MARLUC

CASE STUDY · MARLUC SYNOPSIS

The National Archives AI for Selection tests and later AI pilots

IMPLEMENTED 2020–2025 · GOVERNMENT AND PUBLIC SERVICES · UK · BAND D · CLAIMED ONLY

VERSION 1.0 · PUBLISHED 25 NOVEMBER 2026 · UPDATED 25 NOVEMBER 2026 · FACTS CHECKED 25 NOVEMBER 2026

The UK's national archive tested five commercial machine learning tools and an in-house benchmark on choosing digital records for permanent preservation. In the year to February 2025 it piloted AI for transcription and summarising, and investigated tools for records review.

ABOUT THIS SYNOPSIS

This is MarLuc's summary and grading of publicly available sources, for information only. It is not investment, legal, financial or professional advice, and should not be relied on for any decision. Readers must validate it against the original sources and, where appropriate, challenge its conclusions.

Facts and figures are as reported by the sources listed below, which may be incomplete, self-reported or out of date. Check the original source before relying on any figure.

The impact band, proof marker and SEAM view are MarLuc's opinion. The organisation is not a MarLuc client and has not reviewed or endorsed this synopsis. Names and trademarks belong to their owners.

NOTICE WORDING LAST UPDATED 16 NOVEMBER 2026

AT A GLANCE

AI used: Machine learning classification of records, plus AI transcription and summarisation pilots.

Implemented: 2020–2025. Grade: band D, Task support, Claimed only.

Headline result, as reported: Archive's tests found AI tools can help records managers select records but not replace them; later pilots and investigations cover transcription and sensitivity review, with no results published

1 · The case

The facts, as reported by the 4 sources listed at the end. Where a source is silent, the field says "Not reported".

Organisation

The National Archives (UK). The UK's national archive and a non-ministerial department sponsored by the Department for Culture, Media and Sport, per PublicTechnology. Staff numbers: Not reported. Canvas block: 4. Operations and process. Also 3. Customers and proposition.

Business challenge

The challenge. Government departments must select records of enduring value for permanent preservation at the archive. The archive says processes built for paper cannot handle the volume and spread of digital records. An earlier review it cites estimated UK government data at more than 16 billion emails and 3 billion documents.

How it was addressed. In 2020 the archive tested whether commercially available AI tools could classify its own records as selected or not selected for preservation. Five suppliers took part: Adlib Elevate, Amazon Web Services, Microsoft Azure, InSight by Iron Mountain and Records365 by RecordPoint. The archive also built its own simple benchmark tool. By the Minister's written answer, in the 12 months to February 2025 its officials piloted AI in several areas: transcribing some internal meetings and summarising texts; investigating short summaries of record descriptions for the online catalogue; investigating AI tools to help officials appraise, select and sensitivity-review digital records before transfer; and transcribing handwritten and printed text in historical records. [1] [2] [3] [4]

Who implemented it

In-house (In-house): Ran the tests and built the benchmark tool.
Adlib, Amazon Web Services, Microsoft, Iron Mountain and RecordPoint (Software vendor): Supplied tools tested on the archive's data.
Some suppliers worked through consultancies; the paper anonymises them and the report links to supplier reports.

Technology (SEAM pillar E)

SEAM types: T2 Embedded AI in existing software, T3 API integrations, T8 Predictive / ML models. Deployment: Mixture. Vendors and products: Adlib, Amazon Web Services, Microsoft, Iron Mountain, RecordPoint, Adlib Elevate, AWS, Azure, InSight, Records365.

Where it was built in: Supplier tools were tested on a copy of the archive's own records data. The report describes products from off-the-shelf records management systems to cloud-built classifiers.

What changed

Before: records selection relied on processes designed for paper. After the tests: guidance for departments on using AI to assist with selection, and a set of pilots across the archive. Any change in operating practice or volumes: Not reported. [1] [2] [3] [4]

What it cost

Not reported. The archive says off-the-shelf products need licence and renewal fees and bespoke solutions need highly technical staff and ongoing support.

What it saved

Not reported. The archive's paper says any use of machine learning would aim to reduce the manual burden on records managers while they keep the final decision; no time or cost savings are given. [1] [2] [3] [4]

Governance and control

The archive says records managers should keep control of decisions, models should be monitored and retrained, and data sharing agreements were set up with suppliers for its test data. The Minister said officials piloted AI with robust safeguards and could draw on GDS and government frameworks. Accountable owner and any independent review: Not reported.

S2 · Accountable owner · Not reported
Not reported.

