The application of the ChatGPT language model for automatic generation of structured abstracts

Arkadiusz Pulikowski
ORCID: 0000-0003-1807-8642
Institute of Culture Studies
Faculty of Humanities
University of Silesia in Katowice

Abstract

Purpose/Thesis: The research aimed to evaluate the usefulness of the ChatGPT language model in generating structured abstracts for academic publications.
Approach/Methods: The methodology was qualitative. The study analysed ten articles from the journal Zagadnienia Informacji Naukowej—Studia Informacyjne (5 in Polish, 5 in English). ChatGPT version 4o was used to generate structured abstracts of the selected articles, then compared with the original abstracts to assess whether ChatGPT provided the required information in each section.
Results and conclusions: ChatGPT demonstrated strong capabilities in analysing and summarising documents to create abstracts for scientific publications in the field of information science. The language model performed well for both languages, with only two abstracts exhibiting significant issues in specific sections.
Originality/Value: The study showed the potential of language models, such as ChatGPT, in generating structured abstracts for bibliographic and full-text databases and as a complement to the researcher's workshop.

Keywords:

AI-generated abstracts. ChatGPT. Language models. Structured abstracts.

Text received on 15th of September 2024.

1. Introduction

The release of the ChatGPT tool by OpenAI in 2023 was undoubtedly a groundbreaking event, with an increasing influence on many areas of human activity. The capabilities offered by ChatGPT and similar language models were also quickly recognised within the scientific community. The number of publications employing artificial intelligence (AI) in research across various fields is proliferating. Document summarisation is among the model's key features in natural language

processing. The author decided to leverage this capability to create structured abstracts automatically.

Structured abstracts consist of clearly labelled sections (e.g., Background, Purpose, Methods, Results, Conclusions), which help present information clearly and consistently to readers. While these components are also present in traditional abstracts, they are not explicitly labelled or organised similarly. Headings ensure authors follow a standardised format, reducing the risk of omitting essential elements. Key components such as the purpose, methods, results, and conclusions are generally expected in high-quality abstracts, whether traditional or structured. However, a structured format makes the abstract clearer, easier to read, and easier to search (Pulikowski, 2020, p. 25–26).

The research aims to evaluate the usefulness of the ChatGPT language model for automatically generating structured abstracts. A comparative analysis method was employed to verify the model's utility for abstract generation by comparing author-generated structured abstracts with those generated by ChatGPT. The study utilised research papers published in *Zagadnienia Informacji Naukowej – Studia Informacyjne* (*Issues of Information Science – Information Studies*).

2. Previous studies

Among the numerous publications discussing the benefits and risks of using artificial intelligence in scientific articles, a small but rapidly growing group focuses on abstract generation. A leading topic in this area is the comparison of abstracts generated by language models with original abstracts written by authors, specifically in terms of similarity and distinguishability (blind tests, automatic AI detection, linguistic accuracy, ethical considerations). Most of the publications come from the medical sciences.

The latest papers on the currentl versions of ChatGPT suggest that it has significant potential in generating scientific abstracts, with studies indicating varying levels of quality and accuracy compared to human-written counterparts. In a comparative analysis, original abstracts outperformed those generated by ChatGPT 3.5 and 4.0 in terms of quality; however, ChatGPT-generated abstracts were found to be more readable (Cheng et al., 2023; Hwang et al., 2024). Gravel et al. (2024) report similar findings, noting that while ChatGPT 4.0 does not produce abstracts of higher quality than those written by researchers, it is a valuable tool to help researchers improve the quality of their abstracts. Additionally, experienced reviewers needed help differentiating between AI-generated and human-written abstracts, indicating that ChatGPT can produce convincing content (Holland et al., 2024; Stadler et al., 2024). Despite some concerns regarding hallucinations

and inaccuracies, ChatGPT's ability to summarise and generate concise abstracts suggests it could be a valuable tool in medical research (Hake et al., 2024).

3. Methodology

The study utilised publications in the field of information science, both in Polish and English, which appeared in the journal *Zagadnienia Informacji Naukowej – Studia Informacyjne* during the 2022–2023 period. A total of 10 research papers were selected – 5 in Polish and 5 in English. The purpose of juxtaposing publications in two languages was to assess whether ChatGPT would handle natural language processing equally well in both cases. Table 1 presents descriptions of the articles and assigned identifiers, which will be used primarily to discuss the study's results.

