Educational Support System for Children with Developmental Disabilities through Offline Large Language Models and Retrieval-Augmented Generation

Masaki Kato[1], Makoto Koshino[1], Sakiko Ogoshi[2,*], and Yasuhiro Ogoshi[3]

[1]Department of Electronics and Information Engineering, National Institute of Technology Ishikawa College, Tsubata, Ishikawa, Japan
[2]Department of Electronics and Information Engineering, National Institute of Technology Fukui College, Fukui, Japan
[3]Department of Human and Artificial Intelligent Systems Graduate School of Engineering, University of Fukui, Fukui, Japan
Email: koshino@ishikawa-nct.ac.jp (M.K.); ogoshi@fukui-nct.ac.jp (S.O.); y-ogoshi@u-fukui.ac.jp (Y.O.)
*Corresponding author
Manuscript received February 18, 2024; revised March 4, 2024; accepted March 11, 2024; published July 15, 2024

***Abstract*—In recent years, there has been a growing need for tailored educational and support measures for children with developmental disorders in Japan. Approximately 8.8% of students in regular classes require special support, necessitating close collaboration and information sharing among schools, families, and welfare facilities. However, teachers and supporters face challenges such as instructional skills, workload, and parent interactions, which cannot be resolved simply by increasing years of experience. Collaboration with external specialized institutions is an effective measure, but few schools engage in regular information sharing. This study proposes the use of Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) based on the support information and behavioral history of students accumulated in the Individualized Educational Support System. The proposed approach offers practical solutions to the lack of experience in special education, insufficient collaboration, and workload issues. Experiments with a question-answering system using multiple models confirmed the possibility of providing answers based on content specific to each child. The effectiveness of a summary system in providing individualized support was confirmed through evaluations by actual supporters. The proposed approach demonstrates the potential for addressing the challenges faced in supporting children with developmental disorders in Japan.**

***Keywords*—developmental disabilities, collaborative support, individualized educational support, Large Language Models (LLMs), Retrieval-Augmented Generation (RAG)**

I. Introduction

In Japan, there has been a significant increase in the number of students enrolled in special education classes, which has doubled over the past decade. According to a survey conducted by the Ministry of Education, Culture, Sports, Science, and Technology in December 2022, approximately 8.8% of children, equating to about 540,000 students in elementary schools nationwide, exhibit characteristics that necessitate special support. These characteristics include Attention Deficit Hyperactivity Disorder (ADHD) [1], Autism Spectrum Disorder (ASD) [2, 3], Learning Disability (LD) [4], Specific Learning Disability (SLD), and Developmental Coordination Disorder (DCD). Such a scenario underscores the need for tailored support throughout the various stages of a child's life, from early childhood to adulthood. Furthermore, the Ministry's Course of Study for Special Needs Schools (2017 edition) emphasizes the importance of fostering an environment where families, schools, and support institutions work in collaboration.

Schools are advised to consider the developmental stages of children and students, enhance their language activities, and work toward establishing effective study habits in collaboration with families. The "Vision of Special Needs Education Towards the Establishment of an Inclusive Education System for a Symbiotic Society (Subcommittee on Elementary and Secondary Education in 2012)" by the Central Council for Education highlights the importance of close cooperation between schools and families in supporting children with disabilities. It suggests that schools and families of such children should regularly exchange information about the child's progress and clearly define their respective roles. In line with this, the Triangle Project was launched in 2018, aiming to further promote cooperation among families, education, and welfare sectors, ensuring that children with disabilities and their parents receive seamless support in the community [5].

When preparing individual educational support plans, schools should collaborate with relevant organizations, including medical, welfare, health, labor institutions, and private organizations. Emphasizing the importance of consistent support, the policy states, "It is essential to provide continuous support for students with disabilities, from early childhood through post-graduation, encompassing school life, family life, and community involvement." Therefore, schools must develop individualized educational support plans in collaboration with the student and their parents or guardians. The transition period is particularly critical, with guidelines suggesting that schools should appropriately share information about the student with their next higher education institution or other relevant entities, provided that they have obtained the consent of the individual. Globally, various educational services utilize ICT, examples include systems for sharing knowledge between teachers and students [6], knowledge bases for effective teaching and learning in special education [7], and tools supporting personal learning environments in higher education [8].

