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

1. https://github.com/CopilotKit/CopilotKit 2. https://github.com/OS-Copilot/OS-Copilot 3. https://github.com/ex3ndr/llama-coder 4. https://github.com/Kuingsmile/word-GPT-Plus?tab=readme-ov-file 5.

https://github.com/srikanth235/privy 6. https://github.com/rubberduck-ai/rubberduck-vscode 7. https://github.com/Bin-Huang/chatbox?tab=readme-ov-file

14. https://www.marktechpost.com/2024/05/19/meet-verba-1-0-run-state-of-the-art-rag-locally-with-ollama-integration-and-open-source-models/?amp 15. https://github.com/weaviate/Verba?tab=readme-ov-file 16. https://amp.theguardian.com/technology/article/2024/may/20/microsoft-chatbot-assistant-pc 17. https://amp.theguardian.com/technology/article/2024/may/20/world-is-ill-prepared-for-breakthroughs-in-ai-say-experts 19. https://venturebeat.com/ai/the-future-of-financial-analysis-how-gpt-4-is-disrupting-the-industry-according-to-new-research/ 20. https://www.guru3d.com/story/microsoft-reportedly-readies-billion-bid-to-acquire-valve-steam/ 21. https://www.reuters.com/technology/eu-data-protection-board-says-chatgpt-still-not-meeting-data-accuracy-standards-2024-05-24/ 23. https://www.wired.com/story/anthropic-black-box-ai-research-neurons-features/ 24. https://www.marktechpost.com/2024/05/21/gradient-ai-introduces-llama-3-8b-gradient-instruct-1048k-setting-new-standards-in-long-context-ai/ 25. https://decrypt.co/232299/mistral-ai-7b-upgrade-uncensored-cohere-aya-open-source 26. https://www.phoronix.com/news/Llamafile-0.8.5-Released 27. https://www.bing.com/chat?showconv=1 28. https://huggingface.co/spaces/ehristoforu/dalle-3-xl-lora-v2 29. https://www.marktechpost.com/2024/05/28/nv-embed-nvidias-groundbreaking-embedding-model-dominates-mteb-benchmarks/ 30. https://www.marktechpost.com/2024/05/31/autocoder-the-first-large-language-model-to-surpass-gpt-4-turbo-april-2024-and-gpt-4o-in-pass1-on-the-human-eval-benchmark-test-90-9-vs-90-2/ 31. https://huggingface.co/Bin12345/AutoCoder_S_6.7B 32. https://opencodeinterpreter.github.io/ 33. https://www.marktechpost.com/2024/05/26/finrobot-a-novel-open-source-ai-agent-platform-supporting-multiple-financially-specialized-ai-agents-powered-by-llms/ 34. https://github.com/ai4finance-foundation/finrobot 35. https://ai4finance.org/projects 36. https://github.com/AI4Finance-Foundation/FinGPT 37. https://www.marktechpost.com/2024/06/01/llamafs-an-open-source-self-organizing-file-system-with-llama-3/ 38. https://www.marktechpost.com/2024/06/01/rag-me-up-a-generic-ai-framework-server-uis-that-enables-you-to-do-rag-on-your-own-dataset-easily/?amp 39. https://huggingface.co/mistralai/Codestral-22B-v0.1 40. https://huggingface.co/nvidia/NV-Embed-v1 41. https://www.marktechpost.com/2024/06/03/are-ai-rag-solutions-really-hallucination-free-researchers-at-stanford-university-assess-the-reliability-of-ai-in-legal-research-hallucinations-and-accuracy-challenges/?amp 43. https://venturebeat.com/ai/mistral-launches-fine-tuning-tools-to-make-customizing-its-models-easier-and-faster/ 46. https://github.com/mistralai/mistral-finetune 48. https://venturebeat.com/ai/mistral-announces-codestral-its-first-programming-focused-ai-model/ 49. https://venturebeat.com/ai/from-gen-ai-1-5-to-2-0-moving-from-rag-to-agent-systems/

From:
http://wuff.dyndns.org/ - **Wulf's Various Things**

Permanent link:
http://wuff.dyndns.org/doku.php?id=temp:bookmarks&rev=1717772045

Last update: **2024/06/07 15:54**