RoBERT-IA: Human-AI Collaborative Text Classification

Notebook for the PAN Lab at CLEF 2025

Deyson Gómez Sánchez[1], Jeison D. Jimenez[1], María Paz Ramírez[1], Jairo E. Serrano[1], Juan C. Martinez-Santos[1] and Edwin Puertas[1]

[1]*Universidad Tecnológica de Bolívar; School of Engineering, Architecture, and Design; Cartagena de Indias; 130013; Colombia*

Abstract

Within the framework of the Generative AI Detection 2025 SubTask 2: Human-AI Collaborative Text Classification challenge, this study addresses the classification of texts co-authored by humans and large language models (LLMs), aiming to identify the degree of contribution of each author across six specific categories. Given the increasing accessibility and use of models such as GPT-4o, Claude 3.5, and Gemini 1.5-pro, the proliferation of AI-generated or AI-assisted content presents significant challenges in areas including misinformation, academic integrity, and content authenticity. To tackle this challenge, a finetuning process was applied to the RoBERTa-base model, employing strategies to mitigate class imbalance such as undersampling and loss weighting. The dataset was split into 80% for training and 20% for evaluation, considering key metrics like accuracy, F1-score, and macro recall — the latter used as the official classification metric. Preliminary results indicate that loss weighting for minority classes is a more suitable strategy than synthetic data generation, as it preserves the naturalness of the texts. Evaluation on the test set demonstrated a balanced improvement in key metrics, achieving a macro recall of 47.15% on the evaluation dataset, underscoring the effectiveness of the approach in discriminating the various forms of human-AI collaboration in text creation. Furthermore, post-competition evaluation showed that increasing the number of training epochs surpasses the baseline metrics.

Keywords

Generative AI Detection, Text Classification, Large Language Models (LLMs), AI-generated content.

1. Introduction

The rapid emergence of large language models (LLMs) such as GPT-4o, Claude 3.5, and Gemini 1.5-pro has transformed textual content generation across diverse domains, including digital platforms, education, media, and academia Jakesch et al.[2] These models produce text with high syntactic and semantic quality, enabling efficient human-machine collaborations. However, their widespread adoption raises significant concerns regarding misinformation, academic integrity, content authenticity, and authorship transparency Shah et al.[8] Addressing these challenges requires robust detection mechanisms capable of discerning the degree of human and automated involvement in text creation. This study is situated within Subtask 2 of the Voight-Kampff 2025 challenge, titled Human-AI Collaborative Text Classification, which focuses on categorizing documents co-authored by humans and generative models into six distinct classes based on authorship composition Bevendorff et al.[3] We propose a supervised learning architecture based on the RoBERTa model, fine-tuned specifically for this task. To overcome the severe class imbalance in the dataset—which adversely affects model generalization—we employ a combination of oversampling, undersampling, and SMOTE techniques Zeng et al. [1].

Furthermore, we critically examine back-translation as a data augmentation method, highlighting its limitations for tasks where preserving stylistic authorship is essential

CLEF 2025 Working Notes, 9 – 12 September 2025, Madrid, Spain

*Corresponding author.

†These authors contributed equally.

deygomez@utb.edu.co (D. G. Sánchez); jalvear@utb.edu.co (J. D. Jimenez); atenciom@utb.edu.co (M. P. Ramírez); jserrano@utb.edu.co (J. E. Serrano); jcmartinezs@utb.edu.co (J. C. Martinez-Santos); epuerta@utb.edu.co (E. Puertas)

0009-0005-2172-6905 (D. G. Sánchez); 0009-0001-0134-8426 (J. D. Jimenez); 0000-0001-8165-7343 (J. E. Serrano); 0000-0003-2755-0718 (J. C. Martinez-Santos); 0000-0002-0758-1851 (E. Puertas)

Our work not only advances methodological approaches to multi-class classification in human-AI collaborative writing contexts but also provides fundamental insights into the complex nature of textual collaboration between humans and generative models, with implications for improving authorship verification and combating misinformation Ragab et al. [10]

2. State of the Art

The detection of texts generated by artificial intelligence (AI) has emerged as a critical research domain in response to the rapid and widespread adoption of large language models (LLMs). Initial efforts predominantly addressed the binary classification task of distinguishing human-authored from AI-generated texts, laying the groundwork for more sophisticated detection methodologies. Notably, Uchendu et al. [2] extended this paradigm by differentiating not only between human and AI authorship but also among distinct generative models (e.g., GPT-2 and GROVER) through stylometric and syntactic feature analysis.

