Journal of Dental Sciences 20 (2025) 2460–2466

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.e-jds.com

Original Article

Analyzing the performance of multimodal large language models on visually-based questions in the Japanese National Examination for Dental Technicians

Yuichi Mine [a,b*], Tsuyoshi Taji [c], Shota Okazaki [a,b], Saori Takeda [b], Tzu-Yu Peng [d], Saiji Shimoe [e], Masato Kaku [e], Hiroki Nikawa [c], Naoya Kakimoto [f], Takeshi Murayama [a,b]

[a] *Project Research Center for Integrating Digital Dentistry, Hiroshima University, Hiroshima, Japan*
[b] *Department of Medical Systems Engineering, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan*
[c] *Department of Oral Biology & Engineering, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan*
[d] *School of Dentistry, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan*
[e] *Department of Anatomy and Functional Restorations, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan*
[f] *Department of Oral and Maxillofacial Radiology, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan*

Received 13 February 2025; Final revision received 24 February 2025
Available online 6 March 2025

KEYWORDS
Multimodal large language models;
Dental technician licensing examination;
Image-based questions;
ChatGPT-4o;
OpenAI o1;
Claude 3.5 sonnet

Abstract *Background/purpose:* Large language models (LLMs) offer promising applications in dentistry, but their performance in specialized, image-rich contexts such as dental technology examinations remains uncertain. The purpose of this study was to evaluate the accuracy of three multimodal LLMs, ChatGPT-4o (4o), OpenAI o1 (o1), and Claude 3.5 Sonnet (Sonnet), when presented with questions from the Japanese National Examination for Dental Technicians.

Materials and methods: A total of 240 multiple-choice questions from 2022 to 2024 theory sections of the exam were used. Each question, including its accompanying figures or images where applicable, was presented to the three LLMs in a zero-shot manner without specialized prompting. Correct response rates were calculated overall, as well as by question type (text-only vs. visually-based) and subject area. Statistical comparisons were performed using Cochran's Q test, followed by McNemar's test with Bonferroni correction where indicated.

* Corresponding author. Department of Medical Systems Engineering, Graduate School of Biomedical and Health Sciences, Hiroshima University, 1-2-3 Kasumi Minami-ku, Hiroshima 734-8553, Japan.
E-mail address: mine@hiroshima-u.ac.jp (Y. Mine).

https://doi.org/10.1016/j.jds.2025.02.022

Results: Overall correct response rates were 58.3 % (4o), 67.5 % (o1), and 64.6 % (Claude 3.5 Sonnet). For text-only questions, o1 achieved the highest accuracy (79.1 %), significantly outperforming 4o (68.3 %; $P = 0.017$). In contrast, all models showed reduced accuracy on visually-based questions (44.6—55.4 %), with no significant difference among them.
Conclusion: These results suggest that multimodal LLMs can supplement theoretical dental technology education, although their limited performance on visual tasks indicates the need for traditional hands-on training. Enhanced image interpretation skills may help address workforce challenges in dental technology.

Introduction

Artificial intelligence (AI) has rapidly developed in many fields, including dentistry, due to continuous advances in computing power and algorithmic innovation.[1] Among the various AI-driven technologies, large language models (LLMs) (*e.g.* ChatGPT, Claude and Gemini) have attracted interest for their ability to perform complex natural language processing tasks such as automated question answering, text generation, and content summarization.[2,3] By training on large amounts of text data, these models are able to generate highly consistent, contextually relevant responses with a high degree of fluency.[4] In dentistry, LLMs have demonstrated potential applications in automated dental charting, symptom classification, and dental education.[5] Particularly in dental education, while LLMs show promise in enhancing educational materials and supporting more individualized learning approaches, careful consideration must be given to maintaining critical thinking and problem-solving skills.[5–7]

The Japanese National Examination for Dental Technicians comprehensively tests the knowledge and skills required to fabricate and repair dental prostheses, orthodontic appliances, and other related devices.[8,9] Mastering these diverse subjects presents significant challenges for students, increasing the need for effective and accessible learning tools. Meanwhile, Japan faces a potential shortage of dental technicians due to demographic trends and declining enrollment in dental technician schools.[8] Although the demand for prosthetic restorations may remain steady or change only gradually in an aging society, the reduced supply of newly qualified technicians, coupled with the closure of several dental technician schools, raises concerns about maintaining adequate workforce levels.[8] In addition, a recent Taiwanese study that examined the educational status of dental technician students found a declining enrollment trend, raising concerns about a potential shortage of these professionals.[10] The authors also noted that creating more favorable working conditions could help attract new entrants to the field. Although the context in Taiwan is different, the risk of a shortage is similar to the situation in Japan, demonstrating the importance of proactive strategies to maintain a robust supply of skilled dental technicians.

