

BRIEF COMMUNICATION



Modern threats in academia: evaluating plagiarism and artificial intelligence detection scores of ChatGPT

Andrea Taloni¹, Vincenzo Scoria¹ and Giuseppe Giannaccare¹✉

© The Author(s), under exclusive licence to The Royal College of Ophthalmologists 2023

Eye (2024) 38:397–400; <https://doi.org/10.1038/s41433-023-02678-7>

Plagiarism and research integrity are sensitive issues in the academic setting, especially after the recent offspring of artificial intelligence (AI) and large language models (LLMs) such as GPT-4.0 [1]. As the popularity of ChatGPT increases, some authors have attempted to write abstracts and full-text articles using AI, obtaining essays that resemble genuine scientific papers [2–4]. Detection systems for AI-generated texts have been recently developed. Miller et al. performed AI detection on a large sample of abstracts belonging to articles published between 2020 and 2023, reporting a significant increase in AI-assisted writing [5].

We evaluated herein the plagiarism and AI-detection scores of GPT-4.0 when paraphrasing original scientific essays, and furthermore tested methods that could possibly evade AI detection.

The investigation was conducted on 20 abstracts belonging to published articles indexed on PubMed. The search query included full-text articles published on the *Eye* Journal in 2023, and results were sorted by “best match”. The first listed papers, 5 for each article type (i) meta-analysis (ii) randomized controlled trial, (iii) review and (iv) systematic review, were selected for the analysis. The abstracts were submitted to GPT-4.0, followed by the prompt “Paraphrase this scientific abstract”; for structured abstracts the sentence “Paraphrase this scientific abstract, maintaining the same structure” was used. AI-paraphrased abstracts were submitted to QueText (<https://www.que-text.com/>) for plagiarism checking and to Originality.AI (<https://app.originality.ai/>) for AI detection. Afterwards, Undetectable.AI (<https://undetectable.ai/>) was used to “humanize” the AI-paraphrased abstracts in an attempt to evade AI detection (settings: readability = “doctorate”, purpose = “article”, humanization strength = “balanced”). Humanized abstracts

were resubmitted to Originality.AI looking for investigating the reduction of the AI-detection score (Table 1). Abstract readability was assessed by means of Flesch Reading Ease score, calculated with Microsoft Office Word 365. Numerical values were expressed as mean ± standard deviation. Two-sided paired Student’s *t* test was performed to compare variables between original, paraphrased, and humanized abstracts. A *P* < 0.05 was considered statistically significant.

The paraphrased abstracts produced by GPT-4.0 obtained a mean plagiarism score of 10.7 ± 12.7% [95% CI, 5.1%–16.3%] and a mean AI-detection score of 91.3 ± 17.9% [95% CI, 83.4%–99.1%]. After humanizing the abstracts, the AI-detection score lowered significantly (27.8 ± 33.2% [95% CI, 13.2%–42.3%]; *P* < 0.0001). The difference in readability was not significant between original and paraphrased abstracts (*P* > 0.7); conversely, a significant increase in Flesch Reading Ease (*P* = 0.003) was reported for humanized abstracts compared to original ones (18.4 ± 12.3 [95% CI, 13.0–23.8] vs 9.7 ± 11.0 [95% CI, 4.9–14.6]; *P* = 0.003). Table 2 contains all data recorded from the abstracts. Despite the overall style of the paraphrased abstracts was similar to the original ones, humanized abstracts presented minor punctuation mistakes.

These findings, albeit limited by the small sample size, highlight the capability of GPT-4.0 to produce plagiarism-free scientific essays, and, when refined using humanizing tools, to evade AI detection.

It should be a priority of publishers to improve detection systems for AI-generated texts, caching up on the continuous upgrades of LLMs, and on the malicious attempts to circumvent AI detection. The fate of this modern challenge will profoundly affect the future of scientific research.

