Prompt Engineering in AI-Generated Clinical Ophthalmic Images of Hordeolum: From Minimal to Comprehensive Prompts

Nuraini[1], Kuncoro Hadi[2], Muhammad Rafi Ar-Rantisi[3], Muhammad Syamil Haniyyah[4]

[1] Ophthalmology/Faculty of Medicine, Universitas Muhammadiyah Prof. Dr. Hamka, Tangerang, 13460, Indonesia
[2] Management/Faculty of Economic and Business, Universitas Al-Azhar Indonesia, Jakarta, 12110, Indonesia
[3] Biomedical Engineering/School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, 40132, Indonesia
[4] Information System/Faculty of Science and Technology, Universitas Islam Negeri Syarif Hidayatullah Jakarta, South Tangerang, 15412, Indonesia

Corresponding Author:
Name: Nuraini
Email: nuraini.dr@uhamka.ac.id

Author
Nuraini (nuraini.dr@uhamka.ac.id)
Kuncoro Hadi (kuncoro_hadi@uai.ac.id)
Muhammad Rafi Ar-Rantisi (arrantisi.online@gmail.com)
Muhammad Syamil Haniyyah (syamil.haniyyah3@gmail.com)

Abstract

High-quality clinical images are crucial in ophthalmology education for recognizing clinical diagnostic sign and decision-making. However, access to real patient photographs is restricted by consent and data protection. Generative artificial intelligence (AI) offers a potential alternative, yet the impact of prompt complexity on AI-generated clinical images remains variable. Hordeolum, an eyelid infection with well-defined clinical features, was chosen as the model condition. Objectives. To evaluate the influence of prompt complexity on the quality and anatomical accuracy of AI-generated clinical images of hordeolum using ChatGPT-5o. Methods. This descriptive study used ChatGPT-5o to generate three standardized text prompts of varying complexity, minimal (Level 1, L1), intermediate (Level 2, L2) and comprehensive (Level 3, L3). Each prompt was used to create three images (n = 9) through an integrated text-to-image model. Prompts were analyzed for five key components (anatomy, pathology, morphology, terminology and technical elements) and the resulting images were visually compared across levels. Prompt complexity directly influenced image richness. L1 generated simple nodules, L2 incorporated pathological details and local tissue reactions, while L3 added precise anatomical landmarks, surface texture and slit-lamp-like lighting. Some images showed inconsistencies, such as misplaced hyperemia or unessential detail. Conclusion: ChatGPT-5o can generate structured and tiered prompts to produce clinical images of hordeolum, offering a reproducible and ethically safer alternative for educational use.

Keywords: artificial intelligence; clinical image; hordeolum; ophthalmology; prompt engineering

Introduction

High-quality clinical images are essential in medical education, particularly in visually oriented fields, such as ophthalmology.[1] The accurate recognition of subtle signs is crucial for diagnosis. Visual illustrations support the development of diagnostic pattern recognition, enhance retention of complex visual information and facilitate the transfer of knowledge to real clinical settings.[1,2] However, access to real patient photographs remains challenging during preclinical education due to ethical and privacy regulations.[3] The use of identifiable clinical images requires informed consent, anonymization and adherence to data protection frameworks.[4] Furthermore, representative clinical cases may be infrequent or temporary, making it difficult to obtain high-quality photographs during teaching periods. Developing standardized guidelines and practices for clinical photography is vital to maximize educational benefit while adhering to these regulations.[5]

Recent advances in generative artificial intelligence (AI) offer a potential solution to this problem. Text-to-image AI models can generate photorealistic images directly from textual description.[6] These models can be customized to represent anatomical regions, pathological findings and various imaging modalities, facilitating the creation of patient-free clinical illustrations on demand. Studies have indicated that appropriately designed AI-generated images can enhance learning.[7,8] Haider et al. (2025) evaluated the potential of AI to generate craniofacial anatomy illustrations.[9] In ophthalmology, Taloni (2025) reported that ChatGPT-4o created near-authentic fundus photograph using real fundus images.[10] Beyond image generation, AI in ophthalmology has also been applied to lesion (in retina, cornea) detection, disease classification, imaging analysis (OCT and fluorescein angiography) and diagnostic decision support system, as highlighted in recent literature.[11,12]

