KENSO RESEARCH · METHODOLOGY

Are B2B homepages all starting to sound the same?

The data, the methods, every number and every limitation behind the study.

Cristian Frunze
Founder of Kenso

ABOUT THE DATA

This study uses Kenso internal data: the company, people and web records Kenso uses for sales and outreach, reused by Cristian Frunze for this study.

Aggregates only. Any group smaller than 50 people is suppressed, and no individual is named.

Figures 1 to 3 are the study page's charts, drawn from the same numbers. The methodology after them keeps the full tables.

SUMMARY

Are B2B homepages all starting to sound the same?

Yes, slowly. 57.4% of recent B2B software homepages read more like their industry's typical copy than the median homepage of a decade ago. Sounding different was not linked to faster growth.

57.4%

of 2024-2026 homepages sit closer to their industry's typical copy than the median 2014-2016 homepage did. With no change it would be 50%.

+0.012

rise in how alike hero copy is within an industry, from 0.541 in 2014-2016 to 0.553 in 2024-2026 (95% interval 0.009 to 0.015).

8.62%

of hero copy used "platform" in 2026, up from 2.02% in 2014. "AI-powered" and close variants went from 0.35% in 2022 to 2.41%.

-0.82

percentage points: the gap in median two-year headcount growth between the most distinctive and matched typical companies (95% interval -2.27 to 3.05).

Data and method

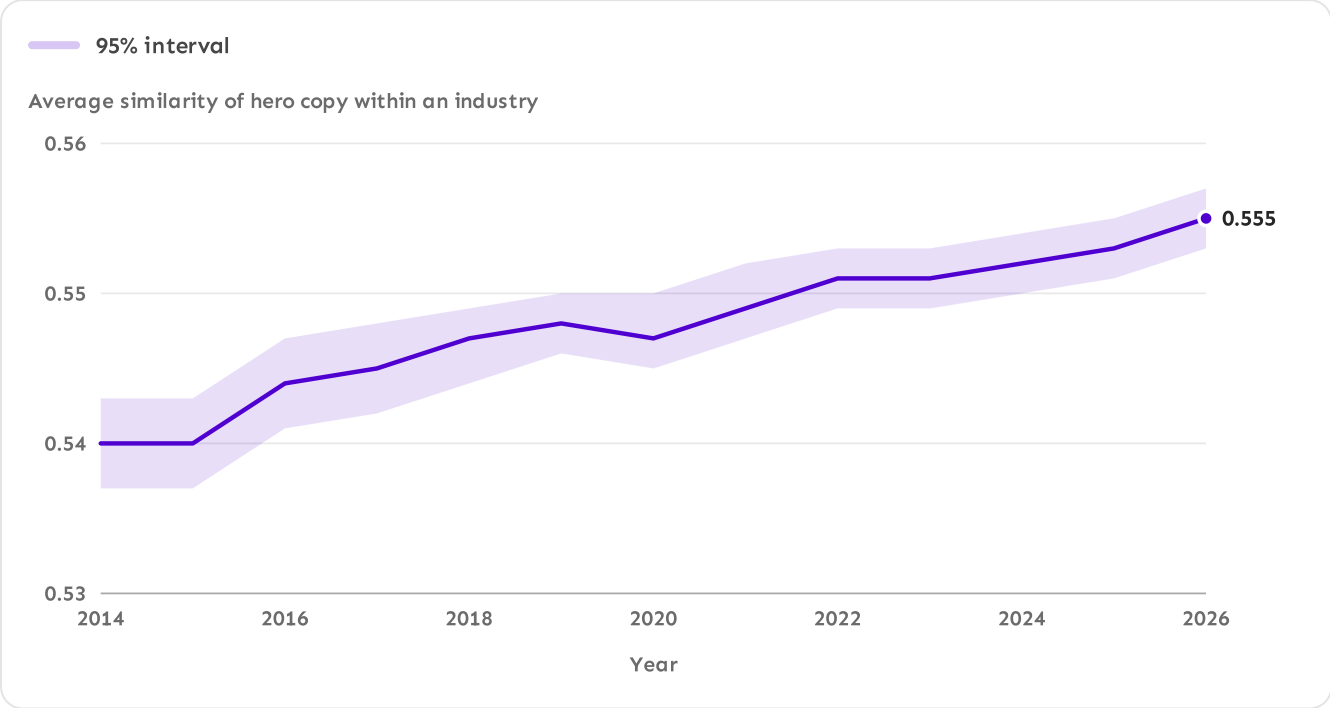
Data	Kenso internal data: Kenso's archive of company websites, one homepage snapshot per company per year, and the company records and work histories Kenso uses for sales and outreach, reused by Cristian Frunze for this study.
Sample	42,045 homepage snapshots from 7,633 B2B software companies, out of a random sample of 10,000 with 11 to 5,000 employees in the US, the UK, Canada, Australia, Ireland and New Zealand.
Date range	One homepage snapshot per company per year, 2014 to 2026. Quarterly headcount from 2015 to March 2026.
Method	Pre-registered. Hero copy compared by meaning (text embeddings) within industry group and year, with same-company and within-company checks. Growth compared by matching distinctive and typical companies. Headline numbers recomputed by a separate analysis.
Excluded	• Parked pages, pages with no usable headline and pages not in English • Template text, and copy shared word for word by three or more companies in a year • Any group, year or cell with fewer than 50 companies
Privacy	Aggregates only. Any group smaller than 50 people is suppressed, and no individual is named.

