



EDITORIAL

Ophthalmology teaching: the journey continues, now with artificial intelligence

Paulo Schor¹

1. Department of Ophthalmology and Visual Sciences, Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, SP, Brazil.

More than a decade ago, an experience in teaching ophthalmology to medical students was described in *Arquivos Brasileiros de Oftalmologia*⁽¹⁾, based on a “simple” idea: for non-ophthalmologist physicians, learning to organize signs, symptoms, and basic knowledge of anatomy and pathophysiology to construct coherent clinical reasoning is more important than memorizing diseases.

A photograph of an eye, a face, or even a representation of a symptom served as the starting point. From it, the students collectively constructed a patient: identification, chief complaint, history, examination, hypotheses, complementary tests, and management. The actual diagnosis behind the photograph mattered little. As stated at the time, “the journey interests us more than the arrival”.

In 2026, this practice was expanded.

As in the original proposal, at the end of the clerkship, an examination based on the same principle was administered. This time, the assessment accounted for 40% of the final course grade. Each student received a clinical image and was asked to construct a complete case based on it, including the chief complaint and its duration, history of present illness, past medical history and medications, ocular examination, complementary tests, hypotheses, differential diagnoses, and management.

The instructions remained faithful to the original proposal: the assessment was not intended to test memorization or the ability to recognize the diagnosis represented in the photograph. Instead, students were required to justify their choices and explicitly describe the anatomical and pathophysiological relationships underlying each step of their reasoning.

It was at this stage, a formal assessment with substantial weight in the final grade, that a new participant was introduced into the process.

After the examination, the instructions to the students, the handwritten answers of 10 students and the corresponding clinical images were provided directly to ChatGPT, using OpenAI’s GPT-5.6 Sol model in a private ChatGPT Business workspace. The material was processed exclusively in that environment, without publication, deliberate sharing with third parties, use for model training, or submission to external internet searches during the analysis. The prompt used for grading was essentially the same as the instruction given to the students.

Nothing was previously typed or organized by the instructor. The answers were not transcribed, the students were not identified, the images were not described, and no diagnosis or answer key was provided. The material presented to the artificial intelligence consisted solely of the student’s instructions (the prompt), photographs of the handwritten examinations and the corresponding clinical images.

The tool was able to recognize the students’ names, interpret the handwriting, reconstruct the answers, and relate them to the respective photographs. The interpretations of the conditions represented in the images were considered accurate by the

<http://dx.doi.org/10.5935/0004-2749.2026-1027>

Submitted for publication:

September 25, 2026

Accepted for publication:

September 28, 2026

Funding:

This study received no specific financial support.

Disclosure of potential conflicts of interest:

The author declares no potential conflicts of interest.

Corresponding author:

Paulo Schor

E-mail: pschor@unifesp.br

Data Availability Statement:

The datasets generated and/or analyzed during the current study are already available.

Edited by

Editor-in-Chief: Newton Kara-Júnior



instructor. More importantly, the analysis was conducted at the level of complexity requested: for each student, the tool reconstructed the proposed case, assessed the pertinence of the history and examination, identified strengths, gaps, and inconsistencies, analyzed the complementary tests and management, and provided an overall assessment.

Thus, it did not limit itself to determining whether the student had “gotten the photograph right”. Instead, it sought to assess how the student had arrived at their conclusion.

Only after this independent analysis were the artificial intelligence’s assessments compared with those of the instructor.

For most students, there was excellent agreement, particularly in identifying the strongest and weakest answers and the reasoning underlying those assessments. In some cases, however, differences emerged. These differences proved to be among the most interesting aspects of the experience.

The artificial intelligence appeared to place particular emphasis on internal coherence, organization of thought, the relationship between findings and anatomical structures, and the prioritization of clinical decisions. The instructor, in contrast, sometimes assigned greater weight to the diagnosis he had initially inferred from the photograph or to specific content emphasized during the clerkship.

One case made this difference particularly clear.

Faced with one of the images, a student constructed a case of ocular chemical burn. The instructor had not interpreted the photograph as depicting a burn and, on first reading, considered the answer unsatisfactory.