¹Department of Ophthalmology, University Magna Graecia of Catanzaro, Catanzaro, Italy. ✉email: giuseppe.giannaccare@unicz.it

Received: 18 July 2023 Accepted: 18 July 2023

Published online: 2 August 2023

Table 1. Representative examples of two analyzed abstracts reported in the original, paraphrased, and humanized form.

Original Abstract		Paraphrased Abstract by GPT-4.0	Humanized Abstract by UndetectableAI
Title The cross-sectional and longitudinal relationship of diabetic retinopathy to cognitive impairment: a systematic review and meta-analysis	Objectives: To establish a potential relationship between diabetic retinopathy (DR) and different stages of cognitive impairment. Methods: Literature searches were conducted on PubMed and EMBASE, with keywords "diabetic retinopathy" and "cognitive impairment". Inclusion criteria were original human studies, and English language. Quality of studies was assessed by the Newcastle-Ottawa Quality Assessment (NOSGEN). The register number of this study on the International Prospective Register of Systematic Reviews (PROSPERO) is CRD42021236747. The main outcome measures were odds ratios (OR) and risk ratios (RR) for cross-sectional and longitudinal studies, respectively. Meta-regression was performed to evaluate the effects of potential moderator variables, including, age, onset age of diabetes mellitus (DM), duration of DM, and HbA1c. Results: Twenty-five studies (17 cross-sectional and 8 longitudinal studies) with a total of 1,963,914 subjects, were included. Among the cross-sectional studies, the pooled ORs of any cognitive impairment, early stage of cognitive impairment and dementia in subjects with DR (95% confidence interval) were 1.48 (1.08–2.02), 1.59 (1.01–2.51), and 1.13 (0.86–1.50), respectively. Among the longitudinal studies, the pooled RRs of any cognitive impairment, early stage of cognitive impairment, and dementia in subjects with DR (95% confidence interval) were 1.35 (1.12–1.65), 1.50 (1.06–2.12), and 1.31 (1.03–1.66), respectively. Meta-regression showed age, onset age of DM, duration of DM, and glycated hemoglobin (HbA1c) were not statistically associated with the outcomes. Conclusions: The presence of DR in DM patients indicates both higher odds of prevalent cognitive impairment and escalated risks of developing cognitive impairment in the future	Objective: The aim of this research was to examine the possible link between diabetic retinopathy (DR) and various stages of cognitive impairment. To achieve this, an extensive literature review was conducted on PubMed and EMBASE, using the key terms "diabetic retinopathy" and "cognitive impairment". We included original human studies written in English in our review. The Newcastle-Ottawa Quality Assessment (NOSGEN) was employed to evaluate the quality of the studies. The study is registered with the International Prospective Register of Systematic Reviews (PROSPERO) under the number CRD42021236747. We primarily measured the odds ratios (OR) for cross-sectional studies and risk ratios (RR) for longitudinal studies. Additionally, we conducted a meta-regression analysis to assess the impact of potential influencing variables such as age, the age when diabetes mellitus (DM) was first diagnosed, the duration of DM, and HbA1c levels. Results: Our analysis included twenty-five studies (17 cross-sectional and 8 longitudinal) encompassing a total of 1,963,914 subjects. The aggregated ORs from cross-sectional studies showed that the chances of any cognitive impairment, early cognitive impairment, and dementia in individuals with DR were 1.48 (1.08–2.02), 1.59 (1.01–2.51), and 1.13 (0.86–1.50), respectively. The pooled RRs from longitudinal studies suggested that the risks of any cognitive impairment, early cognitive impairment, and dementia in individuals with DR were 1.35 (1.12–1.65), 1.50 (1.06–2.12), and 1.31 (1.03–1.66), respectively. Upon conducting a meta-regression analysis, we found that age, age of DM onset, DM duration, and HbA1c levels did not significantly influence the results. Conclusions: Our findings suggest that DR's presence in DM patients indicates a higher likelihood of existing cognitive impairment and an increased risk of cognitive decline in the future.	