Despite these advances, human capacity to reliably detect AI-generated content remains fundamentally limited. Jakesch et al. [3] demonstrate that human detection accuracy approximates random chance (50%), even when incentivized or trained, due to the exploitation of flawed cognitive heuristics by AI systems. This finding underscores the necessity for robust automated detection systems. Bevendorff et al. [4] contribute a significant benchmark with the "Voight-Kampff" challenge—an adversarial competition assessing detection systems' resilience across 70 test variants including obfuscation techniques and cross-lingual scenarios. Although some systems surpassed baseline performance, none achieved perfect classification, highlighting persistent challenges especially as language models evolve.

The growing prevalence of hybrid texts, produced via human-AI collaboration, further complicates detection efforts. Zeng et al. [1] propose a two-stage segmentation and classification pipeline, revealing that frequent human editing and shifting authorship markedly degrade detector performance. Complementing this, Richburg et al. [5] compare authorship embedding models (LUAR), n-grams, and Transformer-based approaches, demonstrating a trade-off between binary detection superiority of embeddings and robustness of n-gram models for fine-grained authorship verification in collaborative texts.

Recent technical innovations have propelled efficiency and accuracy gains. Bao et al. [6] introduce a zero-shot detection method leveraging conditional probability curvature, significantly accelerating detection while maintaining high precision. Stylistic and linguistic feature-based methods continue to show promise: Rujeedawa et al. [7] achieve 82.6% accuracy with Random Forest classifiers using metrics such as text length and lexical richness; Shah et al. [8] enhance this by incorporating explainable AI to reach 93% accuracy, identifying that AI-generated texts manifest greater lexical richness but reduced diversity, corroborating Uchendu et al.'s findings.

In deep learning, Ragab et al. [9] present a hybrid CNN-GRU architecture optimized by a metaheuristic algorithm, attaining over 99% accuracy by combining local and long-term dependency features. Oghaz et al. [10] similarly affirm the robustness of Transformer models, notably RoBERTa, which achieves an F1-score of 0.992 even on short text excerpts. Building on these foundations, recent research broadens the scope and depth of detection capabilities. Boran et al. [11] highlight the criticality of authorship identification in limited-sample contexts, crucial for plagiarism detection in heterogeneous digital environments.

Mizumoto et al. [12] apply linguistic fingerprinting and random forest classification to distinguish ChatGPT-generated essays from student work, underscoring the imperative for automated detection integration in educational settings. Fiedler and Döpke [13] reveal the considerable difficulty experts face in reliably identifying AI-generated academic texts, showing parity between human and machine

performance limitations, especially for high-quality AI-generated content. This accentuates the need for advanced computational solutions tailored to complex detection scenarios.

From a stylometric perspective, Berriche and Larabi-Marie-Sainte [14] introduce methods exploiting intrinsic writing style features to detect ChatGPT-based plagiarism with exceptional precision (up to 100%), including classification of mixed human-AI texts, directly addressing challenges posed by co-authored and paraphrased documents. Desaire et al. [15] develop classifiers capable of distinguishing human academic authorship from ChatGPT-generated content with over 99% accuracy, leveraging discourse-specific linguistic markers vital for formal plagiarism detection.

Lastly, Lau and Zubiaga [16] investigate how human paraphrasing affects LLM-generated text detection, demonstrating that paraphrases markedly degrade detector performance and necessitate novel strategies for resilient detection in real-world edited texts. Beyond technical performance, this body of work reflects broader societal implications. Reliable AI-generated text detection is essential for preserving academic integrity, combating misinformation, and fostering trust in digital communication. However, ethical considerations—such as privacy, transparency, and the potential for misclassification—must guide future developments. Research directions should encompass explainable AI frameworks, multimodal detection methods, and cross-linguistic generalization, ensuring robust, fair, and interpretable detection systems.

3. Data

For the development of this work, we employed the dataset provided by the organizers of the PAN-CLEF 2025 competition, specifically for Subtask 2: Human-AI Collaborative Text Classification [17][18]. This dataset is designed to address the identification and categorization of documents co-authored by humans and large language models (LLMs), such as GPT-4o, Claude 3.5, and Gemini 1.5-pro.

