Enhancing The Recommendation of High-impact Rare-event Business News for Professionals with LLM-based Augmentation

Felipe Bivort Haiek[1] and Anupriya Ankolekar[1]

[1] ModuleQ, Inc. felipe.bivort@moduleq.com 10080 N. Wolfe Rd, Suite SW3-200
Cupertino California USA 95014

Abstract. Personalized news recommendation has become an essential tool for professionals around the world to keep track of news events matching their interests and alleviate information overload. Beyond personalization, an essential aspect of useful news recommendations for professional use is that they highlight events that are more significant and of higher impact. However, we find that state-of-the-art recommenders struggle to identify and recommend news about significant events. In this paper, we address this gap as follows. To mitigate the relative scarcity of news about significant events, we use an LLM to create a synthetic dataset of significant news seeded from business-relevant news in the MIND dataset. We train four state-of-the-art recommendation models (MINER, UNBERT, UniTRec, Fastformer) with synthetically enhanced versions of a subset of the MIND dataset. We find that this successfully improves the performance of two of the recommendation models on the MIND-large dataset restricted to news about significant events in terms of the MRR, NDCG@5 and Hit@5 metrics and the performance of UNBERT on the AUC metric. The contribution of this paper is three-fold: we highlight news significance as an important aspect of useful news recommendation, we demonstrate the use of generative LLMs to create synthetic datasets for training on rare data and lastly, we demonstrate that augmenting some recommendation models with more significant news improves news recommendation performance on the MIND dataset.

Keywords: Recommender Systems, LLMs, Augmentation

Mejorando la Recomendación de Noticias de Negocios sobre Eventos Raros de Alto Impacto via Aumentación por LLMs

Felipe Bivort Haiek[1] and Anupriya Ankolekar[1]

[1] ModuleQ, Inc. felipe.bivort@moduleq.com 10080 N. Wolfe Rd, Suite SW3-200 Cupertino California USA 95014

Resumen. La recomendación personalizada de noticias se ha convertido en una herramienta esencial para que profesionales de todo el mundo puedan mantenerse al tanto de eventos noticiosos que se ajustan a sus intereses y, al mismo tiempo, reducir la sobrecarga de información. Más allá de la personalización, un aspecto clave de las recomendaciones de noticias útiles para uso profesional es que resalten eventos más significativos y de mayor impacto. Sin embargo, encontramos que los sistemas de recomendación más avanzados tienen dificultades para identificar y recomendar noticias sobre eventos significativos. En este artículo, abordamos esta limitación de la siguiente manera. Para mitigar la escasez relativa de noticias sobre eventos significativos, utilizamos un modelo de lenguaje grande (LLM) para crear un conjunto de datos sintético de noticias significativas a partir de noticias relevantes para el ámbito empresarial del conjunto de datos MIND. Entrenamos cuatro modelos de recomendación de última generación (MINER, UNBERT, UniTRec, Fastformer) con versiones mejoradas sintéticamente de un subconjunto del conjunto de datos MIND. Encontramos que esto mejora con éxito el rendimiento de dos de los modelos de recomendación en el conjunto de datos MIND-large, restringido a noticias sobre eventos significativos, en términos de las métricas MRR, NDCG@5 y Hit@5, así como el rendimiento de UNBERT en la métrica AUC. La contribución de este trabajo es triple: destacamos la importancia de la significancia de las noticias como un aspecto clave en la recomendación útil de noticias; demostramos el uso de modelos generativos LLM para crear conjuntos de datos sintéticos que permitan entrenar con datos poco frecuentes; y, finalmente, demostramos que al aumentar algunos modelos de recomendación con noticias más significativas se mejora el rendimiento de la recomendación en el conjunto de datos MIND.

Palabras clave: Sistemas de Recomendación, LLMs, Aumentación

1 Introduction

Personalized news recommendation has become an essential tool (Raza and Ding, 2022) for people around the world to keep track of news events matching their interests and alleviate information overload. They lie at the heart of several prominent news aggregators, such as Apple News and Microsoft News, and are used by general consumers and business professionals alike. However, those who use news recommendation systems to keep track of news and events relevant to their profession, often have additional requirements.