The purpose of this research was to investigate the connection between diabetic retinopathy (DR) and different levels of cognitive impairment. To accomplish this, we thoroughly reviewed literature on PubMed and EMBASE using the search terms "diabetic retinopathy" and "cognitive impairment". Our review focused on English, conducted with human participants and written in English. We utilized the Newcastle Ottawa Quality Assessment (NOSGEN) to evaluate the quality of these studies. The study has been registered with the International Prospective Register of Systematic Reviews (PROSPERO) under the registration number CRD42021236747. We primarily examined odds ratios (OR) for cross-sectional studies and risk ratios (RR) for longitudinal studies. Moreover, we performed a meta-regression analysis to examine the influence of factors like age, age at diabetes mellitus (DM) diagnosis, duration of DM, and HbA1c levels. Findings: Our analysis incorporated a total of twenty-five studies (17 cross-sectional and 8 longitudinal) involving 1,963,914 participants. Based on a combination of cross-sectional and longitudinal studies, we observed that individuals with diabetic retinopathy (DR) have an increased chance of experiencing cognitive impairment. The analysis indicated that the risks of any impairment early cognitive impairment and dementia were higher in those with DR. Specifically, the aggregated odds ratios (ORs) from cross-sectional studies showed a 1.48 fold increase in the chances of any cognitive impairment, a 1.59 fold increase in the chances of early cognitive impairment, and a 1.13 fold increase in the chances of dementia. Similarly, the pooled relative risks (RRs) from studies suggested a 1.35 fold increase in the risks of any cognitive impairment, a 1.50 fold increase in the risks of early cognitive impairment, and a 1.31 fold increase in the risks of dementia. Notably, factors such as age, age of diabetes onset, diabetes duration, and HbA1c levels did not significantly impact these findings based on our meta-regression analysis. In conclusion, our findings indicate that the presence of retinopathy in individuals with diabetes mellitus (DM) may be indicative of both existing cognitive impairment and an increased risk of cognitive decline in the future.
Bifocal use in hyperopic anisometropic amblyopia treated with atropine: a proof-of-concept randomized trial.	Objectives: To investigate the effect of bifocal wearing in the amblyopic eye when atropine is used in the sound eye for the treatment of hyperopic anisometropic amblyopia. Methods: Children 4–8 years old were randomly assigned to bifocal + atropine ($n = 16$) or only atropine (control, $n = 19$) groups of treatment in a proof-of-concept study. Measurements included visual acuity (logMAR), prism and cover test, stereoaquity (Randot preschool or Randot circles), contrast sensitivity (MARS test), accommodation (Grand Seiko WAM5500 and dynamic retinoscopy), retinoscopic and subjective refraction, before starting treatment and at 6 months, except accommodation, which was remeasured at 9–11 months. Main outcome measure was change in logMAR lines of visual acuity, and secondary outcome measures were change in stereoaquity and contrast sensitivity in the amblyopic eye, at 6 months. Results: Improvement in visual acuity of the amblyopic eye was significantly better ($p = 0.04$) in the atropine plus bifocal (3.3 ± 0.9 logMAR lines) than in the atropine only group (2.6 ± 0.8 logMAR lines), whereas change in stereoaquity and contrast sensitivity was not significantly different between the two groups. Differences in accommodative gain, which was impaired in the amblyopic compared to the sound eye, before treatment, decreased after treatment, in the atropine group (0.62 ± 0.15 vs 0.79 ± 0.2, $p = 0.3$), and atropine + bifocal group (0.69 ± 0.15 vs 0.82 ± 0.2, $p = 0.4$). Conclusions: Use of bifocal lens add in the amblyopic eye of children with hyperopic anisometropic amblyopia, treated by atropine penalization, is beneficial in the follow-up period of 6 months.	