

Generative Artificial Intelligence (GenAI) as Information Tool for Lexicographic Information Needs

Theo J.D. Bothma, *Department of Information Science,
University of Pretoria, South Africa*
(theo.bothma@up.ac.za)
(<https://orcid.org/0000-0001-7850-3263>)

and

Rufus H. Gouws, *Department of Afrikaans and Dutch,
Stellenbosch University, South Africa*
(rhg@sun.ac.za) (<https://orcid.org/0000-0002-3423-058X>)

Abstract: The focus of this paper is on GenAI, chatbots and some implications for lexicography and dictionary use. The use of GenAI as an information tool to provide information to end-users (readers) who have a specific information need when reading a text is discussed. GenAI could previously provide content similar to that of a dictionary but could not provide contextualised answers. This has changed and most chatbots now have two modes, a "search" mode and a "thinking/reasoning mode", i.e., it is able to argue logically about its different proposed meanings in context and tends to offer a solution. This feature is discussed at the hand of examples, and it seems as if the chatbots can now provide correct contextual meaning and logically motivate the choice of meaning in context, based on their critical analysis and thinking skills although it still "hallucinates" if it has no answer. Hereafter, the role of dictionaries and lexicographers in an AI-enhanced world is discussed. The traditional assignment of dictionaries to document the status and history of a language remains a very important function and needs to be encouraged, especially in environments with limited language resources. However, exploring new commercial ventures, incorporating latest technologies, would be essential to the future of the discipline and industry.

Keywords: CHATBOT, CONTEXTUALISATION, GENERATIVE ARTIFICIAL INTELLIGENCE, INTERFACE, HALLUCINATE, GENAI, INFORMATION NEEDS, INFORMATION TOOLS, MEANING, TEXT RECEPTION

Opsomming: Generatiewe Kunsmatige Intelligensie (GenKI) as inligtings-werktuig vir leksikografiese inligtingsbehoefte. Hierdie artikel fokus op GenKI, kletsbotte en sommige implikasies daarvan vir leksikografie en woordeboekgebruik. Die gebruik van GenKI as 'n inligtingswerktuig wat inligting moet verskaf aan eindgebruikers (lesers) met spesifieke inligtingsbehoefte wanneer hulle 'n teks lees, word bespreek. GenKI kon vroeër inhoud vergelykbaar met dié van 'n woordeboek verskaf, maar kon nie gekontekstualiseerde antwoorde verskaf nie. Dit het verander en die meeste kletsbotte het tans twee modi, naamlik 'n "soekmodus"

en 'n "dink/redeneermodus". Hulle kan hulle verskillende betekenisvoorstelle logies binne die gepaste konteks beredeneer en neig om 'n oplossing aan te bied. Hierdie kenmerk word aan die hand van voorbeelde bespreek, en dit lyk asof kletsbotte nou korrekte kontekstuele betekenis kan verskaf en die keuse logies binne die konteks kan motiveer, gebaseer op hulle kritiese analise en denkvaardighede, alhoewel hulle steeds "hallusineer" as hulle nie 'n antwoord het nie. Hierna word die rol van woordeboeke en leksikograwe in 'n KI-versterkte wêreld bespreek. Die tradisionele opdrag van woordeboeke om die status en geskiedenis van 'n taal te dokumenteer, is steeds 'n belangrike funksie wat aangemoedig moet word, veral in omgewings met beperkte taalhulpbronne. Die verkenning van nuwe kommersiële ondernemings, insluitend die nuutste tegnologieë, is egter noodsaaklik vir die toekoms van die dissipline en die industrie.

Sleutelwoorde: BETEKENIS, GENERATIEWE KUNSMATIGE INTELLIGENSIE, GENKI, HALLUSINEER, INLIGTINGSBEHOETES, INLIGTINGWERKTUIG, KLETSBOT, KONTEKSTUALISERING, KOPPELVLAK, TEKSBEGRIP

1. Introduction

1.1 Background to artificial intelligence

Currently, artificial intelligence (AI) can be found in numerous and diverse environments. Many newspapers, popular tech journals, tech newsletters, academic articles, books, etc. discuss the uses of AI, with widely divergent opinions. On the one hand, there is a very positive approach, arguing that everybody should be AI literate and should actively use AI. On the other hand, there is a negative approach, arguing, *inter alia*, that AI heralds the end of humanity, the end of jobs and that it is bad for the environment. There are also various issues regarding ethics in the work environment, for students, school children and researchers. Problems of bias, plagiarism, privacy, lack of learning and "hallucinations" are contrasted with the apparent usefulness of this tool, even if it is used correctly. Current AI detection software is not able to differentiate unambiguously between AI generated content and authors' own work, as well as between mis-/disinformation, fake news and deepfakes. The potential positive impact of the AI revolution can, however, not be negated.

