DAFTAR PUSTAKA

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing*. Longman.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives*. David McKay Company.

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*. https://doi.org/10.48550/arXiv.2005.14165

Creswell, J. W., & Creswell, J. D. (2018). *Research design*. SAGE Publications.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2024). *Retrieval-Augmented Generation for Large Language Models: A Survey*. http://arxiv.org/abs/2312.10997

Gopi, S., Sreekanth, D., & Dehbozorgi, N. (2024). Enhancing Engineering Education Through LLM-Driven Adaptive Quiz Generation: A RAG-Based Approach. *2024 IEEE Frontiers in Education Conference (FIE)*, 1–8. https://doi.org/10.1109/FIE61694.2024.10893146

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75–106. https://doi.org/10.2307/25148625

Hwang, G.-J., Xie, H., Wah, B. W., & Gašević, D. (2020). Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence*, *1*, 100001. https://doi.org/10.1016/j.caeai.2020.100001

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of Hallucination in Natural Language Generation. *ACM Computing Surveys*, *55*(12), 1–38. https://doi.org/10.1145/3571730

Johnson, J., Douze, M., & Jegou, H. (2021). Billion-Scale Similarity Search with GPUs. *IEEE Transactions on Big Data*, *7*(3), 535–547. https://doi.org/10.1109/TBDATA.2019.2921572

Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (3rd ed., draft). Stanford University.

Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. (2020). Dense Passage Retrieval for Open-Domain Question Answering. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 6769–6781. https://doi.org/10.18653/v1/2020.emnlp-main.550

Krippendorff, Klaus. (2019). *Content analysis : an introduction to its methodology* (4th ed.). SAGE.

Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2020). A Systematic Review of Automatic Question Generation for Educational Purposes. *International Journal of Artificial Intelligence in Education*, *30*(1), 121–204. https://doi.org/10.1007/s40593-019-00186-y

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2021). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. https://doi.org/10.48550/arXiv.2005.11401

Li, Z., Wang, Z., Wang, W., Hung, K., Xie, H., & Wang, F. L. (2025). Retrieval-augmented generation for educational application: A systematic survey. *Computers and Education: Artificial Intelligence*, *8*, 100417. https://doi.org/10.1016/j.caeai.2025.100417

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, *22*(140), 5–55.

Maity, S., Deroy, A., & Sarkar, S. (2025). *Leveraging In-Context Learning and Retrieval-Augmented Generation for Automatic Question Generation in Educational Domains*. https://doi.org/10.48550/arXiv.2501.17397

Mallen, A., Asai, A., Zhong, V., Das, R., Khashabi, D., & Hajishirzi, H. (2023). When Not to Trust Language Models: Investigating Effectiveness of Parametric and Non-Parametric Memories. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 9802–9822. https://doi.org/10.18653/v1/2023.acl-long.546

Manning, C. D. ., Raghavan, Prabhakar., & Schütze, Hinrich. (2009). *Introduction to information retrieval*. Cambridge University Press.

Merrill, M. D. (2002). First principles of instruction. *Educational Technology Research and Development*, *50*(3), 43–59. https://doi.org/10.1007/BF02505024

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2025). *Large Language Models: A Survey*. https://doi.org/10.48550/arXiv.2402.06196

Nielsen, J. (1993). *Usability Engineering* (1st ed.). Academic Press.

Nogueira, R., & Cho, K. (2020). *Passage Re-ranking with BERT*. https://doi.org/10.48550/arXiv.1901.04085

Novikova, J., Dušek, O., Cercas Curry, A., & Rieser, V. (2017). Why We Need New Evaluation Metrics for NLG. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2241–2252. https://doi.org/10.18653/v1/D17-1238

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *ArXiv Preprint*. https://doi.org/10.48550/arXiv.1706.03762

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3980–3990. https://doi.org/10.18653/v1/D19-1410

Reynolds, L., & McDonell, K. (2021). Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. *Extended Abstracts of the 2021 CHI Conference*

on Human Factors in Computing Systems, 1–7. https://doi.org/10.1145/3411763.3451760

Santhanam, K., Khattab, O., Saad-Falcon, J., Potts, C., & Zaharia, M. (2022). ColBERTv2: Effective and Efficient Retrieval via Lightweight Late Interaction. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 3715–3734. https://doi.org/10.18653/v1/2022.naacl-main.272

Scriven, M. (1967). *The Methodology of Evaluation* (R. W. Tyler, R. M. Gagne, & M. Scriven, Eds.; pp. 39–83). Rand McNally.

Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation Theory, Models, and Applications, 2nd Edition*. Jossey-Bass.

Tessmer, Martin. (2005). *Planning and conducting formative evaluations : improving the quality of education and training*. Routledge.

Vagias, W. M. (2006). Likert-type scale response anchors. *Clemson International Institute for Tourism & Research Development, Department of Parks, Recreation and Tourism Management*.

van der Lee, C., Gatt, A., van Miltenburg, E., Wubben, S., & Krahmer, E. (2019). Best practices for the human evaluation of automatically generated text. *Proceedings of the 12th International Conference on Natural Language Generation*, 355–368. https://doi.org/10.18653/v1/W19-8643

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*, *30*, 5998–6007.

Wang, X., Wang, Z., Gao, X., Zhang, F., Wu, Y., Xu, Z., Shi, T., Wang, Z., Li, S., Qian, Q., Yin, R., Lv, C., Zheng, X., & Huang, X. (2024). *Searching for Best Practices in Retrieval-Augmented Generation*. https://doi.org/10.48550/arXiv.2407.01219

Wei, J., Wei, J., Tay, Y., Tran, D., Webson, A., Lu, Y., Chen, X., Liu, H., Huang, D., Zhou, D., & Ma, T. (2023). *Larger language models do in-context learning differently*. https://doi.org/10.48550/arXiv.2303.03846