


Article

Detecting the Use of ChatGPT in University Newspapers by Analyzing Stylistic Differences with Machine Learning

Min-Gyu Kim and Heather Desaire * 

Department of Chemistry, University of Kansas, Lawrence, KS 66045, USA; mgkim603@ku.edu

* Correspondence: hdesaire@ku.edu; Tel.: +1-785-864-3015

Abstract: Large language models (LLMs) have the ability to generate text by stringing together words from their extensive training data. The leading AI text generation tool built on LLMs, ChatGPT, has quickly grown a vast user base since its release, but the domains in which it is being heavily leveraged are not yet known to the public. To understand how generative AI is reshaping print media and the extent to which it is being implemented already, methods to distinguish human-generated text from that generated by AI are required. Since college students have been early adopters of ChatGPT, we sought to study the presence of generative AI in newspaper articles written by collegiate journalists. To achieve this objective, an accurate AI detection model is needed. Herein, we analyzed university newspaper articles from different universities to determine whether ChatGPT was used to write or edit the news articles. We developed a detection model using classical machine learning and used the model to detect AI usage in the news articles. The detection model showcased a 93% accuracy in the training data and had a similar performance in the test set, demonstrating effectiveness in AI detection above existing state-of-the-art detection tools. Finally, the model was applied to the task of searching for generative AI usage in 2023, and we found that ChatGPT was not used to revise articles to any appreciable measure to write university news articles at the schools we studied.

Keywords: ChatGPT; AI; text detection; generative AI; LLM; supervised classification; journalism



Citation: Kim, M.-G.; Desaire, H. Detecting the Use of ChatGPT in University Newspapers by Analyzing Stylistic Differences with Machine Learning. *Information* **2024**, *15*, 307. <https://doi.org/10.3390/info15060307>

Academic Editor: David Martins de Matos

Received: 17 April 2024

Revised: 20 May 2024

Accepted: 22 May 2024

Published: 25 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, text generators based on artificial intelligence (AI), such as ChatGPT, have garnered the attention of many different fields, especially in academia and education [1]. College students have been early adopters of ChatGPT. In a multinational study on the use of ChatGPT, 53% of students claimed previous use of generative AI, while 47% had heard of ChatGPT [2]. In a related study, students and educators stated that the utilization of ChatGPT could allow students to focus more on creative tasks while providing tools for non-native English-speaking students [3]. However, because of its broad range of uses, including the ability to create written works and answer almost any prompt given [4], educators have expressed their concerns about the decreasing levels of critical thinking and creativity amongst college students [5,6]. Additionally, ethical concerns have risen over the question of content originality, since AI-based text generators such as ChatGPT recycle information from other sources without referencing the original source, prompting a greater need to develop ways to detect the use of AI in written works [5,7].

Additional concerns about the technology are notable if it were to be leveraged in the field of journalism. The potential for the introduction of bias and misinformation [8,9], two well-known, problematic aspects of AI text generators, makes the use of these tools in journalism very risky. News articles and outlets should be trustable by consumers, providing information that is verified to be factual and true. AI-based text generators do not yet deliver error-free, unbiased information [10].

While news consumers have reason to be concerned about AI-generated content, a highly accurate tool for identifying the use of AI in newspaper articles is not yet available.

In fact, the detection of AI-generated text in general is quite difficult, particularly with the increasing parameter sets of the latest models [11]. One research team found that when humans were given a sample of AI-generated scientific abstracts and the original abstracts, only 68% of AI-written abstracts were identified as being written by AI, and 14% of the original abstracts were mistaken to be AI-generated [12]. This example clearly demonstrates the need for a systematic and reliable method of AI detection in writing.

Several different research projects on the detection of AI in written works have been carried out. For example, the RoBERTa detector, a binary classifier that distinguishes human writing from AI, has been used in many different studies due to its large training set (160 GB in total) and its consistently good performance [13]. In one case, the detector was provided with question-and-answer texts from public questionnaire websites, such as Wikipedia and Reddit, and its corresponding human answers, as well as ChatGPT-generated answers. RoBERTa's F1 accuracy when given the entire body of text was 81.89%, and manual differences, such as shorter sentence lengths for humans and less emotional language used by the AI, were found [14].