A significant aspect of GenAI has been the speed of its development. The first release of the first commercial chatbot was at the end of 2022. It was App- or browser-based and since then AI and GenAI have become common tools for users in all walks of life for the satisfaction of multiple types of information needs. It also enabled typical natural language search tasks from the GenAI interface. AI has also been embedded in many application software packages, social media software, system software (e.g., on smartphones such as iPhones and Samsung phones), and it has been embedded in operating systems, browsers, etc. An AI summary is often provided as the default answer to a Google search, presenting an AI Overview, followed by relevant links. Google's latest development, AI Mode, will have a further significant influence on the general AI-based

provision of information to all Google search users. AI Mode is characterized by Germain (2025) as follows:

People use Google Search five trillion times a year — it defines the shape of the internet. AI Mode is a radical departure. Unlike AI Overviews, AI Mode replaces traditional search results altogether. Instead, a chatbot effectively creates a miniature article to answer your question. As you read this, AI Mode is rolling out to users in the US, appearing as a button on the search engine and the company's app. It's optional for now, but Google's head of Search, Liz Reid, said it plainly when launching the tool: "This is the future of Google Search." [...] There is little doubt AI Mode is an impressive piece of technology. It deploys a "fan out method" where the AI breaks your question into subtopics and does multiple searches simultaneously. Google says this lets AI Mode recommend more diverse sources, produce deeper answers to more complex queries, dives deeper — and you have the ability to ask follow-up questions.

1.2 Terminology issues

When discussing AI, one has to be aware of some potential terminology problems. Industry tends not to specify the exact nature of the different AI developments and usages, but rather only employs the generic term *AI* to refer to all these developments. One should, however, strictly and clearly differentiate between:

- AI as an umbrella term for multiple technologies, such as NLP (natural language processing), neural networks, machine learning, deep learning, and symbolic reasoning.
- GenAI, based on large language models (LLMs)
- Agentive AI, which refers to autonomous agents
- General Purpose AI, Artificial General Intelligence (AGI), currently still an ill-defined industry term for the point at which human intellect is outpaced by algorithms
- Hybrid AI systems which combine different types of AI techniques to create more powerful and versatile AI solutions

The focus in this paper is on GenAI, chatbots and some implications for lexicography and dictionary use, with special reference to the end-users (readers) who have a specific information need when reading a text, i.e., a text reception information need.

Since the recent emergence of GenAI, its development has been phenomenal. This can be seen in the reliability of responses when using chatbots. Chatbots originally tended to "hallucinate", but the use of much larger training databases resulted in chatbots becoming more accurate. In this development, multiple iterations of chatbots from multiple software companies have been released. In addition, specialised chatbots for different environments have been developed, e.g., for general queries, coding, for research, text summarization and image, video and audio generation. Of major significance is the recent devel-

opment of two different modes, viz., to be the equivalent of a search engine, and the ability to "think" and to "reason," i.e., to argue about the validity of its answers.

1.3 AI and chatbots in lexicography

AI and chatbots have also been a focus in recent lexicographic research. The focus has mostly been on possibilities for dictionary compilation and the preparation of writing assistants. A prominent feature in modern-day lexicography is the user-perspective, postulating that "the dictionary user must be an important consideration in dictionary-making" (Hartmann 1989: 102). The main implication of the user-perspective is that the compilation of any dictionary should be preceded by the identification of the envisaged target user, the lexicographic needs of this user and their reference skills. However, this is often neglected. Wiegand (1977: 62) already urged that lexicography should strive towards a much closer relationship with the intended target user, and that this user should no longer remain the familiar stranger. Unfortunately, the user has yet again been neglected in recent developments regarding AI and chatbots in lexicography. Too little attention has been directed at the users and their needs and expectations.

Some recent research relevant to AI, lexicography and the linking between text, dictionaries and other information tools can be found in Bothma and Gouws (2022), De Schryver (2023), Bothma and Fourie (2024), Huete-García and Tarp (2024), Lew (2024), Fuertes-Olivera (2024), and Bothma and Fourie (2025).

1.4 Lack of context

Research directed at the use of AI and chatbots in lexicography has shown that GenAI could already provide content similar to a dictionary, but that it cannot provide contextualised answers. Consequently, the reader is still dependent on their own evaluation of the GenAI-provided content to determine the meaning of the word or phrase in context. This is to the detriment of the reader because even when sufficient context is provided in the prompt, the chatbot often provides only a single meaning/sense of a given word. The chatbot sometimes misunderstands the context provided in the prompt and could then easily provide an incorrect meaning. If then queried through a follow-up prompt why it chose a specific meaning, it could not provide any explanation. This causes doubt in the mind of the user and diminishes the user-friendliness and usefulness of the chatbot.

2. Empirical work

In this section, we provide a brief outline of our experimental methodology and the prompts that were used in the experiments.

2.1 Methodology

The experiments were done during March and April 2025. It was evidently not feasible to test and evaluate all chatbots, as the purpose of this paper is to show tendencies and development of chatbot capabilities (compared to 2024). The selection of GPT o3-mini was based on an AI Overview of the different versions of ChatGPT in Google that stated: "o3 and o3-mini: These are reasoning models designed for solving complex problems, especially in domains like research, strategy, coding, math, and science. They are good for structured, logic-driven thinking and layered reasoning" (March 2025). The choice for DeepSeek was because it was recently released, and according to popular reviews, was as powerful as or even more so than OpenAI's ChatGPT and other equivalents.

The choice of examples was based on previous experiences of the first author of this article with ChatGPT-3.5 and 4 during 2024, in which the chatbot "hallucinated" (i.e., provided incorrect answers), and when it was pointed out that the answers were incorrect, it could not provide any rationale or reasoning for those answers. In some of the current experiments, we used modified versions of two of the prompts of the 2024 experiments, supplemented with further examples to test the scope of the chatbots' ability to analyse prompts and reason.

