Machine-Made Media: Monitoring the Mobilization of Machine-Generated Articles on Misinformation and Mainstream News Websites

Hans W. A. Hanley and Zakir Durumeric

Stanford University
hhanley@stanford.edu, zakird@stanford.edu

Abstract

With the increasing popularity of generative large language models (LLMs) like ChatGPT, an increasing number of news websites have begun utilizing them to generate articles. However, not only can these language models produce factually inaccurate articles on reputable websites but disreputable news sites can utilize these LLMs to mass produce misinformation. To begin to understand this phenomenon, we present one of the first large-scale studies of the prevalence of synthetic articles within online news media. To do this, we train a DeBERTa-based synthetic news detector and classify over 12.91 million articles from 3,074 misinformation and mainstream news websites. We find that between January 1, 2022 and April 1, 2023, the relative number of synthetic news articles increased by 79.4% on mainstream websites while increasing by 342% on misinformation sites. Analyzing the impact of the release of ChatGPT using an interrupted-time-series, we show that while its release resulted in a marked increase in synthetic articles on small sites as well as misinformation news websites, there was not a corresponding increase on large mainstream news websites. Finally, using data from the social media platform Reddit, we find that social media users interacted more with synthetic articles in March 2023 relative to January 2022.

1 Introduction

Since the release of ChatGPT on November 30, 2022, Internet users have used the large language model (LLM) to compose letters, write essays, and ask for advice. Now with over 100 million monthly active users (Hu 2023), misuse of this generative model has also been widespread. In one example, CNet, a reputable website that publishes reviews and news on consumer electronics, published articles generated by ChatGPT that were rife with errors (Leffer 2023). However, despite the reported widespread adoption, use, and misuse of ChatGPT and other LLMs for generating news articles, there has not been a study of how much the use of these technologies has increased or changed over time. Similarly, despite LLMs' potential use to spread misinformation (Tang, Chuang, and Hu 2023), it is still largely unclear whether fringe, misinformation, or otherwise unreliable websites utilize them in practice.

In this work, we present a large-scale study of the prevalence of *machine-generated/synthetic* articles sourced from across 3,074 news websites (1,142 misinformation/unreliable websites and 2,252 mainstream/reliable news websites) in order to measure the prevalence of synthetic content on mainstream/reliable and misinformation/unreliable news platforms. To do this, we utilize training data from 19 models, as well as adversarial data from article perturbation/re-writes and paraphrases, to train a DeBERTa-based (He et al. 2021) model to detect English-language synthetic news articles. Benchmarking this classifier across six test sets of machine-generated news articles, including two from real-world companies (Pu et al. 2022), our model achieves an average F-1 score of 0.988 across these datasets. Then, using this newly trained model, we classify over 12.91M articles published between January 1, 2022, and April 1, 2023 from 3,074 news websites.

We find that among reliable/mainstream news websites, synthetic articles increased in prevalence by 79.4% (9.4% of news articles in January 2022 to 16.8% April 2023) while among unreliable/misinformation websites, the prevalence increased by 342% (3.6% of news articles in January 2022 to 16.2% in April 2023). Examining the content of synthetic articles, we find that while mainstream/reliable news websites have largely utilized synthetic articles to report on financial news, COVID-19 statistics, and sports, misinformation/unreliable news websites have reported on a wide range of topics including US Presidential politics and the Russo-Ukrainian War.

Estimating the impact of the release of ChatGPT on the prevalence of synthetic articles, we find that while large mainstream websites did not see a significant (0.19% absolute increase) in synthetic articles, there was an average absolute increase of 2.56% in the percentage of synthetic articles on misinformation websites. Finally, utilizing data from the social media platform Reddit, we analyze users' interactions with synthetic news articles. Between January 1, 2022 and March 31, 2023, we observe a 67.8% increase in the number of synthetic mainstream articles posted to Reddit and a 131% increase in the number of synthetic misinformation news articles. This corresponded with a 281% increase in the number of comments on Reddit submissions featuring a synthetic mainstream article and a 631% increase in the number of comments on Reddit submissions featuring a synthetic misinformation article.

Our work presents one of the first in-depth analyses of the growth of synthetic articles across the news ecosystem. We show that throughout 2022, particularly after the release of ChatGPT, misinformation websites have rapidly increased the amount of synthetic content on their websites. As misinformation websites increasingly utilize synthetic articles, we hope that our work can serve as the basis for helping to identify the misuse of LLMs and for future studies on the spread of misinformation.

2 Background and Related Work

Recent advances in large language models (LLMs) have resulted in impressive performance on a variety of tasks, most notably convincing text generation (Brown et al. 2020; Chowdhery et al. 2022; AI 2022; Zellers et al. 2019). However, despite the their popularity, widespread availability of LLMs can also be problematic. Zeller et al. (2019) perhaps first showed that GPT-2 can create convincing articles that evoke more trust than human-written articles, within the past year, models like Open AI's ChatGPT, Meta's LLaMa, and Google's Bard have largely democratized their use.

Real-World Use of Machine-Generated News Media. While the large-scale democratization of the generative models is new, the use of machine-generated or *synthetic* articles by news websites is not. Since as early as 2019, Bloomberg has used the service Cyborg to automate the creation of nearly one-third of their articles (Peiser 2019). Similarly, since 2019, as reported by the New York Times, other reputable news sources including The Associated Press, The Washington Post, and The Los Angeles Times, have used machine-generation services to write articles on topics that range from minor league baseball to earthquakes (Peiser 2019). However, articles that contain machine-generated content from services such as Cyborg, BERTie, or ChatGPT, while reducing the workload of reporters, have been shown to often contain factual errors (Alba 2023; Leffer 2023). As a result, much research has focused on detecting machine-generated news articles (Zellers et al. 2019; Uchendu et al. 2020; He et al. 2023; Ippolito et al. 2020).

Detecting Machine-Generated Media. Several approaches have been used to detect machine-generated text. BERT-defense (Ippolito et al. 2020) for instance uses a BERT-based (Devlin et al. 2019) model to identify machine-generated texts. DetectGPT (Mitchell et al. 2023) approximates the probabilistic curvature of specific LLMs for zero-shot detection. Mitchell et al. show that if the specific model that is used to generate text is known and can be readily queried to obtain the log probabilities of pieces of text, then it is possible to easily differentiate synthetic articles from human-written news articles. Zhong et al. (2020) propose a graph-based approach that considers the factual structure of articles to detect machine-generated text.

Our work depends on accurately identifying machine-generated articles across news websites. As shown in previous works, however, many machine learning models trained to detect synthetic texts overfit to their training domain, the token distribution of the model used to generate the synthetic texts, and the topics that they were trained on (Mitchell et al. 2023; Uchendu et al. 2020; Lin, Hilton, and Evans 2022). For example, models trained to detect synthetic news articles, often fail to detect shorter machine-generated tweets. Despite these shortcomings, as illustrated by Pu et al. (2022), classifiers focused on only one domain can often perform exceedingly well on datasets seen "in-the-wild." Adversarially training a RoBERTa (Liu et al. 2019) based classifier, Pu et al. achieve an F1 classification score of 87.4–91.4 on a test dataset made up of synthetic news articles purchased from AI Forger and Article Forge. Unlike in other domains, such as tweets or comments, news articles tend to be longer, allowing for greater precision in their classification (Pu et al. 2022).

Reliable and Unreliable News Websites. In this work, we analyze how both reliable/mainstream and unreliable/misinformation news websites have published machine-generated articles throughout 2022 and 2023. Unreliable information from these sites can take the form of *misinformation*, *disinformation*, and *propaganda*, among others (Jack 2017). Within this work, we refer to websites that have been labeled by other researchers as generally spreading *false* information as misinformation/unreliable news websites, including both websites labeled as *misinformation* and *disinformation* within this label. As in prior work, we consider reliable/mainstream news websites as "outlets that generally adhere to journalistic norms including attributing authors and correcting errors; altogether publishing mostly true information" (Hounsel et al. 2020).

