

HEZRON
INSIGHTS

Published, But Can AI Find It?

What Kenyan Organisational Websites Reveal
About Readiness for AI Search

Visual research report | August 2026

Hezron Ochiel | Hezron Insights

244
organisations

9
sectors

7
conditions

30.3%
met all seven

96.1% showed evidence of publishing in 2026

48.1%
explicit
provenance

36.5%
had a high-
confidence risk

From publishing activity to source quality

The headline finding

Kenyan organisations are publishing actively. Far fewer meet the full set of measured source-readiness conditions.

GLOBAL VIEW

96.1%

of organisations with a verifiable latest-content date showed evidence of publishing in 2026.

30.3%

of complete cases met all seven measured content-readiness conditions at the same time.

The gap matters

Fresh publishing does not automatically resolve provenance, duplication, template residue, numerical integrity or entity consistency problems.

AI search changes what organisations compete on

The source itself now matters alongside the page that contains it.

Traditional web discovery

Pages compete to rank, attract a click and persuade a visitor once they arrive.

AI-powered discovery

Content may be retrieved as grounding material, selected as a citation or incorporated into a generated answer.

The organisational question

Can a machine identify who is speaking, understand the structure, reconcile the facts and distinguish the intended source from noise?

This report measures source conditions. It does not measure actual citation or ranking.

What the audit measured

Seven observable source conditions across a frozen cross-sector sample.

01

Meaningful taxonomy
Interpretable organisation of topics and institutional information.

02

Explicit provenance
Named human or identifiable institutional department.

03

No template residue
No placeholder, demo or page-builder noise.

04

Numerical integrity
No coded questionable numerical values.

05

Duplication control
No substantial competing versions of the same information.

06

Entity consistency
Stable organisational identity and factual coherence.

07

Current-year publishing
Latest verifiable content date falls in 2026.

These are analytical indicators - not universal platform requirements.

How to read these numbers

The report keeps missingness visible and avoids turning unresolved observations into failures.

244

organisations

Frozen cross-sector organisational sample

9

sectors

Universities, ministries, agencies, counties, finance, media, health, listed firms and PBOs/NGOs

198

complete cases

Determinate observations on all seven conditions

21

Aug

2026

Common study date and dataset freeze

Determinate denominator

Each condition uses the cases that could be classified confidently.

Complete-case conjunction

The all-seven result includes only organisations with determinate observations on every condition.

Sector comparisons are descriptive because the nine strata use mixed sampling designs.

FINDING 1

Publishing is active.
Source quality is uneven.

Current publishing is common. Full conjunction is not.

Two very different measures tell the story.

96.1%

evidence of current-year publishing

199 of 207 organisations with a verifiable latest-content date

Publishing activity
contains less information
about source quality



30.3%

met all seven measured conditions

60 of 198 complete cases

A recent date is a useful signal of activity. It does not tell us whether the page has strong provenance, clean templates, reconciled numbers, controlled duplication or stable entity information.

The seven conditions are not equally common

Explicit provenance is the clear outlier.



Share of determinate observations meeting each measured condition.

Provenance is the weakest measured condition

The key question is not always whether a named person appears. It is whether responsibility is legible.

103 / 214

determinate organisations exposed either a named human or an identifiable institutional department.

Named human

Useful where readers reasonably expect an author, expert, reviewer or reporter.

Institutional owner

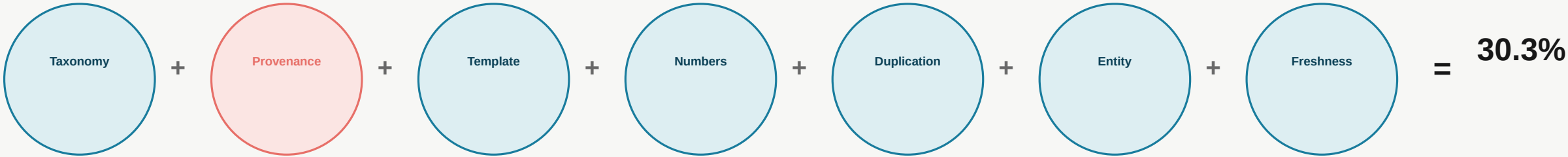
A ministry, department, clinical unit or corporate function can also provide accountable provenance.

