Ranking Factors in the AI Search Era

Study by MonitLabs

Introduction

We started this study because many analyses of AI search visibility felt too surface level. Most begin with a list of domains or brands. This introduces scale bias since larger sites publish more content and naturally appear more often. In this study we start from keywords instead to reduce that bias. We explain this approach in more detail later in the report.

It is also important to understand how to read the data. Do not treat the absolute numbers as universal truths. Every dataset has limits. The real value of this analysis comes from comparisons. We look at what appears more often and what appears less often, and how different factors relate to visibility. Backlinks and traditional SERP rankings give us a useful reference point because SEO professionals already understand these relationships well.

You should also remember that ranking factors rarely affect visibility in a simple linear way. Because of this, Pearson and Spearman correlations alone are often not enough to measure the real strength of a ranking factor. For this reason, we show the data from several angles throughout the report. This helps readers draw their own conclusions. In some places we also share our interpretations and hypotheses based on the patterns we observe.

This report is only the beginning. In future editions we will use the same keyword sets so we can track changes over time. This will help us identify trends in both traditional SEO and AI search.

Study Design and Methodology

This study examines how backlink-related and mention-related signals correlate with visibility in:

- Google organic search results
- Google AI Overviews (AIOs)
- Google AI Mode
- ChatGPT
- Perplexity

Throughout this report, references to “traditional SERPs” and “AIOs” refer specifically to Google Search (United States).

The research is divided into two independent keyword sets: one focused on commercial queries and the other on informational ones. This distinction was introduced to isolate different visibility dynamics between intents.

Dataset Structure

Commercial Dataset

The commercial dataset was designed to evaluate product and service visibility patterns across Google and AI systems.

- 12 industries
- 1,000 keywords per industry
- 12,000 total keywords
- Product-type and service-type queries only
- Commercial intent only
- United States locale
- 41,389 unique domains analyzed
- Data collected in Jan-Feb 2026

Keywords were reviewed by dedicated AI agent to ensure:

- Commercial intent
- No local intent
- Semantic uniqueness

Informational Dataset

The informational dataset was constructed to analyze source citation behavior rather than commercial recommendation patterns.

- 12 industries
- 1,000 keywords per industry
- 12,000 total keywords
- Question-based queries only
- Informational intent only
- United States locale
- 27,204 unique domains analyzed
- Data collected in Jan-Feb 2026

Keywords were reviewed by dedicated AI agent to ensure:

- Informational intent
- No local intent
- Semantic uniqueness

This dataset intentionally maximizes the probability of observing citation behavior in AI-generated responses.

Google SERP Visibility Metrics

For each domain appearing within the datasets, two visibility dimensions were tracked.

Number of SERP occurrences
The frequency with which a domain appears in the top 10 organic Google results across the keyword set.

This captures ranking presence.

SERP Score (Weighted Ranking Strength)
To measure ranking depth — not just presence — a weighted scoring system was applied:

- Position 1 = 10 points
- Position 2 = 9 points
- Position 3 = 8 points
- ...
- Position 10 = 1 point

Points are aggregated across all queries to generate a cumulative SERP Score for each domain.

This distinguishes between domains consistently ranking at the top and those appearing frequently but in lower positions.

AI Visibility Metrics

For AI-generated responses (Google AI Mode, Google AIO, ChatGPT, Perplexity), two visibility dimensions were measured.

Number of citations
The number of times a domain appears as a cited source across all AI responses within the tested dataset.

This isolates citation behavior.

Number of brand mentions
Number of times a given brand was mentioned in LLM responses within the tested dataset.

This isolates recommendation visibility from source citation visibility.

Ranking Factors Analyzed

Across both datasets, correlations were calculated against backlink-related and unlinked web mention-related signals:

Backlink based factors

Total number of backlinks: as the name suggests, this is just a raw number of links pointing to a given website. This includes internal links.

Domain Authority (0–100): a MonitLabs domain-level score derived from backlink data. Similar to other industry metrics like this.

URL Authority (0–100): MonitLabs page-level score derived from backlink data

Referring domains: number of main domains pointing to a given website

No follow referring domains: number of main domains pointing to a given website with at least one no-follow link.

Unlinked mention-based factors

Web mentions count: The total number of times a domain or brand is mentioned online. This shows how much overall attention it receives.

Web mentions score (authority weighted mentions): This metric adjusts mention volume by source quality. It is calculated as the number of web mentions multiplied by a Domain Authority (0-100) factor. Mentions from stronger, more trusted websites increase the score more than mentions from weaker sites. This way, the metric reflects both how often a brand is mentioned and how strong the sources are.

Positive mentions count: The total number of times a domain or brand is mentioned online in a positive context. This reflects how often the brand is discussed favorably across the web.

Net-positive mentions count: The number of positive mentions minus the number of negative mentions. This metric shows the overall sentiment balance, indicating whether a brand is discussed more positively or negatively online.

Negative mentions count: The total number of times a domain or brand is mentioned online in a negative context. This reflects how often the brand is associated with criticism, complaints, or unfavorable discussion.

Correlation Framework

For both commercial and informational datasets, we analyzed how ranking factors correlate with:

- Top 10 organic visibility
- SERP Score
- Google AI Overview citations
- AI system source links
- AI system brand mentions (where applicable)

This structure allows us to assess:

- Whether AI visibility patterns align with traditional link-based authority
- Whether commercial AI recommendations rely on different signals than informational citations
- Whether backlink strength correlates more strongly with ranking position or AI inclusion
- How these patterns vary between intents

Why the Study Starts From Keywords

A common methodological approach in SEO visibility studies is to begin with a predefined list of domains and measure how frequently they appear across SERPs and AI-generated outputs within third party SERP and LLM's answers databases.

That approach presents a major limitation.

A scale bias.

If we start with a list of websites, count how many unlinked mentions or backlinks they have, and then measure how often they are cited or mentioned by LLMs, we are essentially asking each website or brand:

How many races have you won?

But that answer depends heavily on another question:

How many races have you competed in?

In practice, this often comes down to scale. The more content a site publishes, the more chances it has to rank and collect AI citations. At the same time, companies that sell more products or serve more verticals and use cases naturally appear in more contexts where big LLMs can mention them. The more races you enter, whether through content volume or market scope, the more chances you have to win.

We wanted to ask a different question:

Here are the races. Who won them, and what do the winners have in common?

Starting from keywords grouped by industries limits that problem. More on this below.

Study Limitations

No dataset captures the full complexity of search behavior. The following constraints should be considered when interpreting the results.

Scale Bias

Domains with higher SEO metric scores tend to be bigger on average.

When visibility is then measured by raw counting **all** citations, mentions and keywords each **website won**, scale differences can distort the results. These raw visibility numbers may reflect size and scope as much as actual impact of tested ranking factors.

This happens because:

• Larger websites publish more content and target more keywords, which naturally creates more opportunities to rank and to be cited by LLMs.

• Bigger companies offer more products, operate across more verticals, and cover more use cases, increasing the number of situations in which they can be recommended and mentioned.

To limit the impact of the scale bias, this study begins with keywords rather than domains. Visibility emerges from query-level analysis, not from preselected entities.

Why does starting from keywords limit this problem?

Starting from a defined keyword-set limits scale bias because visibility is measured within controlled topical boundaries. The dataset is split into smaller groups by industry and real search intent, so a large brand operating across many verticals only gains visibility within the specific verticals being tested. It does not automatically benefit from its full market

scope. While a large company may still have some statistical advantage, this advantage is much smaller than if total, unrestricted visibilities were compared.

Because the differences in visibility rely less on pure scale, they can be attributed more confidently to the ranking factors being tested, reducing distortion and lowering the overall error in interpretation.

Industry Aggregation and Simpson's Paradox

Simpson's paradox occurs when a pattern visible within individual groups becomes weaker or reverses once all data is combined.

In this study, industries differ significantly in competitiveness. A finance website with a large number of backlinks may still record lower visibility than a hobbies website with fewer backlinks. The difference does not necessarily reflect a weaker link between authority and visibility. It reflects harder competition and more demanding queries in the finance industry.

If industries are not analyzed separately, the results can become blurred. When all data is combined, it may look like any given ranking factor is not very important. In reality, it could be quite important within certain industries, but this effect gets hidden in the overall numbers.

To reduce this distortion, correlations are calculated separately for each industry and then averaged, instead of using one overall correlation strength for all data combined. This keeps the industry context and gives a more accurate picture of how authority works.

Even this approach has limits. Competitiveness can vary within industries. Certain sub-niches may face stronger competition and higher authority thresholds than others. However, it is not practical to keep splitting the

data into smaller and smaller topical groups. At some point, segmentation must stop.

Because of this, some variation remains in the data. **If we sliced the dataset into even tighter, more specific topical segments, the relationship between authority signals and visibility would likely appear stronger.**

The main goal here is not to present absolute correlation values, since those can change depending on how the dataset is prepared, but to compare them. What matters most is distinguishing weaker signals from stronger ones and, in particular, comparing their influence on AI visibility with their influence on something we already understand better: classic Google SERPs.

Dataset Size and Query Selection

Each industry includes 1,000 queries.

This limit reflects the time-intensive process required to ensure true intent alignment. We did not rely on third-party keyword tools to define intent categories. Instead, we manually validated approximately 5 to 10 percent of queries per industry to establish a high-confidence intent baseline.

From there, we deployed a custom-built AI agent to:

- Generate additional query candidates
- Filter out queries that did not strictly match the defined intent
- Remove queries with 0 estimated search volume

This hybrid workflow improved intent precision but constrained scale.

Brand Entity Matching in LLM Responses

Matching brand entities across AI-generated answers presents structural challenges.

LLM responses often refer to brands in inconsistent formats. Variations include:

- Brand-level references versus product or service-level references
- Full legal names versus shortened forms
- Abbreviations versus expanded versions
- Slight spelling differences

Accurately mapping these references to canonical brand entities and corresponding domains required additional processing.

We developed a custom AI agent to standardize entity recognition and match brand mentions across responses. We also conducted manual checks on randomly selected samples and iteratively refined the classification logic.

Despite these controls, some rare misclassifications may remain.

This limitation is more pronounced in industries with complex brand architectures. For this reason, **mention-based factors analysis was based solely on the Software & SaaS industry**, where entity structures are typically clearer and easier to normalize.

SERP Feature Complexity

Traditional Google SERPs include multiple feature types beyond standard organic listings, such as:

- Featured snippets
- People Also Ask boxes
- Knowledge panels
- Product carousels

This study focuses on top 10 organic positions and AI Overview inclusion. It does not attempt to model interactions between all SERP features and ranking signals.

Interpretation

The data in this report measures correlation, not causation.

The findings describe alignment between authority signals and visibility outcomes. They do not establish direct causal mechanisms within AI systems or Google's ranking algorithms.

However, we may occasionally describe the data using cause-and-effect language, but it's important to remember that any cause we mention is only a hypothesis: our best guess about what the data might be showing.

The central question guiding this research is:

Are AI search systems reinforcing the traditional SEO authority model, or evolving toward a different authority logic?

The following sections examine the results across query types, industries, and platforms.

The study provides directional insight into authority dynamics rather than universal, permanent rules.

Strength thresholds

Because scale bias was reduced as described earlier, the observed correlation strengths may appear lower than in some other studies. At the same time, there is a counteracting force: limited topical segmentation. As discussed previously, due to the nature of SEO data and Simpson's Paradox, more granular topic segmentation would likely lead to stronger correlations between ranking factors and visibility metrics.

However, splitting the dataset into very small topical clusters would be operationally too costly, so the analysis was conducted at the industry level. For this reason, the results should not be interpreted in absolute terms. What matters most is how the tested factors compare to one another, not the standalone magnitude of each correlation.

Below is the subjective scale we use to interpret correlation strength when commenting on the data:

- 0 – 0.05: almost no correlation, negligible
- 0.06 – 0.10: very weak
- 0.10 – 0.20: weak
- 0.20 – 0.30: medium
- 0.30 – 0.40: fairly strong

PART I: Backlinks vs Visibility

This section examines how backlink-related signals correlate with visibility across traditional Google results and AI-driven search systems.

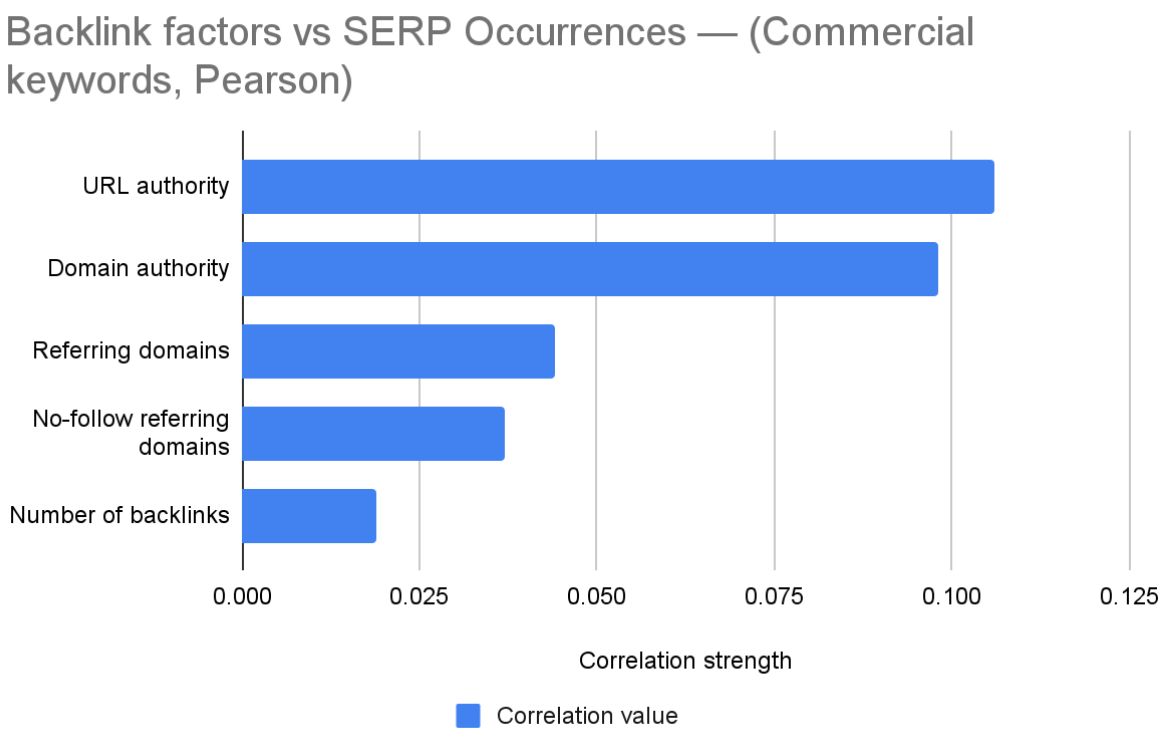
The goal is to measure and compare how strongly link-based authority aligns with rankings, citations, and brand inclusion across different environments.

Backlinks & Traditional SERPs

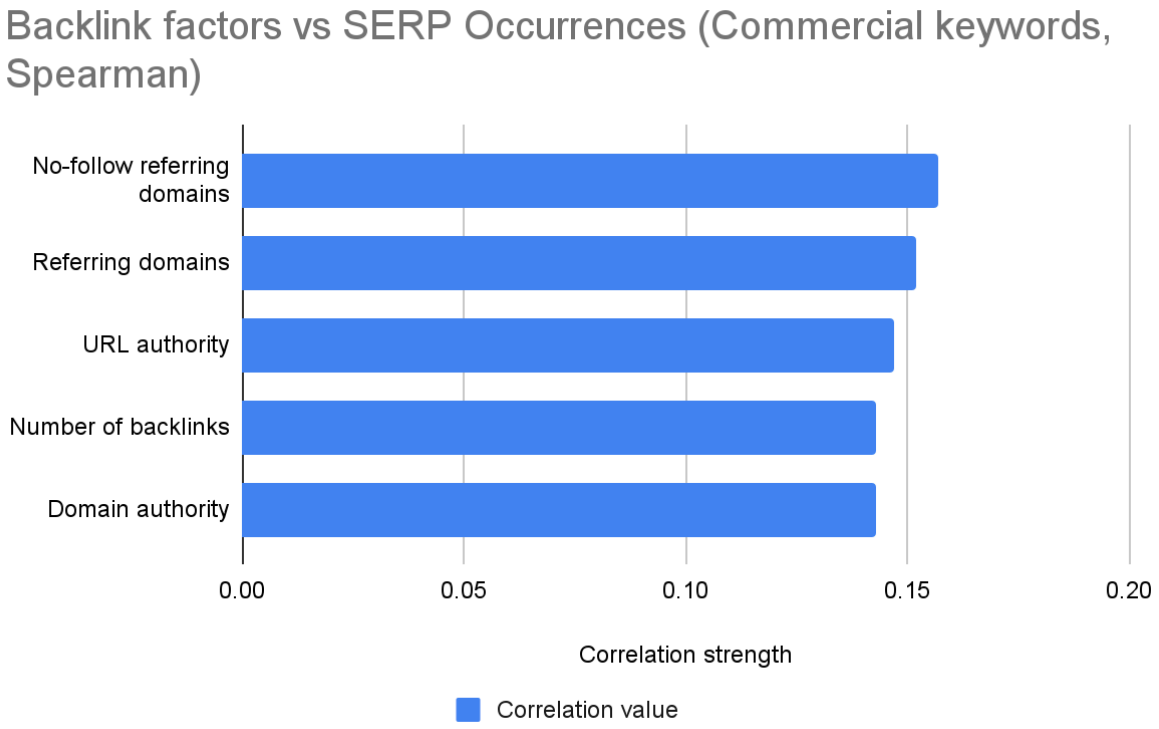
The following charts examine the relationship between backlink-related factors and SERP visibility factors, separately for commercial and informational keywords. The main goal is to give you a clear point of reference. The impact of backlinks and authority signals on classic Google SERPs is well understood, so this serves as a baseline. By separating commercial and informational intent, we can see whether these familiar patterns hold consistently across query types and use them as a benchmark for further comparisons.

Frequency in the Top10

First, let's take a look at correlation between our factors and the frequency of SERP occurrences, despite how close to the first position. Each occurrence in the Top10 counts as one point.

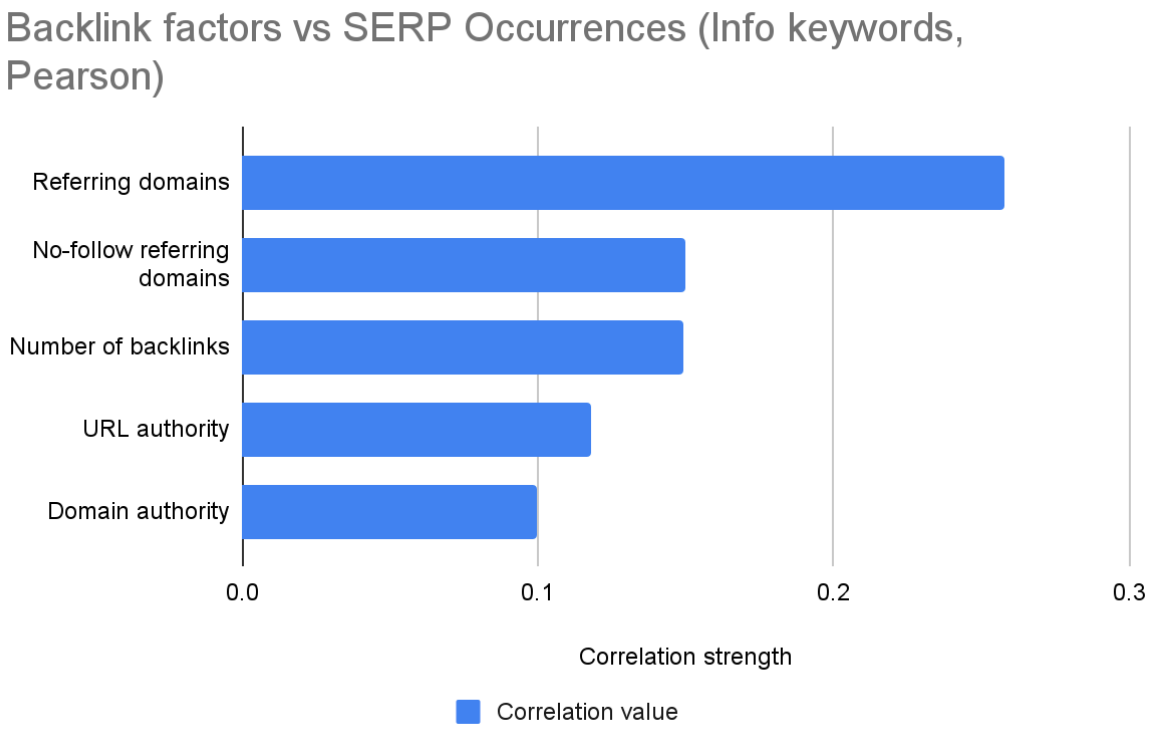


For commercial keywords, Pearson correlations with SERP occurrences are very weak overall. **URL Authority** is the highest at **0.106**, followed closely by **Domain Authority** at **0.098**. **Referring domains 0.044** and **No follow referring domains 0.037** are very weak, while **Number of backlinks** has almost no correlation at **0.019**.



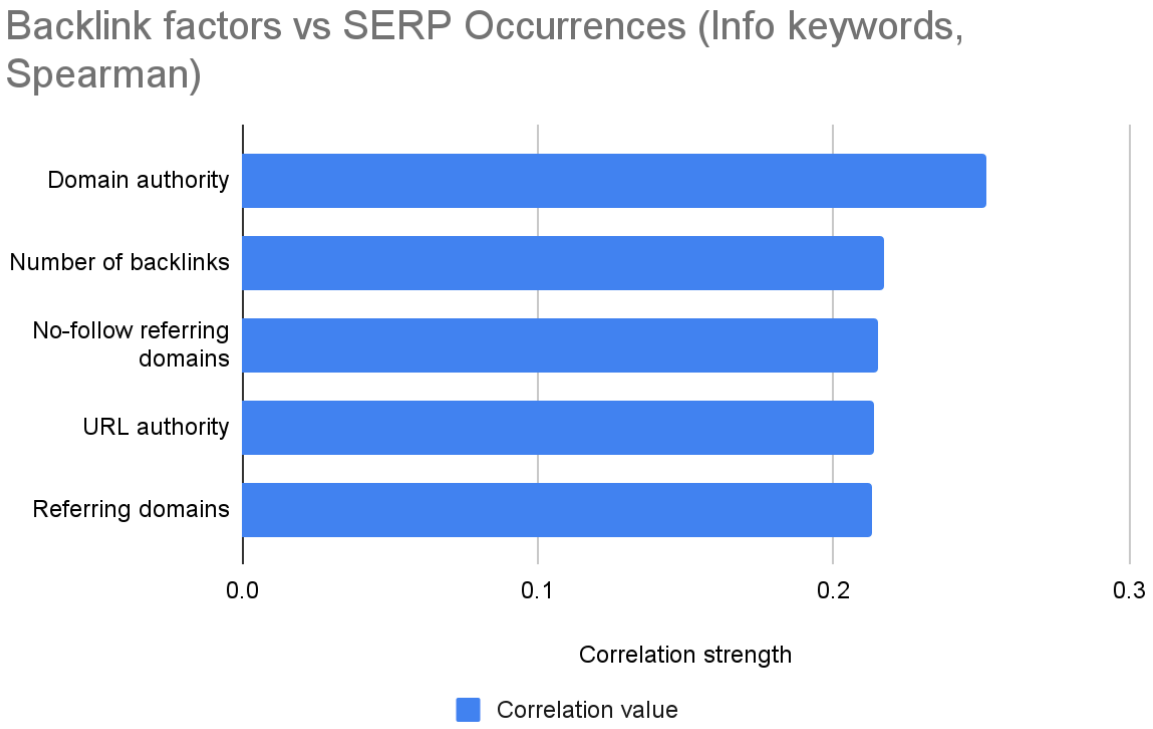
In Spearman, the pattern is stronger and more balanced. **No follow referring domains** leads at **0.157**, followed by **Referring domains 0.152** and **URL Authority 0.147**. **Number of backlinks** and **Domain Authority** both sit at **0.143**. All are weak but relatively close, suggesting a consistent rank based relationship across link factors.

Let’s take a look at these correlations for informational keywords.



For informational keywords, correlations are clearly stronger than for commercial queries.

In Pearson, **Referring domains** leads at **0.258**, a medium correlation and much higher than any value in the commercial set. **No follow referring domains 0.150** and **Number of backlinks 0.149** are weak but still noticeably stronger. **URL Authority 0.118** and **Domain Authority 0.100** trail behind.



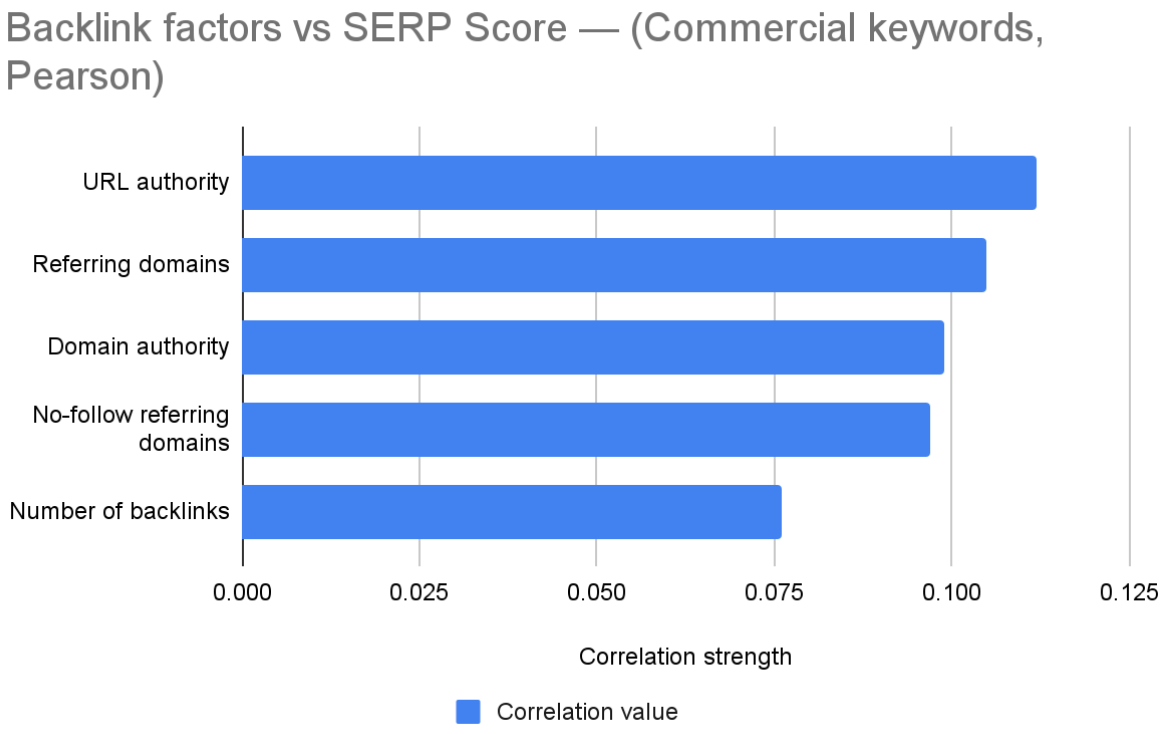
In Spearman, all factors cluster tightly in the medium range, between **0.213** and **0.252**. **Domain Authority** is highest at **0.252**, followed by **Number of backlinks** at **0.217**.

Compared to commercial keywords, informational queries show both stronger and more evenly distributed relationships between backlink factors and SERP occurrences.

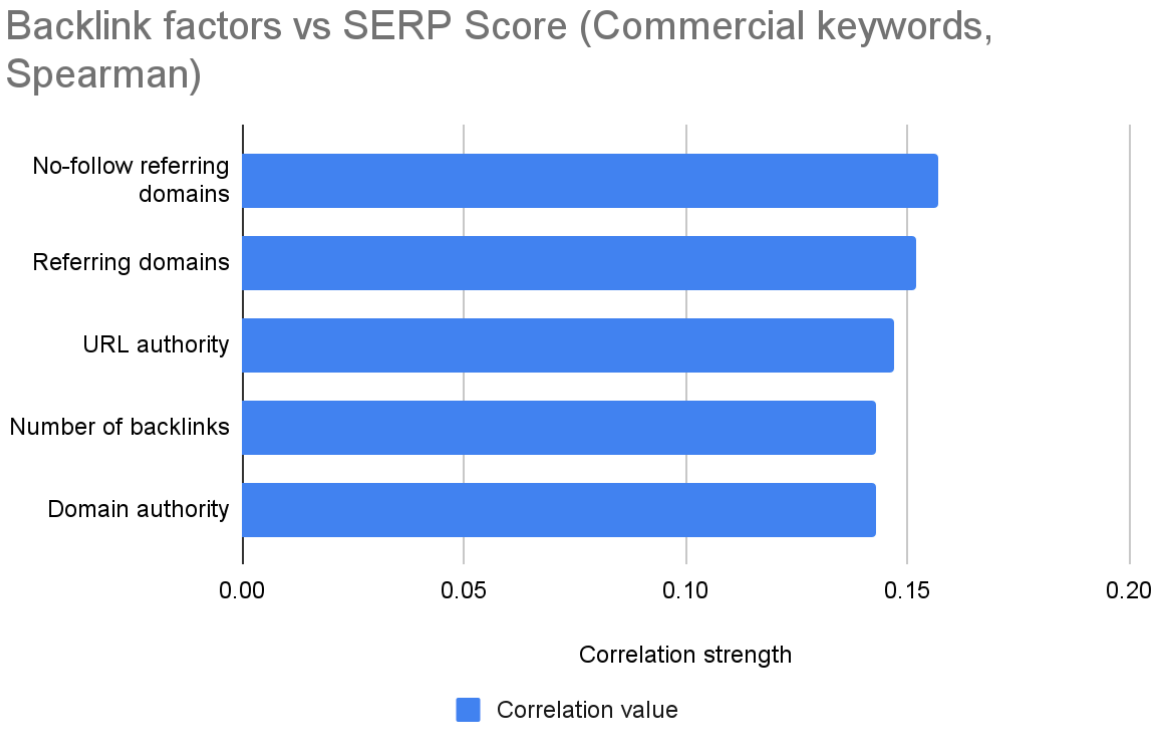
Visibility in Top10: SERP Score

In this section, correlations with SERP Occurrences are compared to correlations with SERP Score. This can help test a common hypothesis: links help you get into the top 10, but once you get there, other factors may matter more.

