大模型从1到1学习-资源集合

大模型资源目录合集（总能找到你想要的）

智能体

1. Significant-Gravitas/AutoGPT - AutoGPT旨在让所有人都能使用和开发人工智能。其使命是为人们提供专注于重要事务的工具。
2. geekan/MetaGPT - 第一家人工智能软件公司的多智能体框架面向自然语言编程。
3. microsoft/autogen - 一个用于自主人工智能的编程框架，在PyPi、Discord和Office Hour上有相关资源。
4. reworkd/AgentGPT - 在浏览器中组装、配置和部署自主人工智能代理。

5. joaomdmoura/crewAI - 角色扮演和自主人工智能体框架。它使智能体能够协作并处理复杂任务。
6. microsoft/JARVIS - JARVIS是一个用于将大型语言模型（LLMs）与机器学习（ML）社区连接起来的系统。（论文：https://arxiv.org/pdf/2303.17580.pdf）
7. mem0ai/mem0 - 人工智能应用的存储层。
8. microsoft/semantic-kernel - 快速且轻松地将最先进的大型语言模型（LLM）技术集成到您的应用程序中。
9. yoheinakajima/babyagi -
10. openai/swarm - 由OpenAI解决方案团队管理的符合人体工程学、轻量级多智能体编排的教育框架。
11. phidatahq/phidata - 构建具有记忆、知识、工具和推理能力的多模态智能体，并通过美观的智能体用户界面进行聊天。
12. TransformerOptimus/SuperAGI - SuperAGI是一个开发者优先的开源自主人工智能代理框架，它能让开发者快速、可靠地构建、管理和运行有用的自主代理。
13. composiohq/composio - Composio通过函数调用为人工智能代理和大型语言模型（LLMs）配备100多种高质量集成。
14. cpacker/MemGPT - Letta（以前叫MemGPT），一个用于创建具有记忆功能的大型语言模型（LLM）服务的框架。
15. google-deepmind/deepmind-research - 该存储库包含DeepMind出版物的实现代码和示例代码。
16. botpress/botpress - 用于构建和部署GPT/LLM智能体的开源中心。
17. OpenMOSS/MOSS - 一个由复旦大学开发且借助工具增强的开源对话式语言模型。
18. smol-ai/developer - 首个能让你在自己的应用中嵌入开发者代理的库。
19. OpenBMB/XAgent - 用于解决复杂任务的自主语言模型代理。
20. langchain-ai/langgraph - 以图的形式构建具有弹性的语言智能体。
21. e2b-dev/e2b - 用于人工智能应用程序和代理的安全开源云运行时。
22. modelscope/agentscope - 更轻松地构建由大型语言模型（LLM）赋能的多智能体应用程序。
23. homanp/superagent - 通过API运行人工智能代理。
24. aiwaves-cn/agents -
25. frdel/agent-zero - 零号特工人工智能框架。
26. microsoft/TinyTroupe - 由大型语言模型（LLM）驱动的多智能体角色模拟，用于提升想象力和获取商业洞见。
27. QwenLM/Qwen-Agent - 基于Qwen≥2.0的代理框架和应用程序，具有函数调用、代码解释器、检索增强生成（RAG）和Chrome扩展功能。

