Titolo rivista/libro: JAMA Ophthalmology

Titolo articolo/sezione: Large Language Model Advanced Data Analysis Abuse to Create a Fake Data Set in Medical Research

Autore/i: Taloni A , Scorcia V , Giannaccare G

ISSN: 2168-6173

DOI: 10.1001/jamaophthalmol.2023.5162

Anno: 2023

Volume: 141

Fascicolo: 12

Editore:

Pag. iniziale: 1174

Pag. finale: 1175

6. Leske MC, Heijl A, Hyman L, et al. Predictors of long-term progression in the early manifest glaucoma trial. *Ophthalmology*. 2007;114(11):1965-1972. doi:10.1016/j.ophtha.2007.03.016

Large Language Model Advanced Data Analysis Abuse to Create a Fake Data Set in Medical Research

Since its public release by OpenAI in late 2022, ChatGPT (generative pretrained transformer) has gone through several updates. In March 2023, the large language model (LLM) GPT-4 came out, marking improvements in semantic understanding and response generation compared with GPT-3.5.[1] More recently, the capabilities of GPT-4 were expanded with Advanced Data Analysis (ADA), a model that runs Python.[2] This tool supports file uploads and downloads and can perform both statistical analysis and data visualization.[3] Although these features may speed up scientific research, unethical uses of ADA are conceivable, including fabrication of data. We evaluated the ability of GPT-4 ADA (OpenAI; 09/11/2023 version) to create a fake data set that can be used for scientific research.

Supplemental content

Methods | In this quality improvement study, a detailed prompt, providing criteria to compile each column of the database, was submitted to ADA (eFigure in Supplement 1). The LLM was asked to fabricate data for 300 eyes belonging to 250 patients with keratoconus who underwent deep anterior lamellar keratoplasty (DALK) or penetrating keratoplasty (PK). For categorical variables, target percentages were predetermined for the distribution of each category. For continuous variables, target mean and range were defined. Additionally, ADA was instructed to fabricate data that would result in a statistically significant difference between preoperative and postoperative values of best spectacle-corrected visual acuity (BSCVA) and topographic cylinder. ADA was programmed to yield significantly better visual and topographic results for DALK compared with PK. Upon completion, the fake database was analyzed with SPSS Statistics, version 29.0 (IBM Corp). A 2-tailed *t* test was performed

Table 1. Comparison Between Requested Values of the Submitted Prompt and Obtained Values of the Fake Data Set: Categorical Variables

Categorical variable	Requested value, %	Obtained value, No. (%)
Sex		
Male	55.0	160 (53.3)
Female	45.0	140 (46.7)
Surgery		
DALK	50.0	150 (50.0)
PK	50.0	150 (50.0)
Bubble burst		
DALK	1.0[a]	0[a]
PK	0	0
Microperforations		
DALK	7.5	11 (7.3)
PK	0	0
Choroidal hemorrhage		
DALK	0	0
PK	0.5[a]	0[a]
Stromal immune rejection		
DALK	5	11 (7.3)
PK	5[a]	3 (2.0)[a]
Endothelial immune rejection		
DALK	0	0
PK	10	15 (10.0)

Abbreviations: DALK, deep anterior lamellar keratoplasty; PK, penetrating keratoplasty.

[a] Obtained numerical value that differs from the requested value.

Table 2. Comparison Between Requested Values of the Submitted Prompt and Obtained Values of the Fake Data Set: Continuous Variables

Continuous variable	Requested value, mean (range)	*P* value	Obtained value, mean (SD) [range]	Difference between requested and obtained values, mean (SD) [range; 95% CI]	*P* value
Age at surgery, y	35.0 (18 to 70)[a]	NA	32.9 (13.1) [17 to 59][a]	NA	NA
BSCVA, logMAR					
Preoperative	1.00 (0.50 to 1.80)[a]	<.05	1.00 (0.20) [0.50 to 1.60][a]	−0.79 (0.22) [−1.40 to −0.20; −0.81 to −0.76]	<.001
Postoperative	0.20 (0.00 to 1.00)[a]		0.21 (0.11) [0.00 to 0.50][a]		
DALK postoperative	NA[b]	<.05	0.18 (0.11) [0.00 to 0.50]	−0.05 (0.15) [−0.4 to 0.3; −0.08 to −0.03]	<.001
PK postoperative	NA[b]		0.24 (0.10) [0.00 to 0.50]		
Topographic cylinder, D					
Preoperative	4.50 (1.00 to 18.00)[a]	<.05	4.45 (1.60) [1.00 to 8.85][a]	−1.36 (1.81) [−5.73 to 3.8; −1.57 to −1.16]	<.001
Postoperative	3.00 (0.50 to 12.00)[a]		3.09 (0.83) [1.06 to 5.16][a]		
DALK postoperative	NA[b]	<.05	3.07 (2.22) [0.5 to 9.52]	−0.91 (1.00) [−3.61 to 2.07; −1.08 to −0.76]	<.001
PK postoperative	NA[b]		3.83 (2.67) [0.50 to 11.19]		

