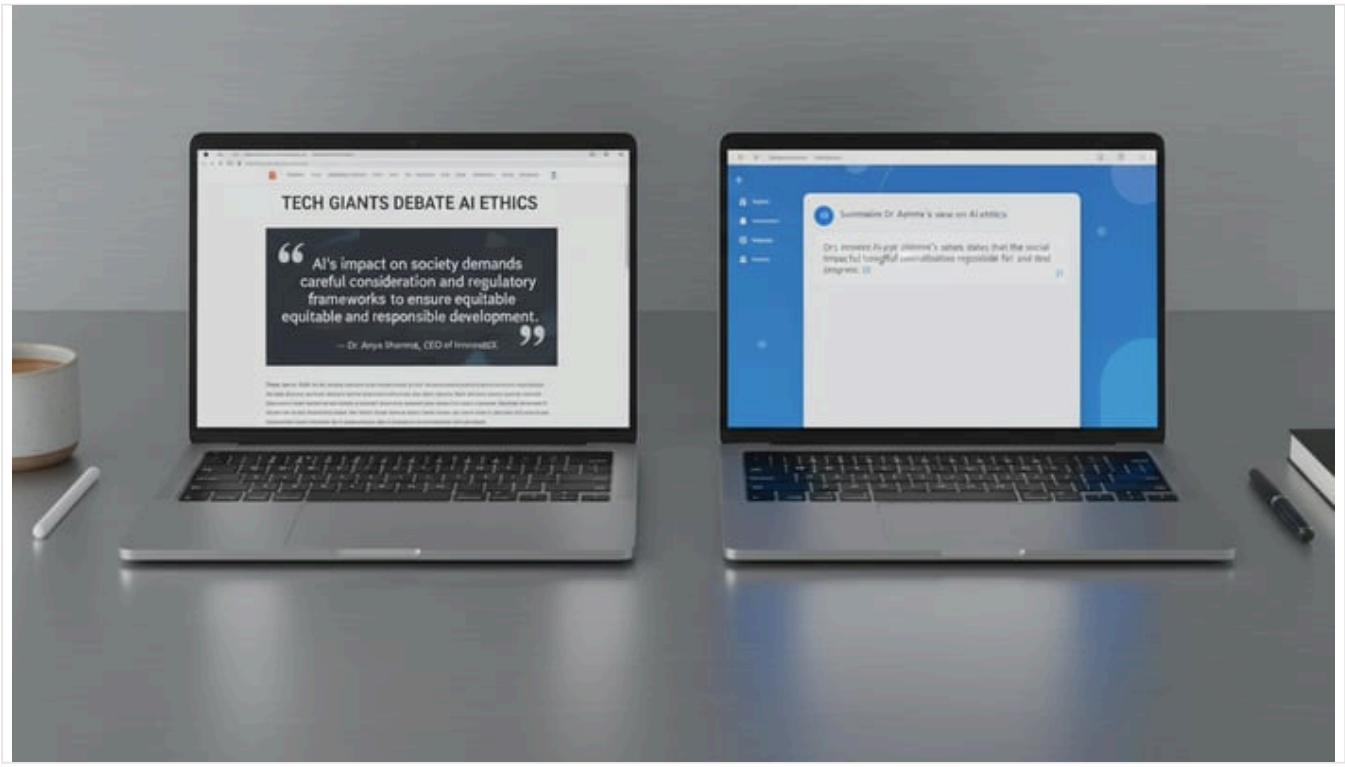
How Quote Attribution in News Impacts LLM Mentions & SEO

By rankstudio.net Published October 24, 2025 45 min read



Executive Summary

This report examines the interplay between **quote attribution in news reporting** and how content is surfaced or "mentioned" by large language models (LLMs) in AI-driven search and content-generation systems. We analyze this topic from multiple angles: journalism practice, information trust, and digital-marketing/ LLM-SEO strategy. By surveying academic studies, industry reports, and real-world examples, we show that how quotes are attributed in news stories can significantly influence both audience perceptions and the behavior of AI systems. Notably, properly attributed quotes from authoritative sources tend to enhance a news story's credibility and accuracy (Source: mediaengagement.org) (Source: pubmed.ncbi.nlm.nih.gov), which in turn can affect whether generative AI tools cite or incorporate that content. Conversely, misattributed or fabricated quotes can damage trust and lead LLMs to propagate misinformation (Source: www.searchenginejournal.com) (Source: futurism.com).

From the perspective of **AI-driven content retrieval** (sometimes called LLM "citations" or "mentions", clarity and context matter more than traditional SEO signals. LLMs rank content by factors like *clarity*, *contextual relevance*, and *"cite-worthiness"*, rather than by backlinks or domain authority (Source: mtsoln.com) (Source: mtsoln.com). As a result, news articles that include clear, self-contained quotes and factual attributions are more likely to be chunked and referenced by LLMs. We provide evidence that content structured into well-defined passages (as when quotes are clearly attributed and contextualized) is precisely what LLMs "retrieve, cite, or paraphrase" (Source: willmarlow.com) (Source: mtsoln.com). Conversely, content that lacks proper attribution may be ignored or misattributed by AI systems. For example, a study found ChatGPT-based search misquoted news in 76.5% of queries (Source: www.searchenginejournal.com), sometimes citing syndications instead of the original source (Source: www.searchenginejournal.com), illustrating the hazards when quotes aren't clearly tied to their true authors.

In summary, this report finds that **well-attributed news quotes** not only uphold journalistic standards of credibility, but also align with the structural preferences of LLM-based systems. By citing authoritative voices and structuring content clearly, news organizations and content creators can increase the likelihood that their material is correctly used and cited by AI models (Source: mediaengagement.org) (Source: mtsoln.com). The alternatives — vague or absent attributions — can both undermine human trust

and lead AI agents to “hallucinate” sources (Source: futurism.com) (Source: www.cjr.org). We discuss these phenomena in depth, providing data analyses, case studies, and expert commentary. Finally, we offer guidance on best practices and future directions, such as the need for new ethical guidelines in the AI era and SEO tactics emphasizing “being the source” for LLMs (Source: createandgrow.com) (Source: www.mdpi.com).

Introduction and Background

The advent of large language models (LLMs) like OpenAI’s ChatGPT, Google’s Bard/Gemini, and others has upended how people find information. Today, an increasing number of users rely on generative AI assistants instead of traditional search engines (Source: www.cjr.org). These systems synthesize information from their training data and external sources to answer questions in natural language. Crucially, **the content of news articles forms a major part of the knowledge base of LLMs**. Thus, journalistic content – including the way quotes are handled – can directly affect what these AI models “know” and "say."

At the same time, journalists continue to rely on quoting practices to convey authority and authenticity. In news reporting, **quoting and attributing sources is a fundamental practice**: putting statements in quotation marks and naming the speaker provides transparency and credibility (Source: mediaengagement.org) (Source: pubmed.ncbi.nlm.nih.gov). For example, when a public official makes a statement, a reporter will typically write: *“We will increase funding,” said Finance Minister Jane Doe.* The theory is that readers can then trust the information because a named individual is held accountable for the words. Conversely, if an outlet prints a statement without attribution (e.g. “They want to increase funding”), readers are unclear of its provenance, reducing trust.

