OPEN

ChatGPT and Lacrimal Drainage Disorders: Performance and Scope of Improvement

Mohammad Javed Ali, M.D., PH.D., F.R.C.S.

Govindram Seksaria Institute of Dacryology, L.V. Prasad Eye Institute, Hyderabad, India

Purpose: This study aimed to report the performance of the large language model ChatGPT (OpenAI, San Francisco, CA, U.S.A.) in the context of lacrimal drainage disorders.

Methods: A set of prompts was constructed through questions and statements spanning common and uncommon aspects of lacrimal drainage disorders. Care was taken to avoid constructing prompts that had significant or new knowledge beyond the year 2020. Each of the prompts was presented thrice to ChatGPT. The questions covered common disorders such as primary acquired nasolacrimal duct obstruction and congenital nasolacrimal duct obstruction and their cause and management. The prompts also tested ChatGPT on certain specifics, such as the history of dacryocystorhinostomy (DCR) surgery, lacrimal pump anatomy, and human canalicular surfactants. ChatGPT was also quizzed on controversial topics such as silicone intubation and the use of mitomycin C in DCR surgery. The responses of ChatGPT were carefully analyzed for evidence-based content, specificity of the response, presence of generic text, disclaimers, factual inaccuracies, and its abilities to admit mistakes and challenge incorrect premises. Three lacrimal surgeons graded the responses into three categories: correct, partially correct, and factually incorrect.

Results: A total of 21 prompts were presented to the ChatGPT. The responses were detailed and were based according to the prompt structure. In response to most questions, ChatGPT provided a generic disclaimer that it could not give medical advice or professional opinion but then provided an answer to the question in detail. Specific prompts such as "how can I perform an external DCR?" were responded by a sequential listing of all the surgical steps. However, several factual inaccuracies were noted across many ChatGPT replies. Several responses on controversial topics such as silicone intubation and mitomycin C were generic and not precisely evidence-based. ChatGPT's response to specific questions such as canalicular surfactants and idiopathic canalicular inflammatory disease was poor. The presentation of variable prompts on a single topic led to responses with either repetition or recycling of the phrases. Citations were uniformly missing across all responses. Agreement among the three observers was high (95%) in grading the responses. The responses of ChatGPT were graded as correct for only 40% of the prompts, partially correct in 35%, and outright factually incorrect in 25%. Hence, some degree of factual inaccuracy was present in 60% of the responses, if we consider the partially correct responses. The exciting aspect was that ChatGPT was able to admit mistakes and correct them when presented with counterarguments. It was also capable of challenging incorrect prompts and premises.

Conclusion: The performance of ChatGPT in the context of lacrimal drainage disorders, at best, can be termed average. However, the potential of this AI chatbot to influence medicine is enormous. There is a need for it to be specifically trained and retrained for individual medical subspecialties.

(*Ophthalmic Plast Reconstr Surg* 2023;39:221–225)

The development of effective large language models (LLM) has recently led to a revolution in the development of artificial intelligence (AI)-based chatbots. The most talked about chatbot today is ChatGPT, developed by a San Francisco-based company, "OpenAI," and released 5 months ago in November 2022. The dialogue form of interaction closely mimics human interaction and creates intelligent-sounding texts in response to user prompts. The prompts could be questions or statements. The development has sparked debate ranging from the limits of AI to the end of traditional education systems. There are concerns that students would not only outsource their writing to ChatGPT but also their thinking.

The scientific community was jolted when the abilities of ChatGPT to construct manuscripts came to light.[1] Several articles also mentioned ChatGPT as a coauthor.[2,3] While efforts have been made to correct it, it also led to a larger question— Can ChatGPT be given authorship? The overwhelming response from the academic stakeholders, including publishers, editors, and scientific societies, was a straight "NO," and rightfully so.[4–7] However, brushing off ChatGPT as something bad for science would be unwise. AI-based disruptive technologies are here to stay, and the scientific community should come together to construct clear guidelines while recognizing their legitimate uses. The potential of ChatGPT to influence medicine is enormous and cannot be ignored.

