

Detecção de Viés Ideológico em Artigos de Notícias Utilizando Aprendizagem Métrica Profunda e Representações Contextuais

1

Abstract. This study investigates ideological bias identification using metric learning with BERT-based models. It demonstrates that Contrastive and Triplet Loss optimize embedding extraction for political classification, outperforming the literature baseline by 12 percentage points in rigorous evaluations. The model proved competitive against zero-shot LLMs, establishing itself as a robust, efficient alternative that captures semantic nuances beyond outlet memorization, ensuring integrity in analyzing unseen data.

Resumo. Este estudo investiga a identificação de viés ideológico via aprendizagem métrica com modelos pré-treinados baseados em BERT. Demonstra-se que o uso de Contrastive e Triplet Loss otimiza a extração de embeddings para classificação política, superando o baseline da literatura em 12 pontos percentuais em avaliações rigorosas. O modelo mostrou-se competitivo frente a LLMs em modo zero-shot, consolidando-se como uma alternativa robusta e eficiente que captura nuances semânticas além da memorização de veículos de imprensa, garantindo integridade na análise de dados inéditos.

1. Introdução

Nos últimos anos, a expansão exponencial de informações online via portais de notícias impõe o desafio de assegurar a imparcialidade dos conteúdos. O viés ideológico em artigos pode distorcer a percepção pública e influenciar decisões políticas e resultados eleitorais [Gentzkow and Shapiro 2006, Chiang and Knight 2011]. Dada a subjetividade inerente ao discurso, a identificação automatizada dessa inclinação é uma tarefa complexa, tornando o desenvolvimento de métodos precisos de detecção uma área de pesquisa altamente relevante para a integridade da informação.

No campo do Processamento de Linguagem Natural (PLN), abordagens buscam automatizar essa detecção via conteúdo textual, hiperlinks e teoria da informação [Spinde et al. 2021, Patricia Aires et al. 2019]. Contudo, as soluções atuais frequentemente limitam-se a cenários puramente polarizados ou dependente de fontes externas, os que comprometem a autonomia dos métodos e dificulta a identificação de nuances ideológicas em contextos onde metadados não estão disponíveis. Além disso, avanços recentes com modelos de linguagem pré-treinados e sistemas baseados em *Large Language Models* (LLMs) [Lin et al. 2024a] enfrentam críticas por manifestarem vieses políticos nativos de suas arquiteturas e disparidades com a percepção humana [Lin et al. 2025].

Para mitigar essas limitações, este trabalho propõe uma metodologia baseada em *Deep Metric Learning*. Diferente de classificadores tradicionais, nossa abordagem foca na extração de características ideológicas a partir de fragmentos textuais locais, permitindo uma classificação agnóstica à fonte. As principais contribuições deste estudo são:

- Aplicação de *Triplet* e *Contrastive Losses* para estruturar espaços vetoriais onde a proximidade reflete afinidade ideológica;
- Método capaz de identificar vieses em fontes inéditas, superando o sobreajuste comum em modelos baseados em arquiteturas de *Transformers*;
- Uma arquitetura especializada com contagem de parâmetros reduzida, garantindo baixo custo de inferência e operação e hardware convencional.

2. Trabalhos Relacionados

A literatura aborda a detecção de viés ideológico sob diversas óticas, desde redes estruturais [Efron 2004] até modelos econômicos de reputação midiática [Gentzkow and Shapiro 2006]. Avanços integraram análise de redes sociais e PLN para quantificar o viés via gráficos de interação [Lin et al. 2011]. Contudo, tais métodos enfrentam limitações severas em documentos com dados esparsos de hiperlinks ou isolados de redes de citação estabelecidas, limitando a capacidade de generalização do modelo.

Alternativamente, metadados de redes sociais, como interações no Twitter [Rao and Spasojevic, 2016; Elejalde et al., 2017] e perfis demográficos de audiência no Facebook [Ribeiro et al., 2018], auxiliam na inferência de inclinação política. Entretanto, essas abordagens são dependentes de APIs de terceiros e restringem-se frequentemente a cenários de alta polarização binária. Tal dependência reforça a necessidade de soluções que operem exclusivamente sobre o conteúdo textual bruto.

