

References

- Acheampong Henry; Chen Wenyu, F. A. N.-M. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artificial Intelligence Review*, 54(8), 5789–5829. <https://doi.org/10.1007/s10462-021-09958-2>
- Ahmad Azuraliza; Yaakub Mohd Ridzwan, S. R. A. B. (2019). A review of feature selection techniques in sentiment analysis. *Intelligent Data Analysis*, 23(1), 159–189. <https://doi.org/10.3233/ida-173763>
- Akbik, A., Blythe, D., & Vollgraf, R. (2018). Contextual string embeddings for sequence labeling. *Proceedings of the 27th International Conference on Computational Linguistics*.
- Alec, R., Jeffrey, W., Rewon, C., David, L., Dario, A., & Ilya, S. (2019). Language Models are Unsupervised Multitask Learners | Enhanced Reader. *OpenAI Blog*, 1(8).
- Bella, G., Helm, P., Koch, G., & Giunchiglia, F. (2024). Tackling Language Modelling Bias in Support of Linguistic Diversity. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 562–572.

-
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *FAccT 2021 - Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. <https://doi.org/10.1145/3442188.3445922>
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in nlp. *ArXiv Preprint ArXiv:2005.14050*.
- Bolukbasi, T., Chang, K. W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems*.
- Borkan, D., Dixon, L., Sorensen, J., Thain, N., & Vasserman, L. (2019). Nuanced metrics for measuring unintended bias with real data for text classification. *The Web Conference 2019 - Companion of the World Wide Web Conference, WWW 2019*. <https://doi.org/10.1145/3308560.3317593>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 2020-December*.
- Buolamwini, J. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification *. *Proceedings of Machine Learning Research, 81*.
- Chang Hsin-Ying; Chen Long-Sheng; Chang Chia-Wei, J.-R. L. (2020). Novel feature selection approaches for improving the performance of sentiment classification. *Journal of Ambient Intelligence and Humanized Computing, NA(NA), 1–14*. <https://doi.org/10.1007/s12652-020-02468-z>

-
- Chaturvedi, S. (2015). A Sociolinguistic study of linguistic variation and code matrix In Kanpur. *Procedia-Social and Behavioral Sciences*, 192, 107–115.
- Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W., Sutton, C., Gehrmann, S., Schuh, P., Shi, K., Tsvyashchenko, S., Maynez, J., Rao, A., Barnes, P., Tay, Y., Shazeer, N., Prabhakaran, V., ... Fiedel, N. (2022). *PaLM: Scaling Language Modeling with Pathways*.
- Cortis Brian, K. D. (2021). Over a decade of social opinion mining: a systematic review. *Artificial Intelligence Review*, 54(7), 1–93. <https://doi.org/10.1007/s10462-021-10030-2>
- Crawford, K. (2022). Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. *Perspectives on Science and Christian Faith*, 74(1). <https://doi.org/10.56315/pscf3-22crawford>
- Das, D., & Bandyopadhyay, S. (2010). Labeling Emotion in Bengali Blog Corpus—A Fine Grained Tagging at Sentence Level. *Workshop on 8th Asian Language Resources COLING2010, August*.
- Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated Experiments on Ad Privacy Settings. *Proceedings on Privacy Enhancing Technologies*, 2015(1). <https://doi.org/10.1515/popets-2015-0007>
- Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing*, 25(4). <https://doi.org/10.1093/llc/fqq018>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, 1.

