



Contents lists available at ScienceDirect

## Journal of Pathology Informatics

journal homepage: [www.elsevier.com/locate/jpi](http://www.elsevier.com/locate/jpi)

# The ChatGPT conundrum: Human-generated scientific manuscripts misidentified as AI creations by AI text detection tool

Hooman H. Rashidi<sup>a,d,\*</sup>, Brandon D. Fennell<sup>c</sup>, Samer Albahra<sup>a,d</sup>, Bo Hu<sup>b,d</sup>, Tom Gorbett<sup>a,d,\*</sup>

<sup>a</sup> Pathology and Laboratory Medicine Institute (PLMI), Cleveland Clinic, Cleveland, OH, United States

<sup>b</sup> Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, United States

<sup>c</sup> University of California, San Francisco – Department of Medicine, San Francisco, CA, United States

<sup>d</sup> PLMI's Center for Artificial Intelligence & Data Science, Cleveland Clinic, Cleveland, OH, United States

## ARTICLE INFO

## Keywords:

Generative AI  
Machine learning  
Text detection  
Human-generated  
Chat-GPT  
GPT  
LLM  
Large Language Model

## ABSTRACT

AI Chat Bots such as ChatGPT are revolutionizing our AI capabilities, especially in text generation, to help expedite many tasks, but they introduce new dilemmas. The detection of AI-generated text has become a subject of great debate considering the AI text detector's known and unexpected limitations. Thus far, much research in this area has focused on the detection of AI-generated text; however, the goal of this study was to evaluate the opposite scenario, an AI-text detection tool's ability to discriminate human-generated text. Thousands of abstracts from several of the most well-known scientific journals were used to test the predictive capabilities of these detection tools, assessing abstracts from 1980 to 2023. We found that the AI text detector erroneously identified up to 8% of the known real abstracts as AI-generated text. This further highlights the current limitations of such detection tools and argues for novel detectors or combined approaches that can address this shortcoming and minimize its unanticipated consequences as we navigate this new AI landscape.

## Introduction

Artificial intelligence (AI) and the subdomain of Machine Learning (ML)<sup>1–3</sup> have progressed rapidly over the past few years, notably in natural language processing<sup>4</sup> and text generation.<sup>5</sup> In November 2022, the world became enamored with ChatGPT, a large language model (LLM) created by OpenAI, with its ability to answer sophisticated questions and to write novel text. To achieve this, ChatGPT (e.g. Chat Generative Pre-trained Transformer) and similar tools are built upon Pretrained Foundational Models (PFMs), which ingest a large corpus of text to create a general representation of language that can be further tuned to specific applications.<sup>6</sup> The tools excel at summarizing, translating, and synthesizing text with numerous possible uses that come with new challenges.<sup>7</sup> The scientific community and educators worldwide face questions of authorship<sup>8–10</sup> and plagiarism as these tools are used in scientific publications and course assignments,<sup>11</sup> while governments grapple with disinformation.<sup>12</sup> This has led to a second wave of innovation to detect the presence of AI-generated text. Academia has led the charge with a diverse set of approaches (e.g., GLTR<sup>13</sup> and RoBERTa-based approaches<sup>14</sup>) reviewed by Jawahar et al<sup>15</sup> and Guerrero and Alsmadi.<sup>16</sup> In addition, numerous companies, including OpenAI, Copyleaks, and Originality.ai, are now selling AI-generated text detection tools. Given the rapidly evolving nature of this field, ongoing

efforts must be made to understand the performance of these tools. Work by Gao et al. on bioRxiv (<https://doi.org/10.1101/2022.12.23.521610>) looked at AI-generated abstract detection by OpenAI's own detection tool. However, we are yet unaware of research focusing on the opposite scenario. How often do these detection tools flag work as AI-generated when it represents original, unaided creative thought from the author? This has broad-reaching implications, with the possibility of career-altering allegations being made against students and researchers when their work was an original creation. We explore this topic by employing a freely available GPT detector API to analyze abstracts from 5 scientific journals from the 1980s forward, well before AI-aided text generation, to assess the detector's false-positive rate.