FIGURES

The study's charts

FIGURE 1

Hero copy has grown more alike almost every year

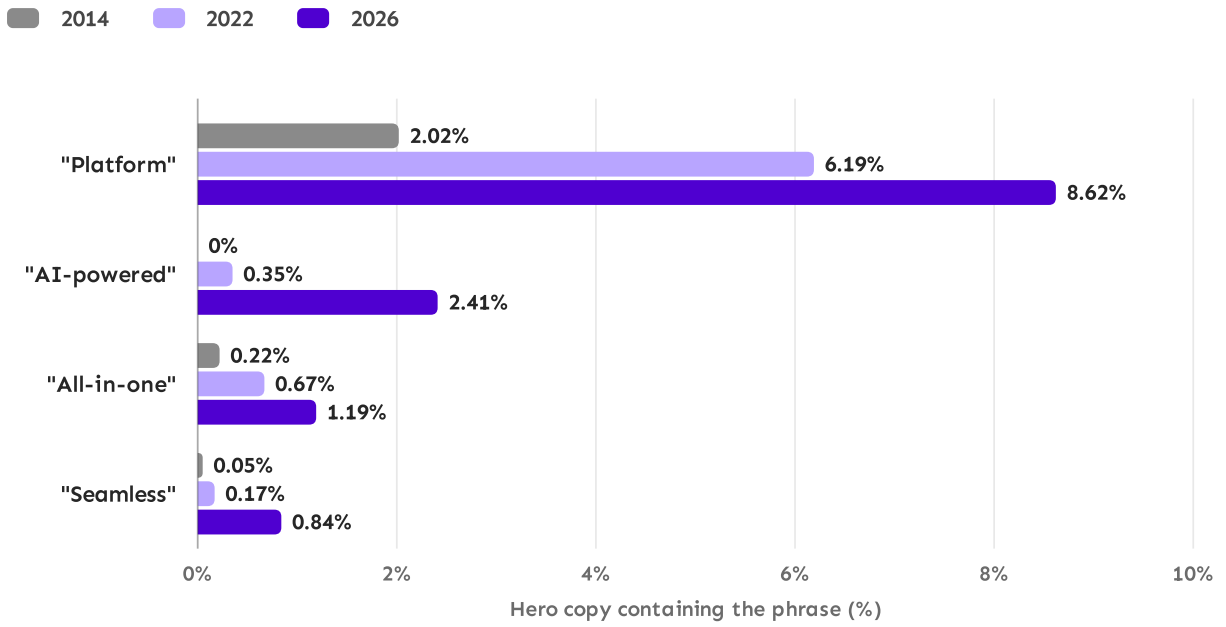


Average similarity in meaning between the hero copy (headline and subheadline) of every pair of B2B software companies in the same industry group, by year. 1 would mean identical meaning, and only the change over time matters. The band is the 95% interval.

Source: Kenso internal data (Kenso's archive of company websites and the company and career records Kenso uses for sales and outreach), analysed by Cristian Frunze.

