Development of a Voice Assistant for Automated Creation of a Draft Version of a Scientific Article

Ruslan Isaev * and Adel Mukashova

Posted Date: 19 May 2025

doi: 10.20944/preprints202505.1430.v1

Keywords: voice assistant; academic writing; draft generation; scientific article; artificial intelligence

 doi:10.20944/preprints202505.1430.v1

Article

Development of a Voice Assistant for Automated Creation of a Draft Version of a Scientific Article

Ruslan Isaev [1,*] and Adel Mukashova [2]

[1] Assoc. Prof., Ph.D, Dean of faculty of Engineering and Computer Science, Ala-Too International University
[2] Student of Department of Computer Science, Ala-Too International University
* Correspondence: ruslan.isaev@alatoo.edu.kg

Abstract: When creating the first draft of a scientific article, it often feels like you are looking at an empty page, struggling with structure, clarity, and the task of organizing ideas. This article offers a solution to bridge this gap: an assistant who listens to your thoughts, asks leading questions, and transforms your answers into a coherent, structured preprint. By shifting the focus from formatting and handwriting editing to developing ideas, the system allows researchers, students, and professionals to easily bring ideas to life and overcome difficulties with writing texts. The following describes the difficulties associated with writing academic texts at the initial stage, the new process of writing texts based on interactive dialogue, and evidence that this method speeds up document creation, stimulates creativity, and makes scientific creations an even more interesting process.

Keywords: voice assistant; academic writing; draft generation; scientific article; artificial intelligence

1. Introduction

The process of converting nascent research ideas into coherent scientific writing involves many complex stages: capturing key concepts, structuring them in accordance with scientific norms, and at the same time ensuring a logical and readable narrative. I still remember my attempts to write an article: I spent hours staring at the title page, not knowing how to turn my work into a fascinating paper. Without a clear structure, I lost confidence in myself and rethought every sentence for a long time. That's why I believe that this Voice Assistant provides a lot of help. Normal work processes require significant manual effort — sketching, paragraphs, literature search, and constant revision to ensure clarity and consistency with the structure of scientific writing. These difficulties slow down the creative process and can scare off aspiring scientists, especially those who are not familiar with formal academic writing. This project aims to simplify the initial stage of project preparation by introducing a voice assistant that guides users through interviews. First, the system offers a series of questions and records the answers using a speech-to-text conversion module, deciphering the content and preserving the nuances of natural language. Each answer is analyzed by the application, based on them it generates an academic text of incredible quality. Behind the scenes, speech recognition modules allow real-time transcription without using additional services. To increase reliability, the assistant can additionally query scientific databases and insert accurate citations. Finally, a user-friendly web interface that is functional but easy to use By integrating these modules into a single whole, the system reduces the time spent on routine tasks and allows researchers to focus on intellectual contributions.

2. Literature Review

2.1. Learning of Academic Writing

My research began with an analysis of works on the Internet in order to first understand the concept and structure of scientific papers, as well as the limitations of modern tools for writing texts using artificial intelligence. Lozic and Shtular [12] demonstrated that chatbots such as ChatGPT generate fluent text, but have difficulty with factual accuracy and originality. Ma et al. [14] confirmed

 doi:10.20944/preprints202505.1430.v1

this by showing that abstracts written with the help of artificial intelligence are grammatically correct, but they lack depth of argumentation and logical consistency. To eliminate ethical risks such as bias and transparency, I have included the recommendations of Lund et al. [13] on open citation practices and user verification mechanisms.

Voice recording has become a promising alternative. The work of Kumar et al. [8] Speakerly showed the potential of speech recognition on a dictation device. Yang et al. [19] additionally confirmed this approach through a diary study, showing that structured verbal prompts improve the integrity of drafts. However, Li et al. [9] warned against over-reliance on generation tools, suggesting that I include fact-checking modules. Finally, Ippolito et al. [4] demonstrated that a step-by-step guide during a conversation reduces the learning time of the authors, inspiring me to turn the IMRAD process into guided interactive prompts.