## Methods

## Data collection

Abstracts were sourced from PubMed using the Requests and BeautifulSoup Python libraries. These tools enabled us to navigate to a PubMed URL that filtered for the specified year range and journal, and iterated through each search result link to pull the abstract from the underlying HTML source code, using BeautifulSoup's find and find\_all methods. The

\* Corresponding author at: 9500 Euclid Ave, Cleveland, OH 44195, United States.  
E-mail addresses: [rashidh@ccf.org](mailto:rashidh@ccf.org) (H.H. Rashidi), [gorbetw@ccf.org](mailto:gorbetw@ccf.org) (T. Gorbett).

<http://dx.doi.org/10.1016/j.jpi.2023.100342>

Received 28 July 2023; Received in revised form 8 October 2023; Accepted 10 October 2023

Available online 17 October 2023

2153-3539/© 2023 The Author(s). Published by Elsevier Inc. on behalf of Association for Pathology Informatics. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

abstracts were then placed into Pandas dataframes in order to analyze and visualize the data, using Matplotlib.

Five journals covering a range of disciplines were included in this study (*Nature*, *Science*, *The New England Journal of Medicine*, *Radiology*, and *Archives of Pathology and Laboratory Medicine*), and abstracts of which were sampled in 5-year clusters starting from 1980 to the present. The final group ranged from 2020 to February 2023. Abstracts were sourced from the first 5 pages of search results for each journal. 1600 random sample abstracts were selected per each 5-year range (1980–2023) within the aforementioned journals (adding up to 14400 total abstracts evaluated). The query also specified a certain abstract length to meet the token restrictions of the model. The distribution of word counts is provided in Fig. 1.

#### AI detection model

Given that a GPT4 or GPT3-based AI detection API was not available to us at the time of this study, the RoBERTa-based detector model was sourced instead from OpenAI's Hugging Face repository ("roberta-base-openai-detector"). The model was downloaded using pipeline from the transformers Python package, and the AutoTokenizer package. The model was then run locally on the full set of aggregated abstracts (as described above) and given a probability score of 0–1.0 (i.e., the likelihood of the text being identified as AI-generated by this GPT-2-based LLM approach). Although no GPT4-based AI detector API was available to us at the time of this study, a very limited subset of the false-positive results (i.e., 10 real abstracts that were called AI generated with >90% probability by the RoBERTa-based detector model) acquired from our main API-based study were also fed into a separate non-API GPT-4-based AI-detector method through the ChatGPT+ user interface. This separate GPT-4-based method was constructed through the following prompt: "Please act as an AI Detector, responding with a score from 0.00 to 1.00 on how likely the given text is generated by an AI, with 1.00 meaning absolutely written by a human and 0.00 meaning absolutely written by an AI." A summary of the results of these very limited runs (10 that were erroneously detected by the GPT2-based method as AI-generated) are shown in Supplementary table 1.

#### Data analysis

The AI detection output was segmented by 5-year time periods starting from 1980 to the present (2023), for each of the aforementioned journals listed above. For each time period, the full-length abstracts were utilized to better represent the true-predictive capability of the AI detection tool on these real abstracts. As noted above, to maintain the same sample size for each year range, 1600 abstracts per each 5-year range were randomly selected using the NumPy Python package. A 50% probability was used

as the base threshold for suspected AI generation, with additional analysis for each of these spanning 10% intervals up to >90% AI text generation detection calls (i.e., >50%, >60%, >70%, >80%, and >90%) as shown in Fig. 2. In addition to the above, a separate text dilution study was also conducted to serve as a relatively predictable positive control (100% of abstract, first 50%, and first 25% of each of the abstracts, by word count), as shown in Fig. 3. The word count was compared by year range and journal using ANOVA. The statistical analysis was performed using R ([cran.r-project.org](http://cran.r-project.org)).

#### Results

Through our AI-text detection approach using the RoBERTa-based detector model, we found that while most of these "real" abstracts were correctly predicted as non AI-generated (i.e., correctly assigned as human-generated), there was a subset of abstracts (~5%–10%) that were being erroneously assigned as AI-generated text with a high degree of confidence (i.e., >80% confidence as seen in Fig. 2).