3 Detecting Machine-Generated Articles

As described in Section 2, several approaches have been developed for identifying synthetic articles, with some of the most successful being transformer-based methodologies (Pu et al. 2022; Gehrmann, Strobelt, and Rush 2019). However, given that past models were trained to detect text from *particular* models (Zellers et al. 2019), are vulnerable to adversarial attacks (Pu et al. 2022), or have unreleased weights (Zhong et al. 2020), we design and benchmark our *own* transformer-based machine-learning classifiers to identify synthetic articles in the wild. We will release the weights of these models at the time of publication.

In addition to training three transformer architectures (BERT, RoBERTa, DeBERTa) on a baseline training dataset (detailed below), we further train these models on datasets generated from two common adversarial attacks (Krishna et al. 2023; Mitchell et al. 2023). Adding to our own benchmarks, we test our new models against datasets of articles generated by two companies, AI Writer and AI Forger, provided to us by Pu et al. (2022). We now describe our training and test datasets, the architectures of our models, and finally our models' performances on our benchmarks.

Baseline Training Datasets. To train a classifier to detect machine-generated/synthetic news articles found in the wild, we require a diverse dataset of articles from a wide array of generative models. Thus, for our baseline training dataset, we take training data of machine-generated/synthetic articles from three primary sources: the Turing Benchmark, Grover, and articles generated from GPT-3.5.

Machine-Generated Training Articles: For much of our training data, we utilize the Turing Benchmark (Uchendu et al. 2021), which contains news articles generated by 10 different generative text architectures including GPT-1 (Radford et al. 2019a), GPT-2 (Radford et al. 2019b), GPT-3 (Brown et al. 2020), CTRL (Keskar et al. 2019), XLM (Lample and Conneau 2019), Grover (Zellers et al. 2019), XLNet (Yang et al. 2019), Transformer-XL (Dai et al. 2019), and FAIR/WMT (Ng et al. 2019; Chen et al. 2020). We note that given the different settings and trained weights provided by the authors of these respective works, the Turing Benchmark altogether includes articles generated from 19 different models. We randomly subselect 1000 articles generated by each of these different models and within the Turing Benchmark as training data.

In addition to the Turing Benchmark training dataset, we use the training dataset of Zellers et al. (2019), which contains realistic, often long-form articles, that mimic the fashion of popular news websites such as cnn.com, nytimes.com, and the washingtonpost.com. Unlike the Grover-generated articles from the Turing Benchmark dataset, which are generated using a prompt of just the title of potential articles, these Grover articles are generated in an unconditional setting and from prompting the Grover model with metadata (*i.e.*, title, author, date, website). As found by Zellers et al., many of the articles produced by their models were convincing to human readers, and we thus include 11,930 machine-generated articles from the base model of Grover (across different Grover decoding settings [*e.g.*, p=1.00, p= 0.96, p=0.92 (nucleus/top-p), k=40 (top-k), *etc*... settings]) in our training dataset.

Finally, given the popularity of the GPT-3.5 model (Hu 2023), with it being the basis of the released version of ChatGPT, and GPT-3.5 being one of the most powerful released models, we add 3,516 articles generated from the `GPT-3.5 davinici` model. To create these articles, we prompt the public API of `GPT-3.5 davinici` with the first 10 words of 3,516 real news articles from 2018(see Section 4; while scraping our news dataset, we acquired several million articles from before 2019). For `GPT-3.5 davinici` model, we use a nucleus decoding setting of p=1.00, p=0.96, and p= 0.92 (some of the most common (Mitchell et al. 2023; Zellers et al. 2019)).

We finally note that as found in prior work (Pu et al. 2022; Uchendu et al. 2021; Zellers et al. 2019), machine-generated news articles are often shorter in length than human-written articles. While training, to ensure that our models do not simply distinguish between longer human-written articles and those generated by generative transformers, we ensure that our machine-generated and human-written articles are of similar lengths (median training synthetic article length of 210 words and median training human article length of 224 words). Similarly, as found by past work, predictions for texts, particularly short texts, tend to be unreliable (Kirchner et al. 2023; Zellers et al. 2019; Pu et al. 2022). As such, for our training and our generated test data (GPT 3.5 dataset), we exclude texts shorter than 1,000 characters (140 words). As a result, we do not use every trained model's articles from the Turing Benchmark. Given that WMT-20/FAIR articles within this dataset are all shorter than 1000 characters, we do not include them within our training dataset. Altogether our training dataset thus includes data from 19 different models (18 from Turing Benchmark and `GPT-3.5 davinici`).

Dataset	**Human Written**	**Machine Generated**
Baseline	33,446	33,446
Pert.	33,446	44,003
Para.	33,446	41,498
Perturb + Para.	33,446	52,055

Table 1: The number of machine-generated and human-written articles within the `Baseline`, `Pert`, `Para`, and `Pert.+Para.` training datasets.

Dataset	**Human Written**	**Machine Generated**
Turing Benchmark	975	18,076
GPT-3.5	1,000	243
GPT-3.5 w/ Pert.	1,000	241
GPT-3.5 w/ Para.	1,000	118
Article Forger	1,000	1,000
AI Writer	1,000	1,000

Table 2: The number of machine-generated and human-written articles within our test datasets.

Human-Written Training Articles: For our set of human-generated articles, as in Zellers et al. (2019), we utilize news articles published before 2019. Specifically, we use 28,446 articles from 2018 from our set of news websites that we later measure (see Section 4; while scraping our news dataset, we acquired several million articles from before 2019), 2,500 articles from the human split of the Grover dataset, and 2,500 articles from the human-train-split within the Turing Benchmark dataset. We present an overview of this complete dataset as the `Baseline` in Table 1.

Baseline Test Datasets. For our test datasets, we use the test splits of the original Grover dataset, the validation split from the Turing Benchmark (the labels from the test split were unavailable to us), and another test dataset consisting of 243 additional GPT-3.5 articles that we created by again prompting `GPT-3.5 davinici`, and 1000 human-written articles from 2018 (see Section 4; as with our training data, while scraping our news dataset, we acquired several million articles from before 2019). Furthermore, to ensure our models generalize and handle articles seen in the wild, we utilize the *In-the-Wild* dataset provided to us by Pu et al. (2022). This dataset consists of news articles created using generative large language models from two independent companies, Article Forger and AI Writer. By testing against these outside datasets, we further validate our approach against articles generated by (1) models not within our dataset and (2) by generative news article services available to the public. We provide details in Table 2.