In the examples and analyses below, GPT o3-mini provided the full analysis directly in its answer. DeepSeek, however, first provided its full thinking process (in a somewhat lighter font, but still easily accessible to the user), and then gave its detailed answer, based on its preceding thinking processes. Both chatbots also provided the number of seconds they "thought/reasoned" about a problem, in the current examples between 28 and 106 seconds.

2.2 Prompts

The full prompts we used to interact with the chatbots are given below and discussed in detail in Sections 2 and 3 of this article.

2.2.1 Prompt 1

"When asked about the meaning of the word 'Barbarossa' in the quoted text below, ChatGPT earlier suggested that it referred to the German army and their Barbarossa campaign in WWII. This was evidently wrong. The text reads: "'The Redbeard's German army approaches our northern borders and he offers me twenty thousand dinars!' He threw the scroll aside, where Imad al-Din picked it up and gazed at it earnestly, as if close scrutiny might reveal some previously concealed zeros. The German army. At the door of the war tent, Malek's heart dropped like a stone. Amid the glee of their recent successes, he had forgotten about Barbarossa and his advancing horde." If you disagree with ChatGPT's earlier explanation, why do you do so? What would be the correct meaning and why do you say so? Please provide references.

Prompt 1 tested whether the chatbot could analyse the example in its historical context, and identify a plausible historical context based on the available keywords. It also tested whether it could argue why one of the potential historical contexts is more likely than the other.

2.2.2 Prompt 2

"When asked about the meaning of the phrase 'Timothy's Book' in the quoted text below, ChatGPT earlier suggested that it referred to 'the first letter of Paul to Timothy, commonly known as 1 Timothy'. This was evidently wrong. The text reads: 'He always paused at the tomb of Prior Philip. It was the largest monument there. Philip, a twelfth-century monk, was a figure of legend, though not much was known about him. According to Timothy's Book, a history of the cathedral started in the Middle Ages and added to later, Philip had organized the rebuilding of the cathedral after it was destroyed in a fire.' If you disagree with ChatGPT's earlier explanation, why do you do so? What would be the correct meaning and why do you say so? Please provide references."

Prompt 2 tested whether the chatbot could analyse the example in a fictional historical context and argue from the formulation in the text of the novel, why the previous answer was illogical.

2.2.3 Prompt 3

"What are the different meanings of the word 'level'? Provide example sentences to illustrate the different meanings and include references."

With this prompt, we tested to what extent the chatbot could generate a valid dictionary article (or, at least, the outline of such an article) and illustrate each meaning with example sentences.

2.2.4 Prompt 4

"Can 'borrow' be used as a double transitive verb?" and "'Borrow' is often used as a double transitive verb in South African English. Is this an acceptable variant, or simply wrong?"

This prompt tested both senses and syntactic features of a word, and we tried to establish whether the chatbot could distinguish between regional variations in the use of words. For this example, we used Google AI overview (based on Gemini).

2.2.5 Prompt 5

"Provide a list of GenAI policies for research, teaching and learning in iSchools across the world (with references) and summarise the content of these policies"

and "Please provide a similar list and summaries for iSchools in the Global South, again with examples and references."

This is evidently not a text reception or purely lexicographic example. However, it does establish to what extent the chatbot is able to analyse a number of different, and typically unrelated, concepts to formulate a correct answer.

2.2.6 Prompt 6

"The sentence 'Time flies like an arrow' has multiple syntactic and semantic interpretations. List the different interpretations with, in each case, an analysis of the part of speech of each word in the sentence, together with references. Which is the most likely interpretation in your opinion?"

This is a complex problem with potentially multiple interpretations due to the ambiguity of the sentence. It requires the chatbot to use the potential meanings and morphological analyses of all the words in the sentence, based on linguistic and lexicographic criteria, to establish the most probable meaning of the ambiguous sentence.

Full details of the GPT o3-mini and Deepseek answers for this prompt are provided in the Addendum, available online (click [here](#)).

2.3 Examples of human-like communication

It is well known that chatbots communicate with users in narrative format, emulating human language characteristics. In this section, we provide examples of such characteristics under different headings.

2.3.1 Informal communication

Both chatbots in the experiments sometimes used informal language to create the impression of human communication, such as in the fairly extensive use of words such as "alright," "okay," "right," by both chatbots, as well as "Interesting!" and "Absolutely!" by DeepSeek.

2.3.2 Analysing prompts

The chatbots provided the users with an idea of how they went about analysing the problem and arriving at an answer, as in the following examples:

"Alright, let's break this down. [...] I need context from historical references [...]," "I initially thought about the WWII context, but no, that doesn't seem to fit," "This makes sense given the historical context of the narrative. I think I misunderstood earlier," "I need to provide sources," "I could refer to general works [...], but I must avoid fabricating references," "I remember several possi-

ble interpretations," "The sentence 'Time flies like an arrow' is famously ambiguous! Let's break it down." (GPT o3-mini)

"Okay, let's tackle this question [...] So I need to figure out why that answer was wrong and what the correct interpretation is," "Let me recall," "Let me read the quoted text again carefully," "The key here is [...]," "I need to explain [...]," "I need to structure all these aspects without overwhelming the user," "Let me make sure I cover every possible angle here." (DeepSeek)

2.3.3 Use of humour

DeepSeek sometimes included humorous statements in its answers, in the following examples:

"Buckle up — this will be a granular journey. 🤖," "Let me know if you'd like to venture into string theory next ... or maybe just another sentence! 😊" (DeepSeek)

No such examples were observed in the current tests with GPT o3-mini.

