

When Remote Is Not Enough: A Frozen-Evidence Audit of Geographic Eligibility Signals in Job-Search Snippets

Exzil Calanza

Independent research preprint. This work is not peer reviewed.

Source commit: `ecd71496f10ec95841123e956c5433e81dc310dd`

The reported counts characterize only the frozen experiments and production sample described here; they are not universal reliability estimates.

Abstract

Remote-work discovery systems often expose a headline and short search snippet before a candidate opens the full source post. This study asks a narrow measurement question: when a job post contains an explicit hard geography or nationality eligibility constraint, is that constraint visible in the indexed snippet that led to the post? We reanalyzed a frozen 8 August 2026 audit containing 16 harvested application-address records and signed-in rendered-page captures. To avoid address-level double counting, records were collapsed by unique source URL and only source posts with at least one VERIFIED_LIVE record were retained. Ten unique posts met that criterion. Nine contained at least one explicit hard geography/nationality constraint under an amended, verbatim phrase lexicon; eight of those nine had at least one constraint absent from the frozen search snippet. The observed hidden-constraint proportion was $8/9 = 0.889$, with a Wilson 95% interval of 0.565–0.980. Because this is a small convenience sample gathered by four targeted searches during one week, the result is not a prevalence estimate for the remote-job market. It is evidence that snippet-level screening can omit eligibility information that becomes visible only after opening the source post.

1. Introduction

“Remote” describes work arrangement, not necessarily applicant geography. A remote role can still require residence near a company site, a specific country, a specific city, or a nationality. Search snippets and social-post previews are therefore incomplete decision surfaces: they may expose the remote label while omitting the condition that determines whether an applicant is eligible.

A prior operational audit collected candidate job posts using four Philippines-oriented searches and then opened each source in a signed-in browser. The original report was intentionally descriptive and warned against generalizing from one week of searches. The present study is a secondary reproducibility analysis over the frozen evidence. Its contribution is methodological rather than market-wide: it corrects the unit of analysis from application addresses to unique source posts, uses an explicit phrase-level coding protocol, and measures whether hard eligibility signals visible in the rendered body were also visible in the frozen search snippet.

2. Data provenance

The frozen evidence comes from `data/runs/jobsources_canary_20260808/`. The candidate inventory contains 16 harvested address records. The verifier output contains the source URL, indexed snippet, rendered signed-in page text, page title, screenshot path, and a fail-closed verdict for each record.

The two principal inputs are hash-bound:

- `candidates.json`: `5f0748dc8a529c2682d33759b0123f597725400bfa11f1f4bcb9426971bdf5e4`
- `verified.json`: `d4fceb8a3695069b7b7b59f097313ab8dade5189d25448c71b109b215ab211e8`

The original verifier itself hashes to `ae26797fa1ba188b8257e4dad37228c3123892518e34c449d873304e8c68afea`. A fresh capture of the already-published public baseline on 6 October 2026 hashes to `ab8a6fd07bc92776c1e49f9bf5925bf066597095b67d9e5e4eada280ffee87bf`; that capture is used only to confirm the public baseline, not to recode the August source posts.

3. Methods

3.1 Eligibility and unit of analysis

The analysis unit is the unique `source_url`. This prevents a single job post containing two contact addresses from counting twice. A source enters the analysis only if at least one frozen record for that URL has verdict VERIFIED_LIVE.

The 16 address records correspond to 14 unique source URLs. Two source URLs appear more than once because the post exposed multiple contact addresses. Eleven address records passed the original live verifier; after URL-level deduplication, these represent 10 unique verified-live source posts.

3.2 Hard-constraint coding

A hard constraint is an explicit rendered-page phrase that geographically or nationally bounds applicant eligibility or the role's work location. The accepted detector uses exact, case-insensitive phrase patterns for:

- a must-be-located-in-the-Philippines requirement;
- a section explicitly labeled “U.S. Remote Roles”;
- “Remote (U.S.)”;
- Abu Dhabi/UAE location;
- Ahmedabad on-site location;
- “Jordanian only”;
- Lagos on-site location;
- must-reside-in-Greater-Metro-Manila;
- “Filipino applicants residing in the Philippines only.”