No âmbito da análise textual direta, as investigações priorizam a escolha lexical e a recorrência temática como indicadores de inclinação ideológica [Dallmann et al. 2015], empregando mecanismos de atenção para capturar o enquadramento (*framing*) em manchetes [Gangula et al. 2019]. Embora modelos baseados na arquitetura *Transformer*, como o BERT (*Bidirectional Encoder Representations from Transformers*), atinjam elevada precisão em fontes previamente mapeadas, tais estruturas demonstram fragilidade ao processar artigos de domínios inéditos [Baly et al. 2020].

Para mitigar essa lacuna, o framework *POLITICS* [Liu et al. 2022] introduziu o aprendizado contrastivo, visando alinhar perspectivas distintas sobre um mesmo evento noticioso. Esse esforço foi aprimorado por volf2025political, que realizaram o ajuste fino (*fine-tuning*) do modelo *POLITICS* utilizando múltiplos datasets unificados. Os autores demonstraram que essa especialização, aliada a uma filtragem prévia de textos sobre assuntos políticos, é essencial para assegurar a robustez e a generalização dos classificadores em diversos domínios.

Contudo, apesar do desempenho robusto desses modelos especializados, tanto as versões refinadas quanto sistemas baseados em modelos de larga escala, como o *IndiVec* [Lin et al. 2024a], estão sujeitos a vieses políticos inerentes ou a falhas de alinhamento com a percepção humana em cenários *zero-shot* [Lin et al. 2025]. Diante dessas limitações, este trabalho propõe o emprego de *Deep Metric Learning* fundamentado na técnica de *semi-hard sampling*. Diferente de abordagens puramente classificatórias, esta estratégia visa otimizar a topologia do espaço de características, forçando o modelo a distinguir nuances ideológicas através de uma amostragem mais informativa. Com isso, busca-se superar os gargalos de generalização dos modelos tradicionais sem incorrer na excessiva complexidade computacional típica de arquiteturas de larga escala.

3. Material e Métodos

A metodologia assume que o viés ideológico se manifesta em padrões discursivos, processando textos via modelos pré-treinados e otimizando *embeddings* por *Contrastive* e *Triplet Loss*. Conforme a Figura 1, o fluxo abrange quatro etapas: definição dos conjuntos de dados, geração de representações por aprendizagem métrica, treinamento de classificadores e análise de desempenho. As seções seguintes detalham cada módulo da arquitetura proposta.

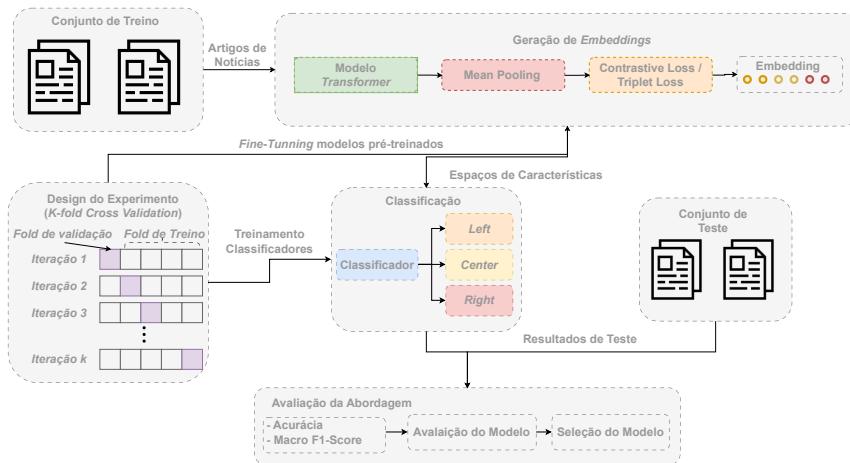


Figura 1. Visão geral do método de detecção de viés ideológico por meio do conteúdo textual de artigos de notícias.