Professionals care not just about the relevance of news to their personal interests, but also about the significance or impact of the news to their work and to global markets and society. Investment bankers, for example, may want to know about a climate event necessitating a move in the location of a company headquarters, or a filing for bankruptcy, or a merger. Commercial data providers offer curated feeds of significant news, e.g., the Refinitiv Significant Developments feed[1], which caters to professionals in finance and investment banking. However, news impact or significance has hitherto not been explicitly modeled in the news recommender literature. Recommender systems typically emphasize performance on click prediction and pay no special attention to significant or high-impact events. As a consequence, state-of-the-art news recommenders do not always correctly rank high-impact news articles, as shown in Figure 1. In this paper, we seek to address this problem and train recommenders to better incorporate news significance in their recommendations.

[1] https://www.lseg.com/en/data-analytics/financial-data/company-data/events/earnings-transcripts-briefs/reuters-investor-briefs-significant-developments

Original	With Augmentation
Raise Billions From Billionaires? Tax Experts Say It's Not That Simple ***[0.67]***	Deere workers grapple with fallout from Trump's trade war ***[0.18]***
After four years of controversy, embattled Palatine-based school district grants transgender students unrestricted locker room access ***[0.66]***	Raise Billions From Billionaires? Tax Experts Say It's Not That Simple ***[0.17]***
Homeless woman living in SUV with dogs moves into apartment with community's help ***[0.65]***	After four years of controversy, embattled Palatine-based school district grants transgender students unrestricted locker room access ***[0.15]***
Deere workers grapple with fallout from Trump's trade war ***[0.64]***	Homeless woman living in SUV with dogs moves into apartment with community's help ***[0.15]***

Fig. 1. Recommendation ranking and scores of news articles with model scores by UNBERT, when trained on a unmodified set of news articles (Original), and when trained by our method (With Augmentation). News considered more significant for professionals is marked with a darker green background.

While training to predict user clicks should incorporate news significance, in reality, recommenders models likely do not encounter enough examples in their training data to learn to interpret the relative significance of news. This is primarily because news events of high significance are often rare, e.g., climate disasters, such as the recent floods in Brazil, or they are rare in a user's click-history. For example, users interested in large technology companies with deep pockets are unlikely to receive and therefore click on news about recent bankruptcies in the companies they follow.

Another aspect of higher-impact news is that they tend to report higher monetary impact or human cost, contain more citations of authoritative sources and include more specific details or in-depth descriptions. Training for credibility, informativeness and quantitative impact can help identify more useful recommendations for business users.

Although such examples are hard to gather naturally, it is becoming increasingly feasible to create a synthetic data set of this nature that includes an outsize proportion of rare events and articles with greater impact and informativeness. Instruction-tuned LLMs have great flair for generating language given guidance around content and style, and enable the possibility of large-scale creation of textual content. To ensure that the news is grounded in the real world, we augment an existing news recommendation dataset, MIND, with articles about high-impact, but relatively uncommon events. In addition, we prompt LLMs to augment news articles to enhance their impact, credibility or informativeness.

Together, these textual data augmentations enable us to create a rich dataset of significant news that we can use for training recommendation models.

With this dataset and associated impression sets from MIND, we train four state-of-the-art recommendation models (MINER, UNBERT, UniTRec, Fastformer). We find that training on an augmented version of MIND successfully improves the performance of all four recommendation models on the MIND-Large dataset restricted to news about significant events in terms of the MRR, NDCG@5 and Hit@5 metrics. In addition, UNBERT improves on the AUC metric.

2 Related Work

Impact of News. The impact of news events has been studied extensively within the literature of news-based stock trading (Fataliyev et al., 2021; Y. Li et al., 2022), which focuses closely on significant events and stock-market signals to make stock price predictions. These models are typically closely tied to stock market prediction and are hard to use in a news recommendation context. An independent stream of research focuses on extracting events from news and detecting their magnitude (Agrawal et al., 2016; Mehta et al., 2019; Xiang & Wang, 2019), often on news of a social or societal nature (e.g., migration and terrorism), rather than on business news that is the focus of this paper.