The artificial intelligence drew attention to another aspect. Once a coherent history of chemical exposure had been constructed, the student correctly recognized that the priority was not to perform additional tests or refine the diagnosis but to initiate immediate and copious ocular irrigation. The student had also logically identified the chemical agent involved, its possible consequences, and the need for referral and follow-up.

The analysis therefore emphasized that the student had understood something more important than the actual diagnosis represented in the photograph: the student had understood the hierarchy and sequence of decisions in an emergency, in which immediate intervention must precede diagnostic refinement.

This observation led the instructor back to the instructions given to the students.

The students had not been asked to determine the actual diagnosis represented in the photograph. Rather, they had been asked to construct a coherent case and make their reasoning explicit.

Under this criterion, the answer was better than it had initially appeared.

The grade was revised.

Perhaps this episode summarizes the experience more effectively than any measure of agreement between the evaluators.

The artificial intelligence did not know the student, had not attended the classes, and did not possess the educational experience accumulated by the instructor. This was an evident limitation. The instructor knows the lived curriculum, understands what was emphasized during the clerkship, and recognizes the relative importance of certain errors.

On the other hand, the machine displayed an interesting characteristic: it received an explicit instruction and applied it without necessarily incorporating the evaluator’s implicit expectations.

In this case, it also served as a reminder for the instructor himself to remain faithful to the prompt he had drafted.

In other answers, the tool highlighted the difference between simply filling in the fields of a prescribed script and effectively establishing relationships among structure, mechanism, clinical manifestation, and clinical decision-making. It also identified situations in which the student’s own hypothesis should have altered the sequence of investigation, as in suspected open-globe injury, in which certain tests are not merely unnecessary but may be inappropriate or even dangerous.

In some of these cases, the second reading also changed the instructor’s initial perception of the answer.

This may be the most promising aspect of an experience that resonates with the possibilities presented by the rapidly expanding field of artificial intelligence in medical education⁽²⁾.

The purpose is not to use artificial intelligence to replace the instructor or to delegate grading decisions to a machine. Its role may be more valuable as a second, independent reader of clinical reasoning.

In science, second readers are familiar. Reviewers do not need to agree completely to be useful: agreement increases confidence, whereas disagreement can compel the explicit examination of criteria, weighting, and assumptions that might otherwise remain intuitive. In open-ended assessments, this exercise is particularly relevant. When the assessment moves beyond a simple right-or-wrong

framework and focuses on reasoning, judgment inevitably becomes part of the process. Two students may arrive at different diagnoses and still construct equally coherent cases; an error expressed in a few words may be conceptually serious, whereas an extensive omission may be of limited importance in a particular case.

More than a decade ago, a photograph and a deliberately unstable case were used to move students beyond their comfort zones and make the construction of their reasoning visible. Artificial intelligence now adds a dimension that was not foreseen in that original experience: by providing a second reading, it not only helps reveal how the student thinks but also makes the criteria and reasoning of the evaluator more visible.

The journey, therefore, continues to matter more than the arrival. Previously, the objective was to observe the path taken by the student. Now, it is also possible to observe how that path is interpreted by the instructor.

One more observer accompanies the journey, and, in doing so, also helps illuminate the perspective of the evaluator.

This editorial was conceived and written by the author, who also verified and reviewed its final content. Generative artificial intelligence tools (ChatGPT, OpenAI) were used solely to assist with the organization, structuring, and formatting of the manuscript, without contributing to its intellectual conception or assuming responsibility for its content.

REFERENCES

1. Chamon W, Schor P. Teaching ophthalmology to the medical student: a novel approach [editorial]. *Arq Bras Oftalmol.* 2012;75(1):5–7.
2. Kara-Junior N, Morinaga CV, Machado DG. The artificial intelligence revolution in medical education. *Arq Bras Oftalmol.* 2026;89(1):e2026-1020.

Complementary reading

3. Meira S, Meira L. The university after generative AI: A field guide for designing AI-native higher education, 2026–2035. Recife: TDS; CESAR School, 2026. [cited 2026 Sep 24]. Available at: <https://biblioteca.tds.company/ebook-the-university-after-generative-ai>.