3.1. Dados Experimentais

Para o desenvolvimento deste estudo, utilizaram-se os conjunto de dados detalhados na Tabela 1. O foco principal recai sobre os dados do *Article Bias Prediction* (ABP) [Baly et al. 2020], composto por 37.554 notícias em inglês rotuladas em três categorias: *left*, *center* e *right*. A confiabilidade desses rótulos provém da plataforma AllSides¹, que emprega auditorias rigorosas e revisões da comunidade, conforme ilustrado na Figura 2.

Tabela 1. Resumo dos conjuntos de dados e volume de amostras.

Conjunto de Dados	Classes	Quantidade
FlipBias [Chen et al. 2018]	Left/Center/Right	3.066
ABP [Baly et al. 2020]	Left/Center/Right	37.554

A robustez metodológica foi validada por dois critérios de particionamento. O primeiro, *media-bias split* (Tabela 2), segregou os artigos de modo que os dados de uma mesma fonte de notícia sejam exclusivos a uma única partição, impedindo que veículos presentes no conjunto de treinamento apareçam na validação ou teste. Essa estratégia é fundamental para mitigar o vazamento de dados e avaliar a capacidade de generalização para fontes não vistas. O segundo, *random split* (Tabela 3), mantém a consistência na distribuição de classes entre treino e validação por meio de uma amostragem aleatória.

¹<https://www.allsides.com/media-bias/media-bias-rating-methods>

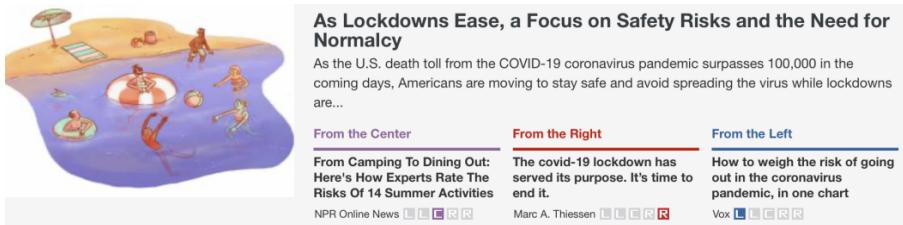


Figura 2. Exemplo de uma amostra de artigo de notícia com o tópico sobre a pandemia de coronavírus [Baly et al. 2020].

Tabela 2. Estatísticas da partição *media-bias split*.

Treino			Validação			Teste		
Viés	Total	%	Viés	Total	%	Viés	Total	%
Left	8.861	33,32%	Left	1.640	69,60%	Left	402	30,92%
Center	7.488	28,16%	Center	618	26,23%	Center	299	23,00%
Right	10.241	38,51%	Right	98	4,15%	Right	599	46,07%

Tabela 3. Estatísticas da partição *random split*.

Treino			Validação			Teste		
Viés	Total	%	Viés	Total	%	Viés	Total	%
Left	9.750	34,84%	Left	2.438	34,84%	Left	402	30,92%
Center	7.988	28,55%	Center	1.998	28,55%	Center	299	23,00%
Right	10.240	36,60%	Right	2.560	36,59%	Right	599	46,07%

Adicionalmente, a generalização foi avaliada com o conjunto *FlipBias* [Chen et al. 2018], que contém 2.781 eventos sob múltiplas perspectivas. Os experimentos seguiram as diretrizes de [Lin et al. 2024b, Lin et al. 2025] para aferir a robustez das representações em dados não expostos durante a etapa de treinamento.

3.2. Tarefa de Geração de *Embeddings*

Para a execução da tarefa, a Figura 3 ilustra o fluxo de processamento adotado. O processo inicia-se com a *tokenização* dos artigos de notícias, seguida pela mineração de exemplos. Esta etapa é fundamental para selecionar amostras informativas que otimizam a convergência e o aprendizado do modelo.

Conforme delineado na arquitetura apresentada, empregaram-se dois modelos fundamentados em *Bidirectional Encoder Representations from Transformers (BERT)* [Devlin et al. 2018], reconhecidos pela eficácia na modelagem de dependências de longo alcance e na extração de relações semânticas granulares [Zhang and Rao 2020]. A seleção recaiu sobre o DistilBERT [Sanh 2019] e o DistilRoBERTa [Liu 2019], variantes destiladas que preservam a robustez das arquiteturas originais, contudo, apresentam reduções substanciais no custo computacional e nos requisitos de memória.

