Important Papers

February 27, 2025

1 Language Model

1. A. Aghajanyan, L. Zettlemoyer, and S. Gupta. Intrinsic dimensionality explains the effectiveness of language model fine-tuning, 2020.

2. Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin. A Neural Probabilistic Language Model. Journal of Machine Learning Research, 3:1137–1155, Mar. 2003.

3. D. Chen, A. Fisch, J. Weston, and A. Bordes. Reading wikipedia to answer open-domain questions. CL, 2017.

4. J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling, 2014.

5. Claude https://www.anthropic.com/news/claude-3-family

6. G. Team, R. Anil, and S. et al. Gemini: A family of highly capable multimodal models, 2024.

7. J. Gehring, M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin. Convolutional sequence to sequence learning. In Proceedings of the International Conference on Machine Learning (ICML), pages 1243–1252, 2017.

8. OpenAI, J. Achiam, S. Adler, et al. Gpt-4 technical report, 2024.

9. A. Graves. Generating sequences with recurrent neural networks, 2014.

10. E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685, 2021.

11. D. Jurafsky and J. H. Martin. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models. 3rd edition, 2025. Online manuscript released January 12, 2025.

12. W. Lai, H. Xie, G. Xu, and Q. Li. Multi-task learning with llms for implicit sentiment analysis: Data- level and task-level automatic weight learning, 2024.

13. R. Lebret and R. Lebret. Word emdeddings through hellinger PCA. CoRR, abs/1312.5542, 2013.

14. P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing, 2021.

15. A. Mnih and G. Hinton. A scalable hierarchical distributed language model. In Proceedings of the 21st International Conference on Neural Information Processing Systems, NIPS'08, pages 1081–1088, USA, 2008. Curran Associates Inc.

16. F. Morin and Y. Bengio. Hierarchical probabilistic neural network language model. In R. G. Cowell and Z. Ghahramani, editors, Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics, volume R5 of Proceedings of Machine Learning Research, pages 246–252. PMLR, 06–08 Jan 2005. Reissued by PMLR on 30 March 2021.

17. F. Morin and Y. Bengio. Hierarchical probabilistic neural network language model. In Aistats, volume 5, pages 246–252. Citeseer, 2005.

18. P. Nguyen, S. Sengupta, G. Malik, A. Gupta, and B. Min. Install: Context-aware instructional task assistance with multi-modal large language models, 2025.

19. M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. Deep contextualized word representations. CoRR, abs/1802.05365, 2018.[17] L. Shang, Z. Lu, and H. Li. Neural responding machine for short-text conversation. CL, 2015.

20. I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the importance of initialization and momentum in deep learning. In S. Dasgupta and D. McAllester, editors, Proceedings of the 30th International Conference on Machine Learning, volume 28 of Proceedings of Machine Learning Research, pages 1139– 1147, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

21. I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the importance of initialization and momentum in deep learning, 2013.

22. I. Sutskever, O. Vinyals, and Q. V. Le. Sequence to sequence learning with neural networks. In Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'14, pages 3104–3112, Cambridge, MA, USA, 2014. MIT Press.

23. H. Touvron, T. Lavril, G. Izacard, X. Martinet, M.-A. Lachaux, T. Lacroix, B. Rozi‘ere, N. Goyal, E. Hambro, F. Azhar, A. Rodriguez, A. Joulin, E. Grave, and G. Lample. Llama: Open and efficient foundation language models, 2023.

24. Y. Wang, X. Li, Z. Yan, Y. He, J. Yu, X. Zeng, C. Wang, C. Ma, H. Huang, J. Gao, M. Dou, K. Chen, W. Wang, Y. Qiao, Y. Wang, and L. Wang. Internvideo2.5: Empowering video mllms with long and rich context modeling, 2025.

25. S. Zhang, L. Dong, X. Li, S. Zhang, X. Sun, S. Wang, J. Li, R. Hu, T. Zhang, F. Wu, and G. Wang. Instruction tuning for large language models: A survey, 2024.

2 Sentiment Analysis

1. G. Brauwers and F. Frasincar. A survey on aspect-based sentiment classification. ACM Comput. Surv., 55(4), Nov. 2022.

2. N. C. Dang, M. N. Moreno-Garc´ıa, and F. De la Prieta. Sentiment analysis based on deep learning: A comparative study. Electronics, 9(3):483, Mar. 2020.