Training and Test Dataset using Perturbations and Paraphrases. Transformer-based classifiers are often particularly susceptible to adversarial attacks, particularly attacks that rewrite sections of the generated article (Mitchell et al. 2023; Pu et al. 2022) and paraphrase attacks (*i.e.*, where a generic

	Turing Benchmark			GPT-3.5			GPT-3.5 w/ Pert			GPT-3.5 w/ Para			Article Forger			AI Writer			Avg.
	F1	Prec.	Recall	F1	Prec.	Recall	F1	Prec.	Recall	F1	Prec.	Recall	F1	Prec.	Recall	F1	Prec.	Recall	F1
OpenAI Roberta	0.717	0.997	0.560	0.092	0.684	0.049	0.022	0.375	0.011	0.302	0.949	0.180	0.888	1.000	0.798	0.748	1.000	0.598	0.462
BERT	0.995	0.996	0.993	0.954	0.928	0.981	0.921	0.941	0.902	0.981	0.967	0.995	0.830	0.737	0.951	0.830	0.851	0.810	0.919
BERT+ Pert	0.996	0.995	0.996	0.929	0.871	0.996	0.936	0.899	0.977	0.952	0.907	1.000	0.806	0.680	0.991	0.829	0.756	0.917	0.908
BERT + Para	0.994	0.993	0.995	0.926	0.881	0.977	0.912	0.880	0.947	0.932	0.873	1.000	0.803	0.677	0.985	0.860	0.808	0.919	0.905
BERT+Para+Para.	0.991	0.995	0.996	0.949	0.907	0.996	0.944	0.906	0.985	0.956	0.916	1.000	0.781	0.664	0.951	0.840	0.785	0.903	0.910
RoBERTa	**0.997**	0.996	0.998	0.956	0.926	0.989	0.945	0.940	0.951	0.976	0.954	1.000	0.890	0.837	0.950	0.945	0.903	0.991	0.952
RoBERTa + Pert.	0.994	0.999	0.989	0.973	0.984	0.962	**0.979**	0.977	0.981	0.978	0.990	0.966	0.840	0.980	0.735	0.895	0.986	0.819	0.943
RoBERTa + Para	0.996	0.999	0.994	0.965	0.949	0.981	0.941	0.950	0.932	0.988	0.981	0.995	0.939	0.901	0.979	0.917	0.877	0.961	0.958
RoBERTa+Pert.+Para.	0.997	**0.999**	0.994	0.970	0.949	0.992	0.972	0.954	0.992	0.993	0.990	0.995	0.918	0.928	0.908	0.945	0.955	0.935	0.965
DeBERTa	0.995	0.998	0.992	0.959	0.935	0.985	0.963	0.949	0.977	0.976	0.954	1.000	0.976	0.958	**0.995**	0.990	0.982	0.998	0.977
DeBERTa + Pert.	0.994	0.991	**0.998**	0.941	0.894	0.992	0.962	0.926	**1.000**	0.958	0.955	**0.998**	0.950	0.913	0.989	0.976	0.955	**0.998**	0.964
DeBERTa + Para.	0.990	0.999	0.983	0.977	**0.996**	0.958	0.946	**1.000**	0.898	0.995	0.995	0.995	0.907	**0.996**	0.833	0.954	**1.000**	0.912	0.962
DeBERTa+Pert.+Para.	0.995	0.998	0.992	**0.985**	0.981	0.989	0.978	0.963	0.992	**0.995**	**0.995**	0.995	**0.981**	0.982	0.980	**0.992**	0.991	0.993	**0.988**

Table 3: F1-Score/Precision/Recall of models on various benchmarks. We bold the best score in each column. As seen, the DeBERTa model performs the best across many of the test datasets, with `DeBERTa+Pert+Para` having the highest average F-1 score across all six datasets.

transformer model is used to paraphrase the output of a different generative model (Krishna et al. 2023)). To guard against these weaknesses, we take two approaches (1) perturbing our set of synthetic articles by rewriting at least 25% of their content using the generic T5-1.1-XL model[1] and (2) paraphrasing each article with the T5-based Dipper model.[2]

Constructing Perturbed Synthetic Articles. To perturb/rewrite sections of our machine-generated articles, as in Mitchell et al. (2023), we randomly MASK 5-word spans of text until at least 25% of the words in the article are masked. Then, using the text-to-text generative model T5-3B (Raffel et al. 2020), we fill in these spans, perturbing our original generated articles. As shown by Mitchell et al. (2023), large generic generative models such as T5 can apply perturbations that roughly capture meaningful variations of the original passage rather than arbitrary edits. This enables us to model divergences from the distributions of texts created by our 19 different generative models (18 from Turing Benchmark and GPT-3.5). We thus utilize T5-3B[3] to perturb the machine-generated articles of our `Baseline` train dataset. In addition, we create a separate test dataset by perturbing our GPT-3.5 test dataset. We note that after perturbing our datasets, we again filter to ensure all articles used for training contain at least 1000 characters. We annotate training and test datasets containing articles perturbed with T5-3B with the suffix `Pert`.

Constructing Paraphrased Synthetic Articles. To paraphrase each of the machine-generated articles within our dataset, we use the approach outlined by Krishna et al. (Krishna et al. 2023). Specifically, as in their work, we utilize Dipper, a version of the T5 generative model fine-tuned on paragraph-level paraphrases, that outputs paraphrased versions of the inputted text. We use the default and recommended parameters[4] as in Krishna et al. to paraphrase the text within our original training dataset as well as our GPT-3.5 test dataset (Krishna et al. 2023). We note that after perturbing our datasets, we again filter to ensure all articles utilized for training contain at least 1000 characters. We annotate training and test datasets containing articles paraphrased with T5-3B with the suffix `Para`.

Detection Models. Having described our training test sets, we now detail our models and evaluate their performance on our 6 test datasets (Turing Benchmark, GPT-3.5, GPT-3.5 w/ `Pert`, GPT-3.5 w/ `Para`, Article Forger, AI Writer). Specifically, we fine-tune three pre-trained transformers, BERT-base (Devlin et al. 2019), RoBERTa-base (Liu et al. 2019), and DeBERTa-v3-base (He et al. 2021)[5,6,7]. For each architecture, we train 4 models to detect machine-generated news articles using our `Baseline`, `Perturb`, `Para`, and `Perturb+Para` training datasets. For each architecture, we build a classifier by training an MLP/binary classification layer on top of the outputted [CLS] token. We use a max token length of 512 (Ippolito et al. 2020; Pu et al. 2022), a batch size of 32, and a learning rate of 1×10^{-5}. Each model took approximately 2 hours to train using an Nvidia RTX A6000 GPU. After training, as in Pu et al. (2022), we determine each model's F1-scores, precision, and recall for each test dataset, and rank each model using its average F-1 score. For a baseline comparison for our trained models, we further test the Roberta-based classifier released by Open AI in 2019 (Solaiman et al. 2019) on each of our test datasets.

Likely due to training our model on synthetic articles from a wide variety of sources, and our model's focus on news articles, as seen in Table 3, we observe that all our trained models perform markedly better than Open AI's 2019 released detection model. We further observe, as aggregated in the `Avg. F1` score column, our set of DeBerta models performs the best in classifying machine-generated/synthetic content, all achieving an average F1 score greater than

[1] https://huggingface.co/google/t5-v1_1-large

[2] https://huggingface.co/kalpeshk2011/dipper-paraphraser-xxl

[3] https://huggingface.co/google/t5-v1_1-large

[4] We use a lexical diversity parameter of 60. For more details on the Dipper model see Krishna et al. (Krishna et al. 2023)

[5] https://huggingface.co/bert-base-uncased

[6] https://huggingface.co/roberta-base

[7] https://huggingface.co/microsoft/deberta-v3-base

0.962. In particular, we observe that our DeBERTa model trained on a dataset that includes our set of adversarial data `Pert + Para`, performs the best at an average F-score of 0.988. This particular model further achieves the best respective F1 scores in classifying the set of articles from Article Forger and AI-writer provided by Pu et al., achieving F1 scores of 0.981 and 0.992 on the two datasets respectively. We note that our model, in addition to performing better than Open AI's Roberta, also outperforms all models benchmarked by Pu et al. (2022) on AI Writer and the AI Forger test datasets, which achieved F-1 scores ranging from 1.6 to 94.9. This illustrates that our model generalizes to other types of machine-generated articles from models not included in our dataset. Given its performance across all six of our datasets, we use our `DeBERTa+Pert+Para` trained model as our detection model for the rest of this work.

4 News Dataset and Classification Pipeline

Having described the DeBERTa-based model that we use to identify machine-generated/synthetic articles, we now describe our datasets of scraped news articles.

