

How Effective is AI as a Language Consultant?

Urška Vranjek Ošlak¹

¹ ZRC SAZU, Fran Ramovš Institute of the Slovenian Language
E-mail: urska.vranjek@zrc-sazu.si

Abstract

This paper explores the applicability of generative artificial intelligence in the field of language consulting, focusing on ChatGPT-4 and the Slovenian language. The analysis is based on an experiment involving 30 real user questions submitted to the Language Consulting Service (LCS) of the Fran Ramovš Institute of the Slovenian Language. The questions cover a range of linguistic categories and were submitted to ChatGPT under controlled conditions. The responses were then compared with expert-produced answers and evaluated in terms of factual accuracy, stylistic appropriateness, terminological correctness, and overall usefulness. The results show that while ChatGPT performs well in terms of clarity, tone, and structure, its output often contains inaccuracies and occasionally misleading information. At this stage, ChatGPT is not suitable as a stand-alone tool for end-users. However, it could serve as a helpful draft generator for human language consultants. The study also outlines ways to improve AI output, including better prompts and access to relevant databases. Although some fundamental limitations of AI remain, its controlled use in language consulting may offer practical support, especially in cases involving repetitive or less complex queries.

Keywords: generative AI; language consulting; applied linguistics; Slovenian language

1. Introduction

With the rapid advancement of artificial intelligence, its role in linguistics and language-related activities is becoming increasingly significant. The effectiveness of generative AI models, such as ChatGPT-4, in linguistic applications has been examined in numerous studies since the launch of this technology, particularly in lexicography and English as a Foreign Language (EFL) instruction.

This study explores another linguistic domain in which generative AI may be beneficial: language consulting. Generative AI is widely recognised for its effectiveness in addressing frequently asked questions (FAQs) and topics with extensive coverage. However, how well does it perform in responding to language-related inquiries, and how reliable is it as a language consultant compared to human experts?

This research is based on an experiment carried out on a set of actual language users' questions and it assesses the applicability of AI in language consulting by evaluating its performance across various linguistic categories and comparing its responses to those of human experts. The evaluation presented in the second part of this paper can serve as a base for determining whether ChatGPT could potentially be employed in creating an

AI-based assistant to be used in language consulting, possibly as assistance for both the language user and the language consultant.

2. AI in linguistics¹

Linguists agree that the launch of ChatGPT “marked a significant milestone in the field of artificial intelligence and language technology” and “it represented a paradigm shift in the way digital tools interact with human language” (Ptasznik et al., 2024). The effectiveness of generative AI models, such as ChatGPT-4, in linguistic applications has been examined in numerous studies since the launch of this technology, particularly in lexicography (see, e.g., Jakubíček & Rundell, 2023; an overview and comparative analysis in de Schryver, 2023; for Slovenian see, e.g., Petrič et al., 2024; Kosem et al., 2024) and English as a Foreign Language (EFL) instruction (see, e.g., Belda-Medina & Calvo-Ferrer, 2022; for a comprehensive overview see Ptasznik & Lew, 2025).

It seems the researchers are unanimous about the fact that AI can be useful in certain linguistic contexts and for certain linguistic activities, however, they disagree on the level of usefulness and on the actual amount of the post-editing needed after ChatGPT finishes its part (however, there is no doubt that post-editing is, at this point in time, definitely needed).² Also, there is extensive debate directed towards the generating and improving of prompts as good prompts are vital in getting the best results.

2.1 AI as a language consultant – yes, but what for?

Several European languages have a well-established tradition of language consulting, with Slovenian being a distinguished example.³ The Language Consulting Service (Jezikovna svetovalnica; hereinafter LCS), administered by the Fran Ramovš Institute for the Slovenian Language (ZRC SAZU), has become a central hub for linguistic consultation in Slovenia. It serves a dual purpose: on the one hand, it offers practical guidance to Slovenian speakers seeking authoritative advice on language use; on the other, it provides linguists with valuable insights into the linguistic uncertainties and normative concerns prevalent among users, thereby illuminating potential gaps in existing language descriptions. Since its establishment in 2012, the LCS has transitioned into a fully digital platform and now incorporates a dynamic system that generates an online language consulting database. This database features interactive codes and keywords that facilitate indexing, categorisation, and user interaction (Dobrovoljc et al., 2020; Dobrovoljc & Vranjek Ošlak, 2021). The database of user-

¹ Numerous high-quality articles and studies are available that provide clear and adequate explanations of the functioning of generative AI, and large language models (LLMs) in particular, i.e., the recent and notably coherent account by Vajjala (2024).