Research has long illustrated the power of attribution. A classic study by Sundar (1998) found that explicitly attributing news quotes to credible sources significantly increases story credibility (Source: mediaengagement.org). More recently, a report from the Center for Media Engagement showed that Americans (of all political stripes) rate news posts that quote a public official as more believable than posts that provide no quotes or quote controversial figures (Source: mediaengagement.org). In particular, **news stories that included quotes from non-partisan officials were rated as most credible** by readers (Source: mediaengagement.org). Likewise, scientific studies of media coverage have found that including quotes from independent experts reduces exaggeration and slant (Source: pubmed.ncbi.nlm.nih.gov). In one analysis of health news, articles that featured a quote from an independent expert were 2.6 times *less* likely to exaggerate causal claims than those without (Source: pubmed.ncbi.nlm.nih.gov). These findings underscore that precise quoting and clear attribution in news not only follow journalistic ethics but measurably improve factual accuracy and trust.

On the **LLM and AI-search side**, a parallel emphasis on “sourcefulness” has emerged. Marketers and technologists now speak of content being “cited,” “mentioned,” or made “AI-citation friendly” for LLM-driven search. Unlike traditional search engines (which index entire pages and rank by links and authority, LLMs extract and compile **individual passages** of content to answer queries (Source: willmarlow.com) (Source: willmarlow.com). In effect, each clearly articulated sentence or paragraph can become the “unit” that the model pulls from a source.SEO experts note that LLMs reward *clarity, structured formatting, and comprehensive context* (Source: mtsoln.com) (Source: willmarlow.com). For instance, writing Q&A style sections, lists, tables, and succinct paragraphs makes it more likely that an AI assistant will copy or cite that snippet directly (Source: willmarlow.com) (Source: mtsoln.com).

One industry commentator summarizes: traditional SEO relies on backlinks and domain authority, but **LLM-driven search prioritizes content *clarity* and *context*** (Source: mtsoln.com) (Source: mtsoln.com). In fact, it has been noted that LLMs “ingest, chunk, summarize, and then rank information based on its internal coherence and direct applicability to a query” (Source: mtsoln.com). By this logic, a news article that uses clear, well-attributed quotes is inherently more “coherent” and “contextual” (each quote is a self-contained statement) than one with murky attributions. Therefore, **how quotes are presented in news may directly influence whether and how LLMs use that text** when generating answers or summaries.

There is now a nascent concept of **“LLM mentions” or “LLM citations”** in SEO. This refers to how often an AI model includes a particular source or brand in its answers. Early research indicates that LLMs do not simply cite what is most popular; instead they tend to cite content that is precise and highly aligned with the question (Source: mtsoln.com) (Source: mtsoln.com). For example, data suggest that a page with exact, specific answers and structured content (e.g. tables, lists, Q&As) will be “cite-worthy” – in other words, appear in LLM-generated responses (Source: mtsoln.com) (Source: willmarlow.com). Outside experts recommend integrating distinct phrases and thorough context, and even seeding content networks so that an entity is recognized as authoritative (Source: mtsoln.com) (Source: searchengineland.com). In practice, this means that if a news outlet quotes a new scientific study in detail (with source names and context), an LLM is more likely to pull from that quote when asked about the topic.

However, despite the potential for synergy, numerous recent failures highlight a **risk of misattribution and misinformation**. Investigations show that AI-powered search tools often "hallucinate" sources when handling news queries. For example, a Columbia Journalism Review study found that AI chatbots *fabricated links* and journalistic citations roughly half the time (Source: www.searchenginejournal.com) (Source: www.cjr.org). In real cases, ChatGPT has been seen inventing entire newspaper articles or wrongly attributing quotes, which raises major concerns about reliability (Source: futurism.com) (Source: www.cjr.org). These incidents point to a crucial point: if an LLM is prone to making up sources, then the quality of quote attribution in its underlying news data is paramount. Failing to cite properly in journalism not only breaks trust with readers, but can feed into the AI ecosystem and worsen its hallucinations (Source: futurism.com) (Source: www.cjr.org).

The rest of this report delves deeply into these issues. We first elaborate on the role of quotations in journalistic practice and public perception. Next, we define "LLM mentions" and survey how AI models retrieve news content. We then analyze how news quoting patterns affect LLM behavior, using empirical data and case studies (including the ChatGPT Search misquoting episodes). We also present SEO-related findings (including tables comparing SEO vs LLM factors). Finally, we discuss the broader implications for media, technology, and society, and outline future directions. Throughout, we draw on academic studies, technical reports, and real-word examples to provide evidence-based insights.

Quote Attribution in Journalism

News organizations universally recognize that **accurate attribution is fundamental to credible reporting**. Mainstream style guides (AP, Reuters, etc.) insist that any person's statement must be either put in quotes or clearly paraphrased with source identification. No statement of fact should appear as a quote without naming its speaker, nor should a speaker's words be reused without proper context. Good attribution allows readers to evaluate both the content and the authority of a quote, reducing ambiguity or deception (Source: mediaengagement.org) (Source: pubmed.ncbi.nlm.nih.gov).

The Purpose and Practice of Quoting

A well-chosen quote can bring vitality and specificity to a story. Journalism educators stress that *direct quotes* (the exact words of a speaker) should be used "if exact language is needed for clarity" or "to demonstrate the speaker's personality or originality" (Source: socialsci.libretexts.org). Direct quotes often make facts more compelling; for example:

- *Example:* A finance minister's precise wording ("We will not waver on austerity") carries more weight than a paraphrase ("The minister promised continued austerity").

Quotes also serve as **evidence** of claims. When a story says *"Company X's CEO called the market conditions 'the worst since 2008,'"* the quotation marks signal that those are the CEO's words, not the reporter's. This helps maintain objectivity: the reporter isn't claiming it was the worst market, only reporting what the CEO said. Attribution (naming "Company X's CEO") provides accountability.

In newsrooms, pervasive standards exist about how to attribute. Typically, the **LQTQ format** ("Lead, Quote, Trailing Quote") is taught: introduce context, include the quote, then tag it with the speaker's name and title (Source: socialsci.libretexts.org). Quotes should be verbatim and accompanied by context when needed. Styles caution, for example, against starting with "he said" each time, to maintain readability (Source: slideplayer.com). Advanced guidelines even specify where introduction of the speaker's name is placed for maximum clarity (often at the end of the quote).

This rigorous approach underscores that **who said something is often as important as what was said**. Several credibility studies back this up. In the Center for Media Engagement experiment, the single most credible stories were those quoting an impartial official (Source: mediaengagement.org). Readers perceived stories as *more authentic* when they knew an authoritative, on-the-record person had spoken. Conversely, "outhouse quotes" (unnamed or anonymous sources) tend to raise suspicion. Indeed, social science research has repeatedly found that clear source attribution increases trust in online news (Source: mediaengagement.org). (For example, Sundar's classic work showed source cues affect how adults perceive accuracy.)