-
- Dixon, L., Li, J., Sorensen, J., Thain, N., & Vasserman, L. (2018). Measuring and Mitigating Unintended Bias in Text Classification. *AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*. <https://doi.org/10.1145/3278721.3278729>
- Dong, Q., Li, L., Dai, D., Zheng, C., Wu, Z., Chang, B., Sun, X., Xu, J., Li, L., & Sui, Z. (2022). A Survey for In-context Learning. *Arxiv*.
- Dwivedi, S., & Ghosh, S. (2022). Subjectivity Identification Through Lexical Rules. *SN Computer Science*, 3(1). <https://doi.org/10.1007/s42979-021-00892-2>
- Dwivedi, S., Ghosh, S., & Dwivedi, S. (2023). *Breaking the bias: Gender fairness in LLMs using prompt engineering and in-context learning*.
- Dwivedi, S., Ghosh, S., & Dwivedi, S. (2024). Navigating Linguistic Diversity: In-Context Learning and Prompt Engineering for Subjectivity Analysis in Low-Resource Languages. *SN Computer Science*, 5(4), 418.
- Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. *ITCS 2012 - Innovations in Theoretical Computer Science Conference*. <https://doi.org/10.1145/2090236.2090255>
- Eidinger, E., Enbar, R., & Hassner, T. (2014). Age and gender estimation of unfiltered faces. *IEEE Transactions on Information Forensics and Security*, 9(12). <https://doi.org/10.1109/TIFS.2014.2359646>
- Emeneau, M. B. (1956). India as a linguistic area. *Language*, 32(1), 3–16.
- Garg, S., Taly, A., Perot, V., Chi, E. H., Limtiaco, N., & Beutel, A. (2019). Counterfactual fairness in text classification through robustness. *AIES 2019 - Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. <https://doi.org/10.1145/3306618.3317950>

-
- Gaut, A., Sun, T., Tang, S., Huang, Y., Qian, J., ElSherief, M., Zhao, J., Mirza, D., Belding, E., Chang, K.-W., & Wang, W. Y. (2020). *Towards Understanding Gender Bias in Relation Extraction*. <https://doi.org/10.18653/v1/2020.acl-main.265>
- Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H., & Crawford, K. (2018). *Datasheets for Datasets*. <http://arxiv.org/abs/1803.09010>
- Gupta, R., Goyal, P., & Diwkar, S. (2010). Transliteration among indian languages using WX notation. *Semantic Approaches in Natural Language Processing - Proceedings of the Conference on Natural Language Processing 2010, KONVENS*.
- Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*.
- Helm, P., Bella, G., Koch, G., & Giunchiglia, F. (2024). Diversity and language technology: how language modeling bias causes epistemic injustice. *Ethics and Information Technology*, 26(1). <https://doi.org/10.1007/s10676-023-09742-6>
- Huang, P.-S., Zhang, H., Jiang, R., Stanforth, R., Welbl, J., Rae, J., Maini, V., Yogatama, D., & Kohli, P. (2020). *Reducing Sentiment Bias in Language Models via Counterfactual Evaluation*. <https://doi.org/10.18653/v1/2020.findings-emnlp.7>
- Hutto, C. J. (2020). *vaderSentiment: VADER Sentiment Analysis*. GitHub.
- Jha, V., Manjunath, N., Shenoy, P. D., & Venugopal, K. R. (2016). HSAS: Hindi Subjectivity Analysis System. *12th IEEE International Conference Electronics, Energy, Environment, Communication, Computer, Control: (E3-C3), INDICON 2015*. <https://doi.org/10.1109/INDICON.2015.7443824>
- Kachru, Y. (2006). *Hindi*.
- Kaity Vimala, M. B. (2020). Sentiment lexicons and non-English languages: a survey. *Knowledge and Information Systems*, 62(12), 4445–4480. <https://doi.org/10.1007/s10115-020-01497-6>