Another relevant study included the use of ChatGPT itself to detect AI-counterpart news articles generated using a variety of news sources, including the *Washington Post* and *CNN*, from the public data set, TuringBench [11,15]. ChatGPT was able to detect AI articles generated using the GPT-1 model 90% of the time, but its performance decreased significantly as the model version increased. When tested on the latest model, GPT-4, ChatGPT identified 97–100% of counterpart articles as AI-generated, but misclassified 95% of human articles as AI-generated, showing ChatGPT's inability to discriminate between AI writing and human writing [11].

A third relevant study used a human-developed feature set of twenty stylistic differences and classical machine-learning tools to distinguish between scientific journals and ChatGPT-generated counterparts [16]. Training/test data sets were developed using scientific research papers from the journal, *Science*, and from AI-generated articles on the same topics. The model had a >99% accuracy on full documents and a 94% accuracy on individual paragraphs [16]. This method has been shown to be useful in developing a highly accurate model compared to current AI detectors available to the public. The "stylistic differences" approach was also recently applied to a larger data set, including 13 different chemistry journals and AI-generated data from multiple prompts, using both ChatGPT and GPT-4. Once again, it demonstrated the ability to distinguish AI- from human-generated writing with high accuracy [8]. Due to the effectiveness of this strategy in detecting AI usage, the method was adapted herein. We developed a similar model to detect the usage of AI, specifically ChatGPT, in university newspapers.

The goals of this study were, firstly, to develop a method to discriminate articles that were fully human-written versus those that included drafts written by humans but with a finished product generated by ChatGPT, using a strategy similar to the one successfully demonstrated previously [8,16], and secondly, to identify whether collegiate journalists are using ChatGPT to edit their university news articles. To accomplish these goals, we collected a training set of student-written news articles from a variety of sources and identified differentiating features, such as grammatical usage and popular words used. These features can be incorporated into a machine learning model. The accuracy of the developed model was tested by comparing its performance to existing state-of-the-art online tools. Finally, we assessed whether students were using ChatGPT in the writing process to produce material for college newspapers.

2. Experimental Details

2.1. Data Set Selection

To accurately portray the diversity of university newspapers and journalism styles incorporated by different colleges, 12 universities' newspapers were chosen to devise the initial training data set. Universities in the training set included: University of Arizona, University of Kansas, University of Missouri, University of Washington, University of

Pittsburg, Auburn University, Florida Gulf Coast University, Wheaton College, Taylor University, Howard University, Johns Hopkins University, and University of Southern California. The criteria for the articles chosen involved selecting 10 news articles from each university newspaper, with articles chosen in numerical order from the date released, starting on 1 September 2022. All training articles were published prior to 30 November 2022, which was the public release date of ChatGPT. Selected articles also had to meet the requirement of having been written by students who attended the respective university. These news articles reported on a variety of topics, such as the implementation of university policies, police reports within the surrounding college town, and overviews of recent college research findings. For each article, the author(s) name(s) and subheading(s) were removed, and the entire article was compressed into a single writing example as part of the data set preparation.

With 120 human-written articles collected, counterpart articles were generated using ChatGPT (Version 3.5) by prompting it with the following phrase and copying the entire human-generated article afterwards: “give me a story in a newspaper style of this story: (insert article)”. The prompt and method of generating the AI-written articles were chosen by taking into consideration the writing process of students and the most likely method that a student would use ChatGPT to generate a news article. These aspects included prompting ChatGPT to generate an article in a specific writing style and providing it with a human-generated draft containing all the details to create a refined article that could be indistinguishable to a human reader. The AI-generated articles varied in length and topic, but generally followed the example article closely. In total, 240 articles made up the training set. Example ChatGPT-generated text can be found in Supplementary Materials, matrix_B.csv, and the prompts used to generate these, along with the links to all the human-written articles are also provided in the Supplementary Materials.

Test set 1 included 50 articles from Cornell University, University of Wisconsin Madison, Syracuse University, University of North Carolina Chapel Hill, and University of Texas at Austin, and the corresponding ChatGPT-generated articles. The news article selection followed the same criteria as the training set. Test set 2, which contained only ChatGPT-generated articles, was built using the same human-written articles from test set 1; however, the following prompt was used: “rewrite this article with all the details in a university newspaper style with all the quotes: (insert article)”.