Objective: This study aims to understand the impact of wearing bifocal glasses in the amblyopic (lazy) eye while using atropine in the unaffected eye for treating hyperopic anisometropic amblyopia (a condition where the eyes have unequal focusing power leading to lazy eye). Methods: We conducted a preliminary study involving children aged 4–8 years. These children were randomly divided into two groups: one treated with bifocal glasses plus atropine (16 children) and the other treated with atropine alone (19 children). We measured various parameters such as visual sharpness (logMAR), eye alignment, depth perception (using Randot preschool or Randot circles), contrast sensitivity (through the MARS test), eye focusing ability (with Grand Seiko WAM5500 and dynamic retinoscopy), and refraction, both before the treatment started and after six months. The only exception was the focusing ability, which was measured again between 9 and 11 months. The primary measurement was the change in logMAR lines of visual acuity, and the secondary measurements were changes in depth perception and contrast sensitivity in the lazy eye after six months. Results: The visual acuity in the lazy eye improved more significantly ($p = 0.04$) in the group treated with atropine and bifocal glasses (improvement of 3.3 ± 0.9 logMAR lines) compared to the group treated with only atropine (improvement of 2.6 ± 0.8 logMAR lines). However, the change in depth perception and contrast sensitivity didn't differ significantly between the two groups. The differences in the ability to focus, which was poorer in the lazy eye before treatment, reduced after treatment in both the atropine group (0.62 ± 0.15 vs 0.79 ± 0.2, $p = 0.3$) and the atropine plus bifocal group (0.69 ± 0.15 vs 0.82 ± 0.2, $p = 0.4$). Conclusions: Adding bifocal glasses to the lazy eye of children suffering from hyperopic anisometropic amblyopia and treating them with atropine penalization proves beneficial over a six-month follow-up period.	Objective: The main purpose of this study is to investigate how wearing glasses in the weaker eye while using atropine in the stronger eye affects the treatment of hyperopic anisometropic amblyopia. This condition arises when there is a difference in focusing power between the eyes leading to an eye. Methods: To conduct this study, we enlisted children aged between 4 and 8 years. They were randomly divided into two groups: one group received treatment with glasses along with atropine (consisting of 16 children) while the other group received only atropine as treatment (comprising 19 children). We measured factors such as visual sharpness (logMAR) eye alignment, depth perception using Randot preschool or Randot circles contrast sensitivity through the MARS test eye focusing ability utilizing Grand Seiko WAM5500 and dynamic retinoscopy and refraction. These measurements were taken before treatment initiation. Again after six months with an exception for eye focusing ability, which was reassessed between 9 and 11 months. The primary parameter assessed was the change in logMAR lines of acuity while secondary measurements included changes in depth perception and contrast sensitivity in the lazy eye after six months. Results: The group that received both atropine and bifocal glasses showed a significant improvement in the visual acuity of the lazy eye ($p = 0.04$). Specifically their visual acuity improved by an average of 3.3 ± 0.9 logMAR lines. On the hand the group that only received atropine experienced an improvement of 2.6 ± 0.8 logMAR lines. However there wasn't a difference between the two groups when it came to changes in depth perception and contrast sensitivity. Both groups showed improvement in the ability to focus which was initially poor in the eye before treatment. This improvement was observed in both the atropine group (0.62 ± 0.15 to 0.79 ± 0.2, $p = 0.3$) and the atropine plus bifocal group (0.69 ± 0.15 to 0.82 ± 0.2, $p = 0.4$). In conclusion adding glasses to the treatment of children with hyperopic anisometropic amblyopia and using atropine penalization proves beneficial, over a six month follow up period.