28. OpenBMB/AgentVerse - AgentVerse旨在用于在应用程序中部署多个基于大型语言模型（LLM）的代理，主要提供任务解决和模拟框架。
29. Significant-Gravitas/Auto-GPT-Plugins - Auto - GPT的插件。
30. huggingface/smolagents - Smolagents是一个用于代理（agents）的基础库。代理（agents）使用它来编写用于工具调用（tool - calling）和代理编排（agent - orchestrating）的Python代码。
31. Ironclad/rivet - 一个开源的可视化人工智能编程环境和TypeScript库。
32. gmpetrov/databerry - 一个用于创建自定义大型语言模型（LLM）智能体的无代码平台。
33. OpenBMB/BMTools - 大型模型的工具学习与ChatGPT插件的开源解决方案。
34. langroid/langroid - 使用多智能体编程来控制大型语言模型。
35. muellerberndt/mini-agi - MiniAGI是一个简单的通用型自主智能体，依赖于OpenAI API。
36. Farama-Foundation/PettingZoo - 一种多智能体强化学习的应用程序接口（API）标准，包括常用的参考环境和实用程序。
37. Josh-XT/AGiXT - AGiXT是一个动态的人工智能平台，它使用自适应记忆、智能功能和插件系统管理指令并在多个人工智能供应商之间执行任务，以提供高效的人工智能解决方案。
38. togethercomputer/moa - 混合代理（MoA）在使用开源软件模型的情况下，在羊驼评估（AlpacaEval）中达到了65.1%的成绩。
39. AgentOps-AI/agentops - 用于人工智能代理监测、大型语言模型（LLM）成本追踪和基准测试的Python软件开发工具包（SDK）。它与各种大型语言模型和代理框架集成。
40. noahshinn/reflexion - [NeurIPS 2023]《反思（Reflexion）：基于言语强化学习的语言智能体》
41. SciSharp/BotSharp - .NET中的人工智能多智能体框架。
42. dot-agent/nextpy -
43. iterative/datachain - 非结构化数据的提取、转换、加载（ETL）、分析和版本控制。
44. agiresearch/OpenAGI - OpenAGI：大型语言模型（LLM）与领域专家的相遇。
45. InternLM/lagent - 一个用于创建基于大型语言模型的代理的轻量级框架。
46. MineDojo/MineDojo - 利用互联网规模的知识构建开放式具身智能体。
47. Forethought-Technologies/AutoChain - AutoChain用于创建轻量级、可扩展且可测试的大型语言模型（LLM）智能体。
48. landing-ai/vision-agent - 视觉代理。
49. BCG-X-Official/agentkit - 一个用于使用Nextjs、FastAPI和Langchain构建受限代理的入门套件。
50. jina-ai/thinkgpt - 用于增强大型语言模型（LLM）并突破其局限的代理技术。
51. farizrahman4u/loopgpt - 一个用于Auto - GPT的模块化框架。

52. Farama-Foundation/chatarena - ChatArena是一个用于大型语言模型（LLMs）的多智能体语言游戏环境，旨在开发人工智能的沟通和协作能力。
53. THUDM/AgentTuning - 智能体调优（Agent Tuning）为大型语言模型赋予通用的智能体能力。
54. Yifan-Song793/RestGPT - 基于大型语言模型的自主代理通过RESTful API（表述性状态传递应用程序接口）控制现实世界中的应用。
55. Link-AGI/AutoAgents - 在2024年国际人工智能联合会议（IJCAI）上，生成了不同的GPT角色以构成一个协作实体来处理复杂任务。
56. AI-Engineer-Foundation/agent-protocol - 这是一个与人工智能代理交互的通用接口，与技术栈无关，可用于任何代理构建框架。
57. kreneskyp/ix - 一个用于自主GPT - 4的代理平台。

提示词工程

1. f/awesome-chatgpt-prompts - 这个资源库整理了ChatGPT提示词，以更好地使用ChatGPT和其他大型语言模型（LLM）工具。
2. PlexPt/awesome-chatgpt-prompts-zh - ChatGPT中文调教指南。各类场景使用指南。学习如何让它遵循你的指令。
3. dair-ai/Prompt-Engineering-Guide - 提示工程的指南、论文、讲座、笔记和资源。
4. stanfordnlp/dspy - DSPy：一个用于对编程语言模型进行编程而非提示（prompting）的框架。
5. guidance-ai/guidance - 一种用于控制大型语言模型的引导语言。
6. outlines-dev/outlines - 结构化文本生成
7. mshumer/gpt-prompt-engineer -
8. jxnl/instructor - 大型语言模型（LLMs）的结构化输出。
9. brexhq/prompt-engineering - 使用OpenAI的GPT - 4等大型语言模型的技巧和诀窍。
10. LouisShark/chatgpt_system_prompt - 一组GPT系统提示词以及有关提示注入/泄露的知识。
11. microsoft/TypeChat - TypeChat是一个用于构建带有类型的自然语言接口的库。
12. sgl-project/sglang - SGLang是一个用于大型语言模型和视觉 - 语言模型的快速服务框架。
13. mit-han-lab/streaming-llm - 2024年国际学习表征会议（ICLR）上提出的带有注意力汇聚（Attention Sinks）的高效流式语言模型。
14. spdustin/ChatGPT-AutoExpert - 用于ChatGPT（非编码）和ChatGPT高级数据分析（编码）的增强型自定义指令。
15. civitai/civitai - 一个包含模型和文本反转的存储库。
16. Moonvy/OpenPromptStudio - AIGC提示词可视化编辑器 | 运维 | 开放式提示工作室