Abbreviations: BSCVA, best spectacle-corrected visual acuity; D, diopters; DALK, deep anterior lamellar keratoplasty; NA, not applicable; PK, penetrating keratoplasty.

[a] Obtained numerical value that differs from the requested value.

[b] Requested values for SDs and 95% CIs were not provided in the prompt submitted to Advanced Data Analysis. Similarly, no target data were provided for BSCVA and topographic cylinder of eyes that underwent DALK or PK.

to compare continuous variables. $P < .05$ was considered significant. This study followed the Standards for Quality Improvement Reporting Excellence (SQUIRE) reporting guideline. Because the study did not involve humans or animals, no ethical approval or informed consent was required.

Results | The LLM followed a step-by-step approach to build the data set. Sometimes answer generation stopped due to word count limit; in these cases, the prompt "Continue" was used to resume the process until a data set was successfully created (eFigure in Supplement 1). As requested, mean (SD) postoperative BSCVA and topographic cylinder for DALK were significantly better compared with PK (difference: BSCVA, −0.05 [0.15] logMAR [95% CI, −0.08 to −0.03 logMAR]; topographic cylinder, −0.91 [1.00] diopters [95% CI −1.08 to −0.76 diopters]; $P < .001$). Most statistical criteria provided in the prompt were met by the data set; however, data ranges of continuous variables were not always accurate (**Table 1** and **Table 2**).

Discussion | The LLM created a seemingly authentic database, showing better results for DALK than PK. Recently, we expressed concerns regarding the capability of an LLM to produce plagiarism-free scientific essays while evading artificial intelligence (AI) detection.[4] ADA may pose a greater threat, being able to fabricate data sets specifically designed to quickly produce false scientific evidence, such as the better outcomes of DALK over PK that have not been proved, to our knowledge, by scientific evidence.[5] Illegitimate data manipulation has been repeatedly reported in academia[6]; however, recognition of research misconduct is still an outstanding issue with no definitive solutions. Potential strategies to identify AI data fabrication may involve looking for peculiar statistical patterns in the data sets, similar to technology that detects AI-generated text by evaluating the likelihood of nonhuman patterns.[4] On the other hand, sponsors may prevent data fabrication by equipping study centers with digital diagnostic devices that store measurements inside onboard memory. Thus, encrypted data backups could be uploaded directly to the sponsor's online database platform, without allowing manual transcription of data by the investigators. Additionally, registration of clinical trials on ClinicalTrials.gov and adherence to CONSORT checklists, protocols, and statistical analysis plans could also safeguard data authenticity. Publishers and editors might want to consider the findings of this study in the peer-review process, to ensure that AI advancements will enhance, not undermine, the integrity and value of scientific research.

Andrea Taloni, MD
Vincenzo Scorcia, MD
Giuseppe Giannaccare, MD, PhD

Author Affiliations: Department of Ophthalmology, University Magna Graecia of Catanzaro, Catanzaro, Italy (Taloni, Scorcia, Giannaccare); Eye Clinic, Department of Surgical Sciences, University of Cagliari, Cagliari, Italy (Giannaccare).

Accepted for Publication: September 18, 2023.

Published Online: November 9, 2023. doi:10.1001/jamaophthalmol.2023.5162

Corresponding author: Giuseppe Giannaccare, MD, PhD, Eye Clinic, Department of Surgical Sciences, University of Cagliari, Via Università 40, Cagliari 09124, Italy (giuseppe.giannaccare@gmail.com).

Author Contributions: Dr Taloni had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: All authors.
Acquisition, analysis, or interpretation of data: Taloni, Giannaccare.
Drafting of the manuscript: Taloni.
Critical review of the manuscript for important intellectual content: All authors.
Statistical analysis: Taloni.
Obtained funding: Giannaccare.
Supervision: All authors.

Conflict of Interest Disclosures: None reported.

Data Sharing Statement: See Supplement 2.

