Barkhausen AI

METHOD NOTE · BA-MN-2

Why time windows matter

A visibility percentage is often reported as though it were a fixed property of an AI engine, but the engines are non-stationary: identical prompts submitted on consecutive days return substantially different answers, overlapping by only 34–42% in the sources they cite and, on the study's brand-qualifying verticals, 45–59% in the brands they mention.

“Every leap begins as a crackle.”

DOCUMENT	BA-MN-2
KIND	Method note
DATE	2026
SUBJECTS	Measurement & statistics
LICENSE	CC-BY-4.0
AUTHOR	Barkhausen AI

ABSTRACT

A visibility percentage is often reported as though it were a fixed property of an AI engine, but the engines are non-stationary: identical prompts submitted on consecutive days return substantially different answers, overlapping by only 34–42% in the sources they cite and, on the study's brand-qualifying verticals, 45–59% in the brands they mention. A percentage measured over one span of dates therefore estimates a quantity that exists only for that span. This note argues that the time window is part of a visibility figure's identity, not metadata attached to it: a percentage reported without its window names no fixed quantity, and two figures drawn from undisclosed windows cannot be compared — the comparison is undefined, not merely imprecise. It sets out how to choose a window (short enough for approximate stationarity, long enough for the repetition a precision target requires), a house notation for stating one, and the perishability disclosure every figure should carry.

CONTENTS

A visibility percentage describes how an AI engine answered a set of prompts during the days those prompts were run, not the engine in the abstract. That distinction would be pedantic if answer engines were stationary — if the same prompt drew the same answer today, next week, and next month — but they are not. The span of dates over which a figure was measured is therefore not administrative metadata recorded for completeness; it is part of what the figure means, in the same sense that a temperature reading means nothing without the moment it was taken. This note sets out why the window belongs to a figure's identity, what follows when it is omitted, and how to state one.

1 Engines are non-stationary

The direct evidence is repetition. When identical prompts are submitted to the same engine on consecutive days, the answers do not repeat. A 2026 study that measured this reported day-to-day overlap of only 34–42% in the sources an engine cited and — on the three verticals where brands were reliably detected — 45–59% in the brands it mentioned [1]. Read plainly: with no change to the prompt and roughly a day between runs, more than half of the cited sources, and roughly half of the mentioned brands, turned over. The same study found that stability differs markedly across engines even within a single day, which is why an instability figure must be reported per engine, never pooled into a single average that describes none of them.

This is not sampling noise that shrinks with more draws. Sampling noise is variation around a fixed quantity; here the quantity itself is moving, because retrieval indexes, ranking, and model versions change underneath the measurement. Averaging more prompts on a single day sharpens the estimate of that day's engine. It does nothing to make that day representative of the following week.

2 A percentage without a window names no quantity

Because the engine changes underneath the measurement, every visibility percentage carries an implicit clause: brand X appeared in 62% of answers *during these dates*. Keep the dates and the number refers to something definite: the engine's behavior over a stated interval. Drop them and it picks out no fixed quantity, because there is no single, time-invariant "percentage of answers" for a system that answers differently from one day to the next.

The consequence for comparison is stronger than a loss of precision. Two visibility figures whose windows are undisclosed cannot be compared at all. This is not the familiar complaint that the comparison is noisy or underpowered; it is that the comparison is undefined. A comparison needs a common quantity that both figures estimate, and two figures drawn from unstated, possibly non-overlapping intervals of a moving system have no such common quantity. A reported jump from 51% to 76% might reflect a real change in the engine, a change in the window, or both, and without the windows there is no way even in principle to tell which — so the difference is not a measurement of anything. The requirement to disclose the window is set out as a minimum in BA-C-4; this note is the reason behind it.

3 Choosing a window

A window is a compromise between two requirements that pull in opposite directions. It must be short enough that the engine is approximately stationary across it, so that the interval names a single regime rather than averaging over a version change midway through. And it must be long enough to accommodate the number of repeated draws the precision target requires — the arithmetic relating sample size to interval width is the subject of BA-MN-1 — spread across enough distinct days that the draws are not all a snapshot of one anomalous afternoon.

These requirements can conflict. If an engine is changing fast enough that only a two-day window is defensibly stationary, and the precision target needs more repetition than two days of sampling can supply, the honest response is to widen the reported interval, not to silently extend the window across a regime change and present the result as a single stable figure. Window length is a stated modeling choice, and the sampling and change-point requirements that constrain it are specified in BA-C-3.

4 Stating a window

A window is only useful if it is stated unambiguously, and a bare month name ("June figures") is not unambiguous — it hides whether the measurement ran for two days or thirty. The house convention states a window as an ISO 8601 date range with an arrow, for example `2026-06-01 → 2026-06-28`, naming the first and last day on which prompts were actually run. The range is the span of collection, not the calendar month it falls within, and it travels with every percentage as an inseparable part of the claim.

5 Limitations

A disclosed window makes a figure interpretable; it does not make the figure durable. Stating the interval is the minimum that lets a reader judge what was measured, but a well-specified window from three months ago still describes a system that has since moved on. This note also treats the window as a single contiguous span; measurements assembled from non-contiguous dates raise further questions of interval statement and pooling that are out of scope here.

Every figure this publication reports therefore closes with the same disclosure, stated verbatim: "These results describe the engines as sampled during [window]; engines change without notice, and results should be assumed perishable."

REFERENCES

1. Schulte, Bleeker, and Kaufmann. *Don't Measure Once: Measuring Visibility in AI Search (GEO)* (2026). https://arxiv.org/abs/2604.07585 Accessed 2026-07-09. [archived]

HOW TO CITE

Barkhausen AI (2026). Why time windows matter. https://barkhausen.ai/notes/why-time-windows-matter/