Continuous professional development and self-learning are crucial for newly qualified dental technicians. While traditional instruction and hands-on training remain essential, young professionals often need additional resources for theoretical reinforcement and problem-solving support in their daily practice. Against this background, the purpose of this study was to evaluate the performance of three multimodal large language models when presented with questions from the Japanese National Examination for Dental Technicians. Specifically, we investigated the accuracy and error patterns of the models in various content areas and discussed the implications for the integration of AI-driven tools in dental technician education and skill development of new dental technicians.

Materials and methods

Dataset

In this study, the Japanese National Examination for Dental Technicians was used for the three-year period from 2022 to 2024. This exam is divided into a theory section and a practical skills section.[8,9] In this study, only questions from the theory section were included. The theory section consists of 80 multiple-choice questions, one of which is selected by the examinee. A total of 240 questions were used as the data set. The questions include dental materials science, oral and maxillofacial, stomatognathic function science, dental technology for removable dentures, dental technology for fixed dental prostheses and restorations, dental technology for orthodontic appliances, dental technology for pedodontic appliances, laws and regulations for dental technicians, and dental laboratory and practice administration. The subjects of the questions were determined based on the National Examination for Dental Technicians Question Book edited by Japan Society for Education of Dental Technology (Tokyo, Japan). The exam includes a variety of photographs, such as dental restorations and appliances, the process of its fabrication, and the instruments used in its fabrication. The characteristics of the dataset are shown in Table 1.

Multimodal large language models and prompting

Three multimodal LLMs were used in this study to examine their performance on the Japanese National Examination for Dental Technicians, which featured questions

Table 1 The number of questions by category and subject in the Japanese National Examination for Dental Technicians.

		2022	2023	2024	Total
Question category	All questions	80	80	80	240
	Text-only questions	46	46	47	139
	Visually-based questions[a]	34	34	33	101
Subject	Laws and regulations for dental technicians Dental laboratory and practice administration	3	3	3	9
	Dental materials science	14	14	14	42
	Oral and maxillofacial anatomy	10	10	10	30
	Stomatognathic function science	5	5	5	15
	Dental technology for removable dentures	20	20	20	60
	Dental technology for fixed dental prostheses and restorations	18	18	18	54
	Dental technology for orthodontic appliances	5	5	5	15
	Dental technology for pedodontics appliances	5	5	5	15

[a] Includes one or more images, figures, or tables.

incorporating figures, tables, and images. The chosen models were ChatGPT-4o (4o; OpenAI Global, San Francisco, CA, USA; released May 13, 2024), OpenAI o1 (o1; OpenAI; released December 5, 2024), and Claude 3.5 Sonnet (Sonnet; Anthropic, San Francisco, CA, USA; updated October 22, 2024). This choice was informed by the results of the MM-Vet v2 benchmark, a comprehensive evaluation framework designed to assess the integrated capabilities of multimodal LLMs.[11] MM-Vet v2 measures seven core capabilities including image-text sequence understanding. According to this benchmark, the Sonnet scored highest, followed by the 4o. o1 was not released at the time of this paper, but was chosen because it is a newer multimodal LLM than 4o. Each LLM can process both textual and visual data, enabling them to handle a diverse array of question formats.

In this study, we used a zero-shot approach,[12] deliberately avoiding any special prompt design or additional instructions for the models. The original Japanese text of each question, along with its set of answer choices, was entered into the prompt window as provided. When questions included figures, tables, or images, these visual materials were presented to the multimodal LLMs in their original form, without any additional explanatory text.

Statistical analysis

All analyses were performed using IBM SPSS Statistics 27 (IBM SPSS, Inc., Armonk, NY, USA). Cochran's Q test was used to compare correct response rates among the three models. If a significant difference was detected, pairwise comparisons were performed using McNemar's test. Because three such comparisons were carried out, the Bonferroni correction was used to keep the overall family-wise error rate at 0.05.