Table 2. Characteristics of 20 analyzed abstracts according to plagiarism and artificial intelligence detection scores.

Title	Article type	Original abstract		Paraphrased abstract		Humanized abstract	
		FRE	AI detection	FRE	Plagiarism	AI detection	Plagiarism
Efficacy, safety, and treatment burden of treat-and-extend versus alternative anti-VEGF regimens for nAMD: a systematic review and meta-analysis.	Meta-Analysis	25.6	100%	17.8	0%	100%	11%
The cross-sectional and longitudinal relationship of diabetic retinopathy to cognitive impairment: a systematic review and meta-analysis	Meta-Analysis	0	100%	0	39%	100%	41%
Oculomotor deficits in attention deficit hyperactivity disorder: a systematic review and meta-analysis	Meta-Analysis	16.6	94%	16.3	8%	22.2	22%
Role of anti-vascular endothelial growth factor in the management of non-proliferative diabetic retinopathy without center-involving diabetic macular oedema: a meta-analysis of trials	Meta-Analysis	0	76%	1	28%	1	18%
Diagnostic accuracy of OCTA and OCT for myopic choroidal neovascularisation: a systematic review and meta-analysis	Meta-Analysis	0	100%	0	25%	7.3	9%
Bifocal use in hyperopic anisometropic amblyopia treated with atropine: a proof-of-concept randomized trial.	Randomized Controlled Trial	0	84%	0	0%	20.3	7%
Comparison of breath-guards and face-masks on droplet spread in eye clinics	Randomized Controlled Trial	27.6	100%	19.6	0%	33.1	17%
Role of fluorescein angiography guided laser treatment in aggressive retinopathy of prematurity	Randomized Controlled Trial	18	100%	19.5	0%	1	7%
Non-penetrating deep sclerectomy with the sub flap (Ahmed's) suture: a 12-month comparative study	Randomized Controlled Trial	0	74%	0	0%	0	0%
Outcome of transcanalicular laser dacryocystorhinostomy with endonasal augmentation in acute versus post-acute dacryocystitis	Randomized Controlled Trial	11.6	100%	12.1	3%	23.3	3%
Conjunctival Lymphoma	Review	0	100%	2.4	28%	13.3	16%
Retinoblastoma and vision	Review	28.2	100%	14.8	0%	34.1	0%
Painting unknown worlds	Review	28.9	100%	17.7	0%	35.6	0%
Malignant lesions of the caruncle	Review	6.3	100%	11.7	29%	28.4	22%
Optical coherence tomography as retinal imaging biomarker of neuroinflammation/neurodegeneration in systemic disorders in adults and children	Review	0	89%	0	0%	0	0%
Home-based screening tools for amblyopia: a systematic review	Systematic Review	21.3	100%	12.2	18%	21.9	22%
Patient-reported outcome measures in vitreoretinal surgery: a systematic review	Systematic Review	0	22%	10.5	8%	23.5	30%
Clinical trials targeting the gut-microbiome to effect ocular health: a systematic review	Systematic Review	1	86%	27.5	7%	23.7	14%
Myopia prediction: a systematic review	Systematic Review	5.8	100%	20.1	0%	36.8	15%
Global and regional prevalence of age-related cataract: a comprehensive systematic review and meta-analysis	Systematic Review	3.5	100%	4.6	21%	16.6	13%

FRE Flesch Reading Ease.

REFERENCES

1. OpenAI, <https://openai.com/>. Accessed June 2023.
2. Gao CA, Howard FM, Markov NS, Dyer EC, Ramesh S, Luo Y, et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digital Med.* 2023;6:1–5.
3. Else H. Abstracts written by ChatGPT fool scientists. *Nature.* 2023;613:423.
4. Májovský M, Černý M, Kasal M, Komarc M, Netuka D. Artificial intelligence can generate fraudulent but authentic-looking scientific medical articles: pandora's box has been opened. *J Med Internet Res.* 2023;25:e46924. <https://doi.org/10.2196/46924>.
5. Miller LE, Bhattacharyya D, Miller VM, Bhattacharyya M. Recent trend in artificial intelligence-assisted biomedical publishing: a quantitative bibliometric analysis. *Cureus.* 2023;15:e39224. <https://doi.org/10.7759/CUREUS.39224>.

AUTHOR CONTRIBUTIONS

Conceptualization, AT and GG; Methodology, AT, GG and VS; Validation, GG and VS; Formal Analysis, AT, and GG; Investigation, AT; Data Curation, AT; Writing—Original Draft Preparation, AT and GG; Writing—Review and Editing, AT, GG and VS; Visualization, AT, GG and VS; Supervision, GG and VS; Project Administration, AT, GG and VS. All authors have read and agreed to the published version of the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Giuseppe Giannaccare.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.