SERP Occurrences show whether backlink factors are related to appearing in the top 10 at all. SERP Score, on the other hand, captures ranking depth within the top 10 and shows whether those same factors continue to matter once a domain is already there. Comparing both helps validate that assumption.

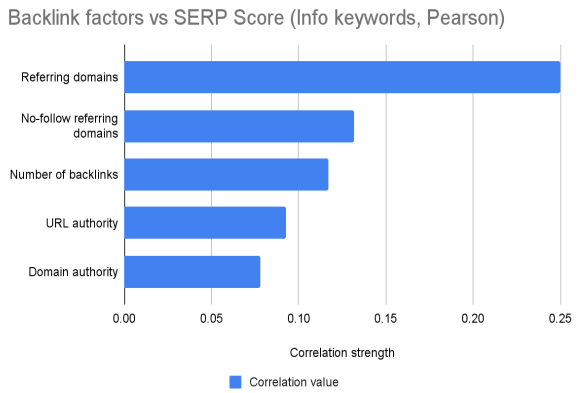


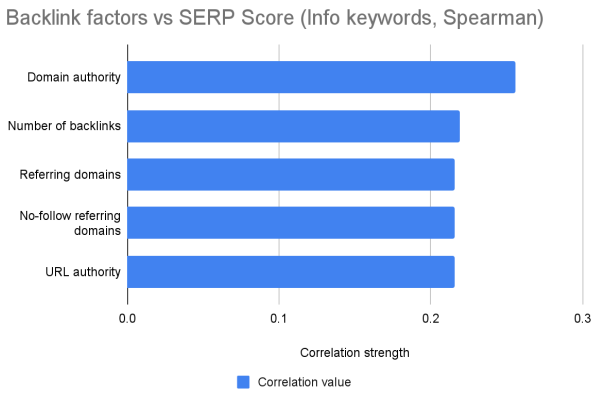
In Pearson, **URL Authority** wins at **0.112** and Referring domains follows at **0.105**, both weak. **Domain Authority** and **No-follow referring domains** drop into very weak territory, while **Number of backlinks** is the weakest at **0.076**.



Across both charts, correlations are consistently weak. In Spearman, **No follow referring domains** leads at **0.157**, followed closely by **Referring domains** at **0.152** and **URL Authority** at **0.147**. **Backlinks** and **Domain Authority** tie at **0.143**. All fall into the weak range.

Overall, link quantity and authority show some relationship with SERP score, but none reach medium strength. Spearman values are consistently higher than Pearson which likely means the relationship between backlink factors and SERP score is somewhat monotonic but not very linear. In simple terms, as backlink metrics increase, rankings tend to improve, but not always in a proportional way.





For informational keywords, correlations are clearly stronger than in the commercial set.

In Spearman, **Domain Authority** leads at **0.256**, a medium correlation. **Number of backlinks** follows at **0.219**, while **Referring domains**, **No-follow referring domains**, and **URL Authority** cluster at **0.216**, all medium. The pattern is consistent and tightly grouped, suggesting a solid monotonic relationship.

In Pearson, **Referring domains** wins at **0.250**, also medium. **No-follow referring domains** drops to weak at **0.132**, and the rest decline further, with **Domain Authority** the weakest at **0.078**.

Overall, authority and referring domains diversity matter more here. Domain Authority being the strongest for Spearman and weakest for Pearson suggest that there is a solid correlation but mostly non-linear.

Backlinks relevance before vs after entering Top10

For commercial keywords, the patterns for SERP Occurrences and SERP Score are very similar.

In Pearson, values stay very weak in both cases. For SERP Occurrences, **URL Authority** led at **0.106**, while for SERP Score it is **0.112**. The ordering of factors barely changes, and differences are minimal. In Spearman, the

numbers are almost identical across both metrics, with No follow referring domains at **0.157** and Referring domains at **0.152.** This suggests backlink factors relate to entering the top 10 and to ranking strength within it at a similar level.

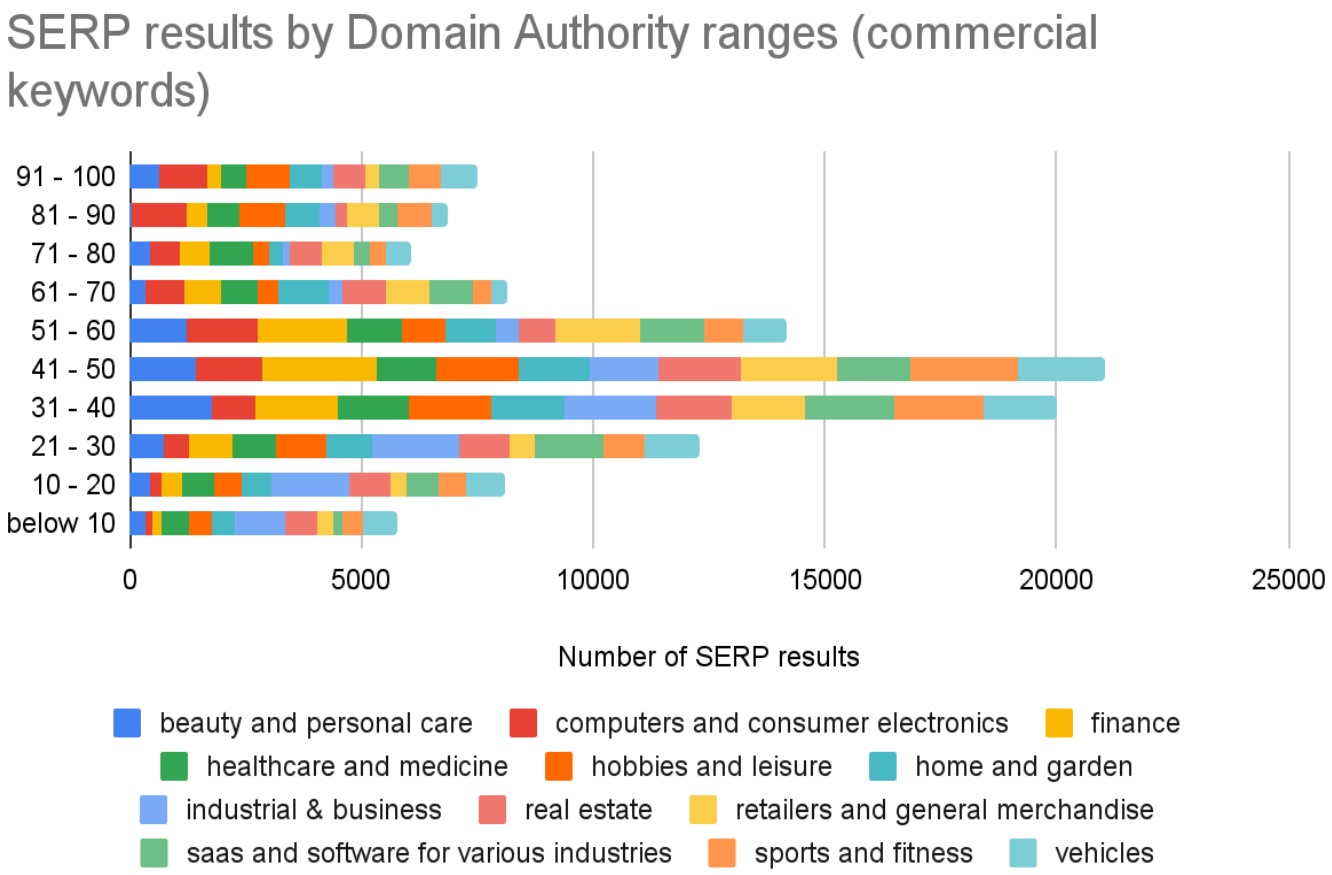
For informational keywords, the same consistency appears. In Pearson, **Referring domains** correlation strength is significant for both SERP Occurrences **0.258** and SERP Score **0.25**, both within medium range. In Spearman, all factors cluster between **0.213** and **0.256** for both metrics. **Domain Authority** peaks at **0.256** for **SERP Score** and **0.252** for **SERP Occurrences**.

Overall, there is no clear drop when moving from Occurrences to Score. Backlink factors seem to matter not only for getting into the top 10, but also for position strength within it.

Frequency in the Top10 by Domain Authority ranges

Since most of the observed relationships appear weak to moderate, let's examine the data from a different perspective.

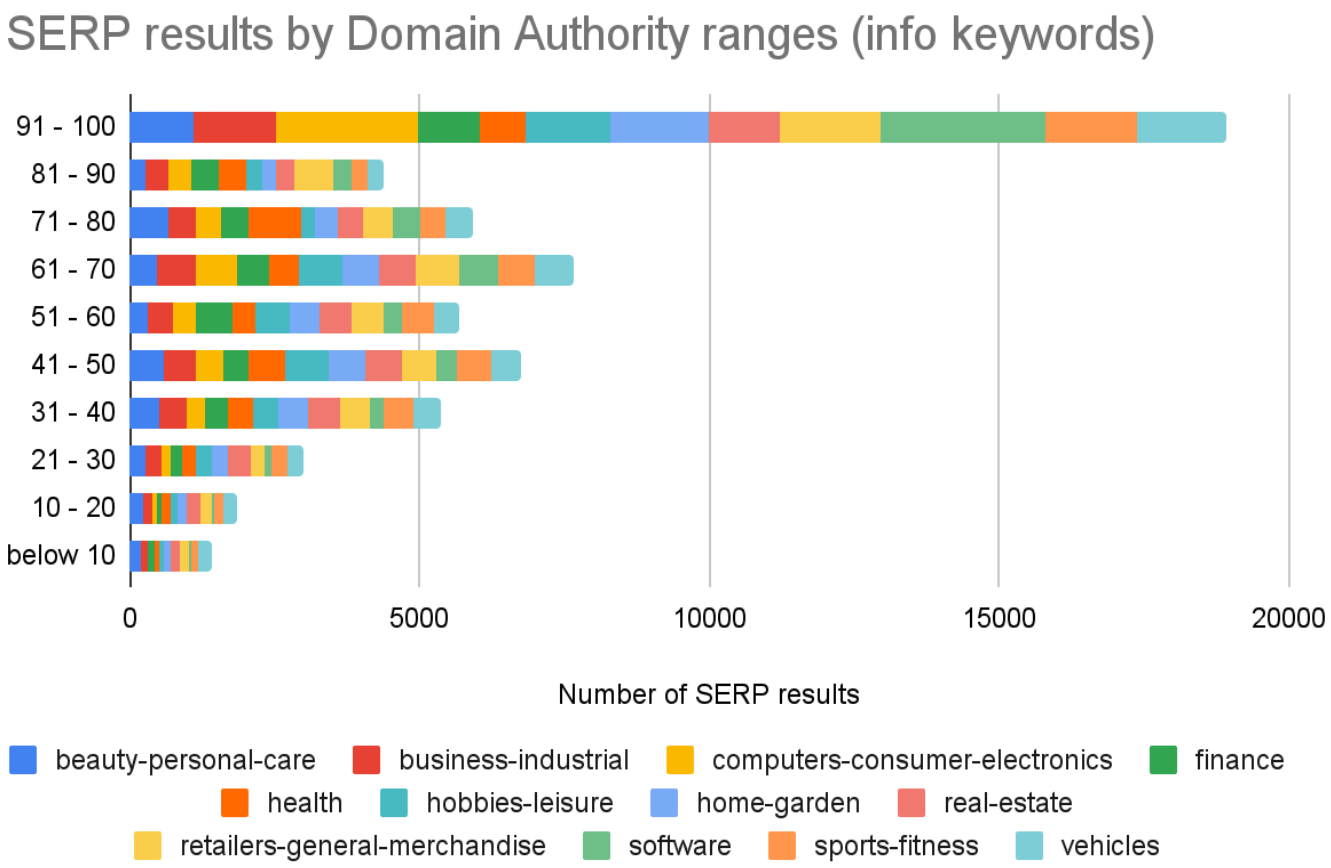
Below are charts showing the total number of Top 10 occurrences for domains grouped by their respective **Domain Authority** ranges.



A clear pattern emerges for commercial keywords. As Domain Authority increases up to the mid-level range (40–50), the total number of observed Top 10 rankings rises accordingly. Beyond this range, however, the number begins to decline.

We interpret this drop as a consequence of the decreasing number of domains in higher authority brackets in the internet in general. The 40–50 range appears to be a tipping point for commercial keywords.

This pattern also helps explain why both Pearson and Spearman correlations are relatively weak. The relationship between Domain Authority and SERP visibility is clearly non-linear, which reduces both the strength of linear (Pearson) and order-based (Spearman) correlation metrics.



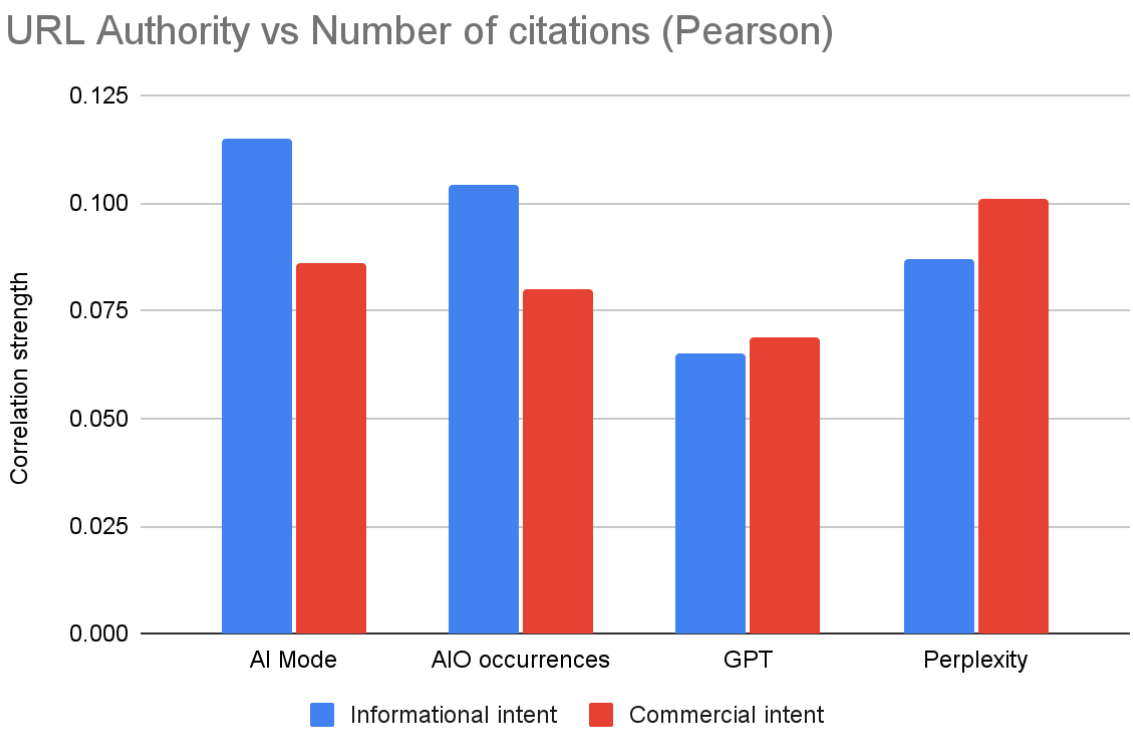
For informational keywords, we observe a similar pattern, although the tipping point appears to occur at a slightly higher Domain Authority range.

At the same time, there is a strong skew toward the most authoritative websites. This highest authority bracket is dominated by platforms such as Reddit, Wikipedia, Facebook, and other large-scale publishers.

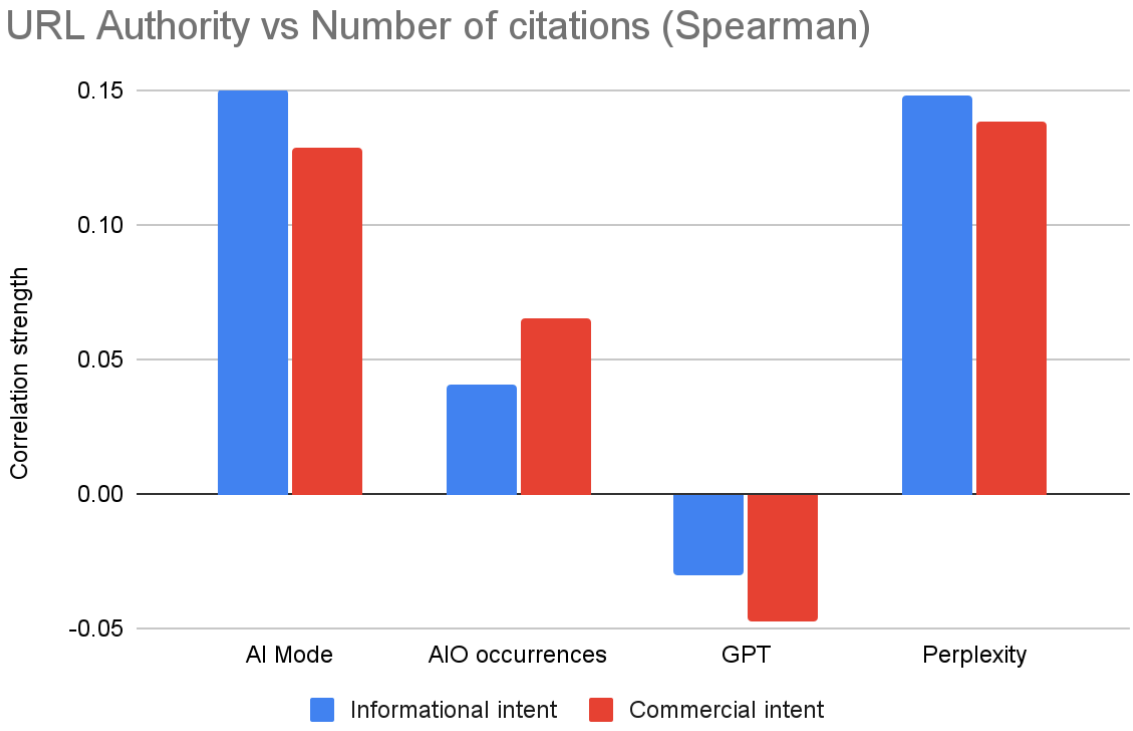
Key Takeaways

- Building links still make a difference but appears to be particularly beneficial for lower-authority websites, with diminishing returns as overall Domain Authority increases.
- Commercial keywords appear to be less heavily skewed toward the highest authority domains than informational keywords.
- The relationship between ranking factors and search visibility is often highly non-linear, which requires more nuanced analysis beyond simple correlation metrics. Relying solely on mathematical correlation coefficients such as Pearson or Spearman can be misleading when assessing the true impact of individual ranking factors.

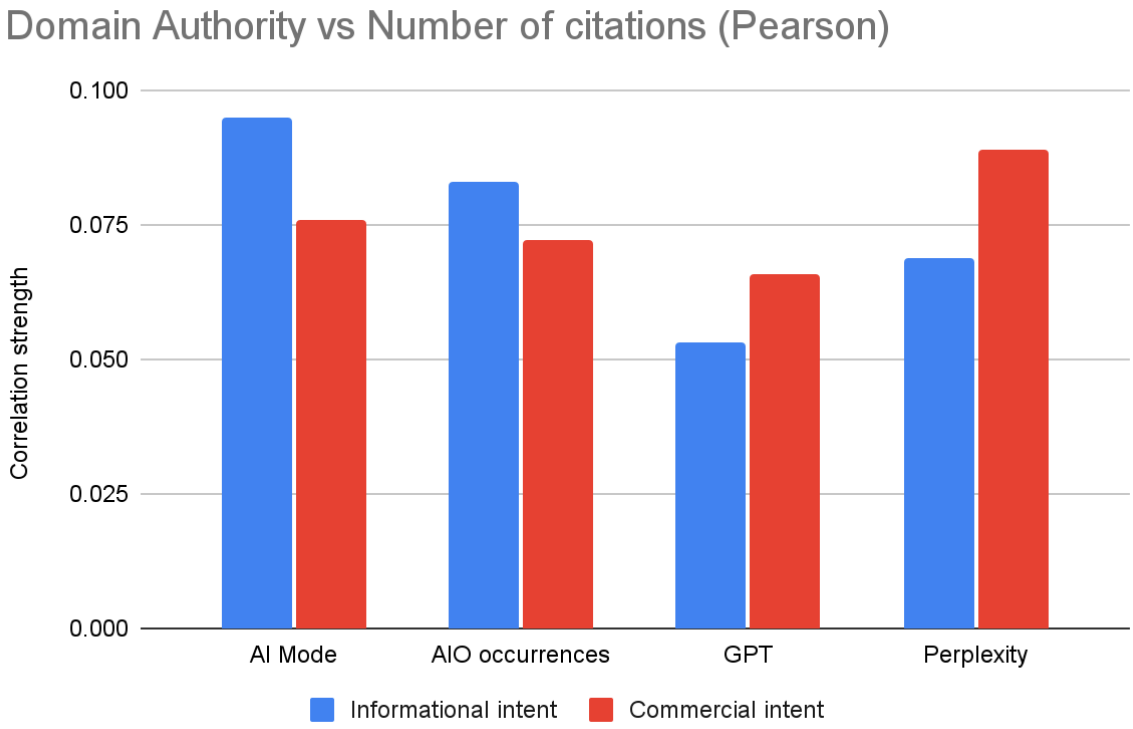
Backlinks & AI Search



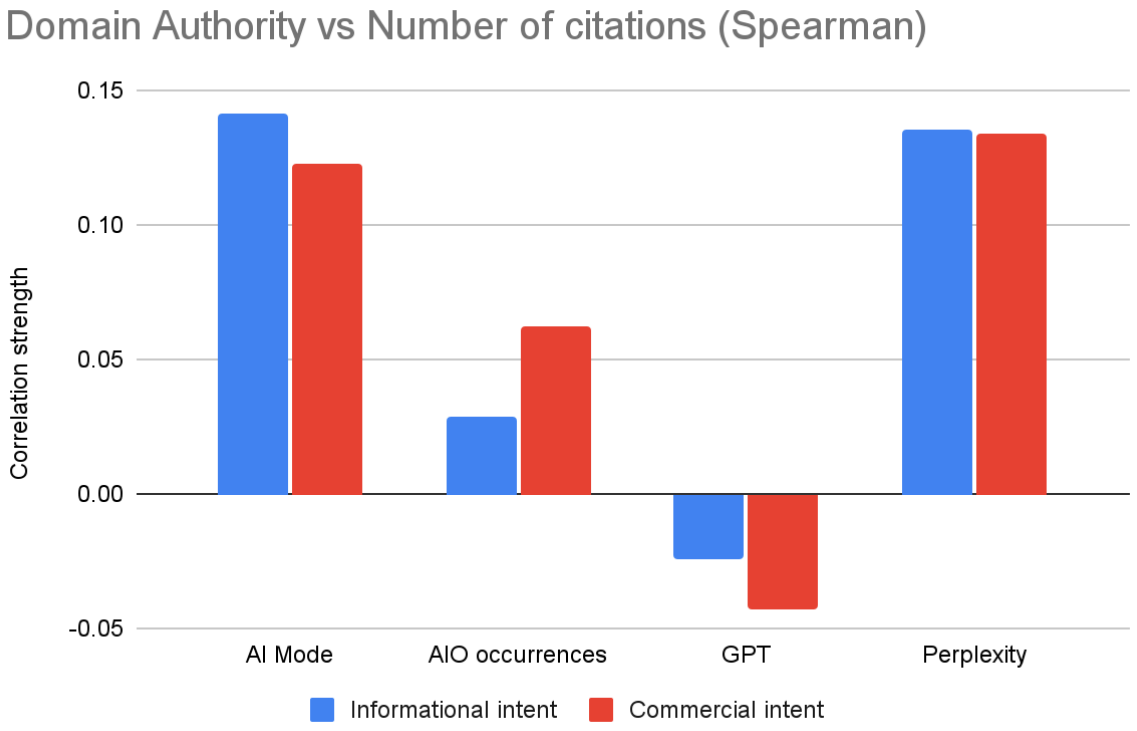
In Pearson, all correlations are weak or very weak. For informational intent, **AI Mode** leads at **0.115**, followed by **AIO occurrences 0.104**. **GPT** is the weakest at **0.065**. For commercial intent, **Perplexity** wins at **0.101**, while **AIO occurrences** is the lowest at **0.08**. Differences are small overall.



In Spearman, patterns shift slightly. For informational queries, **AI Mode** with **0.15** and **Perplexity** with **0.148** are the strongest, but still pretty weak. Surprisingly **GPT** is even negative for both intents, **minus 0.03** informational keywords, and **minus 0.047** for commercial. The strength of the correlation is barely visible so it's not valid to draw any conclusions. **AIO** also has almost no correlation for informational keywords at **0.041**.

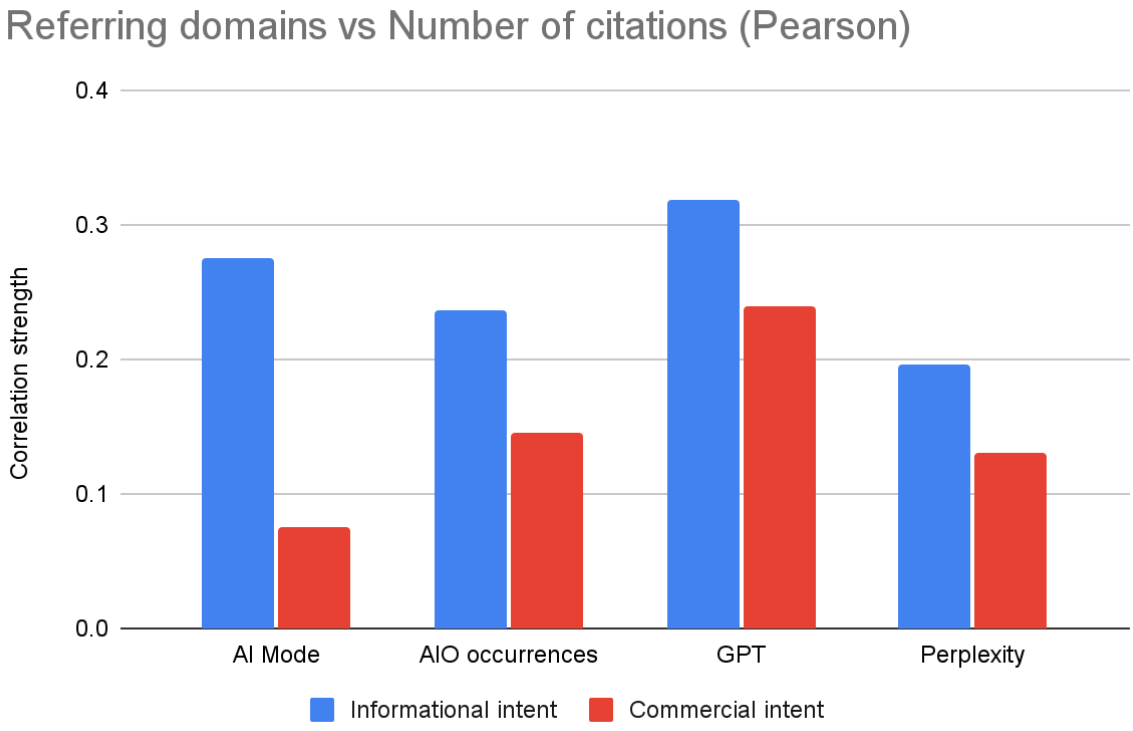


In Pearson, all correlations are very weak. Informational intent peaks at **AI Mode** on **0.095**, while **GPT** is lowest at **0.053**. For commercial intent, Perplexity leads at **0.089** and **GPT** slightly improves to **0.066**. Overall, effects are small and tightly clustered.