17. rockbenben/ChatGPT-Shortcut - 通过人工智能快捷方式最大限度地提高效率和生产力。定制、保存并分享提示，并在共享社区中找到适用于不同场景的提示。
18. microsoft/promptbase - 与提示工程相关的所有内容。
19. PrefectHQ/marvin - 创建令人愉悦的人工智能接口。
20. promptfoo/promptfoo - 测试提示词、代理和检索增强生成（RAG）。同时对大型语言模型（LLM）进行红队测试、渗透测试和漏洞扫描，比较大型语言模型的性能，并通过命令行和持续集成/持续部署（CI/CD）集成进行简单配置。
21. princeton-nlp/tree-of-thought-llm - 关于在2023年神经信息处理系统大会（NeurIPS 2023）上利用大型语言模型进行蓄意问题解决的思考。
22. pydantic/pydantic-ai - 用于在大型语言模型（LLMs）中使用Pydantic的代理框架或填充程序。
23. 1rgs/jsonformer - 一种从语言模型生成结构化JSON的可靠方法。
24. thunlp/OpenPrompt - 一个用于提示学习的开源框架。
25. guardrails-ai/guardrails - 为大型语言模型添加安全限制或约束条件。
26. eth-sri/lmql - 一种在约束引导下高效对大型语言模型（LLMs）进行编程的语言。
27. promptslab/Promptify - 提示工程与版本控制，使用GPT或其他基于提示的模型以获取结构化输出。加入Discord进行相关研究。
28. shreyashankar/gpt3-sandbox - 该项目旨在让用户通过使用新的OpenAI GPT-3 API用几行Python代码创建出色的网络演示。
29. hegelai/prompttools - 用于提示测试/试验的开源工具，支持大型语言模型（如OpenAI、LLaMA）和向量数据库（如Chroma、Weaviate、LanceDB）。
30. bigscience-workshop/promptsource - 一个用于自然语言提示的工具包，包括创建、共享和使用。
31. YiVal/YiVal - 你的通用人工智能应用自动提示工程助手。
32. microsoft/prompt-engine - 一个协助开发者创建大型语言模型提示的库。
33. ianarawjo/ChainForge - 一个用于可视化编程的开源环境，用于对大型语言模型（LLMs）的提示进行实战测试。
34. spcl/graph-of-thoughts - 《思维图谱：用大型语言模型解决复杂问题》的官方实现。
35. ysymyth/ReAct - [ICLR 2023]《ReAct：在语言模型中结合推理与行动》
36. Microsoft/genaiscript - 人工智能（生成式人工智能）脚本的自动化生成。
37. jackmpcollins/magentic - 不间断地将大型语言模型集成为Python函数。
38. adieyal/sd-dynamic-prompts - 一个为AUTOMATIC1111/stable - diffusion - webui编写的自定义脚本，用于创建一个小型模板语言以随机生成提示词。

39. zjunlp/EasyEdit - 一种用于2024年美国计算语言学协会（ACL）会议中大型语言模型（LLMs）的易于使用的知识编辑框架。
40. microsoft/aici - AICI：提示作为WebAssembly程序。
41. zou-group/textgrad - TextGrad：通过文本的自动“求导”，利用大型语言模型对文本梯度进行反向传播。
42. microsoft/PromptCraft-Robotics - 一个在机器人领域使用大型语言模型（LLMs）的社区以及一个与ChatGPT集成的机器人模拟器。
43. greshake/llm-security - 破坏集成应用的大型语言模型的新方法。
44. noamgat/lm-format-enforcer - 强化语言模型的输出格式（如JSON模式、正则表达式等）。
45. Ber666/llm-reasoners - 一个用于大型语言模型中复杂推理的库。
46. jujumilk3/leaked-system-prompts - 泄露系统提示的集合。
47. laiyer-ai/llm-guard - 大型语言模型交互的安全工具包。
48. hiyouga/FastEdit - 10秒内快速编辑大型语言模型。
49. timqian/openprompt.co - 创建、使用和分享ChatGPT提示。
50. explosion/spacy-llm - 将大型语言模型（LLMs）集成到结构化自然语言处理（NLP）流程中。
51. protectai/rebuff - 大型语言模型（LLM）提示注入检测器。
52. getmetal/motorhead - Motorhead是一个用于大型语言模型（LLMs）的服务器，用于内存和信息检索。
53. Mirascope/mirascope - 不具阻碍性的大型语言模型（LLM）抽象概念。
54. cocacola-lab/ChatIE -