Website List. We gather articles published between January 1, 2022 and April 1, 2023, from 3,074 news websites. Our list of websites consists of domains labeled as "news" by Media Bias Fact Check[8] and by prior work (Hanley, Kumar, and Durumeric 2023). Within our list of news sites, we differentiate between "unreliable news websites" and "reliable news websites." Our list of unreliable news websites includes 1,142 domains labeled as "conspiracy/pseudoscience" by mediabiasfactcheck.com as well as those labeled as "unreliable news", misinformation, or disinformation by prior work (Hanley, Kumar, and Durumeric 2023; Barret Golding 2022; Szpakowski 2020). Our set of "unreliable" or misinformation news websites includes websites like realjewnews.com, davidduke.com, thegatewaypundit.com, and breitbart.com. Our set of "reliable" news websites consists of the remaining 2,552 news websites that were labeled as belonging to the "center", "center-left", or "center-right" by Media Bias Fact Check as well as websites labeled as "reliable" or "mainstream" by other works (Hanley, Kumar, and Durumeric 2023; Barret Golding 2022; Szpakowski 2020). This set of "reliable news websites" includes websites like washingtonpost.com, reuters.com, apnews.com, cnn.com, and foxnews.com.

We note that to later understand how websites of varying popularity/size have used machine-generated articles on their websites, we striate our list of websites by their popularity using ranking data provided by the Google Chrome User Report (CrUX) (Ruth et al. 2022). We note that the CrUX dataset, rather than providing individual popularity ranks for each website, instead provides rank order magnitude buckets (*e.g.*, top 10K, 100K, 1M, 10M websites). As such, we analyze our set of websites in the following buckets: Rank $<$ 10K (126 websites), 10K $<$ Rank $<$ 100K (511 websites), 100K $<$ Rank $<$ 1M (1,164 websites), 1M $<$ Rank $<$ 10M (802 websites), and finally Rank $>$ 10M (471 websites).

[8]https://mediabiasfactcheck.com/

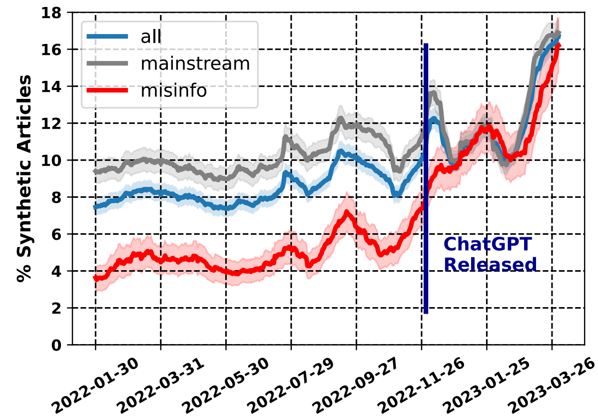

Figure 1: The average percentage of synthetic articles for all, misinformation, and mainstream websites. We provide 95% Normal confidence intervals.

Article Collection. To collect the articles published by news websites, we queried sites' RSS feeds (if available) and scraped the homepages of each website daily from January 1, 2022 to April 1, 2023. Upon identifying newly published articles, we subsequently crawled websites using Colly[9] and Headless Chrome, orchestrated with Python Selenium. To extract article text from each HTML page, we parsed the scraped HTML using the Python libraries `newspaper3k` and `htmldate`.

Given that many of our websites (*e.g.*, cnn.com) have multilingual options, we use the Python `langdetect` library to filter out all non-English articles. Further, to ensure the reliability of our classifications, we only classify news articles that are at least 1000 characters (approximately 140 words) long. Altogether, from our selection of 3,074 websites, we gathered 12.91M articles that were published between January 1, 2022, and April 1, 2023. Finally, we utilize our `DeBERTa+Pert+Para` model to classify each article as either human-written or machine-generated. Classifying all 12.91M articles took approximately 53.4 hours using an Nvidia RTX A6000 GPU.

Ethical Considerations. We collect only publicly available data from our set of websites. We further follow the best practices for web crawling as in Acar et al. (2014).

5 The Use of Machine-Generated Media

Having described our detection model and datasets, in this section, we analyze the changing levels of synthetic content across our set of websites between January 1, 2022 and April 1, 2023. Specifically, we determine (1) whether there has been an increase in the use of synthetic articles, (2) if there has been an increase in their use, which sets of websites are driving this increase, (3) what synthetic articles are topically about, and (4) whether the introduction of ChatGPT has had an effect on the prevalence of synthetic articles.

Large-Scale Trends in Machine-Generated Media. To begin, we plot the average percentage of synthetic news arti-

[9]https://github.com/gocolly/colly

Snippet from Reuters The S&P 500 (.SPX) and Nasdaq (.IXIC) added to losses, while the Dow (.DJI) turned negative on Wednesday after the release of the latest FOMC meeting minutes showed that officials said the central bank may need to raise interest rates sooner than expected and reduce asset holdings quickly.

Figure 2: Example first paragraph of an article classified by our system as machine-generated/synthetic.

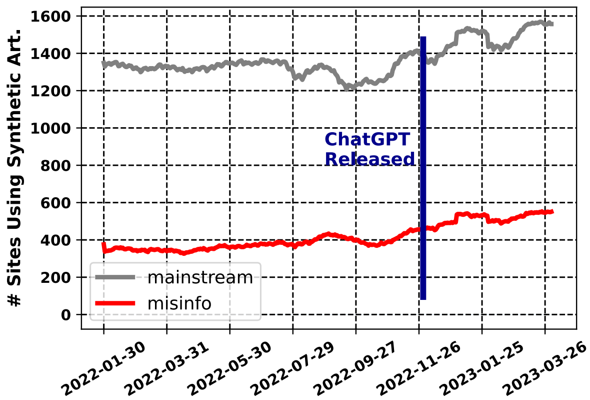

Figure 3: The number of websites that publish at least one synthetic article over a 30-day time span.

cles per website across our dataset between January 1, 2022 and April 1, 2023, in Figure 1. Across all 3,074 sites, we see that 7.9% of articles (101,046 of 1,272,491) published in January 2022 were synthetically generated. However, by March 2023, the fraction of synthetic articles nearly doubled to 16.3% (173,822 of 1,064,353 articles).

We observe that our set of "reliable" websites typically had a greater percentage of synthetic articles at the beginning of 2022 compared with "misinformation" or unreliable news websites. On average, while only 3.6% of articles from our set of misinformation websites were classified as machine-generated in January 2022, 9.4% of articles from our set of mainstream/reliable websites were machine-generated. This result is consistent with prior observations that since 2019 many news websites have begun to use automated services to write quick, often financial-related articles (Section 2). For example, the beginning of one of the articles from Reuters (Figure 2) classified by our system as being machine-generated simply contained simple information about the direction of particular markets and funds.

However, despite reliable/mainstream websites initially having higher levels of synthetic text, misinformation websites had marked increases in levels of machine-generated content during 2022 and 2023 (Figure 4). While between January 1, 2022, and April 1, 2023, reliable/mainstream news websites had a 79.4% relative increase (7.4% absolute percentage increase) in the levels of synthetic content, misinformation websites had a 342% relative increase (12.3% absolute percentage increase). Starting from a lower base, we thus see a substantial increase in the number of synthetic articles from unreliable/misinformation websites.

Similarly, as seen in Figure 3, we further observe that throughout the same period of time, an increasing number of news outlets published a synthetic article within any given

Rank	Misinfo Abs. %	Main. Abs.%
All	+12.3%	+7.44%
Rank < 10K	+6.26%	+5.79%
10K $<$ Rank < 100K	+10.8%	+5.86%
100K $<$ Rank < 1M	+8.20%	+6.61%
1M $<$ Rank < 10M	+12.5%	+11.5%
Rank > 10M	+13.1%	+16.5%

Table 4: Total absolute percentage increase in machine-generated/synthetic articles between January 1, 2022 and April 1, 2023.

