MARLUC

CASE STUDY · MARLUC SYNOPSIS

Consult consultation-analysis tool (i.AI)

IMPLEMENTED 2025–2026 · GOVERNMENT AND PUBLIC SERVICES · UK · BAND C · PARTLY PROVED

VERSION 1.0 · PUBLISHED 15 NOVEMBER 2026 · UPDATED 15 NOVEMBER 2026 · FACTS CHECKED 15 NOVEMBER 2026

Consult uses a language model to group public consultation responses into themes, with officials signing off the themes and checking the mappings. Government-published evaluations on a Scottish Government consultation and a Department for Work and Pensions consultation found it matched human reviewers; it remains in trial.

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Facts and figures are as reported by the sources listed below, which may be incomplete, self-reported or out of date.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: Generative AI: large language models that map consultation responses to themes, with topic modelling to generate candidate themes.

Implemented: 2025–2026. Grade: band C, Process change, Partly proved.

Headline result, as reported: In i.AI's evaluation of a DWP consultation, Consult's theme mapping scored 0.816 F1 against one reviewer, against 0.710 between two reviewers; the tool is still in trial

1 · The case

The facts, as reported by the 5 sources listed at the end. Where a source is silent, the field says "Not reported".

Organisation

Incubator for AI (i.AI), Department for Science, Innovation and Technology, UK Government. UK central government. The Incubator for AI is a technical team within the Government Digital Service, part of the Department for Science, Innovation and Technology (per Nesta). Departments that ran live consultations through the tool include the Scottish Government and the Department for Work and Pensions. Canvas block: 4. Operations and process. Also 5. People and skills, 8. Governance and risk.

Business challenge

The challenge. The government says it runs around 500 consultations a year (press release) or around 600 (i.AI's evaluation), some with more than 100,000 responses. Analysing them takes large amounts of civil service time or is contracted out. The AI Knowledge Hub page says outsourced analysis can cost up to £100,000 per consultation and take months.

How it was addressed. i.AI built Consult, part of the Humphrey suite of tools. It identifies themes for each question from all responses, the policy team signs off the themes, and the tool then maps each response to one or more themes using large language models. Results appear in a dashboard. Officials reviewed the mappings in evaluations on a live Scottish Government consultation on non-surgical cosmetic procedures and a live DWP consultation on health and disability benefits reform and employment support. The code is published as open source. [1] [2] [3] [4] [5]

Who implemented it

In-house (In-house): Built and evaluates Consult within government.
OpenAI (Model or cloud provider): Supplied the language model (GPT-4o at the time of Nesta's assessment).
Nesta (University or research lab): Ran an independent public deliberation on the tool.
The Scottish Government and the Department for Work and Pensions provided live consultations for the evaluations.

Technology (SEAM pillar E)

SEAM types: T3 API integrations. Deployment: Cloud APIs. Vendors and products: OpenAI, Consult, ThemeFinder (open-source package), GPT-4o (at the time of Nesta's assessment).

Where it was built in: A standalone web application with a reviewer interface and a results dashboard. A separate open-source package, ThemeFinder, holds the theme generation and mapping. Not stated to be built into departmental systems.

What changed

Before: teams read a sample to find themes, trained reviewers on a theme framework and then mapped every response by hand. After: in the evaluations, reviewers checked the tool's mapping rather than starting from nothing; in the DWP evaluation the median review took 19 seconds per response. The press release says the tool will soon be used on major consultations without officials reviewing every response individually, with officials always reviewing themes and the sorting through a dashboard. The AI Knowledge Hub says in July 2026 that i.AI is testing Consult with a small number of live consultations and plans wider rollout in 2027. [1] [2] [3] [4] [5]

What it cost

Cost of the build: Not reported. i.AI's developer response in the Nesta report says a recent consultation analysed with the tool cost less than £240 in tokens with OpenAI. The AI Knowledge Hub says initial testing suggested the tool would be 400x cheaper than existing methods.

What it saved

Projected, not measured as a saving: the press release says that across about 500 consultations a year the tool could help save around 75,000 days of analysis, costing £20 million in staffing. Measured in the evaluations: median review time of 23 seconds per response (Scottish Government) and 19 seconds (DWP). [1] [2] [3] [4] [5]

Governance and control

Officials sign off the themes and review the mappings; the evaluations used up to 125 reviewers. Nesta ran 18 deliberative sessions with 144 members of the public on the tool and issued an advisory label with recommendations. i.AI told Nesta it is model agnostic and that its contract with OpenAI means data is not stored or used for training. The DWP evaluation includes an equality analysis by disability status and found no evidence of bias.

S2 · Accountable owner · Claimed
i.AI, within the Government Digital Service, built and evaluates the tool; the policy team running each consultation signs off its themes. No named senior owner is reported. [3] [4]

A1 · Safe use · Outside scrutiny (a regulator, court, auditor or similar body acted)
Nesta ran 18 deliberative sessions with 144 members of the public on the tool and issued an advisory label with recommendations; the people consulted were comfortable with the use because of the human oversight. [4]

A2 · Human in the loop · Evidenced
Reviewers checked the tool's theme mapping in the evaluations; the DWP evaluation recorded that reviewers made no changes for 73% of responses and added or removed themes for the rest. [3]

A3 · Monitoring and assurance · Evidenced
i.AI published evaluations of live use, including an equality analysis by disability status and a theme-stability check, and says further evaluation will precede rollout decisions. [1] [3]

A4 · Vendor AI · Claimed
i.AI told Nesta that its tools are model agnostic and that its contract with OpenAI means data is not stored or used for training. [4]

A5 · Agent controls · Not reported
Consult is described as a theme-mapping tool with human sign-off; no agent use is described.