The final set of documents used in this study were newspaper articles from 2023, where ChatGPT may or may not have been used. News articles from 2023 were collected in four batches. Set 1 included 50 articles from the same universities as test sets 1 and 2. For each of the five universities, 10 articles written sequentially after 31 January 2023 were chosen. Set 2 was acquired in the same way as set 1, except all articles appeared after 30 April 2023. The same method was used to collect set 3 (after 31 August) and set 4 (after 31 October).

2.2. Development of Features

The 240 articles in the training set were manually compared, and several distinct differences were noted. Four main categories of features were identified, which included 1. complexity of the article, 2. types of punctuation present, 3. sentence-level diversity, and 4. common words present. These four types of features had been shown to be effective in classifying scientific journals as human- or AI-written in previous related studies [8,16]. Although the use of several of the distinctions above had already been seen in online AI detectors, the combination of the distinctions and the specifics of the linguistic features used herein had not been reported previously. The distinctions used in the trained model are shown in Table 1.

Table 1. Features used in model.

Feature Number	Feature Type	Short Description	Greater in
1	1	sentences per article	Human
2	1	words per article	Human
3	2	"-" present	ChatGPT
4	2	"," or ":" present	Human
5	2	"?" present	Human
6	2	how many quotation marks present	Human
7	3	standard deviation in sentence length	Human
8	3	length differences in consecutive sentences	Human
9	4	"which" present	Human
10	4	how many "said" present	Human
11	4	how many "but" present	Human
12	4	how many "this" present	Human
13	4	how many "freshman, sophomore, junior, senior" present	Human

Feature types: (1) complexity of article, (2) types of punctuation present, (3) sentence level diversity, and (4) common words present.

2.3. Data Processing

All data processing was completed using RStudio, version 4.3.0. The text from each article initially comprised a single row of a matrix of text. From this matrix, each text example was converted to a numerical 13-feature vector, using the script provided in the Supplemental Materials (Example Code.pdf), and this converted feature matrix was used for all subsequent analyses. Principal components analysis was used to assess the approach and to evaluate the overall variability in the human- and AI-generated articles in the data set. The resulting feature matrix was then used for training and testing using supervised classification with XGBoost 1.7.7.1. The model's parameters are listed below and were not optimized: booster = "gbtree", objective = "multi:softmax", num_class = 2, eta = 0.3, gamma = 0, max_depth = 6, min_child_weight = 1, subsample = 1, colsample_bytree = 1, nrounds = 50, and maximize = F. The training data were evaluated using leave-one-out cross-validation (LOOCV), where each article in question was left out of the training set prior to its evaluation. During testing, all of the articles from the training set were leveraged into the model, and no adjustment to any of the parameters was made. Accuracy was defined as the articles correct over the total number of articles. AUC data was obtained using the package, pROC.

3. Results

The workflow for this experiment and a PCA plot of the converted data is shown in Figure 1. The data demonstrate that the individual articles generally cluster according to the author (human or AI). As seen in the figure below, the human articles formed a wider cluster; thus, they are more diverse, and the ChatGPT articles were found to be in a tighter cluster, and thus, more like one another. Furthermore, the unsupervised data show a fairly good separation between the two article types, providing support that a reasonably accurate supervised model could be developed.