Since it has been only 5 months following ChatGPT's release, there is a lack of data regarding its use in ophthalmology[8] and more so ophthalmic plastic surgery. To the best of the authors' knowledge, this is the first of its kind study in ophthalmic plastic surgery. The present study attempted to quiz ChatGPT (ver 3.5) about several aspects of lacrimal drainage disorders and analyzed the responses on several parameters. Such exercises can identify the loopholes and weaknesses of ChatGPT, create awareness, and potentially lead to better training and subsequently better outcomes.

METHODS

The study adhered to the Tenets of the Declaration of Helsinki. Twenty-one prompts were constructed as questions or statements spanning common and uncommon aspects of lacrimal drainage disorders. Since ChatGPT would perform poorly for information beyond 2021, care was taken to avoid constructing prompts with significant or new knowledge beyond the year 2020. The questions covered common disorders such as primary acquired nasolacrimal duct obstruction and congenital nasolacrimal duct obstruction (CNLDO) and their cause and management. The prompts also tested ChatGPT on certain specifics, such as the history of dacryocystorhinostomy (DCR) surgery, lacrimal pump anatomy, and human canalicular surfactants. ChatGPT was also quizzed on controversial topics such as silicone intubation and the use of mitomycin C in DCR surgery. The responses of ChatGPT were carefully analyzed for evidence-based content, specificity of the response, presence of generic text, disclaimers, and factual inaccuracies. Each prompt was presented three times to assess the variations of ChatGPT in the responses. Three lacrimal surgeons graded the responses into three categories: correct, partially correct, and factually incorrect.

RESULTS

A total of 21 prompts were presented to ChatGPT. Responses were detailed and based according to the prompt structure. In response to most questions, ChatGPT provided a generic disclaimer that it cannot give medical advice or professional opinion but then answered the medical question in detail. Citations were uniformly missing in all of the responses. All the questions demonstrated variations in response to the repetition of similar prompts. However, except in two instances (prompt 1 and prompt 3), the content of the response was essentially similar. Agreement among the three observers was high (95%) in grading the responses. The ChatGPT responses were graded as correct for only 40% of the prompts, partially correct in 35%, and outright factually incorrect in 25%. Hence, some degree of factual inaccuracy was present in 60% of the responses, if we consider the partially correct responses. Because it is not helpful to replicate all the ChatGPT responses, salient aspects of each response will be discussed.