Supporting children with developmental disabilities involves unique challenges due to the evolving nature of their conditions and the highly individualized nature of their needs. This complexity makes it difficult to accurately assess and address each child's specific situation. When offering support to these children, one of the primary challenges is that their conditions vary significantly as they develop, and capturing the state of their condition can be challenging due to its extreme variability from one individual to another.

Consequently, it's vital for teachers, guardians, and supporters to dynamically and closely share information tailored to the individual characteristics of each child. Accumulating such information is essential to understand the behavioral characteristics of the children and to provide support that is tailored to their individual needs. When providing support to children with developmental disabilities, caregivers face challenges due to the dynamic nature of these conditions, which evolve with the children's development and vary significantly from individual to individual. This variability makes it challenging to accurately understand each child's specific condition. Consequently, it is essential for teachers, guardians, and supporters to dynamically and closely share information tailored to the individual characteristics of each child. Accumulating such information is essential to understand the behavioral characteristics of the children and to provide support that is tailored to their individual needs.

In Japan, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) and the Ministry of Health, Labour and Welfare (MHLW) are actively working to promote cooperation among family, education, and welfare sectors. This collaboration aims to support children with disabilities and developmental disabilities, as well as their families. In Fukui Prefecture, a system has been developed and tested in a Proof of Concept (PoC). This system facilitates the connection among schools, families, and specialized institutions, enabling the instant and dynamic sharing of information about children's behavior. It also provides educational support services tailored to the individual characteristics of each child.

In previous studies [9, 10], we developed a system that facilitates immediate and dynamic information sharing on children's behavior among schools, parents, and support organizations, thereby streamlining the support process. This system is designed with a human-centered approach and was developed using a spiral model. During the Proof of Concept (PoC) phase, we have been continuously enhancing the system by adding new features and upgrading versions to meet evolving needs.

On the other hand, addressing the long working hours of teachers in the field is equally essential. It is necessary to utilize digital technology to enhance the efficiency of administrative tasks and reduce teachers' workload.

According to the Survey of Teachers' Working Conditions in 2023 released by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), 77.1% of junior high school teachers and 64.5% of elementary school teachers are reported to exceed the government's 45-hour per month limit for overtime work. MEXT acknowledges that, although there has been a decrease in working hours compared to the past, teachers still face long working hours. In light of this, the demonstration experiments conducted so far have highlighted the need for functions that streamline the preparation of documents related to individual educational support, communication with parents, and the transition of support to new teachers, support staff, and successors based on the individual characteristics of students.

In this study, we propose a novel approach that utilizes Large Language Models (LLMs) [11] and Retrieval-Augmented Generation (RAG) [12, 13]. This approach leverages the support information and behavioral history of students, which are accumulated in the Individualized Educational Support System, to generate personalized support recommendations.

Our study investigates how this automated text generation feature can alleviate the workload of busy teachers and support staff. Additionally, we present survey results regarding the educational support needs of supporters.

In related research, a study has been reported that developed a RAG system using pregnancy-related guidelines from India, focusing on medical education in low- and middle-income countries [14]. Additionally, as an application case in the field of education, research has been reported on the development of an AI tutor system that utilizes RAG to construct a response system for student inquiries [15]. However, upon investigation, no prior research applying RAG or LLMs to support for developmental disabilities was found. In this respect, this study can be considered a pioneering effort.

The use of LLMs is in its nascent stages, with its application advancing across numerous fields. Within educational settings, the benefits of utilizing LLMs have been highlighted; however, concerns such as uncritically accepting generated responses have also been noted. In 2023, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan developed interim guidelines for the use of LLMs in primary and secondary education. According to these guidelines, LLMs should be used merely as a starting point, and ultimately, it is essential for teachers and supporters to review and finalize the content themselves. Upon interviewing teachers and experts, it was suggested that it is more accurate and efficient for the person who was responsible before the transition to use the system, rather than having the new person in charge do so. By conducting experimental demonstrations through this research, we have been able to gain important insights for the future introduction and social implementation of such technologies in educational settings.

Furthermore, the guidelines state that tools such as OpenAI ChatGPT, Microsoft Copilot, Google Bard, Google Gemini, and Anthropic Claude may not always meet the necessary security requirements, thus necessitating caution not to input sensitive information when entering prompts. Consequently, as demonstrated in this research, employing local (offline) LLMs and RAG to avoid external data exposure is preferable.