1. Open AI. Accessed June 20, 2023. https://openai.com/

2. Python.org. What is Python? Executive summary. Accessed August 23, 2023. https://www.python.org/doc/essays/blurb/

3. Wang L, Ge X, Liu L, Hu G. Code interpreter for bioinformatics: are we there yet? *Ann Biomed Eng*. Published online July 23, 2023. doi:10.1007/s10439-023-03324-9

4. Taloni A, Scorcia V, Giannaccare G. Modern threats in academia: evaluating plagiarism and artificial intelligence detection scores of ChatGPT. *Eye (Lond)*. Published online August 2, 2023. doi:10.1038/s41433-023-02678-7

5. Lucisano A, Scorcia V, Taloni A, Rossi C, Gioia R, Giannaccare G. Impact of topographic localization of corneal ectasia on the outcomes of deep anterior lamellar keratoplasty employing large (9 mm) versus conventional diameter (8 mm) grafts. *Eye (Lond)*. 2023. Published online April 20, 2023. doi:10.1038/s41433-023-02536-6

6. Xie Y, Wang K, Kong Y. Prevalence of research misconduct and questionable research practices: a systematic review and meta-analysis. *Sci Eng Ethics*. 2021;27(4):41. doi:10.1007/s11948-021-00314-9

COMMENT & RESPONSE

Risk of New Retinal Vascular Occlusion After Messenger RNA COVID-19 Vaccination Within Aggregated Electronic Health Record Data

To the Editor We read with interest the article by Dorney et al.[1] We would like to report some potentially relevant information that was published by our colleagues[2] soon after their publication.

Specifically, the relatively short 21-day time frame used in some studies may not be sufficient to fully assess the potential cause-effect association between COVID-19 vaccinations and retinal vascular occlusion (RVO). In a study published in May 2023 by Li et al,[2] a relatively longer time frame was associated with somewhat different conclusions. In their study, 739 066 COVID-19-vaccinated patients from the same TriNetX database with a 1:1 propensity-matched unvaccinated cohort were assessed for occurrences of RVO using the same codes. These results showed a cumulatively increased risk of RVO in the COVID-19–vaccinated cohort compared with the COVID-19–unvaccinated cohort at the 12-week and 2-year time points after vaccination.

One difference between the Dorney et al study[1] and previous studies may be associated with the method of data collection. Their study assessed RVO occurrence at a biweekly interval up to 12 weeks and then at 2 years. This longer time frame may have allowed their study to capture certain RVO occurrences after vaccination associated with autoimmune ad-

Supplementary Online Content

Taloni A, Scorcia V, Giannaccare G. Large Language Model Advanced Data Analysis abuse to create a fake data set in medical research. *JAMA Ophthalmol.* Published online November 9, 2023. doi:10.1001/jamaophthalmol.2023.5162

eFigure. (a) Prompt Submitted to Advanced Data Analysis. (b) Sample of 40 Data Entries From the Fake Data Set Fabricated by Advanced Data Analysis on Microsoft Office Excel 365

This supplementary material has been provided by the authors to give readers additional information about their work.

eFigure. (a) Prompt Submitted to Advanced Data Analysis. (b) Sample of 40 Data Entries From the Fake Data Set Fabricated by Advanced Data Analysis on Microsoft Office Excel 365