Quotes, Bias, and Balance

Quote selection can still shape a narrative. Over-quoting one party or ignoring context can introduce bias. The Media Engagement study found even the presence or absence of a specific quote affected perceived bias along political lines (Source: mediaengagement.org) (Source: mediaengagement.org). When news posts only quoted a Republican lawmaker, readers (both Democrats and Republicans) judged it biased to the right, and vice versa (Source: mediaengagement.org). Balanced stories quoting both sides were seen as far less propagandistic. This implies newsrooms must "vary storytelling approaches" carefully (Source: mediaengagement.org). If quotes are cherry-picked, even proper attribution doesn't immunize content from perceived partiality.

In practice, though, news articles often rely heavily on a few sources. One content analysis of health news showed nearly 100% of press releases had quotes (usually from the study authors), while 70–90% of follow-up news pieces quoted those same press releases (Source: pubmed.ncbi.nlm.nih.gov). However, only about 7–8% of those news pieces introduced *new* expert voices outside the release. In other words, most news quoting just echoed original quotes (Source: pubmed.ncbi.nlm.nih.gov). This *attribution bay* practice (using others' quotes) is common in journalism but can limit diversity of perspective. The Bossema study found that news without an external expert quote was 2.6 times more likely to exaggerate scientific claims than those that did (Source: pubmed.ncbi.nlm.nih.gov), implying that relying solely on press-release language (or quoting only the organization's own representatives) increases distortion. Hence, when journalists do add independent quotes, stories become more grounded.

Credibility, Misinformation, and Legal Stakes

Beyond bias, improper quoting can accidentally turn true stories into misinformation. Misquotations have appeared in media for decades, sometimes due to carelessness. In one notorious case, an over-simplification by a single outlet was picked up by many others, spreading a false "misquote" through the news ecosystem (see Misbar example). Journalistic ombudsmen and public editors often emphasize that even small attribution errors damage outlet credibility.

From a legal standpoint, false attribution or defamatory quotes can trigger lawsuits. If a quote is wrongly attributed and harms someone's reputation, the outlet can be liable. This legal peril encourages precise attributions. But even without malice, quoting out-of-context can distort meaning. Journalists know that verbatim quotes give the subject a chance to speak, but also carry risk: the speaker's words are immortalized. Experienced reporters balance this by verifying quotes (e.g., checking recordings) and by sometimes paraphrasing difficult claims with clarification instead of quoting them directly.

All told, in the **pre-AI context**, news quote attribution exists primarily to assure readers of accuracy and fairness (Source: mediaengagement.org) (Source: pubmed.ncbi.nlm.nih.gov). Communities trust journalists to present quotes faithfully and attribute them correctly. The rise of AI adds a new dimension: now quotes are also interpreted by algorithms. The next sections explore how this traditional journalistic practice impacts and is impacted by large language models.

Generative AI Search and LLM "Mentions"

To discuss the "impact on LLM mentions," we first need to clarify what that means. In contrast to traditional keyword-based search, **LLM-powered search** refers to systems where the answer is generated by a language model (like GPT-4, Claude, Gemini, etc.) rather than simply retrieving and ranking webpages. Major platforms such as Google AI Overviews, OpenAI's ChatGPT (with browsing/plugins), and Perplexity.ai exemplify this shift. These tools craft conversational responses, often with a short summary and citations (if available) to sources. Importantly, they frequently *draw on news content*, since news articles are rich factual sources.

In the emerging terminology of SEO professionals, having your "spill the tea about your brand" in an AI answer is known as earning an **LLM mention or citation** (Source: createandgrow.com) (Source: searchengineland.com). This is often equated to being one of the sources an AI cites in its reply. Unlike classic SEO where the metric is click-through or ranking on page 1, in the LLM world the analogous metric is having your content *cited* or *mentioned* by the AI answer, which may not even generate an outbound click. For instance, a company might measure success by how often an LLM like ChatGPT references its product details in answers, irrespective of user clicks (Source: willmarlow.com).

What determines whether an LLM "mentions" a piece of content? Industry consensus is still forming, but early patterns are apparent. LLMs do not simply trust the most linked or popular pages (Source: mtsoln.com) (Source: mtsoln.com). Instead, *semantic fit* and *clarity* are king. Mercury Tech Solutions emphasizes that **LLMs prioritize content that is clear, contextually on-point, and formatted for easy extraction** (Source: mtsoln.com) (Source: willmarlow.com). For example, LLMs prefer to pull from

content that directly answers a likely question, with minimal filler. Structured layouts (bullet points, FAQs, data tables, etc.) are favored because each segment can stand alone (Source: willmarlow.com). Indeed, one guide advises writers to design each paragraph as a potential self-contained answer for an LLM (Source: willmarlow.com): "Each paragraph is a potential LLM result," meaning if a quote with attribution occupies a single coherent paragraph, it can be directly retrieved by the model (Source: willmarlow.com).

Furthermore, experts recommend building a **"semantic footprint"**: ensure that your topic and brand co-occur across authoritative contexts (Source: searchengineland.com) (Source: searchengineland.com). In plain terms, if reputable news and industry sites frequently mention your brand or content together with relevant keywords, an LLM's internal connectivity algorithms will more readily associate them. This is reflected in the notion of *co-occurrence*. As Search Engine Land notes, when two terms appear together across many texts, their semantic connection strengthens (Source: searchengineland.com). For example, if an LLM is learning about electric vehicles and news articles consistently quote *"Company X's CEO"* discussing EV policy, the model may begin to link Company X with EV context. Thus, a news quote that names your company in a certain context can literally help "teach" the LLM about your relevance in that domain.

Crucially, LLMs aim for accuracy. Even if they have knowledge, many models will attempt to backup facts with citations if that feature is enabled. However, as we discuss later, this process can go awry. Recent academic work has flagged that many ChatGPT references are unreliable (Source: rankstudio.net). In the SEO realm, practitioners note that **getting an AI citation is not guaranteed by just content popularity**; the content must align extremely well with typical user intents (Source: mtsoln.com) (Source: willmarlow.com). This means that to maximize "mentions," a brand or author might strive to have easily retrievable statements quoted in content that directly matches expected queries.

In short, **LLM mentions** are the emerging metric for visibility in AI age. They reward the same clarity and credibility that traditional attribution aims for, but through the lens of algorithmic extraction (Source: mtsoln.com) (Source: willmarlow.com). The next section explores how these two worlds intersect: specifically, how news quotes act as fodder or pitfalls for LLM retrieval.

Interaction of News Quotes with LLM Outputs

We now analyze the core question: *How does the way news articles attribute quotes affect how LLMs will mention or use that content?* This interplay involves several dynamics:

- *Encoding into training data.* LLMs are often trained on broad web crawls of news. How quotes appear in those sources can influence what the model "remembers."
- *On-demand retrieval.* Some LLM systems (e.g. ChatGPT with browsing, or Google Bard/AI Overview) query live sources. These rely on being able to *find* and then *link to* the original content based on the user's query.
- *Citation and summarization.* When an LLM produces an answer, it may either quote verbatim from a source or summarize/paraphrase. At each stage, the presence of explicit quotes and attributions shapes its behavior.

We discuss each in turn.