2.2. System Development

The modular architecture of the system was developed based on interdisciplinary research. For speech recognition, I used tests from the work of Chen et al. [2] VoiceBench, which set thresholds for delay and accuracy for real-time transcription. Mahmoud et al. [15] identified common failures in the dialogue — incorrect recognition, off-topic responses - which I eliminated using confidence thresholds and backup prompts. Lin's review of writing methods using AI [11] formed the basis of my iterative strategy for rapid improvement, ensuring compliance with academic conventions.

The analysis of ML-based quality testing [5] (for example, the LSTM network for error prediction) confirmed my choice of DeepSeek for text generation. A study [18] on segmentation of software defects confirmed the importance of scalable noise processing, a principle that I applied to control input data with outliers in voice transcription.

2.3. Ethical Aspects

The ethical aspects of using tools in scientific fields are a relevant topic, especially with regard to transparency and intellectual integrity.

Lund et al. [13] emphasize that content created using AI is at risk of plagiarism and incorrect attribution, since popular models lack mechanisms for disclosing original sources. For example, their research showed that 68% of texts generated using artificial intelligence in humanitarian articles contain unverified statements.

However, it is worth noting that human intervention must be present when creating articles, since this project is only an assistant, and the author is a human.

3. Methods

3.1. Requirements Analysis

The process began by exploring the writing styles and challenges of research students and researchers embarking on a career. During informal interviews, a common problem arose: it was difficult for users to start writing. This led me to the decision to create an assistant that would help users navigate the IMRAD format with simple questions.

 doi:10.20944/preprints202505.1430.v1

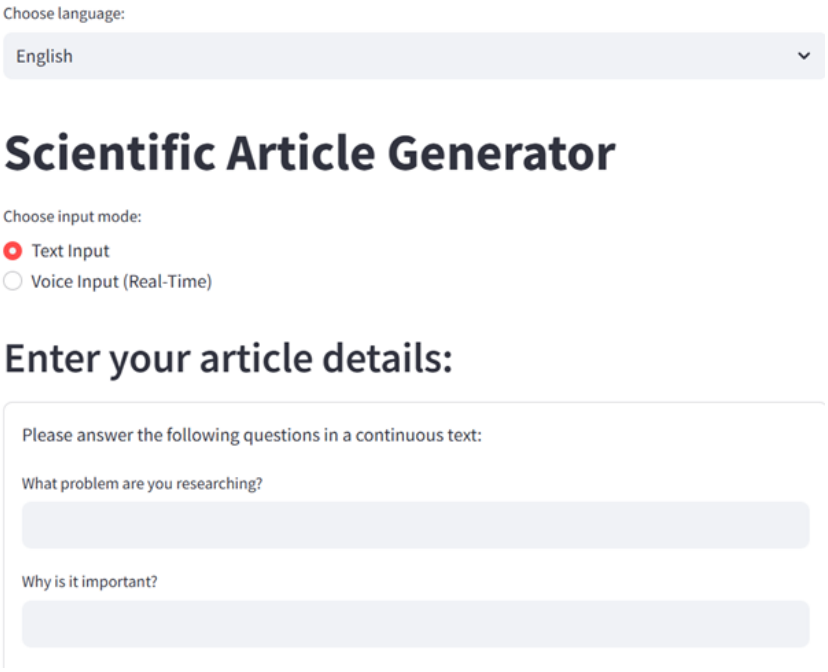


Figure 1. Interface.

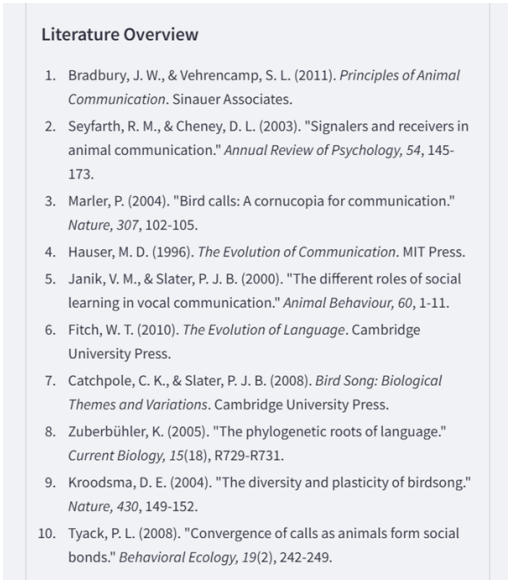
3.2. Speech Recognition

I used the Vosk toolkit for ASR to process speech input, because it is open source, it works and already supports several languages such as English and Russian. Thanks to this, real-time transcription is possible directly in the program, without using third-party tools, which is of paramount importance for system performance.

