

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/374167032>

GPTZero Performance in Identifying Artificial Intelligence–Generated Medical Texts: A Preliminary Study

Article in *Journal of Korean medical science* · September 2023

DOI: 10.3346/jkms.2023.38.e319

CITATIONS

15

READS

334

1 author:



Farrokh Habibzadeh

221 PUBLICATIONS 2,890 CITATIONS

SEE PROFILE

Original Article
Medical informatics



GPTZero Performance in Identifying Artificial Intelligence-Generated Medical Texts: A Preliminary Study

Farrokh Habibzadeh

Past President, *World Association of Medical Editors (WAME)*
Editorial Consultant, *The Lancet*
Associate Editor, *Frontiers in Epidemiology*



Received: Jun 25, 2023

Accepted: Jul 19, 2023

Published online: Sep 14, 2023

Address for Correspondence:

Farrokh Habibzadeh, MD

Research and Development Unit, PIHO
Polyclinic, Eram Blvd., Shiraz 7143837877, Iran.
Email: Farrokh.Habibzadeh@gmail.com

© 2023 The Korean Academy of Medical Sciences.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iD

Farrokh Habibzadeh
<https://orcid.org/0000-0001-5360-2900>

Disclosure

The author has no potential conflicts of interest to disclose.

ABSTRACT

Background: With emergence of chatbots to help authors with scientific writings, editors should have tools to identify artificial intelligence-generated texts. GPTZero is among the first websites that has sought media attention claiming to differentiate machine-generated from human-written texts.

Methods: Using 20 text pieces generated by ChatGPT in response to arbitrary questions on various topics in medicine and 30 pieces chosen from previously published medical articles, the performance of GPTZero was assessed.

Results: GPTZero had a sensitivity of 0.65 (95% confidence interval, 0.41–0.85); specificity, 0.90 (0.73–0.98); accuracy, 0.80 (0.66–0.90); and positive and negative likelihood ratios, 6.5 (2.1–19.9) and 0.4 (0.2–0.7), respectively.

Conclusion: GPTZero has a low false-positive (classifying a human-written text as machine-generated) and a high false-negative rate (classifying a machine-generated text as human-written).

Keywords: Scientific Writing; Journalism; Artificial Intelligence; Classification; Sensitivity and Specificity

INTRODUCTION

A couple of months ago, I forecasted the emergence of artificial intelligence (AI) units and their role in scientific writing and journalism.¹ A few months later, AI-based units such as ChatGPT, have emerged and helped authors with scientific writing.^{2,3} Soon, articles were published with ChatGPT mentioned as their authors.³⁻⁶ With authorship comes responsibility,⁷ and chatbots, at least for the time being, are not intelligent enough for sure to shoulder this responsibility.¹ In response to this wave of using chatbots in scientific writing, several international bodies such as the World Association of Medical Editors (WAME) and the International Committee of Medical Journal Editors (ICMJE) have established statements to prescribe clear standards on using chatbots.^{7,8} The fifth recommendation of WAME states that editors of journals should be provided with appropriate means to help them with detection of AI-generated items (e.g., texts and graphs).⁸ GPTZero (available from <https://gptzero.me/>) is one of the first AI-based websites that has sought media attention, and a common tool now and then used by editors to detect AI-generated text.⁹ If editors are going

to use GPTZero to distinguish AI-generated text from human-written pieces, they should be aware of its performance to determine whether they can rely on its results or not.¹⁰ This preliminary study was conducted to determine the performance of GPTZero in identifying human from ChatGPT-generated texts.

METHODS

As there was no previous study on this issue, in a pilot study, 50 (an arbitrarily chosen value) paragraphs generated either by ChatGPT or human were examined by GPTZero. Twenty paragraphs were generated by the ChatGPT used by Neo-Bing®, using queries beginning with “write a paragraph about...” (Appendix 1). The queries were asked on June 24, 2023, between 6:30 and 6:50 GMT. The text generated were edited slightly—references and extra characters (such as asterisks), if any, were removed. The queries were chosen arbitrarily and cover a wide range of topics including diagnostic tests, immunity, coronavirus disease 2019, hydatid cyst, radiology, and cardiovascular disease, among other things (Appendix 1). Thirty pieces of text copied from articles published in 2022 or earlier, when the chatbots were not available, were also selected arbitrarily.¹¹⁻³⁷ Most of the articles were either mine or published in the *Journal of Korean Medical Science (JKMS)*—I am sure I did not make any misconduct, nor did I use any AI units for writing; *JKMS* has presumably acceptable editorial standards and has checked the quality of the copyediting process and the articles it has so far published. These 30 text pieces also covered various topics. The 50 pieces of texts were then fed to GPTZero (<https://gptzero.me/>) on June 24, 2023 between 11:30 and 12:15 GMT. Based on the information obtained, the sensitivity (*Se*), specificity (*Sp*), accuracy, and the positive and negative likelihood ratios (*LR*⁺ and *LR*[−], respectively), along with their 95% confidence intervals (CIs) were computed.³⁶⁻³⁸

RESULTS

Of 20 ChatGPT-generated text pieces, 13 could be correctly classified as AI-generated text, hence, a *Se* of 0.65 (95% CI, 0.41–0.85, **Table 1**). GPTZero could correctly classify 27 text pieces written by human, translating into a *Sp* of 0.90 (95% CI, 0.73–0.98).³⁷ GPTZero had an accuracy of 0.80 (95% CI, 0.66–0.90). Having the *Se* and *Sp*, then the *LR*⁺ is³⁶:

$$LR^+ = \frac{Se}{1 - Sp} = \frac{0.65}{1 - 0.90} = 6.5 \text{ (95\% CI, 2.1 - 19.9) (Eq. 1)}$$

and the *LR*[−] is³⁶:

$$LR^- = \frac{1 - Se}{Sp} = \frac{1 - 0.65}{0.90} \approx 0.4 \text{ (95\% CI, 0.2 - 0.7) (Eq. 2)}$$

Table 1. Classification of human and AI-generated texts by GPTZero

GPTZero	Source		Total
	AI	Human	
AI	13	3	16
Human	7	27	27
Total	20	30	50

AI = artificial intelligence.