O modelo *Transformer* processa as sequências de entrada, seguido por uma camada de *Mean Pooling* que consolida as representações em um vetor único. O (*fine-tuning*) é regido por estratégias de aprendizagem métrica², utilizando as funções de perda

²Aprendizagem métrica (ou *metric learning*) refere-se ao uso de algoritmos para aprender uma função

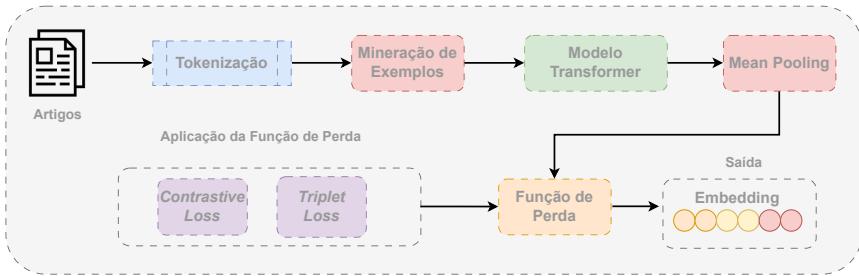


Figura 3. Fluxo de treinamento e geração de características para artigos de notícias.

Contrastive Loss ou *Triplet Loss*. Tal abordagem assegura que os *embeddings* gerados na saída posicionem instâncias contextualmente similares em regiões próximas do espaço de representação, otimizando a discriminação entre as classes.

No que se refere ao pré-processamento, as *stopwords* foram preservadas, visto que a arquitetura BERT demonstra eficácia na extração de nuances contextuais a partir desses elementos. Por fim, o treinamento foi estabelecido com um limite de 100 épocas, utilizando o otimizador Adam com taxa de aprendizado de 10^{-4} e *batch size* de 16. Para mitigar o *overfitting* e assegurar a capacidade de generalização dos modelos, aplicou-se a técnica de *Early Stopping* com paciência de 5 ciclos, monitorando-se a convergência da função de perda no conjunto de validação.

3.2.1. Mecanismos de Aproximação e Distanciamento

Nesta abordagem, os codificadores (DistilBERT e DistilRoBERTa) ajustam os pesos de suas camadas para otimizar a qualidade dos *embeddings* via *Contrastive Loss* e *Triplet Loss*. O objetivo é o aprendizado de representações vetoriais onde instâncias semanticamente similares convirjam no espaço representação, enquanto exemplos dissimilaridades sejam repelidos.

A *Contrastive Loss* é aplicada utilizando a distância Euclidiana sobre pares de exemplos, conforme definido na Equação 1:

$$L = \frac{1}{2}(1 - y)D^2 + \frac{1}{2}y\{\max(0, m - D)\}^2 \quad (1)$$

Onde y representa o rótulo binário (0 para similar, 1 para dissimilar), D denota a distância entre as representações e m é a margem de separação.

Complementarmente, a *Triplet Loss* utiliza triplas compostas por uma âncora (a), um exemplo positivo (p) e um negativo (n). O objetivo, expresso na Equação 2, assegura que a distância entre a âncora e o positivo seja inferior à distância entre a âncora e o negativo por uma margem m :

$$L = \max(0, D(a, p) - D(a, n) + m) \quad (2)$$

de distância que capture a similaridade entre dados.

Para otimizar o aprendizado, empregou-se o *mining* de negativos *semi-hard*. Esses exemplos, que satisfazem a condição $D(a, p) < D(a, n) + m$, fornecem gradientes mais informativos e mitigam o *overfitting* em comparação a negativos *hard* [Kertész 2021]. Esse processo refina a capacidade discriminatória do modelo, permitindo que os *embeddings* capturem relações semânticas profundas, como a ideologia de uma notícia, independentemente da fonte de publicação.