Interpretation: the study does not treat every non-byline page as illegitimate. The sensitivity analysis tests how much the strict result changes when provenance is removed.

Provenance was coded as named human OR identifiable institutional department.

The all-seven result is a screen, not an AI-readiness score

A failed condition does not carry the same consequence for every organisation or content type.



The calculation asks one narrow question: Did this organisation satisfy every measured condition at once?

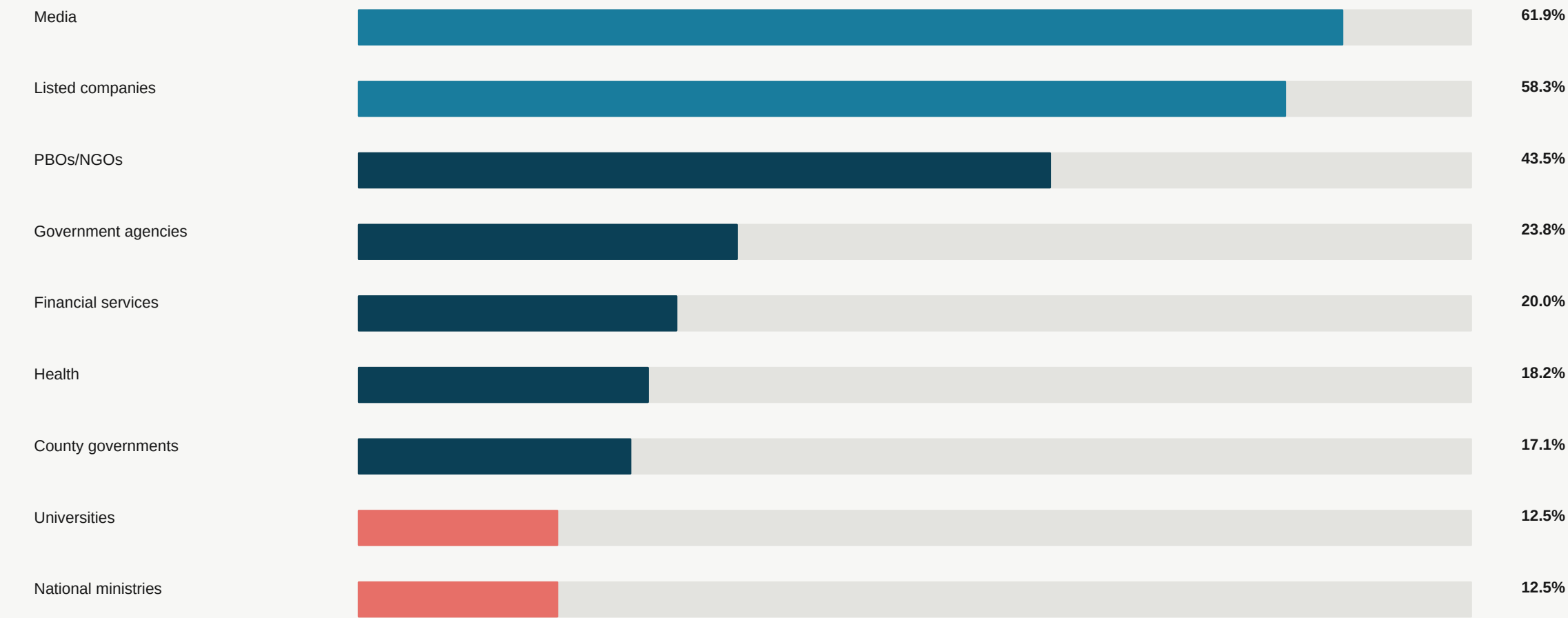
No weights or latent score were applied.

FINDING 2

Sector patterns
are visible - and descriptive.