The selected articles' PDF files were downloaded from the journal's website (http://ojs.sbp.pl/index.php/zin), and the original abstracts were subsequently removed. The research employed the ChatGPT language model, version 4o (Omni), as it was the only model at the time of the study (early June 2024) capable of analysing attached text files. This unique functionality, combined with ChatGPT's widely recognised expertise in language processing, made it the optimal choice for the research.

Table 1. List of publications used in the study.

ID	Article title and author	ZIN No.
PL1	*Analiza struktury leksykalnej tytułów drapieżnych czasopism* Białka N.	2022, 60 (1)
PL2	*Tagowanie zdjęć portretowych w serwisie Instagram* Kosik N.	2022, 60 (1)
PL3	*Budowa i charakterystyka Korpusu Polskich Czasopism Naukoznawczych* Kulczycki E., Mena Y. A. Z., Krawczyk F.	2023, 61 (2)
PL4	*Walki informacyjne w paradygmacie ekosystemów informacyjnych* Materska K.	2023, 61 (1)
PL5	*Modele dojrzałości systemów informacyjnych na przykładzie bibliotek cyfrowych i serwisów danych badawczych* Nahotko M.	2022, 60 (1)
EN1	*Information literacy and information behaviour of disadvantaged people in the COVID-19 pandemic. Case study of beneficiaries of the charitable foundation* Kisilowska-Szurmińska M., Paul M., Piłatowicz K.	2023, 61 (1)
EN2	*Information technology maturity and acceptance models integration: the case of RDS* Nahotko M.	2023, 61 (1)

ID	Article title and author	ZIN No.
EN3	*Research on digital culture (cyberculture) – knowledge domain analysis based on bibliographic data from the Web of Science database* Osiński Z.	2023, 61 (1)
EN4	*Full-Text Search in the Resources of Polish Digital Libraries* Pulikowski A.	2022, 60 (2)
EN5	*How do early career researchers perceive success in their fields? Report on interviews with humanists, theologians, and scientists-artists in Poland* Świgoń M.	2023, 61 (2)

Source: self-authored.

The subject of the analysis was the abstracts generated by ChatGPT based on the attached files containing articles. For each document, ChatGPT was tasked with creating a structured abstract consisting of four sections, corresponding to the sections required by the journal's editorial board as mandatory: Purpose/Thesis, Approach/Methods, Results and Conclusions, and Originality/Value. The abstract generated by the language model was compared with the author's original abstract to determine whether it contained the expected information in each respective section. In cases of uncertainty, the full version of the article was consulted. That was often necessary, as ChatGPT frequently selected information different from the article's author for the individual sections of the abstract. The aim of the comparison was not, as in other studies, to assess whether the AI-generated abstract could be distinguished from a human-written one but rather to examine whether language models could be used to generate informative abstracts, particularly for bibliographic and full-text databases or reference management software. Under this assumption, grammatical and stylistic correctness is of secondary importance but remains relevant for ensuring the accurate and easy comprehension of the text.

The prompt for the articles in Polish was as follows:

> „Zapoznaj się z artykułem naukowym w załączonym pliku i na jego podstawie napisz abstrakt w języku polskim, nie dłuższy niż 200 słów, składający się z czterech akapitów zatytułowanych: Cel badań, Metody badań, Wyniki i wnioski z badań, Wartość poznawcza badań"

In turn, for the articles in English:

> „Read the scientific article in the attached file and, based on it, write an abstract in English, no longer than 200 words, consisting of four paragraphs entitled: Purpose of the research, Research methods, Results and conclusions of the research, Cognitive value of the research"

In both prompts, all labelled sections of the abstract included the word "research" (in Polish: "badań") to enhance the precision of ChatGPT's response. The 200-word limit for the generated abstracts was established based on the analysis of the length of the authors' abstracts. This data is presented in Table 2, together with the word count of abstracts generated by the language model. As can be observed, despite the clearly specified word limit in the prompt, ChatGPT exceeded the 200-word threshold in three cases: PL4, EN3, and EN4.

Table 2. Number of words in authors' abstracts and generated by ChatGPT.

	PL1	PL2	PL3	PL4	PL5	EN1	EN2	EN3	EN4	EN5
Author	177	117	102	109	116	200	229	119	199	157
ChatGPT	185	195	153	203	141	198	173	203	237	142

Source: self-authored.

The prompts for individual articles were entered in new chat sessions to ensure they were not interpreted as related within a single thread. The context was specified within ChatGPT's custom instructions settings to tailor the responses better. In the section "What should ChatGPT know about you to provide better responses?" the following was entered: "I am a university professor. My field of expertise is information science". Meanwhile, in the section "How would you like ChatGPT to respond?" it was specified: "I prefer concise responses written in scientific language".

