MARLUC

CASE STUDY · MARLUC SYNOPSIS

Penda Health AI Consult clinical decision support in Kenya

IMPLEMENTED 2025–2026 · HEALTH AND LIFE SCIENCES · KENYA · BAND D · PARTLY PROVED

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

Penda Health, a network of primary care clinics in Kenya, added an AI tool called AI Consult to its electronic medical record. The tool uses GPT-4o to check clinicians' entries as they work and raises green, yellow or red alerts. A preprint of a quality improvement study reported fewer clinical errors. A later cluster-randomised trial, published in Nature Medicine in June 2026, found better documentation and diagnosis ratings but no statistically significant fall in treatment failure within 14 days.

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Facts and figures are as reported by the sources listed below, which may be incomplete, self-reported or out of date.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 clinical decision support: GPT-4o checks clinicians' entries in the electronic medical record in the background and shows green, yellow or red alerts.

Implemented: 2025–2026. Grade: band D, Task support, Partly proved.

Headline result, as reported: A cluster-randomised trial of 9,691 patients in 16 Kenyan clinics found that GPT-4o clinical decision support improved note and diagnosis quality but did not significantly reduce 14-day treatment failure (2.2% against 2.0%, adjusted odds ratio 0.77, 95% CI 0.55 to 1.08)

1 · The case

The facts, as reported by the 3 sources listed at the end. Where a source is silent, the field says "Not reported".

Organisation

Penda Health, Kenya (trial sponsored by PATH). Penda Health is a Nairobi-based social enterprise founded in 2012. By the preprint's account it runs 16 clinics, records over 1,000 patient visits a day and has more than 100 licensed clinical officers. The trial covered 16 facilities

Business challenge

The challenge. By the trial authors' account, primary care in Kenya is often delivered by clinical officers, mid-level practitioners with a three-year diploma, who make complex diagnosis and treatment decisions without access to senior consultation. They say this can lead to inconsistent adherence to guidelines, diagnostic errors and inappropriate treatment, and that prospective evidence on large language models in low-resource clinics was limited.

How it was addressed. Penda developed AI Consult with its EMR vendor and its clinical quality and IT teams. Version 2 runs GPT-4o in the background at key points in a visit and shows a green (no action), yellow (advisory) or red (requires review) alert. For the trial, clinical officers were randomised to use the EMR with or without the tool. [1] [2] [3]

Who implemented it

In-house (In-house): Built and deployed AI Consult with its clinical quality and IT teams and its EMR vendor..
OpenAI (Model or cloud provider): Supplied GPT-4o and gave cloud compute credits and technical guidance in kind..
PATH (Industry partner): Sponsored the randomised trial, funded by the Gates Foundation..
The EMR vendor is not named in the sources opened. The preprint says OpenAI funded that study and took part in its analysis and reporting; the trial paper says the funder, the Gates Foundation, had no role in design, analysis or publication.

Technology (SEAM pillar E)

SEAM types: T3 API integrations. Deployment: Cloud APIs. Vendors and products: OpenAI, GPT-4o (May 2025 release), AI Consult version 2.0.

Where it was built in: Built into Penda's cloud-based electronic medical record. The system sends the clinician's structured and free-text entries, without patient identifiers, to GPT-4o at key points such as diagnosis and treatment, and returns a green, yellow or red alert. Clinicians do not start a separate query.

What changed

Before: clinical officers recorded notes and orders with no automated second check. After: in the AI arm, GPT-4o reviewed entries at key points and clinicians decided what to do. Roles and structure are not reported as changed. [1] [2] [3]

What it cost

Model cost was US$0.04 per patient. Development, integration and staff time: Not reported. OpenAI gave cloud compute credits and technical guidance in kind.

What it saved

Post hoc, antibiotic spending per patient was US$0.15 lower in the AI arm (95% CI 0.04 to 0.25), which the authors say exceeded the per-patient cost of the model. Consultation time was a median of 11 minutes in both arms. No other cost or time saving is reported. [1] [2] [3]

Governance and control

Ethics, county and national approvals preceded the trial. An independent data and safety monitoring board reviewed it, serious adverse events went to independent review, and an expert panel rated outputs. The device regulator ruled the product outside its oversight scope. Clinicians kept every decision.

S2 · Accountable owner · Evidenced
The trial paper names PATH as sponsor and the Penda clinical quality team ran the tool; the trial is registered with the Pan-African Clinical Trials Registry. [1] [2]

A1 · Safe use · Outside scrutiny (a regulator, court, auditor or similar body acted)
Ethics approval came from the Amref Health Africa Ethical and Scientific Review Committee, with authorisation from two counties and the national science commission; Penda also red-teamed the tool before deployment. [1] [2]

A2 · Human in the loop · Evidenced
Clinicians decided what to do and could disregard the alerts; of 1,000 red-alert encounters reviewed, clinicians fully followed the advice in 19.5% and did not follow it in 23.2%. [1]

A3 · Monitoring and assurance · Outside scrutiny (a regulator, court, auditor or similar body acted)
An independent data and safety monitoring board held scheduled reviews, including one after the first 1,000 participants, and independent review of 33 serious adverse events found no link to the tool. [1]

A4 · Vendor AI · Evidenced
The paper names the exact model release and settings (GPT-4o May 2025, temperature 0.1) and discloses OpenAI's in-kind support of cloud compute credits and technical guidance. [1]

A5 · Agent controls · Not reported
No agents described.