Media and listed companies lead the strict all-seven screen

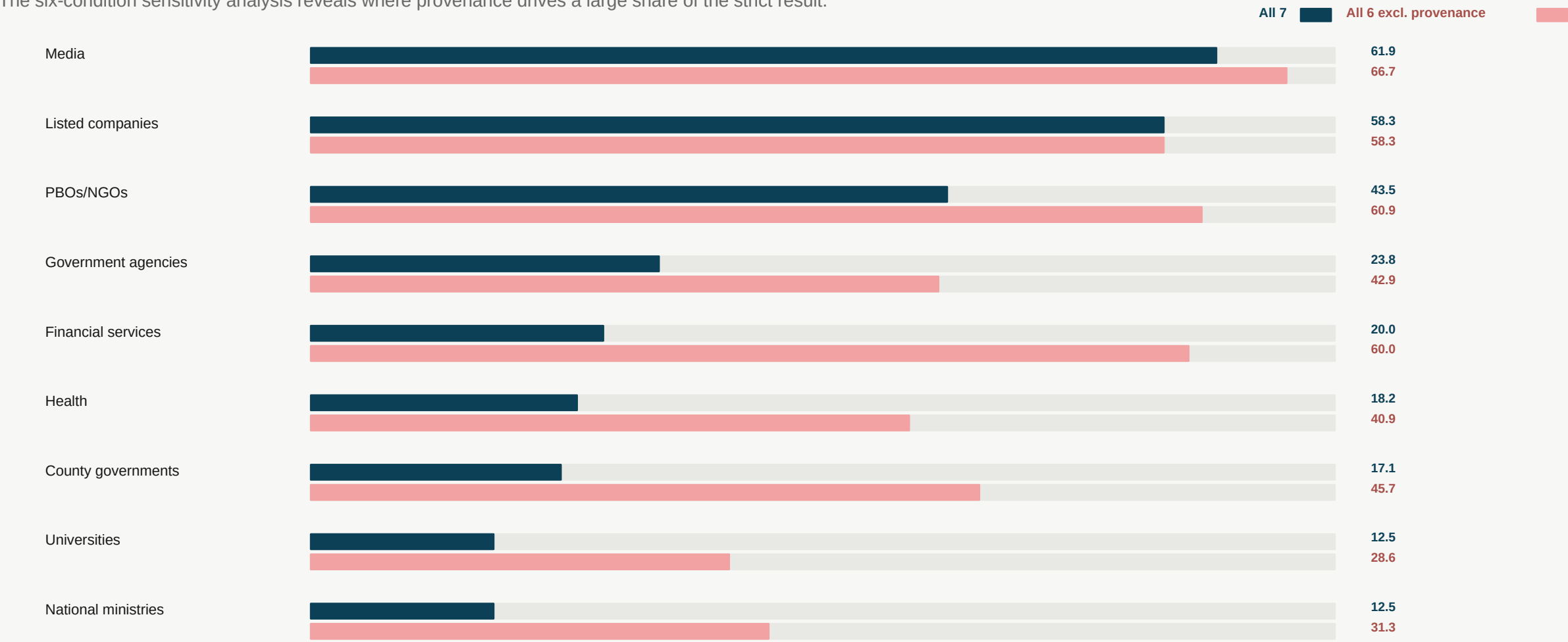
Universities and national ministries sit at the bottom of this descriptive sample.



All 7 uses complete cases within each sector. Mixed sampling designs mean these are descriptive contrasts, not national sector estimates.