Results

Of the 240 questions selected from the 2022—2024 Japanese National Examination for Dental Technicians, 139 were text-only and 101 included one or more images, figures, or tables (Table 1). The three multimodal LLMs showed different levels of accuracy overall and by question type (Table 2). Across all 240 questions, 4o achieved an overall correct response rate of 58.3 % (95 % CI: 51.8—64.6), o1 67.5 % (61.2—73.4), and Sonnet 64.6 % (58.2—70.6). Pairwise comparisons with Bonferroni correction showed that o1 outperformed 4o with statistical significance ($P = 0.019$), while the differences between o1 and Sonnet ($P = 1.000$) and between Sonnet and 4o ($P = 0.134$) were not significant. Considering only the 139 text-only items, the highest correct response rate (79.1 %, 71.4—85.6) was achieved by o1, which significantly exceeded 4o's 68.3 % (59.9—76.0; $P = 0.017$ after Bonferroni correction), while Sonnet fell in between with 71.2 % (62.9—78.6). All three LLMs performed worse on the remaining 101 visually-based questions, achieving correct rates of 44.6 % (34.7—54.8) for

Table 2 Correct response rates (%) and 95 % CIs of the three LLMs.

	4o	o1	Sonnet	o1 vs 4o	o1 vs Sonnet	Sonnet vs 4o
All questions	58.3 (51.8—64.6)	67.5 (61.2—73.4)	64.6 (58.2—70.6)	$P = 0.019$	$P = 1.000$	$P = 0.134$
Text-only questions	68.3 (59.9—76.0)	79.1 (71.4—85.6)	71.2 (62.9—78.6)	$P = 0.017$	$P = 0.178$	$P = 1.000$
Visually-based questions[a]	44.6 (34.7—54.8)	51.5 (41.3—61.6)	55.4 (45.2—65.3)	N.S.	N.S.	N.S.

N.S., Not significant (Cochran's Q test); LLMs, Large language models; 4o, ChatGPT-4o; o1, OpenAI o1; Sonnet, Claude 3.5 Sonnet. Because Cochran's Q test did not indicate a significant difference among the samples, no multiple comparisons were performed.
[a] Includes one or more images, figures, or tables.

4o, 51.5 % (41.3—61.6) for o1, and 55.4 % (45.2—65.3) for Sonnet. Cochran's Q test revealed no significant overall difference between the three LLMs for these visually-based items ($P > 0.05$).

Table 3 and Fig. 1A—C presents analyses by specialty for all questions, text-only questions, and visual-based questions, respectively. Although 4o's correct response rate was moderate overall (58.3 %), its performance varied considerably by specialty. In dental technology for orthodontic appliances (n = 15), it scored only 20.0 % overall, compared to 46.7 % for o1 and 73.3 % for Sonnet.

This gap was particularly significant in text-only orthodontic questions (n = 7), with scores of 14.3 % vs. 71.4 % and 85.7 %, respectively. o1 tended to perform better in areas such as oral and maxillofacial anatomy (n = 30), where it achieved a correct response rate of 76.7 %, outperforming both 4o (56.7 %) and Sonnet (73.3 %), and dental materials science (n = 42), where its 81.0 % accuracy slightly outperformed the other two models. Conversely, Sonnet showed its highest overall accuracy in stomatognathic function science (n = 15), reaching 66.7 %, outperforming 4o (53.3 %) and o1 (46.7 %).

Additional differences were found when text-only and visual questions were examined separately. For the text-only questions, o1 generally outperformed both 4o and Sonnet, although the margin of difference varied by subject. For example, in dental materials science text-only questions (n = 31), o1's accuracy (87.1 %) exceeded that of Sonnet (77.4 %) and 4o (83.9 %), but the gap between the models was relatively modest. All three LLMs had lower percentages of correct responses on the visually-based questions, especially on items that required interpreting multiple images or making fine distinctions based on subtle visual details. In these visually rich scenarios, Sonnet outperformed the other two models in dental materials science (n = 11; 81.8 % vs. 63.6 % and 54.5 %) and orthodontic

Table 3 Comparing correct response rates (%) of three LLMs in different specialties on all questions, text-only questions, and visually-based questions.