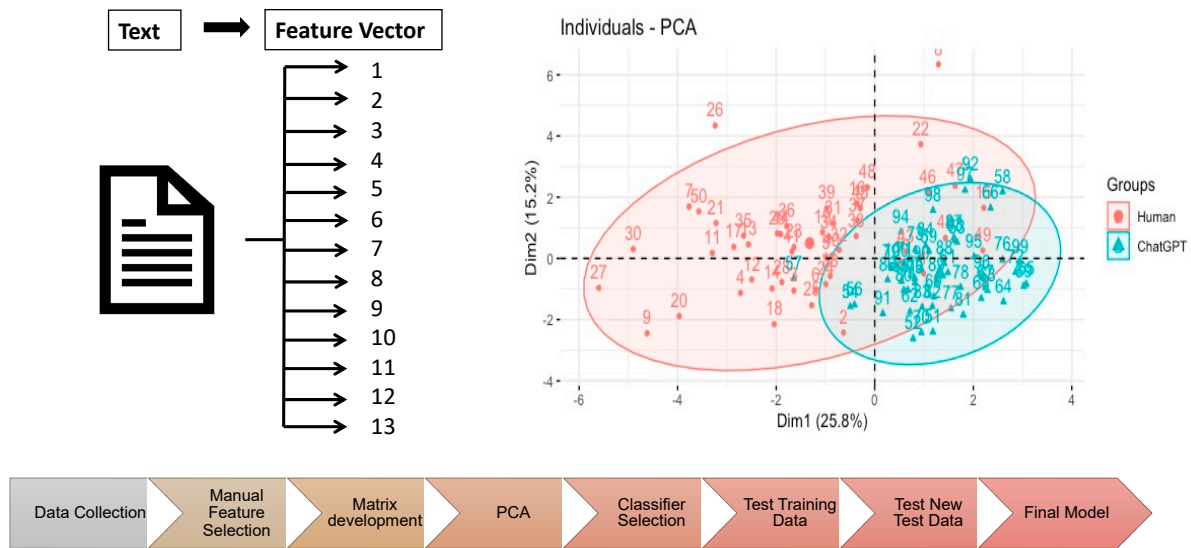


Figure 1. Model training method and binary class assignments of data set. PCA plot of data from the test set is shown above.

After the final set of features had been developed and the PCA plot indicated a good separation of the data, we moved forward with the supervised classification following the strategy described previously [8,16]. The classification accuracy of the training data (assessed using LOOCV) was shown to be satisfactory, as demonstrated in Table 2, so testing on unseen data commenced. The accuracy of the results for the test sets is also shown in Table 2. The model proved adept at determining whether an article was written solely by humans or edited by ChatGPT, both during the training and the testing conditions. We found that 94% of the human-written articles were correctly assigned during training, using LOOCV, with only a small drop-off (of 2%) in the test set.

Table 2. Accuracy of developed model on training and test data sets.

	Article-Level (% Correct)		
	Articles	Percent Correct	AUC
Training Set (Human)	120	94.2%	
Training Set ^a (ChatGPT)	120	92.5%	
		Overall Accuracy: 93.3%	0.933
Test Set 1 (Human)	50	92%	
Test Set 1 ^a (ChatGPT)	50	98%	
		Overall Accuracy: 95%	0.95
Test Set 2 ^b (ChatGPT)	50	78%	
		Overall Accuracy: 85%	0.85

Article level accuracy of detector model with test set 1 (original prompt) and test set 2 (changed, detailed prompt). ^a. Prompt 1: “give me a story in a newspaper style of this story: (human example article)” ^b. Prompt 2: “rewrite this article with all the details in a university newspaper style with all the quotes: (human example article)”.

After assessing the model on human-generated data, test sets for AI-generated data were also analyzed. When the same prompt, which had been used for training, was used to generate data for the AI test set, the percentage of correct assignments was still quite good, measuring at 98%. (See Table 2). While the results for the test set were nominally better than the training set, we attributed this difference primarily to random chance and the size of the data sets. We were also well-aware that testing different prompts is important [8]. Furthermore, others have shown that more detailed prompts produce higher-

quality answers from large language models, especially ChatGPT [1]. Prompts that are clear, contain more controls, and provide previous writing examples and structure, result in answers that are more indistinguishable to humans [8]. To test this model's ability to detect text that had been edited by AI when more detailed prompting was used, a new test set, generated with a new prompt, was produced. The more detailed prompt 2 indicated the use of a specific style (university newspaper) and instructed the model to include all of the quotes provided (the exact prompt is stated at the bottom of Table 2). These texts were somewhat more difficult to classify, as expected. Yet, considering that the model was given a complete article and its changes were expected to be somewhat modest and, therefore, difficult to discriminate from the original text, we deemed the performance to be reasonably good.