Removing provenance changes the picture most in some sectors

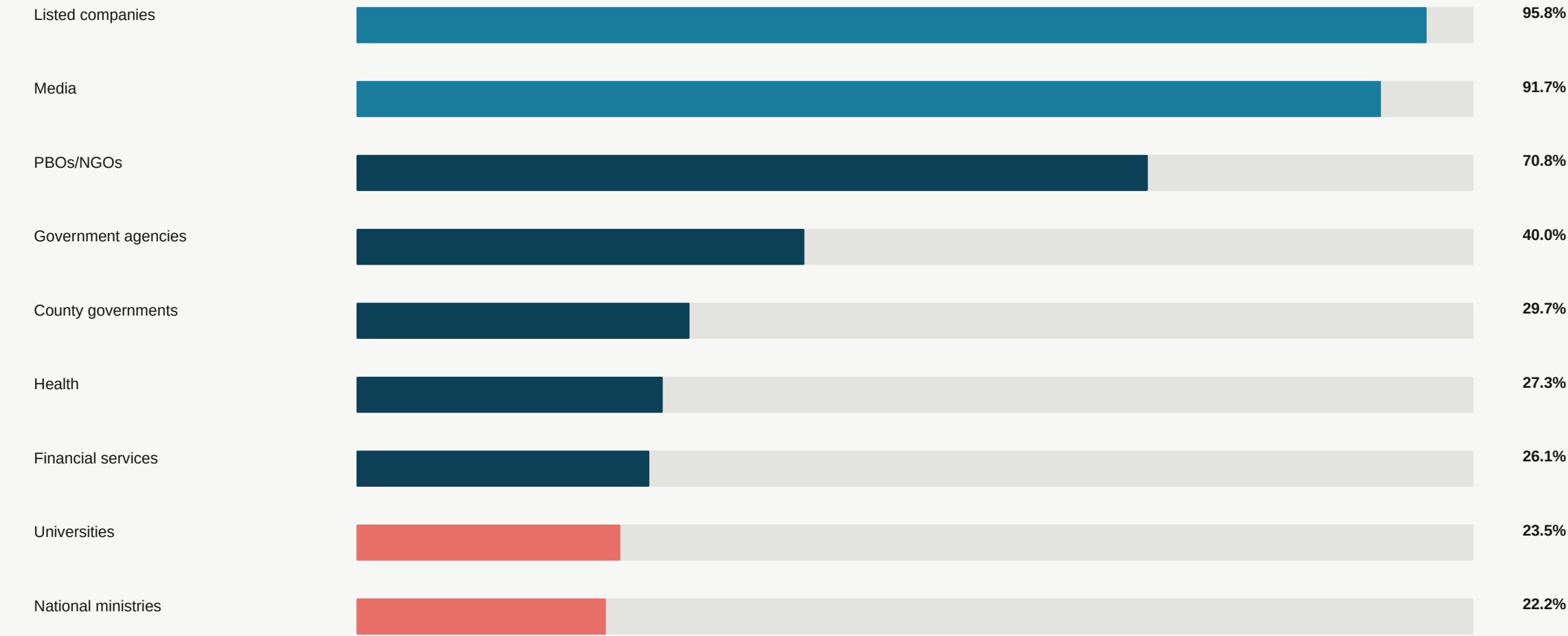
The six-condition sensitivity analysis reveals where provenance drives a large share of the strict result.



All 6 excludes explicit provenance and uses its own complete-case denominator.

Provenance conventions differ sharply by sector

Media and listed companies show the strongest explicit provenance in the audit.



Explicit provenance = named human OR identifiable institutional department; each sector uses its own determinate denominator.

Numerical integrity is relatively strong overall, but uneven by sector

PBOs/NGOs and several public-facing sectors have more room to improve.



Numerical integrity = no coded questionable numerical values; each sector uses its own determinate denominator.

