NHS England test of machine learning for shortlisting job candidates

IMPLEMENTED 2021–2023 · GOVERNMENT AND PUBLIC SERVICES · UNITED KINGDOM · BAND E · CLAIMED ONLY
VERSION 1.0 · PUBLISHED 16 NOVEMBER 2026 · UPDATED 16 NOVEMBER 2026 · FACTS CHECKED 16 NOVEMBER 2026

NHS England's AI Lab Skunkworks and its London Talent team tested whether past shortlisting decisions could train a model to shortlist interview candidates. The historical data showed little variation by sex or ethnicity, but the trained model missed some groups more than others. NHS England concluded that such models need careful thought before use; no source shows the model was used.

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Facts and figures are as reported by the sources listed below, which may be incomplete, self-reported or out of date.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

No measurable effect: tested, no effect.

AI used: Machine learning classifier (word-frequency features with logistic regression), tested for bias.

Implemented: 2021–2023. Grade: band E, No measurable effect, Claimed only.

Headline result, as reported: Model trained on unbiased-looking data still missed candidates unevenly by sex, age and ethnicity, by NHS England's account

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

NHS England. NHS England, the national body leading the NHS in England. The test used NHS England and NHS Improvement vacancies, mostly administrative roles. Staff numbers: Not reported in the sources used. Canvas block: 8. Governance and risk. Also 5. People and skills.

Business challenge

The challenge. Shortlisting NHS job applicants is time-consuming, and decisions vary between people and hiring rounds, per NHS England. The London Talent team wanted to know whether its hiring data held biases a model could learn.

How it was addressed. Working with Bays Consulting and Niaxo, chosen through the Home Office's Accelerated Capability Environment, the team first checked the historical shortlisting data for bias across job bands, sex and ethnicity. It then built a simple, explainable model (word-frequency features with logistic regression) to predict shortlisting, and compared how often it wrongly rejected candidates who had in fact been shortlisted (the miss rate) across groups. [1] [2] [3] [4]

Who implemented it

Bays Consulting (Consultancy): Joint supplier that analysed the data and built the model.
Niaxo (Consultancy): Joint supplier with Bays Consulting.
In-house (In-house): NHS AI Lab Skunkworks: nHS England's AI team that ran the project with the London Talent team.
Suppliers were selected through the Home Office's Accelerated Capability Environment.

Technology (SEAM pillar E)

SEAM types: T8 Predictive / ML models. Deployment: Not reported. Vendors and products: Bays Consulting, Niaxo, TF-IDF and logistic regression shortlisting model (test only).

Where it was built in: Not integrated: trained and tested on historical application data only.

What changed

Nothing in practice. Shortlisting stays with hiring managers. The project produced a published finding on the risk of model bias. [1] [2] [3] [4]

What it cost

Not reported

What it saved

Not reported [1] [2] [3] [4]

Governance and control

The test was designed around bias: explainable models were chosen so bias could be traced, the data was checked before training and predictions were checked by group after. NHS England concluded models for shortlisting should be carefully thought through before use. Who would own any decision to deploy is not reported.

S2 · Accountable owner · Not reported
The London Talent team and the AI Lab Skunkworks ran the test; no owner for any deployment decision is named. [1]

A1 · Safe use · Claimed
Pseudonymised applications for closed vacancies were to be used, per the Skunkworks project page. [2]

A2 · Human in the loop · Not reported
Not reported; the model was not used on live decisions.

A3 · Monitoring and assurance · Claimed
The data was checked for bias before training and the model's miss rate was compared across sex, age and ethnicity after. [1]

A4 · Vendor AI · Not reported
Not reported.

A5 · Agent controls · Not reported
Not reported.

A6 · Standards and regulation · Not reported
Not reported.

What worked, what did not, what is unproven

WORKED

– The test found model bias before any use, by NHS England's account.

– Explainable methods made the uneven results visible and were published openly.

DID NOT

– The model shortlisted some groups more accurately than others, even though the training data showed little variation by sex or ethnicity.

– The cause was not established: NHS England names the make-up of the dataset or differences in language and writing style as possibilities, and says the second was not explored.

UNPROVEN

– Why the model performed unevenly, which NHS England said needs further work.

Timeline

Nov 2021 Skunkworks lead describes the shortlisting bias project for the London Talent team [4]

Apr 2023 NHS England publishes the case study and findings [1]

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: E, No measurable effect. Proof: Claimed only. Level of change: Not graded.

No level assessed: band E, a test with no deployment.

A test of feasibility and bias, with no deployment. It found uneven error rates and was not taken into use in any source found.

The SEAM view and lessons learnt

This is a useful negative result. The team checked its data, chose a model it could explain, measured errors by group, and found uneven performance before anyone relied on it. Clean-looking historical data was not enough to produce a fair model. For a larger organisation, the transferable point is to test error rates by protected group before any AI touches hiring decisions.

1. Test error rates by protected group before use.
2. Choose models whose decisions you can explain.

Evidence

Score 2 of 4. Self-reported: NHS England's own published method and results. Not independently reviewed.

Evaluation method. Analysis of historical data and a model test against a set acceptable error rate, by NHS England's own team and suppliers. Not independently reviewed.

Status. Complete. NHS England's data science pages list the project as complete; no source reports the model in use.

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/nhs-england-recruitment-shortlisting-bias-test

[1] NHS England AI Knowledge Repository: Examining whether recruitment data can, and should, be used to train AI models for shortlisting interview candidates
https://digital.nhs.uk/services/ai-knowledge-repository/case-studies/examining-whether-recruitment-data-can-and-should-be-used-to-train-ai-models-for-shortlisting-interview-candidates

[2] NHS AI Lab Skunkworks: Recruitment shortlisting with the NHS England London Talent team
https://nhsx.github.io/skunkworks/recruitment-shortlisting

[3] NHS England Data Science: AI models for shortlisting interview candidates
https://nhsengland.github.io/datascience/our_work/casestudy-recruitment-shortlisting

[4] Hospital Times: NHS AI Lab Skunkworks, the cautious early adopters (November 2021)
https://hospitaltimes.co.uk/nhs-ai-lab-skunkworks-the-cautious-early-adopters

NHS England website: https://www.england.nhs.uk

Version 1.0 · Published 16 November 2026 · Updated 16 November 2026 · Facts checked against sources 16 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/nhs-england-recruitment-shortlisting-bias-test

VERSION HISTORY

v1.0 16 November 2026 First published