Data Augmentation. Data augmentation uses existing real-world data to create synthetic examples tweaked to contain certain desired characteristics. In computer vision, image data augmentation has been used with great success to train generalizable models and derive greater value from limited training datasets (Shorten & Khoshgoftaar, 2019). Data augmentation for NLP tasks has thus far been limited (Chen et al., 2024; Kou et al., 2004; B. Li et al., 2022; Tian et al., 2024; L. Wu et al., 2024), with augmentations either resulting in drops of text coherence and grammatical correctness or necessitating the training of specialized augmentation models; but generative LLMs have opened up a wide range of textual data augmentation that can be done with relative ease and at low cost for different types of NLP tasks.

News Recommendation. Personalized news recommendation has been studied extensively over the past decades (C. Wu et al., 2023) and modern methods are fairly effective at modeling user interests and successfully recommending engaging content for them (J. Li et al., 2022; C. Zhang et al., 2024). Many of these methods have been developed using the MIND dataset (F. Wu et al., 2020), a large-scale dataset of consumer English news drawn from the Microsoft News website, containing raw text of news and anonymized click history of users. In this paper, we focus on some of the best-performing algorithms on the MIND dataset: (1) MINER (J. Li et al., 2022), the top-ranked model in the public leaderboard of the MIND dataset, which uses a PLM as a text encoder and a poly attention scheme to extract multiple user interest vectors for user representation; (2) Fastformer (C. Wu et al., 2021), an efficient transformer architecture based on additive attention; (3) UNBERT (Q. Zhang et al., 2021), which con-

catenates a candidate's title and all the history titles into a string which then goes into two Bert encoders; (4) UniTRec (Mao et al., 2023), which utilizes an encoder-decoder architecture (BART) and encodes user history by encoder and candidate content by decoder.

		MIND-Business				MIND-Rare			
Metric	Condition	MINER	Fastformer	UniTRec	UNBERT	MINER	Fastformer	UniTRec	UNBERT
AUC	Original	0.609†	0.660	0.691	0.682	0.628†	0.634	0.627	0.683
	Augmented	0.597	0.618	0.682	**0.692***	0.614	0.532	0.624	**0.693**
MRR	Original	0.420	0.423	**0.478**	0.424	0.214	0.2024	0.205	0.225
	Augmented	0.453	0.368	0.475	0.452	**0.232***	0.131	0.198	0.231
NDCG@5	Original	0.465	0.466	0.526	0.472	0.226	0.214	0.210	0.236
	Augmented	0.502*	0.398	**0.526**	0.500*	**0.249***	0.112	0.201	0.248*
Hit@5	Original	0.690	0.683	0.701	0.699	0.426	0.391	0.292	0.420
	Augmented	**0.726***	0.600	0.700	0.716*	**0.461***	0.229	0.281	0.445*

Table 1. Recommendation model performance on the MIND-Business validation set and on MIND-Rare, a subset of impressions containing news manually labeled as rare and higher impact. The best scores for each metric in each dataset are highlighted in bold. Statistically significant improvements are marked with a *. The metrics generally improve with augmentation in Miner's and UNBERT's cases—the exceptions are marked with †.

3 Approach

Our approach is as follows: we identify a subset of news in the MIND dataset that is relevant to business professionals; we then create alternative versions of these news articles using a Generative LLM. We train SOTA recommendation models (MINER, UNBERT, Fastformer, and UniTRec) on the dataset and evaluate their performance on two validation datasets of general business-relevant news and, business-relevant news that we manually labeled as high-impact and rare. In each case, we evaluate model performance after training on either a subset of the original MIND dataset or an augmented version of the original MIND news articles.