3.3. Tarefa de Classificação: Modelos e Parametrização

Após o mapeamento dos *embeddings*, onde a proximidade entre os vetores reflete a similaridade ideológica das notícias. A classificação dos artigos foi realizada por meio de três algoritmos: *K-Nearest Neighbors (KNN)*, *K-Means* e *Multilayer Perceptron (MLP)*. O *KNN* e o *K-Means* foram utilizados para explorar a organização dos dados por vizinhança e agrupamento, respectivamente. Para o *KNN*, aplicou-se um *grid search* sistemático para otimização de hiperparâmetros, variando o número de vizinhos (k) entre 5, 10, 15, 20, 25 e 30. Já o *K-Means* foi configurado com o número de *clusters* equivalente às classes presentes no conjunto de dados ABP.

A rede *MLP* foi estruturada com duas camadas densas (512 e 256 neurônios). Adotou-se a função de ativação *ReLU* para garantir um treinamento mais rápido e estável [Zhang and Rao 2020], enquanto a camada de saída utilizou a *softmax* para a classificação final. O modelo otimizado com o algoritmo *Adam* e a função de perda *Categorical Cross-Entropy*, escolhas consolidadas na literatura para problemas multiclasse [Goodfellow 2016].

A confiabilidade do experimento foi assegurada pela validação cruzada *k-folds*, utilizando 5 *folds*. Esse procedimento garante que a proporção das classes seja mantida em todas as etapas, evitando resultados enviesados e permitindo medir com precisão a capacidade do modelo em classificar novos dados [Brink et al. 2016].

3.4. Avaliação de Desempenho

O desempenho dos modelos de classificação será avaliado pelas métricas de Acurácia e *Macro F1-score*. A Acurácia (Equação 3) fornece uma medida geral da taxa de acerto para o conjunto de classes C . Complementarmente, o *Macro F1-score* (Equação 4) permite uma avaliação equilibrada entre as classes, mitigando distorções causadas por eventuais desbalanceamento no conjunto de dados.

As métricas são formalmente definidas conforme segue:

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3) \qquad \text{Macro } F1 = \frac{1}{|C|} \sum_{c \in C} F1_c \quad (4)$$

Em que $F1_c$ representa a média harmônica entre a Precisão (P_c) e a Revocação (R_c) para cada classe:

$$P_c = \frac{TP_c}{TP_c + FP_c}, \quad R_c = \frac{TP_c}{TP_c + FN_c}, \quad F1_c = 2 \times \frac{P_c \times R_c}{P_c + R_c} \quad (5)$$

Neste contexto, TP , TN , FP e FN representam, respectivamente, os verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos.

A validação da hipótese de pesquisa — de que o discurso textual reflete o viés ideológico — dar-se-á mediante a obtenção de altos índices em ambas as métricas. Espera-se que valores elevados de *Macro F1-score* confirmem a capacidade discriminatória do modelo entre as diferentes vertentes ideológicas, assegurando que o desempenho não seja reflexo de uma classe majoritária no conjunto de dados.

3.5. Ambiente de Execução

A linguagem Python, com as bibliotecas NumPy, Pandas, Scikit-Learn e PyTorch, foi a ferramenta primária para implementação e avaliação dos modelos. Os experimentos ocorreram em um servidor equipado com processador Intel Xeon W-2235, 128 GB de RAM e GPU NVIDIA RTX 8000 (48 GB VRAM), visando a aceleração em hardware.

4. Resultados e Discussão

Os experimentos foram conduzidos seguindo a metodologia proposta, totalizando 24 configurações que integraram os modelos DistilBERT e DistilRoBERTa, funções de perda (*Contrastive Loss* e *Triplet Loss*) e classificadores (*KNN*, *K-Means* e *MLP*). Esta abordagem permitiu uma análise sistemática da sensibilidade dos modelos nos cenários *random split* e *media-bias split*.

4.1. Desempenho no Conjunto de Dados ABP

Os resultados consolidados na Tabela 4 demonstram que a utilização de aprendizagem métrica, especificamente a *Contrastive Loss*, favorece significativamente a separação vetorial no espaço de *embeddings* para a tarefa de viés ideológico.

Tabela 4. Desempenho comparativo (em %) no conjunto de dados ABP utilizando as divisões *media-bias split* e *random split*.