The 1600 random real abstract samples per year range (adding up to 14400 total abstracts) were separated based on their likelihood of being called AI generated text according to the AI detection tool as shown in Fig. 2. The bar chart in Fig. 2 illustrates the trend of AI detection of such real text over time, along with their given likelihood as reported by the AI text detector. For example, in each bar, the "Greater than 90% chance of being AI-generated" represents the abstracts that were greater than the 90% likelihood of being AI generated by the AI detection tool.

Additionally, the same 14 400 sample set was also reviewed based on each of the 5 journals evaluated. Each column adds up to 100% for each journal as shown in Table 1.

As noted earlier, a separate text dilution study was also conducted (Fig. 3) on these 14 400 abstracts to evaluate the effect of word count on the AI detection tool's optimized capabilities. This was conducted by using the full abstract (100% of abstract) versus its first half (50%) and its first quartile (25%) word count as shown in Fig. 3. As expected, the AI detection tool performed best on the full abstract when compared to its diluted abstract counterparts.

We also evaluated the average word count for each journal and year range, as well as the average for each. Based on 2-way ANOVA, the P-values were calculated for comparing each year range's mean word count to other year changes from that year range and the mean word counts between different journals (Appendix Fig. 1).

#### Discussion

In this study, we have demonstrated that up to 8.69% of the "real" abstracts had more than 50% probability of being characterized as AI-generated text, and up to 5.13% of all "real" abstracts being erroneously

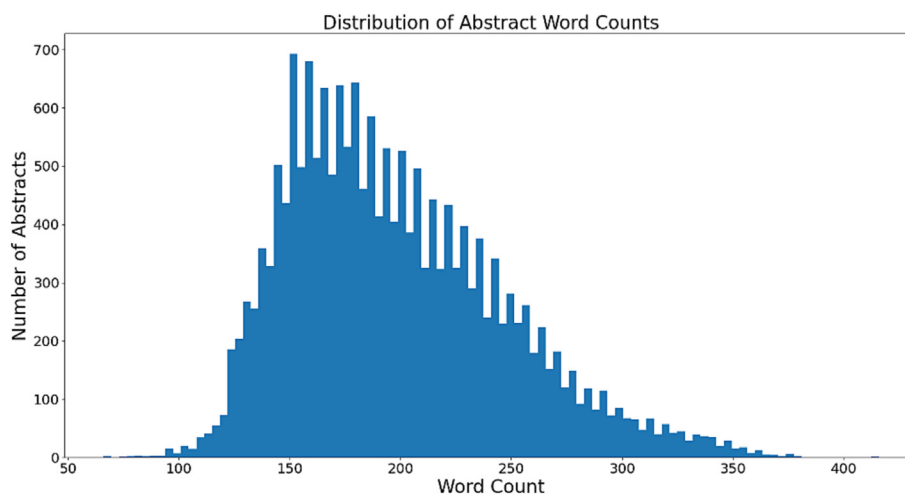


Fig. 1. The overall distribution of the word counts in the abstracts.