3. Q. Gu, Z. Wang, H. Zhang, S. Sui, and R. Wang. Aspect-level sentiment analysis based on syntax-aware and graph convolutional networks. Applied Sciences, 14(2), 2024.

4. W. Lai, H. Xie, G. Xu, and Q. Li. Multi-task learning with llms for implicit sentiment analysis: Data- level and task-level automatic weight learning, 2024.

5. Y. Li, S. Jiang, et al. Multimodal sentiment analysis with image-text correlation modal. In 2023 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics), pages 281–286. IEEE, 2023.

6. M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. Deep contextualized word representations. CoRR, abs/1802.05365, 2018.

7. F. W. Smith and L. Muckli. Nonstimulated early visual areas carry information about surrounding context. Proceedings of the National Academy of Sciences, 107(46):20099–20103, 2010.

8. R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In D. Yarowsky, T. Baldwin, A. Korhonen, K. Livescu,

and S. Bethard, editors, Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 1631–1642, Seattle, Washington, USA, Oct. 2013. Association for Computational Linguistics.

9. I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the importance of initialization and momentum in deep learning. In S. Dasgupta and D. McAllester, editors, Proceedings of the 30th International Conference on Machine Learning, volume 28 of Proceedings of Machine Learning Research, pages 1139– 1147, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.
10. I. Sutskever, J. Martens, G. Dahl, and G. Hinton. On the importance of initialization and momentum in deep learning, 2013.
11. J. Weizenbaum. Eliza-computer program for the study of natural language communication between man and machine. Commun. ACM, 9(1):36–45, Jan. 1966.
12. H. Yan, J. Dai, T. Ji, X. Qiu, and Z. Zhang. A unified generative framework for aspect-based sentiment analysis. In C. Zong, F. Xia, W. Li, and R. Navigli, editors, Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 2416–2429, Online, Aug. 2021. Association for Computational Linguistics.
13. T. Zhu, L. Li, J. Yang, S. Zhao, H. Liu, and J. Qian. Multimodal sentiment analysis with image-text interaction network. Trans. Multi., 25:3375–3385, Jan. 2023.

3 Machine Comprehension

1. D. Chen, A. Fisch, J. Weston, and A. Bordes. Reading wikipedia to answer open-domain questions. CL, 2017.
2. K. M. Hermann, T. Koˇcisk´y, E. Grefenstette, L. Espeholt, W. Kay, M. Suleyman, and P. Blunsom. Teaching machines to read and comprehend, 2015.
3. P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang. SQuAD: 100,000+ questions for machine comprehen- sion of text. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pages 2383–2392, Austin, Texas, Nov. 2016. Association for Computational Linguistics.
4. M. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi. Bidirectional attention flow for machine comprehension, 2018.
5. S. Wang and J. Jiang. Machine comprehension using match-lstm and answer pointer, 2016.

4 Multimodal Learning

1. T. Baltruˇsaitis, C. Ahuja, and L.-P. Morency. Multimodal machine learning: A survey and taxonomy, 2017.
2. F. Dipaola, M. Gatti, A. Giaj Levra, R. Men‘e, D. Shiffer, R. Faccincani, Z. Raouf, A. Secchi, P. Ro- vere Querini, A. Voza, S. Badalamenti, M. Solbiati, G. Costantino, V. Savevski, and R. Furlan. Multi- modal deep learning for covid-19 prognosis prediction in the emergency department: a bi-centric study. Scientific Reports, 13(1):10868, 2023.
3. K. Fisher and Y. Marzouk. Can bayesian neural networks make confident predictions?, 2025.
4. G. Team, R. Anil, and S. et al. Gemini: A family of highly capable multimodal models, 2024.
5. S. Jabeen, X. Li, M. S. Amin, O. Bourahla, S. Li, and A. Jabbar. A review on methods and applications in multimodal deep learning. ACM Trans. Multimedia Comput. Commun. Appl., 19(2s), Feb. 2023.
6. Y. Li, S. Jiang, et al. Multimodal sentiment analysis with image-text correlation modal. In 2023 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics), pages 281–286. IEEE, 2023.
7. P. Nguyen, S. Sengupta, G. Malik, A. Gupta, and B. Min. Install: Context-aware instructional task assistance with multi-modal large language models, 2025.
8. J. Summaira, X. Li, A. M. Shoib, and J. Abdul. A review on methods and applications in multimodal deep learning, 2022.
9. Y. Wang, X. Li, Z. Yan, Y. He, J. Yu, X. Zeng, C. Wang, C. Ma, H. Huang, J. Gao, M. Dou, K. Chen, W. Wang, Y. Qiao, Y. Wang, and L. Wang. Internvideo2.5: Empowering video mllms with long and rich context modeling, 2025.
10. T. Zhu, L. Li, J. Yang, S. Zhao, H. Liu, and J. Qian. Multimodal sentiment analysis with image-text interaction network. Trans. Multi., 25:3375–3385, Jan. 2023.

