Barkhausen AI

METHOD NOTE · BA-MN-1

How to read a visibility percentage

A claim such as "62% of AI answers mention this brand" is only as trustworthy as the sampling behind it, and most published versions of this claim omit the information needed to judge it.

“Every leap begins as a crackle.”

DOCUMENT	BA-MN-1
KIND	Method note
DATE	2026
SUBJECTS	Measurement & statistics
LICENSE	CC-BY-4.0
AUTHOR	Barkhausen AI

ABSTRACT

A claim such as "62% of AI answers mention this brand" is only as trustworthy as the sampling behind it, and most published versions of this claim omit the information needed to judge it. This note sets out the minimum arithmetic a reader needs: what a sample size (n) contributes to a proportion's precision, how the standard error of a proportion behaves as n grows, why proportions near 0% or 100% require a different interval than the standard formula, and which disclosures — engine, version, time window, and query construction — a claim must carry before it is evaluable at all. It closes with a five-item checklist for interrogating any AI-visibility percentage encountered outside a peer-reviewed setting.

CONTENTS

A visibility percentage — "brand X appears in 62% of AI answers to category Y queries" — is a point estimate of a proportion, and every point estimate of a proportion carries an implicit margin of error that shrinks as the number of observations grows. Most published visibility percentages report the point estimate and drop the margin of error entirely, which makes the number impossible to evaluate: a "62%" built from 20 sampled queries and a "62%" built from 2,000 sampled queries are not the same claim, even though they print identically. This note gives the arithmetic needed to tell them apart, and a short checklist for applying it to any percentage encountered in the wild.

1 What the claim must state before it can be judged

A visibility percentage is evaluable only if it discloses, at minimum, the number of observations (n) it was computed from, the AI system and version queried, the time window during which the queries ran, and whether the figure reflects one fixed query phrasing or a distribution of phrasings for the same underlying question. Answer engines are known to give materially different answers to differently worded restatements of the same question, so a percentage built from a single phrasing describes that phrasing, not the underlying topic. Absent these four disclosures, a reader cannot distinguish a well-powered measurement from a screenshot of one lucky (or unlucky) run.

2 The arithmetic behind the percentage

Treat each sampled query as a Bernoulli trial: the brand either appears in the answer or it does not. For an observed proportion $\hat{p}$ from n independent trials, the standard error is

$$SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

and a conventional 95% confidence interval is approximately $\hat{p} \pm 1.96 \times SE$. The width of that interval — not the point estimate alone — is what tells a reader how much to trust the number. Table 1 shows how the interval narrows as n grows, holding $\hat{p} = 0.62$ fixed throughout.

N	95% CI FOR $P = 0.62$	INTERVAL WIDTH
10	[31.9%, 92.1%]	60.2 points
20	[40.7%, 83.3%]	42.5 points
50	[48.5%, 75.5%]	26.9 points
100	[52.5%, 71.5%]	19.0 points
500	[57.7%, 66.3%]	8.5 points
2,000	[59.9%, 64.1%]	4.3 points

A "62%" reported from $n = 20$ carries a 95% interval of roughly 41–83%: the claim is consistent with the brand appearing in anywhere from two-fifths to five-sixths of answers, which is not a precise measurement. The same point estimate from $n = 500$ narrows to roughly 58–66%, a claim precise enough to support a comparison against a competitor or a prior period. The lesson generalizes beyond this specific $\hat{p}$: any percentage reported without n should be read as unspecified-precision until n is supplied, and any n below several hundred should be read as a rough estimate rather than a settled figure.

3 Why proportions near the edges need a different formula

The standard-error formula above degrades badly when $\hat{p}$ is close to 0 or 1. Suppose a claim reports that a brand appeared in every one of 20 sampled answers (100%). The naive interval, $\hat{p} \pm 1.96 \times SE$, collapses to exactly [100%, 100%], which asserts a certainty no finite sample can support. The Wilson score interval [1] corrects this by solving for the interval directly from the binomial likelihood rather than approximating it with a normal curve:

$$\hat{p}_{Wilson} = \frac{\hat{p} + \frac{z^2}{2n}}{1 + \frac{z^2}{n}}, \quad \text{half-width} = \frac{z}{1 + \frac{z^2}{n}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}}$$

For 20 of 20 successes and $z = 1.96$, the Wilson interval works out to approximately [84%, 100%] — a genuinely different and more defensible claim than "100%, full stop." The same correction applies at the low end: a brand that appeared in zero of 20 sampled answers is not proven absent; the Wilson interval for $\hat{p} = 0$, $n = 20$ runs to roughly [0%, 16%]. Any claim of "never appears" or "always appears" built from a small sample should be read through this correction rather than taken as exact.

4 Limitations

The corrections above assume the sampled queries are independent draws from the population of interest. In practice, a batch of near-identical query phrasings submitted in quick succession is not n independent trials in the statistical sense — the effective sample size can be smaller than the query count suggests, and a disclosed n should be read as an upper bound on precision, not a guarantee of it. Separately, none of this arithmetic addresses drift: an interval computed from one sampling window describes the engine as it behaved during that window only, and answer engines change without notice, so a precise interval from three months ago does not certify the same precision today.

5 Checklist: five questions before citing a visibility percentage

1. Does the claim state n, the number of observations the percentage was computed from?
2. Does it state a confidence interval or margin of error, not just the point estimate?
3. Does it state the time window during which the queries were run?
4. Does it name the specific engine and version queried, rather than "AI" generically?
5. Does it disclose whether the figure reflects one query phrasing or a distribution of phrasings for the same question?

A claim that answers all five is evaluable, even if the underlying number turns out to be modest. A claim that answers none of them is not yet a measurement.

REFERENCES

1. E. B. Wilson, Journal of the American Statistical Association. *Probable inference, the law of succession, and statistical inference* (1927). https://doi.org/10.1080/01621459.1927.10502953 Accessed 2026-07-08.

HOW TO CITE

Barkhausen AI (2026). How to read a visibility percentage. https://barkhausen.ai/notes/how-to-read-a-visibility-percentage/

Barkhausen AI · Est. MMXXVI · *Every leap begins as a crackle.*