FINDING 3

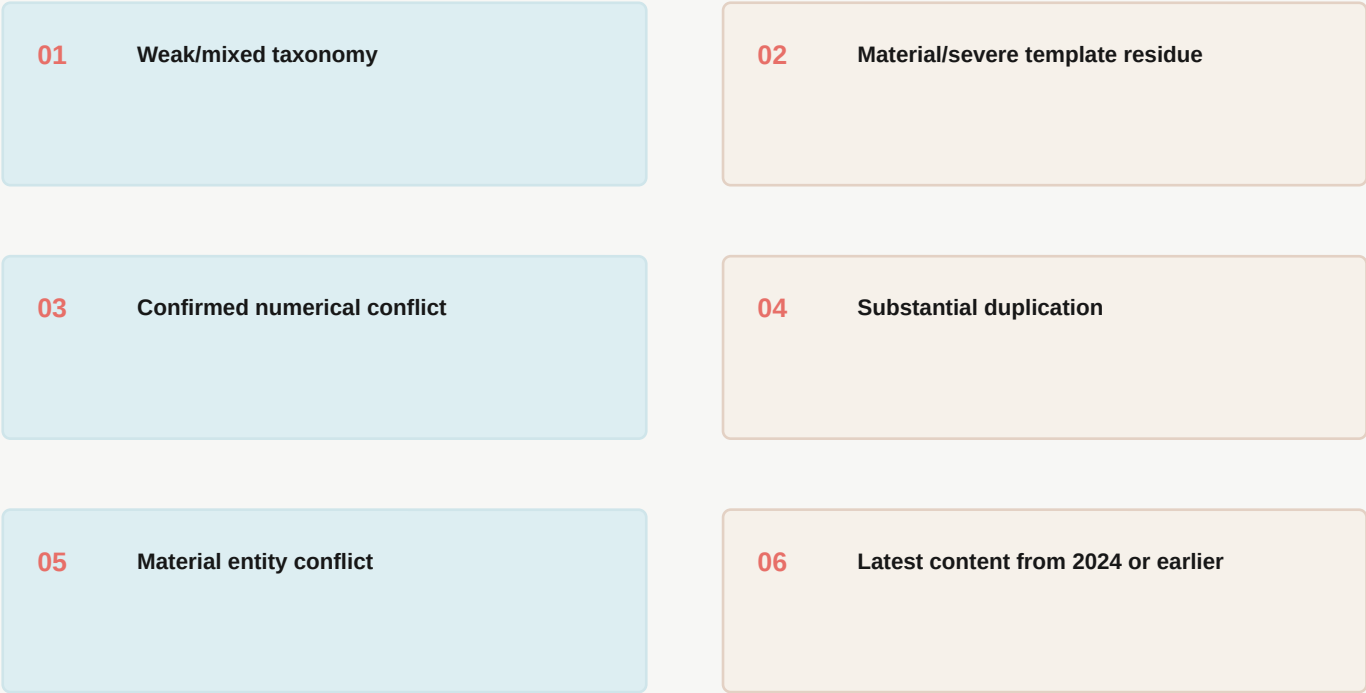
Fresh pages can still
carry machine-facing risks.

More than one in three organisations showed a high-confidence content risk

The risk screen counts six content-facing signals.

36.5%

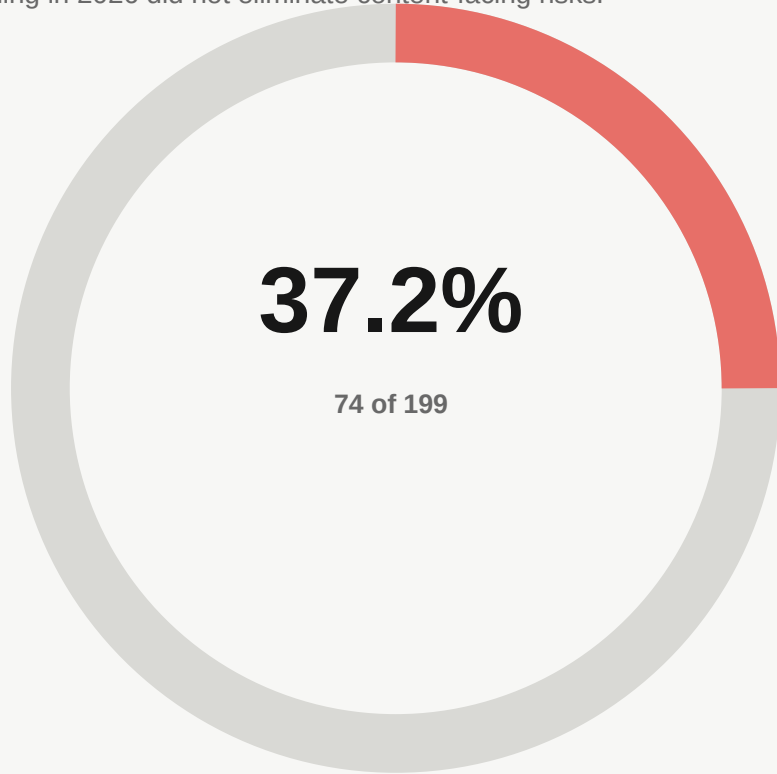
89 of 244 organisations



The risk prevalence screen does not include broken/development links; those are used only in co-occurrence analysis.

