Rotterdam welfare fraud risk-scoring model

IMPLEMENTED 2017–2021 · GOVERNMENT AND PUBLIC SERVICES · NETHERLANDS · BAND F · REVERSED
VERSION 1.0 · PUBLISHED 15 NOVEMBER 2026 · UPDATED 15 NOVEMBER 2026 · FACTS CHECKED 15 NOVEMBER 2026

From 2017 the city used a machine learning model to score its roughly 30,000 welfare recipients for fraud risk and to help choose whom to investigate. The city's audit office warned of biased outcomes in 2021 and the city paused the model. In 2023 journalists, using the city's own model and data, reported that being young, female, a parent or not fluent in Dutch raised a person's score.

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Facts and figures are as reported by the sources listed below, which may be incomplete, self-reported or out of date.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

What went wrong: AI output or malfunction; Weak oversight or controls.

AI used: Machine learning model (risk scoring) trained on past welfare fraud investigations, producing a score for each recipient to help select people for re-examination.

Implemented: 2017–2021. Grade: band F, Failure, Reversed.

Headline result, as reported: City paused its welfare fraud risk-scoring model in 2021 after an audit office warning; a 2023 reconstruction by journalists reported higher scores for women, parents, young people and non-fluent Dutch speakers

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

Gemeente Rotterdam (City of Rotterdam). Dutch municipality. Around 30,000 residents receive welfare benefits from the city each year; the model scored all of them. Canvas block: 8. Governance and risk. Also 4. Operations and process.

Business challenge

The challenge. The city investigates welfare recipients to check payments are correct, up to 6,000 people a year per WIRED. It wanted a way to choose whom to re-examine; the city says the algorithm was meant to burden fewer residents with re-examination where there is no reason for it (NOS, April 2021).

How it was addressed. From 2017 the city used a machine learning model, trained on 12,707 previous investigations, that gave each recipient a risk score from 315 factors, including age, gender, language skills, neighbourhood, marital status and caseworker assessments. Officials ranked recipients and referred the top tenth for investigation, on average the top 1,000 a year (Lighthouse Reports). The city says the algorithm does not decide whether anyone is a fraudster; it is a risk estimate and re-examinations are always done by people. The city shared its model, training data and manuals with journalists. [1] [2] [3] [4] [5]

Who implemented it

Accenture (Consultancy): Originally developed the system before the city took over development in 2018, per WIRED..
In-house (In-house): Took over development in 2018, ran the model and shared its code, data and manuals with journalists in 2022 and 2023..
Lighthouse Reports says the model was built with the help of Accenture; Accenture's own account is not reported in the sources opened.

Technology (SEAM pillar E)

SEAM types: T8 Predictive / ML models. Deployment: Not reported. Vendors and products: Accenture.

Where it was built in: A scoring model in the city's welfare income process: it scored each recipient on 315 inputs and produced a ranking; the top tenth, about 1,000 people on average, were referred for investigation (Lighthouse Reports).

What changed

Before 2017: Not reported in the sources opened. During 2017 to 2021: the model flagged up to a third of those investigated each year, others were chosen by theme (WIRED). After: the city paused the model in 2021 and, as of March 2023, selected recipients at random while it worked on a new version. [1] [2] [3] [4] [5]

Cost of failure

Not reported. No fine, damages or write-off is reported against the city for the model. WIRED reports one resident, whose selection was by theme and not by the algorithm, had €6,000 refunded after winning on appeal in 2021.

What it saved

Not reported. The 2021 audit report criticised the city for not evaluating whether its algorithms did better than the human methods they replaced. [1] [2] [3] [4] [5]

Governance and control

The city's audit office found insufficient attention to ethical principles, no one fully responsible overall, no evaluation against the traditional method and no clear criteria for leaving pilot status. The city largely agreed, with reservations, and paused the model; it says it has added safeguards, including an algorithm register. A 2024 follow-up found more attention to risks, with unclear principles still open.

S2 · Accountable owner · Missing (independent source shows it was absent)
The audit office found there is no central director for the algorithms and nobody is "integraal eindverantwoordelijk" (fully responsible overall), per NOS. [3]

A1 · Safe use · Missing (independent source shows it was absent)
The audit office found it "niet navolgbaar" (not traceable) which choices on ethical questions had been made, and that the data could lead to biased outcomes. [3]

A2 · Human in the loop · Claimed
The city says re-examinations are always carried out by people and the algorithm gives a risk estimate, not whether someone is a fraudster. [2] [3]

A3 · Monitoring and assurance · Missing (independent source shows it was absent)
The audit office found the city did not evaluate whether the algorithms worked better than the traditional method, and the pilot had no defined criteria for permanent use. [2] [3]

A4 · Vendor AI · Not reported
Accenture's role in building the model is reported; vendor controls and testing are not. [1] [2]

A5 · Agent controls · Not reported
No agents described.

A6 · Standards and regulation · Outside scrutiny (a regulator, court, auditor or similar body acted)
Rekenkamer Rotterdam reviewed the city's algorithms in 2021 and again in 2024; the city paused the model after the first report. No court or regulator ruling on the model is reported. [2] [3] [4]

What worked, what did not, what is unproven

WORKED

– The city says human consultants always assessed flagged people and that the model gave a risk estimate, not a verdict on fraud (the city's account, WIRED and NOS).

