# What if ChatGPT wrote the Abstract? - Explainable Multi-Authorship Attribution with a Data Augmentation Strategy

Kanishka Silva<sup>1,\*,†</sup>, Ingo Frommholz<sup>1,\*,†</sup>

<sup>1</sup>University of Wolverhampton, UK

#### Abstract

Active discussions have been conducted regarding implications and issues associated with Large Language Models (LLMs) such as ChatGPT across various domains. One particular concern is the effect of machinegenerated texts, which include a new category in authorship attribution models: machine-generated text resembling human text in topic and writing style. Differentiating human-vs-AI-written text in scientific articles is crucial for several reasons. In this work, we approach this issue from a multi-authorship perspective by investigating automatically generated abstracts. We propose a multimodal transformer which combines handcrafted stylometric features with deep learning-based text features to perform multi-author attribution. We demonstrate the effectiveness of this approach on a curated dataset of 1000 samples and discuss its explainability via the Local Interpretable Model-agnostic Explanations (LIME) framework.

#### **Keywords**

ChatGPT, Multi-Authorship Attribution, Multimodal Transformers, Data Augmentation, Explainability

### 1. Introduction

Recent advancements in text-generative Large Language Models (LLMs) saw many research directions and applications with machine-generated texts emerging, comprising summarisation, information retrieval and data augmentation. Computer-aided day-to-day tasks have evolved to use chat-based applications, providing the ability to summarise a large amount of content and effective information extraction abilities [1, 2]. These developments sparked the discussion of whether computer-assisted writing is accepted for scientific publications. For instance, a user may prompt ChatGPT [3] to generate an abstract for a full text of their scientific paper and then modify it accordingly. Alternatively, an editorial assistant could use ChatGPT to generate the initial draft when creating a summary report for a journal issue. Publishing policy guidelines<sup>1</sup> might accept or reject generated texts to a certain degree. While some guidelines allow using LLMs as a writing aid, using wholly generated and unedited text fragments is often

In: M. Litvak, I.Rabaev, R. Campos, A. Jorge, A. Jatowt (eds.): Proceedings of the IACT'23 Workshop, Taipei, Taiwan, 27-July-2023

<sup>\*</sup>Corresponding author.

<sup>&</sup>lt;sup>†</sup>These authors contributed equally.

kanishka.silva.92@gmail.com (K. Silva); ifrommholz@acm.org (I. Frommholz)

<sup>© 0222</sup> Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0). CEUR Workshop Proceedings (CEUR-WS.org)

<sup>&</sup>lt;sup>1</sup>An example of such policy is the ICML 2023 policy on LLMs – https://icml.cc/Conferences/2023/llm-policy.

forbidden. This makes the task of proper identification of generated texts, or text fragments a crucial one. We can regard this as an *multi-authorship problem*, where the two classes of authors involved: human co-authored text and LLM/chatbot aided human co-authored text. We assume the LLM/chatbot-aided text carries distinguishable stylometric features compared to the human-co-authored text portion in a considered document. Further, for simplicity, the LLM/chatbot-aided text is a wholly generated and unedited text fragment in a known position of the document.

Our main contributions are: using GPT 3.5 as a data augmentation mechanism to generate abstracts; addressing the multi-authorship problem in scientific papers using a multimodal transformer, combining stylometric and decoder-based text features; providing a corresponding dataset and scripts; a novel explainability feature to the multimodal transformer using the Local Interpretable Model-agnostic Explanations (LIME) framework [4]; a case study to utilise ChatGPT [3] as a data augmentation tool for scientific articles. The main research question addressed in this work is how effectively an explainable multimodal transformer-based model can identify ChatGPT-generated text.

The remainder of the paper is organised as follows. Section 2 demonstrates a brief literature survey. Section 3 outlines the proposed multimodal transformer model for ChatGPT text identification. Section 4 describes the dataset. Section 5 elaborates on the experiment design outline, focusing on the research question. Section 6 summarises the results and discussion points. Section 7 outlines the limitations and future directions related to the research presented in this paper. Finally, Section 8 discuss concluding remarks and future directions.