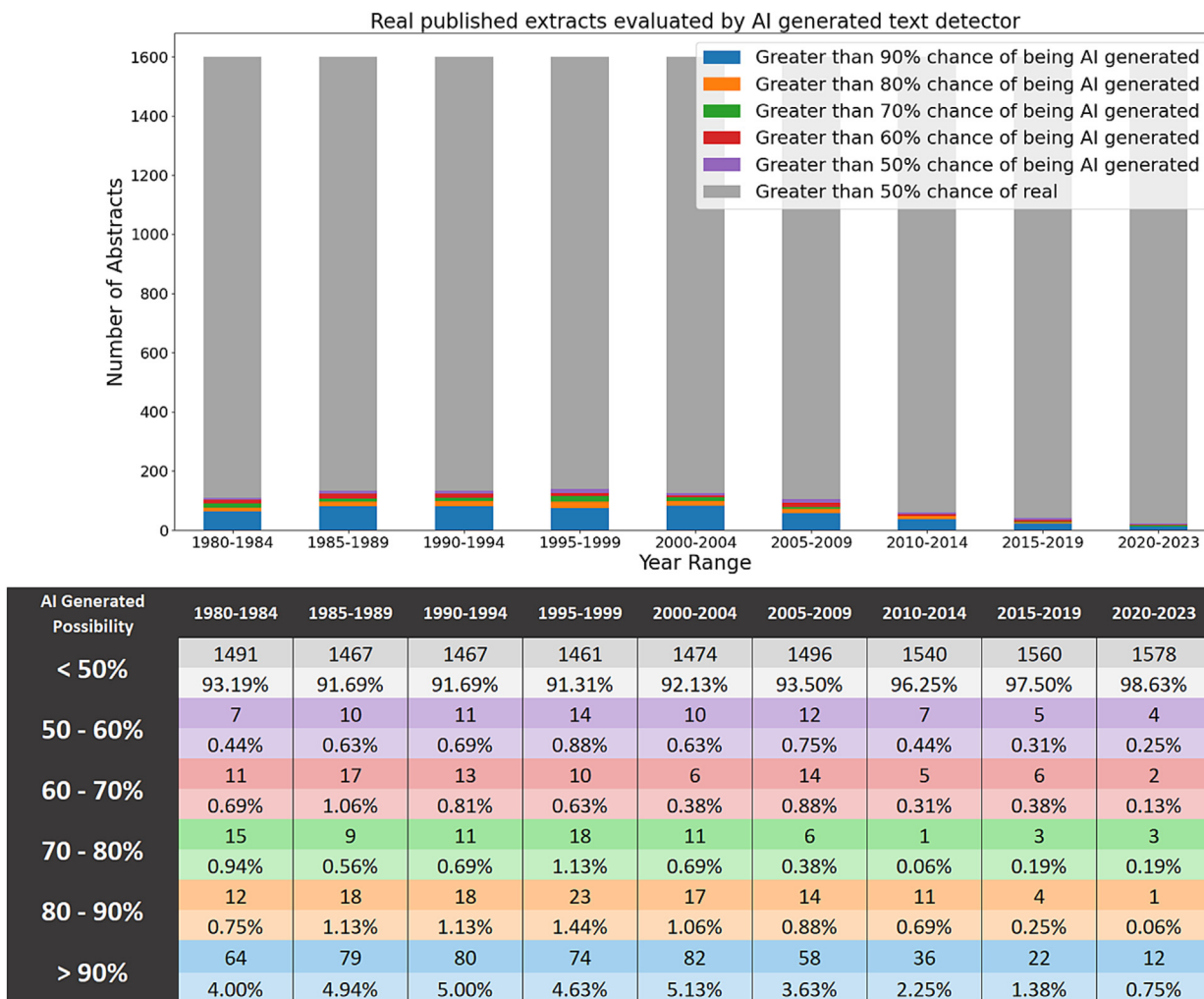


Fig. 2. The full-length abstracts evaluated by the AI-detection tool are shown in the bar chart above over year ranges, with each stack representing the AI detection model's prediction of how likely an abstract is to be predicted as being AI generated. The table beneath the graphs shows corresponding details for each year range.