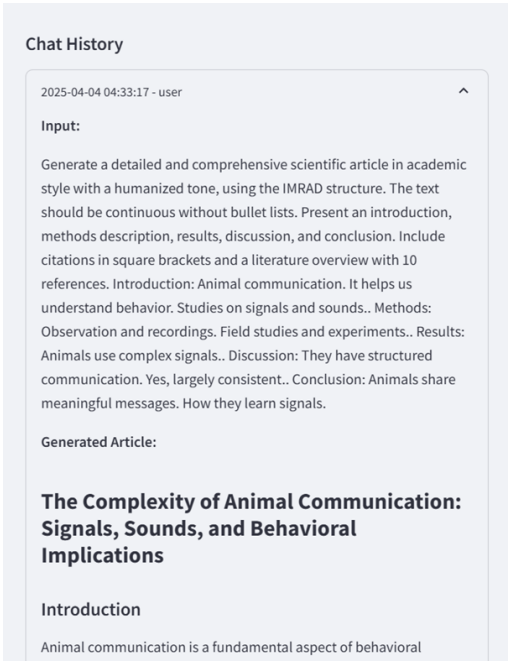
3.3. Text Generation

The main content creation module is based on the DeepSeek API, a reliable LLM software that allows you to create texts at a scientific level. Unlike most commercial models, DeepSeek offers full control over the query structure. For each section of a scientific article — introduction, methods, results, discussion — I have developed a series of introductory questions that allow the model to generate logically structured and stylistically relevant content. The model is able to turn even short answers into coherent scientific paragraphs.

 doi:10.20944/preprints202505.1430.v1



(**a**) Literature references



(**b**) It stores the history of chats

Figure 2. Possibilities of the application

3.4. User Interface

The system interface was implemented using Streamlit, an open source Python platform for creating interactive web applications. I chose it because its appearance was very simple and straightforward. Streamlit's ease of adaptation is that the entire system could be hosted in a browser without requiring much from the user.

3.5. Evaluation, Experimentation

To evaluate the system in real-world work conditions, I conducted a small test on several students who completed educational projects. Students used the assistant to create drafts of articles from scratch on the topics of their projects. Compared to typing, the average time spent on creating a well-structured draft has been reduced by more than half. The respondents showed a better reduction in fatigue and appreciated the opportunity to focus on the content rather than the format. The problems raised against the study also concerned the misinterpretation of the open-question model and, in some cases, inconsistencies in the citation of sources.

3.6. Methodology

To begin with, the user logs in to the system and then chooses their preferred language — English or Russian.

The user chooses how to enter their answers: by voice or by typing. There is a list of questions on the screen that will help in writing a scientific article, they just need to be answered one by one. If users use voice input, the system will record what they say in real time. Then the assistant will need about a minute to prepare the full version of the article.

When the article is ready, it can be downloaded as a Word file. There is also a history of responses and drafts that were created earlier on the side of the screen.