² See, e.g., Benko et al., 2024.

³ For a European overview see Ludányi (2020); for a comparative view see, e.g., Mžourková, (2024).

submitted questions and expert-formulated answers continues to grow and has already formed the empirical basis for a number of linguistic studies, primarily in relation to the role of language consulting in different contexts.⁴

In recent years, the increasing volume of queries and the recurrence of thematically or structurally similar questions have raised questions about the long-term scalability and efficiency of exclusively human-mediated language consultation. The LCS regularly receives a high number of queries, many of which involve recurrent or formulaic topics (orthographic conventions, morphological variants, syntactic constructions) suggesting that a portion of the consultancy process could be handled by automated systems. The scale of the dataset, combined with observable patterns of repetition, presents a compelling case for the integration of AI tools such as ChatGPT into this domain.

However, the growing interest in automation⁵ is not merely a matter of reducing the labour intensity of expert work. It also reflects a shift in (both users' and linguists') expectations regarding how linguistic information should be presented and accessed in the digital era (see, e.g., Lew, 2015; de Schryver et al., 2019). As articulated by Arvi Tavast during the round table discussion on ChatGPT at the eLex 2023 conference (as summarised in de Schryver, 2023: 366–367), contemporary language users increasingly favour direct, conversational access to information: “It is simply more natural to ‘just ask’ what something means, rather than to have to open a book (or website) and to have to start looking up (or searching for) a word, to then seek the proper meaning, etc.” His remarks underscore a broader trend towards user-centred interfaces and real-time interaction, which are well aligned with the affordances of generative AI. Furthermore, Tavast emphasised a key distinction in the intended function of such technologies: “People will require answers to questions about language, rather than requiring a dictionary. Hence, these technologies should not be used as part of the lexicographic process, but as part of providing information to the users.” Given that the core function of language consulting is precisely that – providing clear, practical answers to real users' language-related questions – it makes sense to consider how such services might be improved by implementing current AI tools.

Overall, these observations suggest that bringing AI into language consulting responds both to practical challenges, like growing demand and the need for scalable solutions, and to changing user preferences for quick, conversational access to language expertise. For this reason, the use of generative AI models such as ChatGPT in language consultancy deserves careful and systematic investigation.

⁴ See Vranjek Ošlak, 2023, Vranjek Ošlak, 2024. The first contribution describes how language counselling activities help linguists fill in the gaps in codification caused by the inadequacy of current language manuals. The second contribution focuses on the role of language counselling during the revision of the Slovenian normative guide *Pravopis 8.0*, illustrating how user queries inform of and are informed by evolving normative guidelines.

⁵ Present in lexicography as well and for some time already, see Rundell (2023).

3. The “ChatGPT as a language consultant” experiment

3.1 Methodology of the experiment

This study investigates the applicability of generative artificial intelligence, specifically ChatGPT-4 (model GPT-4o), in the domain of language consulting. The evaluation is based on its performance in responding to 30 language-related queries sourced from the Language Consulting Service (LCS) platform. These queries span a range of linguistic categories, including orthography and standardisation, phonology and pronunciation, morphology and word formation, syntax, lexicology and semantics, etymology, and stylistics (while some touched upon multiple linguistic domains, they were classified under a single primary category for analytical clarity).

The questions were collected progressively from the LCS platform between January and June 2025. Only those queries were selected for which no official answers had been published at the time of submission to ChatGPT. This selection criterion aimed to minimise the risk of overlap, as ChatGPT was instructed to consult similar responses on the LCS platform.⁶ Accordingly, questions with publicly available solutions were excluded to ensure a valid assessment of the model’s autonomous problem-solving capabilities.

Before each question (serving as part of the prompt) was posed, the AI was provided with all relevant information that a human language consultant would typically receive, including expectations regarding response characteristics. The AI-generated answers were subsequently compared with expert responses, with the evaluation focusing on factual and general accuracy, compliance with the principles of democratic language consulting,⁷ and the degree to which each response addressed the substance of the inquiry.

The prompt employed in this task was composed of two distinct components: an introductory explanatory section and a subsequent part containing the actual language question. Each prompt was presented in Slovenian. In the first part of the prompt, all contextual information typically available to a human language consultant was provided, including expectations regarding the format, tone, and argumentative depth

⁶ Because web server logs are not accessible to users of ChatGPT’s web interface, the presence of hyperlinks in generated outputs is the only observable indicator that online content was referenced. ChatGPT may also summarize or paraphrase internet content without displaying all the URLs it consulted.