-
- Kang Seong Joon; Han Dongil, H. Y. (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, 39(5), 6000–6010. <https://doi.org/10.1016/j.eswa.2011.11.107>
- Kiritchenko Xiaodan; Mohammad Saif M., S. Z. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50(1), 723–762. <https://doi.org/10.1613/jair.4272>
- Klare, B. F., Klein, B., Taborsky, E., Blanton, A., Cheney, J., Allen, K., Grother, P., Mah, A., Burge, M., & Jain, A. K. (2015). Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 07-12-June-2015*. <https://doi.org/10.1109/CVPR.2015.7298803>
- Kojima, T., Gu, S. S., Reid, M., Matsuo, Y., & Iwasawa, Y. (2022). Large Language Models are Zero-Shot Reasoners. *Advances in Neural Information Processing Systems*, 35.
- Kusner, M. J., Loftus, J. R., Russell, C., & Silva, R. (2017). Counterfactual fairness. *ArXiv Preprint ArXiv:1703.06856*.
- Lakkaraju, H., Kleinberg, J., Leskovec, J., Ludwig, J., & Mullainathan, S. (2017). The selective labels problem: Evaluating algorithmic predictions in the presence of unobservables. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Part F129685*. <https://doi.org/10.1145/3097983.3098066>
- Larrazabal, A. J., Nieto, N., Peterson, V., Milone, D. H., & Ferrante, E. (2020). Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proceedings of the National Academy of Sciences of the United States of America*, 117(23). <https://doi.org/10.1073/pnas.1919012117>

-
- Liang, P. P., Wu, C., Morency, L. P., & Salakhutdinov, R. (2021). Towards Understanding and Mitigating Social Biases in Language Models. *Proceedings of Machine Learning Research*, 139.
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*, 55(9). <https://doi.org/10.1145/3560815>
- Liu, X., He, P., Chen, W., & Gao, J. (2019). Improving multi-task deep neural networks via knowledge distillation for natural language understanding. *ArXiv Preprint ArXiv:1904.09482*.
- Lorla, S. (2020). TextBlob Documentation. *TextBlob*.
- McCosker, A., & Wilken, R. (2020). Automating vision: The social impact of the new camera consciousness. In *Automating Vision: The Social Impact of the New Camera Consciousness*. <https://doi.org/10.4324/9780429341175>
- Mittal, N., & Agarwal, B. (2013). Sentiment Analysis of Hindi Review based on Negation and Discourse Relation. *Sixth International Joint Conference on Natural Language Processing*.
- Nangia, N., Vania, C., Bhalerao, R., & Bowman, S. R. (2020). *CrowS-Pairs: A Challenge Dataset for Measuring Social Biases in Masked Language Models*. <https://doi.org/10.18653/v1/2020.emnlp-main.154>
- Narayan, D., Chakrabarti, D., Pande, P., & Bhattacharyya, P. (2002). An experience in building the indo wordnet-a wordnet for hindi. *First International Conference on Global WordNet, Mysore, India, 24*.
- Noble, S. U. (2018). Algorithms of oppression. In *Algorithms of oppression*. New York university press.

-
- Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of the Annual Meeting of the Association for Computational Linguistics*.
- Patnaik, B. N. (2014). *Oriya as a typologically disturbed language and some related matters*.
- Prasāda, V., & Jhā, S. (1958). Linguistic survey of the sadar subdivision of manbhum and dhalbhum (Singhbhum). (*No Title*).
- Registrar General India. (2011). Census of India 2011: Provisional Population Totals-India data sheet. *Office of the Registrar General Census Commissioner, India. Indian Census Bureau*.
- Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., Mirza, D., Belding, E., Chang, K. W., & Wang, W. Y. (2020). Mitigating gender bias in natural language processing: Literature review. *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*.
<https://doi.org/10.18653/v1/p19-1159>
- Taboada Julian; Tofiloski Milan; Voll Kimberly; Stede Manfred, M. B. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2), 267–307.
https://doi.org/10.1162/coli_a_00049
- Tatman, R. (2017). *Gender and Dialect Bias in YouTube's Automatic Captions*.
<https://doi.org/10.18653/v1/w17-1606>
- Taylor, A., Marcus, M., & Santorini, B. (2003). *The Penn Treebank: An Overview*.
https://doi.org/10.1007/978-94-010-0201-1_1
- Thomas McCoy, R., Pavlick, E., & Linzen, T. (2020). Right for the wrong reasons: Diagnosing syntactic heuristics in natural language inference. *ACL 2019 - 57th*