4. Study results

The research results were positively surprising in only two cases. ChatGPT generated abstract sections where the content was incorrect or significantly insufficient (PL5 – purpose, EN3 – results). No significant differences were observed in the informational content between abstracts generated for publications in Polish and English. ChatGPT performed equally well in both cases. It is also worth noting that none of the generated abstracts contained fabricated information, a phenomenon known as "hallucination". The research results are presented in Table 3.

Table 3. Correctness of individual sections of abstracts generated by ChatGPT.

	PL1	PL2	PL3	PL4	PL5	EN1	EN2	EN3	EN4	EN5
Purpose	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Methods	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Results	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓
Value	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Source: self-authored.

In some abstracts, sections generated by the language model provided more detailed information than those written by the authors of the publications. That occurred most often when the description generated by ChatGPT was longer than the author's version. However, this issue needs to be revised for a definitive evaluation, and thus, the author decided not to include it in the research results. It is much more feasible to determine whether each abstract section contains the

appropriate information than to evaluate which of the two correctly formulated sections is superior.

Apart from the informational content, it is also important to consider the language of the abstracts generated by ChatGPT. The texts were written clearly and comprehensibly, employing a typical academic style. The stylistic and grammatical correctness of the abstracts is of a high standard. That can be easily assessed in the following example for the EN5 article. A comparison of all author-written and language model-generated abstracts used in the study has been made available in the Zenodo repository (Pulikowski, 2024).

EN5. How do early career researchers perceive success in their fields? Report on interviews with humanists, theologians, and scientists-artists in Poland [Marzena Świgoń, ZIN 2023, 61 (2)]

Purpose/Thesis

Author: This paper presents the Polish researchers' opinions on success perception in science and examines the role of bibliometric and altmetric indices/metrics. It also provides respondents' statements on how their work is evaluated at Polish universities.

ChatGPT: This research explores how early career researchers in the humanities, theology, and arts in Poland perceive success. It examines factors associated with reputation and success, including bibliometric and altmetric indicators, aiming to address gaps in previous studies focused on sciences and social sciences.

Approach/Methods

Author: The data come from in-depth interviews (spring 2023) with early career researchers in Linguistics, Literary Studies, History, Philosophy, Polish studies, Theology, Arts, Music, Plastic/Fine Arts, and Film.

ChatGPT: The study uses longitudinal interviews, with the first round in spring 2023. Interviews were partly structured, recorded, and transcribed for thematic analysis. The responses were manually coded using a heuristic approach and a standardised thematic framework.

Results and conclusions

Author: The respondents' attachment to indices used in the periodical evaluation of scholars at Polish universities, such as the number of publications, ministerial score points, and external grants, was observed. The respondents were critical of citations as the determinant of success and generally not very interested in altmetrics.

ChatGPT: Early career researchers prioritise academic publications, grants, international cooperation, and recognizability. Citations and altmetrics are relevant but less significant. Arts researchers emphasise non-numerical success measures more than those in humanities and theology.

Originality/Value

Author: The importance of this research (National Science Centre grant No 2022/45/B/HS2/00041), which is the Polish continuation of the two international projects (2016–2022), lies mainly in providing a platform to researchers in the arts and humanities, which were not included in earlier studies devoted to the sciences and social sciences.

ChatGPT: This research highlights the unique challenges of early career researchers in humanities, theology, and arts in Poland. It emphasises diverse success indicators beyond bibliometrics, informing policies to support their development and recognition.

Since ChatGPT performed very well in processing Polish while generating abstracts for publications in this language, an additional test was conducted to assess how it would handle generating Polish abstracts based on English articles. This capability could be particularly useful for users of bibliographic or full-text databases. To evaluate ChatGPT's performance in this context, the research was repeated for English publications (EN1–EN5), using the prompt designed for Polish publications. The results presented in Table 4 show that changing the language of the generated abstract did not affect its accuracy. It can be assumed that ChatGPT may be equally effective in many other languages it supports. However, the level of support for those languages may vary depending on linguistic complexity and data availability, making this an assumption that requires further verification.

Table 4. Correctness of Polish abstracts generated by ChatGPT based on English publications

	EN1	EN2	EN3	EN4	EN5
Purpose	✓	✗	✓	✓	✓
Methods	✓	✓	✓	✓	✓
Results	✓	✓	✓	✓	✓
Value	✓	✓	✓	✓	✓

Source: self-authored.