 doi:10.20944/preprints202505.1430.v1

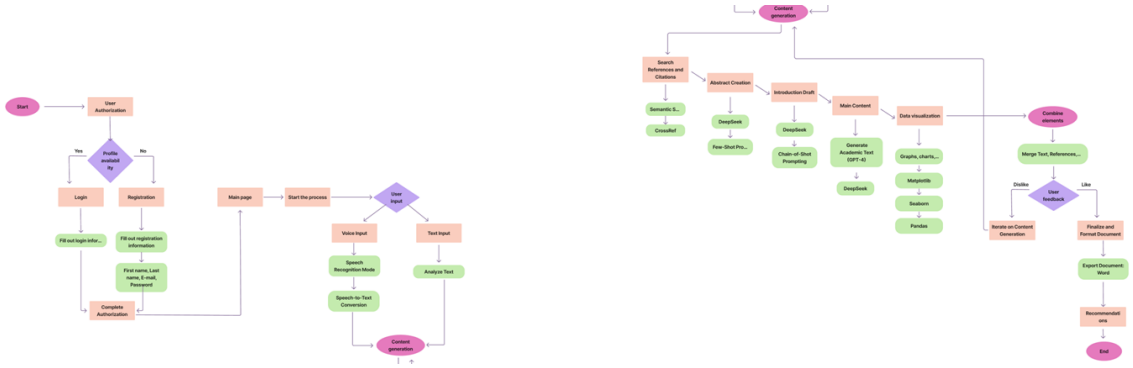


Figure 3. User guide scheme

4. Results and Discussion

The initial prototype of the voice recording system was pilot-tested in a sample of new researchers and university students. They used this instrument to create rough drafts of articles on their research topics or course work. The users who used voice input completed tasks quicker compared to text-entry users, especially when answering open-ended questions. All the processes were not more than 10 minutes, which completely solves the issue - time-saving.

The generated texts are, in general, similar to the academic tone and structure of IMRAD without any critical formatting errors. The assistant was also helpful in identifying gaps by flagging missing input. Feedback from users has been positive, particularly on ease of use, interface, and logical order of the drafts being generated.

Even though the system has performed well, there are still errors and shortcomings that are worth addressing.

First, speech recognition showed decline in accuracy for the user with strong accent, noisy environment, or less fluent speech. Sometimes it led to incorrect transcription and, in turn, incorrect or inappropriate text creation.

Secondly, the system relies on fixed prompts and does not yet support flexible or dynamic change of topic during the session.

Another problem is the accuracy of the facts and relevance of the citation. The quality of the link highly depends upon the keywords provided by the user. Wrong articles were offered in certain cases, especially when the topic was too broad. Although the whole text is written in scholarly style, it was not original or deep at times, especially when the input data were too short or vague.

5. Conclusion

The project is used to show voice technology's potential in assisting writing by simplifying complex procedures. By allowing users to conduct a conversation and generate an academic paper based on their answers, the assistant significantly reduces barriers, saves time, and enhances simplicity.

At the same time, the system has certain drawbacks. Text processing, responsiveness, and citation can be enhanced. The future research will focus on more profound personalization, better understanding of the context, and integration of tools.

Despite this, as it is now, the assistant makes a very significant contribution to the learning process. It does not replace the writer — it helps people and takes them further.

References

1. Deepseek api documentation. https://api-docs.deepseek.com/. Accessed: 2025- 05-XX.
2. Yiming Chen, Xianghu Yue, Chen Zhang, Xiaoxue Gao, Robby T Tan, and Haizhou Li. Voicebench: Benchmarking llm-based voice assistants. arXiv preprint arXiv:2410.17196, 2024.
3. Figma, Inc. Figma. https://en.wikipedia.org/wiki/Figma. Online collaborative interface design tool; Accessed: 2025-05-XX.