FIGURE 2

Stock phrases are several times more common than in 2014

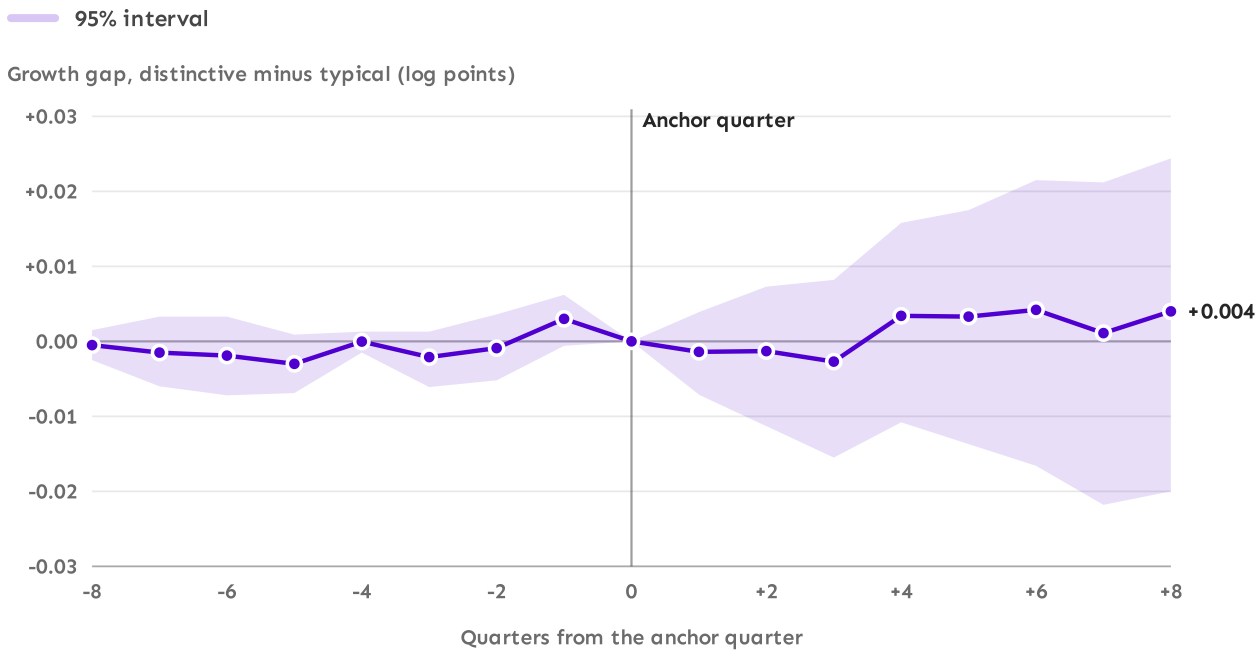


Share of hero copy that contains each phrase in 2014, 2022 and 2026. "AI-powered" also counts AI-driven, AI-native and AI-first. The phrase list was fixed before the data was read.

Source: Kenso internal data (Kenso's archive of company websites and the company and career records Kenso uses for sales and outreach), analysed by Cristian Frunze.

FIGURE 3

Distinctive and typical companies grew alike, before and after



Gap in headcount growth between companies with the most distinctive hero copy and matched companies with typical copy, in log points (0.01 is about 1%). Quarter 0 is the last quarter of the year before the homepage snapshot. 0 means no difference; the bars are 95% intervals.

Source: Kenso internal data (Kenso's archive of company websites and the company and career records Kenso uses for sales and outreach), analysed by Cristian Frunze.

FULL METHODOLOGY

The data, the methods and every number

Question

Has B2B software homepage copy become more alike over the last decade, and do companies with more distinctive copy grow faster?

Data

- **Companies.** 10,000 B2B software companies drawn at random (seed 58) from 51,024 company records in Kenso internal data, Kenso's dataset of public professional profiles, work histories and company records. Records qualify if their industry is a software-product industry (software development; technology, information and internet; computer and network security; data infrastructure and analytics; IT system custom software development; desktop, embedded or mobile software products), their HQ is in the US, UK, Canada, Australia, Ireland or New Zealand, they list a domain not shared with another record, they were founded 1950-2018, and they have 11-5,000 employees today. The panel is shared with other Kenso Research studies (built for the Kenso Research study *Half of B2B software homepages now say AI. Did it pay off?*).
- **Homepages.** One web archive capture of each company's homepage per year, from the crawl nearest mid-year (2014 to 2025), and the September 2026 crawl for 2026. This study uses that one primary crawl per year only.
- **Headcount.** Quarterly headcount per company from 2015 Q1 to 2026 Q1, rebuilt from the start and end dates of roles in Kenso internal data (the shared headcount history, built for the Kenso Research study *Did AI kill the SDR role?*).
- **ken.so.** One fetch of Kenso's own homepage on 2026-10-17, to place Kenso's copy in the distribution. No Kenso campaign data is used.