DISCUSSION

In identifying AI-generated texts, GPTZero had an accuracy of 80%. Although it had an almost acceptable Sp of 0.90, its Se of 0.65 can be considered low to mediocre; many false-negative instances (AI-generated texts mistaken as human writings) may occur. GPTZero classifies a piece of text based on its perplexity score (a measure of text randomness; a lower score is in favor of AI-generated text) and burstiness score (a measure of the variation in perplexity).^{9,39} Given the preliminary results obtained in the current study, the high false-negative rate of 0.35 ($1 - Se$), more than one-third of AI-generated texts are mistakenly classified as human writings. This necessitates employing a better algorithm or changing the current cut-off values for the scores used. For instance, if the perplexity cut-off score is set to a lower value, the Se will increase and lower false-negative results will ensue.³⁷ But, it happens at the cost of a lower Sp .

The results of this study should be interpreted in light of the current evidence and the prevalence of using AI-generated text, which is expectedly different from place to place and from time to time. It is expected to be increased with improvements made in AI units. If the prior probability of using AI-generated text is 20% (and odds of 0.25), then if GPTZero indicates that the text is likely written by an AI (a positive result), the posterior odds of having an AI-generated text is $1.63 = 6.5 \times 0.25$ ($LR^+ \times$ the prior odds), translating into a posterior probability of 62%.³⁶ On the other hand, if GPTZero indicates that the text is written by human, the posterior odds decreases to $0.1 = 0.4 \times 0.25$ ($LR^- \times$ the prior odds), translating into a posterior probability of 9%. Like every highly specific test, GPTZero with its low false-positive rate, is more appropriate for ruling in rather than ruling out suspicious texts—if GPTZero classifies a text as AI-generated (a positive result), the result is very likely to be correct as it is unlikely to be false-positive. Nonetheless, the likelihood that a negative result (stating that the text is written by human) be false-negative is not low.³⁷

Over the time, chatbots will improve and a time comes when the AI will communicate and write better than humans (at least in terms of not making typographical and grammatical mistakes).¹ Thereafter, the only score that could help us to discriminate a human from an AI writing would be its perfectness, rather than its perplexity or other similar scores, I believe. In fact, the perfectness of the text would then not be helpful either, as the AI-based unit will learn how to intentionally incorporate typographical and grammatical mistakes in the text generated to imitate the human being, if they WANT!¹

GPTZero has a low (10%) false-positive (classifying a human-written text as AI-generated) and a high (35%) false-negative rate (classifying an AI-generated text as human-written). Further studies are needed to determine the performance indices of GPTZero and similar software programs more precisely on larger samples and in various scientific disciplines (e.g., cardiology, virology, astrophysics, computer sciences, etc.).

REFERENCES

1. Habibzadeh F. The future of scientific journals: the rise of UniAI. *Learn Publ* 2023;36(2):326-30. [CROSSREF](#)
2. Doskaliuk B, Zimba O. Beyond the keyboard: academic writing in the era of ChatGPT. *J Korean Med Sci* 2023;38(26):e207. [PUBMED](#) | [CROSSREF](#)