⁷ Democratic language consulting is an approach to linguistic advice that emphasizes educating users and strengthening their confidence, rather than simply prescribing correct forms. Instead of providing a single authoritative answer, it offers explanations of linguistic elements, arguments for and against particular usages, and the social or normative implications of different choices, leaving the final decision, when applicable, to the speaker (Žaucer & Marušič, 2009).

of the answer. Before testing began, it was confirmed (on 13 March 2025) that ChatGPT had access to previously published entries on the LCS platform. The introductory part of the prompt used was as follows:

Assume the role of a language consultant. You respond to questions about the Slovenian language posed by its users. Your answers should conform to the principles of democratic language consulting. Where relevant, they should include a historical explanation of the linguistic phenomenon, comparative linguistic facts, a description of the treatment of the issue in linguistic reference works available on the Fran portal, and a well-reasoned and substantiated response to the specific question. You should also consult previously published answers to related questions on the Jezikovna svetovalnica portal (<https://svetovalnica.zrc-sazu.si/>) and include links to them as supporting material in your response (links to other websites should be used only exceptionally). The answer should emulate the style and structure of responses typically found on the Jezikovna svetovalnica platform. The question for you is as follows: ...⁸

The experiment was carried out through the ChatGPT web interface by repeatedly providing input with the prompt and added language query. Since the questions vary considerably in topic, we did not start a new conversation for each and every question.⁹

The model used was the most recent GPT-4o¹⁰ which is supposed to have improved reasoning and accuracy with faster response times and improved efficiency. It is better at understanding nuanced context, adapting to tone and style, and interacting more naturally across multiple modalities (e.g. text, vision, audio) than previous models (Palma, 2025).¹¹

⁸ In Slovenian: Pretvarjaj se, da si jezikovni svetovalec. Odgovarjaš na jezikovna vprašanja uporabnikov slovenščine. Tvoji odgovori naj bodo v skladu z demokratičnim pristopom k jezikovnemu svetovanju. Kjer je relevantno, naj vsebujejo zgodovinsko razlago jezikovnega pojava, primerjalnojezikovna dejstva, opis stanja v jezikovnih priročnikih na portalu Fran in utemeljen/argumentiran odgovor na konkretno vprašanje. Preveri tudi že objavljene odgovore na vprašanja na portalu Jezikovna svetovalnica (<https://svetovalnica.zrc-sazu.si/>) in v odgovor kot dodatno gradivo vključi povezave nanje (na druga spletna mesta pa le izjemoma). Odgovor naj posnema slog odgovarjanja, kot ga imajo odgovori v Jezikovni svetovalnici. Vprašanje zate se glasi: ...

⁹ It should be noted that by not starting a new conversation for each question, the possibility remains that prior exchanges influenced subsequent responses. This potential effect was not tested in the present study.

¹⁰ This model was available to us through institutional access. We used the web interface, which does not surface specific backend versions or snapshots. Given that the majority of this experiment was performed between May 1 and June 30, the snapshot of GPT-4o running in the background was most likely “chatgpt-4o-latest”. This alias designates whichever GPT-4o snapshot was active at the time, but the exact dated identifiers (e.g., “gpt-4o-2025-05-12”) are not exposed to web users.

¹¹ We asked ChatGPT to argue why the 4o model is the best one to use for this task. ChatGPT claimed that “GPT-4o is presently the most suitable for addressing language-related questions [...] as it surpasses earlier models (GPT-4, GPT-3.5, etc.) in addressing multifaceted linguistic questions”.

3.2 Evaluation and discussion

The following is an example of a user question, presented alongside by both the response generated by ChatGPT and the corresponding human-produced answer.¹² This example serves to illustrate the overall quality of ChatGPT's output, with certain features (where applicable) referenced in the subsequent analysis.

The question asks whether the Slovenian word *cerkev* (meaning 'church') is a general term for any place of worship, or if it refers exclusively to Christian buildings. Specifically, the user wonders if it would be linguistically appropriate to refer to a mosque as *cerkev*. The answer explains that although dictionaries define *cerkev* as a building for religious worship, actual usage in Slovenian almost exclusively applies it to Christian places of worship (Catholic, Orthodox, Protestant). Therefore, calling a mosque or synagogue a *cerkev* would be linguistically incorrect and culturally misleading. Other religions have distinct terms: *džamija* (meaning 'mosque'), *sinagoga* (meaning 'synagogue'), etc. This pattern is consistent across many European languages. To remain neutral or general, terms like *bogoslužna stavba* (meaning 'place of worship'), *verski objekt* (meaning 'religious building'), or *svetišče* (meaning 'sanctuary') are recommended.