II. System Overview

To address the issues discussed earlier, we developed a web-based collaborative and individualized educational support system named PicotTo. Fig. 1 presents an overview of the system. Fig. 2 shows an example of the user interface, specifically the screen for filling in the checklist items. The system can be utilized by users in various roles, with accessible functionalities determined by their roles. For example, supporters (such as homeroom teachers) can set checklist items tailored to a child's support goals, and record checks and comments on those items. Guardians can review comments made by supporters and perform checks just like supporters. All data related to checks and comments are stored in a database and can be shared immediately. Additionally, the system allows for the download of

personalized educational support documents as CSV files.

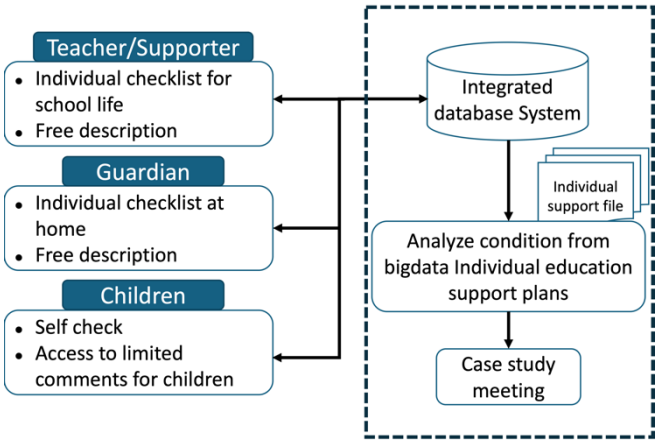

Fig. 1. Diagram of the systems' overview.

The dashed line circled part in Fig. 1 represents the system developed in this study. Before this system was created, approximately every six months, documents compiling the support history, drawing from data stored in PicotTo and the recollections of supporters, were created to formulate educational support plans. However, the production of these documents was found to be a complex and labor-intensive process. In this study, we developed two systems to assist in analyzing the conditions of children based on the history in PicotTo. One system employs Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs) to perform question-answering based on children's past information. The other system uses LLMs to summarize past comments.

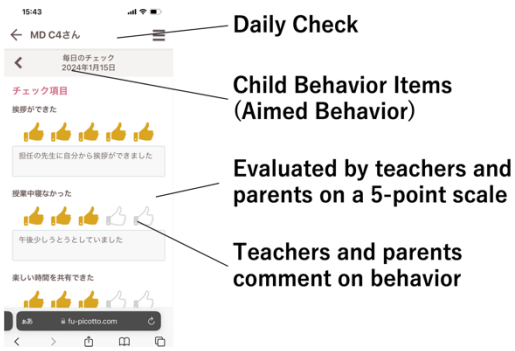

Fig. 2. Example of the user interface.

III. Methods

Large Language Models (LLMs) [11] have been trained using a vast corpus. They learn grammar and knowledge from a tremendous amount of text and can perform various natural language processing tasks, not limited to specific tasks but including translation, summarization, and question-answering.

Retrieval-Augmented Generation (RAG) [12, 13] involves searching for information relevant to a question from external sources before LLMs perform question-answering. RAG has been proven to significantly improve answer accuracy, particularly in knowledge-intensive tasks, and to reduce the frequency of incorrect responses.

First, the target external information is converted into text and split into appropriate units (Fig. 3). In this study, a single unit is defined as the comments stored in PicotTo retrieved on a daily basis. These split texts are then transformed into vectors using a text embedding model. When a user poses a question, it is also converted into a vector using a similar document embedding model. The inner product between the vectors of this query and the split external information is calculated to determine their similarity. Documents with high similarity, along with the query, are used as inputs to perform question-answering with the LLMs (Fig. 4).

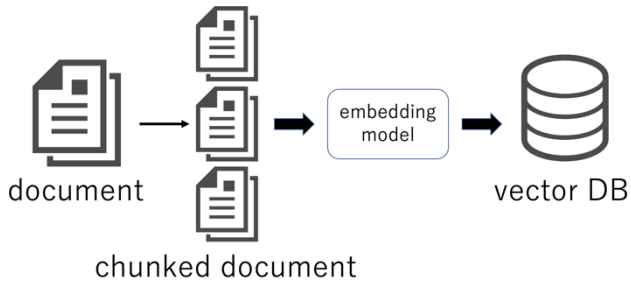

Fig. 3. Creation of a vector database.