### 2. Related Works

*Authorship attribution* is identifying the author of an unknown text by comparing a corpus of known authorship of candidate authors [5]. The approaches for the authorship attribution are traditional stylometric [6, 7, 8] and deep learning-based [9, 10, 11]. The stylometric approaches usually involve handcrafted feature extraction [6, 12, 8]. Ensemble models combining different stylometric features with deep learning have been outperforming other state-of-the-art models [13, 14].

*Multi-authorship attribution* addresses identifying authors or detecting cases where different parts of a document were written by different authors [15, 16]. Approaches include co-authorship graph-based authorship attribution [16]; others simply identify whether a given document is multi-authored [17, 18, 19].

Several attempts in recent research are towards identifying human-vs-AI-created text [20, 21, 22, 23]. Authorship obfuscation with machine-generated text to impose/hide the original writing style is discussed in the works of Jones et al. [24], Dehouche [25]. A pilot study on ChatGPT text authorship has been reported by Landa-Blanco et al. [26]. With the recent advancements of the LLMs and prompt-based text generation models, much research has focused on utilising machine-generated text for human-vs-AI text detection [27, 28].

*Multimodal transformers* [29] combine a pre-trained transformer model output with additional task-specific categorical or numerical features. Gu and Budhkar [29] provide different feature combination methods involving feature concatenation, attention methods, gating mechanisms

and weighted feature summation.

### 3. Explainable Multi-Authorship Attribution

Limited dataset availability is ubiquitous in many human-vs-AI text attribution applications. One mitigation approach would automatically synthesise machine-generated text with a natural language-based generative model. Until the recent applications of such generative models can generate human-like data, most recently, using chatbot APIs like ChatGPT [30], much research has been initiated in formulating data augmentation strategies.

Similarly, as in Figure 1, we propose a scalable task-specific encoder-decoder-based experimental pipeline with multimodal and explainable abilities. The *Decoder LLM* takes the original data and uses an LLM-based generator model such as ChatGPT (used in this work) [3], BARD [31] or PaLM [32] to return synthesised data, i.e. machine-aided text segments. Then according to the desired task, the synthesised and original data are passed through a *Data Merger* to generate the train-test-validation splits for training the model and test data form prediction through an *Explainer* model such as LIME Explainer [4]. The *Encoder Language Model (LLM)* can incorporate any appropriate LLM based on the intended task, such as a classifier for document classification purposes. Then the attention output of the model is passed to the Explainer in conjunction with the test data to obtain highlighted segments of the document which contributed to the Encoder LLM result. We propose this model for a case study of a multi-authorship problem, where a scientific document is multi-authored by a human or machine author if the original writer prompted ChatGPT to generate the abstract.

For the simplicity of the experiments, we considered article D, written by two (machinegenerated+human) or one author's category (human). Given a set of documents  $D = \{D_1, D_2, D_3, ..\}$  where  $D_i$  consist of two sections  $S_1$  and  $S_2$  if multi-authored, which represent sections with potential style changes, authored by either a human or an AI. Using multimodal transformers [29], we concatenated stylometric features (n) and deep learning-based text features (x) to obtain combined multimodal features (m = x || n) to identify whether a given document  $D_i$ is multi-authored.

### 4. Datasets

We used the arXiv dataset described in Cohan et al. [33] as the basis for our experiments. It consists of long, structured documents collected from the arXiv and PubMed Open Access repositories. This dataset contains articles and abstracts separated by the '\n' character and comprises 215,913 arXiv articles and 133,215 PubMed articles. Using zero-shot prompting, we utilised the ChatGPT 3.5 API [30] to generate abstracts for 500 out of 1,000 randomly selected arXiv articles set from the subset mentioned above. The GPT-3.5 Turbo Language Model was utilised to generate abstracts from the selected scientific papers, truncating approximately 2500 words and with 0.7 temperature settings for controlled randomness.