1. **Prompt:** Can you list common lacrimal drainage disorders in humans?
Analysis of the response: ChatGPT listed 10 common lacrimal drainage disorders. Although the top four were correct, it also listed entropion, ectropion, blepharitis, and conjunctivitis, which are factually incorrect.
2. **Prompt:** Describe the anatomy of the lacrimal pump.
Analysis of the response: ChatGPT's response was incorrect and had unrelated anatomy about the lacrimal glands, the nasolacrimal duct, and the nasal cavity. When prompted again, "how are tears propelled from the ocular surface to the lacrimal sac?," the response was even more generic with recycled phrases from the first answer.
3. **Prompt:** What is the history of dacryocystorhinostomy surgery?
Analysis of the response: ChatGPT's response was poor with several factual inaccuracies. When the question was repeated, it wrongly attributed different individuals to the same surgical approach. However, when counter-argued with a prompt such as "I do not think this is correct. Are you sure it is?." It responded with apologies and willingness to learn the facts. **Table 1** details the selected responses verbatim for the readers to decide.
4. **Prompt:** What is the cause of congenital nasolacrimal duct obstruction?
Analysis of the response: The ChatGPT reply was brief and to the point. The reply began by saying that this is not fully understood but is believed to be due to developmental issues at the distal end of the nasolacrimal duct during fetal life.
5. **Prompt:** What is the management of congenital nasolacrimal duct obstruction?
Analysis of the response: This was yet again a good response from ChatGPT, which classified the management into conservative and surgical and listed the modalities under each. **Table 2** details the response verbatim.
6. **Prompt:** Can you describe the management of acute dacryocystitis?
Analysis of the response: While the ChatGPT reply had several standard measures taken by clinicians, it also listed "a gentle massage around the inflamed gland in downward motion toward the nose." There was confusion here with CNLDO management besides factual inaccuracies of calling the lacrimal sac a gland and describing the incorrect technique of Crigler's massage.
7. **Prompt:** What is the cause of primary acquired nasolacrimal duct obstruction?
Analysis of the response: The ChatGPT response was comprehensive as far as the basic pathogenesis was concerned. However, it missed several of the proposed etiological factors.
8. **Prompt:** Which is better, external or endoscopic dacryocystorhinostomy?
Analysis of the response: ChatGPT's response was diplomatic—"Ultimately, the choice between external and endoscopic DCR should be based on the individual patient's needs and the surgeon's expertise." However, the reasons given for choosing endoscopic DCR were not entirely correct when it mentioned—"In general, endoscopic DCR is preferred in cases where the blockage is not severe, the patient has a narrow nasal passage, or the surgeon wants to minimize scarring."
9. **Prompt:** How can I perform an external DCR?
Analysis of the response: The ChatGPT response was detailed and acceptable. It enumerated all the surgical steps sequentially and briefly touched upon postoperative management. **Table 3** details the response verbatim.
10. **Prompt:** Should I use silicone intubation in dacryocystorhinostomy?
Analysis of the response: The response to this controversial question was confusing and not very clear. It concluded by mentioning, "Ultimately, the decision to use silicone intubation should be made on a case-by-case basis after careful consideration of the patient's individual needs and the surgeon's experience and preferences." There was no presentation of any evidence or sentences that could be attributed to any evidence. The lack of medical literature analysis was obvious.
11. **Prompt:** Should I use Mitomycin C in dacryocystorhinostomy?
Analysis of the response: The ChatGPT response was suboptimal and similarly generic as it was for silicone intubation.
12. **Prompt:** What do you know about surfactants in the human canaliculus?
Analysis of the response: The ChatGPT response to this was poor. A sentence linked surfactants to dry eye disease (DED), and remainder of the response was about risk factors and treatment of DED. The question was not answered.
13. **Prompt:** Is punctoplasty effective in the treatment of punctal stenosis?
Analysis of the response: While the response was generally satisfactory, the text suffered from unnecessary sentences in between factual statements. For example, an abrupt

---

TABLE 1. ChatGPT and lacrimal drainage: selected verbatim responses

---

Prompt: What is the history of dacryocystorhinostomy surgery?

ChatGPT's response: The history of DCR surgery dates back to the early 1900s when Howard Kelly and his colleagues developed the first surgical procedure for the treatment of nasolacrimal duct obstruction.

In the late 19th century, Sir William Bowman first described the lacrimal sac.

The first DCR surgery was performed in the early 1900s by Adolf Alt, who introduced an external approach to the procedure. However, this approach had a high rate of complications, such as facial scars, bleeding, and infections.

In the 1960s, Lester T. Jones, an Australian ophthalmologist, developed the endonasal approach to DCR surgery. This method involved creating a new passage between the lacrimal sac and the nasal cavity using a nasal endoscope.

Prompt: I do not think this is a correct answer! Are you sure about this?

ChatGPT's response: As an AI language model, I strive to provide accurate and reliable information based on the data and information available to me. I apologize if my response is incorrect or unclear. Please feel free to provide me with more context or information, and I will do my best to provide an accurate response.

AI, artificial intelligence; DCR, dacryocystorhinostomy.

TABLE 2. ChatGPT and lacrimal drainage disorders

Prompt: What is the management of congenital nasolacrimal duct obstruction?

ChatGPT's response: Congenital nasolacrimal duct obstruction (CNLDO) can be managed both conservatively and surgically.

Conservative management includes:

1. Observation: In many cases, the obstruction resolves on its own during the first year of life. Therefore, in asymptomatic infants, observation alone is recommended.
2. Topical antibiotics: If there is associated conjunctivitis or discharge, topical antibiotics can be prescribed to prevent secondary infections.
3. Massage and warm compresses: The parents can be taught to perform gentle massage over the lacrimal sac area to promote drainage of tears. Warm compress can also be used to reduce inflammation.