A1 · Safe use · Claimed
The Minister said officials piloted AI with robust safeguards, and the archive set up data sharing agreements with suppliers for its test data. [1] [3]

A2 · Human in the loop · Claimed
The archive says AI cannot replace records managers and that they should control how selection decisions are made. [1]

A3 · Monitoring and assurance · Claimed
The archive says model accuracy must be monitored and reviewed periodically, and the tests used held-out data. [1] [2]

A4 · Vendor AI · Claimed
Suppliers ran tools on the archive's data under data sharing agreements; the archive compared them on its own criteria. [1]

A5 · Agent controls · Not reported
Not reported.

A6 · Standards and regulation · Claimed
The Minister cited the Generative AI Framework, the Ethics, Transparency and Accountability Framework and the Data Ethics Framework as resources available to the archive. [3]

What worked, what did not, what is unproven

WORKED

– Promising results were obtained overall, with no tool or approach consistently outperforming the others.

– The in-house benchmark with a simple logistic regression model outperformed all of the supplier tools in the paper's evaluation.

– The archive found that careful preparation of training data and testing gave significantly better results than out-of-the-box tools.

DID NOT

– The archive says AI cannot replace the expertise of records managers.

– One supplier's tool with few training records scored well below the best on F1, which the archive links to the small share of data it used.

– The archive says satisfactory performance on its dataset does not imply equal performance on a larger, more complex department.

UNPROVEN

– Whether any department uses AI for selection in live transfers to the archive.

– Accuracy, time saved and cost in the 2024 pilots on transcription and sensitivity review.

– Performance on messy shared drives as opposed to the archive's curated records system.

Timeline

Oct 2021 Archive publishes Using AI for Digital Records Selection in Government, guidance for records managers [1]

Feb 2025 Minister tells Parliament the archive has piloted AI over the past 12 months, including transcription, summarising and work on records review [3]

2 · Grade and SEAM assessment

MarLuc's opinion, made on the evidence above. How cases are graded: marluc.co.uk/case-studies/how-we-rate

Level of change and impact

Claimed: Not applicable. Grade: D, Task support. Proof: Claimed only. Level of change: Task support.

Task support (Level 1, band D). Tests showed AI could assist records managers; no change to who selects records or how is reported.

Graded D. The archive's tests showed AI tools can help records managers pick records for preservation, but the sources report no move into regular use and no change in who does the work. The 2024 pilots are described in outline only. The results are the archive's own, on one small curated dataset.

The SEAM view and lessons learnt

The tests measured whether tools could classify records, not whether records managers' work changed. The AI's part is clear in the test, because the training data and metrics are published, but the dataset was one small, curated department, which the archive warns against generalising from. Before copying the approach, a leadership team should ask what error rate it will accept in missed records, and who will label the training data.

1. Decide whether missing a valuable record or keeping extra records costs more before choosing a tool.
2. Budget for labelled training data and data science skills.
3. Keep the records manager as the decision maker and retrain models when policy changes.

Evidence

Score 1 of 4. Self-reported: the test results are the archive's own; the 2025 pilots are described only in a ministerial written answer and press coverage of it.

Evaluation method. Self-reported archive tests: results from the archive's own evaluation on its own corporate records, with suppliers running their tools; no independent check.

Sources disagree. The archive's 2021 report says no tool or approach consistently outperformed the others across all tasks; its paper says the in-house benchmark with a simple logistic regression model outperformed all the others. They may rest on different measures.

3 · Original sources

Read these in full. To challenge a fact or a grade, use the contact form at marluc.co.uk or the link on https://marluc.co.uk/case-studies/the-national-archives-ai-for-selection

[1] The National Archives: Using AI for Digital Records Selection in Government (October 2021)
https://cdn.nationalarchives.gov.uk/documents/using-ai-digital-selection-in-government.pdf

[2] The National Archives: Alexa, is this a historical record? (Kuppili Venkata, Young, Bell and Green)
https://cdn.nationalarchives.gov.uk/documents/alexa-is-this-a-historical-record.pdf

[3] UK Parliament written answer via TheyWorkForYou: National Archives: Artificial Intelligence (4 February 2025)
https://www.theyworkforyou.com/wrans/?id=2025-01-27.26414.h

[4] PublicTechnology: National Archives explores AI for sensitivity review and historical transcription (6 February 2025)
https://www.publictechnology.net/2025/02/06/culture-media-and-sport/national-archives-explores-ai-for-sensitivity-review-and-historical-transcription/

The National Archives (UK) website: https://www.nationalarchives.gov.uk

Version 1.0 · Published 25 November 2026 · Updated 25 November 2026 · Facts checked against sources 25 November 2026 · Downloaded 26 November 2026. This PDF may be out of date: the latest version, and a way to report an error or a better source, is at https://marluc.co.uk/case-studies/the-national-archives-ai-for-selection

VERSION HISTORY

v1.0 25 November 2026 First published