A6 · Standards and regulation · Claimed
i.AI told Nesta that an Algorithmic Transparency Recording Standard record for the tool has been published; the AI Knowledge Hub page links to it. The record itself was not read for this synopsis. [4] [5]

What worked, what did not, what is unproven

WORKED

– In the DWP evaluation Consult's mappings were closer to a single reviewer (F1 0.816) than two reviewers were to each other (0.710).

– Reviewers made no changes to the tool's themes for 73% of DWP responses; two reviewers gave identical mappings to each other for 65%.

– Differences between the tool and reviewers had minimal effect on how themes ranked, which is what policy teams mostly use.

– The DWP evaluation found no difference in performance by self-reported disability status.

– In Nesta's deliberative sessions, most participants agreed the tool would benefit the consultation process and were comfortable with it because it has human oversight and does not make decisions.

– In the Scottish evaluation, reviewers added themes 1,671 times and removed them 763 times; reviewers found the theme sign-off process challenging.

– In the DWP evaluation almost three times as many theme mappings were added (19,047) as removed (6,498), and the evaluation says reviewers erred more on the side of including themes than Consult does.

– Nesta's participants worried about an AI model made by a US technology company, and about longer-term effects on the number and diversity of consultation responses.

– The tool is not yet in general use: the AI Knowledge Hub page says it cannot take on new consultations.

UNPROVEN

– The £20 million and 75,000-day figures are government projections across all consultations; no realised saving is reported.

– Performance has been evaluated by i.AI itself; an evaluation by a body outside the developer: Not reported.

– Whether the tool is used without full human review of every response on live consultations: the press release says this is planned; later use: Not reported.

– Which model the tool uses now: Nesta says GPT-4o at the time of its assessment; the current model is Not reported.

Timeline

May 2025 First live use, on a Scottish Government consultation on non-surgical cosmetic procedures; evaluation and press release published on 14 May 2025 [1]

May 2025 Nesta begins 18 deliberative polling sessions with 144 members of the public (27 May to 13 June 2025) [4]

Aug 2025 i.AI evaluation of a live DWP consultation dated 27 August 2025: 87,738 responses reviewed by 125 people [3]

Aug 2025 Nesta publishes its public-opinion report and advisory label (page dated 16 August 2025) [4]

Jul 2026 AI Knowledge Hub page, last updated 23 July 2026, says Consult is in beta, testing on a small number of live consultations, with wider rollout planned for 2027 [5]

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: C, Process change. Proof: Partly proved. Level of change: Process change.

Process change (Level 2, band C). Consult is a redesigned analysis process with sign-off points, but its owners describe it as in trial or beta, so the depth of change is limited to the consultations where it has been used.

The SEAM view and lessons learnt

The design keeps people at the two decisions that matter: which themes to use, and whether each mapping is right. The evaluations compare the tool with a human baseline rather than with perfection, and the second one added a double-reviewed sample to measure how far two people differ. The transferable points for a larger organisation are to measure the human baseline first and to publish the method. The cost and time savings remain projections until a department reports realised figures.

1. Measure how far two people differ before judging the tool.
2. Keep experts in the loop at theme sign-off and at review.
3. Publish the method, the code and the equality analysis with the results.

Evidence

Score 2 of 4. Published evaluations by the developer on live consultations, with a human baseline and an equality analysis in the second; independent public-opinion research by Nesta. No independent audit of accuracy.

Evaluation method. Developer evaluations on live consultations. Scottish Government: one reviewer checked the mapping for every response; no double review. DWP: 17,000 responses double reviewed and 70,738 single reviewed by 125 people; F1 scores with confidence intervals; theme-ranking comparison; equality analysis. Nesta: 18 deliberative polling sessions with 144 people.

Sources disagree. The press release says the government runs around 500 consultations a year; i.AI's DWP report says around 600. The Scottish evaluation counted themes unchanged for 60% of responses and the DWP evaluation 73%, on different consultations and with different review designs. The AI Knowledge Hub gives the Scottish F1 of 0.76 as 'generally considered good'; the DWP report says there is no set definition of a good F1 score.

Status. In trial or beta. As of the AI Knowledge Hub page last updated 23 July 2026, i.AI is testing Consult with a small number of live consultations and plans a wider rollout in 2027.

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/i-ai-consult-consultation-analysis

[1] GOV.UK (DSIT): Government-built “Humphrey” AI tool reviews responses to consultation for first time, in bid to save millions
https://www.gov.uk/government/news/government-built-humphrey-ai-tool-reviews-responses-to-consultation-for-first-time-in-bid-to-save-millions

[2] Incubator for AI: Evaluating Consult: An AI Tool for Enhanced Public Consultation Analysis
https://ai.gov.uk/blogs/evaluating-consult-an-ai-tool-for-enhanced-public-consultation-analysis/

[3] Incubator for AI: Consult Evaluation: DWP's Pathways to Work consultation (27 August 2025)
https://assets.publishing.service.gov.uk/media/68f63fe01c9076042263f0bd/evaluation-report.pdf

[4] Nesta: How does the public feel about AI analysing UK government consultation responses?
https://www.nesta.org.uk/report/ai-sral-consult/

[5] AI Knowledge Hub (DSIT): Consult
https://ai.gov.uk/knowledge-hub/tools/consult/

Incubator for AI (i.AI), Department for Science, Innovation and Technology, UK Government website: https://ai.gov.uk

Version 1.0 · Published 15 November 2026 · Updated 15 November 2026 · Facts checked against sources 15 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/i-ai-consult-consultation-analysis

VERSION HISTORY

v1.0 15 November 2026 First published