5. Study limitations

When analysing the presented results, it is important to consider the research's limitations. First and foremost, it should be noted that ChatGPT exhibits significant variability in the responses it generates. The answers are produced dynamically, incorporating an element of randomness, which means that even when the same question is repeated, the response may be formulated slightly differently – similar in information content, but potentially better or worse. The study shows that the response will still be correct in most cases. Additionally, the model evolves, continuously improving, with new and significantly modified versions being released periodically. All of these factors contribute to a variable and dynamic environment.

Another important limitation to consider when analysing the research results is the focus on a single journal from one discipline – information science – as well as the small number of publications included in the study – 10 in total. Even within the same discipline, there is no certainty that the results would be equally satisfactory for publications with a higher level of content complexity.

Finally, it is important to mention the subjectivity involved in evaluating the correctness of the generated abstracts. Although the Author made every effort

to ensure that the assessment was reliable, it cannot be ruled out that another representative of the discipline, using slightly different criteria, might evaluate the abstracts generated by ChatGPT differently.

6. Conclusions

ChatGPT demonstrated strong capabilities in analysing and summarising documents to create abstracts for scientific publications in information science. The research conducted, along with other studies mentioned in the 'Previous Studies' section, confirms that language models can be successfully used to automatically create structured abstracts, particularly for bibliographic and full-text databases, thereby expanding the existing functionalities of these systems.

A good example of a service that already utilises the capabilities of language models is Scispace (https://typeset.io). It allows users to expand the list of retrieved publications with additional columns containing automatically generated short descriptions based on predefined headings modelled after structured abstract sections (Figure 1). In addition, users can create their custom headings using the "Create new column" button and freely engage in conversations about selected publications using the "Chat with Paper" option.

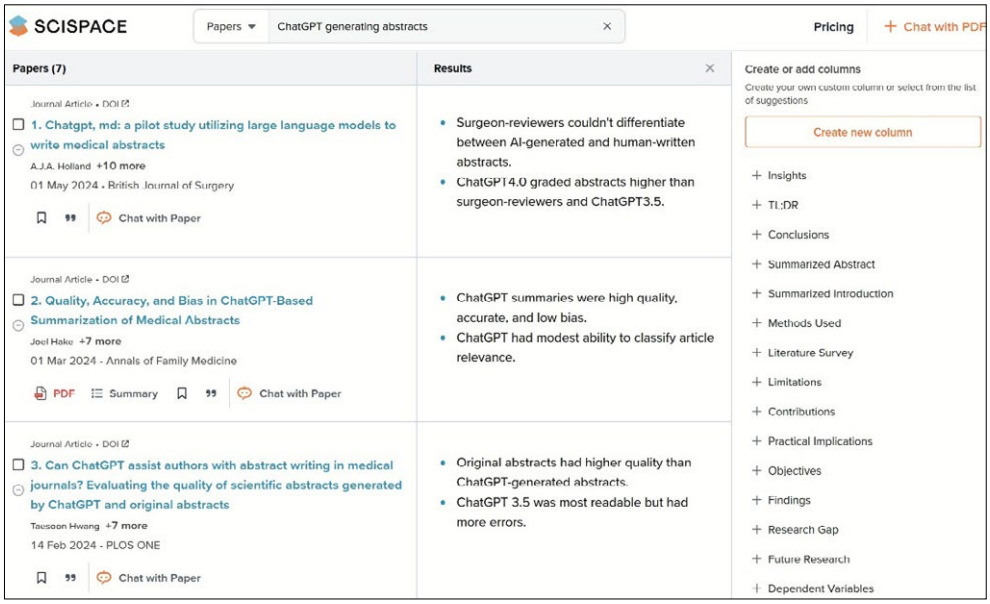

Figure 1. Adding columns to the list of retrieved publications in the Scispace service

Source: https://typeset.io.

In addition to application in bibliographic and full-text databases, language models such as ChatGPT can be effectively used to analyse scientific publications and as a source of inspiration for creating author-written abstracts. When analysing articles, it is possible to generate summaries of a specified length (e.g., 200 or 300 words) containing sections tailored to individual needs and written in the user's chosen language. Users can also ask language models for further clarification or elaboration on specific topics. In the case of personal publications, the abstract writing process can be enhanced by generating one or several versions of summaries using the language model, which can serve as a source of inspiration for further work or to improve an already written abstract.

References

Cheng, S., Tsai, S., Bai, Y. et al. (2023). Comparisons of quality, correctness, and similarity between ChatGPT-generated and human-written abstracts for basic research: Cross-sectional study. *Journal of Medical Internet research*, *25*, e51229. doi: 10.2196/51229.