Cenário	Modelo	Configuração (Repre. / Fun. de Perda / Classificador)	Macro F1-Score	Acurácia
ABP (media-bias split)	Proposto (Exp. 4)	DistilRoBERTa / Contrastive / KNN	45,44	48,89
	Baly et al., (2020)	BERT / - / BERT	33,53	36,75
ABP (random split)	Proposto (Exp. 16)	DistilRoBERTa / Contrastive / KNN	83,90	83,89
	Baly et al., (2020)	BERT / - / BERT	80,19	79,83

No cenário *media-bias split*, o modelo proposto superou o baseline de Baly et al.[2020] em aproximadamente 12 pontos percentuais. Este resultado é particularmente relevante, pois o *media-bias split* exige que o modelo identifique nuances ideológicas em veículos de imprensa nunca vistos durante o treinamento. Contudo, a queda de desempenho em relação ao *random split* (83,90%) evidencia que a presença de fontes comuns facilita a predição por meio da memorização de padrões específicos do veículo, e não necessariamente da ideologia pura.

4.2. Avaliação de Generalização

Para testar a robustez em cenários de estresse rigoroso, o modelo foi submetido ao dataset *FlipBias* sem ajuste fino adicional. Embora os índices de *Macro F1-score* (47,00% no

Exp. 16), apresentados na Tabela 5, tenham sido inferiores aos modelos *in-domain*³ de Lin et al. [2024, 2025], os dados sugerem que a arquitetura captura estruturas ideológicas fundamentais que transcendem o domínio original.

Tabela 5. Desempenho comparativo (em %) no conjunto de dados FlipBias.

Autor	Modelo / Configuração	Estratégia	Macro F1-Score
Lin et al. (2024)	BERT	Fine-tuning	86,20
Lin et al. (2025)	GPT 3.5	Fine-tuning	77,82
Proposto (Exp. 16)	DistilRoBERTa / Contrastive / KNN	Fine-tuning	47,00
Proposto (Exp. 4)	DistilRoBERTa / Contrastive / KNN	Fine-tuning	28,21

4.3. Comparação com Modelos de Linguagem de Larga Escala (LLMs)

Uma análise comparativa foi realizada com uma amostra de 90 instâncias do teste *media-bias split* para situar o modelo frente a LLMs de última geração. Observa-se que o modelo proposto, mesmo sendo uma arquitetura destilada e leve (*DistilRoBERTa*), superou o desempenho *zero-shot* do DeepSeek (16,0%) e aproximou-se da performance do Gemini 3 (34,0%).

Tabela 6. Comparação com LLMs (Amostra Aleatória)

Modelo	Estratégia	Macro F1-Score
Gemini 3	Zero-Shot	34,0%
Proposto (Exp. 4)	Fine-tuning	27,0%
DeepSeek	Zero-Shot	16,0%

É fundamental ressaltar que, embora o desempenho absoluto seja inferior ao Gemini 3, o modelo proposto é ordens de magnitude menor em termos de parâmetros e requisitos computacionais. Enquanto LLMs exigem infraestruturas massivas de nuvem, a arquitetura baseada em aprendizagem métrica aqui apresentada é perfeitamente viável para ambientes com recursos computacionais restritos. Isso valida a eficácia da *Contrastive Loss* em gerar representações discriminatórias competitivas, priorizando a eficiência e a viabilidade prática em cenários limitados.

5. Considerações Finais

Este estudo objetivou avaliar a eficácia da aprendizagem métrica na identificação de viés ideológico, integrando modelos pré-treinados e funções de perda contrastivas. Os resultados demonstram que a *Contrastive Loss* otimiza a separação vetorial, superando o baseline de [Baly et al. 2020] em 12 pontos percentuais no cenário *media-bias split*. A arquitetura mostrou-se competitiva frente a LLMs em modo *zero-shot*, validando a viabilidade de modelos leves e eficientes para a análise de nuances discursivas.

³In-domain refere-se a uma situação em que os dados utilizados para treinar um modelo de aprendizado de máquina e os dados usados para testá-lo vêm da mesma distribuição, contexto ou fonte.

