OpenAccountants

RESEARCH REPORT · OA-RR-2026-01

Machine Access to Primary Tax Law

A measurement of whether an AI system can reach the official tax sources of 133 jurisdictions, and the eight where it cannot.

Published 13 August 2026 · Sources probed 356 · Jurisdictions 133 · Status Reproducible

Summary

The finding in one paragraph

> An AI can reach an official tax source in 79 of 133 jurisdictions. In eight it cannot reach one at all: Israel, Moldova, Cyprus, Mexico, Fiji, Samoa, Maldives and Chad. This does not make those countries' tax figures unknowable, and that is the more troubling result. The figures still circulate freely through commentary, advisory firms and payroll vendors. What is lost is not the number but the ability to check it, so an AI answers confidently from an intermediary while believing it consulted the government.

79
jurisdictions with a readable official source

8
where no official source can be reached

46
where no official source is held yet

18/28
blocks enforced by Cloudflare alone

A limit on every figure in this report: reachability was measured from one vantage point

All probes originated from a single network. A refusal recorded here means *this client, from this origin, was refused* — not that the source is closed to everyone. That distinction was established during review: an assistant running on different infrastructure retrieved a Maltese contribution table as clean markdown from a page that returned 403 to three separate clients here, including two on this machine and one on other infrastructure again.

The mechanism is egress identity — network reputation, TLS fingerprint, or a verified-crawler allowlist — not page rendering. The blocked pages are server-rendered and need no JavaScript, so a headless browser would not change the outcome. What changes it is who is asking. The counts below should therefore be read as **reachability for an unregistered client**, which is the position most self-hosted AI systems are in, and not as a property of the source itself.

Section 1

Why this was measured

Every argument for AI performing tax research rests on a single instruction: *confirm this figure against the official source*. That instruction assumes the source will answer. Where it does not, a language model does not stop and report failure. It answers from memory instead, and the result is indistinguishable from a verified answer.

The assumption had not been tested. This report tests it.

Section 2

The eight closed jurisdictions

In each of these, every official source on record refuses or fails to answer an automated request. This is a property of those countries' infrastructure, not of any particular software.

JURISDICTION	OFFICIAL SOURCE	RESPONSE	ARCHIVED COPY USABLE?
Israel	gov.il	403 CLOUDFLARE	No — archived copy empty
Moldova	sfs.md	403 CLOUDFLARE	No — archive stored the refusal
Cyprus	mof.gov.cy	NO RESPONSE	No — archive stored the refusal
Mexico	sat.gob.mx	TIMEOUT	No — archived copy empty
Fiji	frcs.org.fj	403	YES 84 KB · 29 days old
Samoa	revenue.gov.ws	403	PARTLY 60 KB · 190 days old
Maldives	mira.gov.mv	403	THIN 5 KB · 108 days old
Chad	finances.gouv.td	503	THIN 6 KB · 64 days old

Two major jurisdictions refuse at one door but not another and are therefore **not** on this list. **France** blocks at Legifrance while impots.gouv.fr answers. The **United States** blocks at congress.gov while ten other official sources answer, including the IRS, GovInfo and Cornell. Listing either as closed would have been the more striking claim and the wrong one.

Section 3

The archive is not a substitute

Malta's blocked pages proved readable through the Internet Archive, which suggested a general workaround existed. It does not.

> **When a site blocks automated visitors, it blocks the Internet Archive's crawler too.**
> What is preserved is therefore often the refusal itself. Requesting Malta's archived home page returns 513 characters stating that access is denied: the archive faithfully stored the locked door.

Coverage is page by page and depends on whether the crawler happened to get through on the day it visited. The same Malta host whose home page archived as a refusal has a social security page that archived intact, with the contribution table readable.

Where the archive does work, age erodes it. Tax rates are re-set annually, so a snapshot from 190 days ago (Samoa) or 108 days ago (Maldives) may straddle a change and be confidently wrong. Of the eight closed jurisdictions, **one** has an archived copy both substantial and recent enough to rely on.

Section 3b

A blocked authority launders its own figures

A refusal is selective. Cloudflare's default configuration admits verified search-engine crawlers while turning away everything else, so the pages of a blocked authority are typically *indexed in full* even though no ordinary client can open them. Malta's contribution-rate pages are present in search results, correctly titled, while returning 403 to three independent fetchers tested for this report.

The consequence is not ignorance. It is unattributable knowledge. A query for Malta's 2026 contribution rates returns them readily — 10% from employee and employer, €6.62 weekly up to a €229.44 wage, a fixed €55.93 above €29,084, a maximum of €83.89 — sourced from commentary sites, advisory firms, payroll providers and a salary calculator. The figures are almost certainly right. Not one of them comes from the Maltese government.