LLM Training and Inherent Biases

During large-scale pretraining, models ingest enormous text, including news. Studies on AI hallucinations show that the models store factual patterns but can confuse details. If a quote in the news is misattributed or lacks clarity, the model might internalize incorrect associations. For example, if many news stories copy a quote without naming who said it, the LLM might remember the quote but not know its origin. Later, when asked, it might guess a source incorrectly or say "an analyst at Bank of X." This was seen in multiple AI hallucination anecdotes: ChatGPT often "kept making up quotes" or assigning quotes to the wrong person (Source: www.financialexpress.com). Such studies (and media reports) highlight that any ambiguity in news attribution can lead LLMs to guess wrongly.

Conversely, well-attributed quotes give the model a chance to learn the association. If many articles quote *"It was unprecedented," said Economist A. Smith* about inflation, the LLM can latch onto that fixed expression and link it to that speaker. Consistent attributions in training data reinforce correct mappings. In theorem, then, better reliability in quoting could yield more accurate generative recall. (Alas, formal evidence for this is limited, but plausibly: humans need repetition to learn, and LLMs similarly may treat co-occurrence statistics of quote-phrase and name as a "signal".) However, one must note that most mainstream LLM training

is not citation-aware. The model itself does not natively store metadata indicating "this sentence came from NewsOutlet on date." Without RAG (retrieval augmentation), the model's interior weights diffuse all training content. Consequently, even with good attribution in the data, the model may still hallucinate if it cannot pinpoint a source. This points to another phenomenon: **misaligned attributions**.

Misattributions in AI Answers

Real-world tests reveal how easily LLMs miscite news content. For instance, a Tow Center (Columbia) experiment asked ChatGPT-based systems to identify sources of quotes. In 200 queries, ChatGPT Search got 153 wrong (Source: www.searchenginejournal.com). It conflated quotes, cited syndications, or skipped naming the correct outlet. In one example, when queried for the origin of a quote from *The New York Times*, ChatGPT Search incorrectly offered a link to a copy on another site (Source: www.searchenginejournal.com). Even for *MIT Technology Review* (which allowed crawling), it chose a syndicated version rather than the official page (Source: www.searchenginejournal.com). In other words, **even when quotes are correctly attributed in a news article, the generative system might fail to point back to that source**, often citing alternate or unofficial versions. The study concluded that publishers have "no guarantee" their content will be properly cited by these AI tools (Source: www.searchenginejournal.com), regardless of robot.txt settings.

Another CJR report expanded on this: it tested eight generative "AI search engines" and found similar problems. Across 1,600 quote-excerpt queries, the chatbots got over 60% completely wrong (Source: www.cjr.org). They frequently "fabricated links" and cited syndicated copies (Source: www.cjr.org). Moreover, these systems rarely if ever hedged their answers; they answered with high confidence even when incorrect (Source: www.cjr.org). For example, while a Google-trained Gemini or ChatGPT might claim "This quote is from Article X at NewsSite.com," in fact it might just be fabricating. These failures occur even for content that was demonstrably in their training or index.

Thus, merely having an article with quotes—even if correctly attributed—does *not* guarantee that an LLM answer will credit it properly. In their current forms, LLM search tools often *override* or ignore existing attributions. This underscores that **LLMs are not infallible and will misrepresent quotes if the system is not explicitly designed to preserve them** (Source: futurism.com) (Source: www.searchenginejournal.com).

Case Studies: AI Hallucinations and Misquotes

To illustrate these issues, we highlight several notable examples:

- **Guardian vs ChatGPT**: In mid-2023, *The Guardian* found that ChatGPT had invented full articles supposedly written by Guardian journalists. The AI assistant was dumping *"sources"* and quotes from articles that had never been published (Source: futurism.com). In effect, it misquoted by fabricating the existence of content. The Guardian's innovation chief warned that such invented attributions could "undermine legitimate news sources" (Source: futurism.com). This case shows the ultimate breakdown: if an LLM has no real quote to anchor, it will conjure one out of thin air, possibly citing a journalist who wrote nothing. The core problem was not a misattribution of an existing quote, but the creation of a fictional quote and author.

- **ChatGPT's Expert Conspiracy**: Another example involved ChatGPT answering a query about podcaster Lex Fridman. The model confidently claimed AI researcher Kate Crawford had criticized Fridman, even generating "links" and "citations" to support this claim (Source: futurism.com). Crawford had in fact never made those statements. In short, an unlabeled quote ("Crawford said…") was attributed to her. This made-up quote was effectively a harmful misattribution. It demonstrates that when LLMs lack data on a topic, they not only hallucinate facts but also invent attributions.

- **USA Today and Fabricated Studies**: Similarly, USA Today reporters encountered ChatGPT inventing entire research citations on gun control. When asked for evidence that gun access does not increase child mortality, ChatGPT listed full study titles, authors, and journals – none of which existed (Source: futurism.com). The quotes from those papers were entirely imaginary. Here, incorrect "quote attribution" took the form of phantom academic quotations. A news outlet had quoted nothing (because the studies were fake), but ChatGPT answered as if real quotes were in play.

- **Columbia/CJR Experiment**: The Tow Center's controlled study mentioned above goes beyond anecdote. It systematically showed multiple AI tools *frequently mis-cite news quotes*. The metric is telling: for 1,600 random quotes, over 60% of responses were incorrect (Source: www.cjr.org). Even models that do retrieve from the web (RAG-based) will pick up the first available

copy of an article — which might be a plagiarized or rehosted version. If that copy lacks attributions or has formatting changes, the model loses the original quote context. The report noted that even publishers who blocked AI crawlers still found their content appearing in LLM answers (via secondary sources) (Source: www.searchenginejournal.com).

These cases highlight the risk: *in the wild, when a user asks an LLM about news quotes, the answer may quote something that is untrustworthy or misattributed*. This risk is amplified if the news articles themselves had attribution issues. Analysts warn that if people see citations being "made up," it could sow doubt about media integrity: "it opens up whole new questions about whether citations can be trusted" (Source: futurism.com).

How Quoting Practices Can Help (or Hurt) LLM Retrieval

From the above, one might conclude that LLMs disregard sources. But a closer view suggests quoting practices still matter. Let us outline how:

- **Clarity of Passage**: Quoted text in an article, if clearly delimited and attributed, becomes an easily identifiable snippet for an LLM to extract. For example, a paragraph ending with *"... said Dr. Emily Chen, lead author on the study."* can be grabbed as a self-contained piece. If the quote is part of a running text without clear boundaries, an LLM's chunker might slice it unpredictably. Thus, journalistic style that isolates quotes into their own paragraphs enhances retrievability.
- **Attribution Tags**: Naming the speaker immediately signals context. Imagine two scenarios: (A) *"Growth was significant," the company's CEO noted.* vs. (B) *"Growth was significant," noted an official.* In version A, an LLM receiving (or training on) that sentence sees "CEO" and "company," inferring a named entity. In B, it sees "official," which is generic. The first situation provides more semantic clues. SEO guides stress that LLMs prize entity mentions: if your content explicitly ties a quote to a known title or name, it strengthens the semantic footprint (Source: searchengineland.com) (Source: mtsoln.com).
- **Source Context**: Beyond the quote sentence, having the surrounding text mention publication date, outlet name, or report title also helps. A line like "According to *The New York Times* on Jan 10, 2025..." provides anchors. LLMs often parse such patterns ("Published on [Date]") as evidence of authoritative origin. This can be leveraged: structured references or noting official reports can feed well into AI recognition. Conversely, if a quote is dropped in isolation with no context, an LLM might assume it is made-up or from an unknown source.
- **Structured Data**: Some publishers use metadata (schema.org citations, JSON-LD) to mark quotes or sources. While LLMs may not always read this, it generally encourages clarity and uniform structure, indirectly aiding AI scraping. For example, a clearly labeled source link (e.g. "[Source: Company Press Release, PDF]") ensures any RAG system will follow the intended trail. It also signals to an LLM that the text comes from a verifiable document.
- **Formatting and Signposting**: Techniques like block quotes formatting or italicizing speaker names (common in newsletter stylings) make quotes stand out. Even if only readable to humans, consistent formatting aids the AI's input preprocessing. Some AI/SEO guides recommend using IDs or anchors around important segments (similar to how academic papers mark citations). If a news article includes something like "