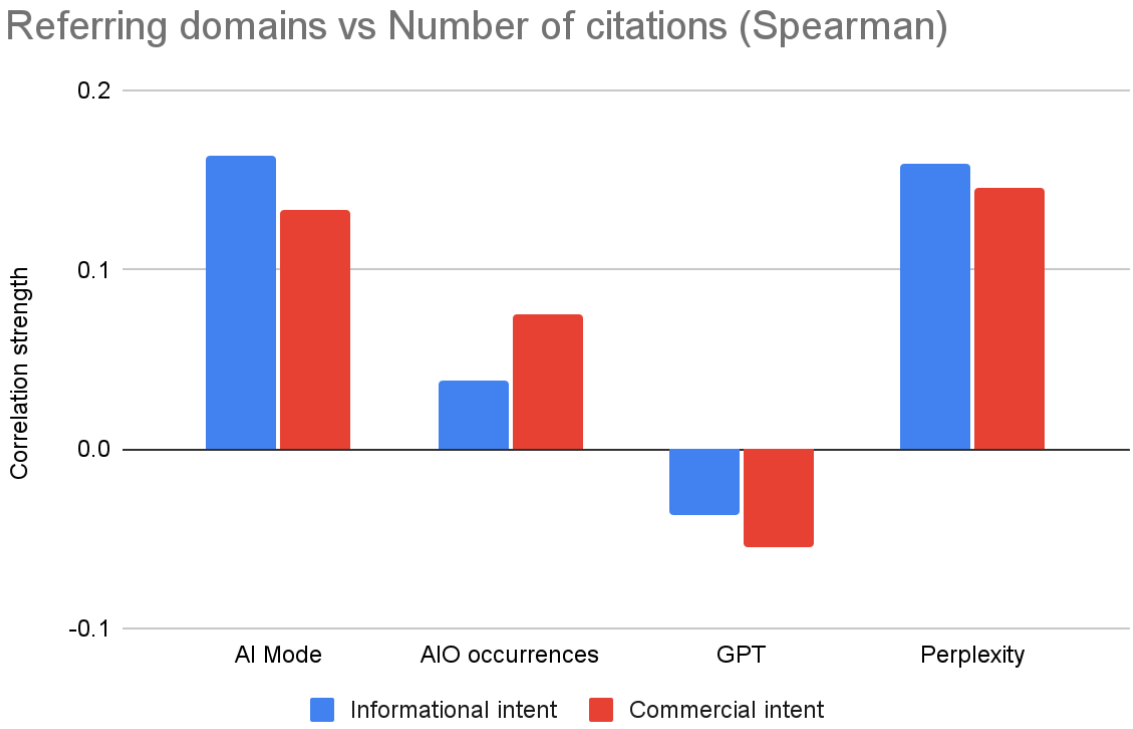


In Spearman, **AI Mode** is **0.141** on informational intent and **0.123** on commercial. Both weak but stronger than Pearson. **Perplexity** is similar at **0.135** and **0.134. GPT** is **negative** again for both intents. **AIO occurrences** have almost no correlation for informational intent at **0.029**.

Compared to Referring domains, Domain Authority shows much weaker linear impact. Unlike Referring domains, it does not produce any medium or strong relationships.



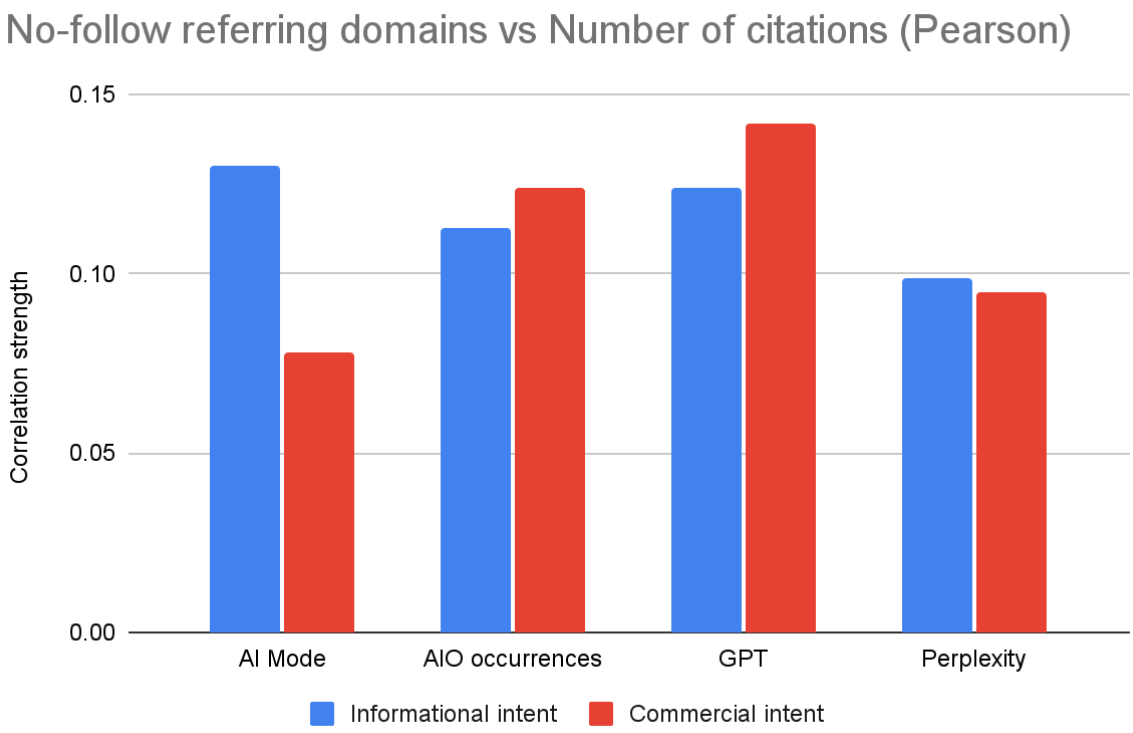
In Pearson, **informational intent shows clearly stronger relationships than commercial**. **GPT** leads at **0.319**, fairly strong, followed by **AI Mode** at **0.276** and **AIO occurrences** at **0.237**, both medium. **Perplexity** is slightly lower at **0.196**, weak to medium. For commercial intent, values drop, **GPT** at **0.239** is medium, **AIO** at 0.145 weak, **AI Mode** at 0.076 very weak.



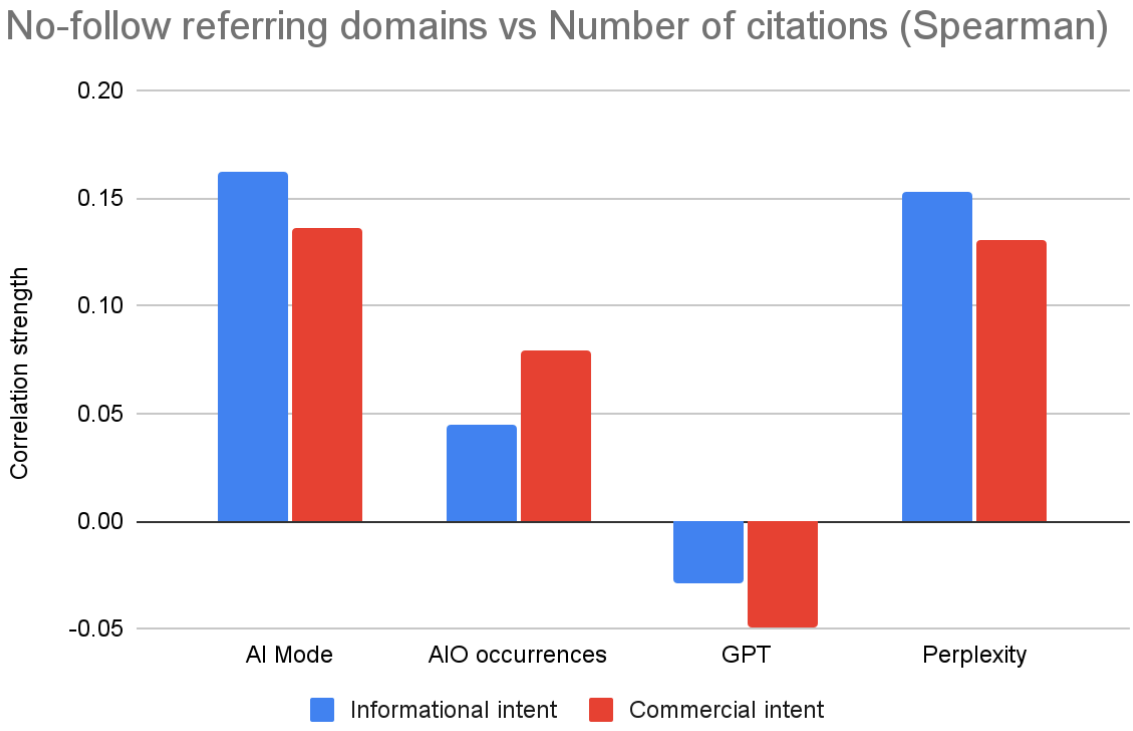
In Spearman, correlations weaken significantly. Informational **AI Mode** at **0.164** and **Perplexity** at **0.159** are weak, while **GPT** turns negative again at **minus 0.037**. Commercial intent shows the same pattern, modest positives except **GPT** at **minus 0.055**.

Compared to URL Authority, Referring domains creates stronger linear effects for informational queries, especially in GPT, and similarly, far less stable order-based (Spearman) patterns.

Now, let's see if number of referring domains with no-follow links only will present weaker correlations than the raw number of referring domains.

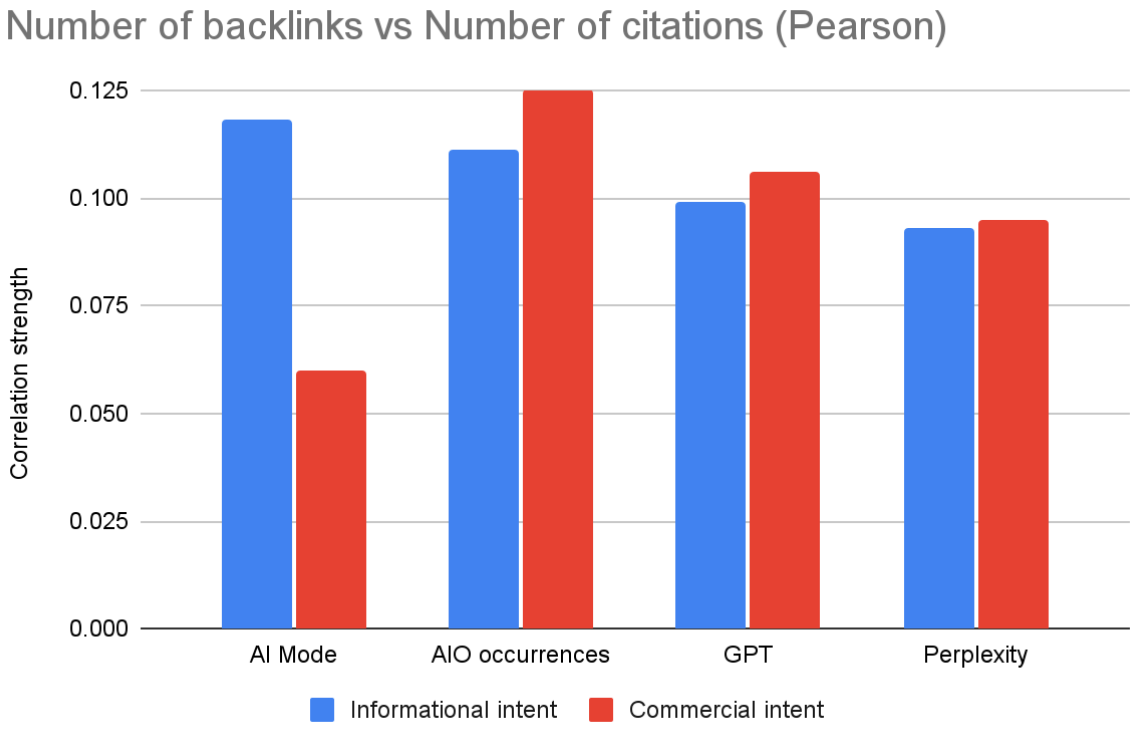


In Pearson, correlations remain very weak to weak. For informational intent, **AI Mode** leads at **0.130**, followed closely by **GPT** at **0.124** and **AIO** at **0.113. Perplexity** is lowest at **0.099.** For commercial intent, **GPT** wins at **0.142** and **AIO** follows at **0.124**, both weak, while **AI Mode** drops to **0.078**.

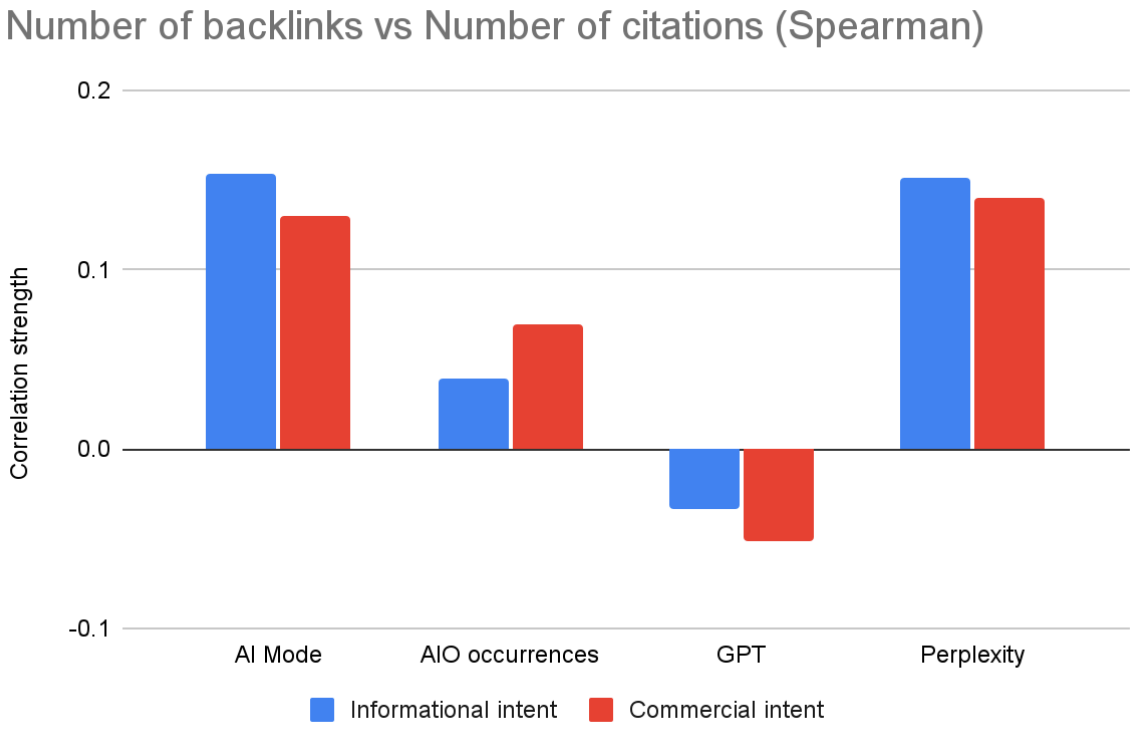


In Spearman, **AI Mode** is at **0.162** for informational intent and **0.136** for commercial. Both correlations are in the weak range but still stronger than for other models. **Perplexity** is close behind. **GPT** is **negative** again for both intents. **AIO** has almost no correlation for informational intent at **0.045** and a bit stronger but still very weak correlation for commercial intent.

Compared to Referring domains, No follow domains show much weaker linear effects. Similar to all other backlink-based factors, relationships are modest and unstable in order-based metric (Spearman).



In Pearson, all relationships are weak or very weak. For informational intent, **AI Mode** leads at **0.118**, with **AIO** at **0.111** and **GPT** at 0.099 close behind. Perplexity is lowest at **0.093**. For commercial intent, **AIO occurrences** wins at **0.125**, **GPT** follows at **0.106**, and **AI Mode** drops to **0.060**.



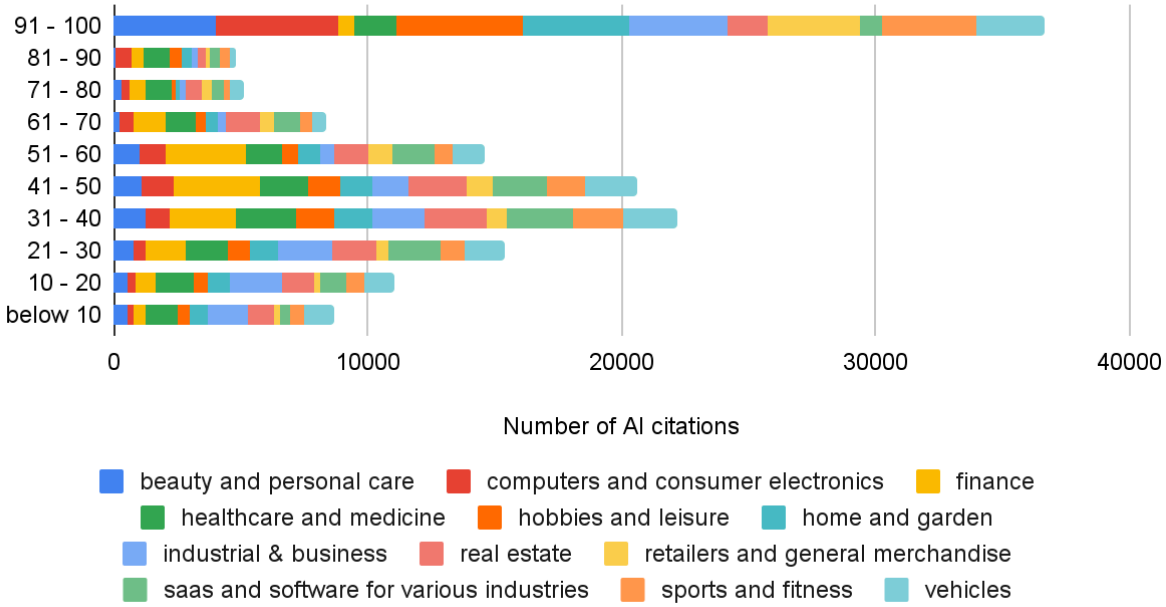
In Spearman, **AI Mode** is at **0.153** for informational intent and **0.130** for commercial one. Both are weak but still the strongest in each group. **Perplexity** is nearly identical. **GPT** is **negative again for both intents**. **AIO** has almost no correlation for informational intent at **0.039**.

Compared to Referring domains, raw Backlink count shows weaker linear impact. **The trend is consistent: Referring domains with do-follow links outperform other factors having its effect more visible on informational keyword sets**, while GPT repeatedly shows an inverse pattern in Spearman but on very low values.

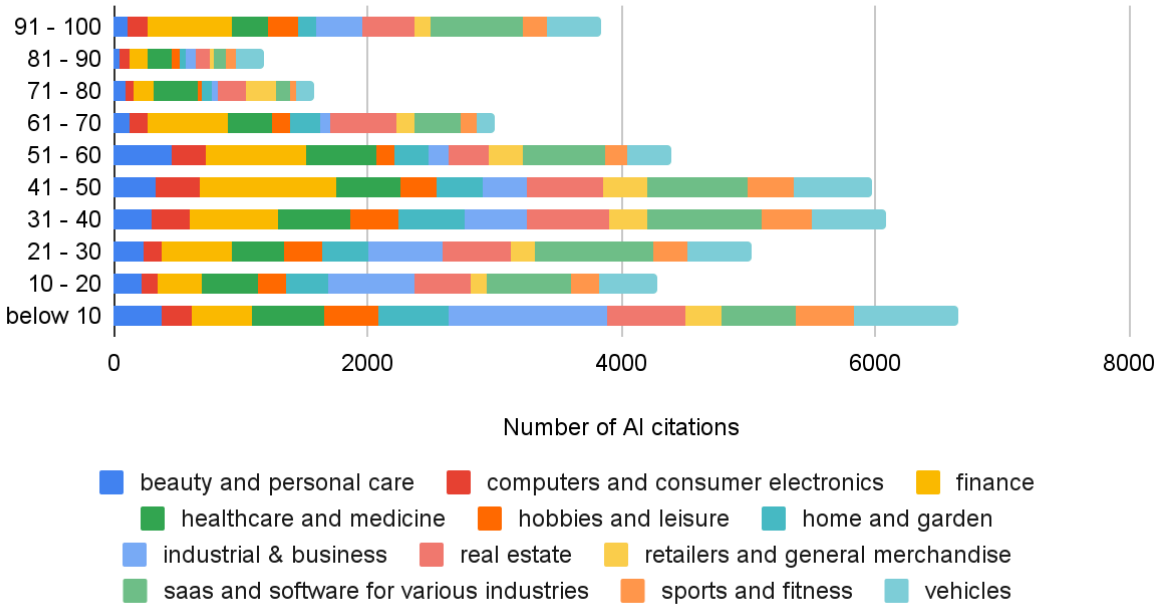
AI Citations by Domain Authority ranges

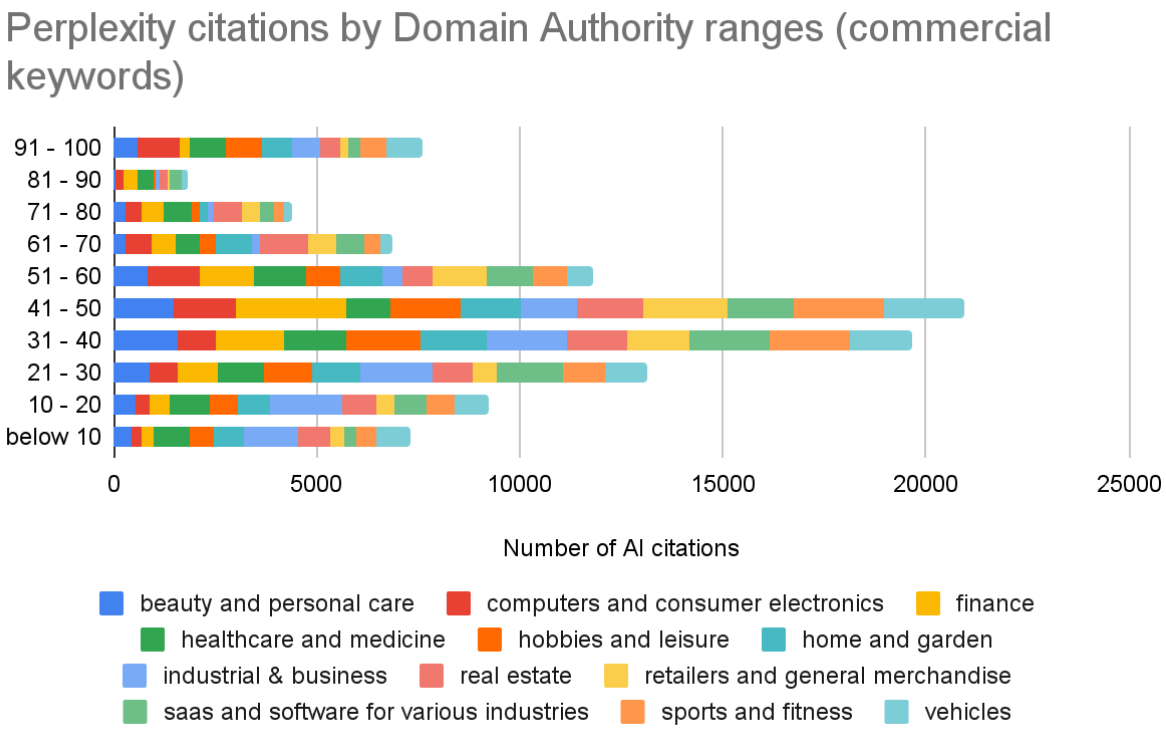
As with traditional SERPs, to account for non-linearity of the relationships, let's now examine the citation data grouped by Domain Authority ranges. First, let's look at commercial keywords across all models:

AI Mode citations by Domain Authority ranges (commercial keywords)



ChatGPT citations by Domain Authority ranges (commercial keywords)

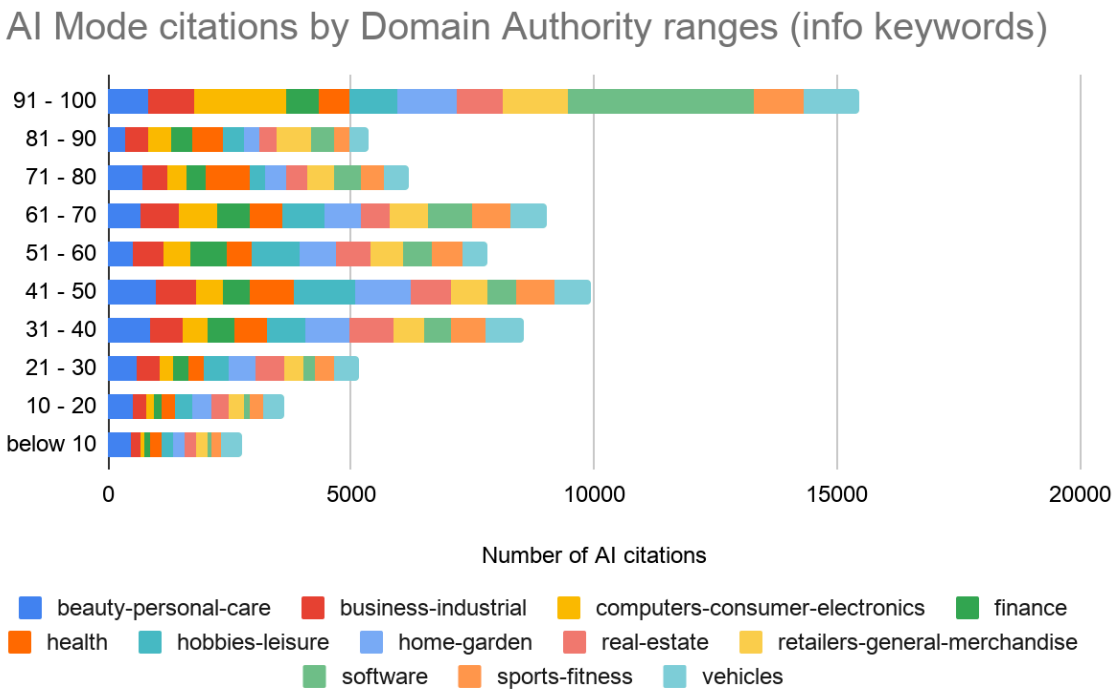


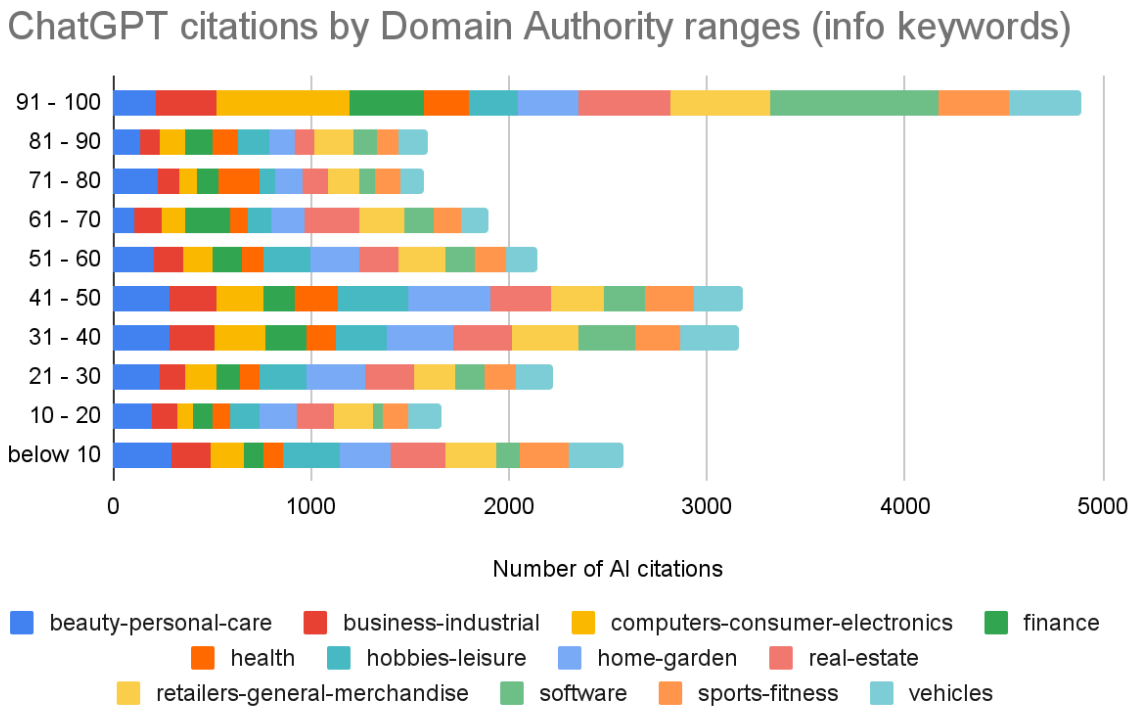


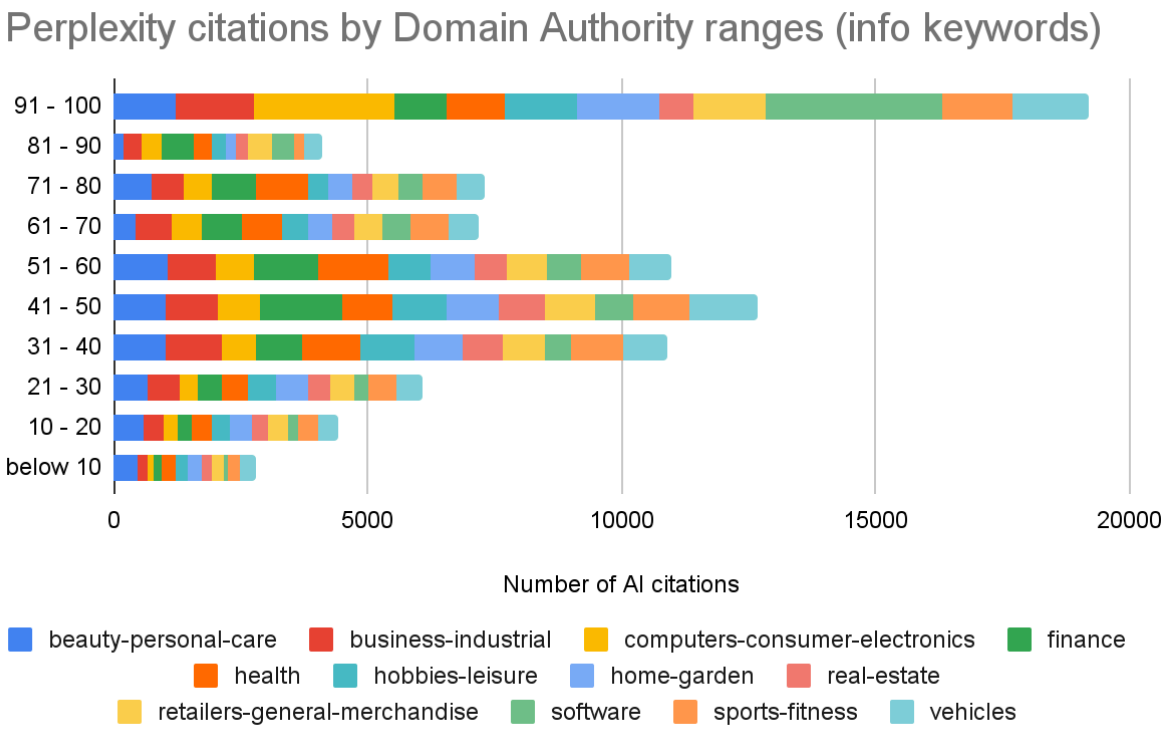
Google AI Mode and Perplexity show a similar pattern between Domain Authority and search visibility. The main difference is that Google relies much more on very high-authority websites. In our US-based data, most of these top citations come from Amazon or from links to Google's own product search results.