Question category	Subject	Questions (n)	4o	o1	Sonnet
All questions	Overall	240	58.3	67.5	64.6
	Laws and regulations for dental technicians, dental laboratory and practice administration	9	55.6	77.8	77.8
	Dental materials science	42	76.2	81.0	78.6
	Oral and maxillofacial anatomy	30	56.7	76.7	73.3
	Stomatognathic function science	15	53.3	46.7	66.7
	Dental technology for removable dentures	60	60.0	61.7	55.0
	Dental technology for fixed dental prostheses and restorations	54	63.0	70.4	63.0
	Dental technology for orthodontic appliances	15	20.0	46.7	73.3
	Dental technology for pedodontics appliances	15	33.3	60.0	33.3
Text-only questions	Overall	139	68.3	79.1	71.2
	Laws and regulations for dental technicians, dental laboratory and practice administration	7	71.4	71.4	85.7
	Dental materials science	31	83.9	87.1	77.4
	Oral and maxillofacial anatomy	20	70.0	85.0	75.0
	Stomatognathic function science	7	71.4	71.4	71.4
	Dental technology for removable dentures	37	70.3	78.4	62.2
	Dental technology for fixed dental prostheses and restorations	22	68.2	72.7	72.7
	Dental technology for orthodontic appliances	7	14.3	71.4	85.7
	Dental technology for pedodontics appliances	8	37.5	75.0	50.0
Visually-based questions[a]	Overall	101	44.6	51.5	55.4
	Laws and regulations for dental technicians, dental laboratory and practice administration	2	0.0	100.0	50.0
	Dental materials science	11	54.5	63.6	81.8
	Oral and maxillofacial anatomy	10	30.0	60.0	70.0
	Stomatognathic function science	8	37.5	25.0	62.5
	Dental technology for removable dentures	23	43.5	34.8	43.5
	Dental technology for fixed dental prostheses and restorations	32	59.4	68.8	56.3
	Dental technology for orthodontic appliances	8	25.0	25.0	62.5
	Dental technology for pedodontics appliances	7	28.6	42.9	14.3

LLMs, Large language models; 4o, ChatGPT-4o; o1, OpenAI o1; Sonnet, Claude 3.5 Sonnet.
[a] Includes one or more images, figures, or tables.