Jan. 2022	**% Syn.**	**CrUX Rank**	**Mar. 2023**	**% Syn.**	**CrUX Rank**
regated.com	67.0%	$>$ 10M+	foreignpolicyi.org	80.6%	$<$10M
egyptianstreets.com	61.9%	$<$ 1M	patriotnewsdaily.com	69.9%	$<$10M
theragingpatriot.org	61.0%	$>$ 10M+	ufoworldnews.com	69.3%	$<$10M
leaderpost.com	60.7%	$<$ 1M	thefrisky.com	64.3%	$<$1M
edmontonsun.com	50.0%	$<$ 1M	thelist.com	60.8%	$<$100K
nationalpost.com	44.8%	$<$ 1M	theunionjournal.com	59.8%	$<$10M
edmontonjournal.com	44.0%	$<$ 1M	egypttoday.com	59.0%	$<$1M
thejakartapost.com	43.0%	$<$ 1M	newsfactsnetwork.com	57.9%	$<$ 10M
blackpressusa.com	42.8%	$<$ 1M	edmontonsun.com	57.9%	$<$1M
thequint.com	42.4%	$<$ 100K	cheddar.com	57.7%	$<$10M

Table 5: Websites with the largest percentage of synthetic content in January 2022 and in March 2023.

30-day time frame. Across our period of study, the number of mainstream websites that published at least one synthetic article increased from 1,340 (52.5% of mainstream websites) in January 2022 to 1,560 (61.1%) in March 2023. Similarly, the number of misinformation websites that published at least one synthetic article increased from 340 (29.8% of misinformation websites) to 550 (48.2%). Together, these results confirm that there *has* been a noted increase in the use of synthetic content generation by our set of news websites in 2022 and 2023.

Trends Among Popular and Unpopular Websites. To understand how popularity and website size correlated with the near doubling of machine-generated content in 2022 and 2023, we plot the percentage of machine-generated/synthetic articles over time in Figure 4 for websites within different rank buckets and striated by whether they are considered "unreliable" or "reliable." As seen in Figure 4, there is a general increase in the amount of machine-generated articles across every popularity stratum.

Examining these increases within particular brackets of popularity, we see (as pictured in Figure 4 and calculated in Table 4) that the least popular websites saw the largest percentage increase in the use of synthetic articles. For both unreliable/misinformation and reliable/mainstream categories, we observe that for websites that rank below 10 million in popularity, the percentage of their articles that were synthetic increased by 13.1% and 16.5% on average, respectively. By contrast, among the most popular misinformation/unreliable websites (*e.g.*, breitbart.com, zerohedge.com) and mainstream/reliable websites (*e.g.*, cnn.com, foxnews.com), synthetic articles enjoyed only a 6.26% and a 5.79% increase overall. Indeed, calculating the absolute percentage increases in machine-generated content, we again observe in Table 5 that the websites that had the largest increases in synthetic content were all small (only

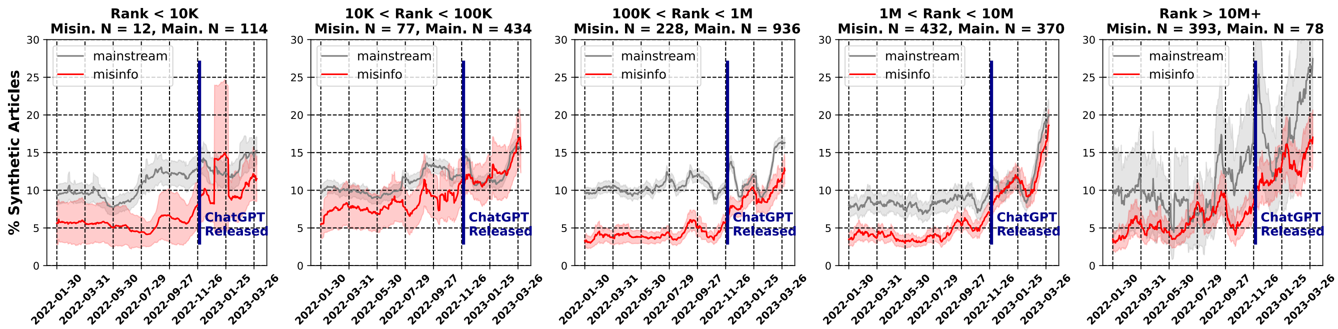

Figure 4: The average percentage of machine-generated/synthetic articles for misinformation/unreliable and mainstream/reliable news websites at different striations of popularity according to Google Chrome User Report (CrUX) from October 2022. All striations of misinformation websites experienced an uptick of machine-generated content around November 30, 2022, the release date of Open AI's ChatGPT.

investing.com in the top 10K websites).

Topics Addressed by Synthetic Articles. While misinformation websites and less popular websites have seen the largest increase in the use of synthetic articles, many reliable and large news websites also heavily use synthetic articles. However, as noted in Section 2, many reliable news sites have acknowledged their use of these machine-generated articles and utilize them in a benign manner. To understand different websites' use of synthetic articles, in this section, we analyze topics addressed by synthetic articles among different types of sites and how this has changed between January 2022 and March 2023.

To identify the key topics within our identified set of machine-generated articles, we apply the topic methodology outlined by Hanley and Durumeric (2023). For each article in our dataset, we embed its constituent paragraphs in a shared embedding space with an MPNet model (Song et al. 2020) fine-tuned for use on semantic similarity tasks. Then utilizing an optimized version of the DP-Means algorithm, we cluster these embeddings to identify topic clusters. Finally, we extract keywords from these topic clusters for human interpretability using the NMPI metric. We use the default settings recommended by Hanley and Durumeric (2023). Clustering the articles of each ranking stratum and bifurcating by whether the website considered misinformation/unreliable or mainstream/reliable news (*e.g.*, articles from websites with rank within each popularity bucket that are misinformation/unreliable or reliable/mainstream news), we determine the top two topics in January 2022 (first month of our study) and in March 2023 (last full month of our study). We do so to determine how the top topics changed between the beginning and end of our study.

We observe that among our mainstream websites, the most prevalent topics discussed in synthetic articles in January 2022 across all striations of popularity concerned either (1) basic statistics about case counts of COVID-19 or (2) information about stocks, the economy, or market trends (Table 6). In contrast, among our misinformation websites, we identify a variety of topics including sports, concerns about tensions between Russia and Ukraine, China, updates about covid cases, and lottery ticket winners. Among the most commonly addressed topics, we thus observe a wider range of topics on misinformation/unreliable news websites.

Examining the set of topics addressed by mainstream websites in March 2023 (Table 6), we see a focus on crime, market trends, and sports (NCAA tournament). However, we note that among less popular news websites, we also identify several articles about former US President Donald Trump's arrest in Manhattan, New York. Among our set of misinformation websites, we find that the major topics included articles about President Trump's arrest, US Florida Governor Ron DeSantis' potential run for the US Presidency, and updates about the Russo-Ukrainian War. Finally, among the least popular misinformation websites, we identify several "scam" articles on potential dating opportunities and payday loans. We thus again see in March 2023 that while many popular mainstream websites have continued to use synthetic content for articles on finance and sports, misinformation websites have used synthetic articles for a wide variety of topics including political news.