Sample size

The panel looked up every company in every primary crawl. Kenso's web archive held a usable capture of the homepage for 2,691 to 4,868 companies a year (53,531 in total). After the filters below, 42,045 company-years from 7,633 companies remain:

Year	Fetched	HTTP 200 HTML	Not parked	Has headline	English	Not template	Not shared copy
2014	2780	2780	2773	1910	1859	1844	1833
2015	2775	2775	2767	2019	1967	1958	1943
2016	2691	2691	2680	2007	1939	1932	1925
2017	3786	3786	3782	3009	2925	2904	2898
2018	4820	4820	4815	3789	3673	3653	3650
2019	4598	4598	4592	3761	3656	3630	3621
2020	4597	4587	4569	3753	3663	3543	3536
2021	4868	4847	4826	4095	3994	3927	3920
2022	4867	4842	4821	4163	4066	4008	4005
2023	4709	4674	4646	4066	3983	3892	3892
2024	4832	4801	4774	4153	4069	3980	3974

Year	Fetched	HTTP 200 HTML	Not parked	Has headline	English	Not template	Not shared copy
2025	4493	4469	4460	3742	3664	3647	3647
2026	3715	3696	3685	3282	3216	3201	3201

The step that drops the most is "has headline": 31% of 2014 captures and 11% of 2026 captures have no usable heading (older sites used headings less, and some sites render text with scripts). The English, template and shared-copy filters together drop 2.5% to 5.8% of captures with a headline, depending on the year.

Hero copy

The hero copy is the homepage headline plus the subheadline under it. The headline is the first H1 heading in the page body, outside navigation, menus, cookie banners and the footer, with 2 to 25 words; when there is none, the first H2 under the same rules. Headings that are section or UI labels ("Featured resources", "Contact us", "Book a call"), have fewer than two words of letters, or are just the company's name (a logo set as a heading) are skipped. The subheadline is the first paragraph or heading after the headline, in page order, with 5 to 60 words, skipping repeats of the headline and UI text. The company's own name and domain are replaced by a fixed token, so names do not make companies look different.

Filters

A capture is kept when it is an HTTP 200 HTML page, not a parked domain, has a usable headline, is in English (a language identifier on the hero copy with title and description, and on the page text, both say English), does not match a list of template and filler copy (site-builder sample text, "coming soon", bot checks and similar), and its hero copy is not shared word for word by three or more companies in the same year. The table under Sample size gives the count left after each filter, by year.

Measures

- **Sameness.** Each hero copy is turned into a 384-number vector by a small open embedding model (BAAI/bge-small-en-v1.5, fixed version, run on Kenso's servers). The similarity of two companies is the cosine of their vectors. For each industry group (software development; technology, information and internet; security, data and other software) and year, sameness is the average similarity over every pair of companies in that group and year. The yearly figure is the company-weighted average over the three groups. Embedding similarities sit high for any pair of texts, so only differences between years mean anything, not the level.
- **Robustness.** The same measure on companies captured in every year of both comparison windows (balanced panel), a within-company trend (each company's distance from its group's average copy, regressed on year with a separate intercept per company), and a word-based version (TF-IDF vectors of the hero copy with one vocabulary across all years).
- **Distinctiveness.** For each company and year: one minus the cosine between its vector and the average vector of the other companies in its group and year.
- **Stock phrases.** Share of hero copy containing each phrase in a list fixed before the data were read, by year, with 95% Wilson intervals; and how concentrated the wording is (share of word pairs that are among the 50 most common, on fixed-size random samples of 500 per year).
- **Growth.** Following the AI homepages study: for each capture year 2018-2024, companies in the most distinctive quarter of their group and year are compared with up to three companies with typical copy (bottom half), matched on industry group, founding period, headcount band and both prior 4-quarter growth rates (within 0.10 log points).

Growth is measured from Q4 of the year before the capture. The event-time chart follows both groups from 8 quarters before to 8 after; a joint test checks that they grew alike before. A regression of 8-quarter growth on distinctiveness, with matching-cell-by-year intercepts and both prior growth rates, is a check. Intervals come from resampling companies 1,000 times.

- **Suppression.** Any group, year, cohort or cell with fewer than 50 companies is not reported.