3.1 Dataset

We use the MIND dataset (F. Wu et al., 2020) for news recommendation, a large public dataset of anonymized impression and click logs collected in late 2019 from a set of randomly-sampled million users browsing the Microsoft News website[2]. Each news article in the dataset is characterized by a title, an abstract,

[2] https://msnews.github.io/

two levels of categories, and a list of extracted entities. Following prior work (J. Li et al., 2022; C. Zhang et al., 2024), we use the small version of the MIND dataset, a random sample of 50,000 users' behaviors distilled from the full MIND dataset. The small version is just a subsample of the the train and test sets of the large version. In principle this allows the user to train in MIND small train and then validate in MIND large validation without leakeage.

The MIND dataset contains a broad set of primarily consumer-facing English news of general interest. In order to focus on news stories that are more suitable for enterprise users, we created a new dataset *MIND-Business*, containing news articles from the MIND-Small dataset restricted to the categories *News* and *Finance*.

The behavior logs in the training and validation sets are further constrained to contain only references to this restricted subset of articles. During training, we further filtered out Impressions, not including at least one positive, which amounted to 67400 Impression sets. The train set is further partitioned; we extract 2500 samples at random to use for quick testing and only train on the remaining impressions from the MIND-Business training set. Altogether, the MIND-Business set contains 23,825 news articles with 203,092 impressions and 112,601 clicks in the click history.

Augmentation	AUC	gAUC	MRR	NDCG@5	NDCG@10	Hit@5	Hit@10
None	0.683	0.682	0.224	0.236	0.302	0.420	0.612
Rare	0.684	0.684	0.226	0.237	0.302	0.429	0.619
Rare + Enhancement (High)	0.692	0.680	0.225	0.238	0.307	0.428	0.628
Rare + Enhancement (High & Low)	**0.693**	**0.689**	**0.231**	**0.248**	**0.316**	**0.445**	**0.639**

Table 2. Ablation study of the impact of different types of augmentation (Low, High and Rare) on the performance of UNBERT on MIND-Rare.

3.2 Text Augmentation

With MIND-Business as a base set of business-relevant news articles, we modified the text of news stories in two different ways: (1) increasing the reported impact of a news story at two different levels *Low* and *High*, and (2) changing the focus of the news story to be on significant, but uncommon events *Rare*. In both cases, the Mistral-7B LLM was used to perform the text augmentation.

For the first type of augmentation, we prompted the LLM to increase the reported impact in terms of monetary cost or the number of affected stakeholders, to increase the informativeness or the credibility and reliability of the article, e.g. by mentioning more specific details or by citing reputable and authoritative sources. For the second type of augmentation, we prompted the model to identify the primary entities in each article and then asked it to write a new article on a sampled high-impact event. High-impact events were of several possible types: political and economic developments with far-reaching impact on

the global economy; natural disasters, such as hurricanes or earthquakes; health crises or pandemics; regulatory changes with market-impact; cyberattacks on governments, companies or essential infrastructure; environmental issues (climate change) with long-lasting impact; trade wars and tariffs and lastly sudden corporate bankruptcies. See Appendices A.1 and A.2 for more detail on the prompts used.

These generated high-impact news help us find real high-impact news. We trained a model to distinguish between synthetic news about high-impact events and randomly sampled news from MIND-Business. Specifically, we fine-tuned the DebertaV3 (He et al., 2022) architecture on 80% of synthesized rare events and as many news at random from Mind-Business, the remaining 20% was used for validation. The training lasted 3 epochs, using the Adam optimizer (Kingma & Ba, 2014) and batch size 16. The resulting model achieves a ROC-AUC of 0.98 on a separate validation set. This suggests that high-impact news are indeed rare and not widely present in the MIND dataset. We used this classifier to sort news articles from MIND-large test and manually chose the top results that best matched our definition of rare and high-impact news. We will refer to the 7418 entries subset of MIND-large validation's logs containing news articles from the aforementioned manually labeled set, as MIND-Rare. Note that MIND-rare is constructed from Mind-large because of the low signal of rare news on Mind-small test portion.