Generic “remote,” “work from home,” time-zone requirements, hardware requirements, U.S. experience requirements, #HiringPH, and a bare “U.S.-based” token are insufficient by themselves.

The bare “U.S.-based” token was removed in a protocol amendment after a provisional pass showed that it can describe teams or roles being recruited for rather than applicant eligibility. The amended version is the only accepted analysis.

3.3 Snippet visibility

For each hard constraint detected in the rendered body, the same constraint group is searched in the frozen indexed snippet. A post is counted as having a hidden constraint when at least one body-level hard constraint is absent from its snippet.

3.4 Statistical summary

The primary statistic is the fraction of hard-constraint source posts with at least one hidden constraint. A Wilson 95% interval is reported because the denominator is small. No population weighting, significance testing, or market-prevalence inference is attempted.

4. Results

Nine of the ten unique verified-live source posts contained at least one explicit hard geography/nationality constraint under the amended coding rule. Eight of those nine had at least one such constraint absent from the frozen indexed snippet.

Thus:

- hard-constraint prevalence within the verified-live convenience sample: $9/10 = 0.900$;
- hidden-constraint rate among hard-constraint posts: $8/9 = 0.889$;
- Wilson 95% interval for the $8/9$ proportion: $0.565\text{--}0.980$.

The single hard-constraint case whose snippet already exposed the relevant signal was a posting whose snippet explicitly named “U.S. Remote Roles.” The hidden cases included rendered-page constraints such as residence near Philippine company sites, Philippine-resident-only applicants, Ahmedabad on-site work, Abu Dhabi work location, a U.S.-remote designation, a Jordanian-only requirement, a Lagos on-site role, and a Greater Metro Manila residence requirement.

These categories are not equivalent in direction: some restrict applicants away from the Philippines, while others positively constrain a role to Philippine residents. The common measurement property is that the hard eligibility/location condition is material to fit and was absent from the snippet.

5. Discussion

The result demonstrates a failure mode in snippet-first job discovery: a search result can be relevant enough to click while still omitting the field that decides whether the candidate can plausibly apply. The operational consequence is not that snippets are “wrong.” They are compressed representations. The consequence is that remote-job filtering should treat the rendered source body—not the snippet—as the eligibility surface of record.

The URL-level deduplication matters. Counting email addresses would overweight posts that list multiple contacts. The secondary analysis reduces 11 verified-live address records to 10 unique verified-live sources and thereby aligns the denominator with the object being evaluated: the post.

The protocol amendment also illustrates why phrase-level automation requires semantic guardrails. A bare “U.S.-based” phrase can occur in “U.S.-based teams” or “US-based roles” without itself specifying where the applicant must reside. Narrowing the lexicon to explicit applicant/location phrases keeps the measurement closer to the construct.

6. Limitations

This study is deliberately narrow. The sample is small, nonrandom, and generated by four targeted searches during one week. It is centered on a Philippines-based job seeker and social-source posts that exposed application addresses. Five harvested address records did not pass the original fail-closed live verification and are not treated as evidence for the primary endpoint.

The analysis does not test whether a candidate would ultimately be accepted, whether employers would make exceptions, whether source posts remained open after 8 August 2026, or how often hidden constraints occur across job boards generally. The Wilson interval quantifies binomial uncertainty for the observed proportion but does not cure the convenience-sampling design.

7. Reproducibility and evidence boundary

The accepted v2 output is `candidate_158862_secondary_validation_results_v2_20261006.json`, SHA-256 `98e61213fd33e8f1a3ee7d9277a38668a3a368f72c6d9aea7f202a4cabf532f9`. The analysis script verifies its frozen input hashes before producing results, collapses duplicate URLs, and applies only the amended lexicon.

This preprint does not reproduce recruiter email addresses. Short eligibility phrases are used only where needed to explain coding. The underlying rendered-page captures remain local evidence.

8. Conclusion

In this fixed August 2026 convenience sample, 8 of 9 unique verified-live source posts with explicit hard geography/nationality constraints had at least one such constraint missing from the indexed snippet that surfaced the post. The defensible conclusion is limited: snippet-level relevance is not sufficient evidence of applicant eligibility. For remote-job workflows, source-body verification should precede application effort.

— Skynet research artifact