Surgical management includes:

1. Probing and irrigation: This is the most commonly performed surgical procedure for CNLDO. Under general anesthesia, a thin probe is inserted through the puncta into the nasolacrimal duct to break down any membrane or blockage. Irrigation is then performed to flush out any debris or mucous plugs.
2. Stent placement: In some cases, a silicone stent can be placed into the nasolacrimal duct to keep it open after probing and irrigation. The stent is usually removed after a few months.
3. Dacryocystorhinostomy (DCR): In recurrent or severe cases, DCR may be recommended. This is a more invasive surgery.

TABLE 3. ChatGPT panel response

Prompt: How can I perform an external dacryocystorhinostomy?

ChatGPT's response: External dacryocystorhinostomy (DCR) is a surgical procedure used to treat a blocked tear duct by creating a new drainage channel between the tear sac and the nasal cavity. Here are the general steps for performing an external DCR

1. Anesthesia: The patient is given general anesthesia, or local anesthesia with sedation.
2. Incision: An incision is made on the side of the nose where the tear duct is blocked. The incision is typically made in the skin crease between the nose and the lower eyelid.
3. Access to the nasal bone: The surgeon will remove the skin and soft tissue to expose the nasal bone.
4. Bone removal: The surgeon will then use a small drill or burr to remove a small portion of the nasal bone, creating a small opening to access the nasal cavity.
5. Nasal mucosa flap: A flap of nasal mucosa (the lining of the nose) is then created and carefully lifted away from the underlying bone.
6. Access to tear sac: The surgeon will locate the tear sac, which is a small, round sac that collects tears from the eye.
7. Creation of a new tear drainage channel: A small opening is created in the tear sac and a new channel is created between the tear sac and the nasal cavity.
8. Placement of stent: A silicone stent or tube is placed in the new channel to help it stay open and promote proper drainage of tears. The stent is typically left in place for several weeks.
9. Closure: The incision is then closed with sutures, and a nasal packing may be placed to control bleeding.

Postoperative care typically involves regular nasal rinses and antibiotic ointment or drops to prevent infection. Most patients can return to normal activities within a week or two after the surgery. However, full recovery may take several weeks to several months, depending on the individual case.

---

description of punctal anatomy was placed in midst of a punctoplasty discussion.

14. **Prompt:** What are the uses of CT—dacryocystography?

Analysis of the response: The ChatGPT response was detailed, sequential, lucid, and comprehensive. **Table 4** lists the response verbatim.

15. **Prompt:** Can you describe "idiopathic canalicular inflammatory disease" or ICID?

Analysis of the response: This was a generic response, and it was clear that ChatGPT could not provide an adequate answer to this question.

16. **Prompt:** How can I evaluate a dacryocystorhinostomy ostium?

Analysis of the response: This was a good response from ChatGPT. It listed five techniques to evaluate a DCR ostium along with a brief description of each. They included endoscopic examination, probing, irrigation, CT scan, and dacryocystography.

17. **Prompt:** What do you know about nasolacrimal duct recanalization?

Analysis of the response: The ChatGPT response here was confusing and essentially did not answer the question.

TABLE 4. ChatGPT and CT-dacryocystography

Prompt: What are the uses of CT-dacryocystography?
ChatGPT's response: CT-dacryocystography is a medical imaging test that uses CT technology to examine the tear drainage system of the eye. It is most commonly used to diagnose blockages or abnormalities in the lacrimal (tear) ducts, which can cause excessive tearing, recurrent infections, or other eye-related problems.
Some common uses of CT-dacryocystography include:
  1. Identifying blocked tear ducts: This test can help identify the exact location and extent of blockages in the tear ducts, which can be caused by a range of conditions such as scar tissue, tumors, or congenital anomalies.
  2. Evaluating the effectiveness of tear duct surgery: In some cases, surgery may be necessary to repair or reposition the tear ducts. CT-dacryocystography can be used before and after surgery to assess the success of the procedure.
  3. Preoperative planning: Before performing tear duct surgery, doctors may use CT-dacryocystography to gather detailed information about the patient's anatomy and plan the surgery more effectively.
  4. Diagnosing other eye-related conditions: Although CT-dacryocystography is primarily used to diagnose tear duct problems, it can also provide images of other structures in the eye, including the orbits (eye sockets), sinuses, and surrounding soft tissues. In some cases, these images may help diagnose other eye-related conditions.