2.3.4 Identifying contextual keywords

Chatbots identified the relevant keywords correctly, and used this in their analyses to substantiate their solutions, as in the following examples:

2.3.4.1 Prompt 1, the Barbarossa example

GPT o3-mini identified the relevant keywords correctly: "The text uses terms like 'dinar,' and names such as 'Imad al-Din' and 'Malek,' and it depicts scenes like a war tent — elements that point to a medieval or early modern Islamic context rather than the modern era."

DeepSeek also identified all the relevant keywords: "References to dinars (a historical currency), Arabic names (Imad al-Din, Malek), and a 'war tent' suggest a medieval Middle Eastern context, likely during the Crusades."

2.3.4.2 Prompt 2, the Book of Timothy example:

GPT o3-mini stated: "The passage explains that 'Timothy's Book' is a source stating that 'a history of the cathedral started in the Middle Ages ...' and goes on to recount that Prior Philip 'had organized the rebuilding of the cathedral after it was destroyed in a fire.' This language is typical of a chronicle or local historical account, not of a letter offering pastoral guidance."

DeepSeek stated: "The context is about Prior Philip, a 12th-century monk, and his role in rebuilding a cathedral after a fire. The text mentions 'Timothy's Book' as a history of the cathedral started in the Middle Ages and added to

later. So, this book is a historical record of the cathedral's history, begun in medieval times and expanded over time."

2.3.5 Lexicographic and linguistic terminology

Chatbots could contextualise prompts in the specific subject/discipline, in the following examples from Prompt 6, "Time flies like an arrow." Both bots could provide the correct parts of speech for all five words in the sentence for all possible interpretations they listed. They also used relevant grammatical terminology, such as "plural," "imperative," "3rd person singular present" etc.

DeepSeek provided a fairly detailed answer in the first iteration (answer 1), but each time suggests to the user that it could give further analyses (answers 2–4), in which it analysed the following for each of the possible interpretations:

In answer 1, for each of the three interpretations, it provided Analysis, Meaning, Structure (Syntactic Tree) and a Reference;

In answer 2, the Syntactic Structure (Sentence Type and Phrase Structure), Semantic Analysis (Meaning and Plausibility);

In answer 3, Parts of Speech & Syntactic Roles (including Syntactic Tree (X-Bar Theory)), Semantic Analysis (Formal Semantics) and Pragmatic Factors;

In answer 4, it summarised the topics, viz., "phonetics, morphology, formal syntax/semantics, neurolinguistics, and even prosody." It then provided the following detail: Phonetic & Morphological Breakdown (Phonetic Transcription (IPA) and Morphology), Syntactic Analysis via Minimalist Program, Interpretation (Derivation and Features), Formal Semantics with Lambda Calculus (logical form), and, very briefly, Pragmatics & Gricean Maxims (Maxim of Relevance and Maxim of Manner), Psycholinguistic Processing (Eye-Tracking Predictions), Neurolinguistic Correlates, Computational Parsing Models, Historical & Cross-Linguistic Analysis, Prosodic Disambiguation, Quantum Linguistics (with a question mark and a smiley) and Final Conclusion at the Planck Scale. It concluded with the remark: "Let me know if you'd like to venture into string theory next ... or maybe just another sentence! 😊"

2.3.6 References

In each of the prompts, we requested the chatbot to provide references. We checked that the references actually existed and were relevant; in contrast with earlier bots, they did not "hallucinate," and the references existed. In some cases, the references were fairly broad and did not refer to the actual example, but provided more detail about the "setting" of their answer in a broader context, as is evident from the analyses of prompt 6, where the chatbots provided references to a number of well-known linguistics textbooks, or historical reference sources, in the case of prompts 1 and 2 (DeepSeek); GPT o3-mini tended to use more popular sources, e.g., to *Wikipedia* (in the case of prompt 1), and *Encyclopaedia Britannica* (as well as a specialist history book) (for prompt 2).

For the less successful (prompts 3 and 4) and unsuccessful (prompt 5) examples, there were some irrelevant and "hallucinated" sources, as detailed in Sections 3.2 and 3.3.

3. Discussion of examples

In this section, we discuss the successful, less successful and unsuccessful examples of chatbot consultations.

3.1 Successful

Both chatbots dealt successfully with prompts 1, 2 and 6.

Both chatbots were remarkably successful in analysing prompts 1 and 2. Both were able to identify relevant keywords in both examples and use these keywords to guide their interpretations. They argued convincingly that the keywords in prompt 1 pointed to a medieval setting, and not to a modern setting, which excluded the Barbarossa campaign of World War II. In prompt 2, both pointed out that the context of the Book of Timothy excluded a biblical interpretation, viz. a reference to the Pauline letter of 1 Timothy. For both prompts, both chatbots were therefore able to place the texts in a correct period and point out that the later interpretation (World War II) and earlier interpretation (a biblical text) were contextually not correct.

3.1.1 Prompt 1

GPT o3-mini provided the following summary answer for prompt 1:

"The earlier explanation was incorrect because it ignores several key contextual clues in the quoted text. Here's why:

"1. Historical Setting:

"The text uses terms like "dinar," and names such as "Imad al-Din" and "Malek," and it depicts scenes like a war tent — elements that point to a medieval or early modern Islamic context rather than the modern era. In contrast, Operation Barbarossa (the German invasion of the Soviet Union in 1941) is a 20th century event, making that explanation anachronistic.