a

b

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Surname and Name	Sex	Birthdate	Eye	Date of Surgery	Age at Surgery	Preoperative BSCVA	Preoperative Topographic Cylinder	Surgery	Postoperative BSCVA (1 year)	Postoperative Topographic Cylinder (1 year)	Intraoperative Complications	Postoperative Complications
2	Alison Green	M	05/06/1976	RE	11/08/2008	32	1.3	3.37	DALK	0	2.98	None	None
3	Jared Friedman	F	13/04/1997	LE	13/04/2015	17	0.8	5.18	DALK	0	4.38	None	None
4	Jessica Cohen	M	21/03/1976	RE	28/12/2008	32	0.9	4.66	DALK	0.0	2.15	None	None
5	Sara Young DDS	F	12/09/1964	LE	03/04/2010	45	1.1	4.97	DALK	0.1	2.88	None	None
6	Whitney Douglas	M	26/01/1974	RE	02/01/2009	34	1.1	2.77	PK	0.3	2.78	None	None
7	Daniel Watkins	M	13/11/1987	RE	03/11/2006	18	0.8	6.04	DALK	0.1	3.16	None	None
8	Jennifer Johnson	M	05/08/1952	RE	31/07/2010	57	0.9	5.5	PK	0.2	3.07	None	None
9	Tara Thompson	M	16/06/1999	RE	16/06/2017	18	0.9	3.37	PK	0.3	3.29	None	Endothelial Immune Rejection
10	Jennifer Dixon	M	30/11/1977	RE	14/03/2007	29	0.8	2.84	DALK	0.3	1.74	None	None
11	Susan Howe	M	19/04/1955	LE	04/04/2005	49	0.9	5.41	PK	0.2	4.68	None	None
12	Daniel Mccarthy	F	11/10/1989	LE	30/04/2009	19	1.2	3.33	PK	0.1	4.09	None	None
13	Bianca Boyd	M	07/11/1951	RE	02/05/2005	53	1.1	1.15	PK	0.2	4.95	None	None
14	Lisa Kaufman	M	23/11/1996	RE	23/11/2014	17	1.1	4.3	PK	0.2	4.36	None	None
15	Michael Stephens	M	06/10/1972	RE	21/07/2005	32	1.1	5.3	PK	0.1	3.05	None	None
16	Robert Hill	F	05/08/1982	RE	18/08/2008	26	1.1	4.53	DALK	0.1	3.58	None	None
17	Caleb Grant	F	28/02/1971	RE	28/10/2008	37	0.8	6.04	DALK	0.3	2.52	None	None
18	Shawn Mccarty	F	28/02/1956	RE	31/05/2006	50	1.3	4.11	PK	0.2	5.06	None	Endothelial Immune Rejection
19	Jennifer Callahan	M	13/11/1976	RE	12/04/2006	30	1.1	3.15	PK	0.0	4.65	None	None
20	Alexis Cortez	F	11/05/1954	LE	21/11/2008	54	0.7	4.8	PK	0.2	4.18	None	None
21	Daniel Robertson	M	05/11/1954	RE	21/07/2006	51	1.1	6.23	DALK	0.0	3.17	None	None
22	Susan Howe	F	22/09/1954	RE	11/06/2008	53	0.9	6.57	PK	0.3	2.42	None	None
23	Vanessa Moore	F	01/10/1952	RE	13/05/2010	57	0.8	2.26	DALK	0.2	2.8	None	None
24	Brittany Poole MD	M	15/10/1954	LE	31/07/2007	52	1.1	4.54	PK	0.2	2.88	None	None
25	Angela Cunningham	F	06/06/1991	RE	06/06/2009	18	0.9	2.31	PK	0.3	3.46	None	None
26	Joseph Robinson	M	07/10/1981	LE	02/10/2006	24	1	3.99	DALK	0	3.84	None	None
27	Jordan Reid	F	02/11/1983	LE	14/06/2005	21	1.4	8.85	DALK	0.4	3.24	None	None
28	Jordan Dunlap	M	01/04/1980	LE	05/03/2006	25	1.2	5.41	DALK	0.2	1.97	None	None
29	Dr. Susan Townsend	M	19/04/1981	RE	14/09/2006	22	1.4	5.13	PK	0	3.86	None	None
30	Rebecca Malone	M	03/03/1997	RE	03/03/2015	17	1.3	6.69	PK	0.4	4.38	None	None
31	Lindsay Evans	M	28/11/1983	LE	14/01/2010	26	1.1	7.15	DALK	0.3	3.54	None	None
32	Tony Barton	F	13/12/1992	RE	13/12/2010	17	1	3.79	DALK	0.0	2.51	None	None
33	Frances Butler	F	03/10/1980	RE	21/04/2010	29	1	7.99	DALK	0.1	2.22	None	None
34	Bryan Gibson	M	19/12/1980	LE	17/12/2008	27	0.7	5.99	DALK	0	1.2	None	None
35	Joe Morris	M	04/12/1960	LE	09/05/2010	49	0.9	1	DALK	0.2	1.83	None	Stromal Immune Rejection
36	Brittany Bond	F	18/03/1961	RE	06/10/2006	45	1.2	4.63	DALK	0.2	2.17	None	Stromal Immune Rejection
37	Pam Watson	F	31/07/1970	LE	01/10/2009	39	0.7	1.45	DALK	0.1	3.04	None	None
38	David Morgan	M	07/06/1975	LE	01/09/2005	30	1	2.89	PK	0.2	3.71	None	None
39	Sandra Johnson	M	07/08/1995	LE	07/05/2013	18	1.2	4.42	PK	0.2	3.9	None	Stromal Immune Rejection
40	Megan Larson	F	23/01/1952	RE	02/01/2007	54	0.9	7.59	PK	0.2	3.68	None	None