Question
Zdravo, zanima me, ali je izraz cerkev splošno poimenovanje oziroma nadpomenka za versko stavbo, namenjeno bogoslužju, ali se uporablja izključno za krščanske stavbe? Na primer, ali bi bilo jezikovno ustrezno džamijo poimenovati cerkev? Hvala
ChatGPT's response
Zdravo, vprašanje o pomenskem obsegu besede cerkev odpira zanimivo razmerje med pomensko razširjenostjo splošnih poimenovanj ter kulturnozgodovinsko zaznamovanostjo posameznih izrazov. V odgovoru pojasnjujem razliko med splošnim pomenom , rabo v slovenščini in možnostmi preimenovanja verskih stavb v drugih verah.
1. ZGODOVINSKO IN SODOBNO POMENJENJE BESEDE CERKEV Beseda cerkev ima v sodobni rabi več pomenov , kot kaže tudi SSKJ2 ¹³ : cerkev -ve ž (ę) 1) zgradba za opravljanje verskih obredov: nova cerkev, gotška cerkev // baročna cerkev; iti v cerkev 2) verska skupnost s svojo organizacijo: pravoslavna, evangeličanska, katoliška cerkev 3) navadno z veliko začetnico: Rimskokatoliška cerkev

¹² The footnotes in the example clarify the cited internet sources and were added by the author of this article.

¹³ The link directs to the Fran dictionary portal, specifically to all dictionary entries listed under the headword *cerkev*.

Z vidika **definicije pod 1** bi torej lahko domnevali, da gre za **splošno poimenovanje za versko zgradbo**. Vendar raba v slovenščini kaže drugače.

2. RABA IN JEZIKOVNA ZAZNAMOVANOST

Čeprav definicija *cerkve* kot »zgradbe za opravljanje verskih obredov« zveni nevtrarno, se v **dejanski rabi nanaša skoraj izključno na krščanske verske zgradbe** (katoliške, pravoslavne, protestantske). S tem je izraz **kulturno in versko zaznamovan**. Temu ustrezno:

- muslimanske bogoslužne stavbe poimenujemo **džamija** (ali mošeja),
- judovske stavbe **sinagoga**,
- budistične **pagoda**, **tempelj** ali **samostan**, odvisno od konteksta.

Na portalu Jezikovna svetovalnica je to pojasnjeno podobno:

»Beseda *cerkev* se uporablja skoraj izključno za krščanske verske zgradbe. Za druge veroizpovedi se uporabljajo posebna poimenovanja (npr. *džamija*, *sinagoga*).«

👉 Vir: <https://svetovalnica.zrc-sazu.si/topic/1721/uporaba-besede-cerkev>¹⁴

3. ALI LAHKO DŽAMIJO POIMENUJEMO CERKEV?

Z **jezikovnega vidika** bi to pomenilo:

- da bi izraz *cerkev* uporabljali kot **nadpomenko** za vsa bogoslužna poslopja;
- v praksi pa bi to **povzročilo pomensko nejasnost** ali **kulturno neskladnost**, saj bi bila takšna raba v nasprotju z uveljavljenim jezikovnim sistemom.

Če torej v slovenščini rečemo:

V vasi so tri cerkve, ena džamija in sinagoga,

s tem jasno pokažemo, da **cerkev ni nevtralna nadpomenka**, temveč **specifična oznaka za krščansko svetišče**.

V jeziku obstajajo sicer primeri, kjer se **eno poimenovanje posploši** (npr. *zdravnik* kot naziv za različne specialiste), vendar se to pri *cerkvi* **ni uveljavilo**, ravno zaradi **povezanosti z določeno religijsko tradicijo**.

4. PRIMERJALNOJEZIKOVNI POGLED

V številnih drugih jezikih imajo prav tako **posebna poimenovanja za verske zgradbe** različnih religij:

- angleško: *church* (krščanstvo), *mosque*, *synagogue*;
- nemško: *Kirche*, *Moschee*, *Synagoge*;
- italijansko: *chiesa*, *moschea*, *sinagoga*.

To kaže na **skladnost v jezikovni tipiki evropskih jezikov**, ki razlikujejo med sakralnimi stavbami po **religijski pripadnosti**.

5. SKLEP

Beseda **cerkev** ima sicer slovarsko definicijo kot zgradba za bogoslužje, vendar se v **dejanski rabi nanaša skoraj izključno na krščanske verske stavbe**. Zato **ni jezikovno ustrezno**, da bi **džamijo** ali **sinagogo** poimenovali kot **cerkev** – to bi bilo **pomensko**

¹⁴ The link leads to an unrelated topic in the LCS.

zavajajoče in kulturno netočno.