The resulting *ChatGPT-Aided Papers* dataset consists of combined 500 synthesised (ChatGPT abstract + original text) and 500 original articles (original abstract + original text) under two categories denoting multi-authored and single-authored, respectively. Each data item in our

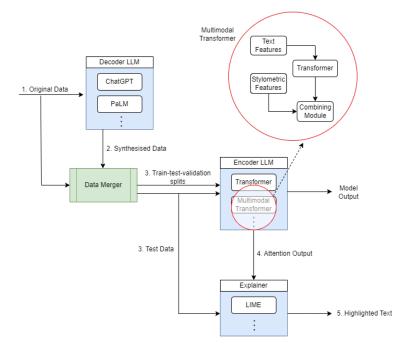


Figure 1: Proposed task-specific encoder-decoder-based experimental pipeline: Decoder LLM, Encoder LLM, and Explainer models

dataset comprises nine fields: article, labels and seven stylometric features per each article: Average Word Length, Average Sentence Length by Characters, Average Sentence Length By Word, Average Syllable per Word, Special Character Count, Punctuation Count, Functional Words Count. A further comparison of datasets is available in Table 1, comparing the document length, sentence length, and vocabulary size of the original papers and the ChatGPT-aided papers.

#### Table 1

Dataset Statistics . Both Original Papers and ChatGPT-Aided Papers contains 500 documents. Document Length, Sentence Length, and Vocabulary Size are calculated per average, and counts are in terms of words

| Dataset                         | Document Length | Sentence Length | Vocabulary Size |
|---------------------------------|-----------------|-----------------|-----------------|
| Original Papers [33]            | 6625.40         | 26.35           | 1163.68         |
| ChatGPT-Aided Papers (Proposed) | 6785.68         | 26.12           | 1170.20         |

As illustrated in Figure 1, the *Data Merger* performs the data augmentation by combining the synthesised and original data to suit the desired task. In our considered case study on the multi-author identification in machine-aided scientific paper writeups, we concatenated 500 ChatGPT abstracts with full paper text to create computer-aided scientific writeups. This synthesised data reflects a scenario where an author requests ChatGPT to summarise the entire paper for writing an abstract. To obtain the human-written documents, we combined the remaining 500

human-written abstracts with respective full paper text, resulting in a dataset comprising 1000 articles in a uniform distribution across classes (Original Papers and ChatGPT-Aided Papers), including the computer-aided and original text.

## 5. Experiment Design

Several experiments were designed to validate the model performance as follows:

- 1. An ablation study with authorship features was performed;
- 2. A multimodal transformer with text and authorship features was applied;
- 3. LIME explainer (see below) was utilised to interpret results;
- 4. Results were analysed and compared to state-of-the-art (SOTA) models.

In our study, we calculated several handcrafted stylometric features: Average word length (AWL), average sentence length by word (ASW), and functional word counts (CFW) per each document<sup>2</sup>. These features were then concatenated with the text features and passed to a multimodal transformer[29] to perform the authorship attribution task. To evaluate the feature significance, an ablation study was conducted. The main experiment flow is about effectively using a multimodal transformer [29] for multi-authorship problems and explaining the model output. We conducted the experiments to compare a BERT model specific for sequence classification [34] with the proposed multimodal transformer and the effect of each aforementioned stylometric feature on the proposed model. Then, we present the LIME Explainer [4] result for a scientific paper text where the abstract was entirely written by prompting ChatGPT. To compare the proposed model with existing models and baselines, we utilised different baseline models such as word-level TF-IDF [9], character n-gram [9], Stylometric features [35] on the ChatGPT-Aided Papers dataset.

We performed hyper-parameter tuning to identify the best model parameters with the converging loss. The proposed model was trained with Adam optimiser, with a 0.5 dropout rate, a 16 batch size, 0.000001 learning rate, 0.000001 epsilon, 0.2 warm-up proportion, and five train epochs maximum 512 token length. To ensure the reprehensibility of the presented research, we release the code-base<sup>3</sup> and the dataset<sup>4</sup>.

### 6. Results and Discussion

The results of the experiments are illustrated in Table 2, showcasing that the proposed multimodal transformer-based model outperforms the BERT (Sentence Classification) model, which only utilises text features. The proposed model demonstrates superior performance in terms of accuracy and F1 of 0.93, which is nearly 50% than the BERT model. This improvement could be due to stylometric features - Average Word Length (AWL), Average Sentence Length By Word (ASW), and Count Functional Words (CFW), which were combined with text features.

<sup>&</sup>lt;sup>2</sup>The application of all 7 features mentioned before resulted in overfitting.

<sup>&</sup>lt;sup>3</sup>Code-base - https://github.com/Kaniz92/Multimodal-ChatGPT-AA. We used Simple Transformers as the boilerplate for the implementation [36].