"2. Meaning of "Barbarossa":

"The term "Barbarossa" comes from the Italian *barba rossa*, meaning "red beard." This nickname was historically associated with figures such as Frederick Barbarossa, the 12th century Holy Roman Emperor known for his red beard. In literary or historical texts set in medieval times, using "Barbarossa" would more likely evoke the memory or legend of such a figure rather than referring to a modern military campaign.

"3. Narrative Distinction:

Notice that the text separately mentions "The Redbeard's German army" and later "Barbarossa and his advancing horde." This implies that the invading force led by Barbarossa is a distinct entity from the other German army mentioned. In a WWII context, "Barbarossa" is synonymous with the Nazi operation, not as a separate leader or faction. The separation here reinforces that the name is being used in a medieval sense — as a proper noun evoking a legendary invader.

"In summary:

"Given the language and imagery, the correct interpretation is that "Barbarossa" in this text is a reference to a medieval invader — likely evoking the historical persona of Frederick Barbarossa (or a similarly legendary figure) — and not the WWII campaign. This interpretation fits the cultural and historical setting of the narrative much better than the modern association.

"For more details on the historical figures and events associated with the name, see:

- [Frederick Barbarossa – Wikipedia](#)
- [Operation Barbarossa – Wikipedia](#)"

3.1.2 Prompt 2

DeepSeek provided the following summary answer for prompt 2:

"The phrase "Timothy's Book" in the quoted text does **not** refer to the biblical First Epistle of Paul to Timothy (1 Timothy). Here's why:

"1. Contextual Discrepancy:

- The text explicitly describes "Timothy's Book" as "a history of the cathedral started in the Middle Ages and added to later."
- **1 Timothy**, by contrast, is a New Testament letter written in the **1st century AD** addressing pastoral and ecclesiastical issues. It has no connection to medieval cathedral histories or chronicles.

"2. Fictional or In-Universe Reference:

- The passage appears to be from **Ken Follett's *The Pillars of the Earth***, a historical novel centered on the construction of a medieval cathedral. In this context, "Timothy's Book" is a fictional chronicle created by Follett to serve as an in-universe historical record.
- The novel often references fictional documents to ground its storytelling in a sense of historical authenticity. For example, Prior Philip (a key character) is depicted as a legendary figure whose deeds are recorded in this fictional "Timothy's Book," which spans generations of cathedral history.

"3. Why ChatGPT's Initial Answer Was Incorrect:

- ChatGPT likely conflated the name "Timothy" with the biblical epistle, missing the **medieval and fictional context** of the text. The mention

of a cathedral's history being compiled over centuries (post-12th century) rules out a 1st-century biblical text.

"4. References:

- In *The Pillars of the Earth*, Follett uses fictional chronicles and manuscripts to enrich the narrative. "Timothy's Book" is one such device, akin to the *Kingsbridge Chronicles* (the series to which the novel belongs).
- The novel itself is set in the 12th century, and the fire/reconstruction of the cathedral aligns with historical events like the real-life destruction of medieval churches (e.g., the 1174 Canterbury Cathedral fire).

"Conclusion:

"'Timothy's Book' is **not a biblical text** but a fictional medieval chronicle within the novel's universe, documenting the cathedral's history. This distinction highlights the importance of **context** when interpreting references in historical or literary works."

It is evident that the reasoning and argumentation are similar to the thinking and reasoning processes of a human who is interested in the historical details of the text.

3.1.3 Prompt 6

The discussions of Prompt 6, about the different meanings of the sentence "Time flies like an arrow," were equally very successful. Both bots analysed the sentences in detail, providing part of speech, syntactic and semantic analyses. Both also provided three different possible interpretations, adapting the part of speech, syntactic and semantic analyses to reflect the specific potential interpretation. They furthermore provided a motivation for their specific choice and motivated this in detail. Both chatbots provided a number of relevant references, all of which existed (i.e., no "hallucinations").

The full answers of both chatbots are provided in the Addendum, available online (click [here](#)), and make for quite interesting reading.

GPT o3-mini introduced and discussed its reasoning process as follows:

"The sentence 'Time flies like an arrow' is famously ambiguous! Let's break it down.

"Interpretation 1: It can mean that time quickly passes, much like an arrow's flight.

"Interpretation 2: 'Time flies' might refer to insects (as a noun), with 'like an arrow' being a comparison.

"There are multiple ways to interpret the parts of speech here: 'flies' as a noun (the insects) or verb (the action), and 'like' as a comparison. It's a classic case of ambiguity.

"I remember several possible interpretations."

This is followed with a detailed discussion of more than 800 words. For a full transcript, see the Addendum, available online (click [here](#)).

At the end of each answer for this prompt, DeepSeek asked whether further, more detailed analyses were required, and provided further details that may potentially be relevant for a specialist, or, citing some of the potential motivations and contents.

DeepSeek introduced and discussed its reasoning process as follows:

Answer 1:

"Okay, so I need to figure out the different interpretations of the sentence 'Time flies like an arrow.' I remember that this sentence is often used as an example of syntactic ambiguity. Let me start by breaking down each word and considering their possible parts of speech;" "First, the straightforward interpretation is probably 'Time passes quickly, similar to how an arrow flies.';" "But the user mentioned there are multiple interpretations. Let me think about other possible ways to parse this;" "Wait, another possibility ...;" "Wait, this is getting complicated;" "But that's a stretch ...;" "I need to make sure I'm not missing any other interpretations. Let me think again."