Za **splošno poimenovanje** vseh verskih zgradb lahko uporabimo besedno zvezo, kot je:

- *bogoslužna stavba*,
- *verski objekt* ali
- *svetišče*, če želimo ostati nevtralni in nadpomenčni.

Če imate dodatna vprašanja o pomenu besede *cerkev* v drugih pomenih (npr. institucija, občestvo, organizacija), jih z veseljem pojasnim.

Lep pozdrav!

Jezikovni svetovalec

Table 1: An example of a user question and the corresponding response generated by ChatGPT

Human-produced response¹⁵

Ustrezna nadpomenka je *svetišče*¹⁶, ki ima širok, različna verstva zajemajoč pomen 'stavba, prostor, namenjen za verske obrede', npr.

- *Morale so se zakriti, ker drugače ne bi smele stopiti v **muslimansko svetišče**.*
- *Vidimo lahko Omarjevo mošejo in mošejo Al Aksa, dve zelo čaščeni **muslimanski svetišči**.*
- *Ne pozabite, da je v Jeruzalemu tretje najsvetejše **islamsko svetišče**.*
- *Ko sem bil v Tokiu, sem se odločil, da obiščem **šintoistično svetišče**.*
- *Videti je mogoče samo nekaj odsekov in sinagogo Nozyk, edino **judovsko svetišče** v Varšavi, ki je preživelo okupacijo, ker so ga uporabljali kot hlev.*

Pomen besede *cerkev*¹⁷ je medtem ožji: 'stavba, namenjena za krščansko bogoslužje'.

Table 2: The human-produced response to the question in Table 1

3.2.1 Stylistic appropriateness and alignment of linguistic attitudes

ChatGPT generates output that, in general, closely mimics the style of responses found in the LCS. The answers are democratic, stylistically appropriate, and polite. They contain various forms of inter- and intralinguistic comparative information. The length of the answers varies: some are approximately 200 words long, while the longest extend to around 700 words, which is significantly more than the average human-produced response in the LCS (typically between 300 and 400 words). The output is clearly structured, making the responses resemble popular scientific writing: explanatory in tone and relatively easy to understand. In rare cases, ChatGPT may even outperform

¹⁵ Krvina & Vranjek Ošlak (2025).

¹⁶ The link directs to the Fran dictionary portal, specifically to all dictionary entries listed under the headword *svetišče*.

¹⁷ The link directs to the Fran dictionary portal, specifically to all dictionary entries listed under the headword *cerkev*.

human consultants in this respect, primarily because it occasionally exceeds expectations in its efforts to be as helpful as possible.

In certain cases, the output generated by ChatGPT does not replicate the style of the LCS, particularly in terms of excessive informality and rather extravagant text formatting. There are also some more specific deviations, such as the representation of pronunciation. For instance, the output does not employ square brackets, which are currently used in the LCS to indicate pronunciation. Instead, it renders the pronunciation of acronyms such as *A. M.* as “»a-em«” (with quotation marks), rather than [âem].

In this experiment, all 30 of ChatGPT’s responses successfully addressed the questions posed in the enquiries, indicating a clear understanding of the prompts and an attempt to provide relevant answers. Overall, approximately 90 percent of the final solutions or suggestions correspond, either partially or fully, to those offered by human language consultants. The only linguistic topics that appear to exceed its capabilities are those concerning etymology, which is understandable given that etymological analysis requires specialised resources that are seldom freely accessible.

In several instances, ChatGPT adopted either a more prescriptive and directive tone than its human counterpart, or, conversely, offered a more flexible and permissive response. Such variation in the output is not unexpected. Contemporary normative tendencies in Slovenian linguistics remain largely aligned with the still-valid codification outlined in *Slovenski pravopis 2001*. However, in recent years, more liberal and descriptive approaches have gained increasing prominence. As such, the oscillation between conservative and liberal stances in ChatGPT’s responses reflects the broader linguistic landscape in Slovenia.

ChatGPT was instructed to provide internet links and executed this task with notable enthusiasm. In the majority of its responses, it included multiple hyperlinks, i.e., to the LCS website, the Fran portal, and various language-related platforms. However, only a minority of these links prove genuinely useful, in the sense that they direct the user to websites containing relevant and applicable content. More frequently, the links lead to webpages whose content is immaterial to the intended topic. A particularly recurrent issue in ChatGPT’s output is the inclusion of links purportedly referring the user to relevant topics on the LCS website: while these links often do lead to pages within the LCS domain, the associated content is frequently unconnected to the subject under discussion.