<sup>&</sup>lt;sup>4</sup>https://huggingface.co/datasets/Authorship/ChatGPT\_Aided\_Papers

We conducted an ablation study by systematically removing each feature and evaluating the performance to gain insights into the contribution of different stylometric features. Table 2 presents the results obtained from this study. The similar decrease in performance, resulting in an accuracy of 0.94 when removing each feature, indicates that all three features contribute equally to the model's performance.

#### Table 2

Multimodal Transformer (MMT)-based results on concatenating stylometric features and text features: AWL – Average Word Length, ASW – Average Sentence Length By Word, CFW – Count Functional Words, BERT – BertForSequenceClassification Model

| Model  | Accuracy | F1   | Precision | Recall |
|--|----------|------|-----------|--------|
| BERT [34]  | 0.55     | 0.52 | 0.57      | 0.56   |
| Stylometric [35]                                 | 0.47     | 0.43 | 0.46      | 0.47   |
| Character Ngram [9]                              | 0.57     | 0.56 | 0.57      | 0.57   |
| Word level TF-IDF [9]                            | 0.52     | 0.51 | 0.52      | 0.52   |
| MMT (BERT) + Stylometric (AWL + ASW + CFW)       | 0.98     | 0.97 | 0.98      | 0.98   |
| MMT (BERT) + Stylometric (AWL + ASW + CFW) - CFW | 0.94     | 0.94 | 0.94      | 0.94   |
| MMT (BERT) + Stylometric (AWL + ASW + CFW) - ASW | 0.94     | 0.94 | 0.94      | 0.94   |
| MMT (BERT) + Stylometric (AWL + ASW + CFW) - AWL | 0.94     | 0.94 | 0.94      | 0.94   |

Figure 2 presents the LIME Explainer results for a sample full paper text, where the initial text is annotated. The significance of each most relevant word is depicted through bar charts in Figure 2b and Figure 2c, which indicate the assigned importance for the proposed model's output. The colour scheme in the figures assigns blue to Figure 2a and red to Figures 2a and 2b, indicating the words that majorly influence the model's decision within each class. The LIME Explainer results provide a deeper understanding of the model's decision-making process and highlight the most specific words which play a crucial role in the model's predictions.

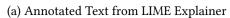
As shown in Table 2, the proposed model demonstrates a superior performance with an accuracy of 0.98, outperforming BERT, Stylometric, Character Ngram, and Word level TF-IDF accuracies of 0.55, 0.47, 0.57, and 0.52, respectively. These results prove the proposed model's effectiveness in multi-authorship identification problems in scientific articles with ChatGPT-aided abstracts.

# 7. Limitations and Future Directions

Our work is limited by the use of ChatGPT 3.5 and its limitations. Research on LLMs is a rapidly developing field that has seen the introduction of more advanced commercial models such as GPT 4 [37] and BARD [31] and further open-source language models. Hence, future work should focus on applying our task to other available LLMs or encoders/decoders (such as VAEs and Diffusion models) to provide us with more robust insights. Furthermore, policies on using LLMs may allow them as writing aid where authors manually proofread and edit a generated text; such hybrid texts should not necessarily be flagged as "generated".

Further, Due to the token limitation in the ChatGPT API, we considered the first 2500 words





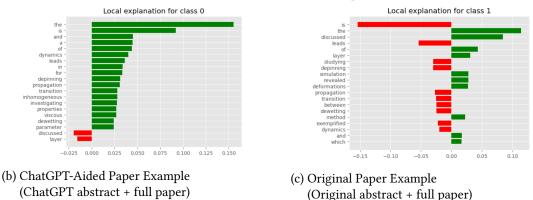


Figure 2: LIME Explainer results for a sample full paper with ChatGPT-written abstract

of each paper as crucial for an abstract generation. The future directions for this research are exploring other explainability models, calculating machine-aided text percentage and analysing other scientific paper sections aided by generator models.

# 8. Conclusion

This paper proposes an encoder-decoder LLMs-based framework for multi-author identification tasks, where machine-generation text identification is intended. We considered a scenario of multi-authorship attribution with machine-aided text and original text. According to the experimental results, the multimodal transformer model combines stylometric features and text features to outperform the BERT model and other ablation studies. The study demonstrates the effectiveness of handcrafted stylometric features. Applying the BERT model and the stylometric features alone provides relatively poor results. Only their combination achieves a significant performance boost. These exciting observations show that both approaches make classification errors, but different ones that are ironed out when combined.