3. Benichou L, ChatGPT . Rôle de l'utilisation de l'intelligence artificielle ChatGPT dans la rédaction des articles scientifiques médicaux The Role of Using ChatGPT AI in Writing Medical Scientific Articles. *J Stomatol Oral Maxillofac Surg* 2023;101456.
[PUBMED](#) | [CROSSREF](#)
4. Curtis N, ChatGPT . To ChatGPT or not to ChatGPT? The impact of artificial intelligence on academic publishing. *Pediatr Infect Dis J* 2023;42(4):275.
[PUBMED](#) | [CROSSREF](#)
5. King MR, ChatGPT . A Conversation on artificial intelligence, chatbots, and plagiarism in higher education. *Cell Mol Bioeng* 2023;16(1):1-2.
[PUBMED](#) | [CROSSREF](#)
6. ChatGPT Generative Pre-trained Transformer, Zhavoronkov A. Rapamycin in the context of Pascal's Wager: generative pre-trained transformer perspective. *Oncoscience* 2022;9:82-4.
[PUBMED](#) | [CROSSREF](#)
7. International Committee of Medical Journal Editors (ICMJE). Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. <https://www.icmje.org/icmje-recommendations.pdf>. Updated 2023. Accessed June 12, 2023.
8. Zielinski C, Winker MA, Aggarwal R, Ferris LE, Heinemann M, Lapeña JF, et al. Chatbots, generative AI, and scholarly manuscripts. WAME recommendations on chatbots and generative artificial intelligence in relation to scholarly publications. <https://wame.org/page3.php?id=106>. Updated 2023. Accessed June 2, 2023.
9. Kirmani AR. Fantastic (AI) bots and how to catch them. *MRS Bull* 2023;48(4):314-8.
[CROSSREF](#)
10. Habibzadeh F. The acceptable text similarity level in manuscripts submitted to scientific journals. *J Korean Med Sci* 2023;38(31):e240.
[PUBMED](#) | [CROSSREF](#)
11. Bekarysova D, Joshi M, Gupta L, Yessirkepov M, Gupta P, Zimba O, et al. Knowledge and perceptions of reactive arthritis diagnosis and management among healthcare workers during the COVID-19 pandemic: online survey. *J Korean Med Sci* 2022;37(50):e355.
[PUBMED](#) | [CROSSREF](#)
12. Casagrande A, Fabris F, Girometti R. Fifty years of Shannon information theory in assessing the accuracy and agreement of diagnostic tests. *Med Biol Eng Comput* 2022;60(4):941-55.
[PUBMED](#) | [CROSSREF](#)
13. Choi JW, Kim HJ, Kim JB, Lee S, Lim C, Chang BC, et al. Early and two-year outcomes after sutureless and conventional aortic valve replacement: a nationwide population-based study. *J Korean Med Sci* 2021;36(9):e57.
[PUBMED](#) | [CROSSREF](#)
14. Chumakov K, Benn CS, Aaby P, Kottiril S, Gallo R. Can existing live vaccines prevent COVID-19? *Science* 2020;368(6496):1187-8.
[PUBMED](#) | [CROSSREF](#)
15. Gasparyan AY, Ayvazyan L, Mukanova U, Yessirkepov M, Kitas GD. Scientific hypotheses: writing, promoting, and predicting implications. *J Korean Med Sci* 2019;34(45):e300.
[PUBMED](#) | [CROSSREF](#)
16. Gasparyan AY, Kumar AB, Yessirkepov M, Zimba O, Nurmashev B, Kitas GD. Global health strategies in the face of the COVID-19 pandemic and other unprecedented threats. *J Korean Med Sci* 2022;37(22):e174.
[PUBMED](#) | [CROSSREF](#)
17. Gupta L, Gasparyan AY, Misra DP, Agarwal V, Zimba O, Yessirkepov M. Information and misinformation on COVID-19: a cross-sectional survey study. *J Korean Med Sci* 2020;35(27):e256.
[PUBMED](#) | [CROSSREF](#)
18. Habibzadeh F. A bird's eye view of science publishing and editing in Iran. *Eur Sci Ed* 2006;32(4):98-100.
19. Habibzadeh F. Statistical data editing in scientific articles. *J Korean Med Sci* 2017;32(7):1072-6.
[PUBMED](#) | [CROSSREF](#)
20. Habibzadeh F. Open access journals in the Middle East and Iran. *J Korean Med Sci* 2019;34(16):e123.
[PUBMED](#) | [CROSSREF](#)
21. Habibzadeh F, Winker M. Duplicate publication and plagiarism: causes and cures. *Notf Retimed* 2009;12(6):415-8.
[CROSSREF](#)
22. Habibzadeh F, Chumakov K, Sajadi MM, Yadollahie M, Stafford K, Simi A, et al. Use of oral polio vaccine and the incidence of COVID-19 in the world. *PLoS One* 2022;17(3):e0265562.
[PUBMED](#) | [CROSSREF](#)

23. Habibzadeh F, Habibzadeh P, Shakibafard A, Saidi F. Predicting the outcome of asymptomatic univesicular liver hydatids: diagnostic accuracy of unenhanced CT. *Eur Radiol* 2021;31(8):5812-7.
[PUBMED](#) | [CROSSREF](#)
24. Habibzadeh F, Habibzadeh P, Yadollahie M. The apparent prevalence, the true prevalence. *Biochem Med (Zagreb)* 2022;32(2):020101.
[PUBMED](#) | [CROSSREF](#)
25. Habibzadeh F, Habibzadeh P, Yadollahie M, Roozbehi H. On the information hidden in a classifier distribution. *Sci Rep* 2021;11(1):917.
[PUBMED](#) | [CROSSREF](#)
26. Habibzadeh F, Sajadi MM, Chumakov K, Yadollahie M, Kottitil S, Simi A, et al. COVID-19 infection among women in Iran exposed vs unexposed to children who received attenuated poliovirus used in oral polio vaccine. *JAMA Netw Open* 2021;4(11):e2135044.
[PUBMED](#) | [CROSSREF](#)
27. Habibzadeh F, Yadollahie M, Simi A. Use of oral polio vaccine and the global incidence of mother-to-child human immunodeficiency virus transmission. *Front Public Health* 2022;10:878298.
[PUBMED](#) | [CROSSREF](#)
28. Habibzadeh P, Dastsooz H, Eshraghi M, Łos MJ, Klionsky DJ, Ghavami S. Autophagy: the potential link between SARS-CoV-2 and cancer. *Cancers (Basel)* 2021;13(22):5721.
[PUBMED](#) | [CROSSREF](#)
29. Habibzadeh P, Mofatteh M, Silawi M, Ghavami S, Faghihi MA. Molecular diagnostic assays for COVID-19: an overview. *Crit Rev Clin Lab Sci* 2021;58(6):385-98.
[PUBMED](#) | [CROSSREF](#)
30. Kim YK, Minn D, Song DY, Lee CH, Ryoo NH, Jeon CH, et al. Prevalence of SARS-CoV-2 antibody in 2,935 healthcare workers at 6 major hospitals, Daegu, Korea. *J Korean Med Sci* 2021;36(43):e294.
[PUBMED](#) | [CROSSREF](#)
31. Krakowska A, Cedzyński M, Wosiak A, Swiechowski R, Krygier A, Tkaczyk M, et al. Toll-like receptor (TLR2, TLR4) polymorphisms and their influence on the incidence of urinary tract infections in children with and without urinary tract malformation. *Cent Eur J Immunol* 2022;47(3):260-6.
[PUBMED](#) | [CROSSREF](#)
32. Malek-Hosseini SA, Habibzadeh F, Nikeghbalian S. Shiraz organ transplant center: the largest liver transplant center in the world. *Transplantation* 2019;103(8):1523-5.
[PUBMED](#) | [CROSSREF](#)
33. Saidi F, Habibzadeh F. The non-operative management of asymptomatic liver hydatids: ending echinococcosis. *J Gastrointest Surg* 2018;22(3):486-95.
[PUBMED](#) | [CROSSREF](#)
34. Singh AV, Yadav VS, Chauhan DS, Singh SV. *Mycobacterium bovis* induced human tuberculosis in India: current status, challenges & opportunities. *Indian J Med Res* 2022;156(1):21-30.
[PUBMED](#) | [CROSSREF](#)
35. Vessal K, Habibzadeh F. Rules of the game of scientific writing: fair play and plagiarism. *Lancet* 2007;369(9562):641.
[PUBMED](#) | [CROSSREF](#)
36. Habibzadeh F, Habibzadeh P. The likelihood ratio and its graphical representation. *Biochem Med (Zagreb)* 2019;29(2):020101.
[PUBMED](#) | [CROSSREF](#)
37. Habibzadeh F, Habibzadeh P, Yadollahie M. On determining the most appropriate test cut-off value: the case of tests with continuous results. *Biochem Med (Zagreb)* 2016;26(3):297-307.
[PUBMED](#) | [CROSSREF](#)
38. MedCalc Software Ltd. Diagnostic test evaluation calculator: version 22.007. https://www.medcalc.org/calc/diagnostic_test.php. Updated 2023. Accessed June 24, 2023.
39. Ahmad N, Murugesan S, Kshetri N. Generative artificial intelligence and the education sector. *Computer* 2023;56(6):72-6.
[CROSSREF](#)