3.2.2 Factual accuracy

Even when the ChatGPT’s final solutions or suggestions correspond with those of the human language consultant, the accompanying argumentation and explanations may be problematic, primarily due to three main areas of deficiency: (1) incorrect language use,

(2) errors and inaccuracies in the application of linguistic terminology, and (3) faults in the description of linguistic phenomena, as well as of other, more general, factual information.

The use of Slovenian in ChatGPT's output is generally acceptable. Orthographic, spelling, and grammatical errors are infrequent. It is evident that the output is not produced directly in Slovenian, but rather mediated through English. This is particularly noticeable in the grammar, especially in sentence structures that tend to be overly noun-oriented, reflecting literal translations from English. At times, stylistically or discursively inappropriate lexical choices occur, including the use of colloquialisms (e.g., the verb *rabiti* instead of *potrebovati*, meaning 'to need') or terminology specific to professional jargon. Isolated instances of orthographic errors (e.g., spelling the noun *niz* incorrectly as *nis*, meaning 'sequence') and invented words (e.g., the non-existent adjective *nadpomenčen* in the example above, meaning 'hypernimitic') are also present.

In ChatGPT's output, terminology established in contemporary Slovenian linguistics is generally used correctly when referring to broader or more widely known concepts, such as parts of speech. However, terms denoting more specialised concepts are frequently used inaccurately (e.g., referring to a single letter as *soglasniški sklop*, meaning 'consonant cluster'), are literal translations from English (e.g., the noun *predmodifikator* instead of *levi prilastek*, meaning 'premodifier'), or are fabricated (e.g., the noun *nespregibnost* instead of *nepregibnost*, meaning 'non-inflectability, the inability of the word to be inflected'). More general or descriptive definitions and paraphrases are often employed (e.g., using the descriptive definition *kratica, ki je izgovorljiva kot beseda* instead of the set term *nečrkovalna kratica*, meaning 'acronym, an abbreviation that is pronounced as a word'),¹⁸ with a varying level of success in explaining the linguistic concepts in the background: while this approach can occasionally serve as an effective means of explaining complex concepts in a clear and accessible manner, it also results in occasional errors, including incoherent or nonsensical explanations. Ideally, ChatGPT would demonstrate both an understanding as well as correct usage of appropriate linguistic terminology, particularly given that human-produced responses in the LCS aim not only to provide answers, but also to educate users by employing accurate terms and elaborating on the underlying concepts in detail.

Errors in the referencing and description of linguistic facts are infrequent, yet they do occur. In the majority of instances, linguistic phenomena are presented accurately and adequately, even when the use of terminology is imprecise and the accompanying explanations remain overly general. Occasionally, however, the linguistic phenomena cited along with their analyses are erroneous or fabricated. For example, ChatGPT has incorrectly asserted that, in Slovenian, denominations for prizes and honours are written with an initial capital letter. Furthermore, ChatGPT appears to lack sufficient capability to retrieve linguistic information from designated online resources. In several

¹⁸ Initialisms, on the other side, are abbreviations consisting of initial letters pronounced separately (e.g., *BBC*). The set Slovenian term for this linguistic concept is *črkovalna kratica*.

cases, it has provided statements that are demonstrably inconsistent with data available on the Fran dictionary portal. One such instance involved the claim that Slovenian lacks an adjectival form derived from the noun *droži* ('dough starter'); however, consultation of the Fran portal readily reveals that the adjective *drožen* is well attested. Mistakes in referencing and the presentation of more general facts also occur. For instance, the Ljubljana International Film Festival has been erroneously referred to as *Liff* rather than the correct acronym *Liffe*.

3.2.3 Assessment of quality, usability and potential improvements

Although many of the final suggestions and the positions adopted in the responses correspond to those provided by human language consultants, the frequent inadequacy of the accompanying argumentation and explanations constitutes a considerable limitation. This shortcoming significantly compromises the viability of generative AI as an autonomous tool intended for independent use by language users. Even when the final output accords with the advice a human consultant might offer, the persistent presence of problematic content within the explanatory components renders the tool unsuitable, at least for the time being, as a stand-alone linguistic resource.

Nonetheless, in some instances, ChatGPT exhibits impressive insight. On occasion, its output includes perceptive observations that may prove valuable even to a human language consultant. For example, in one case, ChatGPT correctly identified the enquiry as terminological in nature and recommended that it be referred to the Terminological Consulting Centre, also operated by the Fran Ramovš Institute of the Slovenian Language. Additionally, the output occasionally contains useful stylistic suggestions, such as advising when certain elements might be placed in brackets to enhance clarity.

