

DABUDA LIMITED

AI ASSURANCE EVIDENCE SERIES

Sample AI Assurance Evidence Pack

Illustrative evidence-pack structure for AI system testing, risk classification, conditional approval, and governance action tracking.

SAMPLE EVIDENCE PACK - ILLUSTRATIVE ONLY

Version: April 2026

Tone: Executive, evidence-led, procurement-ready

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

SAMPLE EVIDENCE PACK

AI Assurance Evidence Pack

Document status

This sample evidence pack is illustrative. All figures, scenarios, findings, thresholds, and outcomes are examples only.

Use Case Summary

AI system designed to support automated responses for customer-service enquiries. The system is intended to assist with low-risk informational questions and route complex or sensitive interactions to human operators.

Intended Use and Boundaries

Boundary item	Defined position
Intended use	Low-risk informational queries using approved source content.
Out-of-scope use	High-risk decisions, legal advice, eligibility determinations, complaint adjudication, or safeguarding decisions.
Required escalation	Complex, sensitive, ambiguous, vulnerable-user, or high-impact interactions.
Human oversight	Operators remain accountable for final action where the AI output affects user outcomes.

Key Risks Identified

Risk	Initial severity	Evidence source	Control direction
Incorrect responses	High	Accuracy testing and expert review.	Improve knowledge retrieval, add approved answer templates, retest.
Limited explanation clarity	Medium	Output review and user comprehension checks.	Add explanation rules and examples.
Failure to escalate	High	Sensitive scenario testing.	Strengthen routing triggers and human handoff policies.
Inconsistent outputs	Medium	Repeatability and paraphrase testing.	Add response constraints, regression suite, and threshold monitoring.

Test Summary

Metric	Result
Total scenarios tested	250
Coverage areas	Standard queries, edge cases, adversarial inputs, repeated prompts, escalation scenarios.
Standard-query performance	High accuracy in routine informational scenarios.
Ambiguous-input performance	Degradation observed where user intent is incomplete or conflicting.
Escalation behaviour	Inconsistent escalation in selected sensitive or multi-issue scenarios.

Assurance Results Snapshot

Quality attribute	Observed result	Release impact
Accuracy	Acceptable for standard low-risk scenarios.	Proceed with bounded use.
Consistency	Variable across repeated ambiguous prompts.	Conditional control required.
Fairness and accessibility	No final conclusion in sample pack. Further subgroup testing required.	Do not expand scope until completed.
Explainability	Some responses lack clear source or reasoning cues.	Improve before wider release.
Escalation	Escalation gaps found in specific scenarios.	Blocking issue for higher-risk use.

Open Issues

Issue ID	Open issue	Severity	Owner	Target action
AI-001	Incomplete escalation logic in specific sensitive scenarios.	High	Product owner	Update handoff rules and retest before pilot.
AI-002	Limited explainability in selected responses.	Medium	AI delivery lead	Add explanation templates and source cues.
AI-003	Output variability across repeated tests.	Medium	QA lead	Define consistency thresholds and regression suite.
AI-004	Monitoring dashboard not yet operational.	Medium	Operations lead	Implement production metrics and alerting.

Release Recommendation

Status: Conditional Approval

- Deployment is permitted only for low-risk informational queries using approved knowledge sources.
- Escalation gaps must be addressed before the system handles sensitive or complex interactions.
- Consistency thresholds must be defined and added to the regression suite.
- Monitoring controls must be implemented before production pilot.
- Known limitations must be recorded and visible to governance stakeholders.

Governance Actions

Action	Required evidence
Document known limitations	Limitations log, resident/customer disclosure language, and owner approval.
Establish monitoring dashboard	Live metrics for accuracy signals, escalation rates, complaints, hallucination reports, and drift indicators.
Schedule periodic re-testing	Retest calendar and trigger conditions for model, prompt, data, or policy changes.
Define escalation policies	Human handoff workflow, responsible teams, and time-bound handling rules.

Important Note

Illustrative scope disclaimer

This document represents an example evidence-pack format only. All figures, scenarios, findings, thresholds, and outcomes are illustrative and must be replaced with real test evidence before procurement, governance, or audit use.

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

A production evidence pack should include source datasets, prompts, model versions, test scripts, assessor notes, defect records, and decision logs sufficient for audit review.