Reference List

1. Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y. K., Wu, Y., & Guo, D. (2024). *DeepSeekMath: Pushing the limits of mathematical reasoning in open language models*. arXiv. https://doi.org/10.48550/arXiv.2402.03300
2. Kalomaze. (2025, March 15). *GRPO judge experiments: Findings & empirical observations* [Blog post]. Bearblog. https://kalomaze.bearblog.dev/grpo-judge-experiments-findings-and-empirical-observations/
3. Unsloth AI. (n.d.). *Unsloth: Open-source fine-tuning for LLMs*. https://unsloth.ai/
4. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., & Christiano, P. (2022). *Training language models to follow instructions with human feedback*. arXiv. https://doi.org/10.48550/arXiv.2212.10522
5. Weights & Biases. (n.d.). *The AI developer platform*. https://wandb.ai/