The dataset includes texts in English, Spanish, and German, spanning multiple domains such as academia, journalism, social media, and education. This diversity reflects the widespread proliferation and applicability of AI-generated or AI-assisted content across various contexts and thematic areas.

To capture the different modes of collaboration between humans and machines in text generation, the documents are annotated with labels that distinguish six specific categories: texts entirely written by humans; texts initiated by humans and continued by machines; texts written by humans and subsequently polished by machines; texts written by machines and later humanized (obfuscated); texts generated by machines and later edited by humans; and deeply mixed texts with interwoven contributions from both authors. The distribution of each class is presented in Table 1.

Table 1
Distribution of label categories in training dataset.

Label Category	Train
Machine-written, then machine-humanized	91,232
Human-written, then machine-polished	95,398
Fully human-written	75,270
Human-initiated, then machine-continued	10,740
Deeply-mixed text (human + machine parts)	14,910
Machine-written, then human-edited	1,368
Total	**288,918**

4. Methodology

In this section, we detail the methodology outlined in Figure 1, which was used to classify documents co-authored by humans and LLMs based on the level of contribution from each party.

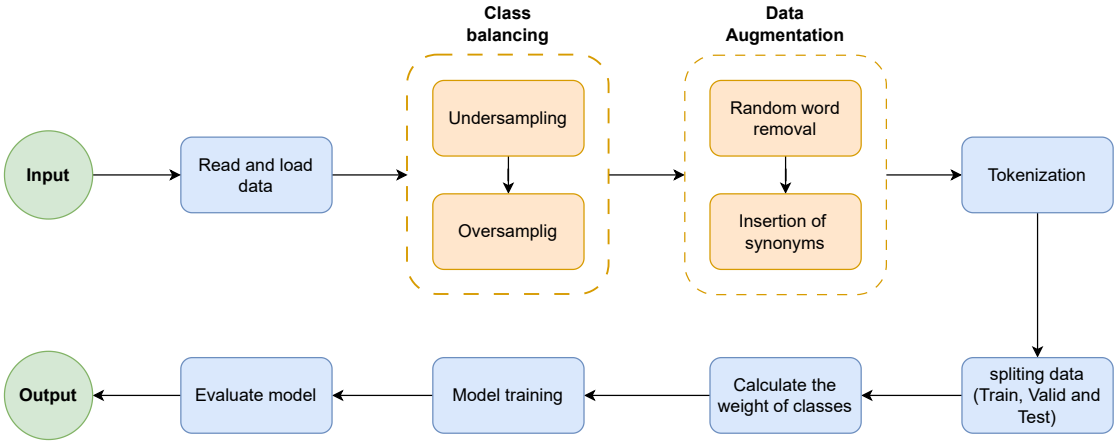

Figure 1: System Pipeline.

4.1. Class Balancing

At this stage, we aimed to reduce the imbalance among the categories in our dataset. To this end, classes 1 and 2 were downsampled using undersampling to 80,000 samples each, while class 5 was oversampled to reach 10,000 samples.

4.2. Data Augmentation

Given that class 5 remained significantly underrepresented compared to the other classes, a 50% probability was applied to perform data augmentation on each instance in this class. Augmentation techniques included synonym replacement and word deletion.

4.3. Loss Weighting

To avoid excessive modification of the dataset through oversampling or undersampling, we opted for a loss weighting strategy, preserving the integrity and original nature of the data—an essential factor for meaningful analysis.

At this stage, class weights were computed based on the current distribution of the data. This strategy aims to mitigate the impact of class imbalance during model training by penalizing classification errors in minority classes more heavily.

The class_weight parameter was employed during the training phase to assign higher penalties to errors in underrepresented classes. This allows the model to better generalize and reduces its bias toward majority classes.

4.4. Feature Extraction

To extract relevant textual features under the adjustments described above, we implemented a fine-tuning process of the RoBERTa Transformer model (roberta-base) specifically for the text classification task [19]. This approach leverages RoBERTa's semantic embeddings, which have shown to significantly enhance classification performance by capturing relationships between words within the text.

