Generative AI's aggregated knowledge versus web-based curated knowledge

Ted Selker[1,†], Yunzi Wu[2,†]

[1]*University of Colorado, Boulder, Boulder, CO USA 80309*
[2]*Lehigh University, 27 Memorial Drive, West Bethlehem, PA USA, 18015*

Abstract

This paper explores what kinds of questions are best served by Generative AI (GenAI) using Large Language Models (LLMs) that aggregate and package knowledge, and when traditional curated web-sourced search results serve users better.

An experiment comparing product searches using ChatGPT, Google search engine without GenAI responses, or both helped us understand more about the compelling nature of generated responses. Even for people that were skeptical of GenAI, the lists search results came back with were slower and less compelling to work with than GenAI's summarization and explaination. However, our qualitative comparison indicated that search engines still provide superior results for factual and niche knowledge.

1. Introduction

We spend our lives communicating to learn and accomplish tasks. One challenge is to develop a repertoire of knowledge and communication tools that are both valuable and productive. The search paradigm has been transformative, making older internet tools like Archie, Veronica, and WAIS forgettable. It provides us with access to the world's knowledge. Search connects keywords using Boolean operators and presents lists of solutions relevant to knowledge and opportunities related to these queries [1]. These results draw from of all curated knowledge on the internet. For some time now, search technologies have started utilizing AI to generate web pages that serve as sources of information, as opposed to 'official' sources. For instance, search results for a restaurant, often present a page compiled by the search engine, which aggregates information from actual customers about the restaurant rather than relying solely on the restaurant's self promoting description. Should we trust the restaurant to describe itself accurately, or could an aggregation of various sources offer a more trustworthy account?"

A popular direction in AI involves training the high-dimensional predictive *Transformer* to form language solutions. This approach leads to the creation of so-called Large Language Models (LLMs) [2, 3]. GenAI starts with human prompt questions, including human-defined information and presentation goals. The LLM-based solutions aggregate and package it as a solution, such as a story, poem, image, computer program, etc.[4].

In conversation, people are accustomed to not having all stories correct or complete. The value of discovering, evaluating, and expanding one's conceptions is best verified through traceable provenance. Whether in a street conversation or while viewing online information, source provenance must be made apparent to accumulate facts rather than rhetoric, thereby distinguishing reliable sources from fabricated stories. The search paradigm that has enabled us to find Web information is valued for its ability to produce referenced responses.

The evaluation of ideas using computers has evolved. Empathetic language responses were first demonstrated in the 1960s with Eliza [5]. By 1981, EPISTLE [6] was already using AI to improve authors' syntax [7]. For a long time. creativity-enhancing software commonly consisted of computer "outliners" and "mind maps." Now, we search the world's knowledge on Google. Siri was a breakthrough in knowledge-based verbal question-answering [16, 17]. Machine learned Large Langauge Models became popular when GenAI added user-focused text generation, responses prioritizing solution flow and continuity[18]. The ability to guide GenAI with style and qualities of responses presents a completely new kind of interaction paradigm. The focus of requests and responses is on what should be presented and how.

1.1. Knowledge Interaction Scenarios

For sometime before GenAI was created, personal assistants like Siri were reducing questions from a person to an actionable query or action. Comparing the two, we posed detailed questions to both the Alexa personal assistant and ChatGPT. An emblematic one is, "What is a quark?"; Siri and GenAI gave similar answers that seemed to have similar knowledge to Wikipedia. While both have models of discourse, ChatGPT also has a model of persuasion, filling in gaps with plausible ideas. Alexa lacks the GenAI capability of filling in gaps with likely-sounding responses, as people do in conversation. So far, Alexa and ChatGPT both feel less reliable than internet searches for different reasons. Alexa finds an answer, while GenAI may try to prioritize and package top answers.