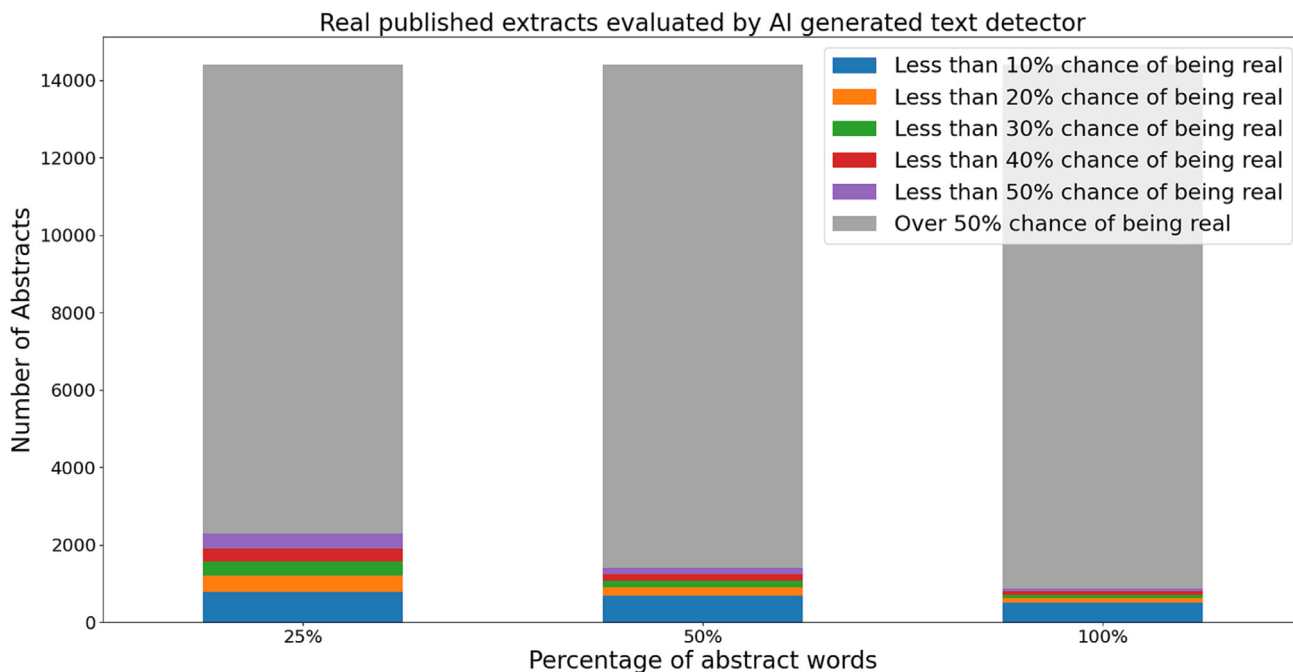
characterized as having a 90% likelihood of being AI-generated. As a more concrete example, from 1990 to 1994, well before ChatGPT or computational capabilities of LLMs, roughly 1 in 20 abstracts were flagged as AI-generated with >90% probability.

However, the data also demonstrates a temporal dependence on its detection. Trending the >90% AI-generated predicted probability by 5-year segment, a peak is seen in 2000–2004 with a steady decline reaching only 0.75% in the 2020–2023 segment. The exact cause of these changes requires further study, but several factors are worth considering. First, the RoBERTa-based OpenAI detector used in this study is fine-tuned with GPT-2 parameters. The corpora used to train RoBERTa combines BookCorpus and English Wikipedia (BERT training set), CC-News, OpenWebText, and Stories.<sup>14</sup> While GPT-2 is trained with WebText, it is created by following all outbound links from Reddit with at least 3 karma points.<sup>17</sup> Thus, the tuning draws on popular ideas shared on a social media site rather than scholarly works. However, this doesn't fully explain the improving accuracy over time. The observed differences may also result from temporal misalignment, a phenomenon nicely summarized by Luu et al., in which the language model training data and testing data differ enough in time that model performance is affected by the subtle evolution of language.<sup>18</sup> However, neither fully explain all the observed trends in the data. If this were the only factor, we would expect even less accurate predictions in the earlier decades. However, in the early 2000s forward, when a more rapid change in accuracy was observed, the adoption of the internet was also rapidly accelerating, potentially catalyzing faster evolution of language or other

additional factors, resulting in more noticeable difference in performance over time.

To further interrogate the model, we also tested the effect of text length on the prediction, serving as a positive control with more text expected to provide more accurate predictions within the token limit of the model. Reassuringly, the accuracy of predictions was demonstrated to be directly related to the abstract lengths when provided with increasing aliquots of the abstracts (25%, 50%, or 100% of text), as shown in Fig. 3.