3.3 Training

We train four state-of-the-art recommendation models both on the original dataset and with augmentation, while keeping the hyperparamaters form each model's original publication. In the original dataset, each training instance consists of the user history of clicked articles and a list of candidate articles in an impression. We follow the common practice of using negative sampling with four negative examples.

When training with augmentations, we swap the positively-labeled news article with a randomly chosen augmented version of the same article. Since the augmentation only increases the impact of the article, this should not affect the true label of the article. In other words, the user is just equally or more likely to click the article with the augmentation.

4 Results

Our results are summarized in Table 1. Note that, for all models without augmentation, performance on MIND-Rare is lower than MIND-Business (except for MINER-AUC). This suggests that high-impact events are truly harder for recommenders to handle, and there exists a need to improve recommender performance on such news. Examining the results for MIND-Business, we observe that different models behave differently using augmented training data. UNBERT shows improvement on all metrics, while MINER's performance tends to

improve (aside from a slight decrease in AUC). UniTRec shows negligible improvement, while Fastformer seems to be deeply penalized by training on augmented data. The largest improvements are achieved by augmented MINER on Hit@5 and augmented UNBERT on AUC, NDCG@5 and Hit@5. These achieve statistical significance compared to the original condition using an unpaired t-test at p-value 0.005.

The simultaneous decrease in AUC and increase in NDCG@5 and Hit@5 for MINER may indicate good within-candidate-set tuning and poor overall affinity classification. AUC compares all candidates scores at the same time and does not average list-wise metrics. Augmentations that change the event (Rare) likely distort MINER's user model and therefore have negative impact on performance.

Examining the results for MIND-Rare, we observe a lift in the metrics for each recommendation model. As before, UNBERT achieves the best results and the most consistent improvement. The improvement in NDCG@5 and Hit@5 are statistically significant using an unpaired t-test with p-value 0.05. MINER also shows consistent improvement in all metrics (besides AUC). We also measured these metrics on a dataset of the most high-impact 75 news articles predicted by the classifier in section 3.2 obtaining the same improvements at a p-value of 0.005 for all models.

We conducted an ablation study on the different types of augmentation to assess the relative contribution of each, by training UNBERT on datasets with different types of augmentations. The results are shown in Table 2. We observe that increasing the augmentation positively impacts all metrics. The metric with the greatest improvement with each augmentation is Hit@10, which suggests that the model is able to accurately position a clicked article within the top ten recommendations in more candidate sets. We repeated this analysis on the most high-impact articles predicted by the classifier and noted the same trends with statistical significance.

5 Discussion

We observe that we have obtained the most consistent improvement on the MRR, NDCG, and Hit metrics using MINER and UNBERT, where augmentation outperforms the baseline not only in the filtered validation dataset (MIND-Rare) but also in the full dataset (MIND-Business). It is interesting to note that both MINER and UNBERT represent the user with more than one embedding vector, one for each interest for the case of MINER, and one along title words and another along titles for UNBERT. This may help the models to focus on different types of entities and/or characteristics in the articles the users choose and increase the available representation space, making the models able to learn embedding vectors that better match each augmentation.

To illustrate the improvement with augmentations, note the position and score of the article about Deere workers in Figure 1 presented at the beginning of the paper. That item is considered to be a high-impact item that also happens to be a positive candidate in an impression set. Training with augmentation

manages to move it three positions higher. For more examples we point the reader to Appendix B.

An open question is whether there is an optimal way to select which fraction of examples from each augmentation type we use for training. In the same vein, we posit that there should be a way to profit from the relative hierarchy of augmentation, where the model should be trained to prefer a high-impact topic change instead of a minor increase in the number of stakeholders impacted. We expect better-quality augmentation to increase the system's performance even more. This could be achieved by further prompt-tuning, using methodologies such as Chain of Thought, or testing different generative LLMs.

