## 已整理頭

Li, R., Allal, L. B., Zi, Y., Muennighoff, N., Kocetkov, D., Mou, C., ... & de Vries, H. (2023). Starcoder: may the source be with you!. *arXiv preprint arXiv:2305.06161*. <https://doi.org/10.48550/arXiv.2305.06161>

Dou, S., Jia, H., Wu, S., Zheng, H., Zhou, W., Wu, M., ... & Huang, X. (2024). What's wrong with your code generated by large language models? an extensive study. *arXiv preprint arXiv:2407.06153*.

<https://doi.org/10.48550/arXiv.2407.06153>

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824-24837.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. *arXiv preprint arXiv:2303.18223*, *1*(2).

## 已整理尾

## 未整理

Mirakhorli, M., & Cleland-Huang, J. (2015). Detecting, tracing, and monitoring architectural tactics in code. *IEEE Transactions on Software Engineering*, *42*(3), 205-220. <https://doi.org/10.1109/TSE.2015.2479217>

Ehsani, R., Pathak, S., & Chatterjee, P. (2025, April). Towards detecting prompt knowledge gaps for improved llm-guided issue resolution. In *2025 IEEE/ACM 22nd International Conference on Mining Software Repositories (MSR)* (pp. 699-711). IEEE. <https://doi.org/10.1109/MSR66628.2025.00107>

Khojah, R., de Oliveira Neto, F. G., Mohamad, M., & Leitner, P. (2025). The Impact of Prompt Programming on Function-Level Code Generation. *IEEE Transactions on Software Engineering*. <https://doi.org/10.1109/TSE.2025.3587794>

Bozkurt, A. (2023). Generative artificial intelligence (AI) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*, *18*(1). <https://orcid.org/0000-0002-4520-642X>

Khazanchi, R. A. S. H. M. I., & Khazanchi, P. A. N. K. A. J. (2024). Generative AI to improve special education teacher preparation for inclusive classrooms. *Exploring new horizons: Generative artificial intelligence and teacher education*, *159*.

Doughty, J., Wan, Z., Bompelli, A., Qayum, J., Wang, T., Zhang, J., ... & Sakr, M. (2024, January). A comparative study of AI-generated (GPT-4) and human-crafted MCQs in programming education. In *Proceedings of the 26th Australasian Computing Education Conference* (pp. 114-123). <https://doi.org/10.1145/3636243.3636256>

Weisz, J. D., He, J., Muller, M., Hoefer, G., Miles, R., & Geyer, W. (2024, May). Design principles for generative AI applications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-22). <https://doi.org/10.1145/3613904.3642466>

Dave, D. M., Mandvikar, S., & Engineer, P. A. (2023). Augmented intelligence: Human-AI collaboration in the era of digital transformation. *International Journal of Engineering Applied Sciences and Technology*, *8*(6), 24-33.

Walter, Y. (2024). Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, *21*(1), 15. <https://doi.org/10.1186/s41239-024-00448-3>

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, *35*, 24824-24837.

Liu, X., Wang, J., Sun, J., Yuan, X., Dong, G., Di, P., ... & Wang, D. (2023). Prompting frameworks for large language models: A survey. *arXiv preprint arXiv:2311.12785*.

Lo, L. S. (2023). The CLEAR path: A framework for enhancing information literacy through prompt engineering. *The Journal of Academic Librarianship*, *49*(4), 102720. <https://doi.org/10.1016/j.acalib.2023.102720>

Nichols, B., Buttlar, D., & Farrell, J. (1996). *Pthreads programming: A POSIX standard for better multiprocessing*. " O'Reilly Media, Inc.".

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., ... & Wu, Y. (2023). How close is chatgpt to human experts? comparison corpus, evaluation, and detection. *arXiv preprint arXiv:2301.07597*.

Cahyawijaya, S., Winata, G. I., Wilie, B., Vincentio, K., Li, X., Kuncoro, A., ... & Fung, P. (2021). IndoNLG: Benchmark and resources for evaluating Indonesian natural language generation. *arXiv preprint arXiv:2104.08200*. <https://doi.org/10.48550/arXiv.2104.08200>

Shi, Z., Mathur, U., & Pavlogiannis, A. (2024, April). Optimistic prediction of synchronization-reversal data races. In *Proceedings of the IEEE/ACM 46th International Conference on Software Engineering* (pp. 1-13). <https://doi.org/10.1145/3597503.3639099>

Chiang, W. L., Li, Z., Lin, Z., Sheng, Y., Wu, Z., Zhang, H., ... & Xing, E. P. (2023). Vicuna: An open-source chatbot impressing gpt-4 with 90%\* chatgpt quality. *See https://vicuna. lmsys. org (accessed 14 April 2023)*, *2*(3), 6.

Ding, N., Chen, Y., Xu, B., Qin, Y., Zheng, Z., Hu, S., ... & Zhou, B. (2023). Enhancing chat language models by scaling high-quality instructional conversations. *arXiv preprint arXiv:2305.14233*. <https://doi.org/10.48550/arXiv.2305.14233>

Cui, G., Yuan, L., Ding, N., Yao, G., Zhu, W., Ni, Y., ... & Sun, M. (2023). Ultrafeedback: Boosting language models with high-quality feedback.