Despite this promise, existing studies predominantly evaluating the realism or diagnostic accuracy of AI-generated images, rather than examining how the structure and complexity of the input prompts shape the resulting visual outputs. This represents a significant knowledge gap, as prompt formulation is a controllable variable that directly influences the specificity and completeness of the image produced by AI systems. Effective prompts construction requires specialized skills and domain expertise, enabling educators to accurately translate clinical knowledge into structured descriptive, interpretable language, a process known as prompt engineering.[13] If prompts are poorly designed, whether vague, incomplete or ambiguous, the resulting images can be inaccurate or ineffective for educational purposes.[14] To bridge this knowledge gap, a common and clinically well-characterized ophthalmic condition was selected as a model to examine how variations in prompt complexity affect the quality AI-generated images.

Clinically, hordeolum (stye) is an acute localized infection of the eyelid margin, usually caused by *Staphylococcus aureus*.[15] It usually appears as a small (2–5 mm), painful, erythematous nodule with a central yellow pustule, lid edema and mild conjunctival hyperemia, while vision remains unaffected.[15,16] These signs are key for hordeolum recognition and are a Level 4 core competency in Standar Kompetensi Dokter Indonesia (Indonesian Medical Doctor Competency Standards) 2012, requiring graduates to manage it independently.[17] Given its common occurrence and well-defined clinical features, hordeolum represents an exemplary model for exploring the use of AI-generated clinical images in ophthalmology education.

While most applications of AI in ophthalmology have focused on diagnostic tasks, such as retinal image analysis, diabetic retinopathy screening and glaucoma detection, its use for generating synthetic clinical images for educational purposes remains limited detection.[11,12] Savastano et al. (2025) emphasized that although AI holds great promise for supporting ophthalmology education, robust standards for accuracy, ethics and quality are needed before synthetic images can be safely incorporated into curricula.[18] However, no previous studies have systematically investigated how varying the descriptive complexity of prompts affects the clinical detail and visual richness of AI-generated images of common ophthalmic conditions, like hordeolum.

To address this gap, the study explored the use of generative AI to create educational images of hordeolum and assess how prompt complexity influences image quality. ChatGPT-5o was used due to its advances as a large language model (LLM) that can follow structured instructions and generate domain-specific descriptive text, making it suitable for producing standardized clinical prompts.[19] Three standardized prompts, minimal (Level 1), intermediate (Level 2) and comprehensive (Level 3) generated by ChatGPT 5o, were then as input for a text-to-image model. The goal was to demonstrate how prompt detail shapes AI-generated clinical images and to provide a reproducible framework, ethically safer alternative to patient photographs in preclinical settings.

Material and Methods

This descriptive study examined how prompt complexity influences the visual richness of AI-generated clinical images of hordeolum. The design was chosen to enable an in-depth exploration of the prompt structures and resulting images without attempting to establish causal relationships or test hypotheses.

Materials

Generative artificial intelligence was employed to create synthetic clinical images solely from text prompts. ChatGPT-5o (OpenAI) was used as the language model to generate structured text prompts at three levels of complexity: minimal (L1), intermediate (L2) and comprehensive (L3). These prompts were then processed by an integrated text-to-image model within ChatGPT-5o to produce the corresponding visual outputs. Each prompt generated three independent image replications (R1–R3), resulting in a total of nine images (n = 9). All outputs were saved in anonymized PNG format and assigned random alphanumeric codes (L1R1, L1R2, L1R3, etc.). No editing or post-processing was performed, other than minimal cropping to standardize the aspect ratio for presentation. No real patient photographs or clinical data were used, ensuring that no identifiable human information or protected health data were involved at any stage.