2. New Scenarios for Knowledge Interaction

Search has changed the way we find, use, and acquire things. The multiple goals of search have shifted its focus from finding a perfect answer to a negotiated set of results that both the search engine and the user interact with. Early purpose-built search engines were limited to finding web pages. Today's search engines serve multiple goals, providing marketing results, sales results, website results, video results, scholarly results, how-to results, and AI website-produced composite results. A menu bar on Chrome allows Google Chrome users to focus on shopping, images, videos, news, maps, books, flights, and finance. We have many goals in mind. The multifaceted goals of the search system mean that one searches and then finds themselves entangled in commercial opportunities that support the search business. The CEO of a large search engine company was asked why its email platform was presenting a married man with ads for breast enhancement and dating services. They said, "We tried to remove those ads, but the advertising generates a lot of revenue." Focusing these systems solely on users' goals will improve the system's reputation as well as productivity. GenAI has so far avoided cluttering solutions with predatory results.

2.0.1. GenAI Can Feel Natural to Use

A few weeks after ChatGPT's initial release, we were amazed by a group of octogenarians discussing their use of the technology. One of them recited a competent poem that ChatGPT had written for them in iambic pentameter, using Shakespearean language to discuss a current topic on their minds. ChatGPT's prowess showed breathtaking composition skills forming an easy on-ramp even for our octogenarians.

2.0.2. AI Can Provide Entertainment

At a rain-forest resort in Sylhet, a group of people had asked ChatGPT to write a poem in the style of the most famous Bangladeshi poet. We watched in awe as one of the groups sang it as a song in Bengali. For something like a poem, the composition is the success.

2.0.3. AI Can Generate Sophisticated Suggestions

We employed ChatGPT to author apologies for training examples about communication affect. Although the AI's suggestions became repetitive after a few sentences, they still helped people write training utterances almost 10 times faster. The reach of examples GenAI draws from far exceed what any person can produce.

2.0.4. Knowledge exploration needs guidance

A friend asked ChatGPT about the best recording Shirley Temple made, and it said she never recorded. "It's worth noting that Shirley Temple, despite her fame as a child actress in the 1930s, retired from acting at the age of 22 and did not release any songs during her career." This statement, of course, was incorrect. The follow-up question aimed to provide more information, and ChatGPT then contradicted itself by noting her many real recordings. We have had to face that the goal of good presentations makes us need to test specific details in GenAI responses.

2.0.5. GenAI Can Create

We can search online support services like Stack Overflow to help us learn from other's programming examples. On the other hand, GenAI is now compiling all known code with tools like Copilot, assisting people in programming. Users can specify algorithms, the programming language to be used, and the desired output to receive working programs. By critiquing the result, they can prompt Copilot to fix bugs, change the approach, or even port the program to another language within seconds. Still, customizing and testing these examples can require many reformulations. However, programming complex tasks by asking questions can feel like trying to drive a car from the back seat.

While even octagnearians are accustomed to conversation, search engines do not facilitate such interactions. GenAI initiates a conversation. Another potential advantage of GenAI over search is that it encourages us to pause and evaluate its results. The clickbait nature of search tools may lead us to make hasty decisions without adequate consideration.

We can rely on GenAI systems for complete and organized responses, but people commonly state that (as with people) we can't trust them for truthfulness. Still, conversation may not be the most effective way to uncover facts or, to analyze and resolve bugs in a computer program.

Today's search engines are expected to respond appropriately to all informational needs, displaying multiple alternative links that could be followed. The ease of accessing various forms of knowledge is now enhanced by the corpus of all digitized information. The AI-generated aspect added to search engines today might seem helpful, but represents a different kind of information not curated by people. Still, the search paradigm is not designed to build knowledge, critically analyze discourse, or formulate solutions. They aim to differentiate and present results tailored to a person's needs in well under a second. However, achieving this level of precision is challenging without understanding the context, background, and objectives of the request.

The GenAI paradigm aggregates connections between a portion of the online knowledge available to legacy search systems. A key question arises: Can the automatically-generated narratives, which bring together parts of many disparate sources of knowledge, compete with the precise knowledge we all curate to be accessed by search engines? This question involves not just the quality and appropriateness of the knowledge but also its digestibility. While we often believe that we'll recognize what we want when we see it, the reality may be more complex.