Estimating the Impact of ChatGPT. As seen in the previous sections, misinformation websites and less popular websites saw the largest increases in the use of synthetic articles. In order to estimate how the introduction of ChatGPT specifically may have affected the levels of synthetic content on news websites, we utilize an ARIMA model (Zhang 2003) to perform an *interrrupted-time-series* analysis. Namely, we examine whether there was a direct jump in the number of synthetic articles above expectation following the release of ChatGPT on November 30, 2022 (OpenAI 2022).

As seen in Table 7, after the release of ChatGPT on November 30, 2022, while we do not see a noticeable increase above expectation in the amount of machine-generated articles among mainstream/reliable news websites, we *do* see a noted jump (2.56%) in the number of synthetic articles from misinformation websites. Every popularity ranking bracket of misinformation websites saw a statistically significant increase in the absolute percentage of their articles that were synthetic, with misinformation websites in the Rank $<$ 10K popularity bracket seeing the highest jump of 4.85%. This was visually seen in Figure 4. We further observe that each popularity stratum of websites saw the rate at which the percentage of synthetic articles increases, also increase (*i.e.*, increase in the rate of increase).

Jan. 2022, Rank $< 10K$

Misinfo. Topics	# Art.	Mainstream. Topics	# Art.
zelensky, biden, dr,	135	covid19, january, cases	854
djokovic, australia, novak	132	watchlist,nasdaq,s&p,	376

Jan. 2022, $10K < $ Rank $< 100K$

Misinfo. Topics	# Art.	Mainstream. Topics	# Art.
ukraine, russia, nato,	434	commerical,kg, commodity	1,491
lotto, win, number	193	yield, kg, cattle	1,167

Jan. 2022, $100K <$ Rank $< 1M$

Misinfo. Topics	# Art.	Mainstream. Topics	# Art.
democrat, trump, republican,	82	covid19,dashboard,sick	613
trump, january, capitol	60	price, watch, invest	533

Jan. 2022, $1M <$ Rank $< 10M$

Misinfo. Topics	# Art.	Mainstream. Topics	# Art.
covid, pfzier, american,	59	astrazeneca, pfizer, mild	72
china, beijing, ccp	36	january, capitol, trump	42

Jan. 2022, Rank $> 10M$

Misinfo. Topics	# Art.	Mainstream. Topics	# Art.
covid19, dr, biden,	101	biden, trump, republican	27
senate, republican, rep	73	bluetooth, bose, headphones	23

March 2023, Rank $< 10K$

Misinfo. Topics	# Art.	Mainstream. Topics	# Art.
zelensky, kremlin, prigozhin,	237	market, watch, stock	2,229
football, game, score	231	degree,murder, sentence	1,193

March 2023, $10K <$ Rank $< 100K$

Misinfo. Topics	# Art.	Mainstream. Topics	# Art.
russia, ukraine, putin,	298	suspect,custody, arrest	1,943
biden, desantis, washington	267	corp, llc, nasdaq	1,785

March 2023, $100K <$ Rank $< 1M$

Misinfo. Topics	# Art.	Mainstream. Topics	# Art.
trump, biden, fox,	277	murder, police, sentence	3,003
trump, fbi, bragg	226	ncaa, basketball, carolina	2,670

March 2023, $1M <$ Rank $< 10M$

Misinfo. Topics	# Art.	Mainstream. Topics	# Art.
ukraine, russia,putin,	201	trump, manhattan, bragg	213
trump, democrats, fbi	159	covid, uk, pfizer	208

March 2023, Rank $> 10M$

Misinfo. Topics	# Art.	Mainstream. Topics	# Art.
craigslist, tinder, italian,	107	ncaa, basketball, game	172
loan, payday, cash	88	trump, manhattan, fox	110

Table 6: The top topics in January 2022, and March 2023, among misinformation and mainstream news websites in every striation of popularity.

Rank	Misin. Abs. % Inc.	Trend Inc.	Main. Abs. % Inc.	Trend Inc.
All	+2.56%***	+0.03%***	+0.19%***	0.04%*
Rank $< 10K$	+4.58%***	0.01%***	+0.02%	+0.01%
$10K <$ Rank $< 100K$	+0.74%***	0.03%***	+0.12%**	+0.01%
$100K <$ Rank $< 1M$	+2.43%***	0.04%***	+0.22%**	+0.03%
$1M <$ Rank $< 10M$	+2.83%***	0.04%***	+0.49%	+0.07%***
Rank $> 10M$	+2.91%***	0.04%***	+2.73%***	+0.06%

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 7: Estimated absolute percentage increase immediately following the release of ChatGPT on November 30, 2022, in machine-generated articles (determined using an ARIMA-based interrupted time series analysis).

The only group of mainstream websites that saw a marked increase in synthetic articles immediately following the release of ChatGPT on November 30, 2022, was mainstream/reliable news websites with a popularity ranking greater than 10 million. However, while there was not a jump in the number of synthetic articles from many popular reliable news websites directly following the release of ChatGPT in 2023, the number of synthetic articles among *all* groups of websites has been increasing and was at its highest levels on April 1, 2023 (Figure 4). We see this mirrored in the overall increase in the trend of mainstream websites' use of synthetic articles (increase in the rate of increase) in Table 7. We note that while this analysis is *not* causal, it illustrates the noticeable increase in the percentage of synthetic articles among misinformation websites immediately following the release of ChatGPT.

6 Reddit Users and Synthetic Articles

Finally, having examined the increase in synthetic articles across the news ecosystem, we examine whether Internet users are *actually* interacting with this machine-generated content more. To do so, we analyze social media users from Reddit's overall interaction with synthetic articles.

Reddit Dataset. To understand Reddit users' interaction with synthetic content, we first gather *all* Reddit submissions/posts and their associated metadata (*e.g.*, date posted, the subreddit of the post, number of comments, *etc.*..) that referenced an article published between January 1, 2022, and April 1, 2023, from one of the news websites in our dataset. To collect this Reddit data, we rely upon Pushshift (Baumgartner et al. 2020), which keeps a queryable replica of Reddit data, getting all available submissions posted between January 1, 2022 and March 31, 2023 (April Reddit data was not available at the time of writing). Altogether, we identify 408,292 Reddit submissions that make use of 281,741 news articles (41,644 from misinformation/unreliable websites; 240,097 from mainstream/reliable news websites) from our list of URLs from 2022 and 2023. Among these articles, we identify 24,085 synthetic articles from mainstream websites and 1,157 synthetic articles from misinformation websites.

Ethical Considerations. We collect only publicly available data from Reddit. In addition, we did not attempt to deanonymize any Reddit user.

User Interaction with Synthetic Articles. In order to examine how Reddit users have interacted with synthetic articles, we plot the percentage of synthetic articles among Reddit submissions that featured an article from our 3,074 news websites (Figure 5). In addition, among the submissions that hyperlinked to an article within our full dataset, we determine the percentage of Reddit comments that were on Reddit submissions that featured a synthetic article rather than a human-written one (Figure 6).

Mainstream Synthetic Articles. As seen in Figure 5, despite the overall percentage increase in the use of synthetic new articles by mainstream/reliable news websites, we do not see a correspondingly large percentage increase in the percentage of mainstream/reliable news submissions on Reddit that are synthetic. Between January 1, 2022, and March 31, 2023, we only observe an overall increase from 8.8% to 11.5% (a 30.8% relative increase). However, while synthetic articles in terms of proportions did not increase significantly among posted mainstream articles, in terms of raw numbers, we find that the daily average of mainstream synthetic article submissions went from 36.8 in January 2022 to 63.4 in March 2023 (up 67.8%). Similarly, we observe a raw increase in the average daily number of comments on these submissions from 3,685 comments to 14,041 comments, a 281% increase (a 16.3% increase on the log scale [given that the number of comments on submissions is exponentially distributed]). As seen in Figure 6, in terms of the percentage of comments that went to submissions that featured synthetic articles, this corresponded to a 1.2% absolute percentage increase.