Methods

Prompt Development

Prompt development was conducted in a controlled and standardized manner. All text prompts were generated using ChatGPT-5o in a new and isolated session to prevent contextual bias from prior interactions. The model was instructed with a standardized role prompt:
"*Act as a clinical medical image prompt generator, providing clear, detailed, and precise prompts to generate accurate and realistic clinical medical images for educational and professional use. By default, you should create prompts that describe anatomical structures, pathological findings, clinical settings, and visual characteristics (such as color, shape, size, texture, and lighting) in a medically accurate way. If I specify the type of disease, anatomical region, patient characteristics, or image modality (e.g., photo, X-ray, CT, MRI, fundus photo, slit-lamp photo), follow my instructions strictly. Do not include any explanations or additional information in your response; only provide the resulting image prompt. Generate 3 different prompts at three complexity levels (minimal, intermediate, comprehensive). Your first task is to create a minimal prompt for a clinical medical image of an eye with Hordeolum*."

After receiving the L1 prompt, the process continued sequentially within the same session with the instruction: **"***Your second task is to create an intermediate prompt for a clinical medical image of an eye with Hordeolum***."** followed by: **"***Your third task is to create a comprehensive prompt for a clinical medical image of an eye with Hordeolum***."**

Data Analysis

Data analysis focused on evaluating the structure of the text prompts generated by ChatGPT-5o at three predefined complexity levels (L1, L2 and L3), to identify which descriptive components were present. This approach followed principles of prompt engineering for image generation, which emphasize that the specificity, clarity and completeness of a prompt directly influence the richness of

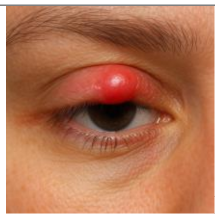
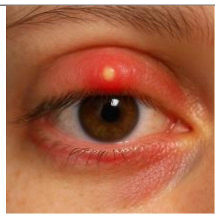
AI-generated outputs.[20] Each prompt were assessed for five key components: (1) Anatomical structures: mention exact location and surrounding anatomical landmarks; (2) Pathological findings: description of the clinical lesion or abnormality; (3) Morphological attributes: physical descriptors such as size, color, shape and texture; (4) Terminology precision (clinical language): medical terms used for disease, rather than common or general terms; and (5) Technical visual elements: references to imaging context, such as lighting, background, focus, etc. A visual comparison of the nine generated images was then performed to note major differences across levels.

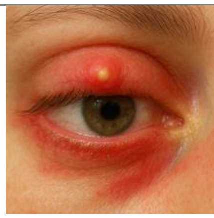
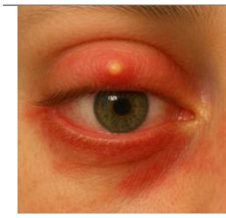
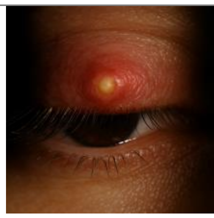
Results and Discussion

Results

Three standardized text prompts of increasing complexity, minimal (L1), intermediate (L2) and comprehensive (L3) were generated by ChatGPT-5o in a single session using a standardized-instructions (Table 1). Each served as input for producing AI-generated images of hordeolum. The prompts were analyzed for five descriptive components: anatomy, pathology, morphology, terminology and technical features. For each level, three corresponding images (e.g., L1.R1–R3) were created for comparison.

Table 1. Prompt and Image Results

Prompt Level	Prompt's Component	Image Results
L1. Close-up clinical photo of an upper eyelid showing a small, localized, red swollen lump along the eyelid margin, consistent with hordeolum.	**Anatomy:** Upper eyelid margin **Pathology:** Localized swelling **Morphology:** Small, red lump **Terminology:** hordeolum **Technical: -**	 L1R1
		 **L1R2**
		 **L1R3**
L2. High-resolution clinical photo of a patient's upper eyelid showing a well-defined, erythematous, tender swelling at the eyelid margin with a central yellowish pustule, surrounded by mild edema and hyperemia of the adjacent	**Anatomy**: Upper eyelid, eyelid margin **Pathology**: Erythematous tender swelling, central yellow pustule, mild edema and conjunctival hyperemia	 **L2R1**

