

Papers timeline

JR

March 13, 2025

Contents

1	2025	2
1.1	March	2
1.2	February	2
1.3	January	2
2	2024	3
2.1	December	3
2.2	September	3
2.3	June	3
2.4	May	3
2.5	February	3
2.6	January	3
3	2023	4
4	2022	5
5	2019	6
5.1	February	6
6	2018	7
6.1	June	7
7	2017	8

1 2025

1.1 March

1.2 February

1.3 January

- [1] DeepSeek-AI *et al.*, *DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning*, arXiv:2501.12948 [cs], Jan. 2025. DOI: 10.48550/arXiv.2501.12948. [Online]. Available: <http://arxiv.org/abs/2501.12948> (visited on 03/14/2025).

2 2024

2.1 December

- [1] Qwen *et al.*, *Qwen2.5 Technical Report*, arXiv:2412.15115 [cs], Jan. 2025. DOI: 10.48550/arXiv.2412.15115. [Online]. Available: <http://arxiv.org/abs/2412.15115> (visited on 03/14/2025).

2.2 September

- [1] N. Muennighoff *et al.*, *OLMoE: Open Mixture-of-Experts Language Models*, arXiv:2409.02060 [cs], Mar. 2025. DOI: 10.48550/arXiv.2409.02060. [Online]. Available: <http://arxiv.org/abs/2409.02060> (visited on 03/14/2025).

2.3 June

- [1] G. Penedo *et al.*, *The FineWeb Datasets: Decanting the Web for the Finest Text Data at Scale*, arXiv:2406.17557 [cs], Oct. 2024. DOI: 10.48550/arXiv.2406.17557. [Online]. Available: <http://arxiv.org/abs/2406.17557> (visited on 03/14/2025).

2.4 May

- [1] A. Grattafiori *et al.*, *The llama 3 herd of models*, 2024. arXiv: 2407.21783 [cs.AI]. [Online]. Available: <https://arxiv.org/abs/2407.21783>.
- [2] A. Gu and T. Dao, *Mamba: Linear-Time Sequence Modeling with Selective State Spaces*, arXiv:2312.00752 [cs], May 2024. DOI: 10.48550/arXiv.2312.00752. [Online]. Available: <http://arxiv.org/abs/2312.00752> (visited on 03/14/2025).

2.5 February

- [1] D. Groeneveld *et al.*, *OLMo: Accelerating the Science of Language Models*, arXiv:2402.00838 [cs], Jun. 2024. DOI: 10.48550/arXiv.2402.00838. [Online]. Available: <http://arxiv.org/abs/2402.00838> (visited on 03/14/2025).

2.6 January

- [1] DeepSeek-AI *et al.*, *DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model*, arXiv:2405.04434 [cs], Jun. 2024. DOI: 10.48550/arXiv.2405.04434. [Online]. Available: <http://arxiv.org/abs/2405.04434> (visited on 03/14/2025).

3 2023

4 2022

5 2019

5.1 February

- [1] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, “Language Models are Unsupervised Multitask Learners,” 2019. [Online]. Available: <https://www.semanticscholar.org/paper/Language-Models-are-Unsupervised-Multitask-Learners-Radford-Wu/9405cc0d6169988371b2755e573cc28650d14dfe> (visited on 03/14/2025).

6 2018

6.1 June

- [1] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, “Improving Language Understanding by Generative Pre-Training,” en,

7 2017