11. https://github.com/topics/natural-language-to-sql 12. https://github.com/FerreroJeremy/ln2sql 13. https://github.com/rupinder1133/ln2sqlmodule 14. https://github.com/maverickjoy/ln2sql 15. https://github.com/joanna-bb/Postgres_LLM/blob/main/Postgres_LLM_Llama_ex1.py 16. https://github.com/bhattbhavesh91/google-gemma-finetuning-n2sql/blob/main/n2sql-google-gemma-finetuning-notebook.ipynb 17. https://github.com/bhattbhavesh91/n2sql-google-gemini 18. https://forum.octopus.energy/t/maximising-oversized-array-or-arguably-undersized-inverter/9750/14 19. https://www.marktechpost.com/2024/05/02/this-ai-paper-introduces-llama-3-8b-instruct-80k-qlora-new-horizons-in-ai-contextual-understanding/?amp 20. https://huggingface.co/namespace-Pt/Llama-3-8B-Instruct-80K-QLoRA 21. https://huggingface.co/namespace-Pt/Llama-3-8B-Instruct-80K-QLoRA-Merged-GGUF/blob/main/Llama-3-8B-Instruct-80K-QLoRA-Merged-Q4_K_M.gguf 22. https://www.unite.ai/decoder-based-large-language-models-a-complete-guide/ 23. https://www.google.com/search?q=llm+lora+tutorial&oq=llm+lora+&gs_lcrp=EgZjaHJvbWUqBwgFEAAYgAQyBggAEEUYOTIHCAEQABiABDIHCAIQABiABDIHCAMQABiABDIHCAQQABiABDIHCAUQABiABDIHCAYQABiABDIICAcQABgWGB4yCAgIEAAYFhgeMggICRAAGBYYHjIICAoQABgWGB4yCAgLEAAYFhgeMggIDBAAGBYYHjIICA0QABgWGB4yCAgOEAAYFhge0gEJMTg2NzVqMGo3qAIUsAIB&client=ms-android-oneplus-rvo3&sourceid=chrome-mobile&ie=UTF-8#ip=1 24. https://zohaib.me/a-beginners-guide-to-fine-tuning-llm-using-lora/amp/ 25. https://xiaosean5408.medium.com/fine-tuning-llms-made-easy-with-lora-and-generative-ai-stable-diffusion-lora-39ff27480fda 26. https://medium.com/data-science-in-your-pocket/lora-for-fine-tuning-llms-explained-with-codes-and-example-62a7ac5a3578 27. https://www.datacamp.com/tutorial/mastering-low-rank-adaptation-lora-enhancing-large-language-models-for-efficient-adaptation 28. https://www.databricks.com/blog/efficient-fine-tuning-lora-guide-llms 29. https://magazine.sebastianraschka.com/p/practical-tips-for-finetuning-llms 30. https://learnopencv.com/sdxl-inpainting/ 31. https://civitai.com/models/176555?modelVersionId=214296 32. https://civitai.com/tag/text 33. https://venturebeat.com/ai/metas-new-multi-token-prediction-makes-ai-models-up-to-3x-faster/ 34. https://venturebeat.com/ 35. https://www.google.com/search?q=how+to+embed+factual+knowledge+into+llm&oq=how+to+embed+factual+knowledge+into+llm&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIHCAEQIRigAdIBCTI3NzcxajBqN6gCFLACAQ&client=ms-android-oneplus-rvo3&sourceid=chrome-mobile&ie=UTF-8 36. https://www.marktechpost.com/2024/05/06/nvidia-publishes-a-competitive-llama3-70b-quality-assurance-qa-retrieval-augmented-generation-rag-fine-tune-model/?amp 37. https://huggingface.co/nvidia/Llama3-ChatQA-1.5-8B/discussions/5 38. https://huggingface.co/QuantFactory/NVIDIA-Llama3-ChatQA-1.5-8B-GGUF 39. https://alphaarchitect.com/2024/05/forecast-equity-risk-premium/ 40. https://github.com/zylon-ai/private-gpt/pull/1825 41. https://huggingface.co/blog/cost-efficient-rag-applications-with-intel 42. https://huggingface.co/Qwen/CodeQwen1.5-7B-Chat 43. https://huggingface.co/deepseek-ai/deepseek-coder-6.7b-instruct 44. https://future.mozilla.org/news/llamafiles-for-embeddings-in-local-rag-applications/ 45. https://www.google.com/search?client=ms-android-oneplus-rvo3&sca_esv=1a5929d447859cf0&sxsrf=ADLYWII7aAUDwsLy9GDbpNCso2Djb-OMnA:1715908662447&q=Pytorch+parallel+inference+on+single+GPU&sa=X&ved=2ahUKEwiu6qrywZOGAxXrV0EAHVj2BZIQ1QJ6BAgeEAE&biw=360&bih=663&dpr=3