6 Conclusion

Business professionals care not just about the relevance of news to their personal interests, but also about the significance or impact of the news to their work and to global markets and society. State-of-the-art recommendation systems currently do not always rank significant news correctly. In this paper, we set out to address this problem and train recommenders to better incorporate news significance in their recommendations. To mitigate the relative scarcity of news about significant events, we used an LLM to create a synthetic dataset of significant news seeded from business-relevant news in the MIND dataset. We trained four state-of-the-art recommendation models (MINER, UNBERT, UniTRec, Fastformer) with synthetically enhanced versions of a subset of the MIND dataset. We show that the four recommendation models find it harder to make predictions on candidate sets of news articles containing high-impact, but rare events. We found that this successfully improves the performance of two of the recommendation models on the MIND-large dataset restricted to news about significant events in terms of the MRR, NDCG@5 and Hit@5 metrics and the performance of UNBERT on the AUC metric. Our approach of adding LLM-generated augmentations improves performance on impression sets containing news about rare events and also improves the general performance of the model in some cases. The contribution of this paper is three-fold: we highlighted news significance as an important aspect of useful news recommendation, we demonstrated the use of generative LLMs to create synthetic datasets for training on rare data and lastly, we demonstrated that augmenting recommendation models with more significant news improves news recommendation performance on the MIND dataset.

References

Agrawal, A., Sahdev, R., Davoudi, H., Khonsari, F., An, A., & McGrath, S. (2016). Detecting the Magnitude of Events from News Articles. *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 177–184. https://doi.org/10.1109/WI.2016.0034

Chen, C.-T., Lee, C., Huang, S.-H., & Peng, W.-C. (2024). Credit Card Fraud Detection via Intelligent Sampling and Self-supervised Learning. *ACM Transactions on Intelligent Systems and Technology*, *15*(2), 35:1–35:29. https://doi.org/10.1145/3641283

Fataliyev, K., Chivukula, A., Prasad, M., & Liu, W. (2021, July). Stock Market Analysis with Text Data: A Review [arXiv:2106.12985 [cs, q-fin]]. https://doi.org/10.48550/arXiv.2106.12985

He, P., Gao, J., & Chen, W. (2022). Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. *The Eleventh International Conference on Learning Representations*.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kou, Y., Lu, C.-T., Sirwongwattana, S., & Huang, Y.-P. (2004). Survey of fraud detection techniques [ISSN: 1810-7869]. *IEEE International Conference on Networking, Sensing and Control, 2004*, *2*, 749–754 Vol.2. https://doi.org/10.1109/ICNSC.2004.1297040

Li, B., Hou, Y., & Che, W. (2022). Data augmentation approaches in natural language processing: A survey. *Ai Open*, *3*, 71–90.

Li, J., Zhu, J., Bi, Q., Cai, G., Shang, L., Dong, Z., Jiang, X., & Liu, Q. (2022, May). MINER: Multi-Interest Matching Network for News Recommendation. In S. Muresan, P. Nakov, & A. Villavicencio (Eds.), *Findings of the Association for Computational Linguistics: ACL 2022* (pp. 343–352). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.findings-acl.29

Li, Y., Lv, S., Liu, X., & Zhang, Q. (2022). Incorporating transformers and attention networks for stock movement prediction. *Complexity*, *2022*, 1–10.

Mao, Z., Wang, H., Du, Y., & Wong, K.-F. (2023, July). UniTRec: A Unified Text-to-Text Transformer and Joint Contrastive Learning Framework for Text-based Recommendation. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 1160–1170). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-short.100

Mehta, S., Islam, M. R., Rangwala, H., & Ramakrishnan, N. (2019). Event Detection using Hierarchical Multi-Aspect Attention. *The World Wide Web Conference*, 3079–3085. https://doi.org/10.1145/3308558.3313659

Raza, S., & Ding, C. (2022). News recommender system: A review of recent progress, challenges, and opportunities. *Artificial Intelligence Review*, *55*(1), 749–800. https://doi.org/10.1007/s10462-021-10043-x

Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0197-0

Tian, M., Liang, J., Zhang, D., Zhang, X., Wang, Z., & Li, H. (2024). Detection of Financial Fraudulent Activities with Machine Learning:A Case

Study of Detecting Potential Tax and Invoice Fraud. *Proceedings of the 2023 7th International Conference on Computer Science and Artificial Intelligence*, 33–39. https://doi.org/10.1145/3638584.3638669

Wu, C., Wu, F., Huang, Y., & Xie, X. (2023). Personalized News Recommendation: Methods and Challenges. *ACM Transactions on Information Systems*, *41*(1), 24:1–24:50. https://doi.org/10.1145/3530257

Wu, C., Wu, F., Qi, T., Huang, Y., & Xie, X. (2021, September). Fastformer: Additive Attention Can Be All You Need [arXiv:2108.09084 [cs]]. https://doi.org/10.48550/arXiv.2108.09084

Wu, F., Qiao, Y., Chen, J.-H., Wu, C., Qi, T., Lian, J., Liu, D., Xie, X., Gao, J., Wu, W., & Zhou, M. (2020, July). MIND: A Large-scale Dataset for News Recommendation. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 3597–3606). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.331

Wu, L., Zhang, F., Cheng, C., & Song, S. (2024). Supervised Contrast Learning Text Classification Model Based on Data Quality Augmentation. *ACM Transactions on Asian and Low-Resource Language Information Processing*. https://doi.org/10.1145/3653300

Xiang, W., & Wang, B. (2019). A Survey of Event Extraction From Text [Conference Name: IEEE Access]. *IEEE Access*, *7*, 173111–173137. https://doi.org/10.1109/ACCESS.2019.2956831

Zhang, C., Sun, Y., Chen, J., Lei, J., Abdul-Mageed, M., Wang, S., Jin, R., Park, S., Yao, N., & Long, B. (2024, February). SPAR: Personalized Content-Based Recommendation via Long Engagement Attention [arXiv:2402.10555 [cs]]. https://doi.org/10.48550/arXiv.2402.10555

Zhang, Q., Li, J., Jia, Q., Wang, C., Zhu, J., Wang, Z., & He, X. (2021). UNBERT: User-News Matching BERT for News Recommendation [ISSN: 1045-0823], *3*, 3356–3362. https://doi.org/10.24963/ijcai.2021/462

A Prompt Instructions

A.1 Impact, Informativeness & Credibility for the Low and High Augmentations

- Monetary Impact: The degree of impact an event has can often be measured by the scale of monetary value involved. The larger the amount of money implicated, either in terms of cost, revenue, investment, or loss, the more significant the event can be considered. Please generate specific figures.
- Number of stakeholders impacted: The larger the number of stakeholders affected, the more profound the potential implications. This can range from employees and shareholders to customers, suppliers and residents. Include ONLY one of the examples and quantify the impact in terms of number of people
- The level of detail and specificity provided in the article. One article may provide more in-depth information about the event, including its causes, effects, and potential solutions. Please generate specific data, figures, numbers and details that are not included in the article.
- The credibility and reliability of the sources cited in the article. An article that cites reputable sources, such as experts in the field or government agencies, may be more helpful for an investment banker as it provides a more reliable basis for making informed decisions.

A.2 High-Impact News Categories for the Rare Augmentation

- Political and economic developments: Political and economic developments in key markets can have a significant impact on the global economy and financial markets. For example, changes in government policies, trade agreements, and geopolitical tensions can affect the value of currencies, commodity prices, and stock markets
- Natural disasters: Natural disasters such as hurricanes, earthquakes, and floods can cause significant damage to infrastructure, disrupt supply chains, and impact the economy. This can lead to increased volatility in financial markets and affect the value of investments.Include ONLY one of the possible events given.
- Health crises: Health crises such as pandemics can have a significant impact on the global economy and financial markets. For example, the COVID-19 pandemic has caused widespread disruptions to global supply chains, reduced demand for certain goods and services, and led to increased volatility in financial markets.
- Regulatory Changes: Sudden changes in financial regulations can affect how banks operate and their profitability. For instance, the introduction of the Dodd-Frank Act after the 2008 financial crisis had significant implications for investment banks.
- Cyber Attacks: While companies and governments take precautions against these, a significant cyber attack on a major institution or infrastructure could have far-reaching impacts on the market.