Gravel, J., Dion, C., Fadaei Kermani, M. et al. (2024). Will ChatGPT-4 improve the quality of medical abstracts? *medRxiv*, 2024–02. doi: 10.1101/2024.02.09.24302591.

Hake, J., Crowley, M., Coy, A., et al. (2024). Quality, accuracy, and bias in ChatGPT-based summarisation of medical abstracts. *The Annals of Family Medicine*, *22*(2), 113–120. doi: 10.1370/afm.3075.

Holland, A., Lorenz, W., Cavanaugh, J., et al. (2024). ChatGPT, MD: A pilot study utilising Large Language Models to write medical abstracts. *British Journal of Surgery*, *111*(5), znae122.039. doi: 10.1093/bjs/znae122.039.

Hwang, T., Aggarwal, N., Khan, P. Z., et al. (2024). Can ChatGPT assist authors with abstract writing in medical journals? Evaluating the quality of scientific abstracts generated by ChatGPT and original abstracts. *Plos One*, *19*(2). doi: 10.1371/journal.pone.0297701.

Pulikowski, A. (2020). The relation between the structure of abstracts in LIS and anthropology journals and their rank. *Zagadnienia Informacji Naukowej – Studia Informacyjne*, *58*(1), 24–39. doi: 10.36702/zin.645.

Pulikowski, A. (2024). The comparison of structured abstracts generated by the ChatGPT language model with the author's original abstracts derived from research papers [Data set]. Zenodo. doi: 10.5281/zenodo.13765256.

Stadler, R. D., Sudah, S. Y., Moverman, M. A., et al. (2024). Identification of ChatGPT-generated abstracts within shoulder and elbow surgery poses a challenge for reviewers. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*. doi: 10.1016/j.arthro.2024.06.045.

Zastosowanie modelu językowego ChatGPT do automatycznego generowania ustrukturyzowanych abstraktów

Abstrakt

Cel/teza: Badanie miało na celu ocenę użyteczności modelu językowego ChatGPT do generowania ustrukturyzowanych abstraktów publikacji naukowych.
Koncepcja/metody badań: Badania miały charakter jakościowy. Analizie poddano 10 artykułów z czasopisma Zagadnienia Informacji Naukowej – Studia Informacyjne, 5 w języku polskim i 5 w języku angielskim. Korzystając z modelu językowego ChatGPT w wersji 4o, dla każdego artykułu wygenerowano ustrukturyzowane abstrakty, które następnie porównywano z abstraktami autorskimi w celu sprawdzenia, czy zawierają poprawne informacje w poszczególnych sekcjach.
Wyniki/wnioski: ChatGPT potwierdził duże możliwości w zakresie analizy i streszczania dokumentów w celu tworzenia abstraktów publikacji naukowych z zakresu informacji naukowej. Model językowy poradził sobie równie dobrze z publikacjami w języku angielskim i polskim. Tylko w przypadku dwóch abstraktów wykryto błędnie wygenerowaną treść pojedynczych sekcji.
Oryginalność/wartość poznawcza: Badanie pokazało potencjał modeli językowych, takich jak ChatGPT, w tworzeniu ustrukturyzowanych abstraktów, zarówno na potrzeby bibliograficznych i pełnotekstowych baz danych, jak i jako uzupełnienie warsztatu badacza.

Słowa kluczowe

ChatGPT. Generowanie abstraktów przez AI. Modele językowe. Ustrukturyzowane abstrakty.

ARKADIUSZ PULIKOWSKI, profesor uczelni w Instytucie Nauk o Kulturze na Wydziale Humanistycznym Uniwersytetu Śląskiego w Katowicach. Zainteresowania badawcze: wyszukiwanie informacji, zachowania informacyjne, infometria, digitalizacja informacji. Wybrane publikacje: Modelowanie procesu wyszukiwania informacji naukowej. Strategie i interakcje (Katowice, 2018), Searching for LIS scholarly publications: a comparison of search results from Google, Google Scholar, EDS, and LISA (Journal of Academic Librarianship, 2021, współaut. A.Matysek), The Relation between the structure of abstracts in LIS and anthropology journals and their rank (Zagadnienia Informacji Naukowej, 2020).

Contact details:
arkadiusz.pulikowski@us.edu.pl
Uniwersytet Śląski w Katowicach
Wydział Humanistyczny
Instytut Nauk o Kulturze
ul. Bankowa 11, 40–007 Katowice

CALL FOR PAPERS #1

AFFECTIVE ASPECTS OF INFORMATION BEHAVIOR, EMOTIONS, INFORMATION WELL-BEING

SUBMISSION DEADLINE:
15TH MARCH 2025