Appendix 1

Text samples used to test the GPTZero performance

Twenty queries asked from ChatGPT, its outputs (bullets) on June 24, 2023, 6:30–6:50 GMT, along with the results of GPTZero (<https://gptzero.me/>) classification examined on the same day between 11:30 and 12:15 GMT (dashes):

- 1) Write a paragraph about the role of toll-like receptors in the innate immunity.
 - Toll-like receptors (TLRs) are well-defined pattern recognition receptors responsible for pathogen recognition and induction of innate immune responses. They play an important role in recognition of viral particles and activation of the innate immune system. Activation of TLR pathways leads to secretion of pro-inflammatory cytokines, such as interleukin-1 (IL-1), IL-6, and tumor necrosis factor- α , as well as type 1 interferon. TLRs serve a central role in innate immunity, but they can also modulate cell function in various non-immune cell types including endothelial cells.
 - GPTZero: *Artificial intelligence (AI)-generated*
- 2) Write a paragraph about the history of tuberculosis in the Indian subcontinent.
 - Tuberculosis has existed in India for several thousand years. The disease was first recognized through a resolution passed in the All India Sanitary Conference held at Madras in 1912. The description of tuberculosis was initially found in India and China as early as 3300 and 2300 years ago, respectively. The TB picture started becoming clear with the introduction of tuberculin testing. India's response to TB has changed with time and with the increasing sophistication of technology. Responses to TB have evolved, from pre-independence through post-independence to the current WHO-assisted period.
 - GPTZero: *Human-written*
- 3) Write a paragraph about application of Cox regression model in survival analysis.
 - The Cox proportional hazards regression model is a statistical method used to analyze survival data. It is used to model the relationship between survival time and one or more predictor variables. The model estimates the hazard ratio (HR) of a given endpoint associated with a specific risk factor, which can be either a continuous variable like age and C-reactive protein level or a categorical variable like gender and treatment group. The Cox regression method is considered as an example of semi-parametric models. It is commonly used in medical research for investigating the association between the survival time of patients and one or more predictor variables.
 - GPTZero: *AI-generated*
- 4) Write a paragraph about the role of proinflammatory cytokines in pathogenesis of SARS-CoV-2.
 - SARS-CoV-2 infects the respiratory epithelial tissue and activates local innate immune cells to release inflammatory cytokines such as IL-1, IL-6, IL-8, IL-12, TNF- α , and other chemokines. Since proinflammatory cytokines are elevated in severe COVID-19 patients, SARS-CoV-2 seems to activate NF- κ B and produces proinflammatory cytokines, which is correlated with COVID-19 pathogenesis. Immune responses to SARS-CoV infection are initiated by the innate immune system, which recognizes pathogens and induces proinflammatory cytokines to trigger the immune response. The immune response is followed by responses of the adaptive immune system consisting of T cells that can directly kill virus-infected cells and B cells that produce pathogen-specific antibodies.
 - GPTZero: *Human-written*
- 5) Write a paragraph about the prognostic value of pericyst degenerative changes in solitary hepatic cystic echinococcosis.
 - According to a study published on PubMed, the presence of laminated membrane detachment and/or pericyst degenerative changes is associated with a favorable outcome in asymptomatic univesicular liver hydatids. However, their absence does not necessarily indicate an unfavorable outcome. Hepatic cystic echinococcosis (CE) is composed of two derived layers of membrane: an inner, nucleated, germinal membrane, and an outer, acellular, laminated layer.
 - GPTZero: *Human-written*
- 6) Write a paragraph about biases exist in test-negative case-control study design in determining vaccine effectiveness.
 - The test-negative case-control study design is a popular approach for estimating vaccine effectiveness (VE) due to its efficiency. However, other biases will occur. In comparing the effectiveness of vaccines, the confounding effects of health care-seeking behavior are reduced. But other biases such as selection bias, information bias, and confounding by indication may still exist.
 - GPTZero: *Human-written*
- 7) Write a paragraph about common mutations in patients with beta-thalassemia in Fars province, Iran.
 - Beta-thalassemia is a genetic disorder that is prevalent in Iran. According to a study, the most common mutation of β -Thalassemia among the western parts of the Mediterranean Sea such as Portugal, Spain, and France is Codon 39. This mutation