Results

Sameness

Version	2014-2016	2024-2026	Change	95% interval
All valid captures (7,633 companies)	0.541	0.553	+0.012	0.009 to 0.015
Balanced panel (310 companies captured in all six window years)	0.544	0.557	+0.013	0.006 to 0.020
Headline only	0.545	0.561	+0.015	0.012 to 0.018
H1 headlines only			+0.012	0.009 to 0.016
Word overlap (TF-IDF)	0.00552	0.00577	+0.00025	-0.00015 to 0.00064

- The yearly figure rose in 10 of 12 year-on-year steps, from 0.540 (2014) to 0.555 (2026). The rise was not steeper after 2022: about 0.0014 a year from 2014 to 2022 and 0.0010 a year from 2022 to 2026.
- By industry group, the change was +0.014 for software development, +0.013 for security, data and other software, and +0.004 for technology, information and internet.
- Within companies: each company's distance from its industry-year average falls by 0.00055 a year (95% interval 0.00037 to 0.00070), over 6,722 companies with two or more years.
- Translation: 57.4% of 2024-2026 company-years are closer to their industry-year average than the median 2014-2016 company-year (median distance 0.2592 then, 0.2504 now). With no change it would be 50%.
- Word overlap barely moved and its interval includes zero. The share of word pairs that are among a year's 50 most common stays near 5% (5.18% in 2014, 4.92% in 2026).
- Hero length grew slightly: median 10 words in 2014-2020, 12 in 2026; the subheadline is found on 57% of valid 2014 captures and 67% of 2026 captures. The headline-only and H1-only versions show the rise does not come from these format changes.

Stock phrases

Phrase	2014	2022	2026
platform	2.02%	6.19%	8.62%
AI-powered (and AI-driven, AI-native, AI-first)	0.0%	0.35%	2.41%
all-in-one	0.22%	0.67%	1.19%
seamless	0.05%	0.17%	0.84%
simple / simplify	1.2%	3.02%	2.22%
streamline	0.27%	0.62%	1.12%
unlock	0.11%	0.6%	0.91%
one place	0.11%	0.27%	0.59%
powerful	0.87%	1.35%	1.12%
end-to-end	0.22%	0.7%	0.44%

Phrase	2014	2022	2026
next-generation	0.33%	0.52%	0.44%
supercharge	0.05%	0.3%	0.12%

Shares are of valid hero copy per year (1,833 in 2014, 4,005 in 2022, 3,201 in 2026); 95% Wilson intervals are not shown here.

Distinctiveness and growth

- 2,757 distinctive company-years (top quarter, capture years 2018-2025) were matched to typical companies (2,038 distinct controls). Per cohort: 2018 283, 2019 276, 2020 275, 2021 370, 2022 379, 2023 375, 2024 418, 2025 381.
- Pre-trend: joint test p = 0.3295, largest pre-period gap 0.003 log points. Passes.
- 8 quarters after the anchor (n = 2,376): median growth 0.0% for distinctive companies and 0.82% for matched typical companies, a gap of -0.82 percentage points (95% interval -2.27 to 3.05). Mean gap 0.004 log points (-0.0181 to 0.0281).
- 4 quarters after (n = 2,757): median gap -0.13 points; mean gap 0.0034 log points (-0.0102 to 0.0171). Most small companies have no headcount change in a year, so medians sit at or near 0.
- Regression check: 8-quarter log growth on the distinctiveness z-score, 14,654 company-years from 3,909 companies: 0.002 log points per standard deviation (95% interval -0.0072 to 0.0112).

ken.so

Kenso's hero copy (fetched 2026-10-17) is at the 92.2nd percentile of distance from the 2026 software-development average, among 2,170 companies: more distinctive than 92.2% of them. It uses none of the pre-registered stock phrases.

Against the pre-registered hypothesis

1. Sameness rises: supported. The change is positive with an interval above zero on all captures and on the balanced panel, and the within-company trend points the same way. The predicted steeper rise from 2023 did not appear.
2. Stock phrases: mostly supported. "Platform" rose steadily, "AI-powered" rose sharply after 2022 and "all-in-one" rose slowly. "Seamless" did not stay flat; it rose from 0.05% to 0.84%.
3. Distinctive copy and growth: not supported. The pre-trend check passes and the matched gap is close to zero with an interval that includes zero, which the plan named as a result that would prove the hypothesis wrong.

Verification

An independent re-analysis, written from the study plan alone (hypothesis, plan and logged decisions) and the same data extracts, without seeing our code, recomputed the headline numbers.