A6 · Standards and regulation · Outside scrutiny (a regulator, court, auditor or similar body acted)
Kenya's Pharmacy and Poisons Board determined the product fell outside its device oversight scope; the paper reports compliance with the Kenya Data Protection Act and the CONSORT-AI reporting extension. [1]

What worked, what did not, what is unproven

WORKED

– Independent physicians rated AI-arm notes as more likely to hold an appropriate diagnosis (adjusted odds ratio 1.74), a comprehensive note (1.68) and an appropriate treatment plan (1.71), in 2,000 reviewed encounters.

– No serious adverse event was judged related to the tool, and independent review found no safety signal.

– In the earlier preprint, physician-rated errors were 16% lower for diagnosis and 13% lower for treatment, and all surveyed clinicians said the tool improved the quality of care.

– The model cost about US$0.04 per patient.

DID NOT

– The primary outcome did not differ significantly: treatment failure in 2.2% with AI against 2.0% in control, adjusted odds ratio 0.77 (95% CI 0.55 to 1.08, P=0.13). The authors conclude any benefit is probably modest.

– There was no difference in correct antibiotic use, antimalarial prescribing, hypertension or malnutrition management, and patient satisfaction scores were identical.

– Of 1,000 red-alert encounters reviewed, clinicians fully followed the AI's advice in 19.5%, partly in 57.3% and not at all in 23.2%.

UNPROVEN

– Whether the tool changes patient outcomes: the authors say detecting modest differences would need more than 100,000 patients, and a co-author told NPR about 139,000.

– Safety for rare severe events: the paper says the trial was not powered for them.

– Whether results hold outside Penda: the preprint calls Penda a particularly strong setting for digital health, with dedicated technical investment.

– Whether the tool is now used across Penda or elsewhere: Not reported; a co-author says he is working toward implementing it globally.

Timeline

Jan 2025 AI Consult version 2 goes live; Penda gives access to a randomly assigned half of its clinicians from 30 January to 18 April (quality improvement study) [2]

Apr 2025 Randomised trial begins enrolling patients across 16 Penda facilities (22 April to 16 July) [1]

Jul 2025 Preprint reports 16% fewer diagnostic errors and 13% fewer treatment errors in the quality improvement study [2]

Jun 2026 Nature Medicine publishes the trial: primary outcome not significantly different, no safety signal found [1]

Jul 2026 NPR reports the result; a co-author says he is working toward implementing the tool globally [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: Partly proved. Level of change: Task support.

Task support (Level 1, band D). Clinicians do the same work with an AI check. The trial shows better note and diagnosis quality and no significant effect on the patient outcome, and the sources report no process or operating model change.

Clinical officers do the same work with an AI second check. The randomised trial found better documentation and diagnosis ratings and no significant effect on patient outcomes. The sources report no change to roles, structure or process, so this is task support with the benefit partly proved.

The SEAM view and lessons learnt

NPR describes this as among the first trials of this kind of AI as an aid to primary care. It reports a null primary result openly. The tool improved what the record says and was judged safe, but the patient outcome chosen was too rare to show a modest effect. The earlier preprint, with a larger error reduction, used a different design and a funder that helped analyse the data. For a

1. Test clinical AI in live care against a control group before wider rollout.
2. Choose a primary outcome common enough to detect, and report process gains separately.
3. Build the tool into the existing workflow and monitor safety independently.

Evidence

Score 3 of 4. Independent: peer-reviewed cluster-randomised trial (Nature Medicine, 2026), funded by the Gates Foundation, with an earlier OpenAI-funded preprint. Limits: one provider and country, declared competing interests (two authors hold Penda stock options; OpenAI gave in-kind support) and a primary outcome that was not significant.

Evaluation method. Randomised trial: pragmatic cluster-randomised trial, randomised by clinical officer across 16 facilities (Pan-African Clinical Trials Registry 202502499779176). The earlier quality improvement study randomly assigned clinicians within clinics.

Sources disagree. NPR describes a 23% decrease in treatment failures; the paper reports raw rates of 2.2% (AI) against 2.0% (control) and an adjusted odds ratio of 0.77 that is not statistically significant. The preprint covers 15 clinics and 39,849 visits and the trial 16 facilities and 9,691 patients, and the paper treats them as separate studies. The trial paper's abstract gives 9,691 patients enrolled while its results text gives 9,702 encounters included.

Status. Complete: trial published June 2026. Penda's rollout after the trial: Not reported.

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/penda-health-ai-consult-kenya

[1] Nature Medicine: Generative AI-enabled clinical decision support system in primary care: a pragmatic, cluster-randomized trial
https://www.nature.com/articles/s41591-026-04503-6

[2] arXiv (Penda Health and OpenAI): AI-based Clinical Decision Support for Primary Care: A Real-World Study
https://arxiv.org/abs/2507.16947

[3] NPR: This AI tool promises a 'second pair of eyes' to clinicians. Did patients benefit?
https://www.npr.org/2026/07/23/g-s1-134929/ai-artificial-intelligence-healthcare

Penda Health, Kenya (trial sponsored by PATH) website: https://www.pendahealth.com

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/penda-health-ai-consult-kenya

VERSION HISTORY

v1.0 15 November 2026 First published