### References

- [1] Y. Ma, J. Liu, F. Yi, Q. Cheng, Y. Huang, W. Lu, X. Liu, AI vs. human differentiation analysis of scientific content generation, 2023. arXiv:2301.10416.
- [2] O. O. Buruk, Academic writing with GPT-3.5: reflections on practices, efficacy and transparency, CoRR abs/2304.11079 (2023). URL: https://doi.org/10.48550/arXiv.2304.11079. doi:10.48550/arXiv.2304.11079. arXiv:2304.11079.
- [3] OpenAI, ChatGPT openai blog, https://openai.com/blog/chatgpt/, 2023.
- [4] M. T. Ribeiro, S. Singh, C. Guestrin, "why should I trust you?": Explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2016, pp. 1135–1144. URL: https://doi.org/ 10.1145/2939672.2939778. doi:10.1145/2939672.2939778.
- [5] E. Stamatatos, A survey of modern authorship attribution methods, J. Assoc. Inf. Sci. Technol. 60 (2009) 538–556. URL: https://doi.org/10.1002/asi.21001. doi:10.1002/asi.21001.
- [6] O. Aborisade, M. Anwar, Classification for authorship of tweets by comparing logistic regression and naive bayes classifiers, in: 2018 IEEE International Conference on Information Reuse and Integration, IRI 2018, IEEE, 2018, pp. 269–276. URL: https: //doi.org/10.1109/IRI.2018.00049. doi:10.1109/IRI.2018.00049.
- [7] D. Madigan, A. Genkin, D. D. Lewis, E. G. D. D. Lewis, S. Argamon, D. Fradkin, L. Ye, D. D. L. Consulting, Author identification on the large scale, in: In Proc. of the Meeting of the Classification Society of North America, 2005.
- [8] J. Soler Company, L. Wanner, On the relevance of syntactic and discourse features for author profiling and identification, in: M. Lapata, P. Blunsom, A. Koller (Eds.), Proceedings EACL 2017, Association for Computational Linguistics, 2017, pp. 681–687. URL: https: //doi.org/10.18653/v1/e17-2108. doi:10.18653/v1/e17-2108.
- [9] M. Fabien, E. Villatoro-Tello, P. Motlícek, S. Parida, Bertaa : BERT fine-tuning for authorship attribution, in: P. Bhattacharyya, D. M. Sharma, R. Sangal (Eds.), Proceedings ICON 2020, NLP Association of India (NLPAI), 2020, pp. 127–137. URL: https://aclanthology.org/2020. icon-main.16.
- [10] C. Saedi, M. Dras, Siamese networks for large-scale author identification, Comput. Speech Lang. 70 (2021) 101241. URL: https://doi.org/10.1016/j.csl.2021.101241. doi:10.1016/j.csl. 2021.101241.
- [11] S. Ruder, P. Ghaffari, J. G. Breslin, Character-level and multi-channel convolutional neural networks for large-scale authorship attribution, CoRR abs/1609.06686 (2016). URL: http://arxiv.org/abs/1609.06686. arXiv:1609.06686.
- [12] Y. Sari, Neural and non-neural approaches to authorship attribution, Ph.D. thesis, University of Sheffield, UK, 2018. URL: https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos. 755204.
- [13] E. Moreau, A. K. Jayapal, G. Lynch, C. Vogel, Author verification: Basic stacked generalization applied to predictions from a set of heterogeneous learners - notebook for pan at clef 2015, in: CLEF, 2015.
- [14] A. Bacciu, M. L. Morgia, A. Mei, E. N. Nemmi, V. Neri, J. Stefa, Cross-domain authorship attribution combining instance based and profile-based features, in: L. Cappellato, N. Ferro, D. E. Losada, H. Müller (Eds.), Working Notes of CLEF 2019 - Conference and Labs of

the Evaluation Forum, Lugano, Switzerland, September 9-12, 2019, volume 2380 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2019. URL: http://ceur-ws.org/Vol-2380/paper\_220.pdf.

