Can open LLMs clear sanctions-screening alerts? What a pre-registered test found

25 November 2026 · 8 min read · blogs.iictp.org/open-llms-sanctions-screening-alert-disposition-turkish-names-explained

Short answer: an arXiv preprint submitted on 21 November 2026 by Mehmet Murat Huyut tests whether open-weight language models that a bank can run on its own servers can do level-1 sanctions-screening disposition. Using 7,103 persons from UK and EU lists and 12,200 synthetic customers, the models sent far fewer innocent customers to level 2 than a date-of-birth rule, but failed to escalate 1.5% to 2.2% of listed persons. The author concludes open models can reduce level-1 workload only behind a deterministic date check, with every name doubt escalated. This is an arXiv preprint. It has not been peer reviewed, and its numbers are the authors' own claims.

TL;DR The short version

- A pre-registered arXiv test checked whether open-weight models a bank can run in-house can clear level-1 sanctions-screening alerts.
- Models sent far fewer innocent customers to level 2 than a strict rule, and stated date-of-birth status correctly in about 98% of alerts.
- Customers were synthetic and the author reports a dataset error; models can cut workload only behind a deterministic date check, with every name doubt escalated.

Read online, listen to it, or change the text size: https://blogs.iictp.org/open-llms-sanctions-screening-alert-disposition-turkish-names-explained

What was tested?

Name screening raises alerts that are mostly false positives, and level-1 analysts clear them by checking date of birth and name against a written policy. The study asks whether open models can do that disposition. It built 12,200 synthetic customers from 7,103 natural persons on UK and EU lists, covering six name conditions (including Turkish spellings of Arabic-origin names and Turkish romanisations of Cyrillic names), two date-of-birth conditions, namesakes, near-name customers and a general population.

How it works

- **Deterministic screen.** A fuzzy and embedding screen raised 7,383 alerts.
- **Three open models** served on domestic infrastructure, and three rule baselines, dispositioned the alerts under the same written policy.
- **Pre-registration.** The test, including a non-inferiority margin, was registered in advance.

What is new compared with earlier approaches

Many reports of AI in compliance use proprietary models and unspecified data. This study uses public lists, local models and a pre-registered design, and reports its own error.

Results reported by the source

- Models sent 0.6 to 16.9 innocent customers per 1,000 screened to level 2, against 24.4 for a rule applying the policy's date clause.
- They stated the date-of-birth status correctly in 98.3% to 98.5% of alerts.
- They failed to escalate 1.5% to 2.2% of the listed persons.
- The pre-registered non-inferiority test was met, but only because of a construction error the author found afterwards: 21 "minor discrepancy" customers had conflicting dates. Without them the rule misses no listed person and no model meets the margin.
- Fuzzy matching found 69% of Turkish spellings of Arabic-origin names; adding embeddings raised this to 74%.

Limitations and open questions

- Customers are synthetic, so real behaviour may differ.
- The error in the dataset construction is reported by the author and weakens the headline result.
- A missed listed person is a regulatory failure, so a 1.5% to 2.2% miss rate is serious. Sources were read on 25 November 2026. Products, rules and standards change, so check the linked originals before you rely on any detail.

How a compliance team could test a tool like this

1. Build a labelled test set from public lists with name variants for your customer base.
2. Keep a deterministic date-of-birth check outside the model.
3. Escalate every name doubt rather than letting the model close it.
4. Measure missed true matches first and false positives second.

What it means for compliance and legal teams

The finding is conditional: models can cut workload, but not as the sole decision-maker. The author's own correction is a useful lesson in checking test design. This is a research result, not a regulatory position; check your supervisor's expectations. See EU AI Act Article 5 for the wider regulatory context.

Frequently asked questions

- **What did the study test?** Whether open LLMs can clear level-1 sanctions-screening alerts.
- **What is the headline risk?** They failed to escalate 1.5% to 2.2% of listed persons.
- **What did the author recommend?** Use open models only behind a deterministic date check and escalate every name doubt.
- **Is it peer reviewed?** No, it is a preprint.

Related reading

- EU AI Act Article 5: prohibited practices
- GlobalHRBench
- AI Practitioner certification

Related credential: AI Practitioner. Explore the exam facts and syllabus.

Sources and further reading

1. Huyut MM. Can Open LLMs Clear Sanctions-Screening Alerts? A Pre-registered Test on Public Sanctions Lists with Turkish Name Variants. arXiv 2611.28914, 21 Nov 2026 (preprint). arxiv.org/abs/2611.28914

Published by IICTP, an independent technical credential body, on 25 November 2026. The current version of this post is at https://blogs.iictp.org/open-llms-sanctions-screening-alert-disposition-turkish-names-explained. More guides: blogs.iictp.org.