The next objective was to compare the quality of the trained detector to other public AI detectors. The Roberta GPT-2 output detector and the ZeroGPT detector were chosen as comparison methods [8,16]. The results were then compared to the results of our detector. They are shown in Figure 2.

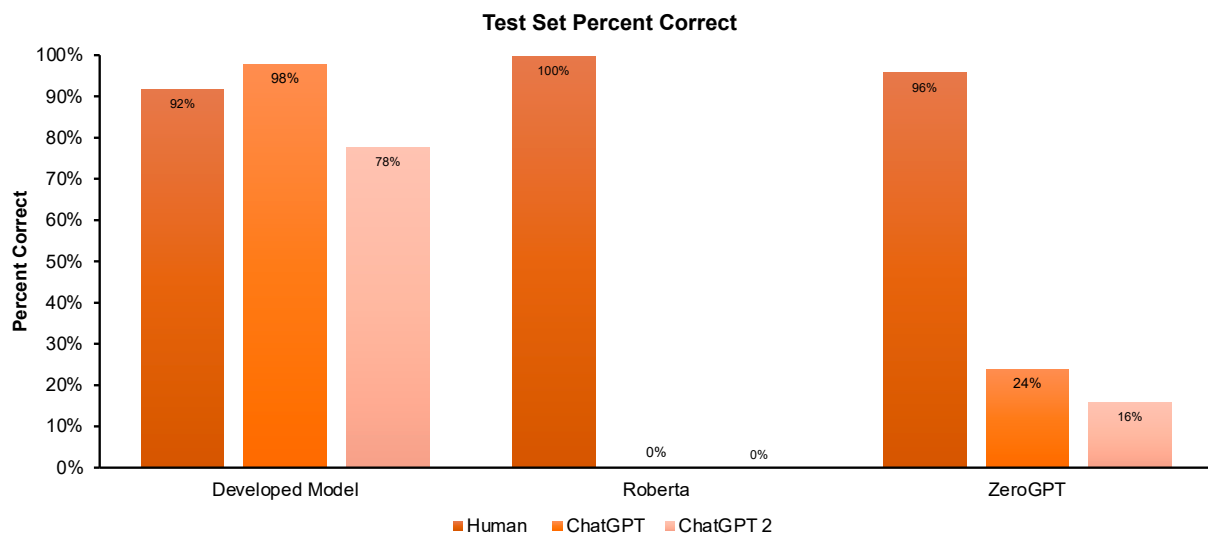


Figure 2. Differences in performance among different models analyzing test sets.

As seen in Figure 2, the developed model correctly identified the presence of ChatGPT used in editing (or not) a high proportion of the time, while the two public AI detectors essentially classified almost all the articles as “human written”, even though the final text was fully AI-generated. In all three sets of data, the Roberta model type was unable to detect any of the ChatGPT-modified articles, while ZeroGPT performed slightly better, correctly detecting 19% of the ChatGPT articles within the data sets. By comparison, the model developed herein was able to correctly identify articles that had leveraged ChatGPT 90% of the time.

Finally, we sought to determine whether texts edited with ChatGPT could be detected in university newspaper articles written in 2023. We acquired 50 articles written after 30 January from the same newspapers as those used in the test set, generated the 13-feature vectors in the same way as before, and used these data to classify the articles as human- or AI-written using the above-described model. Additionally, a second set of 50 articles from the same universities, which had all been written after 30 April, a third set of 50 articles written after 31 August, and a fourth set of 50 articles written after 31 October were also tested. Because of the relatively recent release of ChatGPT, we predicted that the percentage of articles written with the assistance of AI would start out very low and potentially increase slightly post-30 April and beyond. The classification results are shown in Figure 3. As expected, the model predicted that 98% of the unknown articles were written exclusively by humans immediately after 31 January. The April, August, and October test

sets were also classified as human-written, with about the same frequency as was seen in the training and test data. Based on these results, a reasonable conclusion can be made that student journalists were not using ChatGPT in the editing stages to generate university news articles in 2023 (with some caveats, as noted below).