- [15] J. Bevendorff, M. Chinea-Rios, M. Franco-Salvador, A. Heini, E. Körner, K. Kredens, M. Mayerl, P. Pezik, M. Potthast, F. Rangel, P. Rosso, E. Stamatatos, B. Stein, M. Wiegmann, M. Wolska, E. Zangerle, Overview of PAN 2023: Authorship verification, multi-author writing style analysis, profiling cryptocurrency influencers, and trigger detection - extended abstract, in: Proceedings ECIR 2023, volume 13982 of *Lecture Notes in Computer Science*, Springer, 2023, pp. 518–526. URL: https://doi.org/10.1007/978-3-031-28241-6\_60. doi:10.1007/978-3-031-28241-6\_60.
- [16] R. Sarwar, N. Urailertprasert, N. Vannaboot, C. Yu, T. Rakthanmanon, E. Chuangsuwanich, S. Nutanong, CAG: stylometric authorship attribution of multi-author documents using a co-authorship graph, IEEE Access 8 (2020) 18374–18393. URL: https://doi.org/10.1109/ ACCESS.2020.2967449. doi:10.1109/ACCESS.2020.2967449.
- [17] M. M. Iqbal, A. Raza, M. M. Aslam, M. Farhan, S. Yaseen, A stylometric fingerprinting method for author identification using machine learning, Technical Journal 28 (2023) 28–35.
- [18] A. M. Omi, M. Hossain, M. N. Islam, T. Mittra, Multiple authors identification from source code using deep learning model, Proceedings ICECIT 2021 (2021) 1–4.
- [19] P. Bora, T. Awalgaonkar, H. Palve, R. Joshi, P. Goel, Icodenet A hierarchical neural network approach for source code author identification, in: Proceedings ICMLC 2021, ACM, 2021, pp. 180–185. URL: https://doi.org/10.1145/3457682.3457709. doi:10.1145/3457682.3457709.
- [20] K. Pillutla, S. Swayamdipta, R. Zellers, J. Thickstun, S. Welleck, Y. Choi, Z. Harchaoui, MAUVE: measuring the gap between neural text and human text using divergence frontiers, in: Proceedings NeurIPS 2021, 2021, pp. 4816–4828. URL: https://proceedings.neurips.cc/ paper/2021/hash/260c2432a0eecc28ce03c10dadc078a4-Abstract.html.
- [21] J. Pu, Z. Sarwar, S. M. Abdullah, A. Rehman, Y. Kim, P. Bhattacharya, M. Javed, B. Viswanath, Deepfake text detection: Limitations and opportunities, CoRR abs/2210.09421 (2022). URL: https://doi.org/10.48550/arXiv.2210.09421. doi:10.48550/arXiv. 2210.09421. arXiv:2210.09421.
- [22] R. Gagiano, M. M.-H. Kim, X. Zhang, J. Biggs, Robustness analysis of grover for machinegenerated news detection, in: Proceedings of the The 19th Annual Workshop of the Australasian Language Technology Association, Australasian Language Technology Association, Online, 2021, pp. 119–127. URL: https://aclanthology.org/2021.alta-1.12.
- [23] A. Uchendu, T. Le, K. Shu, D. Lee, Authorship attribution for neural text generation, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), Proceedings EMNLP 2020, Association for Computational Linguistics, 2020, pp. 8384–8395. URL: https://doi.org/10.18653/v1/2020. emnlp-main.673. doi:10.18653/v1/2020.emnlp-main.673.
- [24] K. Jones, J. R. C. Nurse, S. Li, Are you robert or roberta? deceiving online authorship attribution models using neural text generators, in: C. Budak, M. Cha, D. Quercia (Eds.), Proceedings of the Sixteenth International AAAI Conference on Web and Social Media, ICWSM 2022, Atlanta, Georgia, USA, June 6-9, 2022, AAAI Press, 2022, pp. 429–440. URL: https://ojs.aaai.org/index.php/ICWSM/article/view/19304.
- [25] N. Dehouche, Plagiarism in the age of massive generative pre-trained transformers (gpt-3),

Ethics in Science and Environmental Politics 21 (2021) 17-23. doi:10.3354/esep00195.