Number	Study	Re-analysis	Match
Valid company-years	42,045	42,045	yes
Sameness 2014-2016 / 2024-2026 / change	0.5413 / 0.5534 / +0.0121	0.5413 / 0.5534 / +0.0121	yes
Balanced panel change (companies)	+0.0126 (310)	+0.0126 (310)	yes
Headline-only change	+0.0154	+0.0154	yes
"platform" 2014 / 2026	2.02% / 8.62%	2.02% / 8.62%	yes

Number	Study	Re-analysis	Match
"AI-powered" group 2014 / 2026	0.00% / 2.41%	0.00% / 2.41%	yes
"all-in-one" 2014 / 2026	0.22% / 1.19%	0.22% / 1.19%	yes
"seamless" 2014 / 2026	0.05% / 0.84%	0.05% / 0.84%	yes
Late homepages more typical than early median	57.4%	57.4%	yes
ken.so percentile	92.2	92.2	yes
Matched distinctive / with 8-quarter outcome	2,757 / 2,376	2,757 / 2,376	yes
8-quarter median gap	-0.82 pp	-0.82 pp	yes
8-quarter mean gap	0.0040 log points	0.0040 log points	yes
Largest pre-period gap	0.0030	0.0030	yes
Pre-trend joint test p (bootstrap)	0.33	0.31	yes, within bootstrap noise; both pass
8-quarter median gap 95% interval (bootstrap)	-2.27 to 3.05	-2.28 to 3.19	yes, within bootstrap noise

The first pass differed on the growth gap. Two causes were found and fixed before this table: the study broke ties in match distance by row order, which a reader cannot reproduce (now by company id), and the brief for the re-analysis averaged controls in percent rather than log points. The median gap is sensitive to both because most small companies have exactly 0% growth; the mean gap and the regression are not.

Alternative explanations

- **A change in which sites are captured.** Kenso's web archive captures more sites in later years. Ruled out as the main driver: the balanced panel of 310 companies gives the same rise, and each company's own copy drifts toward its industry average.
- **A change in page format, not wording.** Later pages more often carry a subheadline and fewer fall back to an H2. Ruled out: headline-only (+0.015) and H1-only (+0.012) give the same rise.
- **Survivor bias.** The panel is today's company records, so early years hold only companies that survived. Survivors may have been more conventional to begin with, which would bias the rise down, not up; it remains a limitation.
- **The embedding model.** A model trained on recent web text may place recent marketing language closer together. Partly addressed: word overlap shows no rise, so the finding is about meaning, as the model measures it. Named as a limitation.
- **No growth link because headcount is coarse.** Headcount is a slow outcome and is rebuilt from profiles; a link with revenue or win rates would not show here. Named as a limitation.

Limitations

- Kenso's web archive does not capture every site every year, and small sites are captured less often. The balanced panel guards against changes in which sites are captured.
- The panel is today's company records, so companies that died before 2026 are missing, which affects early years most.
- Industry labels do not separate B2B from B2C software.
- Headcount is rebuilt from profiles that exist today; older years undercount, and staff outside six countries are missing. Comparisons are between matched companies, not absolute numbers.

- The hero rules are heuristics on HTML; pages that render their text with scripts give no headline and are dropped.
- The growth link is a matched comparison and event study. It shows correlation and timing, not cause.
- Distinctiveness is distance in meaning, not quality. A clear, specific homepage and a confusing one can be equally far from the average.
- The embedding model's similarity values sit high for any pair of texts, so the level means little; only changes over time and differences between groups are read.
- Matching on size and growth leaves many very small companies with no headcount change, so medians are often 0. The mean and the regression give the same answer.
- The balanced panel is small (310 companies), because few small sites were captured in all six window years.
- Two pieces were added after the plan was registered and before any result was seen (the headline-only check), or after the main results (the translation, H1-only check, hero length and per-group changes). Neither changes a headline number.

Sources

- Kenso's web archive, monthly web crawls June 2014 to September 2026 (one per year).
- Kenso internal data: company records and work histories (the shared homepage panel, built for the AI homepages study, and the shared headcount history, built for the SDR study).
- Xiao, S., Liu, Z., Zhang, P. and Muennighoff, N. (2023). C-Pack: Packaged resources to advance general Chinese embedding. arXiv:2309.07597 (the BGE embedding models; model BAAI/bge-small-en-v1.5, fixed revision).
- Growth method: the AI homepages study.