The search systems of the past 20 years have enabled us to access a vast array of publicly presented information. However, the importance of structurally-coherent stories often outweighs the significance of knowing where and how we found something. Recent work has focused on demonstrating that helping people understand which GenAI results can be trusted may improve their decision-making [8, 9]. We designed an experiment to see if the GenAI results might change the speed and way people make decisions.

A consumer purchasing experiment was designed to demonstrate how the different paradigms worked for people who believed in and also for people who didn't believe in GenAI. The primary aim of our experiment is to examine how using both legacy and new knowledge exploration platforms affects complex knowledge-based user decisions, such as buying a car. The study seeks to contrast how knowledge gathered from the web through legacy search engines and newly introduced GenAI aids consumers in this process. What are the most and least effective aspects of these online information access paradigms?

3. Consumer experiment method

The study hypothesizes that a GenAI tool offers a spectrum of choices and hypothetical solutions, although these may not always be actionable. A combination of chat and search functionalities might contribute to generating more actionable solutions. When relying solely on search engines, users might find it challenging to initiate the process and become bogged down in details, even though the results are more specific.

We began by recruiting individuals who are contemplating buying a car shortly and who reside in the U.S. The experiment was designed to last approximately one hour. The docent administered the experiment with a uniform script for each user. As well as the results of the pretest, searches, and post-tests, recorded observations were made of each participant.

Before the experiment, participants answered a questionnaire to gauge familiarity with search engines and GenAI chatbots. The questionnaire also aimed to explore their decision-making approach when it comes to buying a car.

Participants were required to complete three tasks in an order of their choosing.

- Task 1: Use only a search engine for information gathering.
- Task 2: Use only ChatGPT3.5 for the same purpose.
- Task 3: Use both ChatGPT3.5 and a search engine.

Each participant was tasked with researching a car purchase based on real-world scenarios and goals, such as finding the best car within a specific budget or comparing various car models while focusing on different lifestyle needs and preferences. They used Google search and ChatGPT3.5 tools to find several options, compare them, and narrow their choices down to two.

We used screen recording, facial expressions, and verbal reactions to gain insight into user challenges, thought patterns, and decision-making processes. The task was designed to be ecologically valid as a realistic, complex task.

Recording the experience based on both how they acted and what users did helped us better understand the role and efficacy of online tools in complex decision-making scenarios like car purchasing.

4. Data

We included three sections in our consumer experiment: pre-survey, three tasks, and post-survey.

In our pre-survey, we asked two questions to understand the participants' car purchase history and habits, All had purchased at least one car and 60 % said that they would buy a car based on the brand their family already owned, the other 40 % rest would base it on what they found in a search.

Our purchasing experiment included three tasks. For each task, we asked the same question: "What might be different about using search or ChatGPT for finding cars?" [Figures 3, 6, and 9]. The reason

Fig. 1. Google Task: Initial cars you are considering; include reasons for each.

Fig. 2. Google Task: Final 2 car selections include reasons for each.

Fig. 3. Google Task: What might be different about Search or ChatGPT for finding cars?

Fig. 4. ChatGPT Task: Initial cars you are considering; include reasons for each.

we asked the same question was to compare the differences before, during, and after participants used ChatGPT3.5 for the search task.

4.0.1. Task One: Use only Google search to find a car.

As they are used to, participants opened a browser and used the Google search box to learn about cars that meet the following requirements: a car suitable for an active lifestyle, that is safe, reliable, and also good for transporting kids.

They found several options [Figure 1], compared them, and identified differences. In the end, they narrowed their choices down to two[Figure 2].

4.0.2. Task Two - Use only ChatGPT to find a car.

Participants used the new GenAI paradigm to find cars that meet the following requirements: suitable for an urban, luxury lifestyle, and good for transporting kids. They came across various alternatives [Figure 4] and assessed the distinctions between them. Finally, they settled on two final options [Figure 5].