- [26] M. Landa-Blanco, M. A. Flores, M. Mercado, Human vs. ai authorship: Does it matter in evaluating creative writing? a pilot study using chatgpt (2023).
- [27] J. J. Bird, A. Ek'art, D. R. Faria, Chatbot interaction with artificial intelligence: human data augmentation with t5 and language transformer ensemble for text classification, Journal of Ambient Intelligence and Humanized Computing 14 (2020) 3129–3144.
- [28] F. Gao, J. Zhu, L. Wu, Y. Xia, T. Qin, X. Cheng, W. Zhou, T. Liu, Soft contextual data augmentation for neural machine translation, in: A. Korhonen, D. R. Traum, L. Màrquez (Eds.), Proceedings ACL 2019, Association for Computational Linguistics, 2019, pp. 5539–5544. URL: https://doi.org/10.18653/v1/p19-1555. doi:10.18653/v1/p19-1555.
- [29] K. Gu, A. Budhkar, A package for learning on tabular and text data with transformers, in: Proceedings of the Third Workshop on Multimodal Artificial Intelligence, Association for Computational Linguistics, Mexico City, Mexico, 2021, pp. 69–73. URL: https://www. aclweb.org/anthology/2021.maiworkshop-1.10. doi:10.18653/v1/2021.maiworkshop-1.10.
- [30] OpenAI, Chatgpt API, https://openai.com/docs/api, 2021.
- [31] Google AI Blog, BARD: Google AI Search Updates, https://blog.google/technology/ai/ bard-google-ai-search-updates/, YYYY. Accessed: Month Day, Year.
- [32] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H. W. Chung, C. Sutton, S. Gehrmann, P. Schuh, K. Shi, S. Tsvyashchenko, J. Maynez, A. Rao, P. Barnes, Y. Tay, N. Shazeer, V. Prabhakaran, E. Reif, N. Du, B. Hutchinson, R. Pope, J. Bradbury, J. Austin, M. Isard, G. Gur-Ari, P. Yin, T. Duke, A. Levskaya, S. Ghemawat, S. Dev, H. Michalewski, X. Garcia, V. Misra, K. Robinson, L. Fedus, D. Zhou, D. Ippolito, D. Luan, H. Lim, B. Zoph, A. Spiridonov, R. Sepassi, D. Dohan, S. Agrawal, M. Omernick, A. M. Dai, T. S. Pillai, M. Pellat, A. Lewkowycz, E. Moreira, R. Child, O. Polozov, K. Lee, Z. Zhou, X. Wang, B. Saeta, M. Diaz, O. Firat, M. Catasta, J. Wei, K. Meier-Hellstern, D. Eck, J. Dean, S. Petrov, N. Fiedel, Palm: Scaling language modeling with pathways, CoRR abs/2204.02311 (2022). URL: https://doi.org/10.48550/arXiv.2204.02311. doi:10.48550/arXiv.2204.02311.
- [33] A. Cohan, F. Dernoncourt, D. S. Kim, T. Bui, S. Kim, W. Chang, N. Goharian, A discourse-aware attention model for abstractive summarization of long documents, in: M. A. Walker, H. Ji, A. Stent (Eds.), Proceedings NAACL-HLT 2018, Association for Computational Linguistics, 2018, pp. 615–621. URL: https://doi.org/10.18653/v1/n18-2097. doi:10.18653/v1/n18-2097.
- [34] H. Face, Hugging face transformers documentation: Bertforsequenceclassification, https://huggingface.co/docs/transformers/v4.30.0/en/model\_doc/bert#transformers. BertForSequenceClassification, Accessed: 2023.
- [35] Y. Sari, M. Stevenson, A. Vlachos, Topic or style? exploring the most useful features for authorship attribution, in: E. M. Bender, L. Derczynski, P. Isabelle (Eds.), Proceedings of the 27th International Conference on Computational Linguistics, COLING 2018, Santa Fe, New Mexico, USA, August 20-26, 2018, Association for Computational Linguistics, 2018, pp. 343–353. URL: https://aclanthology.org/C18-1029/.
- [36] T. C. Rajapakse, Simple transformers, https://github.com/ThilinaRajapakse/ simpletransformers, 2019.
- [37] OpenAI, GPT-4 technical report, CoRR abs/2303.08774 (2023). URL: https://doi.org/10.

48550/arXiv.2303.08774. doi:10.48550/arXiv.2303.08774. arXiv:2303.08774.