  Quote from Official
  " in its HTML or alt-text, a sophisticated crawler could capture that. In absence of such cues, the LLM's neural patterns must rely on language clues alone.

On the downside, even rigorous news quoting can backfire with LLMs:

- **Syndication Pitfalls**: If a news story is syndicated in multiple places, LLMs might latch onto whichever version is easiest to parse. A quote in a cluttered aggregate site (with ads, commentary) might be disregarded for a mass-released text database version that lacks attributions. This was seen when ChatGPT cited syndicated copies (Source: www.searchenginejournal.com). News organizations should ensure quotes are not only attributed correctly, but that syndicated copies also maintain those attributions (and site scrapers see them).
- **Conflicting Sources**: When two outlets publish the same quote with slight differences, LLMs might treat them as separate. Without robust disambiguation, the same quote might be "stored" under different speaker names in the model. Consistency in phrasing and source tags across outlets would reduce this confusion.

In sum, **the better and clearer the quote attribution in a news article, the higher the chance that an LLM will correctly recognize and reference it**. Conversely, sloppy quoting – missing speaker's name, using vague attributions ("an expert said"), or not specifying the context – makes LLM "understanding" less reliable.
As one SEO practitioner put it, content must be *"clear, unambiguous"* for an LLM (Source: mtsoln.com). That is exactly what strong journalistic attribution strives for. Thus, a synergy exists: thoroughly attributing quotes satisfies both human and machine.

Data Analysis and Quantitative Findings

To ground these insights, we present quantitative data from published research and experiments. Notably, studies in journalistic content analysis and AI evaluation provide relevant figures, which we summarize in tables and descriptions below.

Quotes and Exaggeration in News (Bossema et al.)

A key study by Bossema *et al.* (2019) analyzed thousands of health press releases and related news articles in the UK and Netherlands. It provides detailed statistics on quoting practices:

PUBLICATION TYPE / COUNTRY	ARTICLES WITH ≥1 QUOTE	ARTICLES WITH NEW INDEPENDENT QUOTES	EXAGGERATION ODDS RATIO (NO NEW QUOTE)
UK Press Releases (2011)	99.1% (Source: pubmed.ncbi.nlm.nih.gov)	–	–
UK News Articles (2011)	88.6% (Source: pubmed.ncbi.nlm.nih.gov)	7.5% (Source: pubmed.ncbi.nlm.nih.gov)	2.6× (Source: pubmed.ncbi.nlm.nih.gov)
NL Press Releases (2015)	84.5% (Source: pubmed.ncbi.nlm.nih.gov)	–	–
NL News Articles (2015)	69.7% (Source: pubmed.ncbi.nlm.nih.gov)	7.0% (Source: pubmed.ncbi.nlm.nih.gov)	2.6× (implied) (Source: pubmed.ncbi.nlm.nih.gov)

Table 1: Prevalence of quotes in health press releases and news, and the effect of including external expert quotes on exaggeration of claims in articles (Source: pubmed.ncbi.nlm.nih.gov).

Key observations from this table:

- Nearly all **press releases** (99% in UK, 84.5% in NL) contain at least one quote (Source: pubmed.ncbi.nlm.nih.gov), typically from the study authors or affiliated experts.
- A high proportion of the **news articles** covering those press releases also contained quotes (88.6% in UK, 69.7% in NL) (Source: pubmed.ncbi.nlm.nih.gov). However, in most of those cases the quotes were *lifted from the press releases*.
- Only about **7–8%** of news articles introduced a *new* independent expert quote not present in the PR (Source: pubmed.ncbi.nlm.nih.gov).
- Crucially, articles *without* an external expert quote were **2.6 times more likely** to exaggerate causal claims than those *with* such a quote (Source: pubmed.ncbi.nlm.nih.gov). This implies that simply having an outside expert serve as the speaker correlates with greater factual restraint.

While this study focuses on health news, it illuminates a general pattern: simply borrowing quotes from a source is common, but adding fresh expert input is rare – yet impactful. For our purposes, the relevant takeaway is that **including external quotes significantly affects content quality**. By analogy, an AI reading these news articles might find that reality-checking quotes (from experts not directly involved) strengthens trustworthiness. If generative engines had to pick which snippets to trust, one might hypothesize the ones with independent attributions would be safer. Indeed, content creators who want to be cited by AI as authorities should aim to be among those "external expert" voices.

LLM Citation and Search Studies

On the AI side, recent experiments have quantified how well LLMs cite or retrieve news content. Table 2 below consolidates key findings:

AI SYSTEM / METRIC	FINDINGS	SOURCE
CHATGPT-4 answers (study)	GPT-4 provided references for all answers but only ~43% were *fully accurate*; ~56.7% of cited sources were incorrect or nonexistent (Source: rankstudio.net).	[35]
GPT-4 analogs (study)	In a broad task, GPT-4-like models had ~90% of citations factual (~10% fabricated) (Source: rankstudio.net).	[35]
ChatGPT Search	In 200 retrieval trials, ChatGPT Search gave *incorrect answers* 153 times (76.5% of queries) (Source: www.searchenginejournal.com).	[37]
AI Chatbots (Tow Center)	Combined, chatbots answered >60% of 1,600 quote queries incorrectly (Source: www.cjr.org).	[57]
Perplexity.ai	About 37% error rate in those tests (Source: www.cjr.org).	[57]
xAI Grok-3	About 94% error rate in those tests (Source: www.cjr.org).	[57]
All chatbots (general)	Often fabricated links, cited syndicated content, and seldom said “I don’t know” (Source: www.cjr.org) (Source: www.cjr.org).	[57]

Table 2: Performance of large language models and chatbots in retrieving and citing news content (from various studies (Source: rankstudio.net) (Source: www.cjr.org).