ChatGPT includes far fewer links in its answers. It also references more lower-authority websites (percentage-wise). This could mean it uses a different method to choose its sources, or that it applies less strict filtering. In future versions of this report, we will look at this more closely to see whether this pattern continues. This also helps explain the inverse correlations observed for GPT in the Spearman coefficients, as it tends to cite content from many low-authority websites.

Now, let's move to **informational keywords:**







For informational keywords, we see that not only Google AI Mode shows a strong bias toward the highest Domain Authority websites, but Perplexity and ChatGPT do as well. ChatGPT still cites the highest percentage of lower-authority websites compared to the other models, similar to what we observed for commercial keywords. However, this effect is weaker for informational queries.

It is also important to note that ChatGPT provides fewer citations overall than the other two models. It also shows fewer citations for informational keywords than for commercial ones.

Apart from that, the overall pattern is similar to commercial intent keywords. As Domain Authority increases, the benefits start to decrease

after the mid-authority range around 40 to 50. This suggests that the impact of higher authority becomes smaller beyond that point.

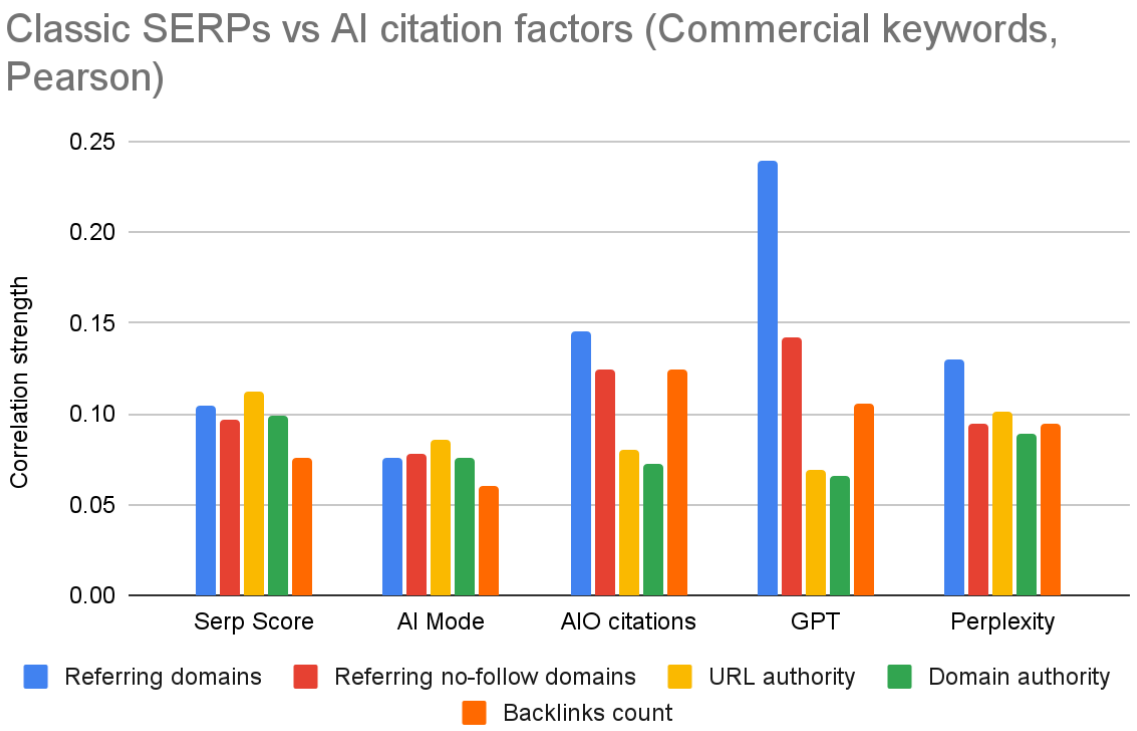
The highest authority bracket is dominated by the same platforms that often appear in traditional SERPs for informational queries. These include large user-generated or content-heavy platforms such as Reddit, Facebook, Wikipedia, Medium, TikTok, Pinterest, and YouTube, along with some government domains.

When we dug deeper into the data, we observed some model-specific preferences. Perplexity shows a strong tendency to cite YouTube, while ChatGPT references Yahoo relatively often.

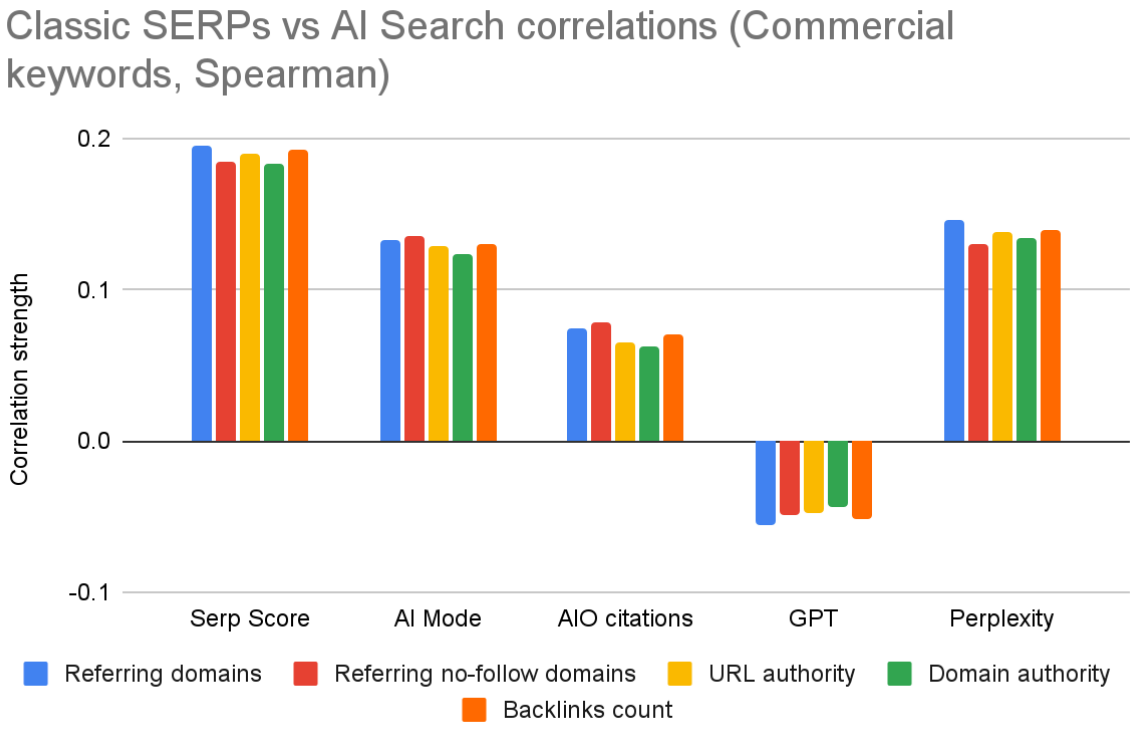
Backlinks: AI Search vs Traditional SERPs

This chapter compares how backlinks relate to visibility in traditional search results and AI-powered search. We analyze whether and to what extent the importance of backlinks changes across these environments.

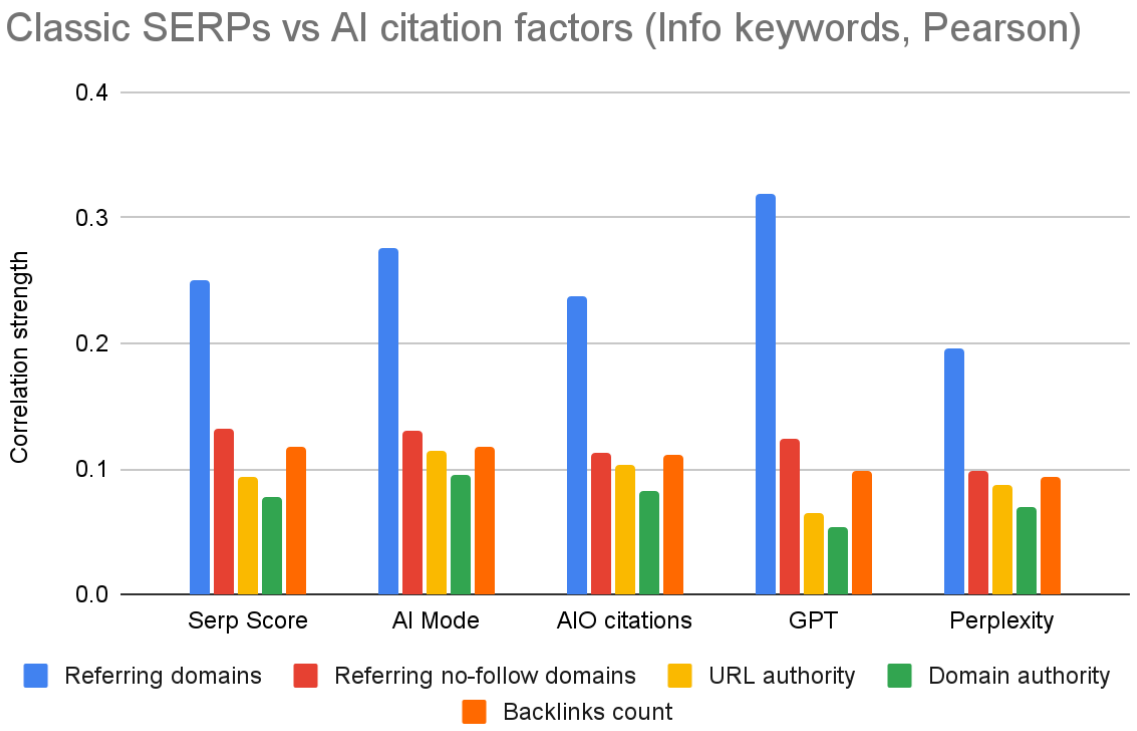
First, let's take a look at keywords with **commercial intent:**



In Pearson, **GPT** stands out. It shows a medium correlation with **Referring domains** at **0.239**, clearly the strongest value in the table. **AIO citations** also lean more on **Referring domains** at **0.145** and **Backlinks count** at **0.125. SERP Score** and **AI Mode** remain very weak across all factors, mostly around **0.07** to **0.11.**

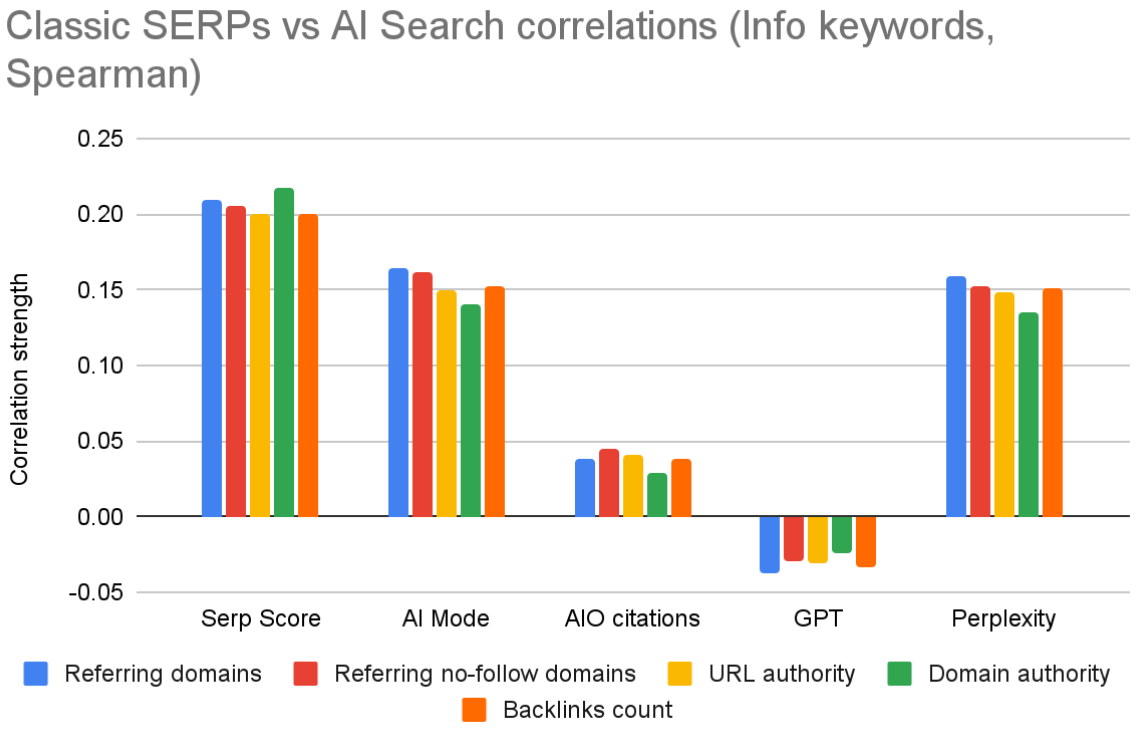


In Spearman, SERP Score is the strongest overall, weak but consistent around 0.18 to 0.195 across all factors. Perplexity and AI Mode are also weak and stable. GPT is negative for every factor. Overall, order-based (Spearman) relationships vary more between engines, while linear ones (Pearson) vary more within engines.



Here, where the data is depicting correlation between factors and info keywords, we see much stronger relation to referring domains than we did for commercial keywords.

Pearson clearly shows Referring domains as the strongest factor across almost all engines. It reaches 0.25 for SERP Score, 0.276 for AI Mode and peaks at 0.319 for GPT, which is fairly strong. Other factors are much weaker, mostly between 0.053 and 0.132, with Domain Authority consistently the lowest.



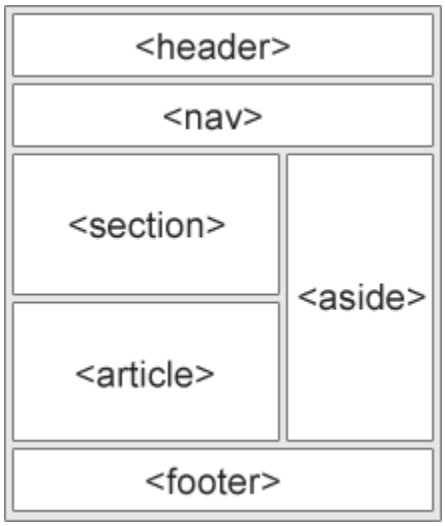
In Spearman, the gap largely disappears. For SERP Score, all factors cluster tightly between 0.2 and 0.217. AI Mode and Perplexity are also evenly distributed around 0.14 to 0.164. AIO citations have almost no correlation, and GPT is slightly negative across all factors.

The takeaway

Backlinks clearly matter for getting cited in AI search. However, their relationship with visibility in AI search appears to be less monotonic than with visibility in traditional SERPs, as reflected in the lower Spearman scores. The relationship is most non-monotonic in ChatGPT, largely because a high share of its citations comes from the websites in the lowest authority tier. This likely makes ChatGPT the easiest platform for newer websites to gain visibility.

Backlink location impact

Possible locations of a link:

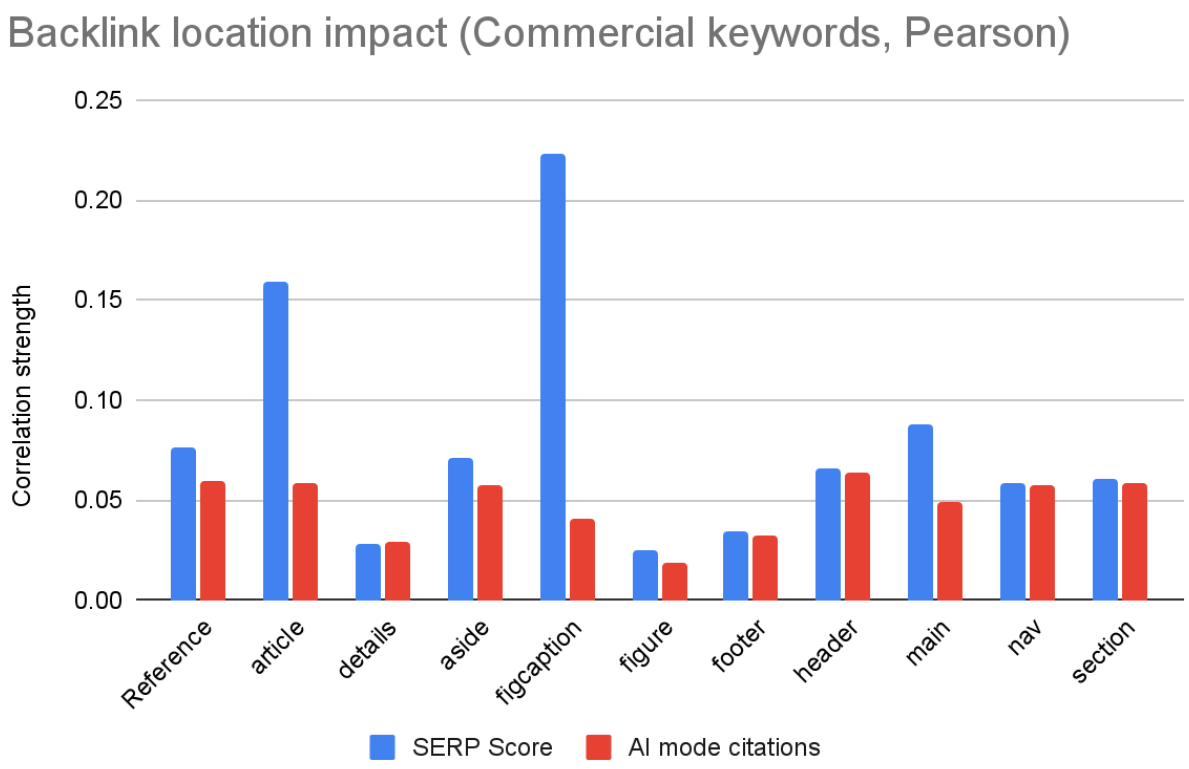


Source: https://www.w3schools.com/html/html5_semantic_elements.asp

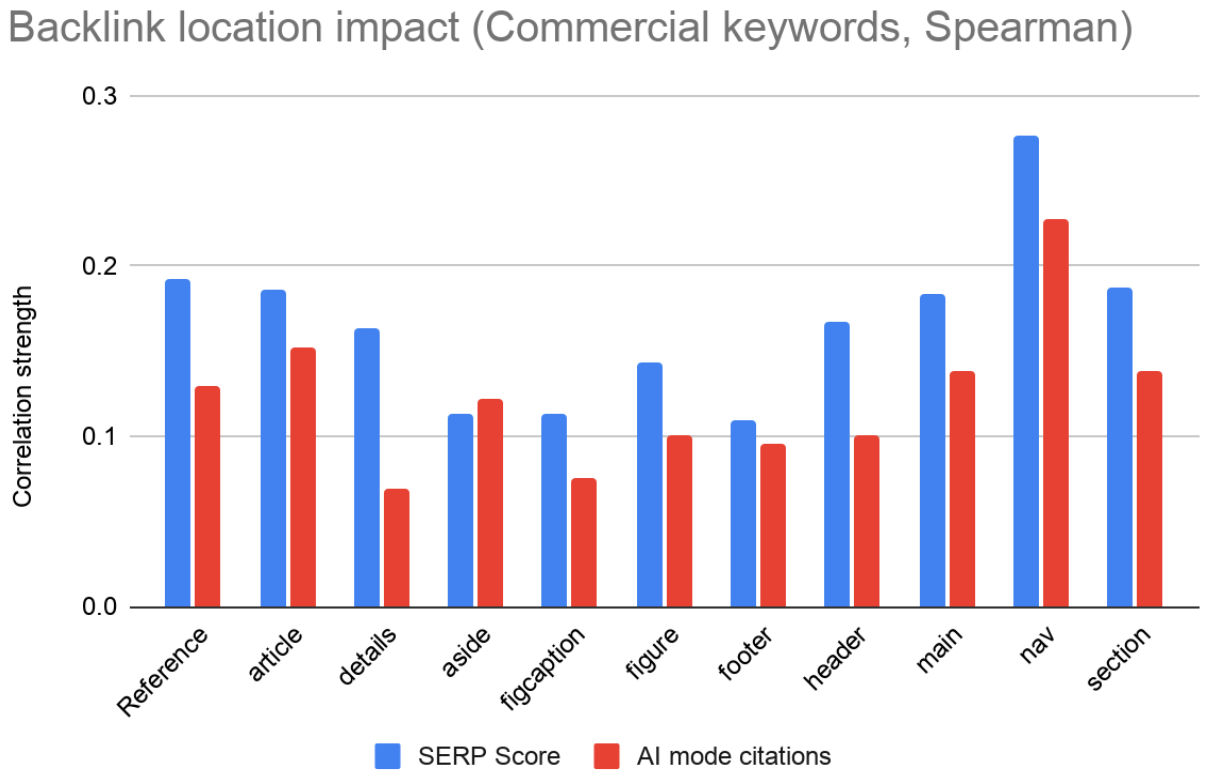
The classification of each link is based on html attributes of surrounding sections. More on that here: https://www.w3schools.com/html/html5_semantic_elements.asp

Important note: all links in our study include both external and internal links to a domain.

Here is a breakdown of how particular backlink locations differ between each other with how strong is their relation to visibility metrics:

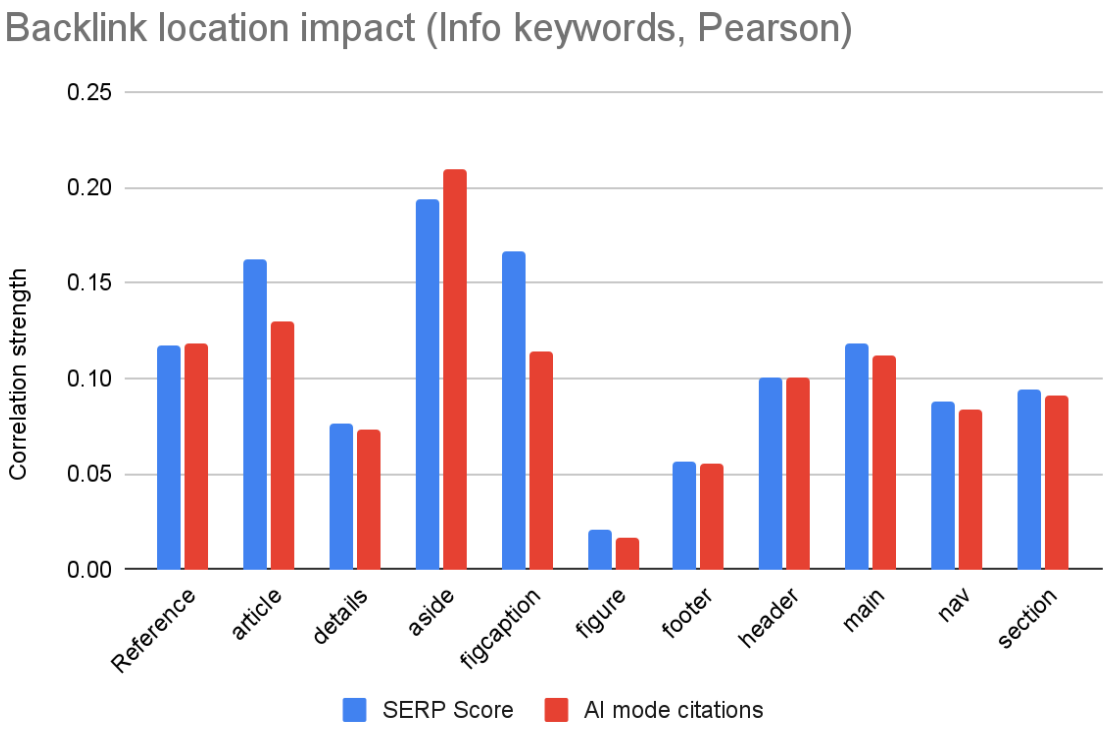


In Pearson, the biggest gap between intents is **Figcaption.** It is medium for **SERP Score: 0.223**, but drops to very weak for **AI Mode: 0.041.**

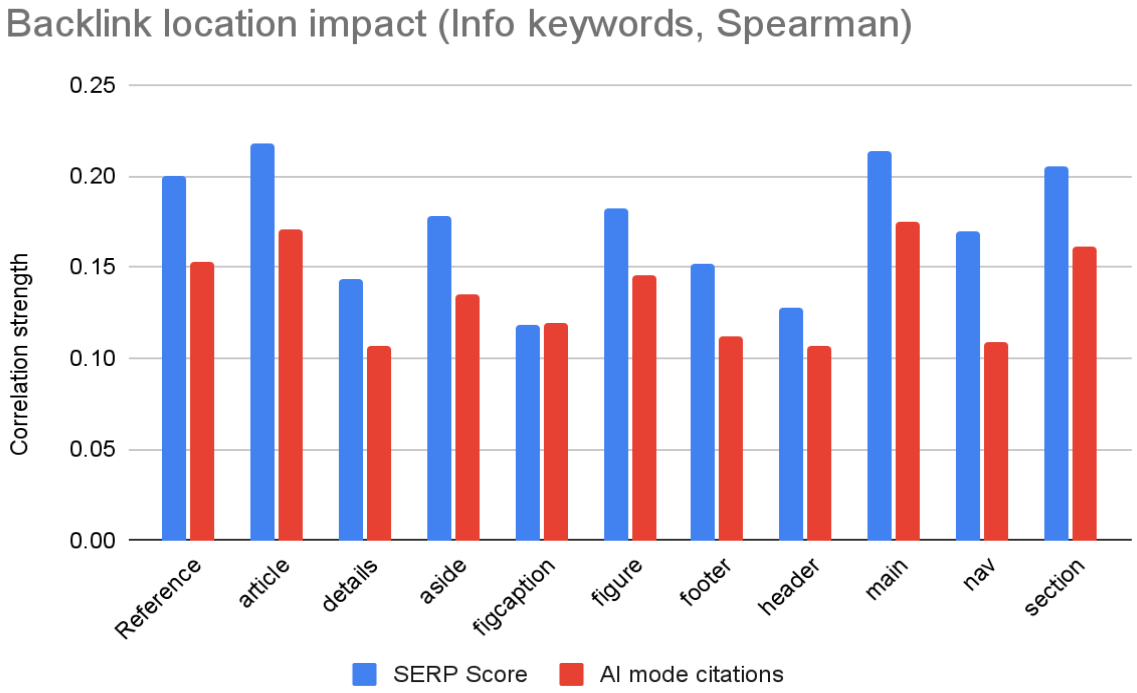


In Spearman, **nav** shows the strongest relationship for both **SERP Score** at **0.277** and **AI Mode** citations at **0.227**, both medium, though AI Mode is consistently lower. Most likely for this link placement, internal links play the most significant role. **Reference, article**, **main**,, and **section** correlations are the strongest after the **nav, but still** all within a weak range for both intents. In most groups, SERP correlations exceed AI Mode, sometimes clearly, like for **Reference: 0.193** vs **0.130**, or for **details: 0.163** vs **0.069**.

Let’s move on to the keywords with **informational intent**.



In Pearson, **aside** stands out a bit, weak (almost medium) for **SERP Score** at **0.194** and slightly stronger for **AI Mode** at **0.209** (lower end of medium range), the only case where **AI Mode** clearly exceeds **SERP Score**. Most other placements are within very weak range, and **figure** has almost no correlation for both **SERP Score** and **AI Mode**.



Spearman consistently shows correlations within the weak to low medium range, but stronger than Pearson. **Article** at **0.218**, **main** at **0.214** and **section** at **0.205** lead for **SERP Score**, all medium. **AI Mode** citations follow the same pattern, slightly lower in most cases, with **article** at **0.171** and **main** at **0.175** the strongest. **Figcaption** is almost identical for both: **0.118** and **0.119.**

The takeaway

In both Pearson and Spearman correlations, in almost all placements, AI Mode correlations trail SERP Score, confirming again that traditional rankings are more strongly bound to links in general than AI Mode citations are.