In addition to the increase in comments and submissions featuring synthetic articles, we find that this increase is moderately correlated with websites' publication of these articles. We calculate a $\rho = 0.617$ Pearson correlation between the overall percentage of synthetic news articles published by mainstream/reliable news websites and the percentage of Reddit submissions that featured synthetic mainstream/reliable articles. We similarly observe a $\rho = 0.341$ correlation between the percentage of Reddit comments that went to submissions featuring synthetic articles and the percentage of synthetic news articles published by mainstream/reliable news websites. This indicates that as the number of synthetic articles published by mainstream/reliable news websites, Reddit users' interactions with them increased in tandem.

Finally, we determine whether these machine-generated articles tend to receive more or less interaction from Reddit users. Performing a pairwise comparison on a domain basis of the number of comments on human-written articles against synthetic articles, we determine the Cohen's D effect size. We find, after controlling for the particular website, that *on average* synthetic articles from mainstream websites tend to receive approximately 1.10 fewer comments from Reddit users than human-written articles (Cohen's D = 1.09).[10] Thus, while there has been a slight increase in the percentage of synthetic articles on Reddit from our set of mainstream websites, machine-generated articles from these mainstream/reliable websites were *on average* less popular on Reddit than human-written articles.

Misinformation Synthetic Articles. As seen in Figure 5, there was a marked increase in the percentage of Reddit submissions that feature synthetic news articles from misinformation/unreliable news websites. Notably, we observe an increase from 5.0% in January 2022, to 9.2% in March 2023, an 85.1% relative increase. In terms of raw numbers, we observe that while on average 1.9 synthetic articles are featured in Reddit submissions each day in January 2022, this increased to an average of 4.47 in March 2023, a 131% increase. Similarly, we observe a raw increase in the average total number of comments on these submissions each day from 85.0 comments to 621.7 comments, a 631% increase (a 44.7% increase on the log scale [given that the number of comments on submissions is exponentially distributed]). This corresponded with a 2.2% absolute percentage increase in the percentage of Reddit comments on synthetic articles (Figure 6).

For misinformation websites, we find that the increase in Reddit comments and submissions featuring synthetic articles is highly correlated with misinformation websites' publication of synthetic articles. Namely, we calculate a $\rho = 0.909$ Pearson correlation between the overall percentage of synthetic news articles published by misinformation/unreliable news websites and the percentage of Reddit submissions that featured synthetic misinformation/unreliable articles. Furthermore, there is a $\rho = 0.799$ Pearson correlation between the number of comments on submissions featuring synthetic articles from unreliable/misinformation news websites and the percentage of synthetic news articles published by misinformation/unreliable news websites. Even more so than for mainstream/reliable news websites, we thus find a deep connection between the number of synthetic articles published by misinformation websites and Reddit users' posting and relative interaction with these synthetic articles.

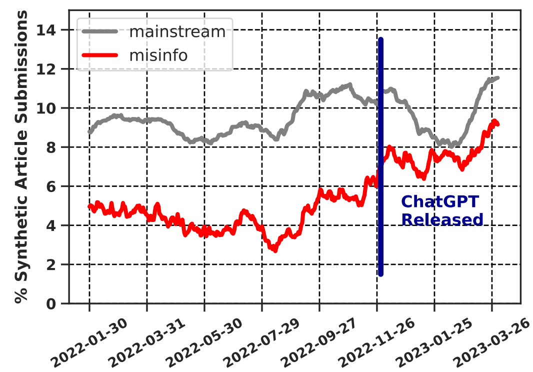

Figure 5: Percentage of Reddit submissions featuring news articles from between January 1, 2022, and March 31, 2023, that went to synthetic/machine-generated articles.

We again determine the Cohen's D effect size for the difference in users' comments on synthetic and human-written articles. After controlling for the particular website, human-written content from our set of misinformation/unreliable news websites tends to receive approximately 1.41 more comments than synthetic content (Cohen's D = 1.36).[11] This shows, again, on the whole, that human-written articles *tend* to see more engagement.

7 Discussion and Conclusion

In this work, we implemented a DeBERTa-based model to classify 12.91 million articles from 3,074 news websites as human-generated or synthetic. We find that between January 1, 2022 and April 1, 2023, the percentage of synthetic articles produced by mainstream/reliable news increased by 79.4% while the percentage produced by misinformation/unreliable news websites increased by 342%. Estimating the effect of ChatGPT, we observe a noticeable jump in the percentage of synthetic articles from misinformation websites around its release. Finally, using Reddit data, we show that social media users have interacted with synthetic articles more often in the past year, especially with synthetic

[10] We apply a Mann–Whitney U-test and find this difference to be statistically significant (*i.e.*, $p \approx 0$).

[11] We again apply a Mann–Whitney U-test and find this difference to be statistically significant (*i.e.*, $p \approx 0$).

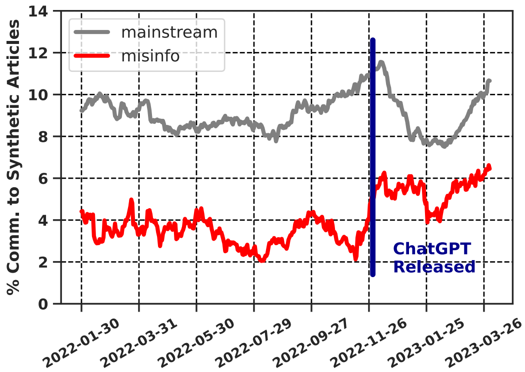

Figure 6: Percentage of Reddit comments on news submissions between January 1, 2022, and March 31, 2023, that went to synthetic/machine-generated articles submissions.

articles from misinformation websites. We now discuss several limitations and implications of this work.

Limitations. We note that while we sampled our dataset from a large set of 3,074 news websites and gathered over 12.91M articles, we did not gather articles from *every* news website and focused on English-language media. As such, our results largely do not apply to non-English media. Similarly, because we used pre-defined lists of misinformation websites, our work largely misses the *probable* existence of new misinformation websites that appeared since the launch of ChatGPT. However, we note that many of these new websites, given their novelty, would likely be small, limiting their effect on our analysis. We further note that within this work, we use a binary label of whether an article is machine-generated or human-written and do not make a distinction for articles that were largely machine-generated and then human-edited. We leave the detection of this specific type of synthetic article to future work.

Detection of Machine-Generated Media. We find that by training on data from a wide variety of generative models, we were able to outperform Open AI's released RoBERTa detector as well as several other released detectors (Pu et al. 2022). Furthermore, we find, as in prior works (Gagiano et al. 2021; Pu et al. 2022), that including data from common attacks can increase overall detection accuracy. We argue that future detectors applied to real-world data should account for these techniques.

The Rise of Synthetic Misinformation. We found that throughout 2022 and 2023, as large language models became more widely accessible, the percentage of machine-generated content on misinformation sites has had a 342% relative increase. While at the beginning of 2022, a lower percentage of misinformation/unreliable news websites' content was synthetic (3.6% vs. 9.4%), we find that by March 2023, across all popularity brackets examined, misinformation websites had begun to close this gap (16.2% vs. 16.8%). Unlike mainstream websites, misinformation websites experienced a noticeable jump in synthetic content after the release of ChatGPT (as determined by our *interrupted-time-series* analysis). Given the rapid adoption of the use of synthetic articles by misinformation websites, in particular, and given that these websites do not utilize synthetic articles for sports, finance, and statistical trackers (*e.g.*, COVID-19) like mainstream websites, we argue for future study of how misinformation websites, in particular, have utilized these technologies to spread information on social media and the broader Internet.