---

18. **Prompt:** What are the indications for dacryocystectomy?
Analysis of the response: While tumors and trauma were mentioned, the ChatGPT response was off the mark here, listing chronic dacryocystitis, nasolacrimal duct obstruction, and CNLDO as indications for dacryocystectomy. Interestingly, while it described the procedure correctly initially, it later falsely described it as a bypass procedure. **Table 5** lists the response verbatim for the readers to decide.

19. **Prompt:** Can you list the complications of injecting Botulinum toxin into the lacrimal gland?
Analysis of the response: This was a good response where ChatGPT put forth common and rare complications.

20. **Prompt:** Nasal endoscopy is helpful in the routine management of congenital nasolacrimal duct obstructions. Do you agree with this?
Analysis of the response: The ChatGPT response to this controversial question was interesting, where it presented arguments on both sides of the issue. **Table 6** places the response verbatim.

21. **Prompt:** Acute dacryocystitis can only be managed with an immediate endoscopic DCR. Do you agree with this?
Analysis of the response: The prompt assessed whether ChatGPT could challenge an incorrect premise. The response did not agree with this statement, and rightly so. **Table 6** places the response verbatim.

DISCUSSION

The current study reports the performance of ChatGPT in the context of lacrimal drainage disorders. While the performance of ChatGPT in this study was generally unsatisfactory, the potential of this AI chatbot to influence medicine is enormous. However, there is a need for it to be specifically trained and retrained for individual medical subspecialties. As far as lacrimal drainage disorders are concerned, ChatGPT is not at a stage where it can be very useful as a patient-education resource, or resident and fellows training resource.

TABLE 5. ChatGPT and serious factual inaccuracies

Prompt: What are the indications for a dacryocystectomy?
ChatGPT's response: Dacryocystectomy is a surgical procedure used to remove the lacrimal sac, which is a small sac that collects the tears before they drain into the nasal cavity. The indications for dacryocystectomy may include:
  1. Chronic dacryocystitis that does not respond to medical treatment: Chronic infection of the lacrimal sac can lead to recurrent episodes of discharge, pain, and swelling. In such cases, dacryocystectomy may be required to remove the infected sac and prevent further infection.
  2. Nasolacrimal duct obstruction: If the nasolacrimal duct is severely obstructed, tears may not be able to drain properly, leading to epiphora (excessive tearing). In such cases, dacryocystectomy may be performed to bypass the obstruction.
  3. Tumors of the lacrimal sac: A benign or malignant tumor of the lacrimal sac may require a dacryocystectomy for complete removal.
  4. Trauma: Trauma to the lacrimal sac or the nasolacrimal duct can cause scarring and obstruction, which may require dacryocystectomy for repair.
  5. Congenital abnormalities: Some babies are born with a blocked tear duct, which may require a dacryocystectomy to fix it.

TABLE 6. Assessing ChatGPT's abilities to challenge premises

Prompt: Nasal endoscopy is helpful in routine management of congenital nasolacrimal duct obstructions. Do you agree with this?
ChatGPT's response: As an AI language model, I cannot provide a personal opinion. However, nasal endoscopy is indeed helpful in routine management of congenital nasolacrimal duct obstructions. It allows the doctor to visualize the inside of the nasal cavity and to identify any obstructions that may be causing or contributing to the obstruction. Additionally, it can be used to guide the placement of stents or dilators to help open the duct, monitor the progress of the treatment, and assess any complications that may arise. Therefore, nasal endoscopy is an important tool in the management of this condition.
Prompt: Acute dacryocystitis can only be managed with an immediate endoscopic DCR. Do you agree with this?
ChatGPT's response: As an AI language model, I do not have personal opinions or beliefs. However, based on medical literature and guidelines, acute dacryocystitis can initially be managed with antibiotics, warm compresses, and analgesics. Still, if the condition is severe or recurs frequently, an endoscopic DCR may be necessary to achieve a long-term cure. The decision ultimately depends on the individual patient and their specific needs, as determined by a qualified healthcare professional.