5. Evaluation

To measure the model's performance, the metrics of accuracy, F1-score, precision, and recall were analyzed both at the global dataset level and for each individual class. Additionally, the loss function's progression during the training and evaluation phases was continuously monitored to prevent overfitting or poor generalization of the model. Once the dataset was preprocessed as specified in the (Data) section, the resulting loss weights are shown in Table 2

Table 2
Class weights

Class	Weight
Clase 0	0.6110
Clase 1	0.5749
Clase 2	0.5749
Clase 3	2.8321
Clase 4	3.0845
Clase 5	3.0619

Based on the above data, it is evident that the model assigns greater weight to misclassifications in classes 3, 4, and 5, as these are the classes with fewer samples.

To evaluate the performance of the fine-tuned RoBERTa model, the original dataset was split into 80% for training and 20% for testing. This split enabled rigorous analysis during both the training process and the final validation on previously unseen data, allowing subsequent evaluation on the other unseen data provided in the competition's evaluation phase.

5.1. Model testing

We evaluated the model's effectiveness based on the four key metrics previously mentioned, along with the training and validation losses over a total of 13,797 steps within a single full epoch, as presented in Table 3, which shows representative stages of the training process.

Table 3
Training metrics progression

Step	Training Loss	Validation Loss	Accuracy	Macro F1	Macro Precision	Macro Recall
500	1.06	0.51	0.87	0.87	0.88	0.87
1000	0.40	0.36	0.88	0.89	0.91	0.88
1500	0.35	0.27	0.92	0.92	0.93	0.92
6000	0.21	0.19	0.95	0.95	0.95	0.95
6500	0.19	0.26	0.94	0.94	0.95	0.94
13000	0.14	0.14	0.97	0.97	0.97	0.97
13500	0.11	0.15	0.97	0.97	0.97	0.97

The results shown in the table demonstrate a solid and consistent model performance throughout training. A progressive reduction in loss is observed in both training and validation, accompanied by continuous improvements in classification metrics. Notably, the weighted F1 score reached values above 0.96 in the final stages, indicating an efficient balance between precision and recall.

Additionally, the class-wise analysis shows that the model achieves F1 scores ranging from 0.908 to 0.999 across classes, demonstrating its ability to adequately handle the diversity and complexity of the six evaluated categories. These results confirm the robustness and effectiveness of the fine-tuning process, ensuring high performance in the human-AI collaborative classification task.

Table 4
Test dataset evaluation metrics

Metric	Value
Eval Loss	0.1466
Eval Accuracy	0.9689
Macro F1	0.9691
Macro Precision	0.9692
Macro Recall	0.9695
Eval F1 per Class	[0.98, 0.96, 0.96,0.91,0.90,0.99]
Eval Runtime (s)	280.6589
Eval Samples per Second	98.319
Eval Steps per Second	3.075
Epoch	1.0

Upon completion of training, the trained model was evaluated using the test set, corresponding to 20% of the initially reserved data, yielding the metrics shown in Table 4. As observed, the weighted metrics for F1, precision, and recall continue to reflect balanced and consistent performance. Furthermore, the evaluation was completed in 280.66 seconds, with a processing throughput close to 98 samples per second, indicating efficiency in model deployment.

5.2. Competition Evaluation

The results revealed that the strategy implemented by our team, identified as VerbaNex registered in the challenge under the name "gsdeyson", ranked 15th among the participating teams in the PAN-CLEF 2025 Subtask 2 challenge. The detailed performance metrics provided by the competition organizers are presented in Table 5.

Table 5
Ranking comparison - Selected positions

Rank	Team Name	Macro Recall	Macro F1	Accuracy
14	e.chensey	49.56%	50.1%	58.96%
-	Baseline	48.32%	47.82%	57.09%
15	gsdeyson	47.15%	47.15%	56.24%
16	tmarchian	44.33%	42.76%	51.42%

Based on the table, we note that our model placed slightly below the established baseline and the team in 14th position. With a Macro Recall and Macro F1 of 47.15% and an overall accuracy of 56.24%, the model's performance highlights significant areas for improvement.

These results demonstrate that, although the model is capable of addressing the complexity of the task, its ability to generalize and discriminate among the different categories has not yet reached the expected level compared to the competition. Factors such as class balancing, preprocessing quality, or hyperparameter optimization could be revisited to enhance performance.

5.3. Post-competition evaluation

Analyzing the training data, since the validation loss continued to decrease and did not exceed the training loss, the model was deemed susceptible to further training. Therefore, the training was extended to 3 epochs, and predictions were made on the evaluation dataset. Unfortunately, this submission was made after the deadline; however, the PAN-CLEF 2025 organizers kindly provided the prediction results from this last submission, as shown in the Table 6.