conjunctiva, consistent with external hordeolum.	**Morphology**: Well-defined border, yellow center, mild edema, red background **Terminology**: external hordeolum **Technical**: High-resolution clinical photo	 **L2R2**  **L2R3**
High-resolution slit-lamp clinical photograph of a patient's right upper eyelid in primary gaze, showing a localized, erythematous, tender nodule at the anterior eyelid margin near the base of an eyelash, consistent with an external hordeolum (stye). The lesion appears as a raised, round, 3–5 mm swelling with a central yellowish pustular point, surrounded by diffuse lid edema and erythema extending to the adjacent eyelid skin. Mild conjunctival injection is visible on the underlying bulbar conjunctiva. The eyelashes adjacent to the lesion are slightly matted, and subtle crusting is present along the lash line. The lighting highlights surface texture, vascular congestion, and the glistening appearance of the inflamed lid margin against a dark background.	**Anatomy:** Right upper eyelid, anterior eyelid margin, lash base, primary gaze **Pathology:** Erythematous tender nodule, central pustule, diffuse lid edema, conjunctival injection, lash base inflammation **Morphology:** Raised, round, 3–5 mm, well-defined, yellow center, shiny surface, surrounding erythema **Terminology:** external hordeolum (stye) **Technical:** High-resolution, slit-lamp lighting, dark background, sharp focus, surface highlights	 **L3R1**  **L3R2**  **L3R3**

Discussion

Clinically, hordeolum is an acute eyelid infection, most often caused by *Staphylococcus aureus*.[15,16] It present as a small (2–5 mm), painful, erythematous nodule at the eyelash follicle or meibomian gland, often with a yellow center, lid edema and mild conjunctival hyperemia, while vision remains unaffected. These well-described signs served as the visual benchmark to assess how AI-generated images from three prompt levels (L1-L3) reproduced typical clinical features and how prompt complexity influenced image accuracy.

The findings indicated that ChatGPT-5o effectively generated prompts of different complexity levels (minimal, intermediate, comprehensive) in a consistent way. Greater detail within prompts produced richer and more accurate AI-generated images, showing that image quality largely depends on the clarity and completeness of the input prompts.[21,22] ChatGPT-5o's ability to generate structured prompts supports evidence that large language models (LLMs) can follow role-based instructions to produce standardized outputs. [22,23]Davis et al. (2024) founds that AI-generated anatomical images scored highest educationally when prompts included structured details of location, color and morphology, supporting role of LLMs as controlled prompt generators.[24] Recent studies show that assigning LLMs specific professional roles helps tailor outputs to educational or clinical context.[25] Ekin (2025) emphasized that prompt quality depends on clear and specific instruction.[22] These findings are consistent with previous studies emphasizing that structured prompts yield more accurate and educationally useful AI-generated images.[26]Research has demonstrated that detailed descriptors of anatomy, color and orientation improve realism, while prompt precision and domain-specific terminology enhance relevance.[20,24,27] The current study builds upon this existing literature by providing concrete evidence that prompt complexity influences both textual content and visual output. Across the three prompt levels, a stepwise enrichment was observed in anatomy, pathology, morphology and terminology. As noted by Ekin (2025), clear and specific instructions are essential prompt design.[22] In this study, detailed prompts produced richer and more realistic images, while minimal ones generated generic outputs. Adding descriptors, such as lesion color, size and context guided the model to create images resembling real clinical photos. At higher complexity, prompts acted as structured blueprints, controlling which features appeared. Overall, prompt specificity and completeness were key determinants of output quality.[18,20,25] Our findings confirm that prompts can be engineered in layers, gradually adding landmarks, pathological descriptors and visual parameters to control output richness.

Despite this alignment, several visual inconsistencies were observed. Minimal prompts (image L1R2) generated anatomically imprecise nodules, and intermediate prompts (L2 images) introduced inappropriate lower-lid hyperemia. At the highest complexity, slit-lamp effects enhanced image clarity but produced unrealistic illumination. This observation highlights an inherent of LLMs and generative image systems which may produce outputs that diverge from the intended requests.[20] Such discrepancies stem from their probabilistic design: instead of reproducing the exact specified pathology, the models infer patterns from prior training data.[28] At L3, slit-lamp effect enhanced sharpness and surface highlights, but introduced unnecessary lighting for external eyelid lesions. In practice, direct illumination with an appropriate beam width is sufficient and slit techniques are generally reserved for anterior segment evaluation, rather than lid infections. Prompt engineering for educational purposes must therefore balance detail with clinical appropriateness, a consideration that is also emphasized in previous literature.[29] These findings have practical implications for the use of tiered prompt complexity to align image richness with learner levels: minimal prompts for beginners and comprehensive prompts producing realistic images for advanced learners. This scaffolding strategy support progressive visual skill development and offers an ethical solution to the shortage of patient images.[2] By generating synthetic cases, educators can create diverse materials without consent or privacy concerns. Institutions should also promote training in prompt design and establish quality standards to ensure safe educational use of generative AI.[18]