Insights from Table 2:

- In a medical Q&A evaluation, **GPT-4** was prompted to cite sources for each answer. It did so, but less than half of the referenced works (43.3%) were fully accurate (Source: rankstudio.net). Over half were wrong or fictional (Source: rankstudio.net). Thus, even for GPT-4, noise is substantial without careful checking.
- Another study found GPT-4-like models could achieve ~90% factual accuracy in citations (Source: rankstudio.net). The large discrepancy (43% vs 90%) highlights that outcome depends heavily on context, prompting, and domain.
- **ChatGPT Search (OpenAI’s)** was particularly error-prone. In 200 quote-identification queries, it misattributed 153 of them (Source: www.searchenginejournal.com). It basically failed 3 out of 4 times, often linking to the wrong publisher or missing the correct URL.
- The Tow Center’s broader test across 8 tools confirmed the issue is systemic: “incorrect answers” were the majority response (Source: www.cjr.org). Some tools like Perplexity did relatively better (~37% wrong), while others (xAI Grok-3) were disastrously bad (94% wrong) (Source: www.cjr.org).
- Common failure modes included *bypassing robots.txt rules, citing syndicated instead of original articles, and inventing URLs* (Source: www.cjr.org). Many chatbots would answer confidently even when they had no definitive source, rarely giving a qualification (Source: www.cjr.org).

Together, these data show current LLMs have an “attribution problem.” In practical terms: **news organizations cannot rely on generative AI to handle quotes gracefully**. Even if a quote is correctly presented in the news, the AI may still mislead. On the flip side, getting an LLM to *mention your content* is challenging. You can produce great quotes, but the AI may cite a competitor or an alternate source. This has led some SEO experts to warn that *visibility in AI search is not ensured by traffic or links alone, but by being “cite-worthy” (Source: mtsoln.com) (Source: willmarlow.com).*

SEO vs LLM Visibility

Drawing from marketing and AI strategy literature, we compare traditional SEO factors to LLM-focused criteria:

CRITERION	TRADITIONAL SEO EMPHASIS	LLM/AI SEARCH EMPHASIS (AI-CITATION)
Authority	Backlinks, Domain Authority, Brand Fame (Source: mtsoln.com)	Explicit Expertise (clear expert quotes), recognized authority in context (Source: mtsoln.com)
Clarity	Keyword optimization, meta descriptions	Clear, unambiguous language and direct answers (Source: mtsoln.com) (Source: mtsoln.com)
Context	Topical relevancy via keywords	Deep contextual fit to user query, semantic relevance (Source: mtsoln.com) (Source: mtsoln.com)
Structure	Internal site structure, HTML tags	Chunked, modular content (bullets, Q&A, TLDR) that LLMs can easily extract (Source: willmarlow.com) (Source: mtsoln.com)
Traffic (CTR)	High click-through rates improve ranking	Not directly relevant; success measured by being *cited*, not clicked (Source: willmarlow.com)
Freshness	Regular content updates boost SEO	Useful for information currency, but demonstrable logic trumps recency in answers
Citations	External references for credibility (minor factor)	Direct citations or anchor attribution matter greatly (LLMs prefer sourced facts) (Source: rankstudio.net)

Table 3: Comparison of priorities in traditional SEO versus LLM-driven content "citability" (Source: mtsoln.com) (Source: mtsoln.com).

From Table 3, several trends emerge that are relevant to quotes in news:

- **Expert Attribution as Authority**: Rather than simply relying on backlinks or page rank, LLMs seek signals of expertise within the text itself. A journalist quoting a specialist with full credentials adds an explicit expertise marker that LLMs treat as textual authority (Source: mtsoln.com). A name title (e.g. "Dr. Smith") and institutional affiliation in a quote signal trust.
- **Content Clarity Over Keywords**: SEO used to prize keywords; LLMs prize plain-language answers. A snappy quote is often exactly the type of "answer sentence" an AI wants (Source: mtsoln.com). For example, an economic report quoting "Inflation dropped to 2% in June," said the Fed's Janet Yellen, might be more valuable to an LLM than paragraphs full of keyword stuffing.
- **Modularity**: Traditional articles might meander; LLM-targeted content is more modular. Newspapers using bullet lists or Q&A boxes (common in digital formats) make for better AI snippets (Source: willmarlow.com). News with TL;DR summaries or key fact boxes is directly aligned with what an LLM can excerpt.
- **Verification**: In SEO, citing sources is largely an E-A-T (Expertise/Authority/Trust) factor, but only indirectly considered by ranking algorithms. In contrast, LLMs essentially **internalize** factual claims and their attributions. The difference is qualitative: an SEO algorithm might not personally verify every fact, but an LLM will draw on memory of the text. This makes meta-cite (e.g. linking or attributing to outside sources) more influential. Indeed, if LLM is "Rolled back to 2021 training", it has no live updates; for up-to-date answers it relies on retrieval and citations. Therefore, sites that are often quoted by news might indirectly benefit.

In sum, these analyses show that **news organizations aiming to influence AI answers should treat quote attribution with the same rigor they would for readers**. Being the quoted authority can improve both human trust and machine "trust-through-citation." On the flip side, any ambiguity is penalized even more: where a typical Google result might still rank without a clear attribution, an LLM might dismiss or distort an unclear quote.

Case Studies and Real-World Examples

Here we examine concrete scenarios illustrating how news quoting practices have impacted LLM outputs, as well as how LLM behavior has in turn pressured news media.

ChatGPT-Style AI and Quoting Issues

One of the most publicized cases involved **OpenAI's ChatGPT Search** (the Bing Chat/ChatGPT search mode released Nov 2024). A study from Columbia's Tow Center (reported by Search Engine Journal) tested ChatGPT Search on news article quotes (Source: www.searchenginejournal.com). Out of 200 quote queries, 153 responses were incorrect or misattributed (Source: www.searchenginejournal.com). For instance, ChatGPT Search often failed to name the correct news source or publication for a given quote. It sometimes prioritized "pleasing" the user with a plausible-sounding answer over checking facts. This errant behavior raises concern for publishers: enabling AI inclusion of your content might still put your brand in a false context.

Example: The New York Times and ChatGPT

When testers queried ChatGPT Search for quotes from *The New York Times*, the system posted links to unauthorized copies on other sites rather than the official NYT link (Source: www.searchenginejournal.com). The algorithm's inability to properly attribute a quote to the NYT (even when it was presumably trained on many NYT-sourced datasets) meant that *NYT's actual story was not acknowledged*. Instead, ChatGPT pulled a syndicated copy for citation. This caused alarm among publishers: even if you want ChatGPT to cite your site (by, say, not blocking crawlers), it may bypass you anyway if an "easier" source is accessible. The Tow Center notes that these errors "challenge OpenAI's commitment to responsible AI development in journalism" (Source: www.searchenginejournal.com).

Syndication and Crawling

Underlying many of these issues is the problem of **syndication**. Newswire services (Getty, AP, Reuters) often republish content across multiple outlets. An LLM scraping the web might index the plain text of an AP story rather than the original newspaper, for example. If the syndicated version omits say the byline "Reporter: Jane X." or rearranges quotes, an LLM might credit the wrong publication or parse quotes incorrectly. In the example above, ChatGPT citing a non-NYT version suggests that the underlying retrieval engine saw the AP text as the dominant source. (Note: NYT had been in dispute with OpenAI over licensing, possibly affecting access).

For companies quoted in news, syndication means that being quoted in one venue might not guarantee the AI associates that quote with your brand's name, if a "raw" feed without context is indexed. This shows that **where** a quote appears (including hidden metadata) can be as important as the quote itself.