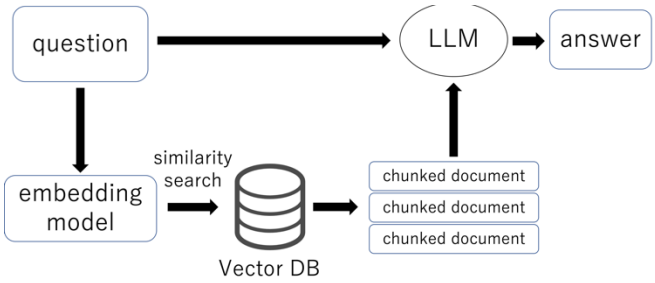

Fig. 4. System diagram of RAG with improved information retrieval.

In PicotTo, comments shared among supporters and comments directed towards the children are stored. These comments are segmented by date and by supporter and saved in a vector database. In this study, data from a six-month period pertaining to a specific child were utilized.

Due to the potential inclusion of personal information in the data used for supporting children with developmental disabilities, services such as the OpenAI API cannot be employed for LLMs or embedding models. Therefore, it is necessary to utilize LLMs and embedding models that can operate in a local environment. Several LLMs capable of local operation have been released, among which Llama2 is widely known [16]. In this study, we use models that have been fine-tuned with Llama2, including Xwin-LM-70B-V0.1-GPTQ, Xwin-LM-13B-V0.2, Swallow-70B-instruct-GPTQ, and Swallow-13B-instruct-hf. These models have demonstrated improved capabilities in generating Japanese text due to the fine-tuning. Xwin-LM-70B-V0.1-GPTQ and Swallow-70B-instruct-GPTQ are models that have been subjected to a quantization technique called GPTQ [17], applied respectively to Xwin-LM-70B-V0.1 and Swallow-70B-instruct-hf. GPTQ is a quantization technology that can compress the bit-width per model weight to approximately 4 bits while retaining the performance of the original model. This allows the models to be used in environments with limited computational resources. For the text embedding model, "multilingual-e5-large" is utilized.

IV. Question-Answering Based on Children's Past Information with RAG

To address the following six anticipated questions,

experiments were conducted using data from six months of records. These records, entered by teachers, parents, and supporters, detailed the behavior of a 15-year-old male student with a developmental disability. The observations and interventions documented by these records were critical for supporting the student. An evaluation of these data was carried out by a supporter who had been closely assisting the student.

1) When does the child's condition worsen?
2) What kind of support is effective for the child in question?
3) In what situations does the child fall asleep during class?
4) Does the child experience panic attacks?
5) When does the child exhibit optimal performance?
6) What are the differences between the child's good and bad days?

Table 1 presents the evaluation results for each model. A response was deemed correct if it was customized to the individual child, based on the provided external information.

Table 1. Evaluation of each model for every question

	Q1	Q2	Q3	Q4	Q5	Q6	Accuracy
Xwin-LM-70B -V0.1-GPTQ	○	○	○	○	○	○	100%
Swallow-70B -instruct-GPTQ	○	x	○	○	○	x	66%
Xwin-LM -13B-V0.2	○	x	○	x	x	x	33%
Swallow-13B -instruct-hf	x	x	x	x	x	x	0%

Xwin-LM-70B-V0.1-GPTQ achieved the highest accuracy rate. Other models often provided advice applicable to general cases or answered questions with misunderstood intentions, leading to incorrect responses. Examples of question responses from Xwin-LM-70B-V0.1-GPTQ are presented below.

Q1. When does the child's condition worsen?

While there are various situations that can cause discomfort, we can identify specific scenarios from the provided information. For example, on a certain day, when subject P1 went to a movie, the bright screen and loud sound made them feel unwell. This suggests that the child may be sensitive to situations with intense light or noise. However, it's important to note that the causes of discomfort can vary individually. Identifying the exact situations that cause discomfort requires considering the child's own experiences and the context they're in.