5 Attention

1. G. Brauwers and F. Frasincar. A survey on aspect-based sentiment classification. ACM Comput. Surv., 55(4), Nov. 2022.

2. Q. Gu, Z. Wang, H. Zhang, S. Sui, and R. Wang. Aspect-level sentiment analysis based on syntax-aware and graph convolutional networks. Applied Sciences, 14(2), 2024.

3. M. Luong, H. Pham, and C. D. Manning. Effective approaches to attention-based neural machine translation. CoRR, abs/1508.04025, 2015.

4. M. Seo, A. Kembhavi, A. Farhadi, and H. Hajishirzi. Bidirectional attention flow for machine compre- hension, 2018.

5. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vish- wanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems, volume 30. Curran Associates, Inc., 2017.

6. P. Shaw, J. Uszkoreit, and A. Vaswani. Self-attention with relative position representations. CoRR, abs/1803.02155, 2018.

6 Question Answering

1. D. Chen, A. Fisch, J. Weston, and A. Bordes. Reading wikipedia to answer open-domain questions. CL, 2017.

2. T. Kwiatkowski, J. Palomaki, O. Redfield, M. Collins, A. Parikh, C. Alberti, D. Epstein, I. Polosukhin, J. Devlin, K. Lee, K. Toutanova, L. Jones, M. Kelcey, M.-W. Chang, A. M. Dai, J. Uszkoreit, Q. Le, and S. Petrov. Natural questions: A benchmark for question answering research. Transactions of the Association for Computational Linguistics, 7:452–466, 2019.

3. M. Pa¸sca. Open-domain question answering from large text collections. MIT Press, 2003.

4. M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. Deep contex- tualized word representations. CoRR, abs/1802.05365, 2018.

7 Instruction Tuning

1. S. Min, M. Lewis, L. Zettlemoyer, and H. Hajishirzi. Metaicl: Learning to learn in context, 2022.

2. V. Sanh, A. Webson, C. Raffel, S. H. Bach, L. Sutawika, Z. Alyafeai, A. Chaffin, A. Stiegler, T. L. Scao, A. Raja, M. Dey, M. S. Bari, C. Xu, U. Thakker, S. S. Sharma, E. Szczechla, T. Kim, G. Chhablani, N. Nayak, D. Datta, J. Chang, M. T.-J. Jiang, H. Wang, M. Manica, S. Shen, Z. X.

Yong, H. Pandey, R. Bawden, T. Wang, T. Neeraj, J. Rozen, A. Sharma, A. Santilli, T. Fevry, J. A. Fries, R. Teehan, T. Bers, S. Biderman, L. Gao, T. Wolf, and A. M. Rush. Multitask prompted training enables zero-shot task generalization, 2022.

3. Y. Wang, Y. Kordi, S. Mishra, A. Liu, N. A. Smith, D. Khashabi, and H. Hajishirzi. Self-instruct: Aligning language models with self-generated instructions, 2023.

4. J. Wei, M. Bosma, V. Y. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le. Finetuned language models are zero-shot learners, 2022.

5. C. Xu, Q. Sun, K. Zheng, X. Geng, P. Zhao, J. Feng, C. Tao, and D. Jiang. Wizardlm: Empowering large language models to follow complex instructions, 2023.

6. S. Zhang, L. Dong, X. Li, S. Zhang, X. Sun, S. Wang, J. Li, R. Hu, T. Zhang, F. Wu, and G. Wang. Instruction tuning for large language models: A survey, 2024.