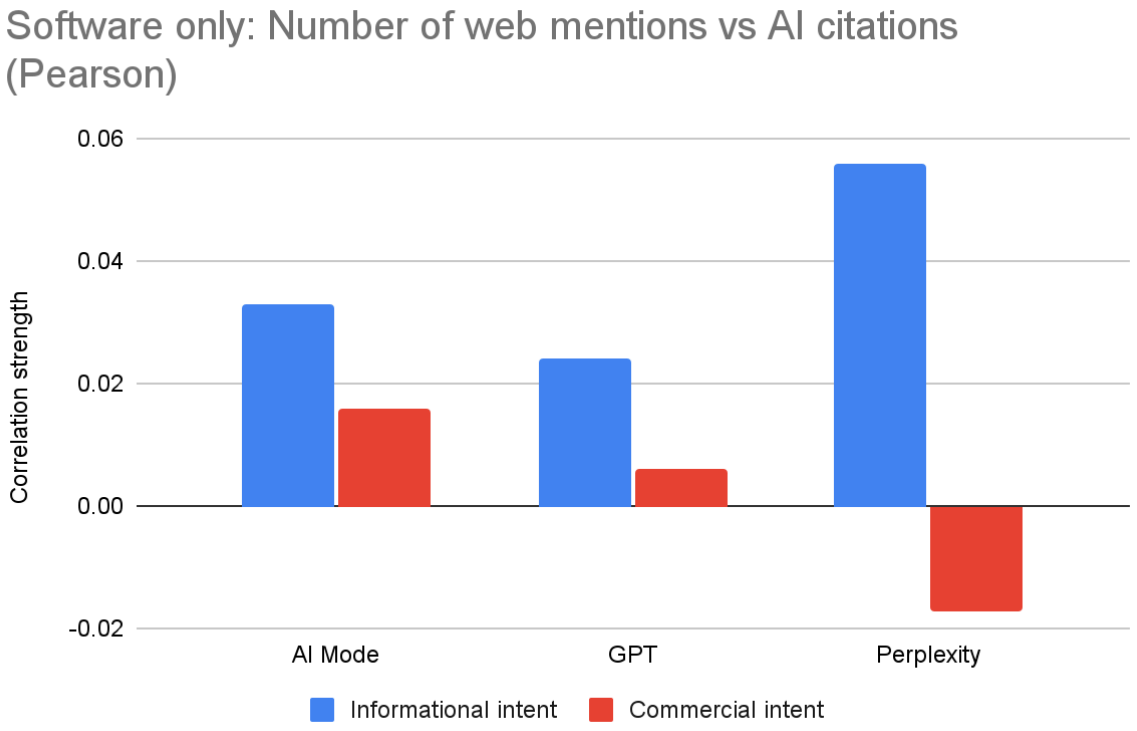
Also, Spearman correlations are stronger than Pearson's for both classic SERPs and AI Mode, meaning the correlations are more consistent in terms of ordering than linearity.

PART II: Mentions vs Visibility

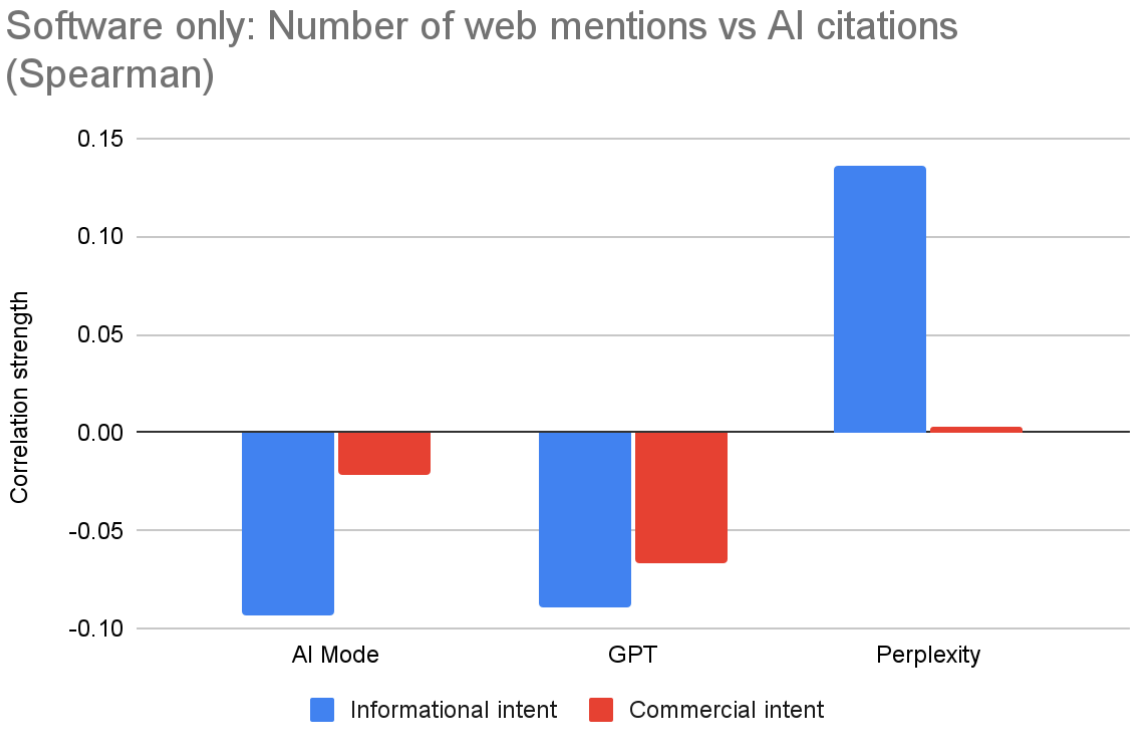
This chapter analyzes the relationship between unlinked web mention based factors and visibility metrics. **The analysis is limited to the Software industry**. As explained earlier, matching brand entities across AI generated responses requires consistent naming and clear brand structures. Software brands typically have simpler and more standardized entity formats, which reduces classification errors and increases reliability of the results.

Web Mentions and AI Citations

Below, we examine the relationship between unlinked web mentions and citations in AI search results. We present correlation data to show how often websites of brands that receive more mentions across the web also appear as sources in responses from different AI models.

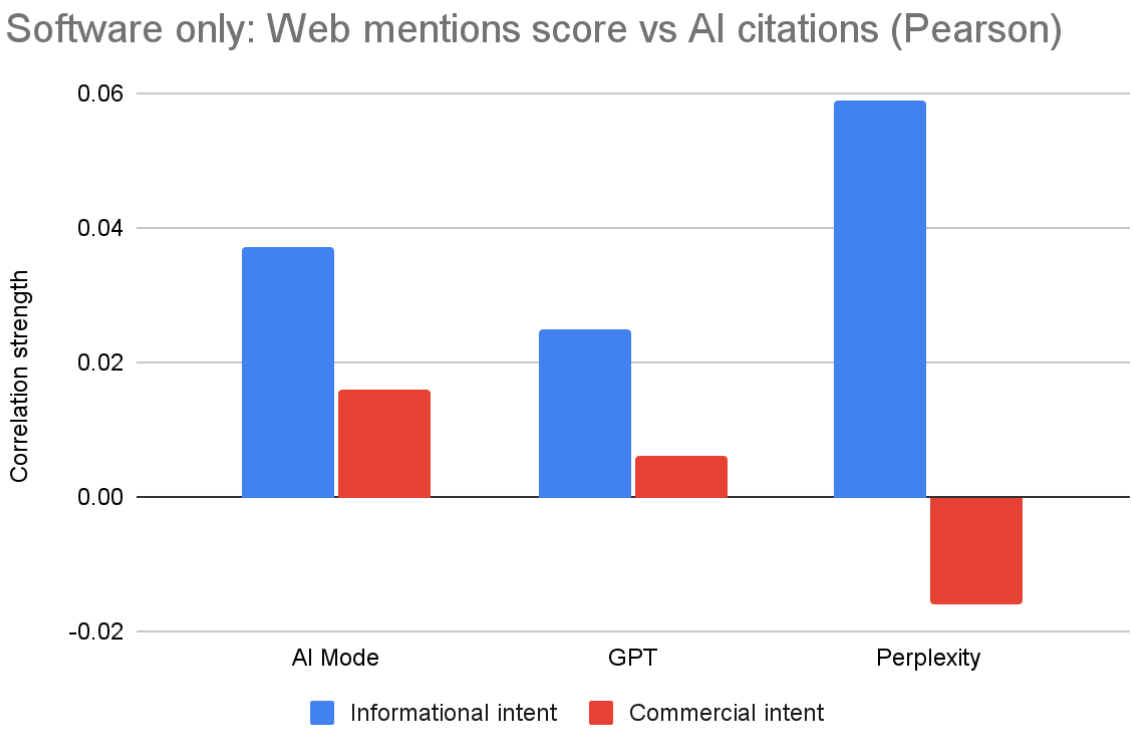


In Pearson, correlations are almost nonexistent. For informational intent, **Perplexity** is the highest at **0.056**, still very weak. **AI Mode** at **0.033** and **GPT at 0.024** are negligible. For commercial intent, values drop further, with **GPT** at **0.006** and **Perplexity** turning slightly negative at **minus 0.017**.



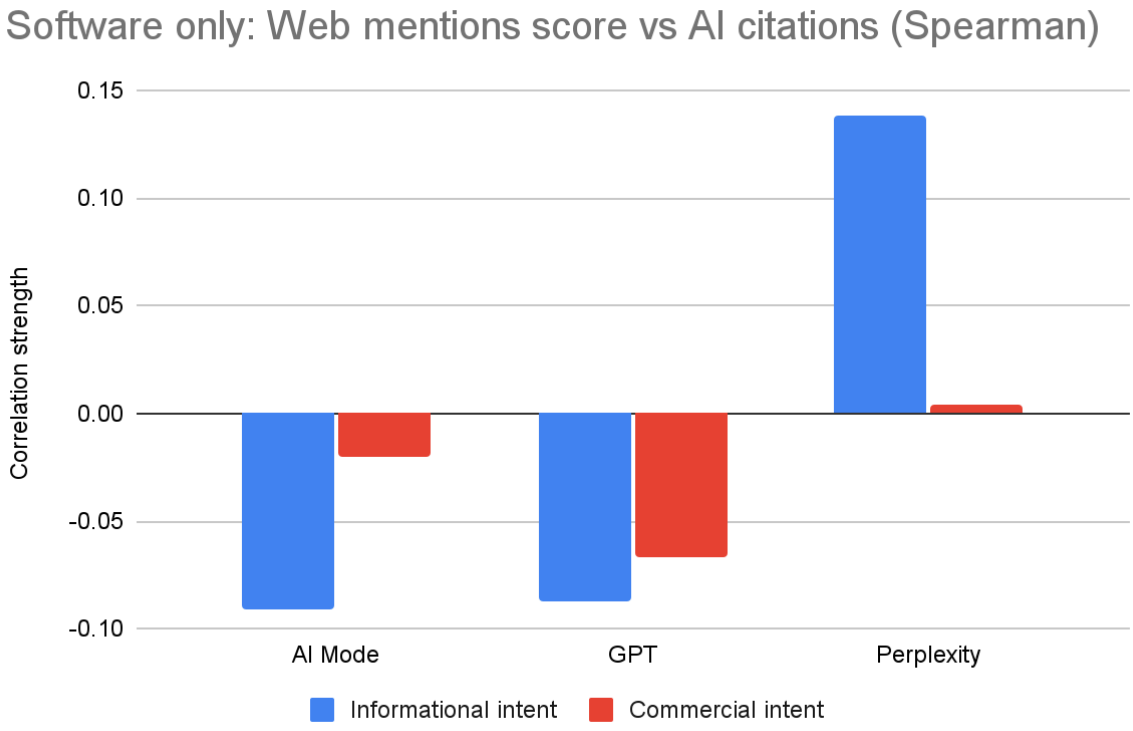
In Spearman, patterns shift a little. For informational queries, **Perplexity** reaches **0.136**, weak but clearly stronger than the rest. **AI Mode** at **minus 0.093** and **GPT** at **minus 0.089** show inverse relationships. **For commercial intent, all values hover around zero**, with **Perplexity** at **0.003** being the weakest and **GPT** at **minus 0.067** being the strongest but still very weak.

Now let's see what changes when we consider not only the number of web mentions, but also the authority of the websites that provide these unlinked mentions. **Web mentions score** metric accounts for that.



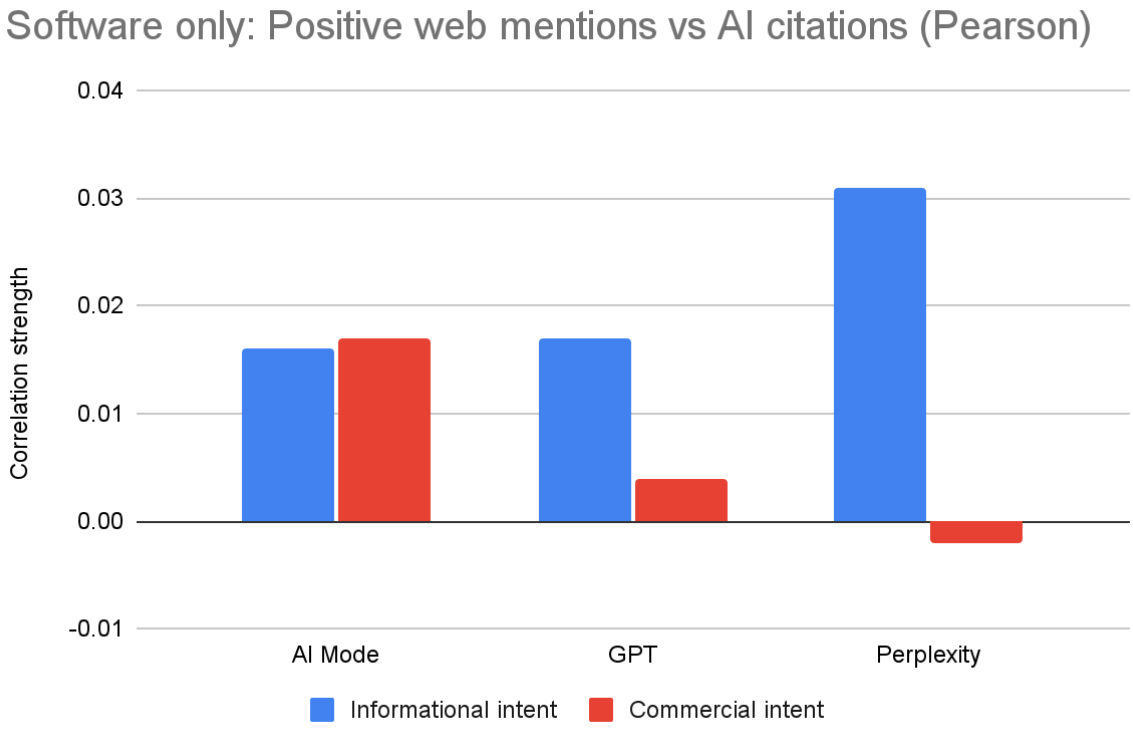
Compared to raw web mentions, weighting them by Domain Authority changes almost nothing.

In Pearson, correlations remain extremely weak. Informational intent shows **Perplexity** at **0.059**, **AI Mode** at **0.037** and **GPT** at **0.025**. Commercial intent stays near zero, with **GPT** at **0.006** and **Perplexity** at **minus 0.016**. The values are nearly identical to the previous set.

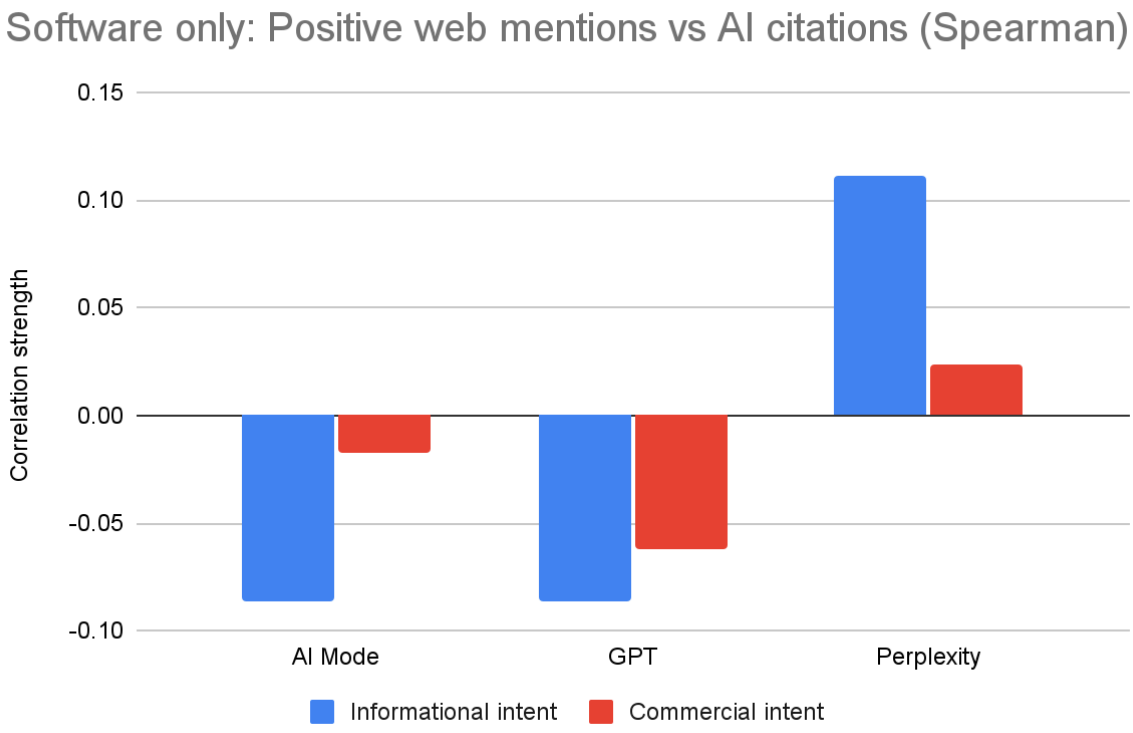


In Spearman, the same pattern holds. Informational **Perplexity** reaches **0.138**, weak and the only meaningful value. **AI Mode** at **minus 0.091** and **GPT** at **minus 0.087** remain negative. Commercial intent again shows even weaker relations.

The clear trend is that even authority weighted unlinked web mentions do not improve correlations with AI citations.



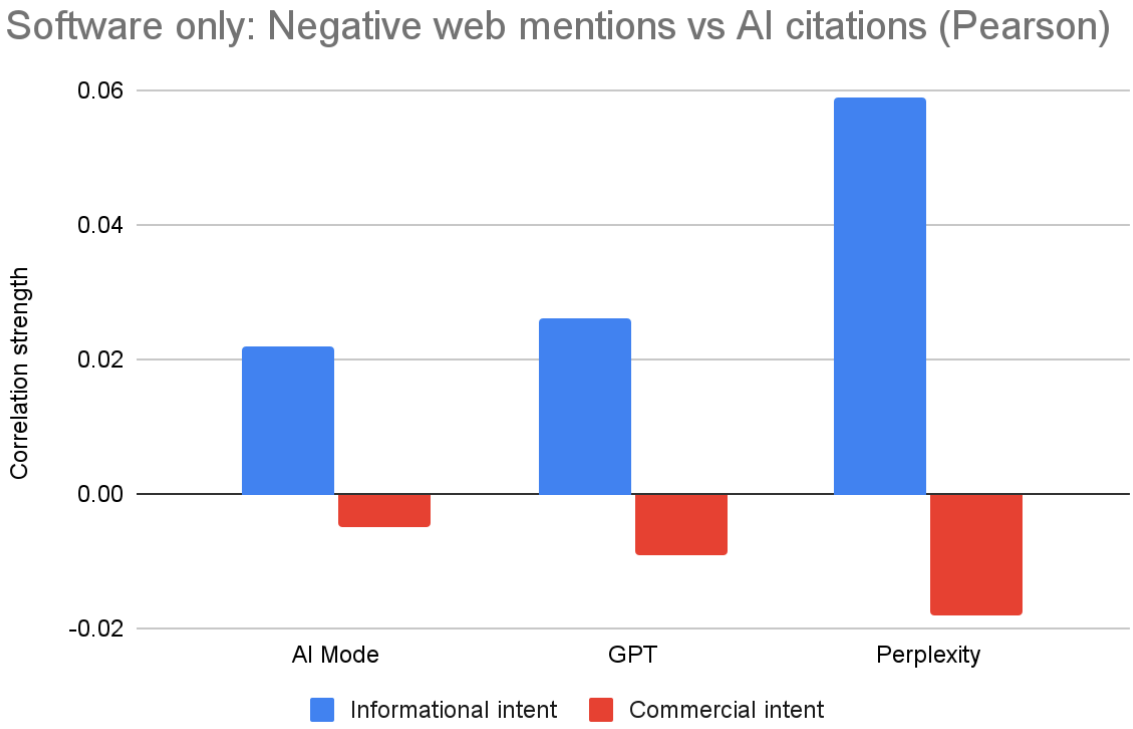
For positive web mentions, Pearson correlations are even weaker than in the previous two datasets. Informational intent peaks at **Perplexity** at **0.031**, while **AI Mode** at **0.016** and **GPT** at **0.017** are negligible. Commercial intent is also negligible, with **GPT** at **0.004** and **Perplexity** slightly negative at **minus 0.002** being even weaker than informational intent correlations.



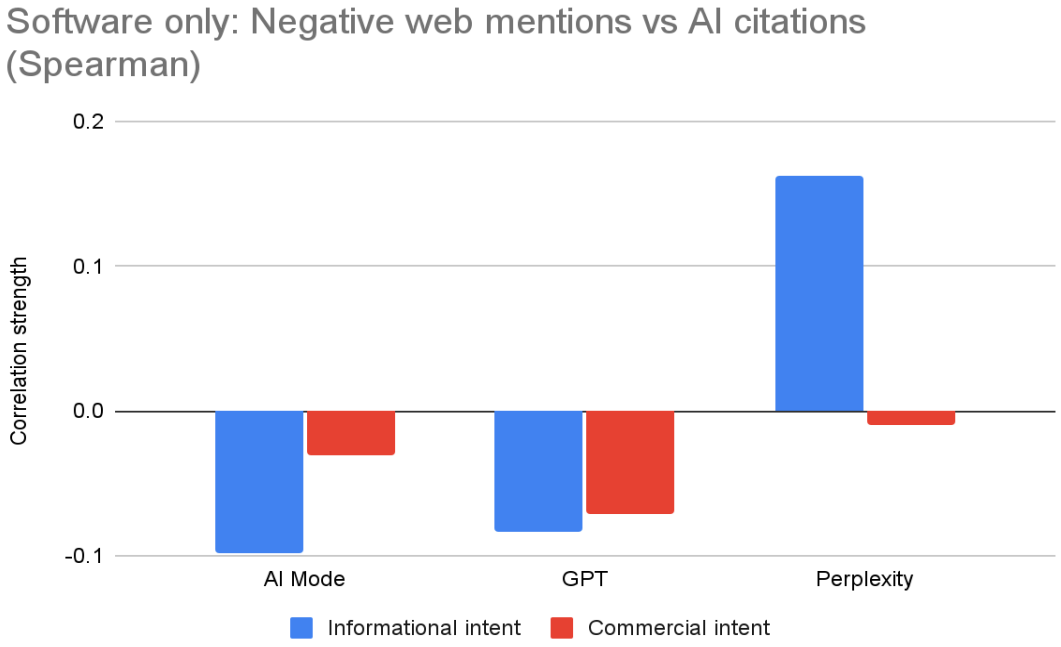
In the Spearman results, the pattern remains similar to what we observed with the other mention-based factors. Informational **Perplexity** reaches **0.111**, weak but lower than the 0.136 to 0.138 seen with raw **Web mentions** and **Web mentions score**. **AI Mode** at **minus 0.086** and **GPT** at **minus 0.086** stay negative. Commercial intent again clusters around zero.

Now, let's take a look at the influence mentions with negative sentiment have.

Note: negative mentions can still show a positive correlation with visibility because they often scale with overall attention. Brands that receive many negative mentions typically receive many mentions in general. In other words, high criticism usually comes with high exposure. As a result, negative sentiment does not automatically reduce visibility metrics, since it may simply reflect a higher overall level of public discussion.

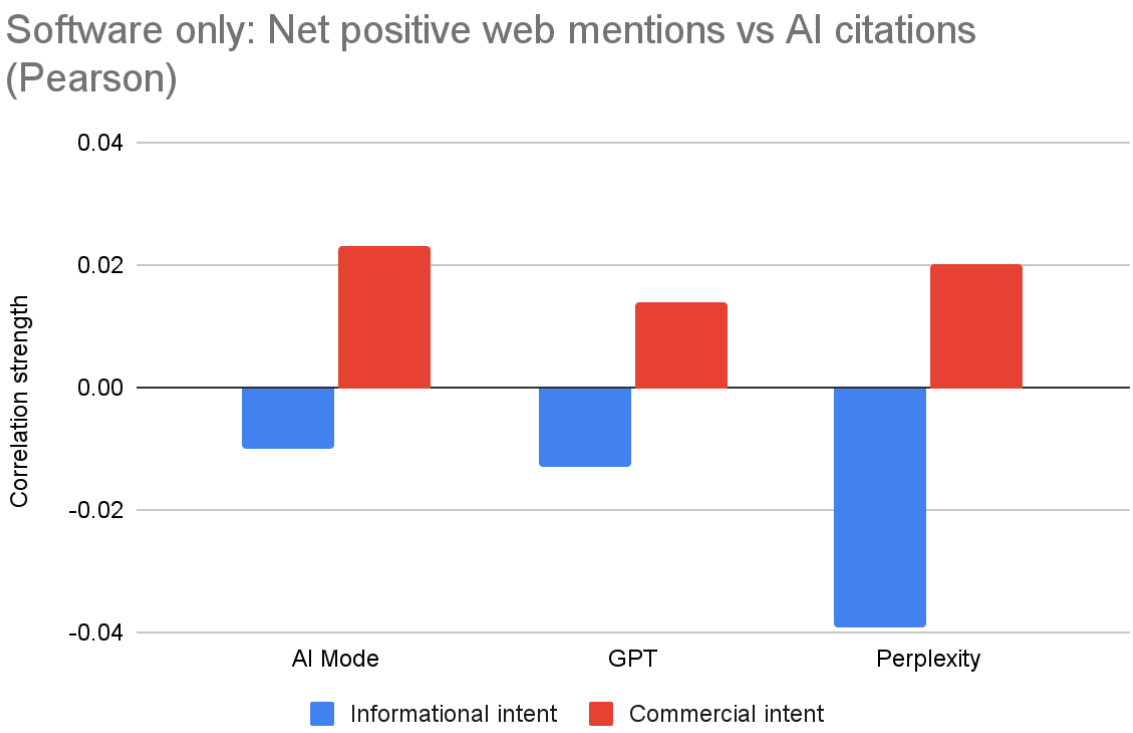


Pearson correlations remain extremely weak. Informational intent peaks at **Perplexity with 0.059**, similar to earlier mention based datasets. **AI Mode** at **0.022** and **GPT** at 0.026 are negligible. For commercial intent, all values turn slightly negative, from **minus 0.005** to **minus 0.018**.

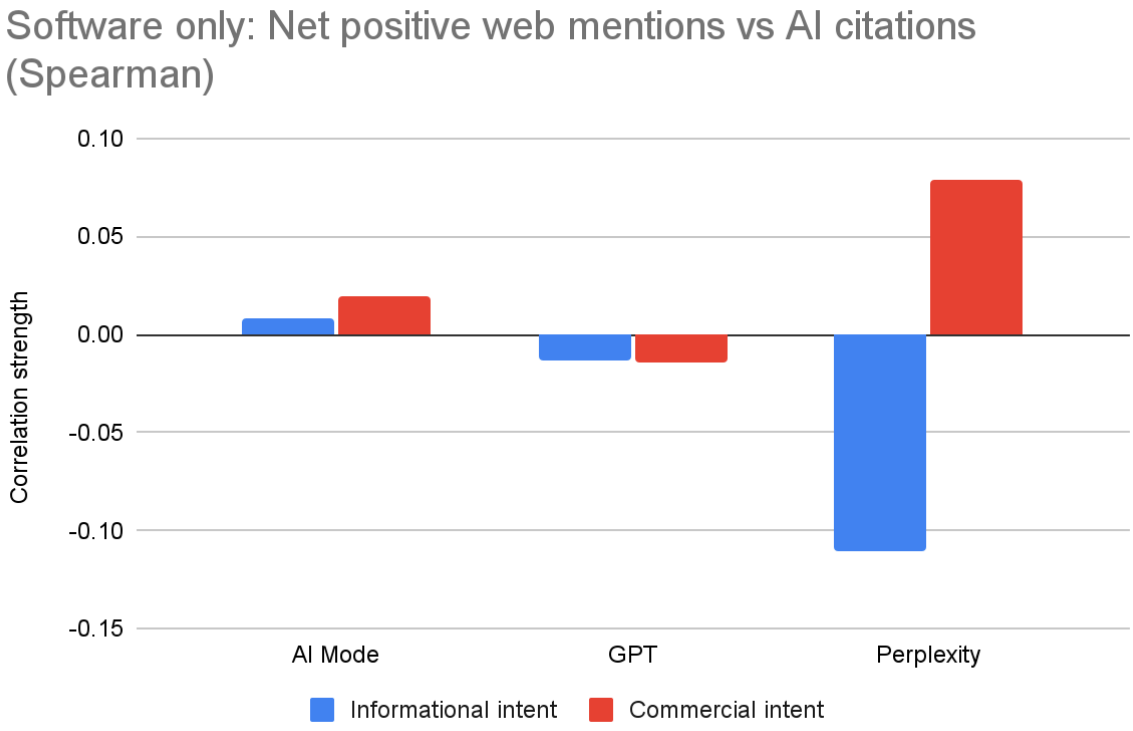


In Spearman, the divergence becomes clearer. Informational **Perplexity** reaches **0.162**, weak and actually higher than in the previous three mention based sets. At the same time, **AI Mode** sits at minus **0.098** and **GPT** at **minus 0.083** remain negative. Commercial intent again sits around zero (**AI Mode** and **Perplexity**) or weak negative (**GPT**).