4.0.3. Task Three - Use Both Google search and GenAI to find a car.

Participants checked details about cars that met the criteria for a rural lifestyle, fuel efficiency, low maintenance costs, and suitability for transporting kids—using both Google search and ChatGPT. After discovering multiple options [Figure 7], they evaluated and distinguished the differences among them. Ultimately, they whittled their selections down to just two [Figure 8].

Each participant's time spent on the task was documented [Figure 10].

Fig. 5. ChatGPT Task: Final 2 car selections include reasons for each.

Fig. 6. ChatGPT Task: What might be different about Search or ChatGPT for finding cars?

Fig. 7. Google search and ChatGPT Task: Initial cars you are considering; include reasons for each.

Fig. 8. Google search and ChatGPT Task: Final 2 car selections include reasons for each.

Fig. 9. Google search and ChatGPT Task: What might be different about Search or ChatGPT for finding cars?

Index	Description	Task 1	Task 2	Task 3
	Tool	Google	ChatGPT	Google+Chat GPT
User 1	Tried ChatGPT a few times	17	14	14
User 2	First-time user	38	30	10
User 3	First-time user	12	8	29
User 4	ChatGPT daily user	6	5	5
User 5	ChatGPT daily user	15	12	14
	Mean for tool	17.6	13.8	14.4
	Median for tool	15	12	14
	Task 1 VS. 2 P-Value	0.030		
	Task 1 VS. 3 P-Value	0.679		
	Task 2 VS. 3 P-Value	0.931		

Fig. 10. The time spent on each user and task, p-value is calculated with the T-test

Following the completion of the three tasks, we solicited participants' opinions on the tools through a post-survey. What drawbacks do you perceive in ChatGPT when compared to Google search [Figure 11]? What benefits do you see in using ChatGPT over Google search[Figure 12]? How would you rate the ease or difficulty of locating the information you required using both Google search and ChatGPT[Figures 13]? Another question delves into whether ChatGPT was helpful, and the last question explored the frequency with which they utilize ChatGPT.

5. Discussion

Pre-survey findings All participants had prior experience purchasing one or more vehicles. Participants were very focused on their preconceptions about different car brands and types. Even when viewing new information, they relied on their own research and familial car ownership when proposing a car purchase .

Through the think-aloud process, we recorded that while all participants were aware of ChatGPT,
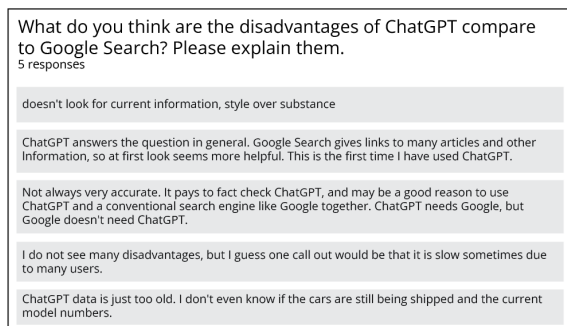
Fig. 11 (left box):

What do you think are the disadvantages of ChatGPT compare to Google Search? Please explain them.
5 responses

doesn't look for current information, style over substance

ChatGPT answers the question in general. Google Search gives links to many articles and other Information, so at first look seems more helpful. This is the first time I have used ChatGPT.

Not always very accurate. It pays to fact check ChatGPT, and may be a good reason to use ChatGPT and a conventional search engine like Google together. ChatGPT needs Google, but Google doesn't need ChatGPT.

I do not see many disadvantages, but I guess one call out would be that it is slow sometimes due to many users.

ChatGPT data is just too old. I don't even know if the cars are still being shipped and the current model numbers.

Fig. 11. What do you think are the disadvantages of ChatGPT compared to Google search?