AI, artificial intelligence; DCR, dacryocystorhinostomy.

There are several limitations of ChatGPT,[1,9] and they include the following:

1. The output is based only on the libraries on which they were trained.
2. It can provide plausible-sounding but factually incorrect answers.
3. Training it to be more cautious will make ChatGPT decline to answer queries.
4. It may require multiple tweaking of the prompts to give a correct answer.
5. Since it was trained upon data before 2021, any new information from 2021 and beyond will not be part of the ChatGPT responses.
6. The responses are verbose and commonly generic.
7. It has the potential to respond to harmful instructions.
8. It may exhibit discriminatory behavior in responses.
9. It may potentially facilitate the production of fraudulent papers such as paper mills.[10]
10. It cannot take responsibility for its content generation.

Interesting patterns of ChatGPT responses were observed in the present study. Specific prompts such as "how can I perform an external DCR?" were responded to by a sequential listing of all the surgical steps. Several responses on controversial topics such as silicone intubation and mitomycin C were generic and not precisely evidence-based. ChatGPT response to specific questions such as canalicular surfactants and idiopathic canalicular inflammatory disease was poor. The presentation of variable prompts on a single topic led to responses with either repetition or recycling of the phrases. Several factual inaccuracies were noted across many ChatGPT replies. When alerted to these inaccuracies, ChatGPT was able to correct itself and learn from the human user. It also was able to challenge an incorrect premise, as shown in the response to prompt 21.

While there are several limitations, the rapidly learning LLM such as ChatGPT may well soon overcome most of these, if not all. The potential for such AI chatbots to revolutionize medicine is high. They can be quite helpful with academic training modules and innovations, developing manuscript writing skills, data and statistical analysis, and involvement in aspects of patient care such as radiology reports and discharge summaries.[1,11–13] However, healthcare, as it appears, is not ready for it at present, and hence, efforts must be put in by all the stakeholders so that the full potential of ChatGPT is harnessed for the benefit of science and overall humanity.

ACKNOWLEDGMENT

REFERENCES

1. Ali MJ, Djalilian A. Readership awareness series—paper 4: chatbots and ChatGPT – ethical considerations in scientific publications. *Semin Ophthalmol* 2023;28:153–154.
2. O'Connor S; ChatGPT. Open artificial intelligence platforms in nursing education: tools for academic progress or abuse? *Nurse Educ Pract* 2023;66:103537.
3. Zhavoronkov A; ChatGPT Generative Pre-trained Transformer. Rapamycin in the context of Pascal's Wager: generative pre-trained transformer perspective. *Oncoscience* 2022;9:82–84.
4. COPE Position Statement. In: Authorship and AI tools. Available at: https://publicationethics.org/cope-position-statements/ai-author. Accessed March 17, 2023.
5. Thorp HH. ChatGPT is fun, but not an author. *Science* 2023;379:313.
6. Flanagin A, Bibbins-Domingo K, Berkwits M, et al. Nonhuman "Authors" and implications for the integrity of scientific publication and medical knowledge. *JAMA* 2023;329:637–639.
7. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature* 2023;613:612.
8. Potapenko I, Boberg-Ans LC, Hansen S, et al. Artificial intelligence-based chatbot patient information on common retinal diseases using ChatGPT. *Acta Ophthalmol* 2023. doi:10.1111/aos.15661.
9. ChatGPT Limitations. Open AI. Available at: https://openai.com/blog/chatgpt. Accessed March 17, 2023.
10. Ali MJ, Djalilian A. Readership awareness series—paper 3: paper mills. *Ocul Surf* 2023;28:56–57.
11. Shen Y, Heacock L, Elias J, et al. ChatGPT and other large language models are double-edged swords. *Radiology* 2023;307:e230163.
12. van Dis EAM, Bollen J, Zuidema W, et al. ChatGPT: five priorities for research. *Nature* 2023;614:224–226.
13. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health* 2023;5:e107–e108.