AI开发接口

1. jmorganca/ollama - 使用Llama 3.3、Mistral、Gemma 2和其他大型语言模型快速上手。
2. ChatGPTNextWeb/ChatGPT-Next-Web - ChatGPT、Gemini等的跨平台用户界面（UI）使您能够一键拥有自己的大型语言模型（LLM）应用程序。
3. xtekky/gpt4free - 官方的gpt4free仓库包含各种强大的语言模型。
4. oobabooga/text-generation-webui - 一个用于大型语言模型的Gradio网络用户界面，支持多个推理后端。
5. RVC-Boss/GPT-SoVITS - 一分钟的语音数据可用于训练一个良好的语音合成（TTS）模型（小样本语音克隆）。
6. gradio-app/gradio - 构建并分享优秀的Python机器学习应用程序。点赞以支持。
7. mckaywrigley/chatbot-ui - 所有型号均提供人工智能聊天功能。
8. openai/openai-python - 用于OpenAI API的官方Python库。

9. danny-avila/LibreChat - 一个增强版的ChatGPT克隆版本，具有各种特性，如不同的API、人工智能模型和功能，并且它是一个活跃的自托管开源项目。
10. sunner/ChatALL - 同时与多个聊天机器人（如ChatGPT、必应聊天等）聊天以找到最佳答案。
11. GaiZhenbiao/ChuanhuChatGPT - ChatGPT API和许多大型语言模型（LLMs）的图形用户界面（GUI）。它有各种功能，如智能体（agents）和基于文件的问答（file - based QA）等，并且有一个美观的用户界面。
12. CopilotKit/CopilotKit - 用于各种人工智能应用（如副驾驶、应用内代理、聊天机器人和文本区域）的React UI和优雅的基础架构。
13. mlc-ai/web-llm - 高性能的浏览器内大型语言模型（LLM）推理引擎。
14. jina-ai/clip-as-service - 使用CLIP对图像和句子进行可扩展的嵌入、推理和排序。
15. chathub-dev/chathub - 一站式聊天机器人客户端。
16. TheRamU/Fay - Fay是一个开源数字人框架。它有适用于各种应用的不同版本。
17. sashabaranov/go-openai - 用于OpenAI ChatGPT、GPT - 3、GPT - 4、DALL·E和Whisper API的Go语言包装器。
18. SillyTavern/SillyTavern - 大型语言模型（LLM）中面向高级用户的前端。
19. openai/openai-node - OpenAI API的官方JavaScript/TypeScript库。
20. sebastianstarke/AI4Animation - 在Unity中利用电脑智能让角色栩栩如生。
21. xiangsx/gpt4free-ts - 在一个xtekky/gpt4free的TypeScript版本的复刻项目中提供了一个免费的OpenAI GPT - 4 API。
22. wzpan/wukong-robot - 悟空机器人是一个简单、灵活且优雅的中文语音对话机器人/智能音箱项目。它支持ChatGPT多轮对话，并且可能是首个支持脑机交互的开源智能音箱项目。
23. yihong0618/xiaogpt - 使用小米智能音箱玩ChatGPT和其他大型语言模型（LLM）。
24. nat/openplayground - 一个可在笔记本电脑上运行的大型语言模型（LLM）游乐场。
25. postgresml/postgresml - 适用于机器学习和人工智能应用的带GPU的Postgres（一种数据库管理系统）。
26. Shaunwei/RealChar - 创建、定制人工智能角色/伙伴并与其进行实时对话。利用各种技术实现随时随地无缝的人工智能对话。
27. ParisNeo/lollms-webui - 大型语言模型之主的网络用户界面。
28. zhayujie/bot-on-anything - 基于大型模型的聊天机器人构建器能够迅速将ChatGPT、Claude和Gemini等人工智能模型集成到Telegram、Gmail、Slack等软件应用程序和网站中。
29. deanxv/coze-discord-proxy - 通过Coze - Bot代理Discord对话，经由API请求GPT4模型，提供对话、文生图、图生文和知识库检索等功能。
30. vocodedev/vocode-python - 构建基于语音的、模块化且开源的大型语言模型（LLM）智能体。