Q4. Does the child experience panic attacks?

Yes, there are instances when panic occurs. Panic can sometimes manifest in children who are capable of handling highly challenging situations. For example, in the episode on a certain day, the child felt unwell during a music class and asked to go to the bathroom. In such situations, the child may experience panic.

However, in this case, the child appeared calm after going to the toilet, suggesting recovery from the panic state. Furthermore, when a supporter asked the child, "Are you okay?", the child responded, "I feel a bit calmer now." This indicates that the child is able to understand their condition and communicate it appropriately.

Even if the child experiences panic, it's possible to support them in returning to a stable state by the supporter remaining calm and communicating effectively with the child.

It's important to note that, in the system, not all questions will receive correct answers, and sometimes irrelevant external information might be referenced. Therefore, it's crucial to inform actual users about the possibility of incorrect answers and to encourage them to verify the external information referenced to assess the validity of the responses.

A possible direction is to create a RAG system that utilizes comments on specific check items for each child and texts documenting actual support provided, enabling more detailed information sharing.

V. Create Summaries of Comments with LLMs

To enhance the efficiency of analysis related to children and reduce the workload of teachers, we summarize comments stored in PicotTo using LLMs.

We retrieve six comments and input them into LLMs to generate summaries. Subsequently, we integrate these summaries and input them into LLMs again to generate a monthly summary.

We conducted an experiment using comments shared among supporters over a six-month period for a specific child. The model used was Xwin-LM-70B-V0.1-GPTQ.

We show a portion of the generated summaries.

Positive aspects:

- The ability to consider others' intentions and situations is evident, as seen in spontaneous expressions of greetings, such as saying "Good morning."
- Engagement in non-academic activities, like cooking and enjoying the illustrations in picture books, is commendable.
- The cooperation among supporters to provide activities tailored to individual interests and preferences proves to be effective.

Areas needing improvement:

- If the content of group moral lessons is challenging to comprehend, alternative methods or formats of communication should be explored.
- Encouraging interaction with other children and participation in group activities is important, even when the focus is intense.
- Attention to health management and providing physical and mental support are necessary when dealing with instability caused by hay fever.

Monthly summaries of the child's condition were generated to provide an overview of the most notable events and progress during each month as follows: Notable events include the student spontaneously saying, "Good morning," showing consideration for others' intentions and situations. His engagement in non-academic activities, such as enjoying cooking activities and picture book illustrations, has been positively evaluated. Additionally, the cooperation among supporters to provide activities aligned with the student's interests and preferences is commendable. On the other hand, areas needing improvement include exploring alternative methods or formats of communication when the content of a group moral lesson is unclear to the student. It's also important to encourage interaction with other children and group activities. Furthermore, when the student is unstable due to hay fever, attention to health management and physical and mental support is necessary.

The summaries of the six comments use specific events as a basis to compile the child's behavior. Moreover, the monthly summaries effectively highlight the most notable events of each month. As demonstrated in the following chapters, the need for using LLMs for tasks like summarizing text is increasing. Summaries like those generated in this experiment could be valuable for individualized education plans, transferring teaching plans, and reducing the workload of teachers. Moving forward, we plan to conduct interviews with supporters and experts.

VI. Interview Survey

The implementation of Large Language Models (LLMs) with Retrieval-Augmented Generation (RAG) for educational support in the field of education is still in its nascent stages. Consequently, understanding teachers' perceptions of using such AI for educational support becomes crucial, especially since comprehensive knowledge in this area is still limited. To gain insights into these perceptions, we conducted an interview survey with experienced teachers and supporters of children with developmental disabilities who have participated in the use of the system during the Proof of Concept (PoC) of this research development.

Outlined below are the findings from the interview survey, which focused on teachers responsible for special support in schools and supporters in welfare facilities. The survey specifically addressed their views on the automatic sentence generation function of AI within PicotTo, an individualized education support system.

We conducted an interview survey focusing on the following four aspects:

1) Do you believe an AI-based automatic text generation feature should be integrated into PicotTo?
2) In which situations would you find an AI-based automatic text generation feature useful within PicotTo?
3) Should PicotTo include an AI-based chatbot functionality?
4) If PicotTo were to implement an AI-based messaging summarization feature for parents, what capabilities do you think it should encompass?
    a) Generating positive sentences.
    b) Using polite language.
    c) Avoiding discriminatory language.
    d) Crafting concise sentences.
    e) Summarizing growth areas.