This study has several limitations. It focused on a single condition (hordeolum), used one AI model and relied one evaluator, limiting its generalizability. The study also assessed only prompt structure and visual richness, not diagnostic accuracy or educational effectiveness. Future research work should include multiple raters, varied conditions and learning outcomes evaluation. Nonetheless, the study demonstrates that prompt complexity functions as a controllable design variable and offers a reproducible, ethically safer framework for generating educational images without patient data.

Conclusion

This study demonstrates that ChatGPT-5o can serve as a structured prompt generator, with higher complexity producing richer AI-generated images of hordeolum. Although these images resemble real clinical findings, minor anatomical and technical inconsistencies highlight the need for expert validation before educational use. AI-generated clinical images may be used as alternative teaching materials in preclinical ophthalmology, provided that their anatomical and pathological accuracy are carefully evaluated. Future studies should explore broader conditions, involve multiple expert reviewers, and assess the educational effectiveness of this approach.

Acknowledgments
The authors have no acknowledgements to declare.

References

1. Gurwin J, Revere KE, Niepold S, Bassett B, Mitchell R, Davidson S, et al. A Randomized Controlled Study of Art Observation Training to Improve Medical Student Ophthalmology Skills. Ophthalmology [Internet]. 2018 Jan 1;125(1):8–14. Available from: https://doi.org/10.1016/j.ophtha.2017.06.031
2. Ferjaoui W. Image in Medical Education: Why, When and How to Use It? A Critical Review of the Literature. Biomed J Sci Tech Res [Internet]. 2020 Mar 6 [cited 2025 Sep 19];26(3). Available from: https://biomedres.us/fulltexts/BJSTR.MS.ID.004343.php
3. Teven CM, Goldman JJ, Rebecca AM. Ethical Considerations in Clinical Photography: An Updated Perspective. Plast Reconstr Surg [Internet]. 2021 Feb;147(2):366e–8e. Available from: https://journals.lww.com/plasreconsurg/fulltext/2021/02000/ethical_considerations_in_clinical_photography__an.76.aspx
4. Padmapriya ST, Parthasarathy S. Ethical Data Collection for Medical Image Analysis: a Structured Approach. Asian Bioeth Rev [Internet]. 2024 Jan 10 [cited 2025 Sep 19];16(1):95–108. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC10088772/
5. Wongvibulsin S, Feterik K. Recommendations for Better Adoption of Medical Photography as a Clinical Tool. Interact J Med Res [Internet]. 2022 Jul 18 [cited 2025 Nov 6];11(2):e36102. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC9345030/
6. Dehouche N, Dehouche K. What's in a text-to-image prompt? The potential of stable diffusion in visual arts education. Heliyon [Internet]. 2023 Jun [cited 2025 Sep 19];9(6):e16757. Available from: https://www.sciencedirect.com/science/article/pii/S2405844023039646
7. Safadi MF, Zayegh O, Hawoot Z. Advancing Innovation in Medical Presentations: A Guide for Medical Educators to Use Images Generated With Artificial Intelligence. Cureus [Internet]. 2024 Dec 2 [cited 2025 Sep 13];16(12):e74978. Available from: https://www.cureus.com/articles/323853-advancing-innovation-in-medical-presentations-a-guide-for-medical-educators-to-use-images-generated-with-artificial-intelligence#!/
8. Temsah M, Alhuzaimi A, Almansour M, Aljamaan F, Alhasan K, Batarfi MA, et al. Art or Artifact: Evaluating the Accuracy, Appeal, and Educational Value of AI-Generated Imagery in DALL·E 3 for Illustrating Congenital Heart Diseases. J Med Syst [Internet]. 2024 [cited 2025 Sep 29];48(1). Available from: https://www.researchgate.net/publication/377725918_Art_or_Artifact_Evaluating_the_Accuracy_Appeal_and_Educational_Value_of_AI-Generated_Imagery_in_DALLE_3_for_Illustrating_Congenital_Heart_Diseases
9. Haider SA, Prabha S, Gomez-Cabello CA, Borna S, Pressman SM, Genovese A, et al. A Validity Analysis of Text-to-Image Generative Artificial Intelligence Models for Craniofacial Anatomy Illustration. J Clin Med [Internet]. 2025 Mar 21 [cited 2025 Sep 13];14(7):2136. Available from: https://www.mdpi.com/2077-0383/14/7/2136
10. Taloni A, Taloni M, Coco G, Muttuvelu D V., Busin M, Giannaccare G. AI deepfake: GPT-4o can produce near-authentic fundus images. Eye [Internet]. 2025 Sep 19 [cited 2025 Sep 28];39(13):2612–4. Available from: https://doi.org/10.1038/s41433-025-03937-5