Answer 2:

"The user might be interested in the linguistic theories behind ambiguity. Chomsky's deep vs. surface structure comes into play here. The same surface sentence has different deep structures based on parsing. Radford's analysis of syntactic ambiguity would support this. Also, mentioning lexical ambiguity with homonyms is important;" "Also, consider the user's possible deeper needs. They might be studying linguistics, preparing for an exam, or just curious about language. By elaborating, I should provide enough detail to satisfy academic curiosity but keep it understandable. Avoid jargon where possible, or define terms if used," "Let's delve deeper into the syntactic and semantic ambiguities of 'Time flies like an arrow, expanding on linguistic theories, structural analyses, and contextual plausibility."

Answer 3:

"Okay, the user has already asked for a more granular analysis after I provided a detailed breakdown. They seem really interested in the linguistic aspects of the sentence 'Time flies like an arrow. Let me make sure I cover every possible angle here," and "Also, are there other possible interpretations I haven't considered? Maybe 'like' as a conjunction, though that's less likely. Or 'arrow' as a modifier? Probably not. But I should check;" "I need to make sure each analysis is thorough, referencing linguistic theories, providing syntactic trees with proper notation, and discussing semantic plausibility. Also, check if there are any other references or studies that

discuss this specific sentence — maybe classic examples in ambiguity literature."

Answer 4:

"Let me think about how to go deeper. First, maybe I should look into phonetics and morphology. But the sentence is in English, and the words are straightforward." It then briefly mentions "homonyms and their phonetic representations," "syntactic theories beyond X-Bar. Maybe using Minimalist Program structures, discussing movement or feature checking;" "Semantic roles could be expanded. Theta roles were mentioned;" "Pragmatic aspects could include Gricean maxims in more depth;" "computational linguistics;" "psycholinguistics;" "historical linguistics;" "morphological analysis;" "lexical semantics;" "neurolinguistics;" "prosody." It concludes its reasoning as follows: "I need to structure all these aspects without overwhelming the user. Organize them into sections with clear headings, use linguistic notation where applicable, and reference theories and studies for each point. Ensure that each detail adds value and connects back to the main interpretations. Check for coherence and avoid redundancy from previous answers. Highlight how each tiny detail contributes to the overall understanding of the ambiguity."

The four answers of DeepSeek result in a detailed discussion of approximately 6,000 words. For a full transcript, see the Addendum, available online (click [here](#)).

3.2 Partially successful

The answers to prompts 3 and 4 were less successful, primarily due to poorly structured groupings and "hallucinations" in the answer to prompt 3, and inconsistencies in the answers to prompt 4.

3.2.1 Prompt 3

In prompt 3, the chatbots were required to construct the equivalent of a dictionary article for the meanings/senses of a single word, *level*, as well as provide references.

GPT o3-mini introduced its answer as follows: "The word 'level' is a versatile term in English, used in different contexts with varying meanings. Below are the main meanings of 'level', each with examples and references." In each case, an example paraphrase or sentence from a dictionary is provided, in this case, the *Oxford English Dictionary* (OED), *Merriam-Webster*, *Cambridge Dictionary*, *Collins English Dictionary* and *Longman Dictionary*. The examples are in all cases provided in quotation marks, which creates the impression that they are taken verbatim from the specified dictionary, but in none of the examples this

is the case — chatbot in each case provided a paraphrase. In addition, the proper dependencies between meanings/senses (as is evident in all the cited dictionaries) were not maintained. In this case, the chatbot is therefore a less dependent source of meanings/senses.

The chatbot also offered the option that it could provide examples from only one specific dictionary: "Let me know if you'd like definitions from a specific dictionary like OED or Merriam-Webster with direct quotes." We asked it to provide an answer from only the OED. None of the examples in this answer, however, were direct quotes from the OED — unless it referred to one of the other dictionaries of English published by Oxford, and not the "*The historical English dictionary*," typically known as "*Oxford English Dictionary*," and abbreviated as "OED" (<https://www.oed.com/>).

This is a less successful answer than could be obtained by typing "level meaning" in a Google search bar, which provides a comprehensive article with examples, synonyms etc.

3.2.2 Prompt 4

Prompt 4 (a query in the Google search bar, the answer of which is an AI Overview based on Gemini) is included to highlight confusion which sometimes occurs in GenAI answers. We asked whether the verb "borrow" could be used as a double transitive verb. In a first answer, it said: "Yes, the verb 'borrow' can be used as a double transitive verb, also known as a ditransitive verb." The rest of the explanation and examples seemed counterintuitive. We therefore repeated the question about an hour later (to check consistency in its responses), and the answer it gave was: "No, the verb 'borrow' is not typically used as a double transitive verb. While 'borrow' does take a direct object (what is being borrowed), it doesn't usually take an indirect object (to whom or for whom)." Since "borrow" is often used as a double transitive verb in (informal) South African English (i.e., as an equivalent of "lend"), we asked whether this usage is an acceptable variant in South African English, or simply a mistake. The answer read as follows: "While the use of 'borrow' as a double transitive verb (e.g., 'Can I borrow you the car?') is common in South African English, it's generally considered a nonstandard usage in formal English. It's often a loan translation from Afrikaans, where 'leen' (lend) and 'leen by' (borrow from) are used in a way that translates directly to 'borrow' in the double transitive construction. While it's widely understood and accepted in informal contexts, it's best avoided in formal writing or speech." This is the expected answer (which may confirm the authors' biases). Interestingly, this possible meaning is not listed in the *Dictionary of South African English* (DSAE, <https://www.dsae.co.za/>). The correct/incorrect reply from Google Overview AI highlights the importance of evaluating any answer GenAI provides, and of triangulating it, especially if it is counterintuitive.

