References:

[1], [2], [3], [4], [5], [6],

[1] K. Shah, H. Patel, D. Sanghvi, and M. Shah, “A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification,” *Augment. Hum. Res.*, vol. 5, no. 1, p. 12, Mar. 2020, doi: 10.1007/s41133-020-00032-0.

[2] E. Wulczyn, N. Thain, and L. Dixon, “Ex Machina: Personal Attacks Seen at Scale,” in *Proceedings of the 26th International Conference on World Wide Web*, in WWW ’17. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, Apr. 2017, pp. 1391–1399. doi: 10.1145/3038912.3052591.

[3] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza, “Power to the People: The Role of Humans in Interactive Machine Learning,” *AI Mag.*, vol. 35, no. 4, Art. no. 4, Dec. 2014, doi: 10.1609/aimag.v35i4.2513.

[4] L. Dixon, J. Li, J. Sorensen, N. Thain, and L. Vasserman, “Measuring and Mitigating Unintended Bias in Text Classification,” in *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, in AIES ’18. New York, NY, USA: Association for Computing Machinery, Dec. 2018, pp. 67–73. doi: 10.1145/3278721.3278729.

[5] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, “On Calibration of Modern Neural Networks,” in *Proceedings of the 34th International Conference on Machine Learning*, PMLR, Jul. 2017, pp. 1321–1330. Accessed: Jun. 29, 2025. [Online]. Available: https://proceedings.mlr.press/v70/guo17a.html

[6] R. Snow, B. O’Connor, D. Jurafsky, and A. Ng, “Cheap and Fast – But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks,” in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, M. Lapata and H. T. Ng, Eds., Honolulu, Hawaii: Association for Computational Linguistics, Oct. 2008, pp. 254–263. Accessed: Jun. 29, 2025. [Online]. Available: https://aclanthology.org/D08-1027/