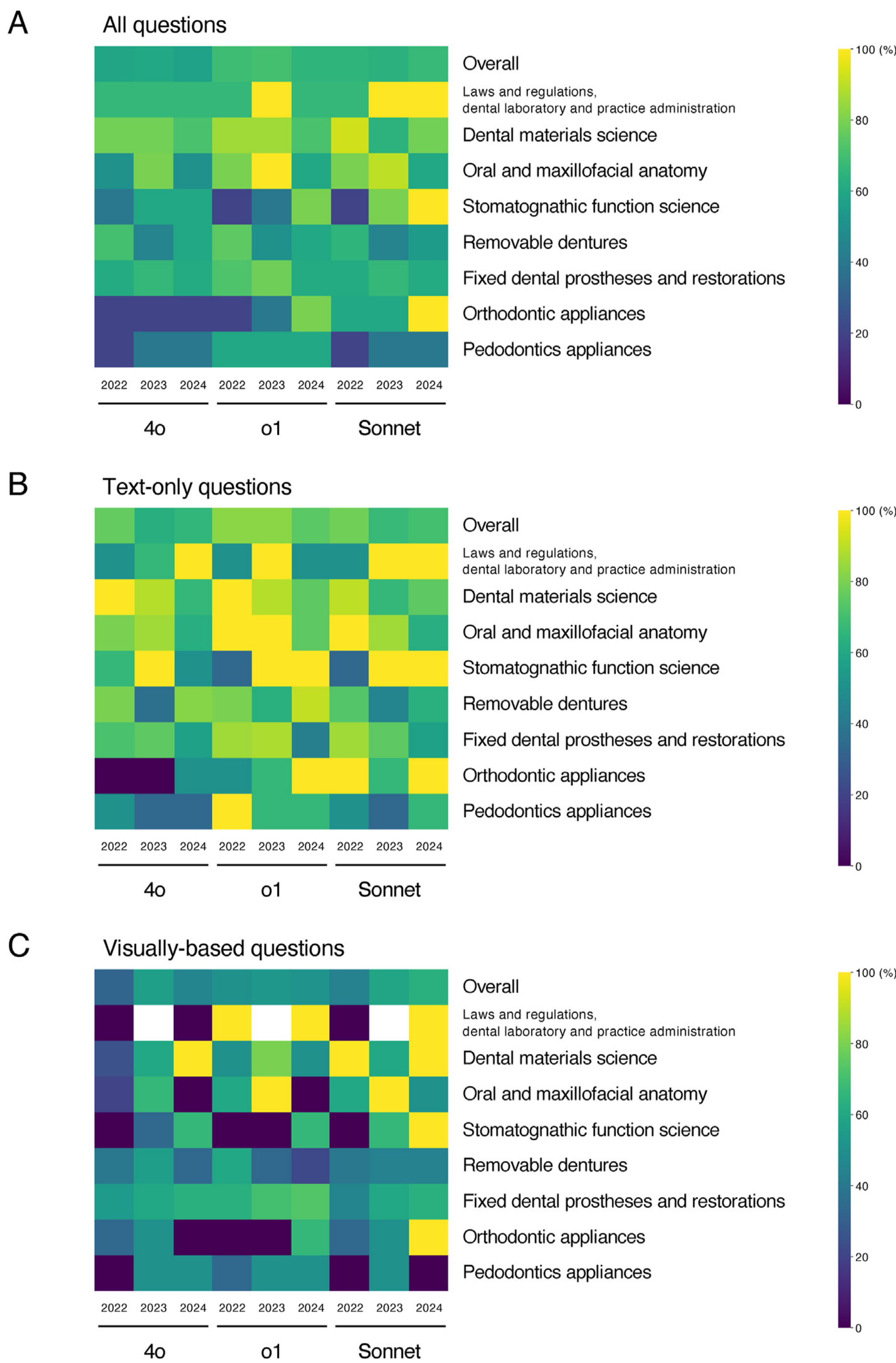

Figure 1 Heatmaps of correct response rates among the three LLMs across the 2022—2024 Japanese National Examination for Dental Technicians. (A) All questions (B) Text-only questions (C) Visually-based questions (white cells indicate that no question was available for that subject area in a given year). Rows represent subject areas, and columns correspond to each LLM. Warmer colors (yellow) indicate higher accuracy, while cooler colors (purple) indicate lower accuracy. LLMs, Large language models; 4o, ChatGPT-4o; o1, OpenAI o1; Sonnet, Claude 3.5 Sonnet.

appliances (n = 8; 62.5 % vs. 25.0 % each for 4o and o1), although its lead was narrower or reversed in other specialties.

The heat maps in Fig. 1 visually summarize these trends for all questions (A), text-only questions (B), and visually-based questions (C) from 2022 to 2024. The color gradients indicate that o1 was the most consistent high performer across many subject areas in the text-only questions, while 4o's performance was generally average, but particularly weak in orthodontics. Sonnet's results were somewhat heterogeneous, with strong performance in stomatognathic function and orthodontics contrasting with more moderate accuracy in other areas. These results show how the strengths and weaknesses of each model vary according to both the specialty in question and the presence or absence of visual materials.

Discussion

Dental technicians play an important role in dental care with dentists, dental hygienists, and dental assistants. As a member of the dental team, the activities of dental technicians are aimed at repairing defects in the dentition and providing services through prosthetic structures, orthodontic appliances and splints.[9] So far, LLM performance has been reported for dentists[7,13–15] and dental hygienist examinations,[6,16] and detailed insights are expected in the context of dental technician education. This study evaluated the performance of three multimodal LLMs on the Japanese National Examination for Dental Technicians with visually-based questions, providing insight into their capabilities and limitations in dental technology education.