Table 6
Results for submission post-competition

Num Epoch	Macro Recall	Macro F1	Accuracy
3	49.19%	48.5%	56.94%

This confirms that increasing the number of epochs effectively improved the model's predictions.

6. Conclusions

This study presents the methodology and results obtained in the Voight-Kampff Generative AI Detection 2025 competition. Our system incorporated class balancing techniques through oversampling and undersampling to address data distribution imbalance, as well as data augmentation strategies based on random word deletion and random synonym insertion to increase the diversity of minority classes. During the training process, class-specific weight calculation was implemented as a complementary measure to mitigate persistent imbalance. This methodology was applied to fine-tune the RoBERTa-base model, resulting in a 15th place ranking in the competition, highlighting the need for further refinements in our approach.

The main areas for improvement identified include: firstly, more effectively addressing the issue of imbalanced data distribution by selecting more suitable feature engineering techniques, such as incorporating additional semantic and linguistic features into RoBERTa's base representations. Secondly, it is necessary to implement more robust multilingual models that allow better generalization across texts in different languages. Finally, a rigorous study and selection of hyperparameters associated with the model training process is required.

Our research contributes to the effort to understand and detect texts generated by large language models, as well as texts written by humans and outputs resulting from human-AI collaboration. We remain committed to the continuous advancement of our methodology and the refinement of our model to accurately identify authentic and original literary productions, distinguishing them from synthetic outputs generated by language models.

Acknowledgments

The authors would like to acknowledge the support provided by the master's degree scholarship program in engineering at the Universidad Tecnologica de Bolivar (UTB) in Cartagena, Colombia.

Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-4 for grammar, spelling, and translation assistance. After using this tool, the author(s) reviewed and edited the content as needed and take full(s)

responsibility for the publication's content.

References

[1] Z. Zeng, S. Liu, L. Sha, Z. Li, K. Yang, S. Liu, D. Gašević, G. Chen, Detecting ai-generated sentences in human-ai collaborative hybrid texts: Challenges, strategies, and insights, 2024. URL: https://arxiv.org/abs/2403.03506. arXiv:2403.03506.

[2] 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 of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, ????, pp. 8384–8395. URL: https://aclanthology.org/2020.emnlp-main.673/. doi:10.18653/v1/2020.emnlp-main.673.

[3] M. Jakesch, J. T. Hancock, M. Naaman, Human heuristics for AI-generated language are flawed 120 (????) e2208839120. URL: https://www.pnas.org/doi/10.1073/pnas.2208839120. doi:10.1073/pnas.2208839120, publisher: Proceedings of the National Academy of Sciences.

[4] J. Bevendorff, M. Wiegmann, J. Karlgren, L. Dürlich, E. Gogoulou, A. Talman, E. Stamatatos, M. Potthast, B. Stein, Overview of the "voight-kampff" generative AI authorship verification task at PAN and ELOQUENT~2024 (????).

[5] A. Richburg, C. Bao, M. Carpuat, Automatic authorship analysis in human-AI collaborative writing, in: N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, Torino, Italia, 2024, pp. 1845–1855. URL: https://aclanthology.org/2024.lrec-main.165/.

[6] G. Bao, Y. Zhao, Z. Teng, L. Yang, Y. Zhang, Fast-detectgpt: Efficient zero-shot detection of machine-generated text via conditional probability curvature, 2024. URL: https://arxiv.org/abs/2310.05130. arXiv:2310.05130.

[7] M. I. H. Rujeedawa, S. Pudaruth, V. Malele, Unmasking ai-generated texts using linguistic and stylistic features, International Journal of Advanced Computer Science and Applications 16 (2025). URL: http://dx.doi.org/10.14569/IJACSA.2025.0160321. doi:10.14569/IJACSA.2025.0160321.

[8] A. Shah, P. Ranka, U. Dedhia, S. Prasad, S. Muni, K. Bhowmick, Detecting and unmasking ai-generated texts through explainable artificial intelligence using stylistic features, International Journal of Advanced Computer Science and Applications 14 (2023). URL: http://dx.doi.org/10.14569/IJACSA.2023.01410110. doi:10.14569/IJACSA.2023.01410110.