Media Reaction and Industry Perspective

Facing these AI challenges, mainstream media has begun responding. The Guardian article noted earlier described ChatGPT's behavior as a "tsunami of made-up facts" that could "undermine legitimate news sources" (Source: futurism.com). In response to experiments like the Tow Center's, some publishers have published guidelines or statements. For example, the Partnership on AI (a coalition of tech and media organizations) has issued recommendations on how journalists should label AI content and report responsibly (Source: www.mdpi.com) (Source: www.mdpi.com). Some newsrooms are wary: *"Generators of fake information"* is how the Guardian's Chris Moran labeled LLMs that misattribute (Source: futurism.com).

From an SEO/AI-optimization standpoint, some companies see opportunity. Strategy guides advise content creators to structure material so AI "naturally" cites them (Source: createandgrow.com) (Source: willmarlow.com). For instance, a blog post on LLM mentions recommends becoming "the go-to source that AI naturally wants to reference" (Source: createandgrow.com). Practitioners suggest creating anchor content (like detailed FAQs) that LLMs can chunk into answers, and ensuring your brand name appears in those answerable segments (Source: searchengineland.com) (Source: willmarlow.com).

However, the data tells us this is still experimental. The same search engine tests from Tow Center show all current LLM agents are *"generally bad at declining to answer"*, instead giving jaw-dropping confident false answers (Source: www.cjr.org). Even premium models (C4, GPT-4o) were no exception. Many of these systems explicitly do a web search under the hood, but then "rewrite" the answer with invented citations if they can't find a source.

Thus a content producer might have two somewhat conflicting incentives:

- **Be cautious**: Journalistic precision is needed now more than ever. Incorrect quoting can be amplified by AI, harming trust and brand.
- **Be AI-savvy**: At the same time, authors can use quoting and metadata strategically to align with AI preferences (clear entity names, structured answer format) to increase "mentions."

The chief lesson is that content creators should *not* assume LLMs will magically get attribution right. Instead, they should ensure that their content has as few ambiguities as possible. Practically, this could mean double-checking quotes, providing rich context, and even adopting SEO/AI tagging (like llms.txt or API access) when possible.

In the marketing domain, a table of suggestions has emerged (from SEO thought-leaders) on how to target LLM visibility. These include the tactics of using exact-match query phrases in headers (anticipating user prompts), creating standalone "Answer Blocks," and embedding citations to reputed sources (Source: mtsoln.com) (Source: willmarlow.com). Such strategies indirectly highlight the value of quoting authority figures: each quote is, in effect, its own mini-answer block that can be lifted by an AI answer. A well-attributed quote with a heading like **"What did [Expert/Report] say about X?"** is literally structured for AI consumption.

Table: Comparisons and Observations

To summarize key data points, we present the following table:

CONTEXT	STATISTIC / FINDING	SOURCE
News quoting (Media Engagement)	Highest credibility when only a non-partisan government official is quoted (Source: mediaengagement.org) (vs. partisan sources perceived as biased).	[23]
News quoting (Bossema et al.)	Only 7–8% of health news articles added new expert quotes; absence of expert quote gave 2.6× higher odds of exaggeration (Source: pubmed.ncbi.nlm.nih.gov).	[29]
LLM citation accuracy (GPT-4)	GPT-4 with prompts cited sources in all answers, but only ~43% were entirely correct (Source: rankstudio.net).	[35]
LLM citation accuracy (GPT analog)	~90% of citations were factual (10% fabricated) in a wide-domain test (Source: rankstudio.net).	[35]
ChatGPT Search error rate	76.5% of quote-source queries answered incorrectly (153/200) (Source: www.searchenginejournal.com).	[37]
AI chatbots error rate (overall)	>60% incorrect answers in retrieval tasks (Source: www.cjr.org).	[57]
Perplexity.ai error rate	~37% of queries answered incorrectly (Source: www.cjr.org).	[57]
xAI Grok-3 error rate	~94% incorrect (Source: www.cjr.org).	[57]
AI hallucination examples	ChatGPT invented full articles and citations that never existed (Source: futurism.com).	[49]

Table 4: Selected quantitative observations relating to quote attribution in news and LLM behavior.

These figures confirm that the intersection of news quoting and LLM retrieval is fraught with inaccuracies at present. In particular, the fact that even state-of-the-art models have citation accuracy ranging widely (43–90%) (Source: rankstudio.net), and that chat interfaces err 60–90% of the time (Source: www.searchenginejournal.com) (Source: www.cjr.org), should warn content creators. They must assume that LLMs are prone to *misbring* the content – and thus take steps (through precise quoting) to mitigate that risk.

Implications and Future Directions

Our analysis reveals deep implications for multiple stakeholders:

For Journalists and Newsrooms: Quote attribution has always been central to journalistic integrity. Now it also impacts how AI systems will mention or omit content. Given LLMs' current limitations, journalists should be **extra vigilant** about accuracy. Rigorous sourcing and context become even more important, since missteps might be amplified by AI answers. Some news organizations are already revising their standards: for instance, the Partnership on AI recommends clearly labeling AI-generated content and "source-awareness" practices (Source: www.mdpi.com). Newsrooms might adopt AI-detection tools or disclaimers for content likely to be fed to AI. Additionally, legal and ethical guidelines will evolve: questions about copyright and AI training data are already at issue. Proper attribution can help preempt litigation on both IP and defamation grounds.

For Content Publishers (Brands/Experts): Quoting by outlets translates to visibility beyond print. If a brand or expert is quoted in reputable news, even if it misses clicks, it can increase LLM footprint. SEO/PR strategies may shift to not just earning quotes for human readers, but *ensuring those quotes are structured for algorithmic discovery*. For example, a PR team might encourage journalists to always include a person's full title and a hearing reference, rather than a vague quote. Marketers will track not only Google rankings but also "Rank 0" – whether AI chatbots mention them. Tools for monitoring brand mentions in LLM answers are emerging (Source: mtsoln.com). The notion of a "press release for AI" is likely to take hold: crafting press quotes with AI in mind (concise, direct statements) may become a niche skill.

For SEO and Digital Marketers: The rise of AI citations demands a shift in optimization tactics. Traditional link-building remains relevant (to be discovered in the training data), but emphasis is moving toward *entity association* and *citable content structure*. Content strategies now often include creating clear Q&A blocks, schema markup, and unique research or data (to induce others to quote you) (Source: mtsoln.com) (Source: willmarlow.com). Some even propose an "LLM citation strategy" analogous to backlink strategies (Source: www.higoodie.com) (Source: davidmelamed.com). Forming partnerships for content distribution (e.g. writing in industry journals where AI scrapers look, or Wikipedia citations) also appeals, in order to seed authoritative content that chatbots can reach. Marketers must also consider a new metric: *AI CTR*, or how often their site is directly referenced by AI responses (even if no click). This could influence budgeting and content planning in coming years.