The Japanese National Examination for Dental Technicians is characterized by its extensive use of photographs and visual materials. Many exam questions include images of technical instruments, cross sections of fabricated items, and arrangements of artificial teeth, with an emphasis on practical knowledge. Because the questions are designed to ask the examinee about precise technical procedures and the final state of a finished product, they often focus on images that illustrate specific tasks. Another feature is the focus on testing practical skills that are directly related to real-world tasks, such as the correct process for creating dentures, selecting appropriate materials, and properly handling instruments. For example, test items may require examinees to identify problem areas in a pictured dental appliance or select the proper fabrication workflow. In addition, there is a strong tendency to test whether examinee can accurately explain specific steps in the fabrication process using photographs or diagrams and demonstrate a correct understanding of each step. These visuals often use color-coding or symbols such as red lines or marks to highlight important areas, and exam questions typically refer back to these marked sections, asking candidates to identify or discuss the relevance of the details shown.

The response results showed differences in how these models processed text-based versus visually-based questions. For text-only questions, the LLMs achieved correct response rates of 68.3 %, 79.1 %, and 71.2 % for 4o, o1, and Sonnet, respectively. This performance suggests potential applications in supporting theoretical aspects of dental technology education. In visually-based questions, the LLMs showed lower performance (44.6 %, 51.5 %, and 55.4 %, respectively), especially in questions requiring interpretation of prosthetic fabrication processes and morphological assessments. Previous studies reported that the accuracy rates for image-based examination questions in the dental and medical fields were relatively low, with particularly poor performance observed in clinically practical questions requiring complex image interpretation and decision-making.[14,15,17]

It has been reported that there is a difference in exam performance between medicine and dentistry using LLMs. Several factors could account for this difference, one of which is the amount of training data available.[18] According to the Web of Science database, general medicine includes about 20 million publications, whereas "Dentistry, Oral Surgery & Medicine" has only about 500,000.[18] This imbalance is likely to affect how well the LLMs learn. Dental image recognition also presents unique challenges. A previous study found low accuracy (25.0 % and 9.1—27.3 %) in orthodontics and prosthodontics, where multiple images must be analyzed together.[14] Our analysis revealed patterns in LLM performance across dental specialties. Orthodontic appliance questions were particularly challenging, with 4o achieving only 20.0 % accuracy overall and 14.3 % on text-only questions in this specialty, while Sonnet performed relatively well (73.3 %). Questions on pedodontic appliances were similarly challenging for both 4o and Sonnet (33.3 % each overall), with o1 performing better (60.0 %). These specialty-specific differences were widened on visually-based questions, where correct response rates for all models declined on visually-based orthodontic questions, with 4o and o1 achieving only 25.0 % accuracy compared to Sonnet's 62.5 %. The models showed relative strength in dental materials science (76.2—81.0 % overall) and laws and regulations (55.6—77.8 % overall). These results suggest that each LLM has weaknesses, depending on the subfield, when detailed image interpretation or multi-step fabrication processes are required; such patterns are consistent with previous reports indicating that visual complexity contributes significantly to error rates.[14,15,17]

The lower performance in visual recognition tasks can be attributed to several factors. Technical limitations in recognizing details of dental prostheses, interpreting three-dimensional structures from two-dimensional images, and processing variations in lighting and perspective may have affected the results. In addition, the quality and quantity of visual information may also affect accuracy. Nakao et al. found that adding more images resulted in lower accuracy.[17] This is an obstacle in clinical practice, where multiple images are needed for diagnosis and treatment planning. While these results indicate the potential of LLM as an educational tool to complement theoretical content, they also suggest that visual assessment has limitations.

While AI has the potential to support theoretical aspects of dental technician education, one of the most critical challenges facing the field is the future shortage of dental technicians and the high early-career attrition rate among newly licensed technicians.[8,19] Although the difficulty of the national examination itself does not appear to be a major barrier, given its consistently high pass rate (around 95 % over the past five years) in Japan,[9] many newly qualified dental technicians may struggle with skill development and career sustainability after entering the workforce. Studies have shown that AI-assisted assessment tools and interactive digital training platforms can support skill acquisition and long-term professional competency.[20] The multimodal LLMs may be expected to serve as a self-study tool for these young dental technicians.

Another limitation of this study is that each test question was presented only once to each LLM. Repeated testing of the same items might have provided insight into the consistency and stability of the models' performance. In addition, although our results showed significantly lower accuracy for visually-based questions, it remains unclear whether these errors are primarily due to deficiencies in image interpretation or difficulties in processing the associated textual descriptions. Further studies would need to investigate these differences more rigorously.