-
- Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference.* <https://doi.org/10.18653/v1/p19-1334>
- Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). *LLaMA: Open and Efficient Foundation Language Models.*
- Tsarfaty, R., Seddah, D., Kübler, S., & Nivre, J. (2013). Parsing morphologically rich languages: Introduction to the special issue. *Computational Linguistics*, 39(1). https://doi.org/10.1162/COLI_a_00133
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 2017-December.*
- Venugopal, J. P., Subramanian, A. A. V., Sundaram, G., Rivera, M., & Wheeler, P. (2024). A Comprehensive Approach to Bias Mitigation for Sentiment Analysis of Social Media Data. *Applied Sciences*, 14(23), 11471.
- Wadawadagi Veerappa B., R. S.; P. (2020). Sentiment analysis with deep neural networks: comparative study and performance assessment. *Artificial Intelligence Review*, 53(8), 6155–6195. <https://doi.org/10.1007/s10462-020-09845-2>
- Wang, X., & Fu, G. H. (2010). Chinese subjectivity detection using a sentiment density-based naive Bayesian classifier. *2010 International Conference on Machine Learning and Cybernetics, ICMLC 2010*, 6. <https://doi.org/10.1109/ICMLC.2010.5580700>
- Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review.* <https://doi.org/10.1007/s10462-022-10144-1>

-
- Wei, X., Kumar, N., & Zhang, H. (2025). Addressing bias in generative AI: Challenges and research opportunities in information management. *ArXiv Preprint ArXiv:2502.10407*.
- Whittlestone, J., Alexandrova, A., Nyrupe, R., & Cave, S. (2019). The role and limits of principles in AI ethics: Towards a focus on tensions. *AIES 2019 - Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*.
<https://doi.org/10.1145/3306618.3314289>
- Xia Yitai; Pan Xiaoting; Zhang Zuopeng Justin; An Wuyue, H. Y. (2019). Sentiment analysis for online reviews using conditional random fields and support vector machines. *Electronic Commerce Research*, 20(2), 343–360.
<https://doi.org/10.1007/s10660-019-09354-7>
- Xu, P., Zhu, X., & Clifton, D. A. (2023). Multimodal Learning With Transformers: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
<https://doi.org/10.1109/TPAMI.2023.3275156>
- Yadav Dinesh Kumar, A. V. (2019). Sentiment analysis using deep learning architectures: a review. *Artificial Intelligence Review*, 53(6), 4335–4385.
<https://doi.org/10.1007/s10462-019-09794-5>
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2019). XLNet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, 32.
- Zhao, J., Zhou, Y., Li, Z., Wang, W., & Chang, K. W. (2018). Learning gender-neutral word embeddings. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*. <https://doi.org/10.18653/v1/d18-1521>

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., ... Wen, J.-R. (2023). *A Survey of Large Language Models*.



Appendix 1: List of Published Research Papers

- Dwivedi, Satyam, and Sanjukta Ghosh. "Subjectivity identification through lexical rules." *SN Computer Science* 3.1 (2022): 32.
- Dwivedi, Satyam, Sanjukta Ghosh, and Shivam Dwivedi. "Breaking the bias: Gender fairness in LLMs using prompt engineering and in-context learning." *Rupkatha Journal on Interdisciplinary Studies in Humanities* 15.4 (2023).
- Dwivedi, Satyam, Sanjukta Ghosh, and Shivam Dwivedi. "Navigating Linguistic Diversity: In-Context Learning and Prompt Engineering for Subjectivity Analysis in Low-Resource Languages." *SN Computer Science* 5.4 (2024): 418.



Appendix 2: Data and Code

All data, code, results, plots, and supplementary materials are available in the relevant GitHub repositories at: <https://github.com/dwisatyam>.