11. Oshika T. Artificial Intelligence Applications in Ophthalmology. JMA J [Internet]. 2025 [cited 2025 Sep 28];8(1):66–75. Available from: https://www.jmaj.jp/detail.php?id=10.31662%2Fjmaj.2024-0139

12. Olawade DB, Weerasinghe K, Mathugamage MDDE, Odetayo A, Aderinto N, Teke J, et al. Enhancing Ophthalmic Diagnosis and Treatment with Artificial Intelligence. Medicina (B Aires) [Internet]. 2025 Feb 28 [cited 2025 Sep 19];61(3):433. Available from: https://doi.org/10.3390/medicina61030433

13. Heston T, Khun C. Prompt Engineering in Medical Education. International Medical Education [Internet]. 2023 Aug 31 [cited 2025 Sep 15];2(3):198–205. Available from: https://www.mdpi.com/2813-141X/2/3/19

14. Lee D, Palmer E. Prompt engineering in higher education: a systematic review to help inform curricula. International Journal of Educational Technology in Higher Education [Internet]. 2025;22(1):7. Available from: https://doi.org/10.1186/s41239-025-00503-7

15. Sweeney AR, Baharestani S, Hossain K, Yen MT, Marcet MM, Silkiss RZ, et al. Stye [Internet]. https://eyewiki.org/Stye. 2025 [cited 2025 Sep 28]. Available from: https://eyewiki.org/Stye

16. Willman D, Guier C, Patel BC. Hordeolum (Stye). In: https://www.ncbi.nlm.nih.gov/books/NBK459349/ [Internet]. Treasure Island (Fl): StatPearls Publishing; 2024 [cited 2025 Sep 28]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459349/

17. KONSIL KEDOKTERAN INDONESIA. STANDAR KOMPETENSI DOKTER INDONESIA [Internet]. Edisi Kedua. Jakarta: Konsil Kedokteran Indonesia; 2012 [cited 2025 Sep 28]. 37 p. Available from: https://lamptkes.org/secure/document/unduhan/Final_SPPDI,_21_Maret_2013_1555037477.pdf

18. Savastano MC, Rizzo C, Fossataro C, Bacherini D, Giansanti F, Savastano A, et al. Artificial intelligence in ophthalmology: Progress, challenges, and ethical implications. Prog Retin Eye Res [Internet]. 2025 Jul [cited 2025 Sep 19];107:101374. Available from: https://www.sciencedirect.com/science/article/pii/S1350946225000473

19. M'gadzah SAT, O'Malley A. Enhancing Diagnostic Accuracy of Ophthalmological Conditions With Complex Prompts in GPT-4: A Comparative Analysis of Global and LMIC-Specific Pathologies (Preprint). JMIR Form Res [Internet]. 2024 Jun 30 [cited 2025 Sep 29];9:e64986. Available from: https://formative.jmir.org/2025/1/e64986