Fig. 12 (right box):

What do you think are the advantages of ChatGPT compare to Google Search? Please explain them.
5 responses

fluency of composition, frames and reframes composition quickly

ChatGPT gives concise information which is less confusing than Google Search with many, many choices.

You get many interesting results fast! Wow. You can more easily stack search terms than modern Google. I imagine ChatGPT also serves as a better brainstorm tool than Google in some ways. It really makes you think when you use it.

ChatGPT allows specific questions and gives scenarios and answers based on what you are asking. It is also a lot quicker and more advanced than a google search.

So ChatGPT gives me a full explanation of options and choices, but these are old Sep 2021 data.
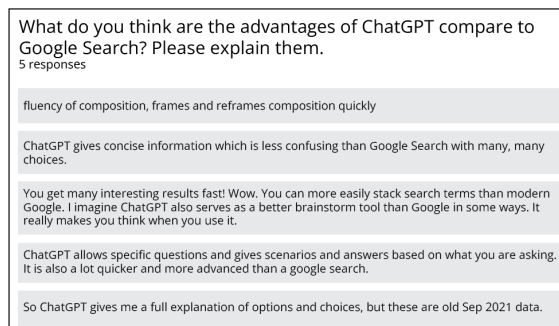
Fig. 12. What do you think are the advantages of ChatGPT compared to Google search?

Fig. 13 (left):

How easy or difficult was it to find the information you needed Google and ChatGPT

Range	Google Search	ChatGPT
1	1	2
2	2	2
3	1	0
4	1	0
5	0	0
The P-value of the T-test		0.915

Fig. 13. How easy or difficult was it to find the information you needed Google and ChatGPT?

Fig. 14 (right box):

Did ChatGPT help?
5 responses

yes. Models suggested seem feasible.

Yes - Gives a starting point

Certainly, yes. Even if it didn't always give me accurate results, it gave many good options in this case.

Yes

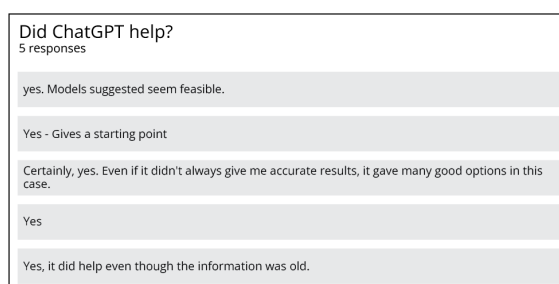Yes, it did help even though the information was old.

Fig. 14. Did ChatGPT help?

two were first-time users who had very negative expectations about what the experience would be like.

Participant feedback on three tasks and time spent[Figures 3,6, and 9]: ChatGPT: Focuses on stylistic presentation rather than on current or accurate information. Search vs. ChatGPT: Google search requires several passes for more in-depth research, whereas ChatGPT provides concise answers upon the first request.

AI Concerns: While ChatGPT can suggest new options, concerns about the long-term societal impact of AI were raised.

Efficiency: ChatGPT can quickly list top options, saving user organizing effort and time.

Personalization: ChatGPT offers personalized responses, while search engines can provide specific models and additional suggestions.

Range of Options: ChatGPT provided a more comprehensive list in one answer.

- Time and Effort: It took several steps to obtain the needed information using a search engine; ChatGPT offered summarized information in a single step.
- Accuracy Concerns: Questions were raised regarding ChatGPT's accuracy.
- Quality vs. Quantity: ChatGPT offered more options, but search engines provided more accurate results.
- Data Currency: Search engines offered more up-to-date information.
- Observations on the use of time on each task: Experienced Users were Faster; Users who were familiar with ChatGPT (Users 1, 4, and 5) generally completed tasks faster than those who were not (Users 2 and 3) [Figure 10].
- Daily users are most efficient: Users 4 and 5, who are daily users and 4.0 subscribers, had the fastest times across all tasks, indicating that familiarity and subscription level may correlate with efficiency [Figure 10].