8 Prompt Engineering

1. Prompt engineering guide, https://www.promptingguide.ai/introduction/basics.

2. H. He, H. Zhang, and D. Roth. Rethinking with retrieval: Faithful large language model inference, 2022.

3. J. Huang and K. C.-C. Chang. Towards reasoning in large language models: A survey. In A. Rogers, J. Boyd-Graber, and N. Okazaki, editors, Findings of the Association for Computational Linguistics: ACL 2023, pages 1049–1065, Toronto, Canada, July 2023. Association for Computational Linguistics.

4. T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, and Y. Iwasawa. Large language models are zero-shot reasoners, 2023.

5. P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing, 2021.

6. V. Sanh, et al. Multitask prompted training enables zero-shot task generalization, 2022.

7. J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. Chi, Q. Le, and D. Zhou. Chain-of- thought prompting elicits reasoning in large language models, 2023.

8. Z. Zhu, Y. Xue, X. Chen, D. Zhou, J. Tang, D. Schuurmans, and H. Dai. Large language models can learn rules, 2024.

9 Code generation

1. M. Chen, J. Tworek, H. Jun, Q. Yuan, H. P. de Oliveira Pinto, J. Kaplan, H. Edwards, Y. Burda, N. Joseph, G. Brockman, A. Ray, R. Puri, G. Krueger, M. Petrov, H. Khlaaf, G. Sastry, P. Mishkin, B. Chan, S. Gray, N. Ryder, M. Pavlov, A. Power, L. Kaiser, M. Bavarian, C. Winter, P. Tillet, F. P. Such, D. Cummings, M. Plappert, F. Chantzis, E. Barnes, A. Herbert-Voss, W. H. Guss, A. Nichol, A. Paino, N. Tezak, J. Tang, I. Babuschkin, S. Balaji, S. Jain, W. Saunders, C. Hesse, A. N. Carr, J. Leike, J. Achiam, V. Misra, E. Morikawa, A. Radford, M. Knight, M. Brundage, M. Murati, K. Mayer, P. Welinder, B. McGrew, D. Amodei, S. McCandlish, I. Sutskever, and W. Zaremba. Evaluating large language models trained on code, 2021.

2. J. Cruz-Benito, S. Vishwakarma, F. Martin-Fernandez, and I. Faro. Automated source code generation and auto-completion using deep learning: Comparing and discussing current language model-related approaches. AI, 2:1–16, 01 2021.

3. Z. Feng, D. Guo, D. Tang, N. Duan, X. Feng, M. Gong, L. Shou, B. Qin, T. Liu, D. Jiang, and M. Zhou. CodeBERT: A pre-trained model for programming and natural languages. In T. Cohn, Y. He, and Y. Liu, editors, Findings of the Association for Computational Linguistics: EMNLP 2020, pages 1536– 1547, Online, Nov. 2020. Association for Computational Linguistics.

4. T. H. M. Le, H. Chen, and M. A. Babar. Deep learning for source code modeling and generation: Models, applications and challenges. CoRR, abs/2002.05442, 2020.

5. E. Nijkamp, B. Pang, H. Hayashi, L. Tu, H. Wang, Y. Zhou, S. Savarese, and C. Xiong. Codegen: An open large language model for code with multi-turn program synthesis, 2023.

6. P. Shojaee, A. Jain, S. Tipirneni, and C. K. Reddy. Execution-based code generation using deep rein- forcement learning, 2023.

7. Y. Wang, H. Le, A. D. Gotmare, N. D. Q. Bui, J. Li, and S. C. H. Hoi. Codet5+: Open code large language models for code understanding and generation, 2023.

8. Y. Wang, W. Wang, S. R. Joty, and S. C. H. Hoi. Codet5: Identifier-aware unified pre-trained encoder- decoder models for code understanding and generation. CoRR, abs/2109.00859, 2021.

9. D. Zan, B. Chen, D. Yang, Z. Lin, M. Kim, B. Guan, Y. Wang, W. Chen, and J.-G. Lou. Cert: Continual pre-training on sketches for library-oriented code generation, 2022.

10 Distillation

11 Quantization

12 Pruning