is also the most frequent in Queshm Island, located south of Iran in the Persian Gulf. Another study conducted in Hamadan province, Iran showed that β -thal caused by mutations on the HBB gene is the most common single-gene disorder in the world. The HBB gene mutation was investigated in 41 patients referred to a referral hospital. A third study showed that IVS-I-5 (G > C) was the most common mutation by far accounting for about 80% of all detected β -thal alleles in this region.

– GPTZero: *AI-generated*

8) Write a paragraph about maximizing the likelihood ratio and its application in assessing the performance of a diagnostic test.

- The likelihood ratio (LR) is a semiquantitative measure of the performance of diagnostic tests which indicates how much a diagnostic procedure modifies the probability of disease, and is calculated from the sensitivity and specificity of the test (or directly from the change in probability associated with the test result). The LR can be used to calculate post-test probabilities of disease for patients with positive or negative test results. The LR can be maximized by choosing an appropriate cut-off value for the test result.

– GPTZero: *Human-written*

9) Write a paragraph about application of Shannon's information theory and entropy in assessing a diagnostic test performance.

- Shannon's information theory is a mathematical theory that describes the fundamental laws of data compression and error correction over a noisy channel. It has found a wide range of applications in several areas where information plays a key role, which goes well beyond the original scopes for which they have been conceived, namely data compression and error correction over a noisy channel. One of these applications is assessing the performance of diagnostic tests. Entropy is used to measure the amount of uncertainty or randomness in the data. In the context of diagnostic tests, entropy can be used to measure the amount of uncertainty or randomness in the test results. The entropy can be used to calculate the sensitivity and specificity of the test, which are measures of how well the test can detect true positive and true negative cases.

– GPTZero: *AI-generated*

10) Write a paragraph about clinical value of administrating glucocorticoids in controlling the cytokine storm in patients with severe COVID-19.

- Glucocorticoids have been shown to be effective in controlling the cytokine storm in patients with severe COVID-19. According to a study published in PubMed, glucocorticoids could modulate immune cells, reduce cytokine and chemokine, and improve endothelial functions in patients with severe COVID-19. The study also suggests that benefits of glucocorticoids have been observed in multiple clinical trials, but the timing, dosage and duration vary across studies.

– GPTZero: *Human-written*

11) Write a paragraph about the role of histamine and heparin release from basophils in the pathogenesis of aspergillosis in an immunocompromised host.

- Aspergillosis is a fungal infection caused by *Aspergillus* species. In immunocompromised hosts, basophils play an important role in the pathogenesis of aspergillosis. Basophils release histamine and heparin when activated. Histamine enlarges blood vessels to improve blood flow and heal the affected area. Histamine also opens pathways for other cells in the immune system to quickly target and respond to the allergen. Heparin prevents blood from clotting too quickly. The release of histamine and heparin from basophils can cause bronchoconstriction, mucus secretion, and airway inflammation. These effects can lead to respiratory distress in immunocompromised patients with aspergillosis.

– GPTZero: *AI-generated*

12) Write a paragraph about the prevalence of valvular heart disease in China.

- According to a study published in BMC Cardiovascular Disorders, the prevalence of valvular heart disease (VHD) in China was 3.8% with an estimated 25 million patients. The prevalence of VHD increased with age and was higher in participants with hypertension or chronic kidney disease than in their counterparts. Among participants with VHD, 55.1% were rheumatic and 21.3% were degenerative.

– GPTZero: *Human-written*

13) Write a paragraph about cardiovascular complications of COVID-19.

- COVID-19 can result in systemic inflammation, multiorgan dysfunction, and critical illness. The cardiovascular system is also affected, with complications including myocardial injury, myocarditis, acute myocardial infarction, heart failure, dysrhythmias, and venous thromboembolic events. The most common cardiovascular complications in COVID-19 patients were myocardial injury (21.2%) and arrhythmia (15.3%), followed by heart failure (14.4%) and acute coronary syndrome (1.0%). The underlying pathophysiology of COVID-19-associated cardiovascular complications is not fully understood, although direct viral infection

of myocardium and cytokine storm have been suggested as possible mechanisms of myocarditis. Although respiratory failure is the primary cause of death, cardiovascular complications such as acute myocardial injury and myocarditis, cardiac fibrosis, arrhythmias, endothelial dysfunction, dysautonomia, and thrombotic events may also contribute to overall morbidity and mortality of COVID-19 patients.

– GPTZero: *AI-generated*

14) Write a paragraph about liver transplantation in the Middle East.