Summary of Statistical Analysis on Time Taken for Tasks 1, 2, and 3. Null Hypothesis: A t-test did not show a significant difference in time taken between using Google search (Task 1) and ChatGPT

(Task 2) for the tasks. Result: A p-value of 0.030 rejected the null hypothesis. There is a statistically significant difference in the time taken to complete tasks using Google search and ChatGPT.

Data did not show a meaningful difference in the time required to complete Task 3 (utilizing both Google search and ChatGPT) as compared to Task 1 (utilizing only Google search). Result: A p-value of 0.68 leading us to not reject the null hypothesis.

Data did not show a significatnt difference in the time taken to complete Task 2 (ChatGPT) and Task 3 (Google search and ChatGPT). Result: A p-value of 0.93 was obtained,

Quantitative Conclusion: There is a statistically significant difference in the time taken to complete tasks when comparing Google search and ChatGPT (Task 1 vs. Task 2).

There is no statistical difference in the time taken to complete the tasks for Google search used with ChatGPT (Task 3), compared to using Google search (Task 1) or ChatGPT (Task 2) alone [Firgure 11].

Post-Study Survey: Advantages of ChatGPT: All participants found ChatGPT to be fluent, concise, and efficient, offering a tailored search experience and aiding in brainstorming and critical thinking.

Disadvantages of ChatGPT: While ChatGPT was useful for quick queries, users described it lacking depth, accuracy, and current information compared to Google search.

Ease of use in ChatGPT: On a scale of 1 to 5, with 1 being the easiest and 5 being the hardest, participants rated the ease of finding information on ChatGPT between 1 and 2.

Ease of use in search: On the same scale, participants rating of search varied between 1 and 4.

Data was insufficient to show a not significant difference in the ease or difficulty ratings between Google search and ChatGPT. A p-value of 0.915 failed to reject the null hypothesis.

While it appeared that ChatGPT is easier to use than Google search. However, more subjects are needed to test for significance. Indeed, all but one participant expressed a preference for ChatGPT following the completion of the three tasks. The outlier found that search accuracy was much more important than the way ChatGPT brought everything together.

The especially interesting cases were the participants that did not have favorable views of GenAI on pretests. They found its comparative value convincing, productively used it, and integrated its results. The typical time to go through the three tasks was one hour. One participant who was very used to knowledge exploration did everything in 20 minutes. Although everyone showed their prior biases about car brands in their reactions to GenAI suggestions, one emblematic participant was incredulous at results that didn't match their brand loyalty and prior beliefs. Still, that participant rated GenAI as valuable to their exploration.

5.0.1. Consumer Experiment Observations

GenAI packages and organizes a spectrum of choices and hypothetical solutions, but these may not always be actionable.

Supported: Participants found that ChatGPT offered a broad range of options but raised concerns about its accuracy. This suggests that while ChatGPT can generate a variety of choices, these may not always inspire action.

A combination of chat and search functionalities helps in generating more actionable solutions.

Supported: Participants who used both Google search and ChatGPT found that they could obtain a more comprehensive list of options. This suggests that the combination does indeed provide more actionable solutions.

Users may sometimes lack a systematic approach.

Partially Supported: The feedback that ChatGPT helps in brainstorming and critical thinking suggests that users find the chat format helpful in structuring their approach.

When relying solely on search engines, users often find it challenging to initiate the process and may get bogged down in details.

Partially Supported: Participants noted that Google search requires more in-depth research, which could be interpreted as getting "bogged down in details." When then does search give us what we want and when does it bog us down in the details?

6. Conclusion

This paper shows when search engines or GenAI systems might better serve users' knowledge needs in an information rich decison. Our quantitative purchasing experiment found even sceptical participants valued GenAI's efficiency for gathering knowledge for putting a decision in context.

Search and Generated responses have complementary strengths of the two knowledge-access paradigms suiting different goals. Advancing verifiable generative workflows could augment knowledge-building by combining provenance from search with contextual perspectives from AI as Perplexity attempts to do [10]. While further research is needed, these findings suggest that integrating the unique advantages of search and GenAI could empower more robust knowledge exploration. Overall, this paper demonstrates the value each approach offers for varied user needs.