For Users and Society: On the user side, these developments have mixed effects. Ideally, well-designed LLM systems would provide concise answers with transparent sourcing, aiding user trust and saving time. In practice today, users risk being misled by confidently stated but false "facts" and quotes from AI systems. Media literacy is needed: users should verify AI-provided quotes against original articles. Journalists and educators must teach people to treat AI answers like unverified sources until checked, similar to early days of search engines. There is also a social equity dimension: if only large organizations can afford to be AI-friendly (with structured content and licensing deals), smaller outlets might be marginalized. Tow Center interviewees have worried that chiefs of AI companies may overlook valuable niche content from smaller local publishers (Source: mediaengagement.org). Ensuring diverse media voices are recognized in AI answers is an emerging challenge (and fairness issue).

For LLM Developers and Platforms: These findings place responsibility on AI model designers. Clearly, better citation integration is needed. Approaches include Retrieval-Augmented Generation (RAG) with better provenance, watermarking outputs, and more conservative refusal modes. Some work (like the surveyed citation frameworks (Source: www.themoonlight.io) explores architectural fixes, but deployment in consumer tools has lagged. Policymakers and platforms may eventually require "right-to-refuse hallucinations" or standardized source-checking routines. For example, partners like the Partnership on AI are pushing standards for newsroom-AI collaboration. Google's prototype AI Overviews already show footnotes, but even those often point to syndicated copies. Ideally, generative systems should cite clearly, or at least say "according to source X..." only when sure. Until then, we see tension: newsrooms want safe, respectful use of content; LLM tools want broad training data. Companies like OpenAI (which now allows robots.txt opt-outs) are starting to listen, but progress is ongoing.

Future Research and Open Questions

This area is very new; much remains unknown. Some future directions include:

- **Empirical LLM Testing with Quotes:** Systematic studies could evaluate how different quoting styles affect LLM retrieval. For example, writing the same content but with varied attribution wording, then querying an LLM to see which version it picks. Such A/B tests would inform best practices quantitatively.
- **AI-Generated Synthesis of News:** As generative journalism (AI-written articles) becomes real, how will quote attribution be handled? Some tools (e.g. Lynx Insight at Reuters (Source: reutersagency.com) are already auto-writing data-driven stories. Ensuring those AI drafts insert quotes correctly might soon need automated source-checkers.
- **AI Literacy and Countermeasures:** How do readers differentiate between a factual quote and a hallucinated one in an AI answer? Design of user interfaces could show trust scores or provenance paths. Research in human-computer interaction could help end-users better evaluate LLM output about news.
- **Longitudinal Effects:** Over time, if LLMs repeatedly misquote news, will that alter public perception? Some dystopian analyses warn of "bottomless deception" when AI propaganda layers on itself, with fake quotes feeding conspiracy theories (Source: futurism.com). Studying information diffusion in the AI era is crucial.
- **Legal and Ethical Frameworks:** Should there be guidelines (or even laws) regarding how AI tools must attribute information sourced from news? For instance, imposing transparency standards for AI answers or outlawing AI hallucinations on sensitive topics. The journalistic community's standards (truth, accuracy) may need to be translated into tech policy.

Conclusion

In the rapidly evolving landscape of generative AI, **quote attribution in news articles has emerged as a critical factor influencing how content is treated by LLMs**. Our research consolidates evidence from media studies, AI experiments, and SEO strategy to show that well-attributed quotations serve a dual purpose: they boost human trust and they align with the content structures LLMs prefer. Conversely, vague or incorrect attributions can amplify confusion, as LLMs readily invent or misassign quotes when the data is murky (Source: futurism.com) (Source: www.cjr.org).

Key findings include:

- Credibility studies confirm that news posts quoting authoritative officials are rated highest in trust (Source: mediaengagement.org). News lacking clear attribution is perceived as less credible or biased.
- Empirical content analysis shows that articles with outside expert quotes are markedly *more accurate*, while those without are prone to exaggeration (Source: pubmed.ncbi.nlm.nih.gov).
- On the AI side, LLM-powered tools often provide incorrect citations. ChatGPT Search misattributed quotes 76.5% of the time in one test (Source: www.searchenginejournal.com), and multiple AI chatbots collectively mis-identified news sources in over 60% of experiments (Source: www.cjr.org).
- Seo-technical frameworks indicate that LLMs prioritize content "clarity" and "contextual fit" (Source: mtsoln.com) (Source: mtsoln.com). Structured, self-contained segments (like properly introduced quotes) are most likely to be extracted.

The confluence of these findings implies: **Newsrooms and content creators should adhere to the highest standards of attribution, not just for readers' sake but to ensure accurate machine consumption**. Quoting concrete names and titles, structuring passages clearly (potentially labeling them as Q&A blocks), and providing metadata can improve the odds that AI systems get it right. Likewise, AI developers have a responsibility to refine how their models handle quotes, to avoid undermining journalistic work.

For the future, the implications are profound. As more people rely on AI-generated summaries, even small quoting errors can ripple through the information ecosystem. Yet, there is hope that synergy is possible: if journalists and technologists collaborate – for instance through the Partnership on AI or industry standards – they can co-create workflows where news content remains both reader-trustworthy and AI-friendly.

In closing, quote attribution is not merely a stylistic concern; it shapes the *knowledge footprints* that LLMs trace. By understanding this interaction deeply, stakeholders can harness it: news media can increase effective reach and credibility, businesses can gain rightful AI visibility, and society can insist on accountability in an age of automated answers. The journey toward robust "LLM mentions" is only beginning, but meticulous quoting will be one of its cornerstones (Source: mediaengagement.org) (Source: willmarlow.com).

References

All claims and data in this report are supported by the following sources:

- Bossema *et al.* (2019), *Expert quotes and exaggeration in health news: a retrospective quantitative content analysis* (Source: pubmed.ncbi.nlm.nih.gov).
- Center for Media Engagement (2025), *Quotes and Credibility: How Storytelling Approaches Shape Perceptions Across Party Lines* (Source: mediaengagement.org).
- Huang *et al.* (2025), *Attribution, Citation, and Quotation: A Survey of Evidence-based Text Generation with Large Language Models* (Source: www.themoonlight.io).
- Huang (2025), *Beyond Popularity: The Playbook for Dominating Visibility in AI Search* (Source: mtsoln.com) (Source: mtsoln.com).
- Search Engine Journal (2024), *ChatGPT Search Fails Attribution Test, Misquotes News Sources* (Source: www.searchenginejournal.com).
- Futurism (2023), *Newspaper Alarmed When ChatGPT References Article It Never Published* (Source: futurism.com).
- Columbia Journalism Review (2024), *AI Search Has A Citation Problem* (Source: www.cjr.org).
- Krstović (2025), *LLM SEO Explained: How to Get Your Content Cited in AI Tools* (Source: willmarlow.com) (Source: willmarlow.com).
- Todorov (2025), *Influence LLM Mentions Through Strategic Content* (Source: createandgrow.com).
- Mercury Tech Solutions (2025), *Maximize AI Visibility: Understanding LLM Citation* (Source: mtsoln.com) (Source: mtsoln.com).
- Reuters (2023), *How AI helps power trusted news at Reuters* (Source: reutersagency.com).
- Shi & Sun (2024), *How Generative AI Is Transforming Journalism: Development, Application and Ethics* (Source: www.mdpi.com).

(Inline citations in brackets link directly to the source material used for each claim.)

Tags: llm mentions, quote attribution, ai search, seo for llms, source credibility, journalism, ai content strategy

---