> **The real damage of a closed authority**
>
> Blocked sources do not silence a figure, they strip it of provenance. The number survives the journey; its date, its authority and any accountable party do not. An AI then presents an intermediary's copy with the confidence appropriate to a primary source, and no reader can tell the difference.

This is measurable in our own corpus, where 65% of citations were found to be non-primary. Some of that is carelessness. In the eight closed jurisdictions it is structural: there is no primary citation available to make.

Section 4

Who is holding the door shut

Twenty-eight sources actively refuse automated access. **Cloudflare enforces eighteen of them.** A single infrastructure provider therefore stands between AI systems and a material share of the world's published tax law, in all likelihood as an incidental consequence of default configuration rather than deliberate policy.

Nine of the refusing hosts are governments in their own right, including France's Legifrance, the United States Congress, Indonesia's regulations portal, Cayman Islands legislation, and the revenue authorities of Malta, Maldives, Samoa, Chad and Barbados.

A secondary finding: the polite convention is barely used.

Three sites of 356 use robots.txt to request that automated agents stay away. Twenty-eight enforce a refusal at the firewall. There is almost no middle ground: authorities either leave the door open or bolt it, and the mechanism designed for exactly this negotiation goes unused.

`Section 5`

What our own records got wrong

The probe was pointed at our own registry as well. It disagreed with reality 28 times in 356.

- **Seventeen sources believed reachable now refuse.** The world changed and nothing reported it.
- **Eleven recorded as blocked in fact answer.** This was a live defect: a source marked blocked with no recorded alternative is discarded when a Guide is served, so those jurisdictions were supplied with no official source at all while a working one sat unused.

Both have been corrected against measurement. The correction moves a flag only on conclusive evidence: a timeout is never recorded as a block being lifted, or the next automated visitor walks into a closed door with no warning.

`Section 6`

Method, and two errors caught before publication

Sources were requested with a standard browser identity, following redirects, on a 25-second timeout. robots.txt was read separately from the page itself, so a request to stay away is never mistaken for an enforced refusal. Blocks are attributed to a provider only where response headers or body identify one.

A success code is not a source

Eleven sources return success and no content. Argentina's revenue service replies with 274 bytes; Chile's with 166. Any checker trusting the status code alone would record these as verified. They are recorded separately as insubstantial.

A government cannot be identified by its web address

An initial pass classified authorities by domain pattern and reported that the United Kingdom, Ireland, the Netherlands, Sweden, Denmark and Norway held no government source, because their revenue authorities do not use `.gov` addresses. This produced a confident and incorrect headline. Every figure in this report uses curated classification instead.

Readable sources responded in 558 milliseconds at the median and none exceeded ten seconds, so where access is granted, latency is never the constraint. The dataset and the probe are committed to the repository and the measurement can be reproduced in full.

Section 7

What this means

For **79 jurisdictions**, the instruction to verify against a primary source is achievable, and an AI equipped with a sound method can genuinely perform tax research. That covers the majority of the world and the large majority of economic activity.

For **eight**, an unregistered client is refused. The limit is not the model's reasoning, its context window or its instructions, and it is not solved by a better prompt. It is a server declining to answer this particular caller. Two remedies exist and they are of very different quality: register as a verified crawler and be let in under a declared identity, or accept the figures second-hand from whoever republished them.

> **The professional consequence**
> Where the authority refuses, the figure does not disappear — its provenance does. An answer can still be produced, but nobody can say which source it came from, when it was read, or who stands behind it. That is the gap a named professional closes, and no amount of model capability closes it instead.

Section 8

Recommendations

1. **Register as a verified crawler before assuming any door is shut.** The refusals recorded here are aimed at anonymous clients, not at crawlers with a declared public identity. Registration is free, is the mechanism these blocks exist to encourage, and would convert a large share of this report's negative findings. For an organisation whose entire proposition is attribution, being a named agent rather than an unidentified address is also the position of principle.

2. **Record an alternative route for every source that stays blocked.** Fifty-one sources are flagged as blocked and only one has an alternative recorded, so fifty currently resolve to nothing at all. Malta demonstrates the pattern: where the revenue authority is closed, the statute repository is frequently open.

3. **Separate our gaps from the world's.** Forty-six jurisdictions have no official source on file. Some genuinely publish nothing accessible; for others we have simply not looked. Counting them together overstates the closure.

4. **Repeat quarterly.** Twenty-eight changes since the previous records were written answers the question of whether a single measurement suffices.

5. **Publish.** No comparable measurement appears to exist for tax authorities. It is of direct interest to the profession, and it is the clearest available evidence for why a named professional remains necessary.

Every figure in this report is reproducible from the committed probe and dataset. Where a result was uncertain it is marked uncertain rather than resolved into a cleaner narrative: eight closed jurisdictions is a smaller and less dramatic number than an earlier cut of the same data suggested, and it is the number the evidence supports.