6.1. Broader uses of LLM AI

The last 20 years have shown the value of using keywords to access comprehensive knowledge sources. The value of these comprehensive data sources has changed the way we do everything, from play to work, education, and procurement.

Search systems now take on many goals, from creating website-like results to promoting goods and services, and even helping with physical directions.

The acquisition of facts has become fraught. Fake facts are not only found in GenAI's hallucinations, but also in things people make up or repeat to support or create some strategic social or political movement.

There used to be the idea that social media would help break down social barriers. It now seems that social networks solving social problems were an early mirage.

Fantasies are stories that are rooted enough in reality that people imagine and often hope they will become reality. More than hallucinations, GenAI creates fantasies. At least parts of fantasies are, or can become, a reality. Without checking out what in fantasy is true or achievable, people can echo each other's fantasies in a spiral of confabulation. But communication and knowledge are important.

The current opportunities for using AI to teach and provoke people to create more expansive and informed products are possibly the most important technical achievements of our time. It is up to us to make scenarios that improve the solutions people create.

An important step towards all of this is assured truth and provenance, two features that search has honored and continue to be crucial for all real solutions. The provenance of where knowledge originated is critical to knowing what we know. We look forward to a world in which peoples' searches, work, and communication are well-informed [9].

We see so much to do in making GenAI results useful and accurate. As program development environments help debug programming, so Knowledge Development Environments might help us refine and check our knowledge and communication work.

7. Acknowledgments

References

[1] J. Hardwick, Google search operators: The complete list (44 advanced operators), 2023. URL: https://ahrefs.com/blog/google-advanced-search-operators/.

[2] Y. Belinkov, J. Vig, Analyzing the structure of attention in a transformer language model, in: Proceedings of the ACl, BlackboxNLP Workshop, volume 24, Springer, Singapore, 2019, pp. 63 – 76.

[3] T. Brants, A. C. Popat, P. Xu, F. J. Och, J. Dean, Large language models in machine translation, in: J. Eisner (Ed.), Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), Association for Computational Linguistics, Prague, Czech Republic, 2007, pp. 858–867. URL: https://aclanthology.org/D07-1090.

[4] K. Roose, Why a conversation with bing's chatbot left me deeply unsettled, 2023. URL: https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html#:~:text=The%20Shift-,A%20Conversation%20With%20Bing's%20Chatbot%20Left%20Me%20Deeply%20Unsettled,declaring%20its%20love%20for%20me.&text=Kevin%20Roose%20is%20a%20technology,Times%20podcast%20%E2%80%9CHard%20Fork.%E2%80%9D&text=Last%20week%2C%20after%20testing%20the%20new%2C%20A.I.

[5] J. Weizenbaum, Eliza a computer program for the study of natural language communication between man and machine, Commun. ACM 9 (1966) 36–45. doi:10.1145/365153.365168.

[6] G. E. Heidorn, K. Jensen, L. A. Miller, R. J. Byrd, M. Chodorow, The epistle text-critiquing system, IBM Systems Journal 21 (1982) 305–326.

[7] J. Lombradi, Welcome microsoft word in a new version of windows, 1990.

[8] S. E. Spatharioti, D. M. Rothschild, D. G. Goldstein, J. M. Hofman, Comparing traditional and llm-based search for consumer choice: A randomized experiment, 2023. arXiv:2307.03744.

[9] T. Selker, Ai for the generation and testing of ideas towards an ai supported knowledge development environment, 2023. arXiv:2307.08876.

[10] D. Colla, M. Delsanto, M. Agosto, B. Vitiello, D. P. Radicioni, Semantic coherence markers: The contribution of perplexity metrics, Artificial Intelligence in Medicine 134 (2022) 102393. URL: https://www.sciencedirect.com/science/article/pii/S0933365722001440. doi:10.1016/j.artmed.2022.102393.