Future research is expected to include investigation of clinical applications, analysis of AI decision-making processes in dental technology contexts, and development of methodologies for AI integration in dental technology education. As AI technology develops, improvements in visual recognition capabilities and knowledge integration may enhance its utility in dental technology education, potentially helping address the workforce challenges identified in the introduction while maintaining educational quality. This research adds to our understanding of how current AI technologies can support dental technology education while identifying areas needing development. The results suggested that while AI tools can support theoretical learning, the visual and practical skills required in dental technology remain good developed through traditional hands-on training approaches.

Declaration of competing interest

All authors declare no conflicts of interest.

Acknowledgments

This study was partially supported by SmaSo-X Challenge Project Young Researchers Research Grant from the Graduate School of Innovation and Practice for Smart Society, Hiroshima University to Y.M.

References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436–44.
2. Yang J, Jin H, Tang R, et al. Harnessing the power of LLMs in practice: a survey on chatGPT and beyond. *ACM Trans Knowl Discov Data* 2024;18:1–32.
3. Lyo S, Mohan S, Hassankhani A, Noor A, Dako F, Cook T. From revisions to insights: converting radiology report revisions into actionable educational feedback using generative AI models. *J Imaging Inform Med* 2024:1–15.
4. Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. *Lancet Digit Health* 2023;5:e333–5.
5. Büttner M, Leser U, Schneider L, Schwendicke F. Natural language processing: chances and challenges in dentistry. *J Dent* 2024;141:104796.
6. Yamaguchi S, Morishita M, Fukuda H, et al. Evaluating the efficacy of leading large language models in the Japanese national dental hygienist examination: a comparative analysis of ChatGPT, Bard, and Bing Chat. *J Dent Sci* 2024;19:2262–7.
7. Uehara O, Morikawa T, Harada F, et al. Performance of ChatGPT-3.5 and ChatGPT-4o in the Japanese national dental examination. *J Dent Educ* 2024 (in press).
8. Oshima K. Current status of supply of and demand for dental technicians in Japan: evaluation and countermeasures against the decrease in the number of dental technicians. *Jpn Dent Sci Rev* 2021;57:123–7.
9. Hsu MS, Yeh CL, Cheng SJ, Lin CP. Integrating digital technologies in dental technician education: a comparative study of national examination in Asian countries. *J Dent Sci* 2025;20:28–35.
10. Shih YH, Cheng FC, Lin YC, Lin WC, Chiang CP. Overview of the education system for dental technicians in Taiwan. *J Dent Sci* 2024;20:971–9.
11. Yu W, Yang Z, Ren L, et al. Mm-vet v2: a challenging benchmark to evaluate large multimodal models for integrated capabilities. *arXiv* 2024;2408:00765.
12. Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. *Adv Neural Inf Process Syst* 2020;33:1877–901.
13. Chau RCW, Thu KM, Yu OY, Hsung RT, Lo ECM, Lam WYH. Performance of generative artificial intelligence in dental licensing examinations. *Int Dent J* 2024;74:616–21.
14. Morishita M, Fukuda H, Muraoka K, et al. Evaluating GPT-4V's performance in the Japanese national dental examination: a challenge explored. *J Dent Sci* 2024;19:1595–600.
15. Kim W, Kim BC, Yeom HG. Performance of large language models on the Korean dental licensing examination: a comparative study. *Int Dent J* 2025;75:176–84.
16. Song ES, Lee SP. Comparative analysis of the response accuracies of large language models in the Korean national dental hygienist examination across Korean and English questions. *Int J Dent Hyg* 2024 (in press).
17. Nakao T, Miki S, Nakamura Y, et al. Capability of GPT-4V(ision) in the Japanese national medical licensing examination: evaluation study. *JMIR Med Educ* 2024;10:e54393.
18. Liu M, Okuhara T, Huang W, et al. Large language models in dental licensing examinations: systematic review and meta-analysis. *Int Dent J* 2025;75:213–22.
19. Teng TY, Wu JH, Lee CY. Acceptance and experience of digital dental technology, burnout, job satisfaction, and turnover intention for Taiwanese dental technicians. *BMC Oral Health* 2022;22:342.
20. Lin PY, Tsai YH, Chen TC, et al. The virtual assessment in dental education: a narrative review. *J Dent Sci* 2024;19:S102–15.