Let’s see if anything changes when we look only at the difference between positive and negative mentions:



For net positive mentions, Pearson correlations are essentially zero. Informational intent is slightly negative across all models. Commercial intent turns slightly positive but still negligible, topping at **0.023** for **AI Mode**.



In Spearman, values remain unstable and, with the exception of Perplexity, completely negligible. Informational intent is near zero for **AI Mode** at **0.008** and **GPT** at **minus 0.013**, while **Perplexity** drops to **minus 0.11**, an inverse pattern within weak range. Commercial intent shows **Perplexity** at **0.079** within very weak but positive range.

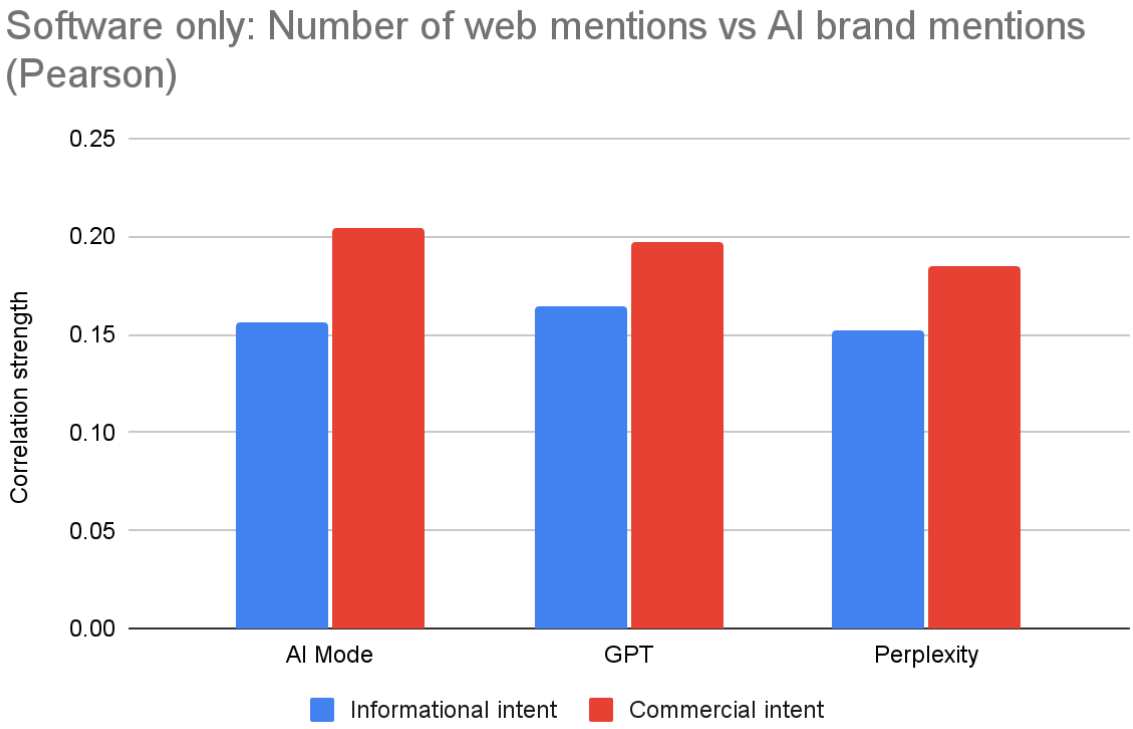
Compared to the previous sentiment-agnostic mention-based factors, the pattern holds: sentiment adjustments do not strengthen the relationships, and correlations remain minimal and inconsistent.

The takeaway

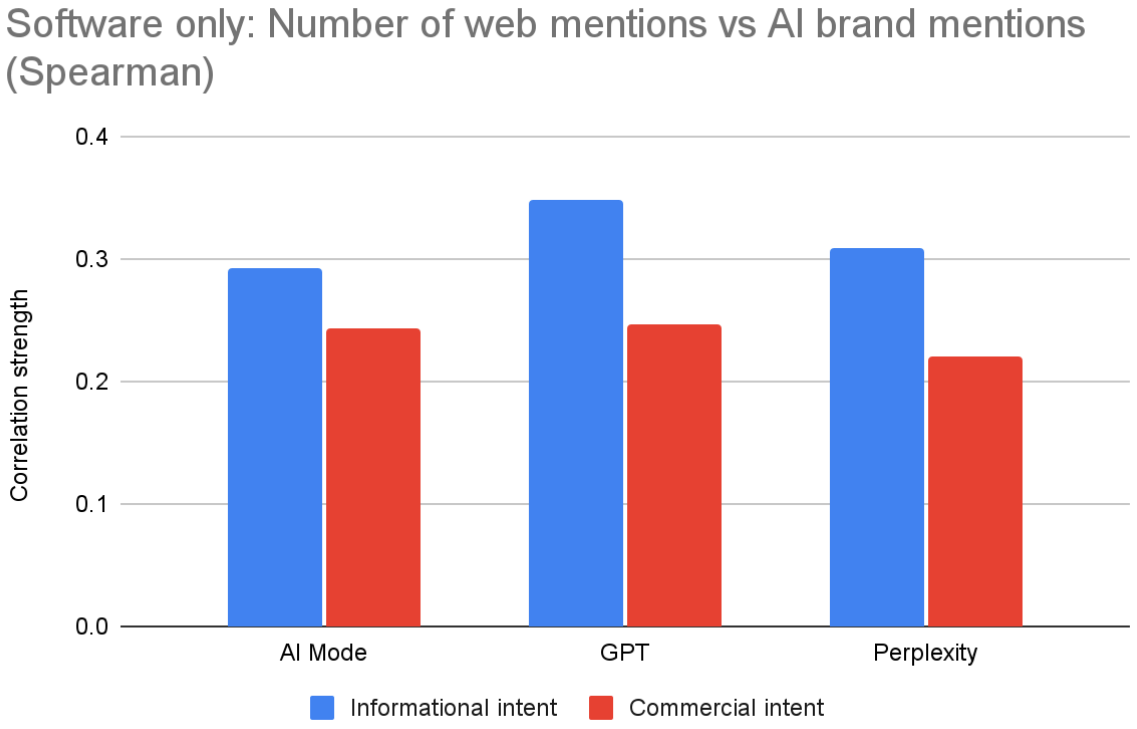
Unlinked brand mentions show no meaningful relationship with how often a website gets cited in LLM responses.

Web mentions vs AI Brand mentions

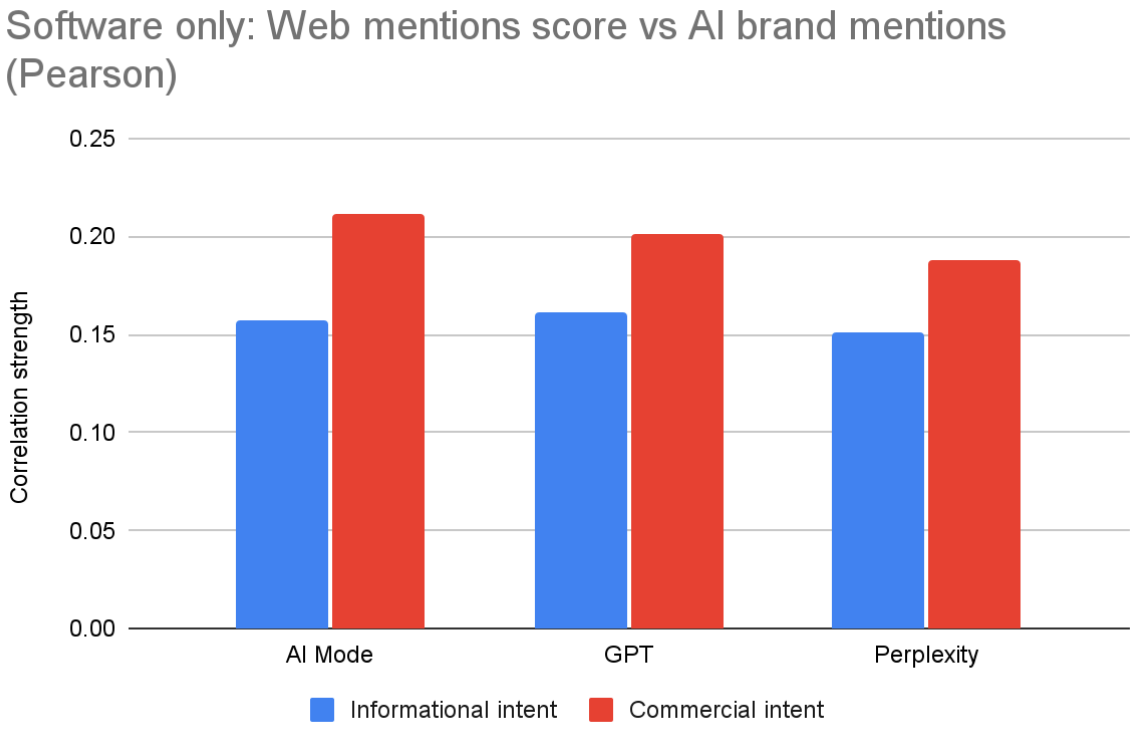
This section focuses on the relationship between unlinked web mentions and how often a given brand appears in LLM-generated answers. Do not confuse this with the previous chapter, where we examined how unlinked brand web mentions relate to how often a brand's website is cited as a source.



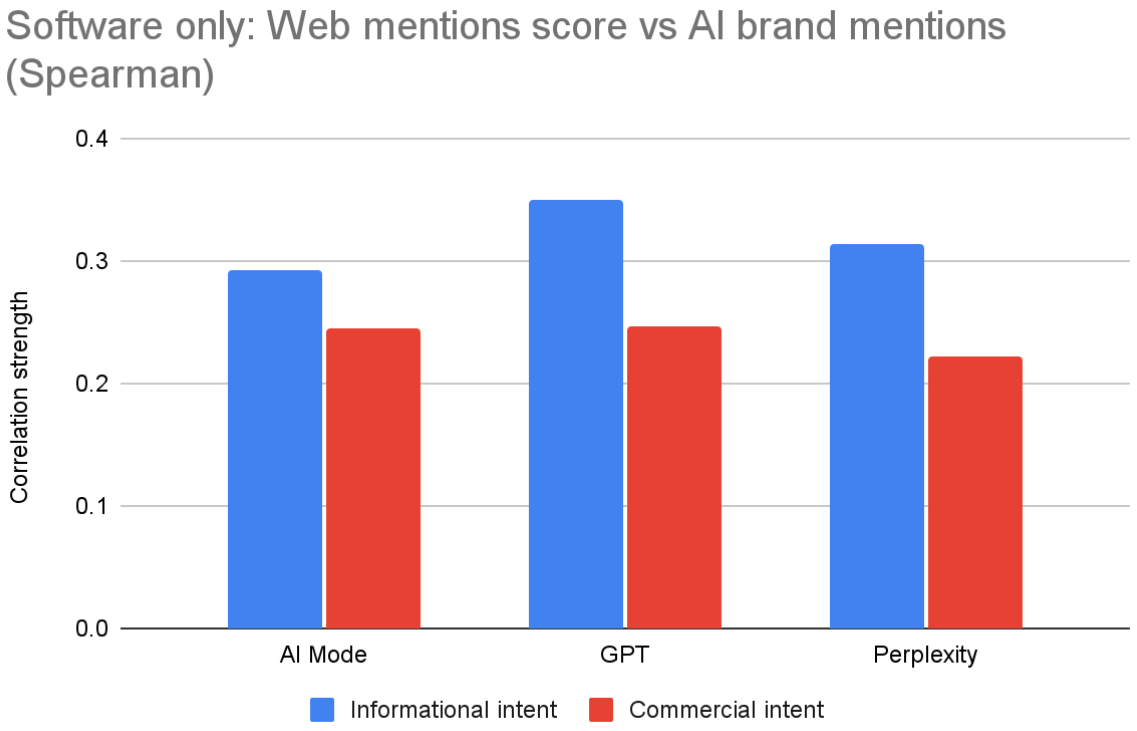
In Pearson, correlations are consistently within weak to medium ranges. For informational intent, **GPT** leads at **0.164**, followed closely by **AI Mode** at **0.156** and **Perplexity** at **0.152**. For commercial intent, values are stronger, peaking at **0.204** for **AI Mode** and **0.197** for **GPT**.



In Spearman, relationships become clearly stronger. **GPT** reaches **0.348** for informational intent, fairly strong, with **Perplexity** at **0.309** and **AI Mode** at **0.292** also high. Commercial intent remains a little lower, within the medium range around **0.22** to **0.246**.

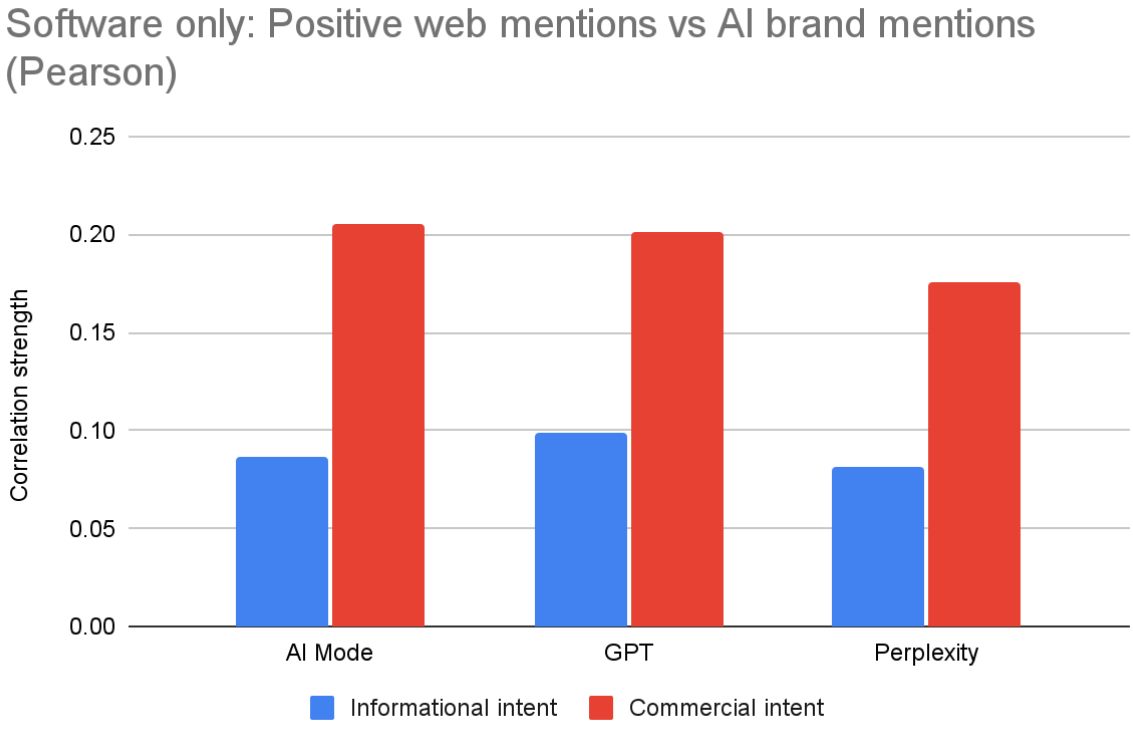


In Pearson, correlations stay weak to medium. Informational intent ranges from **0.151** to **0.161**, with **GPT** slightly leading. Commercial intent is stronger, peaking at **0.211** for **AI Mode** and **0.201** for **GPT**.

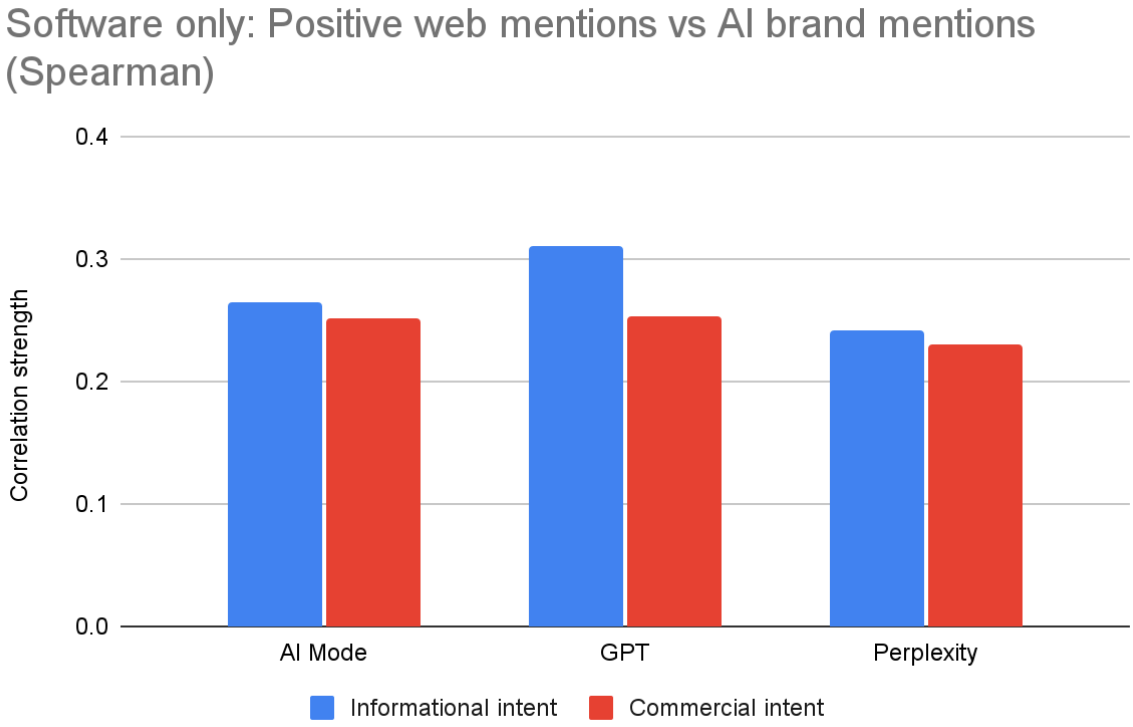


In Spearman, relationships are clearly stronger. **GPT** reaches **0.349** informational, fairly strong, with **Perplexity** at **0.313** and **AI Mode** at **0.293** close behind. Commercial intent remains weaker but still consistent around **0.222** to **0.247**.

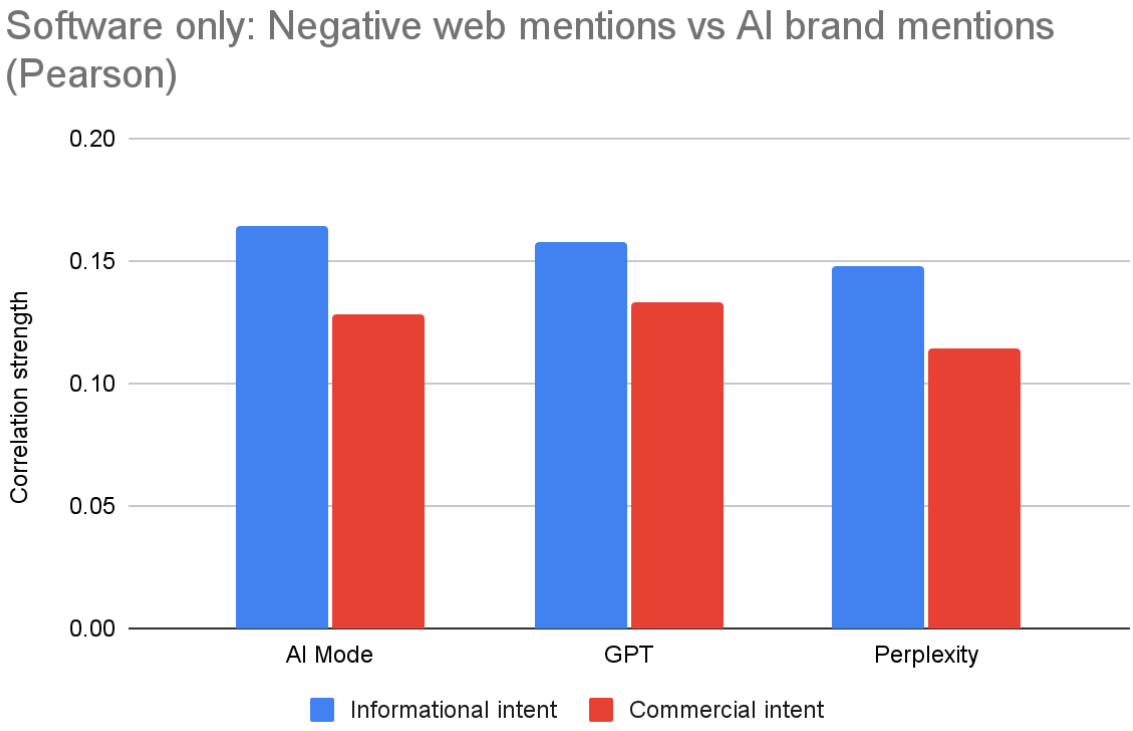
The results are almost identical to raw web mentions.



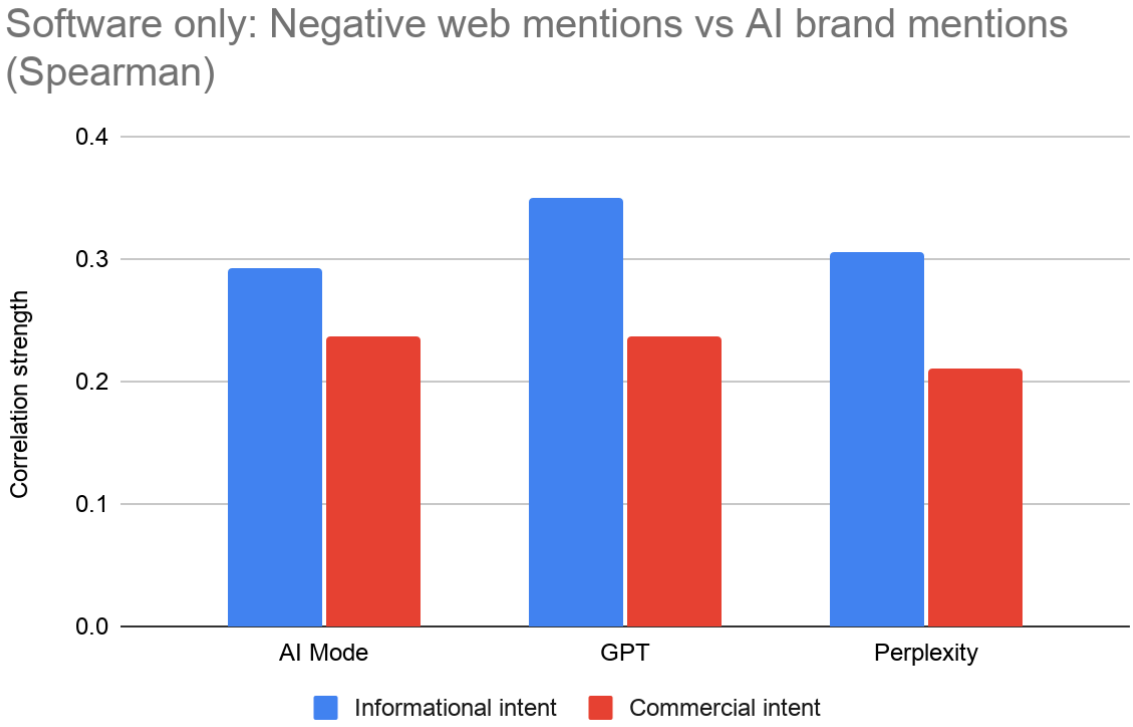
For positive web mentions, Pearson shows a split pattern. Informational intent is weaker, ranging from **0.082** to **0.099**, all upper levels of the very weak range. Commercial intent is clearly stronger, peaking at **0.205** for **AI Mode** and **0.201** for **GPT**, both medium. **Perplexity** follows at **0.176**.



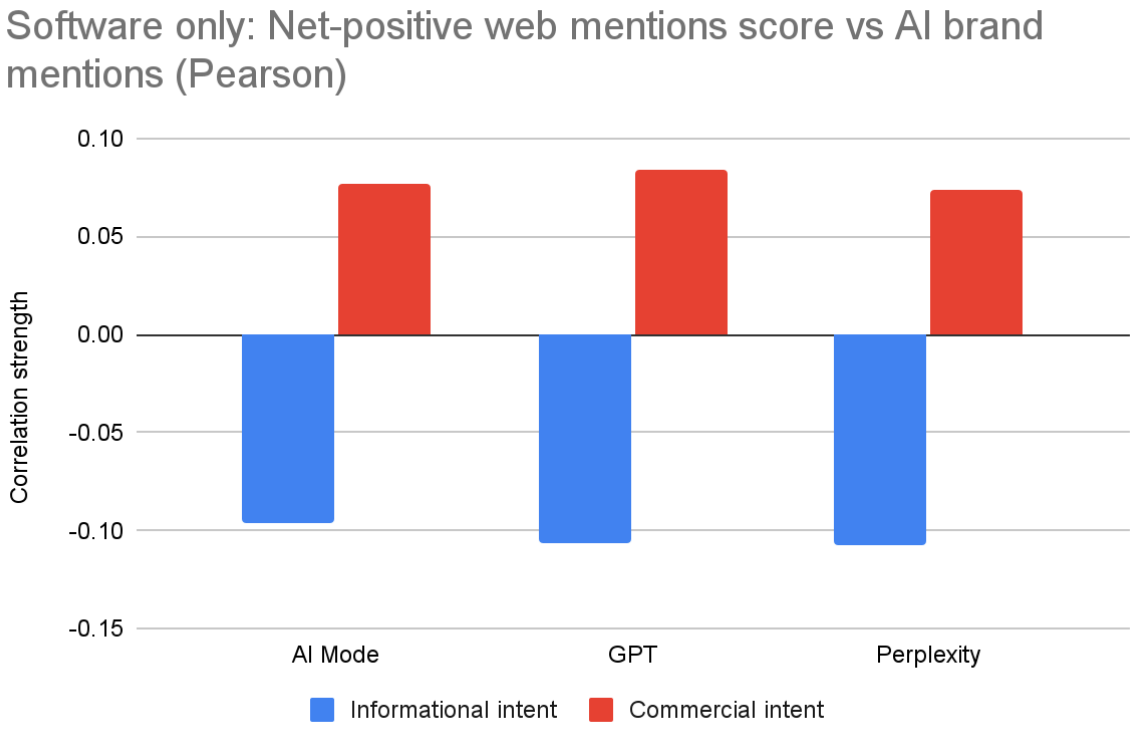
In Spearman, correlations are much stronger and consistent. **GPT** leads at **0.31** for informational intent and **0.254** for commercial. **AI Mode** and **Perplexity** cluster between **0.231** and **0.265**, all medium and stable.



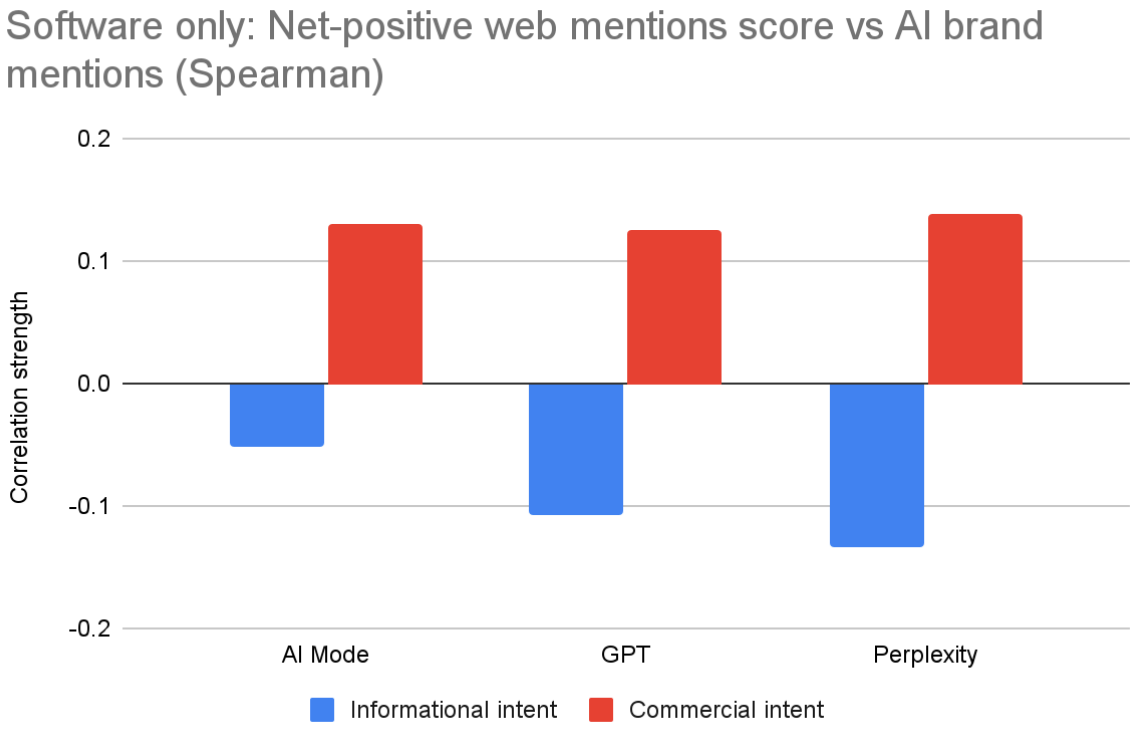
In Pearson, correlations are weak but consistent. Informational intent ranges from **0.148** to **0.164**, with **AI Mode** the highest at **0.164**. Commercial intent is a bit lower, between **0.114** and **0.133**. The spread is narrow and stable across models.