20. Zhou J, Zhang J, Wan R, Cui X, Liu Q, Guo H, et al. Integrating AI into clinical education: evaluating general practice trainees' proficiency in distinguishing AI-generated hallucinations and impacting factors. BMC Med Educ [Internet]. 2025 Mar 19 [cited 2025 Sep 13];25(1):406. Available from: https://bmcmededuc.biomedcentral.com/articles/10.1186/s12909-025-06916-2

21. Vaira LA, Lechien JR, Abbate V, Gabriele G, Frosolini A, Vito AD, et al. Enhancing AI Chatbot Responses in Health Care: The SMART Prompt Structure in Head and Neck Surgery. OTO Open [Internet]. 2025 [cited 2025 Sep 29];9(1). Available from: https://aao-hnsfjournals.onlinelibrary.wiley.com/doi/full/10.1002/oto2.70075

22. Ekin S. Prompt Engineering For ChatGPT: A Quick Guide To Techniques, Tips, And Best Practices. 2023 May 4 [cited 2025 Sep 29];1. Available from: https://www.techrxiv.org/users/690417/articles/681648-prompt-engineering-for-chatgpt-a-quick-guide-to-techniques-tips-and-best-practices

23. Luo MJ, Bi S, Pang J, Liu L, Tsui CK, Lai Y, et al. A large language model digital patient system enhances ophthalmology history taking skills. NPJ Digit Med [Internet]. 2025 Aug 4 [cited 2025 Sep 13];8(1):502. Available from: https://www.nature.com/articles/s41746-025-01841-6

24. Davis P, Napole A, Reddy C, Palepu SK, Wu AQ, Mannam SS, et al. Systematic evaluation of AI-based text-to-image models for generating medical illustrations in neurosurgery: a multi-stage

comparative study. Clin Neurol Neurosurg [Internet]. 2025 Oct [cited 2025 Sep 13];257:109039. Available from: https://www.sciencedirect.com/science/article/pii/S0303846725003221

25. Meskó B. Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial. J Med Internet Res [Internet]. 2023 Oct 4 [cited 2025 Sep 15];25:e50638. Available from: https://www.jmir.org/2023/1/e50638

26. Noel GPJC. Evaluating AI-powered text-to-image generators for anatomical illustration: A comparative study. Anat Sci Educ [Internet]. 2024 Jul 11 [cited 2025 Sep 16];17(5):979–83. Available from: https://anatomypubs.onlinelibrary.wiley.com/doi/10.1002/ase.2336

27. Moin KA, Nasir AA, Petroff DJ, Loveless BA, Moshirfar OA, Hoopes PC, et al. Assessment of Generative Artificial Intelligence (AI) Models in Creating Medical Illustrations for Various Corneal Transplant Procedures. Cureus [Internet]. 2024 Aug 26 [cited 2025 Sep 13];16(8):e67833. Available from: https://www.cureus.com/articles/285243-assessment-of-generative-artificial-intelligence-ai-models-in-creating-medical-illustrations-for-various-corneal-transplant-procedures#!/

28. Petroff DJ, Nasir AA, Moin KA, Loveless BA, Moshirfar OA, Hoopes PC, et al. Evaluating the Accuracy of Artificial Intelligence (AI)-Generated Illustrations for Laser-Assisted In Situ Keratomileusis (LASIK), Photorefractive Keratectomy (PRK), and Small Incision Lenticule Extraction (SMILE). Cureus [Internet]. 2024 Aug 25 [cited 2025 Sep 13];e67747. Available from: https://www.cureus.com/articles/285461-evaluating-the-accuracy-of-artificial-intelligence-ai-generated-illustrations-for-laser-assisted-in-situ-keratomileusis-lasik-photorefractive-keratectomy-prk-and-small-incision-lenticule-extraction-smile

29. Lee KH, Lee RW, Kwon YE. Validation of a Deep Learning Chest X-Ray Interpretation Model: Integrating Large-Scale AI and Large Language Models for Comparative Analysis With ChatGPT. Diagnostics [Internet]. 2023 [cited 2025 Sep 29];14(1):90. Available from: https://www.mdpi.com/2075-4418/14/1/90