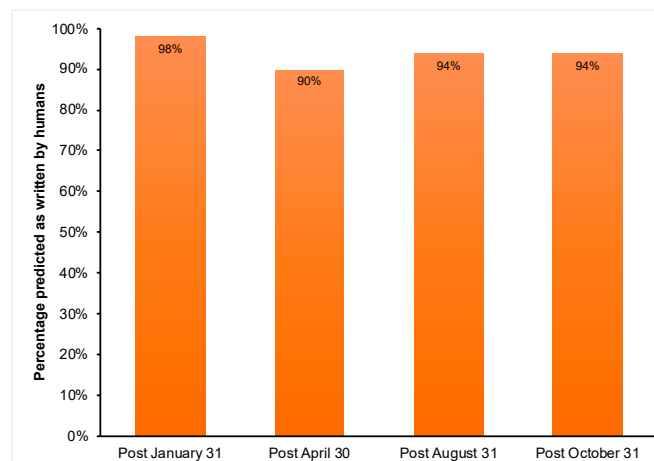


Figure 3. Percentage of articles predicted by trained model as written by humans after public release of ChatGPT.

4. Discussion

Although the method of utilizing machine learning and linguistic differences is not a new concept in the realm of AI text detection [6], this is the first example focusing on the field of student journalism. We found that collegiate journalists tended to use popular words such as “said”, “but”, and “this”, with a more diverse sequence of sentences and words used per sentence throughout the entire article compared to ChatGPT. For example, in the training set, the word “said” was used 285 times in 120 human articles, while it was used only 14 times in 120 articles from ChatGPT. The differences we observed are likely uniquely suitable for distinguishing this particular type of writing; the model should not be applied outside the domain of student journalism. Within this domain, though, the approach works well.

With the use of manually identified stylistic differences and an XGBoost-based classifier, the AUC of the training set was 0.933, with similar results on the test set (when the same prompt was used). We noted that this model’s performance was somewhat lower than a previous, similar study focusing on academic science papers [16]. However, the analysis task undertaken herein, comparing human-written documents to AI-edited documents, was more challenging than simply comparing purely human-written versus purely AI-generated documents, which was the task of the prior study. The ChatGPT-edited stories were often very similar to those produced by students (see Figure 4 for a comparative example). We chose the more challenging task of detecting edited documents for this project because this use of ChatGPT matches how students leverage the tool. In a survey of 94 student users of ChatGPT, researchers found that 47% of the students used it to paraphrase and 45% used it to summarize [17]. Far fewer students used it to write an entire assignment from scratch [17]. When considering the similarity of the writing examples, as shown in Figure 4, and the accuracy of the resulting model, this case study provides additional evidence that a classification strategy based on stylistic features proves valuable for detecting generative AI. By using a larger set of training data, putting more effort into feature engineering, and optimizing hyperparameters, the model’s accuracy could potentially be improved beyond what was demonstrated here.

<p>Now in its fourth year, the Campus Cupboard continues to serve food to students at the Kansas Union, ensuring that university members have access to nutritious food. The Campus Cupboard looks to decrease food insecurity on campus by supplying free, nutritional food to any student, employee or faculty member. The cupboard offers a range of fresh fruits and vegetables, nonperishables and hygiene items. The Campus Cupboard is located on the fourth floor of the Kansas Union and is open Monday-Friday, 9 a.m. to 5 p.m. Currently, the program serves about 400 students a month, graduate assistant for Campus Cupboard Alex Currier said, and the numbers are increasing as more people hear about it. The program was established in 2017 following research done by former student Ike Urie. The study found that about 54% of undergraduate students are considered food insecure. Food insecurity is described by the USDA as “reports of reduced quality, variety or desirability of diet,” and is still a prevalent issue on campus.</p>	<p>In its fourth year of operation, the Campus Cupboard remains committed to combating food insecurity among university members by providing access to nutritious food. Located on the fourth floor of the Kansas Union, the Campus Cupboard offers free, high-quality food to any student, employee, or faculty member in need. The initiative aims to alleviate food insecurity on campus by offering a wide range of options, including fresh fruits and vegetables, nonperishable items, and hygiene products. Open from Monday to Friday, 9 a.m. to 5 p.m., the Campus Cupboard ensures that individuals have a convenient and reliable resource to obtain essential food items. Currently, the program serves approximately 400 students per month, but the number is steadily increasing as more people become aware of its existence. According to Alex Currier, the graduate assistant for Campus Cupboard, the initiative was established in 2017 following research conducted by former student Ike Urie. The study revealed that 54% of undergraduate students experience food insecurity. Food insecurity, as defined by the USDA, refers to a lack of access to an adequate and nutritious diet. Unfortunately, it remains a significant issue on campus, making initiatives like the Campus Cupboard essential in addressing the needs of the university community.</p>
--	--