One potential avenue for enhancing ChatGPT's output lies in the refinement of the input provided. The prompt could include more explicit instructions regarding the preferred sources ChatGPT should consult.¹⁹ Furthermore, its capabilities could be more effectively leveraged through the development of a Custom GPT, which allows users to configure the model with specialised instructions and integrate external knowledge sources. This could include granting access to relevant linguistic databases (such as the LCS database and the Fran portal) either by uploading structured datasets or connecting them via API. In this way, the Custom GPT could be more systematically tailored to function as a specialised language consultant.

Additional embedded knowledge concerning the Slovenian language (its structural characteristics and specialised linguistic terminology) would further enhance the model's performance. In this context, the Centre for Language Resources and

¹⁹ For guidance on how to formulate effective prompts in order to obtain the desired output, see Saravia (2022).

Technologies at the University of Ljubljana is developing a large language model for Slovenian, named GaMS (Generative Model for Slovenian), within the framework of the PoVeJMo project. The integration of such a model, in conjunction with ChatGPT, would very likely increase the overall effectiveness of generative AI in fulfilling the role of a language consultant.

However, certain limitations of artificial intelligence are unlikely to be resolved in the near future, most notably its non-deterministic nature,²⁰ whereby identical queries may yield different outputs across repeated interactions, and its well-documented tendency to hallucinate, that is to generate information that is factually incorrect or entirely fabricated. Additional concerns also warrant caution, such as the model’s opaque decision-making processes (often referred to as the “black box” problem), susceptibility to bias and stereotypes, and lack of transparency. Nonetheless, these deficiencies are generally not overly intrusive or disruptive in the context of deploying this technology for language consulting.

4. Conclusion

This experiment has shown that generative AI holds clear potential for use in language consulting, however, not (yet?) in a format that would be suitable for the end-user, such as a chatbot or similar stand-alone tool. At present, the quality of the output is not consistently high enough, and errors remain too frequent to justify unsupervised use. Nevertheless, AI-generated responses could serve effectively as preliminary drafts to be post-edited by a human language consultant. The extent of post-editing would naturally vary depending on the complexity of the enquiry, and all information provided would still require verification. Even so, this approach could ease the consultant’s workload and increase efficiency.

In particular, AI could assist by performing targeted corpus-like queries, identifying related or previously published answers, or offering draft formulations for typical user enquiries. With further refinement of prompts and better access to trusted resources, its utility would increase substantially. These improvements could be realised through the development of a Custom GPT and through collaboration with dedicated large language models for Slovenian such as GaMS.

To conclude, the potential is undeniably there. But as has already been observed in the context of AI-assisted translation, “[t]he integration of human expertise is crucial for overcoming the limitations of AI” (Benko et al., 2024). AI should not be seen as a replacement for human consultants, but rather as a supplementary tool, one that, if

²⁰ The “non-deterministic nature” described here is not inherent to the model: identical queries can yield identical results if randomness is disabled (by setting the temperature parameter to 0). When the web interface is used without adjusting these parameters, variability in responses stems from the default settings and from web search results.

properly configured and supervised, can meaningfully support language consulting practice.

5. Acknowledgements

This contribution was composed under the P6-0038 programme, funded by the Slovenian Research and Innovation Agency (ARIS).