Even current publishers can have material source-quality problems

Publishing in 2026 did not eliminate content-facing risks.



organisations with verifiable 2026 content still exhibited at least one high-confidence risk.

Current publishing and source quality are separate dimensions.

A team can publish frequently while legacy content, placeholders, conflicting statistics or duplicated pages remain available for retrieval.

The risks often cluster around publishing workflow problems

Template residue repeatedly appears alongside broader web-governance signals.

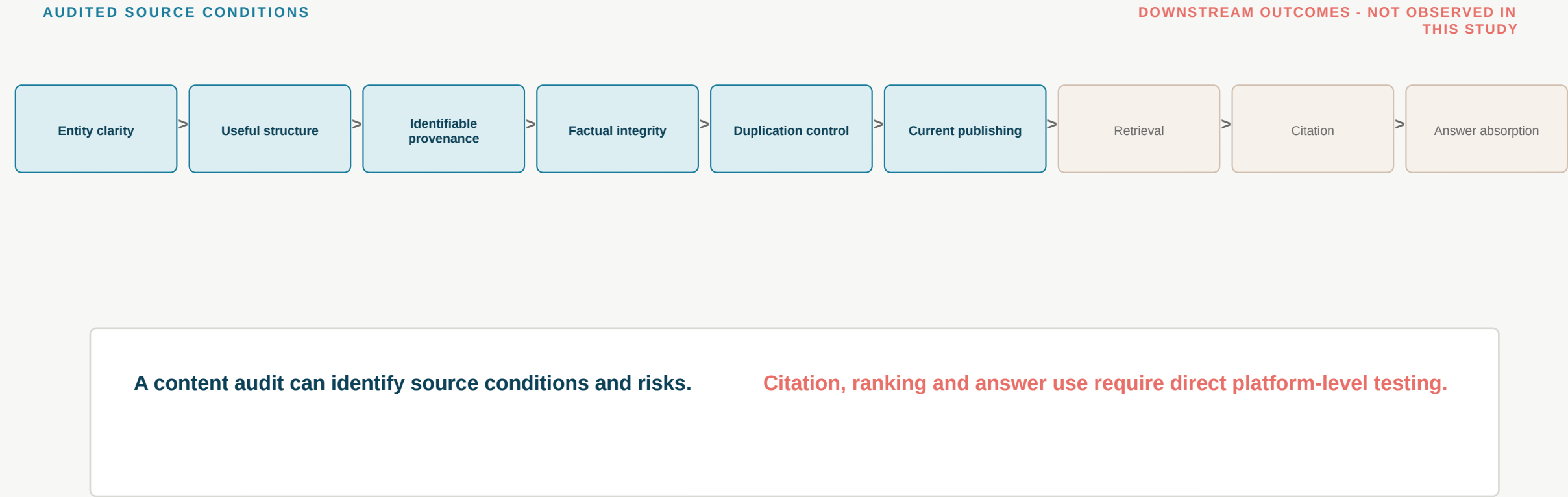


The practical signal: content defects are often symptoms of workflow and governance issues, not isolated editing mistakes.

Co-occurrence analysis includes broken/development links as supporting web-governance evidence.

A practical content-readiness model

Fix upstream source conditions before expecting downstream AI-search outcomes.



Six practical priorities for communications and digital teams

The strongest opportunities sit inside editorial governance.

01 **Stabilise entity identity**

Keep major organisational facts consistent across About, service, news and report pages.

02 **Use useful structure**

Give pages descriptive headings, meaningful categories and clear information hierarchy.