Para trabalhos futuros, pretende-se explorar técnicas de *Domain Adaptation* para otimizar a transferência de conhecimento e elevar o desempenho em dados de fontes não expostas durante o treinamento. Além disso, planeja-se investigar o comportamento de LLMs sob estratégias de *Few-Shot* e *Fine-Tuning*, permitindo uma comparação mais consistente com a arquitetura proposta. Tais avanços visam mitigar a memorização de padrões específicos de veículos e consolidar a robustez do modelo em cenários de generalização rigorosa.

Referências

- Baly, R., Da San Martino, G., Glass, J., and Nakov, P. (2020). We can detect your bias: Predicting the political ideology of news articles. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4982–4991.
- Brink, H., Richards, J., and Fetherolf, M. (2016). Real-world machine learning. manning publications.
- Chen, W.-F., Wachsmuth, H., Al Khatib, K., and Stein, B. (2018). Learning to flip the bias of news headlines. In *Proceedings of the 11th International conference on natural language generation*, pages 79–88.
- Chiang, C.-F. and Knight, B. (2011). Media bias and influence: Evidence from newspaper endorsements. *The Review of economic studies*, 78(3):795–820.
- Dallmann, A., Lemmerich, F., Zoller, D., and Hotho, A. (2015). Media bias in german online newspapers. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pages 133–137.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Efron, M. (2004). The liberal media and right-wing conspiracies: using cocitation information to estimate political orientation in web documents. In *Proceedings of the thirteenth ACM international conference on Information and Knowledge Management*, pages 390–398.
- Gangula, R. R. R., Duggenpudi, S. R., and Mamidi, R. (2019). Detecting political bias in news articles using headline attention. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 77–84. Association for Computational Linguistics.
- Gentzkow, M. and Shapiro, J. M. (2006). Media bias and reputation. *Journal of political Economy*, 114(2):280–316.
- Goodfellow, I. (2016). Deep learning.
- Kertész, G. (2021). Different triplet sampling techniques for lossless triplet loss on metric similarity learning. In *2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pages 000449–000454.
- Lin, L., Wang, L., Guo, J., and Wong, K.-F. (2025). Investigating bias in LLM-based bias detection: Disparities between LLMs and human perception. In Rambow, O., Waner, L., Apidianaki, M., Al-Khalifa, H., Eugenio, B. D., and Schockaert, S., editors,

Proceedings of the 31st International Conference on Computational Linguistics, pages 10634–10649, Abu Dhabi, UAE. Association for Computational Linguistics.

Lin, L., Wang, L., Zhao, X., Li, J., and Wong, K. (2024a). Indivec: An exploration of leveraging large language models for media bias detection with fine-grained bias indicators (no. arxiv: 2402.00345). arxiv.

Lin, L., Wang, L., Zhao, X., Li, J., and Wong, K.-F. (2024b). Indivec: An exploration of leveraging large language models for media bias detection with fine-grained bias indicators.

Lin, Y.-R., Bagrow, J., and Lazer, D. (2011). More voices than ever? quantifying media bias in networks. In *Proceedings of the international AAAI conference on web and social media*, volume 5, pages 193–200.

Liu, Y. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Liu, Y., Zhang, X. F., Wegsman, D., Beauchamp, N., and Wang, L. (2022). POLITICS: Pretraining with same-story article comparison for ideology prediction and stance detection. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1354–1374, Seattle, United States. Association for Computational Linguistics.

Patricia Aires, V., G. Nakamura, F., and F. Nakamura, E. (2019). A link-based approach to detect media bias in news websites. In *Companion proceedings of the 2019 world wide web conference*, pages 742–745.

Sanh, V. (2019). Distilbert, a distilled version of bert: Smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Spinde, T., Plank, M., Krieger, J.-D., Ruas, T., Gipp, B., and Aizawa, A. (2021). Neural media bias detection using distant supervision with BABE - bias annotations by experts. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1166–1177. Association for Computational Linguistics.

Zhang, Y. and Rao, Z. (2020). Deep neural networks with pre-train model bert for aspect-level sentiments classification. In *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, pages 923–927. IEEE.