Considering this, the data also was found to be variably influenced by the journal type, regardless of the time period. Specifically, 4 of the journals representing either specific disciplines (*Radiology*, *Archives of Pathology & Laboratory Medicine*) or a diverse array of subjects (*Science*, *Nature*) showed false-positive rates between 1.89 and 3.89%, while *The New England Journal of Medicine (NEJM)* was shown to be an outlier at 10.24% for unknown reasons. We further analyzed the average word count of each set of journal abstracts and employed ANOVA to evaluate the statistical difference. When the average word count of all abstracts was compared with the average word count of the individual journals, each journal was found to be relatively different ( $P < .05$ ). *Nature* was closest to the average with a  $P$ -value of .0415, with the remaining journals showing  $P$ -values of  $< .001$  (S1 Fig. 1). This suggests that similar performance is observed with statistically significantly different abstract lengths, and this likely is not the reason for the difference in the performance of *NEJM*. The format and structure of the abstracts is journal dependent, with *NEJM*, *APLM*, and *Radiology* all using various subsections (i.e., context, objective, design, results, and conclusions), while *Nature* and *Science* employ a single paragraph form. Still,



AI Generated Possibility	25%	50%	100%
< 50%	12136 84.13%	13028 90.31%	13558 93.98%
50 - 60%	390 2.70%	168 1.16%	80 0.55%
60 - 70%	339 2.35%	165 1.14%	84 0.58%
70 - 80%	356 2.47%	165 1.14%	77 0.53%
80 - 90%	426 2.95%	220 1.53%	120 0.83%
> 90%	779 5.40%	680 4.71%	507 3.51%

Fig. 3. The effect of diluting the abstracts (i.e., comparing the full length abstract 100% to 50% which represents the first half of the abstract and 25% representing the first quarter of the abstract length) on the AI-detection tool's outcomes. As expected, the AI-detection tool's capabilities of accurately assigning the real text as real, diminishes through each dilution.

this does not distinguish *NEJM* from the other journals. While the underlying cause remains unclear, it also highlights the importance of context for

Table 1

The percentages of each journal separated by likelihood of being called AI-generated text. *Arch Pathol Lab Med* and *NEJM* noted above are the journals of *Archives of Pathology & Laboratory Medicine* and the *New England Journal of Medicine*, respectively.

AI generated possibility (%)	<i>Arch Pathol Lab Med</i> (%)	<i>Nature</i> (%)	<i>Radiology</i> (%)	<i>Science</i> (%)	<i>NEJM</i> (%)
<50	93.98	96.69	95.72	93.89	84.69
50-60	0.60	0.31	0.75	0.65	0.84
60-70	0.75	0.31	0.58	0.56	1.12
70-80	0.64	0.31	0.44	0.19	1.30
80-90	0.94	0.50	0.58	0.83	1.81
>90	3.11	1.89	1.94	3.89	10.24

the tool's performance and demonstrates the potential for similar performance across multiple subject domains and abstract lengths.

The confidence in the probabilities also merits further discussion. The data demonstrate not only the identification of certain real text as AI-generated but with a confidence exceeding 90%. While these tools exist in an academic realm, thoughtful interpretation and skepticism will likely remain the norm; however, as they begin to transition into a broader consumer market with multiple corporations offering this technology, the notion of such strong prediction will warrant additional concern.

One fundamental limitation of this study is that it only evaluates a single AI-text detection model, which is not the most current technology available. OpenAI and other companies continue to develop more sophisticated methods to detect AI-text, but they have not, at the time of authoring this article, offered API access to these tools to facilitate large-scale analysis. Additionally, due to the expense of these tools, we were limited to using the freely available model. Ongoing assessment is needed especially given GPT-3 and GPT-4's much more extensive corpus<sup>19,20</sup> compared with

GPT-2, leading to more sophisticated text generation while also allowing training and tuning of more sophisticated detectors. Notably, a small sample of the real abstracts (10 random abstracts, 2 from each of the 5 journals) that were being called as AI-generated by the GPT-2-based model, were also evaluated in a GPT-4 non-API-based method through ChatGPT + UI which showed some interesting results as noted in our Supplementary Table 1. Although, at first glance, this GPT4-based AI-detector method was able to accurately identify the real abstracts as human and non-AI-generated, it also was shown to be rather non-specific for this task since a separate set of AI-generated abstracts fed into it were all also being called human generated (see Supplementary table 1). It is important to note that this GPT-4-based non-API approach used within this very limited subset of abstracts is not a validated method and more sophisticated trained and validated AI detection LLM approaches in the future (e.g., GPT-4 or others) may be able to address many of the aforementioned shortcomings.