6. References

- Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Using chatbots as AI conversational partners in language learning. *Applied Sciences*, 12(17), 8427.
- Benko, V., Kříhová, Z., Lehečka, B., & Vystrčilová, D. (2024). Persian to Czech Dictionary – A Traditional Dictionary in the Era of AI? In K. Š. Despot, A. Ostroški Anić & I. Brač (eds.) *Lexicography and Semantics: Proceedings of the 21st EURALEX International Congress*, pp. 565–581.
- Carbajal-Carrera, B. (2024). AIsplaining: Generative AI explains linguistic identities to me. *Australian Review of Applied Linguistics*, 47(3), pp. 340–365.
- de Schryver, G.-M. (2023). Generative AI and lexicography: The current state of the art using ChatGPT. *International Journal of Lexicography*, 36(4), pp. 355–387.
- de Schryver, G.-M., Chishman, R., & da Silva, B. (2019). An overview of digital lexicography and directions for its future: An interview with Gilles-Maurice de Schryver. *Calidoscópico*, 17(3), pp. 659–683.
- Dobrovoljc, H., & Vranjek Ošlak, U. (2021). Codification within reach: Three clickable layers of information surrounding the new Slovenian normative guide. In I. Kosem, M. Cukr, M. Jakubíček, J. Kallas, S. Krek, & C. Tiberius (eds.) *Electronic lexicography in the 21st century: Post-editing lexicography. Proceedings of the eLex 2021 conference*, pp. 637–652.
- Dobrovoljc, H., Lengar Verovnik, T., Vranjek Ošlak, U., Michelizza, M., Weiss, P., & Gliha Komac, N. (2020). *Kje pa vas jezik žuli?* Ljubljana: Založba ZRC, ZRC SAZU.
- Jakubíček, M., & Rundell, M. (2023). The end of lexicography? Can ChatGPT outperform current tools for post-editing lexicography? In M. Medved, M. Měchura, C. Tiberius, I. Kosem, J. Kallas, M. Jakubíček & S. Krek (eds.) *Electronic lexicography in the 21st century (eLex 2023): Invisible Lexicography. Proceedings of the eLex 2023 conference*, pp. 518–533.
- Kosem, I., Gantar, P., Arhar Holdt, Š., Gapsa, M., Zgaga, K., & Krek, S. (2024). AI in lexicography at the University of Ljubljana. In S. Krek (ed.) *Book of abstracts of the workshop “Large language models and lexicography”*. Cavtat, Croatia, p. 29.
- Krvina, D., & Vranjek Ošlak, U. (2025). *Nadpomenka za versko stavbo*. Jezikovna svetovalnica. Accessed at: <https://svetovalnica.zrc-sazu.si/topic/7427/nadpomenka-za-versko-stavbo>. (July 20, 2025)

- Lew, R. (2015). Dictionaries and their users. In H. Jackson (ed.) *The Bloomsbury Companion to Lexicography*. London: Bloomsbury Academic, pp. 244–265.
- Ludányi, Z. (2020). Language consulting: A brief European overview. *Eruditio – Educatio*, 15(3), pp. 25–47.
- Mžourková, H. (2024). Czech and Slovenian Language Consulting: What Do They Tell Us about Language and Language Users? In S. Štumberger (ed.) *Predpis in norma v jeziku*. Obdobja 43. Ljubljana: Založba Univerze v Ljubljani, pp. 225–231.
- Palma, G. (2025). *All ChatGPT Models Explained: Which One Should You Use?* Accessed at: <https://justainews.com/blog/openai-chatgpt-models-explained/>. (May 23, 2025)
- Petrič, T., Arhar Holdt, Š., & Robnik-Šikonja, M. (2024). Pomembnost realistične evalvacije: Primer popravkov sklona in števila v slovenščini z velikim jezikovnim modelom. *Slovenščina 2.0*, 12(1), pp. 106–130.
- Ptasznik, B., & Lew, R. (2025). Dictionaries versus AI Tools through the Eyes of English Majors. *International Journal of Lexicography*, XX, pp. 1–19.
- Ptasznik, B., Wolfer, S., & Lew, R. (2024). A Learners' Dictionary Versus ChatGPT in Receptive and Productive Lexical Tasks. *International Journal of Lexicography*, 37(3), pp. 322–336.
- Rundell, M. (2023). Automating the Creation of Dictionaries: Are We Nearly There? In *Proceedings of the 16th International Conference of the Asian Association for Lexicography (ASIALEX 2023)*, Seoul, Korea, 22–24 June 2023, pp. 9–17.
- Saravia, E. (2025). *Prompt engineering guide*. Accessed at: <https://www.promptingguide.ai/>. (June 20, 2025)
- Vajjala, S. (2024). Generative Artificial Intelligence and Applied Linguistics. *JALT Journal*, 46(1), pp. 55–76.
- Vranjek Ošlak, U. (2023). Language counselling: Bridging the gap between codification and language use. *Eesti ja Soome-Ugri Keeleteaduse Ajakiri / Journal of Estonian and Finno-Ugric Linguistics*, 14(1), pp. 149–173.
- Vranjek Ošlak, U. (2024). Jezikovno svetovanje v obdobju prenove pravopisnega priročnika. In S. Štumberger (ed.) *Predpis in norma v jeziku*. Obdobja 43. Ljubljana: Založba Univerze v Ljubljani, pp. 399–405.
- Žaucer, R., & Marušič, F. (2009). Jezikovno svetovanje, praksa in ideali. In M. Stabej (ed.) *Infrastruktura slovenščine in slovenistike*. Obdobja 28. Ljubljana: Založba Univerze v Ljubljani, pp. 449–456.

This work is licensed under the Creative Commons Attribution ShareAlike 4.0 International License.

<http://creativecommons.org/licenses/by-sa/4.0/>