Additionally, we welcome other opinions regarding AI-based features.

We conducted interviews with four professionals from educational and welfare institutions, including a skilled supporter from a welfare institution, an experienced teacher, a specialist, and a new supporter from a welfare institution.

1) All interviewees expressed the desire for PicotTo to include an AI-based automatic sentence generation function. The primary reason cited was workload reduction.
2) The responses to the second question included:
- Utilization in creating individualized educational support plans.
- Reference for parent communication and for record-keeping.
- For children newly receiving support, it is beneficial to understand their characteristics based on past records from other institutions.
- A feature that summarizes behaviors and changes over specific periods would be advantageous.

3) All respondents were in favor of incorporating an AI-based chatbot function into PicotTo.
4) Regarding the AI-based message summarization functions for parents, see Table 2.

Table 2. AI-based message summarization functions for parents

Function	#
Generates sentences using positive language	3
Concise sentence construction	3
Excluding discriminatory language	3
Polite language	1

A need was identified for a function focused on summarizing the child's growth and identifying their strengths. Additional feedback suggested that it would be beneficial if reports could be used for individual educational support plans and for instructing younger staff. It was also mentioned that since all staff members utilize the system, it should not be apparent who is adept at writing and who is not. One interviewee, recalling their time supporting at a child daycare center, noted that considerable time was spent completing contact sheets and facility records, which detracted from engaging with the children. On days when time was short, they would have to complete these tasks after the children had left. An automatic sentence generation feature that could craft statements from daily evaluations or brief summaries of activities and outcomes would be highly valuable. Furthermore, it was suggested that the chatbot function could assist new supporters or those seeking to understand a child's characteristics or progress to date, with queries such as wanting to know about the child's previous experiences or the outcomes of similar past activities

VII. Conclusions

This study developed two systems to facilitate information coordination among supporters of children with developmental disabilities and provide personalized support. The first system utilizes Retrieval-Augmented Generation (RAG) for a question-answering system based on individual student information, while the second system employs Large Language Models (LLMs) for summarizing individual student information. Experiments with the question-answering system, using sample questions and multiple models, confirmed the possibility of providing answers based on content specific to each child. Evaluations by actual supporters confirmed the effectiveness of the summary system in providing individualized support.

The survey results indicated a need for a function that utilizes AI to automatically summarize behavioral history, particularly for skilled teachers and supporters who have been involved in the Proof of Concept (PoC) to date. There was significant anticipation for the application of RAG to independently segregate and utilize data, and to facilitate the transfer of information to subsequent entities. Moreover, there were high expectations for such technology to reduce the workload involved in preparing individual educational support plans. In Japan, where individualized educational support plans and instructional plans are mandatory,

streamlining this part of the workload could significantly alleviate the burden on teachers. The survey also revealed a strong demand for system usage and a keen interest among teachers regarding its utilization and operation. The future operation mode of the system is an area of focus, necessitating ongoing adaptations to reflect teachers' needs and discussions with the Board of Education and other stakeholders.

A limitation of this study is the difficulty in conducting experiments with multiple children and students. This constraint is primarily due to two important factors. The first is the need for ethical approval and adequate consent for experiments involving data on children and students. The second is the challenge of efficiently collecting and recording information on children and students, due to the lack of widespread systems, making it difficult to obtain diverse experimental data. An additional critical limitation of this study arises from the propensity of LLMs to produce hallucinations [18], which are instances where the generated text includes inaccurate information not grounded in user inputs or existing facts. This phenomenon significantly undermines the reliability of text generation, posing a substantial challenge to achieving completely accurate and trustworthy outputs.

As a future research direction, it is proposed to construct a system that not only incorporates PicotTo's comments but also utilizes documents recording the actual support provided, enabling more detailed information sharing.