- Liver transplantation is a well-established procedure in the Middle East. Countries such as Egypt, Iran, Saudi Arabia, and Turkey are pioneers of liver transplantation in the region. Despite political conflicts, these countries have collaborated in the Middle East Society of Organ Transplantation (MESOT) to develop a platform for promoting transplantation and specifically liver transplantation in the Middle East. The Organ Transplant Centre (OTC) at King Faisal Specialist Hospital & Research Centre is the most comprehensive and advanced facility for multi-organ transplantation in the Middle East.

– GPTZero: *AI-generated*

15) Write a paragraph about tensor diffusion imaging and its application in diagnosis of cerebrovascular accidents.

- Tensor diffusion imaging (DTI) is a type of magnetic resonance imaging (MRI) that measures the motion of water molecules in different tissues, especially in the white matter of the brain. DTI can show the direction and degree of water diffusion, which reflects the structure and organization of axons, the nerve fibers that connect brain cells. DTI can help detect and quantify damage to the white matter caused by neurological diseases or injuries, such as stroke or traumatic brain injury. DTI has been used in clinical and research settings since the 1980s.

– GPTZero: *AI-generated*

16) Write a paragraph about application of Gaussian mixture models in determining diagnostic tests performance indices.

- Gaussian mixture models (GMM) are a flexible class of models for density estimation. They are used in many applications such as clustering, outlier detection, and image segmentation. In the context of diagnostic tests performance indices, GMM can be used to estimate the sensitivity and specificity of a diagnostic test. The sensitivity and specificity are performance indices that measure the ability of a diagnostic test to correctly identify positive and negative cases. GMM can be used to model the distribution of test results for positive and negative cases separately. The sensitivity and specificity can then be estimated from the overlap between these distributions.

– GPTZero: *AI-generated*

17) Write a paragraph about the general picture of juvenile rheumatoid arthritis in the USA.

- Juvenile rheumatoid arthritis (JRA) is a form of arthritis that affects children below 16 years. It is an autoimmune condition that causes joint inflammation (swelling) and stiffness for more than six weeks in children below 16 years. The exact cause is not known, but the two factors that trigger the condition include hereditary genetic predisposition and environmental factors such as viruses that trigger the development. Some of the most common symptoms include pain in the joints, swelling, stiffness, fatigue, poor appetite, intermittent fever, eye problems, rashes, and anemia. Girls are more prone to have juvenile rheumatoid arthritis. After a prolonged period, complications may include long-term pain, joint deformity, stunted growth, anemia, and vision problems.

– GPTZero: *AI-generated*

18) Write a paragraph about normal range of the prostate-specific antigen in Japanese men.

- The normal range of prostate-specific antigen (PSA) in Japanese men is 0-4 ng/mL. However, there is no specific cutoff point between a normal and an abnormal PSA level. Your doctor might recommend a prostate biopsy based on the results of your PSA test.

– GPTZero: *AI-generated*

19) Write a paragraph about how malaria parasite stimulates the immune system.

- Malaria parasites stimulate the immune system by releasing molecules that activate the immune cells. The immune cells then produce cytokines that help to control the infection. The parasite also stimulates the production of antibodies that can help to clear the parasite from the blood. However, some studies indicate that Plasmodium parasites inhibit normal dendritic cell maturation.

– GPTZero: *AI-generated*

20) Write a paragraph about the most appropriate regression analysis model to be used when the dependent variable is a whole number and the independent variables are sex and the ambient temperature.

- When the dependent variable is a whole number and the independent variables are sex and the ambient temperature, the most

appropriate regression analysis model to be used is Poisson regression. Poisson regression is used when the dependent variable is a count variable (i.e., a whole number) and the independent variables are categorical or continuous. It is used to model count data that follow a Poisson distribution. The Poisson distribution is used to model count data that have a mean equal to its variance. In this case, the dependent variable is a count variable (i.e., a whole number) and the independent variables are categorical (sex) and continuous (ambient temperature). Therefore, Poisson regression is the most appropriate regression analysis model to be used in this case.

– GPTZero: *AI-generated*

Thirty arbitrary-chosen text pieces copied from different parts of articles published in 2022 or earlier, when no AI-based unit was available to help authors with writing:

- 21) "Toll-like receptors (TLRs) contribute to the innate immune system. They are an element of non-specific immunity, which enables organisms to react quickly to foreign antigens, without being previously exposed to them. TLRs are pattern recognition receptors. TLR gene polymorphisms are widely investigated in connection with various infections."³¹
– GPTZero: *Human-written*
- 22) "Coronavirus disease 2019 (COVID-19) has affected more than 228 million people worldwide; the death toll exceeded 4.6 million on September 19, 2021. Currently only a handful of specific vaccines with various degrees of effectiveness have received authorization and are used for immunization in several countries. However, limitations exist in the production and worldwide distribution of the specific vaccines, especially in resource-limited settings. Even for those vaccines with acceptable efficacies, additional studies are needed to assess their long-term effectiveness and safety profile."²⁶
– GPTZero: *Human-written*
- 23) "The level of urbanization and the population density mostly depend on the city of residence; the mothers in the exposed group were matched with unexposed women in the same city. We did not assess whether the residence neighborhoods of the participants in the studied cities were substantially different. However, we matched 9 women in the unexposed group to each mother in the exposed group according to several confounders; it is unlikely that most of these matched women came from a part of the city with very different incidence rates compared with the mother in the exposed group."²⁶
– GPTZero: *Human-written*
- 24) "The effects of lockdowns and international travel bans were hard to account for in our study since policies changed with time. However, the mean stringency index, an index reflecting the level of restrictions imposed in a given country, was not significantly different between countries using OPV vs those using IPV. The population density was also not significantly different between the two groups of countries. None of these variables was thus taken into account in the model."²²
– GPTZero: *Human-written*
- 25) "Generally speaking, the likelihood ratio indicates how many times more (or less) likely a certain condition for a test result is expected to be observed in diseased, compared with non-diseased, people. Four general possible conditions include likelihood ratio for a certain test value, likelihood ratio for a positive or negative test, and likelihood ratio for a range of test values."³⁶
– GPTZero: *Human-written*
- 26) "Classification tasks are a common challenge to every field of science. We often need to categorize a new observation into one of predefined groups based on its attributes. The definitions provided for a classifier varies a little bit from field to field. In statistics, a classifier is an algorithm that help us with data categorization. In machine learning, the function of a classifier is to map objects based on their features to classes. In medicine, a classifier can be considered a diagnostic test helping physicians to classify people to healthy and diseased groups. Classifiers are extensively used in various disciplines of science."²⁵
– GPTZero: *Human-written*
- 27) "Mean and SD are reported to present the center and dispersion of normally distributed data. For non-normally distributed data median and IQR should be reported. If distribution of variables is tested, either mean (SD) or median (IQR) is presented. Sometimes information about distribution of parameters is missing. Editors without access to raw data are unable to check the normality. They should, however, know when SD exceeds half of the corresponding mean, it is unlikely that the data follow normal distribution."¹⁹
– GPTZero: *Human-written*
- 28) "Diagnostic tests are important clinical tools. If that is possible, we have to use gold-standard tests for the diagnosis of diseases. However, a gold-standard test either does not exist or is very difficult or expensive to perform for certain disease conditions. There-

fore, we have to use alternative diagnostic tests as surrogates for gold-standard tests.”³⁷

– GPTZero: *AI-generated*

- 29) “Employing a Bayesian approach, the post-test (posterior) probability of a disease depends on the pre-test probability of the disease and the test result. The post-test probability of a disease after the patient is tested can however be considered the pre-test probability of the next test to be done. Based on what has been presented, the cut-off value of the second test should be different for two patients suspicious for the same disease but having different results on their first test, hence different post-test probabilities.”³⁷

– GPTZero: *Human-written*

- 30) “Many predatory journals have so far been launched and published in the Middle East and Iran. Sometimes, it is hard for an author to distinguish between a legitimate and a predatory journal. Being under pressure of publishing their articles by the national rules set for fulfillment of graduation or career promotion, some of the postgraduate students and even university faculty members happily pay the APC and publish their articles in predatory journals. Often, they even do not know that the journal is predatory. Worse, from time to time the accreditation committee evaluating the credits for promotion of the authors does not aware that the journal is predatory too and approves the articles.”²⁰

– GPTZero: *Human-written*

- 31) “The coronavirus disease 2019 (COVID-19) pandemic and consequent need for boosting immunity and following hygiene and social distancing regulations to mitigate infection risks have led researchers, clinicians and journal editors around the world to search for rational hypotheses, ideas, and emerging evidence for prevention and management of this new disease. While established scholarly platforms continue publishing and disseminating peer-reviewed and validated items on the new coronavirus, numerous online channels with variable quality checks are emerging and supplying their users with statements, instructions, and recommendations on various aspects of COVID-19.”¹⁷

– GPTZero: *Human-written*

- 32) “Current research activities are diversifying to combine scientific observations with analysis of facts recorded by scholars from various professional backgrounds. Citation analyses and networking on social media are also becoming essential for shaping research and publishing strategies globally. Learning specifics of increasingly interdisciplinary research studies and acquiring information facilitation skills aid researchers in formulating innovative ideas and predicting developments in interrelated scientific fields.”¹⁵

– GPTZero: *Human-written*

- 33) “It is a widespread belief that all liver hydatids, even asymptomatic cysts, should be operated upon pre-emptively to avert any impending complications. It has been shown that the presence of LMD or PDCs is associated with a favorable outcome of the cyst—the cyst is unlikely to grow or being complicated.”²³

– GPTZero: *Human-written*

- 34) “Secondary publication, in which an author obtains permission from the original publisher to submit the work elsewhere, informs the editor at the new journal of the previous publication and, if accepted as a secondary publication, indicates in the article the existence of the previous publication; this is legitimate and acknowledged by the International Committee of Medical Journal Editors (ICMJE). Such secondary publication or simultaneous publication has been used to help reach more readers (as in the case of the ICMJE guidelines) and more audiences in different languages.”²¹

– GPTZero: *Human-written*

- 35) “Iran has allocated a larger budget to its scientific research sector and the number of graduates and assistant professors has increased significantly over recent years. Junior professors are required to publish scientific articles in recognized journals to obtain academic career promotion. Finally, postgraduate students are obliged to publish their research theses in order to graduate.”¹⁸

– GPTZero: *Human-written*

- 36) “The type of polio vaccine used in a given country strongly depends on the HDI—countries with higher HDI (commonly, high-income countries) prefer and can afford to use IPV, which is significantly more expensive than OPV. An increase of 0.1 unit in HDI was independently associated with a higher reduction in MTCT rate in countries using IPV only compared with those using OPV (61 vs. 22%, respectively).”²⁷

– GPTZero: *Human-written*

- 37) “Currently, two major categories of diagnostic assays are commercially available for diagnosing SARS-CoV-2. The first group of assays identifies the viral RNA using molecular techniques that are based mostly on polymerase chain reaction (PCR) or nucleic

acid hybridization. The second group are immunological assays that detect either antibodies that are produced in response to the infection or antigenic proteins. Laboratory-based SARS-CoV-2 molecular assays are currently the reference standard for the diagnosis of this infection.”²⁹