– The city disclosed the model, training data and manuals when asked, which made independent analysis possible (Lighthouse Reports).

– After its audit office's report the city paused the model and, per a 2024 Rekenkamer follow-up, paid more attention to algorithm risks.

DID NOT

– Rekenkamer Rotterdam found the model was trained on biased data, because it used mainly data from recipients who had already been re-examined (NOS, translated).

– The audit office said indirect proxies such as Dutch speaking level, whether someone needs civic integration and neighbourhood could point to non-Dutch origin (NOS, translated).

– WIRED reported that being a parent, a woman, young, not fluent in Dutch or struggling to find work increased a person's score, and that experts who reviewed the findings expressed serious concerns that the system may have discriminated; Lighthouse Reports concluded that it discriminates on ethnicity, age, gender and parenthood.

– Experts quoted by Lighthouse Reports say the model performs little better than random selection; this is their assessment, not a measured result published by the city.

– Whether any person lost benefits or was wrongly investigated because the model scored them: WIRED reports individual accounts but the sources opened give no count.

– Whether the replacement model the city said it was developing was ever deployed: not reported in the sources opened.

– The city's own accuracy figures for the model: Lighthouse Reports says it received evaluations of performance, but they are not quoted in the sources opened.

– Whether the model was lawful under Dutch and EU rules: no court or regulator finding is reported.

Timeline

2017 City begins using a machine learning model to score welfare recipients for fraud risk [2]

2018 City takes over development of the model from consulting firm Accenture [2]

Apr 2021 Rekenkamer Rotterdam (the city's audit office) publishes its report on algorithm use and warns of biased outcomes [3]

2021 City pauses use of the risk-scoring model [2]

Mar 2023 Lighthouse Reports, WIRED and Dutch partners publish their reconstruction of the model [1]

May 2024 Rekenkamer follow-up (Kleur bekennen) finds the city has paid more attention to algorithm risks, with open points remaining [4]

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: F, Failure. Proof: Reversed. Level of change: Not graded.

No level assessed: band F, a deployment paused after an audit warning and later reported as biased. The model was used to select people for investigation, so its effects fell on residents.

Band F: a live deployment that was paused and not restored. The city paused the model in 2021 after its audit office reported a likelihood of biased outcomes; by the city's account people flagged were always assessed by human consultants. The city did not claim transformation. No court or regulator has ruled on the model, so the marker is Reversed rather than Proved. The bias findings in 2023 come from journalists' reconstruction with outside expert review, which the city did not dispute in the sources opened.

The SEAM view and lessons learnt

The model sat between data and decisions: scores shaped who was examined, while humans took the final decision. The audit office's concerns were at the join: no single owner, no comparison with the old method, and training data drawn from people already selected. Those faults are not specific to welfare. Any organisation that uses scores to choose whom to check needs an owner, a baseline to beat and a test for proxies for protected characteristics before the

1. Name one accountable owner for any model that picks people for checks.
2. Test the model against the old method and for proxies for protected characteristics before use.
3. Tell the people affected how selection works and publish the variables used.

Evidence

Score 2 of 4. Mixed: an official audit office report (read through press coverage) and an investigative reconstruction with expert review; no court or regulator finding. The city's own accuracy evaluations were not seen.

Evaluation method. Official audit office report (April 2021, read via NOS) plus an investigative reconstruction using the city's model, data and manuals with external expert review of the method (Lighthouse Reports, 2023). Quotes from NOS, Binnenlands Bestuur and Joop are translated from Dutch.

Sources disagree. Joop reports the city has stopped the algorithm and is building a new version meant to avoid discrimination; WIRED says the city paused its use in 2021 and was working on a new version, with recipients selected at random meanwhile. Alderman Richard Moti said in 2021 that there had been no discrimination in the city's use of algorithms (Joop, translated); the 2023 reconstruction reports otherwise.

Status. Paused since 2021; a replacement model's status is not reported in the sources opened.

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/rotterdam-welfare-fraud-risk-scoring

[1] Lighthouse Reports: Suspicion Machines
https://www.lighthousereports.com/investigation/suspicion-machines/

[2] WIRED: This Algorithm Could Ruin Your Life
https://www.wired.com/story/welfare-algorithms-discrimination/

[3] NOS: Rekenkamer Rotterdam: risico op vooringenomen uitkomsten door gebruik algoritmes
https://nos.nl/artikel/2376810-rekenkamer-rotterdam-risico-op-vooringenomen-uitkomsten-door-gebruik-algoritmes

[4] Binnenlands Bestuur: Rotterdam heeft meer oog voor risico's algoritmes
https://www.binnenlandsbestuur.nl/digitaal/ondermijning/rotterdam-heeft-meer-oog-voor-risico-s-algoritmes

[5] Joop (BNNVARA): Alleenstaande vrouwen en anderstaligen eerst: Rotterdam gebruikte discriminerend algoritme om bijstandsfraude op te sporen
https://www.bnnvara.nl/joop/artikelen/alleenstaande-vrouwen-en-anderstaligen-eerst-rotterdam-gebruikte-discriminerend-algoritme-om-bijstandsfraude-op-te-sporen

Gemeente Rotterdam (City of Rotterdam) website: https://www.rotterdam.nl

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/rotterdam-welfare-fraud-risk-scoring

VERSION HISTORY

v1.0 15 November 2026 First published