Currently, it's crucial for the social implementation of the functions highlighted in this report to validate their effectiveness in facilitating effective student support while alleviating the workload of teachers and parents. Gaining an understanding of the use of LLMs in educational and welfare support is fundamental. We believe that incorporating an automatic sentence generation function using LLMs into an ICT-based individualized education support system will significantly contribute to realizing more effective education in today's fast-paced society. We aim to develop this system for broader social implementation, fostering a collaborative environment between the educational field and the system. This collaboration is key to achieving a society where no student or parent is left behind, ultimately benefiting all involved.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Masaki Kato developed the system and authored the paper. Makoto Koshino oversaw the research and provided revisions to the manuscript. Sakiko Ogoshi and Yasuhiro Ogoshi have been involved in the long-term development of both the system and the research. All authors have approved the final version of the manuscript.

Acknowledgment

This work was supported by KAKENHI (Grant-in-Aid for Scientific Research (C): 15K01752, 15K04553, 18K02896, 19K12245, 22K12283, 22K02786) and SCOPE: 162305008, Hitachi Solutions East Japan, Ltd.

References

[1] G. Polanczyk, M. S. Lima, B. L. Horta, J. Biederman, and L. A. Rohde, "The worldwide prevalence of ADHD: A systematic review and metaregression analysis," *The American Journal of Psychiatry*, vol. 164, no. 6, pp. 942–948, 2007.

[2] I. Hertz-Picciotto and L. Delwiche, "The rise in autism and the role of age at diagnosis," *Epidemiology*, vol. 20, no. 1, pp. 84–90, 2009.

[3] N. Nassar *et al*., "Autism spectrum disorders in young children: effect of changes in diagnostic practices," *International Journal of Epidemiology*, vol. 38, no. 5, pp. 1245–1254, 2009.

[4] M. Altarac and E. Saroha, "Lifetime prevalence of learning disability among US children," *Pediatrics*, 119 Supplement, pp. S77–S83, 2007.

[5] Recent Policy and Status on Special Needs Education in Japan, *NISE Bulletin*, vol. 18, 2019.

[6] S. Rajalakshmi and R. S. D. Wahidabanu, "Sharing and capturing tacit knowledge in higher education—The Info-Ca-SH," *IJCTE*, vol. 3, no. 3, pp. 365–368, 2011.

[7] S. K. Chettiar, C. J. Venkateswaran, and A. Clementking, "Special need education tool design using cubical database," *IJCTE*, vol. 3, no. 3, pp. 457–462, 2011.

[8] N. M. Tuah, D. S. Ajor, and N. Jamil, "Assistive tools towards personal learning environment in higher education," *IJCTE*, vol. 7, no. 5, pp. 398–401, 2015.

[9] Y. Ogoshi, S. Ogoshi, T. Saitoh, Y. Itoh, M. Wakamatu, T. Kanno, H. Hongo, "Proof of concept: ICT system PicotTo for individual education support—A case study of junior high school," presented at IEEE ICBDA2022 7th International Conference on Big Data Analytics, 2022.

[10] S. Ogoshi, Y. Ogoshi, T. Saitoh, Y. Itoh, M. Wakamatu, T. Kanno, A. Suzuki, and A. Nakai, "Development of an educational support system using ICF codes and investigation of system usage," presented at IEEE ICBDA2023 8th International Conference on Big Data Analytics, 2023.

[11] X. Zhao *et al*., "A survey of large language models," arXiv:2303.18223, 2023.

[12] P. Lewis *et al*., "Retrieval-augmented generation for knowledge-intensive NLP tasks," *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 2020.

[13] Y. Gao *et al*., "Retrieval-augmented generation for large language models: A survey", arXiv:2312.10997, 2023.

[14] Y. Ghadba *et al*., "Transforming healthcare education: Harnessing large language models for frontline health worker capacity building using retrieval-augmented generation," medRxiv, https://doi.org/10.1101/2023.12.15.23300009, 2023.

[15] D. Chenxi. "How to build an ai tutor that can adapt to any course and provide accurate answers using large language model and retrieval-augmented generation", arXiv: 2311.17696, 2023.

[16] H. Touvron *et al*., "Llama 2: Open foundation and fine-tuned chat models," arXiv:2307.09288, 2023.

[17] E. Frantar *et al*., "GPTQ: Accurate post-training quantization for generative pre-trained transformers," arXiv:2307.09288, 2023.

[18] Y. Zhang *et al*., "Siren's song in the AI ocean: A survey on hallucination in large language models," arXiv: 2210.17323, 2023.

Copyright © 2024 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).