03 **Make responsibility legible**

Identify responsible authors, departments or institutional owners where appropriate.

04 **Reconcile consequential numbers**

Attach evidence to important statistics and review conflicting figures across pages.

05 **Control duplicate and development content**

Keep legacy, staging and duplicate versions from competing with the intended source.

06 **Review evergreen pages**

Check old claims, dates, contacts and service information on a defined cycle.

Practice implications are derived directly from the study findings and discussion.

What this report does not prove

The boundary is as important as the findings.

Not a national prevalence estimate

The sample describes 244 audited organisations; several strata are purposive or balanced pilots.

Not a universal AI-readiness index

The seven-condition conjunction is a transparent descriptive screen, not a validated certification standard.

Not proof of citation or ranking

The study did not test ChatGPT, Google AI Mode/Gemini, Copilot, Claude or Perplexity at query level.

Not a causal estimate of provenance

The 30.3% to 48.8% change uses different complete-case denominators and is a sensitivity analysis.

These limitations are stated in the manuscript and should travel with any public summary of the findings.

Methodology at a glance

A secondary analysis of a frozen organisational website audit.

1

Sample

244 Kenyan organisations across nine strata.

2

Coding

Structured, AI-assisted and evidence-linked first-pass categorical coding against a project codebook.

3

Missingness

Indeterminate observations excluded from substantive denominators and reported separately.

4

Analysis

Condition prevalence, descriptive sector profiles, strict all-seven conjunction and provenance sensitivity analysis.

5

Integrity checks

Scripted audit confirmed unique IDs, valid categories, common study date and evidence URL retention.

6

Reliability boundary

No blinded second human coder independently recoded a stratified subsample; no inter-coder reliability coefficient is reported.

Limitations to keep in view

The report is designed to support disciplined interpretation, not stronger claims than the data can carry.

- 1 Secondary analysis rather than a new page-level corpus sampled specifically for AI-search content research.
- 2 No independent second-coder reliability test for judgement-dependent categories.
- 3 Mixed sampling designs mean sector comparisons are descriptive.
- 4 Websites are dynamic; findings represent the 21 August 2026 audit date.
- 5 The seven-condition conjunction is not a validated universal AI-readiness index.
- 6 **No direct measurement of retrieval, citation selection, ranking or answer absorption across AI-search platforms.**

The manuscript also notes that shared use of the frozen audit dataset across related studies should be disclosed to editors where relevant.

The bigger conclusion

AI-search preparation is an information-governance challenge before it is a platform tactic.

Current publishing activity + **Machine-facing source quality** = **Readiness to be tested**

The next empirical step is direct platform testing: do these source conditions predict which Kenyan organisational pages AI-search systems retrieve, cite and incorporate into answers?

That is the point where source-readiness evidence can become citation evidence.

Study, selected references and author

The visual report is a public-facing companion to the full academic manuscript.

STUDY

From Publishing to Citability: AI-Search Content-Readiness Signals Across Kenyan Organisational Websites

Hezron Ochiel | Hezron Insights, Nairobi, Kenya | Study date: 21 August 2026

SELECTED REFERENCES

- Aggarwal et al. (2024). GEO: Generative Engine Optimisation. ACM SIGKDD.
- Google Search Central (2025a, 2025b, 2026). Helpful content and AI-search guidance.
- Microsoft Bing Webmaster Blog (2026). AI Performance in Bing Webmaster Tools.
- Wang & Strong (1996). Beyond accuracy: What data quality means to data consumers.
- Moreau & Missier (2013). PROV-DM: The PROV Data Model. W3C.
- Hogan et al. (2021). Knowledge graphs. ACM Computing Surveys.
- Lombard et al. (2002); Krippendorff (2004). Inter-coder reliability guidance.

HEZRON OCHIEL

Founder, Hezron Insights

Hezron Insights publishes research, frameworks and practical analysis on AI visibility, digital authority, strategic communication, reputation and public relations.

Full manuscript and supporting materials should accompany any academic or technical use of these findings.

hezroninsights.com