Figure 4. Text on the left depicts a section of a student-written article. Text on the right shows the ChatGPT-generated counterpart. Similarity of the two texts showcases AI-text generation ability. Human-written text is reproduced with permission of the University Daily Kansan.

Another unique aspect of this work is its application for testing articles written after the release of ChatGPT to test for undisclosed AI usage. We predicted that the percentage of articles predicted to be written by humans would decrease over time as students began to adopt the technology and use it in their journalism submissions. However, this hypothesis was not confirmed. The percentage of articles predicted to be written by humans was similar, and in some cases higher, than the accuracies obtained in the training and test sets, where all articles were verified to be human-written. Based on the data, it can be concluded that student journalists were most likely not using ChatGPT to generate news articles in 2023. If they were, their usage was probably far less than 10% of the time. It would be difficult to speculate about ChatGPT usage—or not—beyond this threshold, considering the accuracy of the model.

5. Conclusions

The study herein utilized classical machine-learning tools to develop a detection model that could accurately differentiate university newspaper articles that were solely human-written from those that leveraged ChatGPT to generate the final content. Thirteen stylistic differences—between newspaper articles written by humans or ChatGPT—were identified, and a supervised classification model using XGBoost and the identified features was developed. The model differentiated the two types of articles with acceptable performance and showed superiority to online detectors for this particular task. Finally, we used the tool to determine that university students from five different universities were not leveraging ChatGPT, to any appreciable extent, to produce the final copies of their news articles.

Limitations and future studies. The most important limiting consideration in this study was the fact that the types of prompts explored were limited. It is possible that students could have still used ChatGPT on their stories but evaded detection by this method if they had designed a prompt that would introduce a minimal number of edits. For example, if prompted with a statement like “make only necessary changes to correct the grammar in this finished story. . .”, we do not expect that this model would have prevailed with high accuracy. However, some may argue that such a use of ChatGPT in the writing process is not much different than using the grammar-editing functions already present in word-processing programs, so it is less important to detect this kind of usage.

Moving forward, this study provides a model for testing for the use of ChatGPT or other generative AI tools in documents that do not disclose their usage. We would anticipate

that future models would benefit from larger training sets, which should ultimately lead to a model with higher accuracy, so that the lower limits of detection of AI usage could be achieved. Researchers may also wish to restrict the variability within the human-generated content by setting tight boundaries on the content to be tested. This approach could be another way to boost the model's performance, since wider variability within the data set necessarily makes the classification challenge more difficult.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/info15060307/s1>: University newspapers.pdf (document of university newspaper sites and links, prompts included); Example Code.pdf (example code for extracting each feature building the feature matrix); matrix_B.csv (ChatGPT articles from training set, matrix).

Author Contributions: Conceptualization, M.-G.K. and H.D.; methodology, M.K and H.D.; software, M.-G.K. and H.D.; validation, M.-G.K.; formal analysis, M.-G.K.; data curation, M.-G.K.; writing – original draft preparation, M.-G.K.; writing – review and editing, M.-G.K. and H.D.; visualization, M.-G.K.; supervision, H.D.; project administration, H.D.; funding acquisition, H.D. All authors have read and consented to the final version of the manuscript.

Funding: This research was supported by funding from the Keith D. Wilner Professorship to H.D.

Data Availability Statement: The raw data and original code used in this article are provided in the Supplementary Material section.