In light of these results, it is worth noting that the detection rate provided by these tools, especially in the 2020–2023 abstracts, must be interpreted with caution. While our assumption that the AI generator technologies could not have been used in earlier cases, the assumption no longer applies in this time frame and thus, the true value of the false detection rate is less than or equal to the quoted value of 0.75%. Finally, as noted above, temporal misalignment likely amplifies the false-positive rate especially within the older abstracts.

Looking forward, the role and impact of LLMs remain in their infancy. As the scientific and educational community grapples with these changes, distinguishing human versus AI creations will remain an important subject as AI-generated text’s potential for misuse and abuse looms large.<sup>21,22</sup> There now exists an “arms race” not only to push the envelope of possibility but to continue to distinguish humans from AI.<sup>23</sup> Drawing on lessons from medical diagnostic testing, we envision new methods emerging with a highly sensitive screening test to meet the needs of identifying content of concern which can then be combined with highly specific diagnostic tests to confirm that suspicion and boost the overall performance. Additionally,

the importance of standardizing and benchmarking the performance of these models in different domains and contexts also remains important as in the work by Uchendu et al in creating TuringBench.<sup>24</sup> While this innovation accelerates, it is essential not to lose sight of the pitfall of labeling something as AI when it is, in fact, human. The observations presented here is only the beginning of this narrative.

**Contributions**

Hooman Rashidi (HR) and Tom Gorbett (TG) are the lead investigators and have developed and validated the above study whose results are highlighted within this manuscript. Brandon Fennel (BF), Bo Hu (BH), and Samer Albahra (SA), along with HR and TG, significantly contributed to the writing, review, statistical analysis, and editing of the manuscript. All authors reviewed the manuscript.

**Data availability**

The datasets collected and analyzed during the current study will be made available on Github via a public repository upon publication of this work.

**Declaration of interests**

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dr. Hooman Rashidi serves as an Associate Editor for JPI. Dr. Brandon Fennel has served as a reviewer for prior work published in JPI.

**Appendix A. Appendix**

*A.1. Word counts and p-values by year range and journal*

Year Range	Arch Pathol Lab Med Word Count	Arch Pathol Lab Med p-value	Nature Word Count	Nature p-value	Radiology Word Count	Radiology p-value	Science Word Count	Science p-value	NEJM Word Count	NEJM p-value	Total Word Count	Total p-value
1980-1984	151.183	<.0001	183.2965	<.0001	169.3113	<.0001	164.3077	<.0001	183.6531	<.0001	177.9794	<.0001
1985-1989	172.8927	<.0001	187.0545	<.0001	165.8025	<.0001	173.4929	<.0001	228.3395	0.006104	194.0494	<.0001
1990-1994	194.3539	0.0944	170.7029	<.0001	173.6226	<.0001	155.0492	0.014261	260.7025	<.0001	188.6125	<.0001
1995-1999	193.1922	0.2407	177.2495	<.0001	209.7798	<.0001	136.6727	0.040423	259.3189	<.0001	199.4075	0.0441
2000-2004	188.0734	0.1931	181.2178	<.0001	227.1258	0.000652	144.7121	0.650383	256.0905	<.0001	200.18	0.1661
2005-2009	195.3571	0.0173	210.6511	<.0001	224.2932	0.103883	140.7531	0.045832	245.7133	0.000158	204.7425	0.0114
2010-2014	198.9042	<.0001	217.1418	<.0001	236.8529	<.0001	135.2258	0.000221	236.2857	0.592972	206.5175	<.0001
2015-2019	201.5607	<.0001	217.1077	<.0001	284.5431	<.0001	133.6424	<.0001	236.7447	0.592972	223.7962	<.0001
2020-2023	220.8028	<.0001	207.2953	<.0001	290.4152	<.0001	135.928	0.001511	198.5714	<.0001	220.5469	<.0001
<b>Total</b>	<b>193.4598</b>	<b>0.0003</b>	<b>197.8223</b>	<b>0.0415</b>	<b>220.6326</b>	<b>&lt;.0001</b>	<b>146.1878</b>	<b>&lt;.0001</b>	<b>224.3741</b>	<b>&lt;.0001</b>	<b>201.7591</b>	

Fig. A1. Average word counts and p-values are graphed by year range and journal, using the random sample of 14,400 abstracts.