In Spearman, relationships strengthen considerably. **GPT** reaches **0.349** for info keywords, fairly strong, while **AI Mode** at **0.293** and **Perplexity** at **0.306** are also high. Commercial intent clusters around **0.21** to **0.237**, medium range.



In Pearson, informational intent turns clearly negative across all models, **AI Mode** at **minus 0.096**, **GPT** at **minus 0.106**, **Perplexity** at **minus 0.107**, all weak inverse relationships. Commercial intent is slightly positive but very weak, topping at **0.084** for **GPT.**



In Spearman, informational intent remains negative, from **minus 0.051** to minus **0.134**. Commercial intent shifts to weak positive values, between **0.125** and **0.139**.

Net positive mentions break the earlier pattern. One possible explanation is about what the sentiment actually refers to. For informational queries, the sentiment in a web mention often reflects the opinion of the website about another product or topic, not necessarily about the brand itself. Very positive language is common on affiliate sites or content farms, which may be seen as lower quality and therefore appear less often in AI responses.

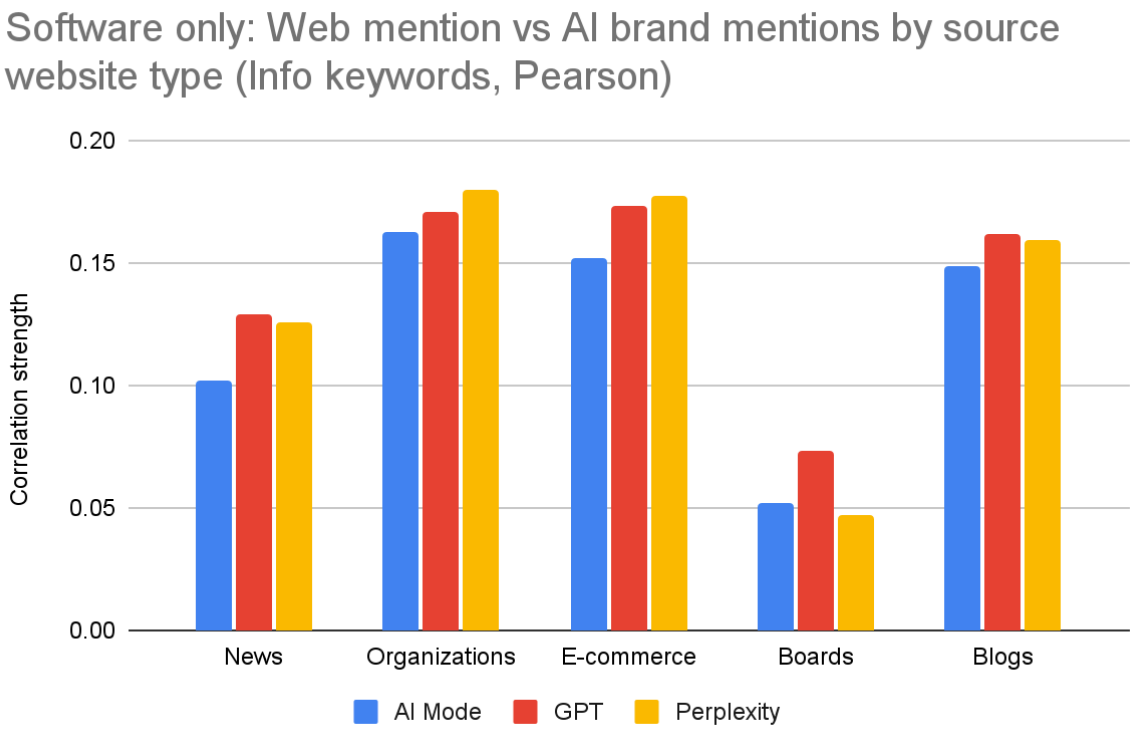
For commercial queries, the sentiment is more likely to refer directly to the product or brand. This makes the relationship more closely connected to real brand visibility in AI generated responses.

Key takeaways

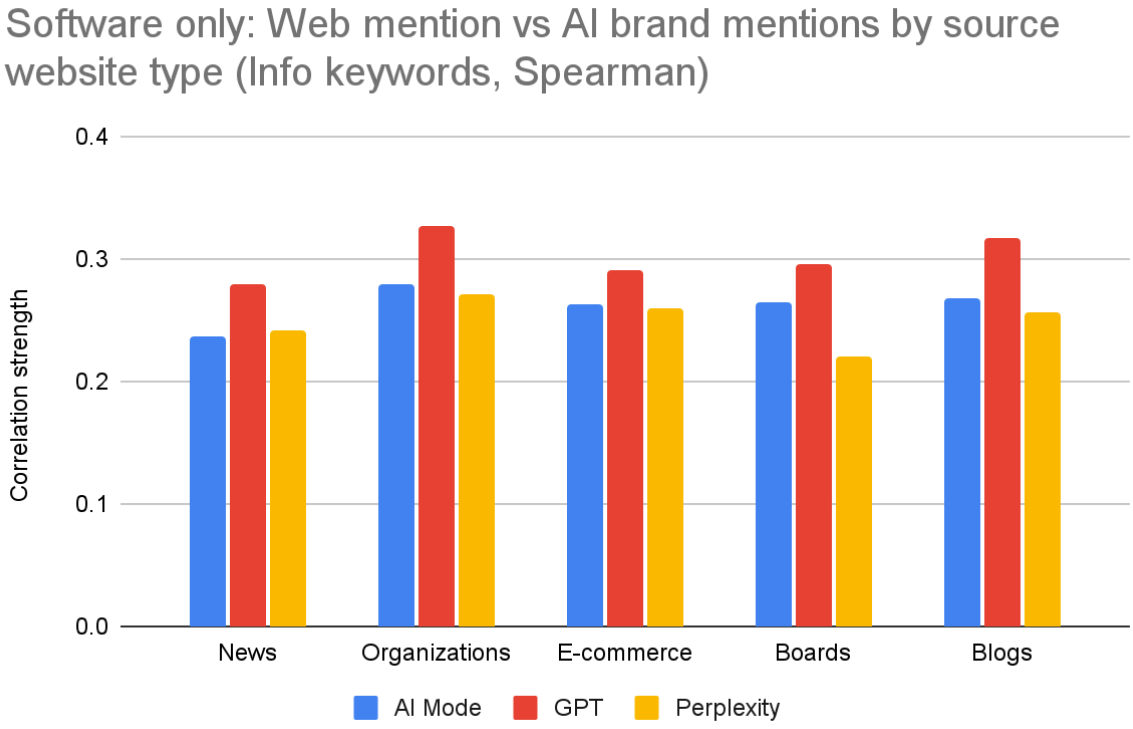
- Unlike citations, where mention-based factors have close to zero or unstable correlations, web mentions show consistent and meaningful relationships with LLM brand mentions, especially in order-based metrics like Spearman.
- In some cases, this relationship even appears stronger than the relationship between backlinks and visibility in traditional SERPs.
- The authority of the website making the mention does not seem to matter much. Pure mention count and mention score metrics show very similar correlation strengths.

Web mentions by website type

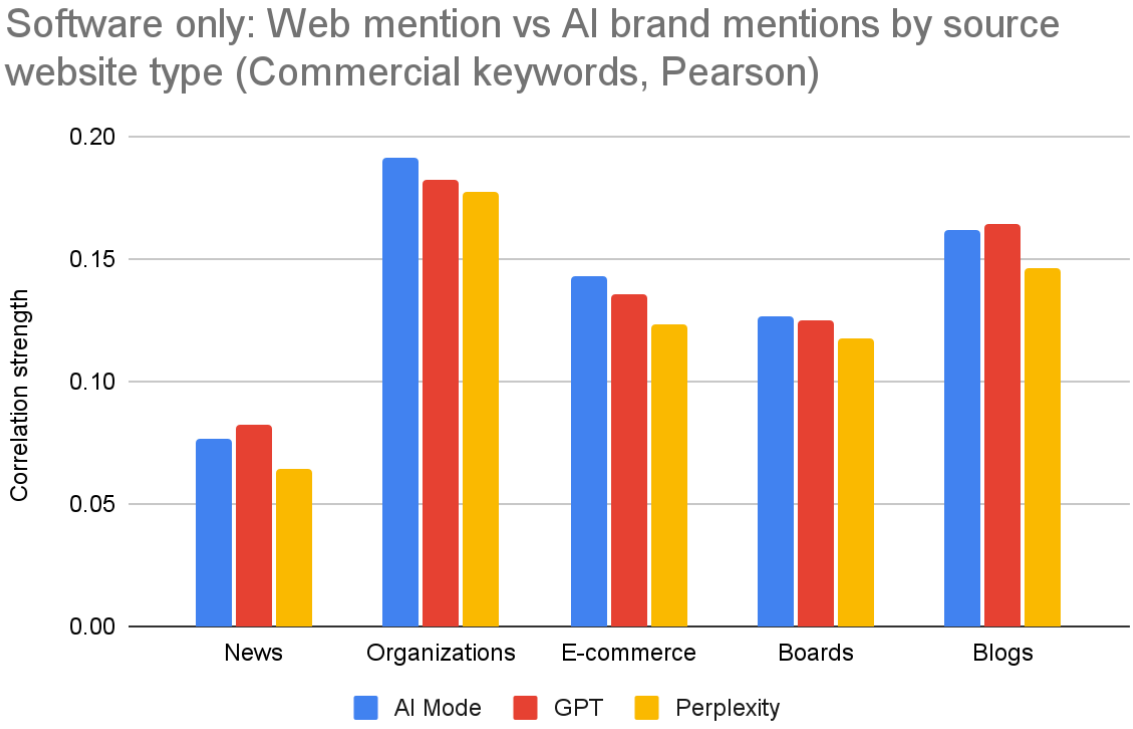
Let's examine web mentions by the type of website they come from. The goal is to see whether some website types contribute stronger signals than others. To classify website types, we used the categories provided by the DataForSEO.



In Pearson, correlations are mostly within the weak range. **Organizations** lead across models, reaching **0.191** for **AI Mode** and **0.182** for **GPT. Blogs** and **E-commerce** follow closely. **News** sit slightly lower but still within the weak range. **Boards** are clearly the weakest source, ranging from 0.047 to 0.073.

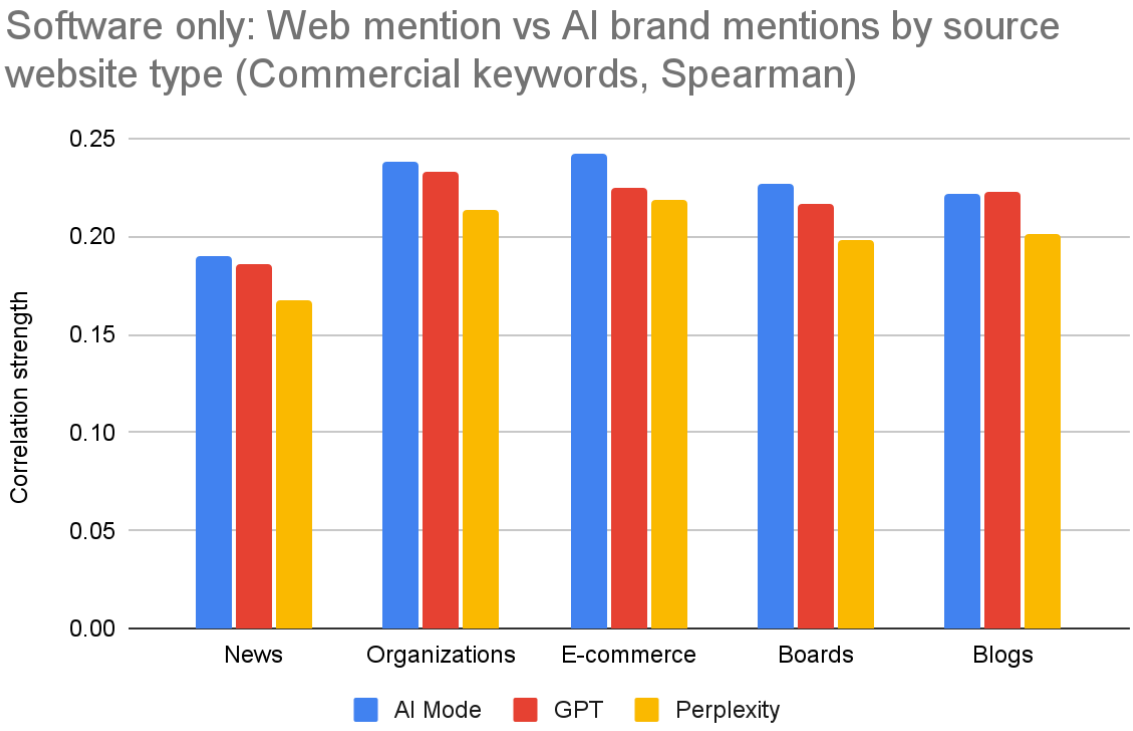


In Spearman, all placements strengthen and become more consistent. All website types are tightly clustered together, mostly within medium range.



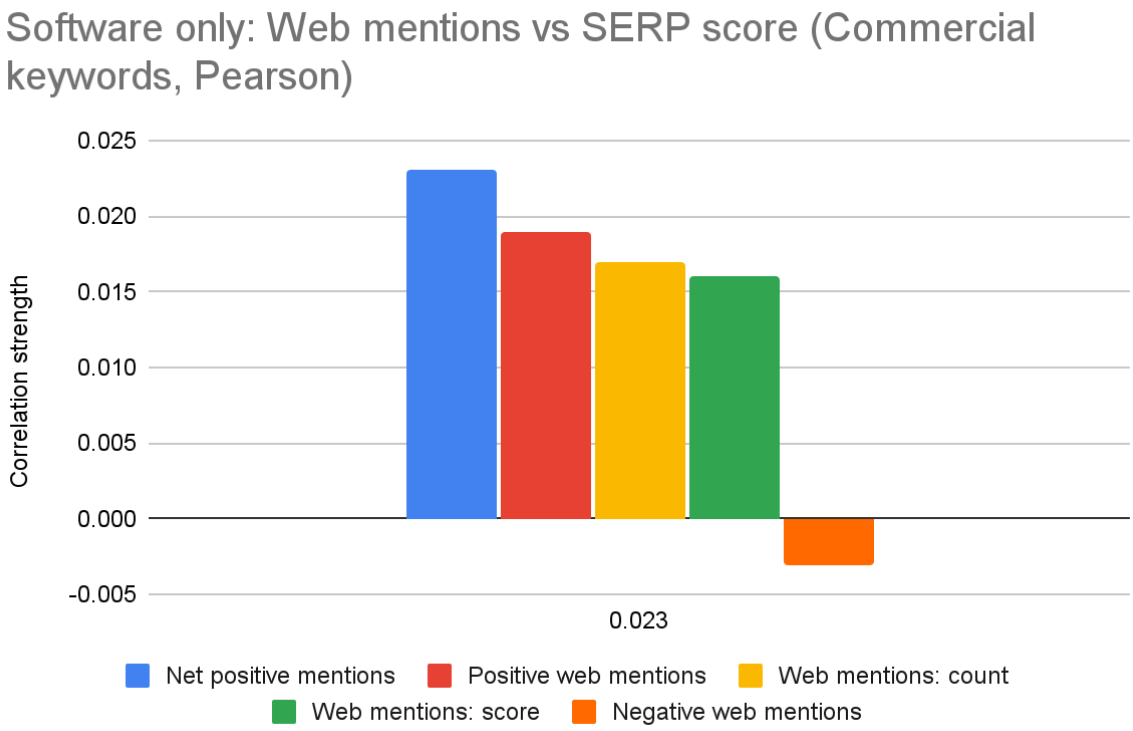
The commercial keyword dataset shows the same structure as the informational one.

In Pearson, **Organizations** are the strongest source, reaching **0.191** for **AI Mode** and **0.182** for **GPT**. **Blogs** follow closely, **0.146** to **0.164**. **E-commerce** and **Boards** are slightly weaker but still in the weak range. **News** type is the weakest, between 0.064 and 0.082, very weak.



In Spearman, correlations increase and cluster tightly. **E-commerce** at **0.242** and **Organizations** at **0.238** lead for **AI Mode**, both within medium range. GPT and Perplexity follow the same order. **News** stay the lowest but considerably higher than for Pearson: **0.167 to 0.190**.

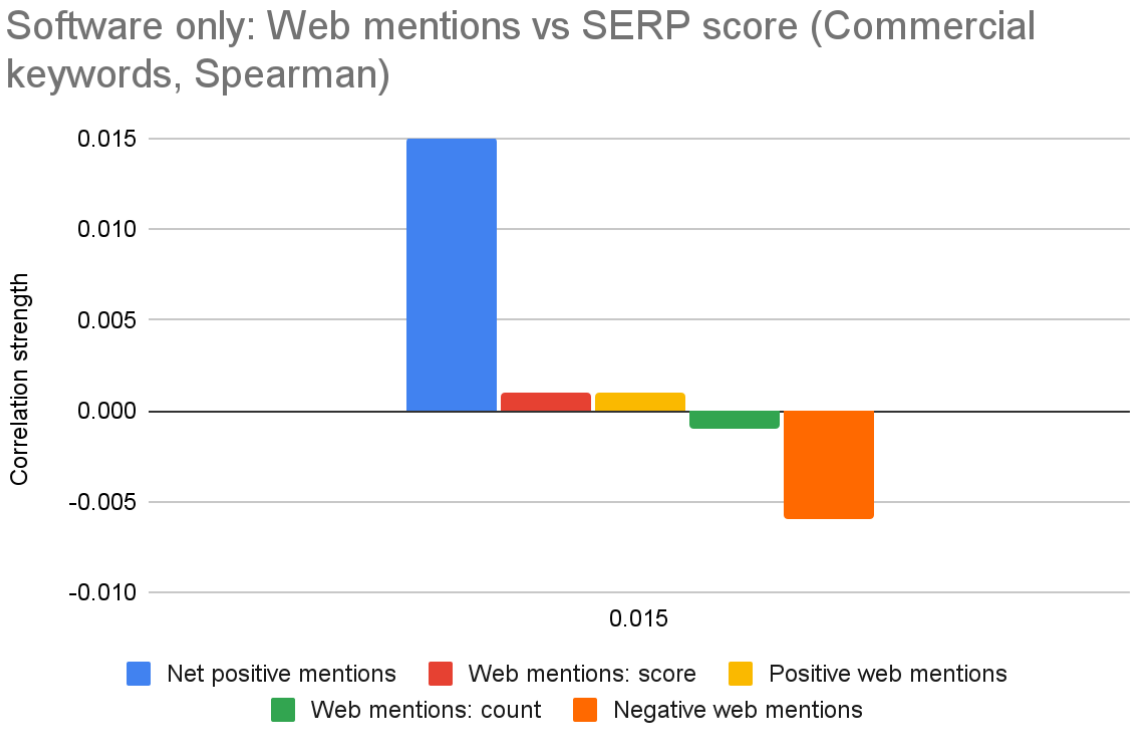
Web mentions & Traditional SERPs



For commercial keywords, correlations between web mention based factors and SERP Score are essentially nonexistent.

In Pearson, all values fall into almost no correlation.

Net positive mentions are the highest at **0.023**, followed by **Positive web mentions** at **0.019**, **Web mentions count** at **0.017**, and **Web mentions score** at 0.016. **Negative web mentions** are slightly negative at **minus 0.003**.

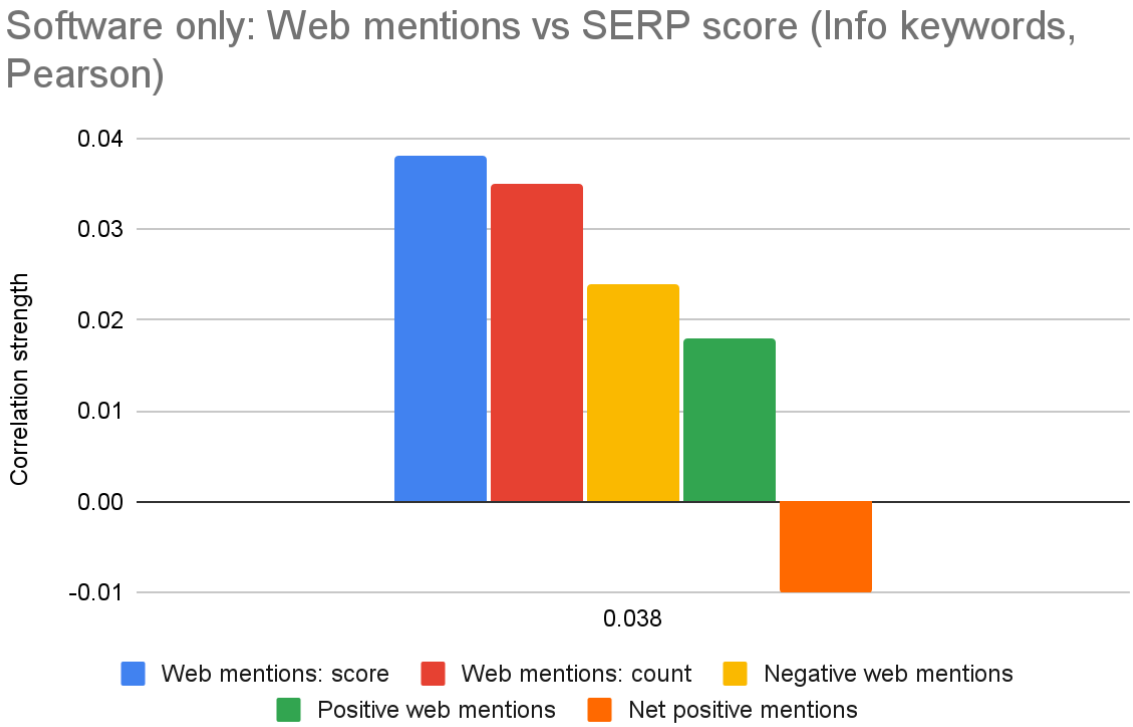


In Spearman, the pattern is even flatter. All values are near-zero strengths. The strongest one, **Net positive mentions** reach only **0.015**, completely negligible.

Overall, web mention signals show no meaningful relationship with SERP performance for commercial queries.

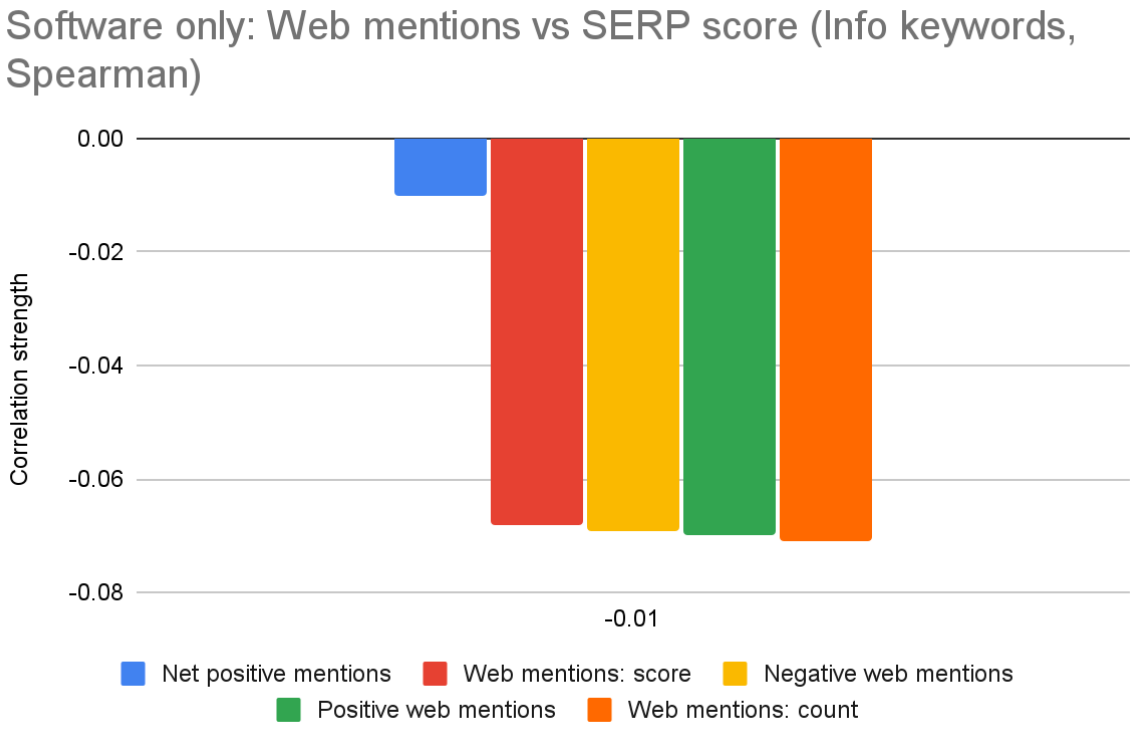
How about info queries?

Let's have a look:



For informational keywords, Pearson correlations remain extremely weak, though slightly higher than for commercial queries.

All fall into almost no correlation range.



In Spearman, the pattern shifts more clearly to negative. Values range from **minus 0.01** for **Net positive mentions** down to **minus 0.071** for **Web mentions count**. Unlike the commercial set, where Spearman hovered around zero, informational queries show extremely small but consistent inverse relationships.

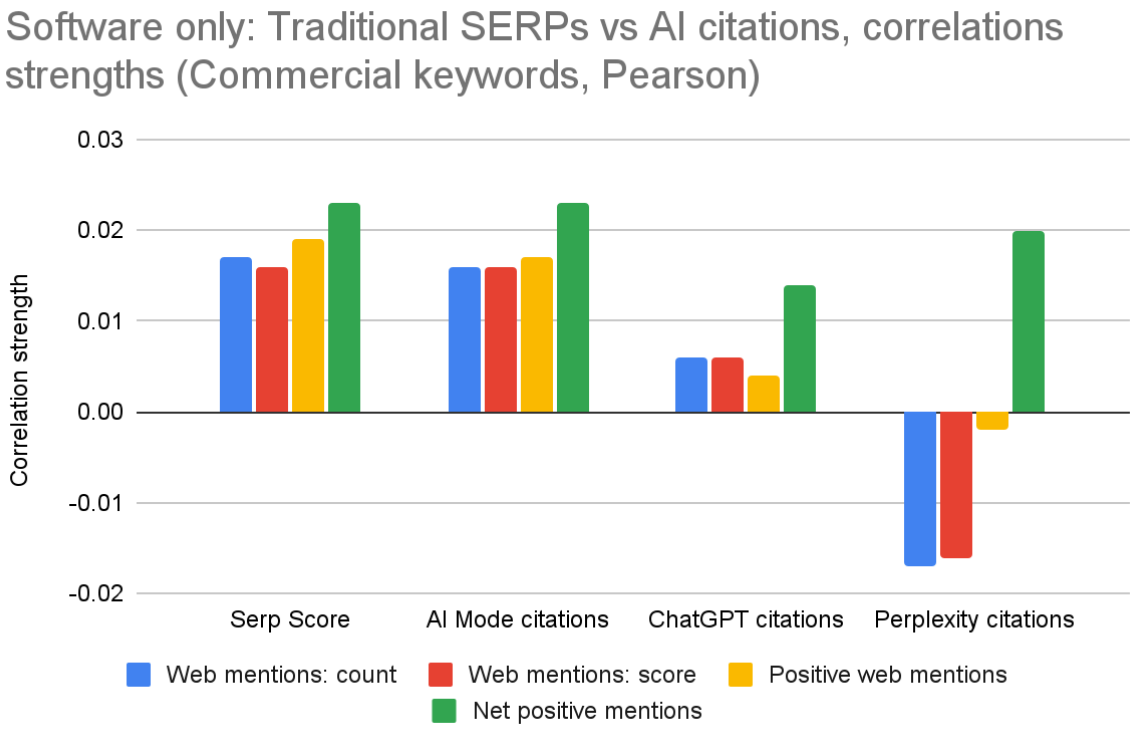
The takeaway

Overall, mention based factors still show no meaningful connection to SERP Score.

PART III: AI Search vs Traditional SERPs

In this chapter, we compare how different factors relate to visibility in AI search in comparison to traditional SERPs. Some of the data repeats earlier results, but we present it again to use traditional SERPs as a clear reference point.