Acknowledgments: Some of the code for extracting features was written with the assistance of ChatGPT. The authors take responsibility for the accuracy of this code.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. King, M.R. A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education. *Cel. Mol. Bioeng.* **2023**, *16*, 1–2. [[CrossRef](#)] [[PubMed](#)]
2. Abdaljaleel, M.; Barakat, M.; Alsanafi, M.; Salim, N.A.; Abazid, H.; Malaeb, D.; Mohammed, A.H.; Hassan, B.A.R.; Wayyes, A.M.; Farhan, S.S.; et al. A multinational study on the factors influencing university students' attitudes and usage of ChatGPT. *Sci. Rep.* **2024**, *14*, 1983. [[CrossRef](#)] [[PubMed](#)]
3. Ibrahim, H.; Liu, F.; Asim, R.; Battu, B.; Benabderrahmane, S.; Alhafni, B.; Adnan, W.; Alhanai, T.; AlShebli, B.; Baghdadi, R. Perception, performance, and detectability of conversational artificial intelligence across 32 university courses. *Sci. Rep.* **2024**, *13*, 12187. [[CrossRef](#)] [[PubMed](#)]
4. Alasadi, E.A.; Baiz, C.R. Generative AI in Education and Research: Opportunities, Concerns, and Solutions. *J. Chem. Educ.* **2023**, *100*, 2965–2971. [[CrossRef](#)]
5. Iskender, A. Holy or Unholy? Interview with OpenAIs ChatGPT. *Eur. J. Tour. Res.* **2023**, *34*, 3414. [[CrossRef](#)]
6. Mitrovic, S.; Andreoletti, D.; Ayoub, O. ChatGPT or Human? Detect and Explain. Explaining Decisions of Machine Learning Model for Detecting Short ChatGPT-generated text. *arXiv*, 2023; arXiv:2301.13852.
7. Cingillioglu, I. Detecting AI-generated essays: The ChatGPT challenge. *Emerald Insight* **2023**, *40*, 259–268. [[CrossRef](#)]
8. Desaire, H.; Chua, A.E.; Kim, M.G.; Hua, D. Accurately detecting AI text when ChatGPT is told to write like a chemist. *Cell Rep. Phys. Sci.* **2023**, *4*, 101672. [[CrossRef](#)] [[PubMed](#)]
9. Al-Smadi, M. ChatGPT and Beyond: The Generative AI Revolution in Education. *arXiv* **2023**, arXiv:2311.15198.
10. Lund, B.D.; Wang, T.; Mannuru, N.S.; Nie, B.; Shimray, S.; Wang, Z. ChatGPT and a New Academic Reality: AI-Written Research Papers and the Ethics of the Large Language Models in Scholarly Publishing. *JASIST* **2023**, *74*, 570–581. [[CrossRef](#)]
11. Bhattacharjee, A.; Liu, H. Fighting Fire with Fire: Can ChatGPT Detect AI-generated Text? *arXiv* **2023**, arXiv:2308.01284. [[CrossRef](#)]
12. Gao, C.A.; Howard, F.M.; Markov, N.S.; Dyer, E.C.; Ramesh, S.; Luo, Y.; Pearson, A.T. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *Nature* **2023**, *6*, 75. [[CrossRef](#)] [[PubMed](#)]
13. Liang, G.; Guerrero, J.; Zheng, F.; Alsmadi, I. Enhancing Neural Text Detector Robustness with μ Attacking and RR-Training. *Electronics* **2023**, *12*, 1948. [[CrossRef](#)]
14. Guo, B.; Zhang, X.; Wang, Z.; Jiang, M.; Nie, J.; Ding, Y.; Yue, J.; Wu, Y. How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection. *arXiv* **2023**, arXiv:2301.07597.
15. Uchendu, A.; Ma, Z.; Le, T.; Zhang, R.; Lee, D. TURINGBENCH: A benchmark environment for Turing test in the age of neutral text generation. *arXiv* **2021**, arXiv:2109.13296.

16. Desaire, H.; Chua, A.E.; Isom, M.; Jarosova, R.; Hua, D. Distinguishing academic science writing from humans or ChatGPT with over 99% accuracy using off-the-shelf machine learning tools. *Cell Rep. Phys. Sci.* **2023**, *4*, 101426. [[CrossRef](#)] [[PubMed](#)]
17. Patekar, J. Writing with AI: University Students' Use of ChatGPT. *J. Lang. Educ.* **2023**, *9*, 128–138. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.