## A.2. Information to Access Supporting Code on Github

Github repository link (placeholder) will be published and provided upon publication.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jpi.2023.100342>.

## References

- Rashidi HH, Tran NK, Betts EV, Howell LP, Green R. Artificial intelligence and machine learning in pathology: the present landscape of supervised methods. *Acad Pathol* 2019;6.237428951987308.
- Rashidi HH, Bowers KA, Reyes Gil M. Machine learning in the coagulation and hemostasis arena: an overview and evaluation of methods, review of literature, and future directions. *J Thromb Haemost* JTH 2022;S1538-7836(22):18293. <https://doi.org/10.1016/j.jtha.2022.12.019>.
- Rashidi HH, Albahra S, Robertson S, Tran NK, Hu B. Common statistical concepts in the supervised machine learning arena. *Front Oncol* 2023;13.
- Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. *J Am Med Inform Assoc JAMIA* 2011;18:544–551.
- Zhou C, et al. A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT. Preprint at: <https://doi.org/10.48550/arXiv.2302.09419>; 2023.
- Bommasani R, et al. On the Opportunities and Risks of Foundation Models. *arXiv.org* <https://arxiv.org/abs/2108.07258v3>, <https://doi.org/10.48550/arXiv.2108.07258>; 2021.
- Shen Y, et al. ChatGPT and other large language models are double-edged swords. *Radiology* 2023;230163. <https://doi.org/10.1148/radiol.230163>.
- Stokel-Walker C. ChatGPT listed as author on research papers: many scientists disapprove. *Nature* 2023;613:620–621.
- Lee, J. Y. & Huh, S. Can an artificial intelligence chatbot be the author of a scholarly article?
- Thorp HH. ChatGPT is fun, but not an author. *Science* 2023;379:313.
- Morrison R. Large Language Models and Text Generators: An Overview for Educators. Online Submission: <https://eric.ed.gov/?id=ED622163>; 2022.
- Stiff H, Johansson F. Detecting computer-generated disinformation. *Int J Data Sci Anal* 2022;13:363–383.
- Gehrmann S, Strobelt H, Rush A. GLTR: statistical detection and visualization of generated text. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. Association for Computational Linguistics; 2019. p. 111–116. <https://doi.org/10.18653/v1/P19-3019>.
- Liu Y, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. Preprint at: <https://doi.org/10.48550/arXiv.1907.11692>; 2019.
- Jawahar G, Abdul-Mageed M, Lakshmanan LVS. Automatic Detection of Machine Generated Text: A Critical Survey. Preprint at: <https://doi.org/10.48550/arXiv.2011.01314>; 2020.
- Guerrero J, Alsmadi I. Synthetic Text Detection: Systemic Literature Review. Preprint at: <https://doi.org/10.48550/arXiv.2210.06336>; 2022.
- Radford, A. et al. Language models are unsupervised multitask learners.
- Luu K, Khashabi D, Gururangan S, Mandym K, Smith NA. Time Waits for No One! Analysis and Challenges of Temporal Misalignment. Preprint at: <http://arxiv.org/abs/2111.07408>; 2022.
- Brown TB, et al. Language Models are Few-Shot Learners. Preprint at: <http://arxiv.org/abs/2005.14165>; 2020.
- OpenAI. GPT-4 Technical Report. Preprint at: <https://doi.org/10.48550/arXiv.2303.08774>; 2023.
- Solaiman I, et al. Release Strategies and the Social Impacts of Language Models. Preprint at: <https://doi.org/10.48550/arXiv.1908.09203>; 2019.
- Kreps SE, McCain M, Brundage M. All the News that's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation. SSRN Scholarly Paper at: <https://doi.org/10.2139/ssrn.3525002>; 2020.
- Heidt A. 'Arms race with automation': professors fret about AI-generated coursework. *Nature* 2023. <https://doi.org/10.1038/d41586-023-00204-z>.
- Uchendu A, Ma Z, Le T, Zhang R, Lee D. *TURINGBENCH: A Benchmark Environment for Turing Test in the Age of Neural Text Generation*. 2021.