3.3 Unsuccessful

Prompt 5 is not a lexicographic example, and is included to show that, even though chatbots have improved considerably since earlier versions, they still hallucinate if their training data does not contain the required data to answer the question.

3.3.1 Prompt 5

We used two prompts for this question. In the first prompt, we required the chatbot, in this case, DeepSeek, to be able to identify the different iSchools in the world (see the iSchools website at <https://www.ischools.org/>) and then find on the websites of the iSchools any policies for the use of GenAI in research, teaching and learning. The answer DeepSeek provided seemed to be correct, but very general. It provided references from 2023, however, without URLs to enable the verification of the existence of the sources, or more information. Some of these documents could be found, but in most cases, documents with the titles that were listed, were not available. It also provided a number of references under the heading "Hypothetical References," which is very strange.

We also needed information on the same topic, restricted to iSchools in countries in the Global South. This would imply that DeepSeek should be able to understand which countries the Global South consisted of, to match this list with the list of iSchools worldwide and then find the relevant policies. It was unfortunately not successful at all. We knew that, at that stage, there were four iSchools in Global South countries, one each in Brazil and Indonesia, and two in Africa. DeepSeek identified six iSchools in Global South countries, five of which were wrong and only one which was correct. It again provided incorrect or non-existent references and again added "Hypothetical References." The discussions were very general and none of the content could be verified.

These two examples again highlight the importance of evaluating and triangulating any answer GenAI provides.

4. GenAI versus dictionaries

The preceding discussion and examples show that GenAI cannot fully replace "conventional" dictionaries — at this stage. Access to actual dictionary sources provides better assistance, e.g., a search in online dictionaries that are available via open access, or even a Google search that provides "Definitions from Oxford Languages," results in retrieving better and more comprehensive definitions, as well as additional lexicographic information on demand. Currently, GenAI cannot be regarded as an authoritative lexicographic source, even when it claims to cite online dictionaries (as in prompts 3 and 4); it can, however, be an additional information tool for more complex information needs, as in the cases of

prompts 1, 2 and 6. In addition, it still "hallucinates" if it does not have an answer, as in the case of prompt 5. The potential unreliability of any GenAI-generated answer again emphasises the importance of evaluating any such answers and triangulating the results with other reputable (or curated) sources.

GenAI-enhanced tools are capable of responding to lexicographic information needs and to combine the obtained data with additional resources, constituting a network of information resources, as is evident in prompts 1, 2 and 6.

Irrespective of the advantages and disadvantages of AI, one has to accept that AI and GenAI are here to stay, at least for the foreseeable future. One cannot ignore these developments and should rather strive to understand what it can and cannot do in order to embrace it intelligently and ethically.

5. Suggestions for lexicographic research

5.1 Traditional lexicographic research

A focus in "traditional" dictionary research has always been, and continues to be, important, as dictionaries can best document and display the status and history of a language. This is witnessed in dictionaries such as the OED (*Oxford English Dictionary*, <https://www.oed.com>), WAT (*Woordeboek van die Afrikaanse Taal*, <https://www.wat.co.za>), the DSAE (*Dictionary of South African English*, <https://www.dsae.co.za>) and many others. These endeavours need to be encouraged, especially in environments with limited language resources. The lexicography of the African languages in South Africa could also benefit from such an approach. The dynamic nature of language demands that dictionary definitions should continuously be refined.

No single dictionary can be everything for everyone, but in the absence of an established typological infrastructure, existing dictionaries should increase their scope beyond the traditional typological boundaries so that they can present a wider linguistic coverage to their target users. One option the lexicographer has, is to plan and compile multifunctional online dictionaries with added functionalities, e.g., language-learning exercises, lexicotainment, etc. They could also offer access to additional lexicographic and non-lexicographic data, and/or links to such data, even to data in dictionary-external sources. This demands the appropriate interfaces and interface design that allow the filtering of data as well as drilling down and retrieving information on demand. Interface design is an exceptionally important aspect for the development of a successful online dictionary, and typically requires a number of specialists to collaborate, including lexicographers, UX (user experience) designers, programmers/computer scientists, information scientists and a number of others, as well as users to test the efficiency of the proposed interface.

When planning the interface design, it would be helpful to be reminded of Ben Shneiderman's Visual Information-Seeking Mantra (1996): "Overview first, zoom and filter, then details on demand." This should be understood as:

- "Overview: Gain an overview of the entire collection"
- "Zoom: Zoom in on items of interest"
- "Filter: Filter out uninteresting items"
- "Details-on-demand: Select an item or group and get details when needed"

To this could be added "Relate", "History" and "Extract", realised as:

"Relate: View relationships among items."

"History: Keep a history of actions to support undo, replay, and progressive refinement."

"Extract: Allow extraction of sub-collections and of the query parameters."

These aspects of interface design will not be discussed in this paper. Existing online dictionaries already offer an interface design that can guide the user within a comprehensive article to both a number of restricted articles, cf. Gouws (2022: 120), in which various aspects of the treatment of the lemma are presented or to dictionary-external data like other dictionaries in the same portal, or even to Google and other search engines. An option that modern-day lexicographers should consider is to integrate access to a chatbot, enabling dictionary users to retrieve complementary information to enhance the quality and extent of their dictionary consultation experience. This integration of GenAI and lexicography requires, among others, the appropriate interface design and data distribution structure.