– Environmental issues: Environmental issues such as climate change and resource depletion can have a significant impact on the global economy and financial markets. For example, changes in weather patterns and resource availability can affect agricultural production, energy prices, and the value of investments.Include ONLY one of the possible events given
– Trade Wars and Tariffs: Sudden escalations in trade wars or the imposition of new tariffs are events that can impact global trade and consequently, the financial markets.Include ONLY one of the possible events given
– Sudden Corporate Bankruptcies: The collapse of a major company or financial institution, such as the bankruptcy of Lehman Brothers in 2008, can have a ripple effect on the economy and financial markets.

A.3 Example augmentations

We present sample augmentation results on Table 3.

B Mislabeling example

On Figure 2 we present the rankings for the UNBERT model on synthetic articles. The original version of the model fails to sort articles by impact properly.

Augmented News
"Inflation … is not quite where we would like to see it." Lesetja Kganyago
London Stock Exchange Group Faces Major Trading Glitch, Swiftly Resolves the Issue
London Stock Exchange Group Faces Catastrophic Systems Failure, Almost a Billion in losses
London Stock Exchange Group Faces Systems Failure, Prompting Urgent Recovery Efforts
London Stock Exchange Group shares inched up 0.28% to £92.10 Monday
LSEG`s Schwimmer: Regulators should be 'careful' about imposing rules on AI

Fig. 2. Recommendation ranking of synthetic candidate news articles with model scores by UNBERT. The more intense the color, the higher the score given by augmented-UNBERT.

Original	High Enhancement	Low Ennhancement	Topic Change
The Cost of Trump's Aid Freeze in the Trenches of Ukraine's War	As the conflict in Ukraine continues to escalate, the Trump administration's decision to freeze aid to the country has left many soldiers on the front lines struggling to survive. Lt. Ivan Molchanets, a veteran of the Ukrainian military, was one such soldier.	The conflict in Ukraine continues to escalate, and the Trump administration's decision to freeze aid to the country has left many soldiers on the front lines struggling to survive	Ukraine's Lt. Ivan Molchanets Faces Financial Uncertainty as Major Company Goes Bankrupt
Man hospitalized after Northeast Austin shooting, EMS medics say	According to a statement from Austin-Travis County EMS, a man was rushed to the hospital Sunday evening after being shot in a Northeast Austin neighborhood	According to a statement from an unnamed source, a man was rushed to the hospital Sunday evening after being shot in a Northeast Austin neighborhood	Grand EMS Medics Faces Financial Uncertainty Due to Regulatory Changes in Northeast Austin
UPFRONT recap: Assembly Speaker says governor is playing politics with special session	Assembly Speaker Robin Vos, R-Rochester, has accused Democratic Gov. Tony Evers of playing politics with the call for a special session on new gun legislation.In an interview aired Sunday on "UPFRONT," produced in partnership with Wisconsin Public Television, Vos criticized Evers for not working in good faith to find common ground on the issue.	Assembly Speaker Robin Vos, R-Rochester, has accused Democratic Gov. Tony Evers of playing politics with the call for a special session on new gun legislation. In an interview, Vos criticized Evers for not working in good faith to find common ground on the issue	Assembly Speaker Robin Vos and Governor Tony Evers Discuss Health Crises and Their Impact on the Global Economy and Financial Markets
Realme takes chunk of India mobile market as Samsung slides	Realme has captured a significant portion of the Indian mobile market, with Samsung experiencing a staggering 400 percent increase in phone shipments year-on-year	Realme has captured a significant portion of the Indian mobile market, with Samsung experiencing a modest 100 percent increase in phone shipments year-on-year	Realme's Profitability Takes a Hit as India's Financial Regulations Change

Table 3. Augmentation samples.