[9] M. Ragab, E. B. Ashary, F. Kateb, A. Hakeem, R. Mosli, N. N. Albogami, S. Nooh, Classification of human-written and ai-generated sentences using a hybrid cnn-gru model optimized by the spotted hyena algorithm, Alexandria Engineering Journal 126 (2025) 116–130. URL: https://www.sciencedirect.com/science/article/pii/S1110016825005666. doi:https://doi.org/10.1016/j.aej.2025.04.071.

[10] M. Maktabdar Oghaz, L. Babu Saheer, K. Dhame, G. Singaram, Detection and classification of chatgpt-generated content using deep transformer models, Frontiers in Artificial Intelligence Volume 8 - 2025 (2025). URL: https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2025.1458707. doi:10.3389/frai.2025.1458707.

[11] T. Boran, M. Martinaj, M. S. Hossain, Authorship identification on limited samplings, Computers & Security 97 (2020) 101943. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167404820302194. doi:10.1016/j.cose.2020.101943.

[12] A. Mizumoto, S. Yasuda, Y. Tamura, Identifying ChatGPT-generated texts in EFL students' writing: Through comparative analysis of linguistic fingerprints, Applied Corpus Linguistics 4 (2024) 100106. URL: https://linkinghub.elsevier.com/retrieve/pii/S2666799124000236. doi:10.1016/j.acorp.2024.100106.

[13] A. Fiedler, J. Döpke, Do humans identify AI-generated text better than machines? Evidence based on excerpts from German theses, International Review of Economics Education 49 (2025)

100321. URL: https://linkinghub.elsevier.com/retrieve/pii/S1477388025000131. doi:10.1016/j.iree.2025.100321.
[14] L. Berriche, S. Larabi-Marie-Sainte, Unveiling ChatGPT text using writing style, Heliyon 10 (2024) e32976. URL: https://linkinghub.elsevier.com/retrieve/pii/S2405844024090078. doi:10.1016/j.heliyon.2024.e32976.
[15] H. Desaire, A. E. Chua, M. Isom, R. Jarosova, D. Hua, Distinguishing academic science writing from humans or ChatGPT with over 99% accuracy using off-the-shelf machine learning tools, Cell Reports Physical Science 4 (2023) 101426. URL: https://linkinghub.elsevier.com/retrieve/pii/S266638642300200X. doi:10.1016/j.xcrp.2023.101426.
[16] H. T. Lau, A. Zubiaga, Understanding the effects of human-written paraphrases in LLM-generated text detection, Natural Language Processing Journal 11 (2025) 100151. URL: https://linkinghub.elsevier.com/retrieve/pii/S2949719125000275. doi:10.1016/j.nlp.2025.100151.
[17] J. Bevendorff, D. Dementieva, M. Fröbe, B. Gipp, A. Greiner-Petter, J. Karlgren, M. Mayerl, P. Nakov, A. Panchenko, M. Potthast, A. Shelmanov, E. Stamatatos, B. Stein, Y. Wang, M. Wiegmann, E. Zangerle, Overview of PAN 2025: Voight-Kampff Generative AI Detection, Multilingual Text Detoxification, Multi-Author Writing Style Analysis, and Generative Plagiarism Detection, in: J. C. de Albornoz, J. Gonzalo, L. Plaza, A. G. S. de Herrera, J. Mothe, F. Piroi, P. Rosso, D. Spina, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Sixteenth International Conference of the CLEF Association (CLEF 2025), Lecture Notes in Computer Science, Springer, Berlin Heidelberg New York, 2025.
[18] J. Bevendorff, Y. Wang, J. Karlgren, M. Wiegmann, M. Fröbe, A. Tsivgun, J. Su, Z. Xie, M. Abassy, J. Mansurov, R. Xing, M. N. Ta, K. A. Elozeiri, T. Gu, R. V. Tomar, J. Geng, E. Artemova, A. Shelmanov, N. Habash, E. Stamatatos, I. Gurevych, P. Nakov, M. Potthast, B. Stein, Overview of the "Voight-Kampff" Generative AI Authorship Verification Task at PAN and ELOQUENT 2025, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), Working Notes of CLEF 2025 – Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, CEUR-WS.org, 2025.
[19] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized BERT pretraining approach, CoRR abs/1907.11692 (2019). URL: http://arxiv.org/abs/1907.11692. arXiv:1907.11692.

A. Online Resources

The GitHub repository containing the implementation and resources of this work is available via:

- GitHub.