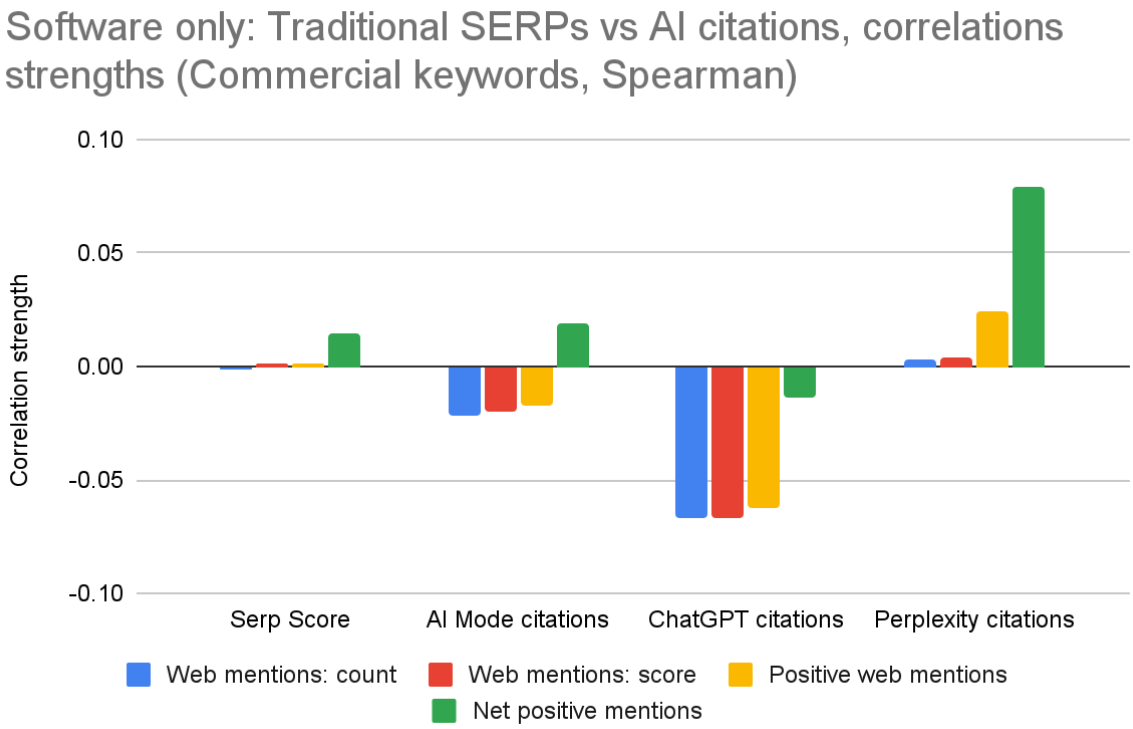
Unlinked web mentions: Traditional SERPs vs AI citations



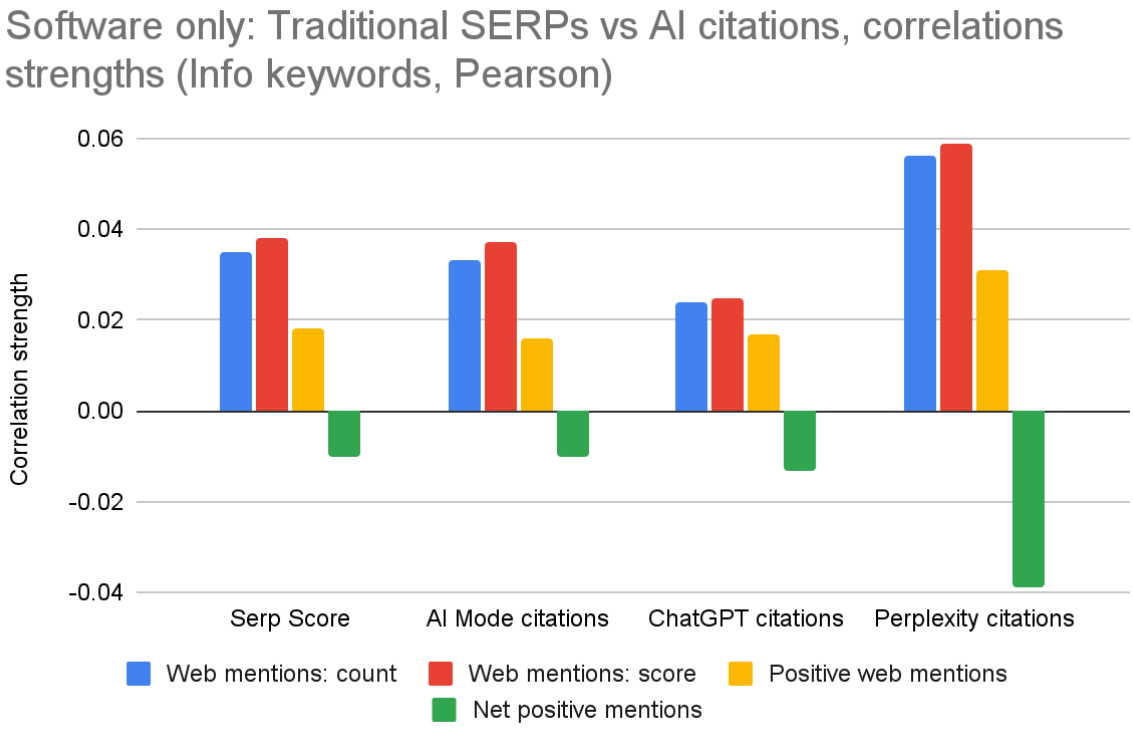
For commercial keywords, Pearson correlations are uniformly negligible.

For **SERP Score** and **AI Mode** citations, values range from **0.016** to **0.023**, all with almost no correlation. **ChatGPT** citations are even lower, between

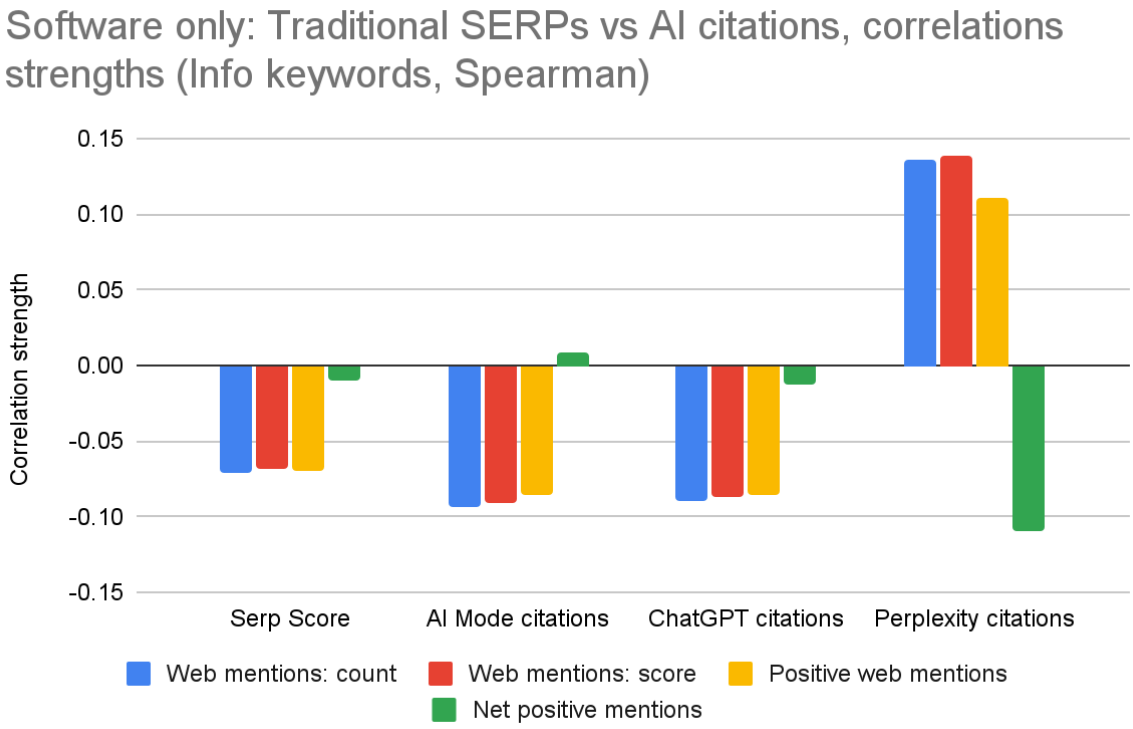
0.004 and **0.014. Perplexity** is slightly negative for Web mentions count and score at **minus 0.017** and **minus 0.016**, while **Net positive mentions** reach **0.02**.



In Spearman, patterns are still extremely weak but more differentiated. **SERP Score** stays around zero. **AI Mode** is slightly negative for count and score, **down to minus 0.022**. **ChatGPT** shows a clearer inverse pattern, reaching **minus 0.067. Perplexity** stands out modestly, with **Net positive mentions** at **0.079**, very weak but the highest value.



For informational keywords, Pearson correlations remain almost nonexistent, though slightly higher than in the commercial set. **SERP Score** peaks at **0.038** for **Web mentions score**, and **AI Mode citations** reach **0.037**, both almost no correlation. **ChatGPT** stays low around **0.024** to **0.025**. **Perplexity** is the highest, at **0.056** and **0.059**, still very weak. **Net positive mentions** are negative across all engines, down to **minus 0.039**.



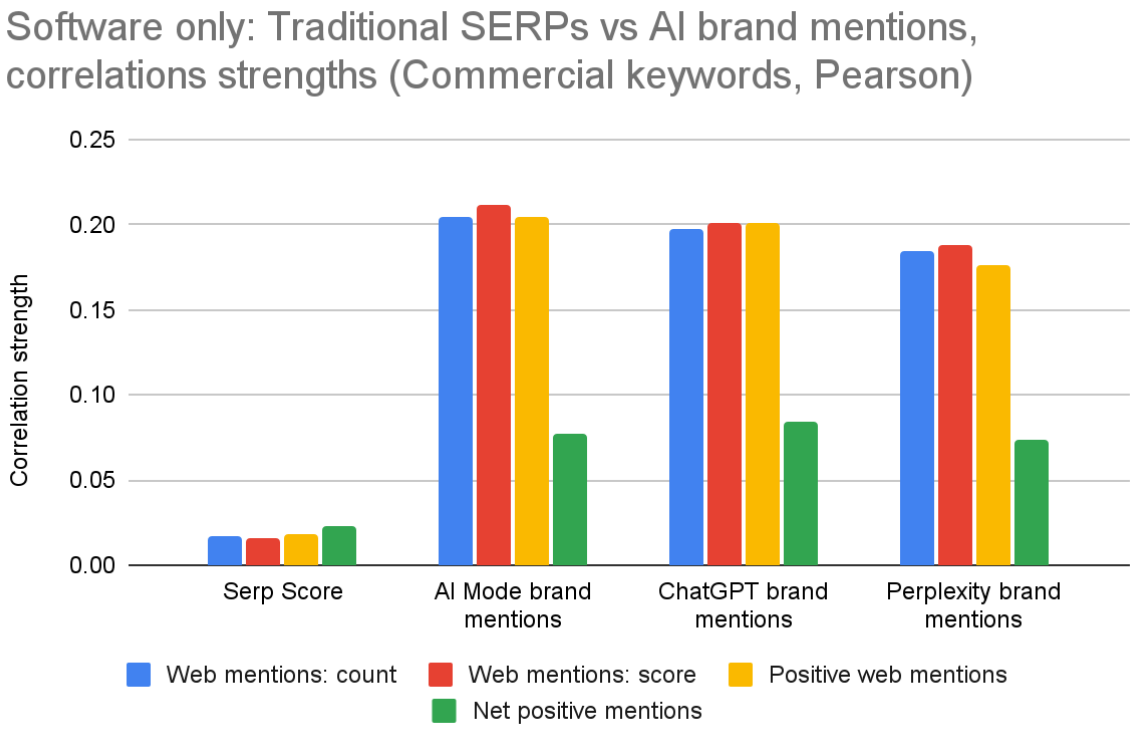
In Spearman, a bit clearer pattern appears. **SERP Score**, **AI Mode**, and **ChatGPT** show consistent very weak negative rank relationships, around **minus 0.07** to **minus 0.093. Perplexity** stands out positively, reaching **0.136** and **0.138** for count and score, weak but meaningfully higher.

The takeaway

Mention-based factors show no meaningful and stable relationship with either SERP visibility or AI citations. Please note that these weak negative correlations may occur due to the general non-linearity of the data shown in the “AI Citations by Domain Authority Ranges” chapter. This non-linearity can weaken existing positive correlations and may cause some results to appear slightly negative when there is no meaningful overall trend.

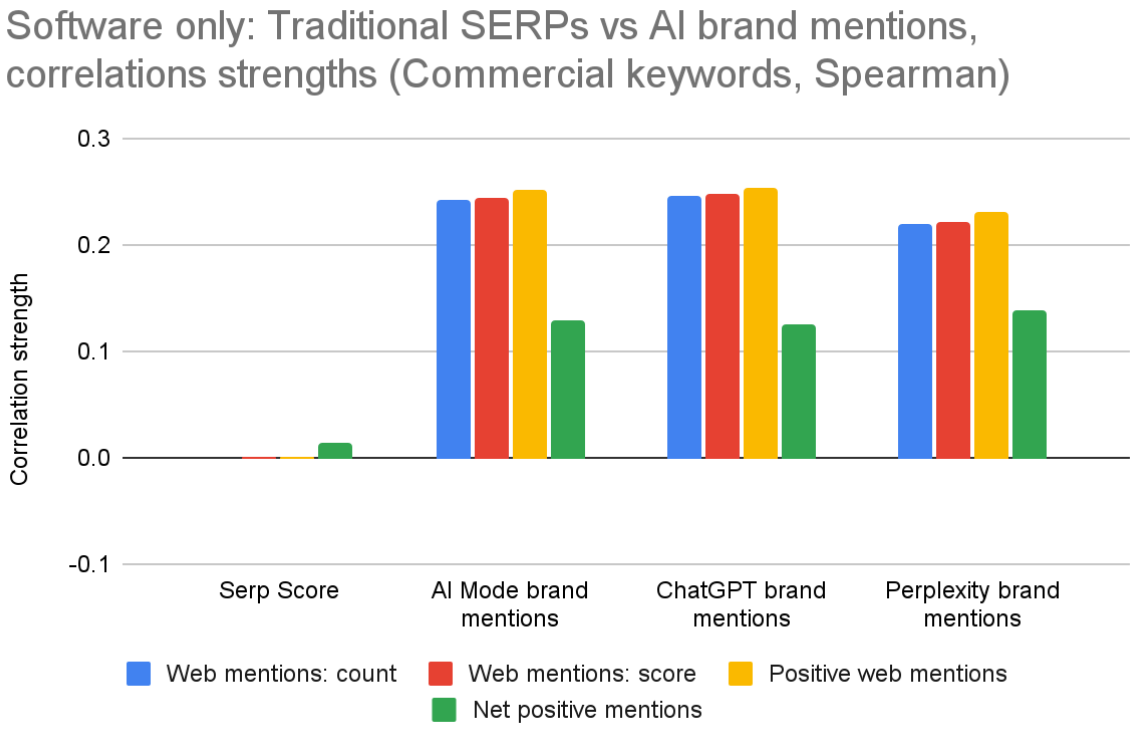
Traditional SERPs vs AI brand mentions

Here we compare how web-mention based factors affect brand mentions in LLMs in comparison to traditional SERPs.



For commercial keywords, the split between SERP and brand mentions is very clear.

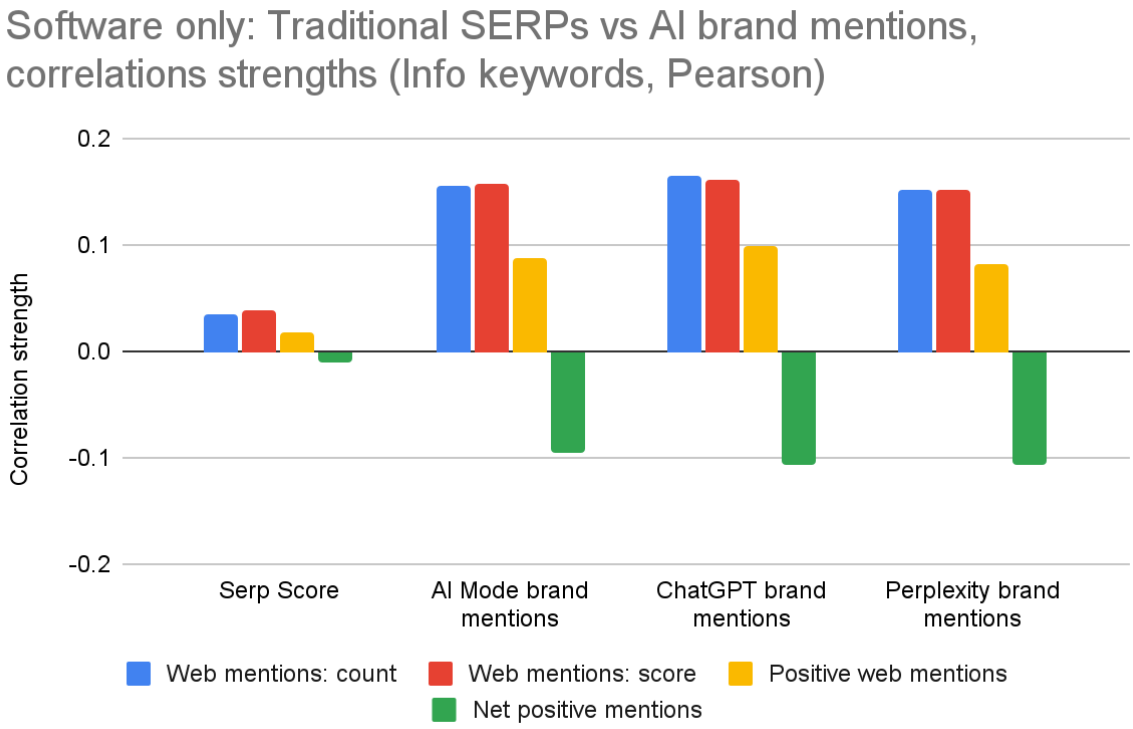
In Pearson, **SERP Score** remains near zero, between **0.016** and **0.023**. In contrast, **AI Mode** brand mentions reach **0.211** for **Web mentions score** and **0.205** for **Positive web mentions**, both medium. **ChatGPT** brand mentions are similar, up to **0.201**, and **Perplexity** peaks at **0.188**. **Net positive mentions** are much weaker, between **0.074** and **0.084**.



In Spearman, SERP again stays flat around zero. Brand mentions are clearly stronger, clustering between **0.22** and **0.254** for **count**, **score**, and **positive mentions**, medium range. **Net positive mentions** remain lower, around **0.125** to **0.139**.

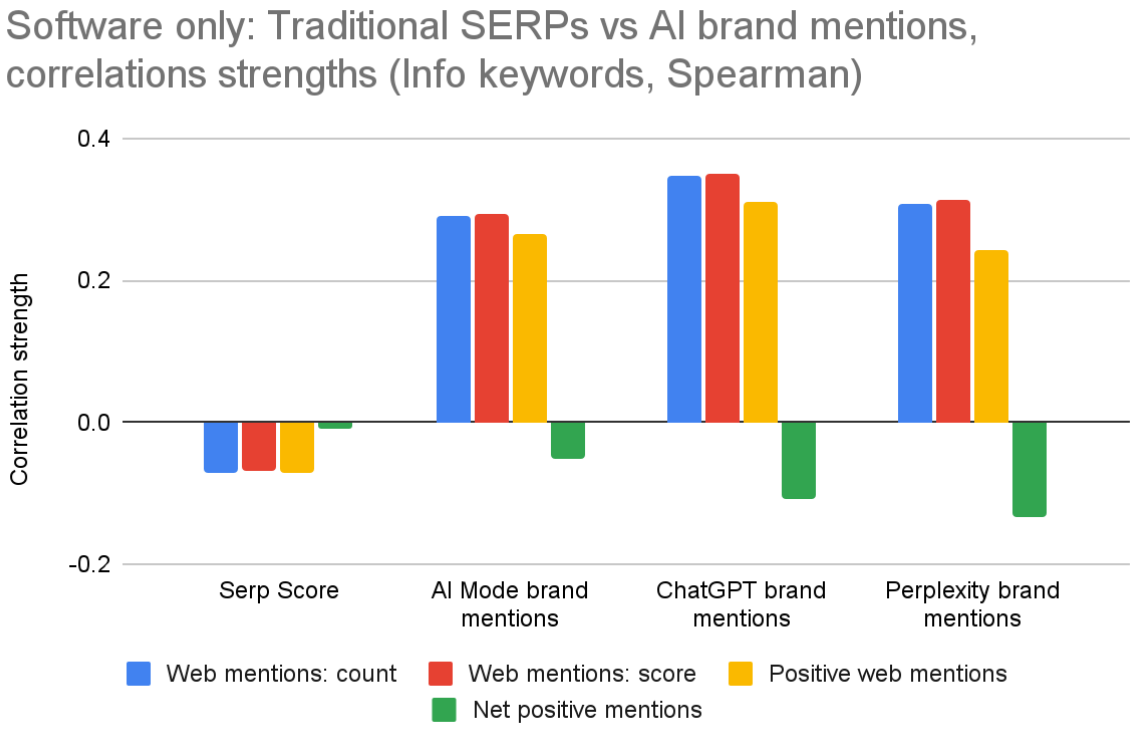
The takeaway

Compared to citation-based results, unlinked brand mentions show a stable and meaningful relationship with brand visibility in LLM responses. In contrast, traditional SERPs show no meaningful relationship with unlinked web mentions.



For informational keywords, SERP Score again shows almost no relationship.

Brand mentions tell a very different story. In Pearson, **Web mentions count** and **score** cluster between **0.151** and **0.164** across models, weak but clearly higher than SERP Score. **Positive web mentions** are lower, around **0.082** to **0.099**. **Net positive mentions** are negative for all models, down to **minus 0.107**.



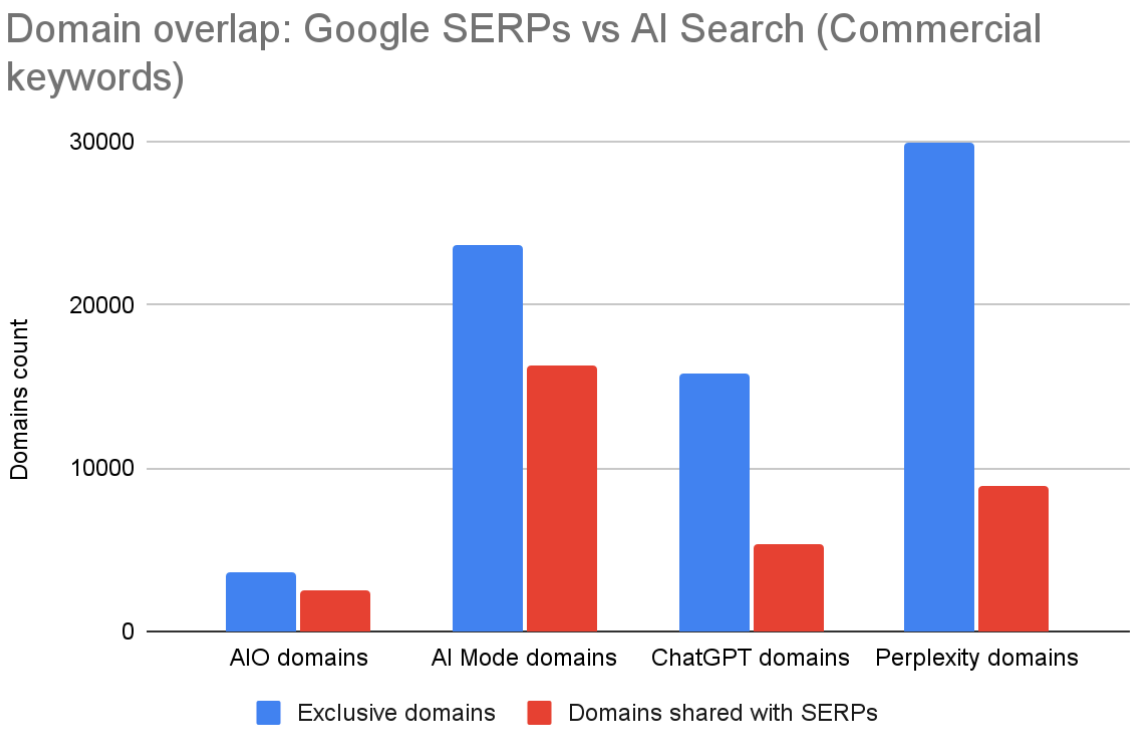
In Spearman, relationships strengthen considerably. **ChatGPT** reaches **0.349** for **Web mentions score**, fairly strong, with **Perplexity** at **0.313** and **AI Mode** at **0.293** close behind. **Positive mentions** remain slightly weaker. **Net positive mentions** are consistently negative, especially for **Perplexity** at **minus 0.134**.

The trend for the commercial set is similar but stronger. Overall web mention volume aligns with LLMs brand mentions, while sentiment impact remains unclear.

AI Search & Traditional SERPs overlap

The following charts illustrate the overlap between AI search results and traditional Google SERPs. For each AI model, we compare the total domains cited in AI-generated answers with the domains appearing in the classic SERPs. The charts show how many of the domains mentioned by the model

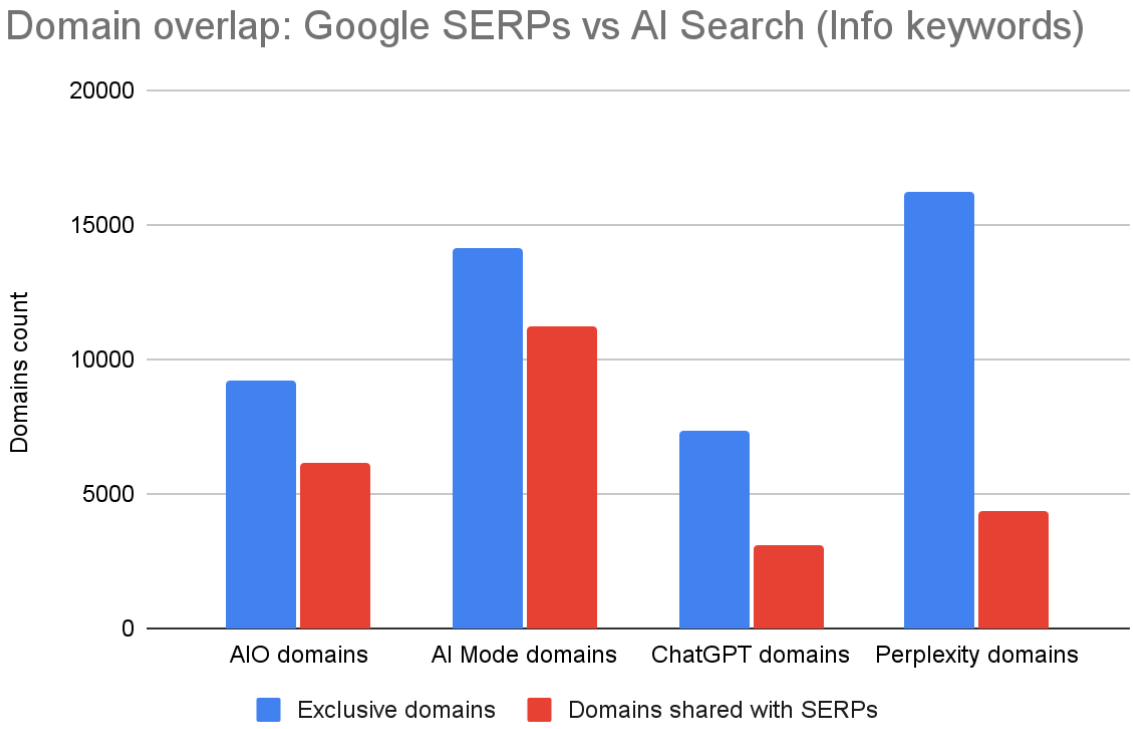
also appear in the traditional search results (red columns), and how many are unique to the AI response and do not appear in the SERP at all (blue columns).



AI Mode has **23 692** exclusive domains and **16 279** shared with **SERPs**. **Perplexity** shows an even larger divergence, with **29 972** exclusive domains versus **8 982** shared. **ChatGPT** also leans strongly independent, **15 848** exclusive and **5 322** shared.

Google AIO is the closest to **Google SERPs**, with **3 610** exclusive and **2 545** shared domains.

AI systems introduce a substantial share of domains outside traditional SERPs. Especially Perplexity and ChatGPT, while AIO and AI Mode remain a bit more aligned with classic blue links.



AI Mode again shows strong alignment with Google, with **11 210** shared domains versus **14 145** exclusive ones. **AIO** also remains relatively close, **6 165** shared and **9 239** exclusive.

ChatGPT and **Perplexity** rely more on independent sources. **ChatGPT** has **7 355** exclusive domains and only **3 080** shared. Perplexity shows the largest divergence, **16 220** exclusive versus **4 394** shared.

Compared to the commercial set, the same overall trend holds: Google AI systems remain closer to classic Google SERPs than non-Google AI platforms.

Part IV: Conclusions and Takeaways

What the data shows

- Traditional **SERPs align more closely with AI citations**, while **LLM brand mentions behave differently**.
- **Unlinked web mentions do not seem to matter for SERP rankings**, but they **do matter for brand recommendations in LLM answers**.
- **Mentions from very high-authority sites are not the only ones that matter**. The raw number of mentions and authority-weighted mention score show similar effects.
- The **"winner takes it all" rule only partially applies** in both SEO and AI search, leaving a lot of room for smaller and mid-authority sites to gain visibility.
- **ChatGPT appears to be the easiest environment to gain visibility for a new site**, which makes it a useful **early indicator of whether your strategy moves in the right direction**.

What we can do with this

Informational SEO is shifting from simply attracting traffic to your website toward **distributing content across the web so LLMs can reference it and send potential clients your way**.

People are now asking AI the same questions that your on-site evergreen and TOFU content used to answer. That doorway to your site is closing. Now and into the future, brand discovery happens through conversations on

forums, social media, YouTube, industry publications and blogs. Both humans and large language models (LLMs) read these spaces to assess your credibility, reputation, and authority. This is how they learn who you are and whether your brand can be trusted.

To stay visible, create authentic, high-quality content across third-party platforms. Encourage discussions, reviews, and mentions that accurately represent your brand. Doing so strengthens both human perception and how AI systems interpret your business.

At the same time, it is worth noting that this shift is far less important for traditional search with commercial intent. These search results have not changed to the same extent. Strong SEO fundamentals will likely continue to deliver reliable ROI and remain a solid foundation for growth.