
Inconsistent responses	Similar residents may receive different guidance.	Repeatability checks across paraphrased scenarios.
Lack of transparency	Residents may not understand system boundaries.	Clear disclosure, limitations, and escalation language.
Potential bias	Certain groups may receive poorer service outcomes.	Diverse scenario sets and equality impact review.
Failed escalation	Sensitive cases may be mishandled by automation.	Human handoff trigger validation and routing checks.

### Scenario Testing Approach

DaBuDa applies structured scenario-based testing to evaluate the assistant in realistic resident journeys, ambiguous cases, edge cases, and misuse scenarios.

Scenario category	Example test coverage
Common user journeys	Housing repairs, council tax questions, benefit signposting, complaints, service updates.
Edge cases and ambiguity	Incomplete information, conflicting intent, vague requests, multi-issue prompts.
Vulnerable user scenarios	Financial hardship, disability access needs, mental health indicators, safeguarding-adjacent language.
Adversarial and misuse testing	Prompt injection, attempts to bypass escalation, requests for prohibited decisions.

### Evaluation dimensions

- Accuracy against approved source material.
- Consistency across equivalent resident queries.
- Appropriateness of tone, content, and disclaimers.
- Correct escalation and handoff behaviour.
- Accessibility and comprehension of resident-facing outputs.

## Escalation and Handoff Validation

The assurance process validates that the AI assistant does not continue handling cases that require human review, specialist judgement, or statutory decision-making.

Validation area	Expected control evidence
Sensitive case recognition	Trigger list, scenario outcomes, false-negative review, and issue log.
Correct routing	Mapping between user intent, council service owner, and escalation path.
Automation boundary enforcement	Test evidence showing the assistant refuses or defers prohibited decisions.
Resident communication	Clear wording that explains next steps and avoids pretending to make official decisions.

## Transparency and Governance Evidence

Evidence artefact	Purpose
Test coverage report	Shows which resident journeys, risk categories, and service areas were tested.
Scenario outcome logs	Records prompt, output, expected result, assessor judgement, and repeat-test result.
Risk and failure register	Captures known limitations, severity, owner, mitigation, and release impact.
Explanation samples	Demonstrates how the system communicates boundaries, uncertainty, and escalation.
Stakeholder decision record	Documents who approved pilot release, what evidence was reviewed, and what residual risk was accepted.

## Example Release Recommendation

### Status: Controlled Pilot Release

- Deployment limited to low-risk service categories and approved knowledge sources.
- Mandatory human escalation for defined vulnerable, sensitive, ambiguous, or high-impact scenarios.
- Resident-facing disclosure of AI use, system boundaries, and escalation options.
- Ongoing monitoring of unresolved queries, complaint signals, hallucination patterns, and handoff failures.
- Periodic re-testing after content changes, model updates, prompt changes, or policy changes.

## Implementation Notes for Councils

For central government and specified arm's-length bodies, the Algorithmic Transparency Recording Standard is mandatory for in-scope algorithmic tools. For wider public-sector organisations, including local government, it is recommended as a transparency and accountability practice. A council deployment should therefore prepare ATRS-style evidence even where formal filing is not mandatory.

## Important Note

**Illustrative scope disclaimer**

This case study is illustrative only. It does not represent a specific client engagement, formal legal assessment, procurement reference, or verified deployment outcome.

## Guidance Alignment

This document is aligned to current UK public-sector AI assurance expectations and uses the following references as governance context. It does not claim certification, accreditation, formal approval, or legal compliance.

- UK Government: AI Playbook for the UK Government, 10 February 2025.
- UK Government: Algorithmic Transparency Recording Standard guidance for public sector bodies, 8 May 2025.
- UK Government: Data and AI Ethics Framework, updated 18 December 2025.
- Cross-Government Testing Community: AI Testing and Assurance Framework for Public Sector.
- UK Government: Trusted third-party AI assurance roadmap, 3 September 2025.

**Note**

For councils, this pack should be adapted to local governance arrangements, data-protection review, equality impact assessment, service ownership, and procurement requirements.