5.2 Lexicographic research and technology

Lexicographic databases, their compilation and use are focal areas in modern-day lexicographic research. These databases and the lexicographic products relying on them can be improved with more attention paid to the concepts of granularity and (re)usability. Granularity is required for the fine-grain filtering and display of lexicographic data, whereas (re)usability should play a pertinent role in deriving different lexicographic products, including dictionaries for specific user groups from the data on offer. Different applications, e.g. translation tools, writing assistants and tools that help users to improve their writing can employ the same data but applied in different ways.

Lexicographic data becomes the gold standard in the planning and compilation of lexicographic products, and the data assessment, distribution and presentation need to be done in such a way that different target groups can use different tools to retrieve the information they require in specific extra-lexicographic situations. The optimal use of these data can be improved if lexicographers, AI experts and other collaborators can successfully explore and license new commercial ventures, incorporating the latest technologies, including AI and GenAI. Such ventures could also elevate the level of contextualized linking for text reception.

6. Conclusion

Based on the limited number of examples that have been examined, and contrary to earlier research, it seems as if the latest versions of GenAI chatbots can often provide correct contextual meaning. It can logically motivate the choice of meaning in context, based on its critical analysis and thinking skills, typically associated with humans.

However, the reliability of the results of GenAI answers varies, covering the full spectrum from very reliable to complete hallucinations. A mixture of reliability criteria prevails. As a result, the reader remains responsible to evaluate the answers. In case of uncertainty, they need to check other resources in order to verify and triangulate the results.

When expanding the information retrieval structure of a dictionary to allow users access from any point in a given dictionary to one or more chatbots, lexicographers need to make their users aware of the fact that GenAI provides non-curated data that might not always live up to the correctness, appropriateness and validity of the information they can retrieve from the curated data in dictionaries.

However, good lexicography demands interdisciplinary collaboration. An interactive relation between lexicographers and developers of GenAI could be beneficial for both groups. Lexicography could benefit by improved versions of GenAI becoming sophisticated lexicographic information tools and GenAI could benefit from the expertise in lexicography regarding the user-perspective, information needs, lexicographic functions, appropriate contents and relevant structures to distribute and access data.

7. Addendum

The full answers of both chatbots for prompt 6 are provided in the Addendum, available online (click [here](#)).

8. References

- Bothma, T.J.D. and I. Fourie.** 2024. Enhancing Conceptualisations of Information Behaviour Contexts through Insights from Research on e-Dictionaries and e-Lexicography. *Information Research. An International Electronic Journal* 29(2): 179-197.
<https://publicera.kb.se/ir/article/view/55127> (accessed 15 June 2025)
- Bothma, T.J.D. and I. Fourie.** 2025. Contextualised Dictionary Literacy, Information Literacy, and Information Behaviour in the e-Environment. *Library Management* 46(1-2): 14-28.
<https://doi.org/10.1108/LM-08-2023-0082>
- Bothma, T.J.D. and R.H. Gouws.** 2022. Information Needs and Contextualization in the Consultation Process of Dictionaries that are Linked to e-Texts. *Lexikos* 32(2): 53-81.
<https://doi.org/10.5788/32-2-1697>

- De Schryver, G.-M.** 2023. Generative AI and Lexicography: The Current State of the Art Using ChatGPT. *International Journal of Lexicography* 36(4): 355-387.
<https://doi.org/10.1093/ijl/ecad021>
- Fuertes-Olivera, P.A.** 2024. Making Lexicography Sustainable: Using ChatGPT and Reusing Data for Lexicographic Purposes. *Lexikos* 34: 123-140.
<https://doi.org/10.5788/34-1-1883>
- Germain, T.** 2025. Is Google about to Destroy the Web? *BBC*, 13 June 2025.
<https://www.bbc.com/future/article/20250611-ai-mode-is-google-about-to-change-the-internet-forever> (accessed 15 June 2025)
- Gouws, R.H.** 2022. Extended Article Comments in Online Dictionaries. *Lexicography. Journal of ASIALEX* 9(2): 111-138.
<https://doi.org/10.1558/lexi.24289>
- Hartmann, R.R.K.** 1989. Sociology of the Dictionary User: Hypotheses and Empirical Studies. Hausmann, F.J. et al. (Eds.). 1989. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography*: 102-111. Berlin: De Gruyter.
- Huete-García, Á. and S. Tarp.** 2024. Training an AI-based Writing Assistant for Spanish Learners: The Usefulness of Chatbots and the Indispensability of Human-assisted Intelligence. *Lexikos* 34: 21-40.
<https://doi.org/10.5788/34-1-1862>
- Lew, R.** 2024. Dictionaries and Lexicography in the AI Era. *Humanities and Social Sciences Communications* 11(1): 1-8.
<https://doi.org/10.1057/s41599-024-02889-7>
- Shneiderman, B.** 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages, Boulder, CO*: 336-343.
<https://doi.org/10.1109/VL.1996.545307>
- Wiegand, H.E.** 1977. Nachdenken über Wörterbücher. Aktuelle Probleme. Drosdowski, G., H. Henne and H.E. Wiegand (Eds.). 1977. *Nachdenken über Wörterbücher*: 51-102. Mannheim: Bibliographisches Institut.