References

Acar, G.; Eubank, C.; Englehardt, S.; Juarez, M.; Narayanan, A.; and Diaz, C. 2014. The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. In *ACM Conference on Computer and Communications Security*.

AI, O. 2022. ChatGPT: Optimizing Language Models for Dialogue. http://web.archive.org/web/20230109000707/https://openai.com/blog/chatgpt/.

Alba, D. 2023. AI Chatbots Have Been Used to Create Dozens of News Content Farms - Bloomberg. https://www.bloomberg.com/news/articles/2023-05-01/ai-chatbots-have-been-used-to-create-dozens-of-news-content-farms.

Barret Golding. 2022. Iffy Index of Unreliable Sources. https://iffy.news/index/.

Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; and Blackburn, J. 2020. The pushshift reddit dataset. In *AAAI conference on web and social media*.

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33.

Chen, P.-J.; Lee, A.; Wang, C.; Goyal, N.; Fan, A.; Williamson, M.; and Gu, J. 2020. Facebook AI's WMT20 News Translation Task Submission. In *Proceedings of the Fifth Conference on Machine Translation*, 113–125.

Chowdhery, A.; Narang, S.; Devlin, J.; Bosma, M.; Mishra, G.; Roberts, A.; Barham, P.; Chung, H. W.; Sutton, C.; Gehrmann, S.; et al. 2022. Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311*.

Dai, Z.; Yang, Z.; Yang, Y.; Carbonell, J. G.; Le, Q.; and Salakhutdinov, R. 2019. Transformer-XL: Attentive Language Models beyond a Fixed-Length Context. In *57th Annual Meeting of the Assoc. for Computational Linguistics*.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies)*.

Gagiano, R.; Kim, M. M.-H.; Zhang, X. J.; and Biggs, J. 2021. Robustness analysis of grover for machine-generated news detection. In *19th Annual Workshop of the Australasian Language Technology Association*.

Gehrmann, S.; Strobelt, H.; and Rush, A. M. 2019. Gltr: Statistical detection and visualization of generated text. *arXiv preprint arXiv:1906.04043*.

Hanley, H. W.; and Durumeric, Z. 2023. Partial Mobilization: Tracking Multilingual Information Flows Amongst

Russian Media Outlets and Telegram. *arXiv preprint arXiv:2301.10856*.

Hanley, H. W.; Kumar, D.; and Durumeric, Z. 2023. A Golden Age: Conspiracy Theories' Relationship with Misinformation Outlets, News Media, and the Wider Internet. *ACM Computer-Supported Cooperative Work And Social Computing*.

He, P.; Liu, X.; Gao, J.; and Chen, W. 2021. DeBERTa: Decoding-enhanced BERT with disentangled attention. In *International Conference on Learning Representations*.

He, X.; Shen, X.; Chen, Z.; Backes, M.; and Zhang, Y. 2023. MGTBench: Benchmarking Machine-Generated Text Detection. *arXiv preprint arXiv:2303.14822*.

Hounsel, A.; Holland, J.; Kaiser, B.; Borgolte, K.; Feamster, N.; and Mayer, J. 2020. Identifying Disinformation Websites Using Infrastructure Features. In *USENIX Workshop on Free and Open Communications on the Internet*.

Hu, K. 2023. ChatGPT sets record for fastest-growing user base - analyst note — Reuters.

Ippolito, D.; Duckworth, D.; Callison-Burch, C.; and Eck, D. 2020. Automatic Detection of Generated Text is Easiest when Humans are Fooled. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.

Jack, C. 2017. Lexicon of lies: Terms for problematic information. *Data & Society*, 3(22): 1094–1096.

Keskar, N. S.; McCann, B.; Varshney, L. R.; Xiong, C.; and Socher, R. 2019. Ctrl: A conditional transformer language model for controllable generation. *arXiv preprint arXiv:1909.05858*.

Kirchner, J. H.; Ahmad, L.; Aaronson, S.; and Leike, J. 2023. New AI classifier for indicating AI-written text. *OpenAI blog*.

Krishna, K.; Song, Y.; Karpinska, M.; Wieting, J.; and Iyyer, M. 2023. Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense. *arXiv preprint arXiv:2303.13408*.

Lample, G.; and Conneau, A. 2019. Cross-lingual language model pretraining. *arXiv preprint arXiv:1901.07291*.

Leffer, L. 2023. CNET's AI-Written Articles Are Riddled With Errors. https://gizmodo.com/cnet-ai-chatgpt-news-robot-1849996151.

Lin, S.; Hilton, J.; and Evans, O. 2022. TruthfulQA: Measuring How Models Mimic Human Falsehoods. In *60th Annual Meeting of the Assoc. for Computational Linguistics*.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Mitchell, E.; Lee, Y.; Khazatsky, A.; Manning, C. D.; and Finn, C. 2023. Detectgpt: Zero-shot machine-generated text detection using probability curvature. *arXiv preprint arXiv:2301.11305*.

Ng, N.; Yee, K.; Baevski, A.; Ott, M.; Auli, M.; and Edunov, S. 2019. Facebook FAIR's WMT19 News Translation Task Submission. In *Fourth Conference on Machine Translation*.

OpenAI. 2022. Introducing ChatGPT. https://openai.com/blog/chatgpt.

Peiser, J. 2019. The Rise of the Robot Reporter - The New York Times. https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html.

Pu, J.; Sarwar, Z.; Abdullah, S. M.; Rehman, A.; Kim, Y.; Bhattacharya, P.; Javed, M.; and Viswanath, B. 2022. Deepfake Text Detection: Limitations and Opportunities. *arXiv preprint arXiv:2210.09421*.

Radford, A.; Narasimhan, K.; Salimans, T.; and Sutskever, I. 2019a. Improving Language Understanding by Generative Pre-Training. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I.; et al. 2019b. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8): 9.

Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; and Liu, P. J. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(1).

Ruth, K.; Kumar, D.; Wang, B.; Valenta, L.; and Durumeric, Z. 2022. Toppling top lists: Evaluating the accuracy of popular website lists. In *ACM Internet Measurement Conference*.

Solaiman, I.; Brundage, M.; Clark, J.; Askell, A.; Herbert-Voss, A.; Wu, J.; Radford, A.; Krueger, G.; Kim, J. W.; Kreps, S.; et al. 2019. Release strategies and the social impacts of language models. *arXiv preprint arXiv:1908.09203*.

Song, K.; Tan, X.; Qin, T.; Lu, J.; and Liu, T.-Y. 2020. Mpnet: Masked and permuted pre-training for language understanding. *Adv. in Neural Information Processing Systems*.

Szpakowski, M. 2020. Fake News Corpus. https://github.com/several27/FakeNewsCorpus/.

Tang, R.; Chuang, Y.-N.; and Hu, X. 2023. The science of detecting llm-generated texts. *arXiv preprint arXiv:2303.07205*.

Uchendu, A.; Le, T.; Shu, K.; and Lee, D. 2020. Authorship attribution for neural text generation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Uchendu, A.; Ma, Z.; Le, T.; Zhang, R.; and Lee, D. 2021. TURINGBENCH: A Benchmark Environment for Turing Test in the Age of Neural Text Generation. In *Findings of the Association for Computational Linguistics: EMNLP*.

Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R. R.; and Le, Q. V. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32.

Zellers, R.; Holtzman, A.; Rashkin, H.; Bisk, Y.; Farhadi, A.; Roesner, F.; and Choi, Y. 2019. Defending Against Neural Fake News. *Advances in Neural Information Processing Systems*, 32.

Zhang, G. P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50.

Zhong, W.; Tang, D.; Xu, Z.; Wang, R.; Duan, N.; Zhou, M.; Wang, J.; and Yin, J. 2020. Neural Deepfake Detection with Factual Structure of Text. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*.