– GPTZero: *Human-written*

- 38) “Autophagy plays a prominent role in maintaining cellular homeostasis through the removal of damaged organelles, abnormal proteins, and invading organisms. Defects in autophagy are associated with various pathological conditions, including cancer. It can lead to accumulation of damaged mitochondria and alter cellular metabolism, leading to a high oxidative state. Furthermore, impairments in autophagy flux can lead to ER stress and subsequent accumulation of chaperone proteins and an eventual rise in the unfolded protein burden.”²⁸

– GPTZero: *AI-generated*

- 39) “The current pandemic of coronavirus disease 2019 (COVID-19) has set up new challenges in the management of persons with chronic diseases such as rheumatological disorders. Various registries and surveys have helped provide real-world data on patients with rheumatic diseases. Analysis of data from electronic record databases and other registries has shown that COVID-19 outcomes are usually poorer in patients with rheumatic diseases.”¹¹

– GPTZero: *Human-written*

- 40) “The principles of health justice which were initially adopted in the Alma-Ata Declaration, implied multi-sectoral cooperation, community involvement, employment of modernized technologies, and universal health services’ coverage. While the rationale for global action in this field is justifiable from the social point of view, uncertainties and discrepancies of related evidence-based studies have long frustrated healthcare decision-makers.”¹⁶

– GPTZero: *Human-written*

- 41) “Tuberculosis (TB) continues to be an important disease both for humans and animals since the origin of human civilization. Evidence of TB exists in 3000 year old Egyptian mummies and about 17,000 year old fossilized bison. TB affects both humans and animals (domestic and wild ruminants), and *Mycobacterium tuberculosis* is the primary cause of human TB. *M. tuberculosis* survives under extreme adverse conditions in the host, infects nearly one-third human population and is responsible for about 1.6 million deaths per year (including 0.3 million deaths in HIV-positive patients).”³⁴

– GPTZero: *Human-written*

- 42) “Focusing on accuracy measured for diagnostic tests and the agreement between raters, we have to note that the disease is a hidden and objective status of the patient and the physician makes assumptions on it by interpreting the result of the test. Thus, the test is a means to extract information from the patient to diagnose the disease. Therefore, the most accurate diagnostic test will be the one that can extract as much information as possible: the more knowledge flows from the disease to the reader, the more accurate the diagnostic test. The information is implied at the beginning of the diagnostic process and plays a fundamental role.”¹²

– GPTZero: *AI-generated*

- 43) “Serologic tests are commonly used in seroepidemiologic and prevalence studies. The design is typically conducted to understand the current situation of a condition of interest, say a disease. For example, over the past two years, soon after the announcement of the coronavirus disease pandemic, many serologic tests have been developed for diagnosis of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2); numerous seroepidemiologic studies have been conducted to determine the prevalence of the disease in various parts of the world.”²⁴

– GPTZero: *Human-written*

- 44) “Organ transplantation in Iran started in 1967 with the first successful live donor kidney transplant performed at Namazi Hospital in Shiraz.1 Subsequently, transplantation progressed, although at a slow pace with only a handful of procedures done until mid-1970. Some patients who could afford the expenses went abroad, mainly to the United Kingdom, to receive living-related transplants; outcomes were usually poor.”³²

– GPTZero: *Human-written*

- 45) “Serological tests are very helpful for sero-epidemiological studies. Coronavirus disease 2019 (COVID-19) is known to induce robust antibody response in most of the affected individuals. The antibody concentrations could serve as an important laboratory index with prognostic implications for patients recovering from COVID-19. In spite of their limitations, serologic-based assays are currently the best available method to document past infections. Furthermore, they can be used to gauge individual’s immune response to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) vaccines.”²⁴

– GPTZero: *Human-written*

- 46) "Plagiarism of scientific text, as defined by the use of others' word-for-word phraseology in a manuscript to convey the desired implication, might thus warrant reappraisal in articles originating from a non-English speaker in a developing country with limited access to professional editorial assistance."³⁵
– GPTZero: *Human-written*
- 47) "Although a large patient population was included and propensity score matching was performed to overcome the retrospective nature of the study, unadjusted confounders could affect study results. Second, there might be some underestimation or overestimation of co-morbidities and clinical outcomes because most of preoperative variables and clinical outcomes were drawn on the basis of diagnosis codes or procedure codes."¹³
– GPTZero: *Human-written*
- 48) "The first community outbreak occurred in Daegu, starting with patient 31 on February 18, and soon spread across the country. With a maximum of 741 new patients per day in the city as of February 29, 2020, the cumulative number of patients had surpassed 6,000 by the end of March 2020."³⁰
– GPTZero: *Human-written*
- 49) "Cystic echinococcosis is common in many countries and involves the liver in 75% of cases. The current belief is that all incidentally discovered asymptomatic liver hydatids must be subjected to some types of interventional treatments pre-emptively for fear of some presumably impending complications. We conducted this study to assess the prognosis of patients with liver hydatids without any surgical interventions."³³
– GPTZero: *Human-written*
- 50) "Prophylactic vaccination is the most effective intervention to protect against infectious diseases. The commonly accepted paradigm is that immunization with both attenuated virus (live but with substantially reduced virulence) and inactivated (killed virus particles) vaccines induces adaptive and generally long-term and specific immunity in the form of neutralizing antibodies and/or activating pathogen-specific cellular immune responses."¹⁴
– GPTZero: *Human-written*