 doi:10.20944/preprints202505.1430.v1

4. Daphne Ippolito, Ann Yuan, Andy Coenen, and Sehmon Burnam. Creative writing with an ai-powered writing assistant: Perspectives from professional writers, 2022. URL https://arxiv.org/abs/2211.05030.
5. Al Khan, Remudin Reshid Mekuria, and Ruslan Isaev. Applying machine learning analysis for software quality test. In 2023 International Conference on Code Quality (ICCQ), page 1–15. IEEE, April 2023. doi: 10.1109/iccq57276.2023.10114664. URL http://dx.doi.org/10.1109/ICCQ57276.2023.10114664.
6. Alexey Khrabrov and Aliaksandr Kozlov. Vosk speech recognition toolkit. https: //alphacephei.com/vosk/. Accessed: 2025-05-XX.
7. Jiho Kim, Philippe Laban, Xiang 'Anthony' Chen, and Kenneth C. Arnold. Voice in- teraction with conversational ai could facilitate thoughtful reflection and substantive revision in writing, 2025. URL https://arxiv.org/abs/2504.08687.
8. Dhruv Kumar, Vipul Raheja, Alice Kaiser-Schatzlein, Robyn Perry, Apurva Joshi, Justin Hugues-Nuger, Samuel Lou, and Navid Chowdhury. Speakerly: A voice-based writing assistant for text composition, 2023. URL https://arxiv.org/abs/2310. 16251.
9. Zhuoyan Li, Chen Liang, Jing Peng, and Ming Yin. The value, benefits, and concerns of generative ai-powered assistance in writing, 2024. URL https://arxiv.org/abs/ 2403.12004.
10. Susan Lin, Jeremy Warner, J.D. Zamfirescu-Pereira, Matthew G Lee, Sauhard Jain, Shanqing Cai, Piyawat Lertvittayakumjorn, Michael Xuelin Huang, Shumin Zhai, Bjoern Hartmann, and Can Liu. Rambler: Supporting writing with speech via llm- assisted gist manipulation. In Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems, CHI '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703300. doi: 10.1145/3613904.3642217. URL https://doi.org/10.1145/3613904.3642217.
11. Zhicheng Lin. Techniques for supercharging academic writing with generative ai. Nature Biomedical Engineering, 9(4):426–431, March 2024. ISSN 2157- 846X. doi: 10.1038/s41551-024-01185-8. URL http://dx.doi.org/10.1038/ s41551-024-01185-8.
12. Edisa Lozic and Benjamin Stular. Fluent but not factual: A comparative analysis of chatgpt and other ai chatbots' proficiency and originality in scientific writing for humanities. Future Internet, 15(10), 2023. ISSN 1999-5903. doi: 10.3390/fi15100336. URL https://www.mdpi.com/1999-5903/15/10/336
13. Brady D. Lund, Ting Wang, Nishith Reddy Mannuru, Bing Nie, Somipam Shim- ray, and Ziang Wang. ¡scp¿chatgpt¡/scp¿and a new academic reality:¡scp¿artificial intelligence-written¡/scp¿research papers and the ethics of the large language mod- els in scholarly publishing. Journal of the Association for Information Science and Technology, 74(5):570–581, March 2023. ISSN 2330-1643. doi: 10.1002/asi.24750. URL http://dx.doi.org/10.1002/asi.24750.
14. Yongqiang Ma, Jiawei Liu, Fan Yi, Qikai Cheng, Yong Huang, Wei Lu, and Xiaozhong Liu. Is this abstract generated by ai? a research for the gap be- tween ai-generated scientific text and human-written scientific text. arXiv preprint arXiv:2301.10416, 2023.
15. Amama Mahmood, Junxiang Wang, Bingsheng Yao, Dakuo Wang, and Chien-Ming Huang. User interaction patterns and breakdowns in conversing with llm-powered voice assistants. International Journal of Human-Computer Studies, 195:103406, 2025. ISSN 1071-5819. doi: https://doi.org/10.1016/j.ijhcs.2024.103406. URL https://www.sciencedirect.com/science/article/pii/S1071581924001897.
16. Hua Shen, Chieh-Yang Huang, Tongshuang Wu, and Ting-Hao Kenneth Huang. Convxai: Delivering heterogeneous ai explanations via conversations to support human-ai scientific writing. In Companion publication of the 2023 conference on computer supported cooperative work and social computing, pages 384–387, 2023.
17. Streamlit, Inc. Streamlit – a faster way to build and share data apps. https:// github.com/streamlit/streamlit. Open-source Python framework for interactive web apps; Accessed: 2025-05-XX.
18. Akbiike Toktosunova, Alymbek Ergeshov, Gulzada Esenalieva, Andrei Ermakov, and Ruslan Isaev. Developing an artificial intelligence tool for image generation using a unique dataset with image-to-image functionality. In Proceedings of the International Conference on Computer Systems and Technologies 2024, CompSysTech '24, page 132–136, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400716843. doi: 10.1145/3674912.3674936. URL https://doi.org/10.1145/ 3674912.3674936.
19. Xuyu Yang, Wengxi Li, Matthew G. Lee, Zhuoyang Li, J. D. Zamfirescu-Pereira, and Can Liu. Rambler in the wild: A diary study of llm-assisted writing with speech, 2025. URL https://arxiv.org/abs/2502.05612.

 doi:10.20944/preprints202505.1430.v1

disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.