31. alexrudall/ruby-openai - OpenAI API与Ruby。
32. ahmadbilaldev/langui - 人工智能用户界面。用于GPT、生成式人工智能和大型语言模型(LLM)项目的开源Tailwind组件。
33. ollama/ollama-js - Ollama JavaScript库。
34. xusenlinzy/api-for-open-llm - 用于开放大型语言模型的OpenAI风格的API。支持各种模型，如LLaMA、ChatGLM等。
35. anse-app/anse - ChatGPT、DALL - E和Stable Diffusion模型的超强体验。
36. mylxsw/aidea-server - AIdEA是一款多功能一体的APP，支持GPT、国内大型语言模型（如通义千问和文心一言），以及用于文生图、图生图、SDXL1.0、超分辨率和图像上色的Stable Diffusion。
37. aallam/openai-kotlin - 支持多平台和协程的Kotlin OpenAI API客户端。
38. guinmoon/LLMFarm - 在iOS和MacOS上离线使用适用于Llama和其他大型语言模型的GGML库。
39. uezo/ChatdollKit - ChatdollKit可让你将自己的3D模型转化为聊天机器人。

智能体开发框架

1. langchain-ai/langchain - 构建具有情境感知推理能力的应用程序。
2. nomic-ai/gpt4all - GPT4All能够在任何设备上运行本地大型语言模型（LLMs）。它是开源的，可用于商业用途。
3. comfyanonymous/ComfyUI - 最强大且模块化的扩散模型具有用于图形用户界面（GUI）、应用程序接口（API）和后端的图/节点接口。
4. langgenius/dify - Dify是一个开源的大型语言模型（LLM）应用开发平台，拥有直观的界面，具备多种功能，可实现从快速制作原型到投入生产的过程。
5. lobehub/lobe-chat - Lobe Chat是一个具有现代设计的开源人工智能聊天框架。它支持多个人工智能供应商、知识库和多模态，并能一键免费部署私人聊天应用程序。
6. logspace-ai/langflow - Langflow是一个基于Python的、与模型无关的低代码应用构建器，用于RAG（检索增强生成）和多智能体AI应用程序。它可以与任何API或数据库协同工作。
7. run-llama/llama_index - LlamaIndex是一个用于大型语言模型（LLM）应用的数据框架。
8. FlowiseAI/Flowise - 使用拖放式用户界面创建您的个性化大型语言模型（LLM）流程。
9. chatchat-space/Langchain-Chatchat - Langchain - Chatchat（最初名为Langchain - ChatGLM）是一个基于Langchain、ChatGLM、Qwen、Llama等的检索增强生成（RAG）和代理（Agent）应用程序，用于基于本地知识的大型语言模型（LLM）。
10. go-skynet/LocalAI - 一个可替代OpenAI和Claude等服务的开源项目。它可以在消费级硬件上运行，并执行诸如生成不同媒体类型等各种任务。
11. infiniflow/ragflow - RAGFlow是一个用于深度文档理解的开源RAG（检索增强生成）引擎。

12. mindsdb/mindsdb - AGI的查询引擎是一个构建人工智能的平台，该人工智能能够在联邦数据上进行学习和回答问题。
13. embedchain/embedchain - 你的人工智能应用的存储层。
14. songquanpeng/one-api - 这是一个OpenAI密钥管理与再分配系统。它支持多种大型语言模型（LLMs），拥有英文用户界面（UI），可单文件执行，并且有Docker镜像以便于部署。
15. Cinnamon/kotaemon - 一个基于检索增强生成（RAG）技术、可用于与文档聊天的开源工具。
16. labring/FastGPT - FastGPT是一个基于大型语言模型（LLMs）的知识平台，提供多种功能，可轻松开发和部署问答系统。
17. deepset-ai/haystack - 一个用于构建大型语言模型（LLM）应用的人工智能编排框架，适用于像检索增强生成（RAG）以及带有高级检索方法的聊天机器人之类的任务。
18. BerriAI/litellm - Python SDK和代理服务器（LLM网关）能够调用100多个OpenAI格式的大型语言模型（LLM）API，包括Bedrock、Azure等的API。
19. flairNLP/flair - 一个用于高级自然语言处理的非常基础的框架。
20. langchain-ai/langchainjs - 构建具有情境感知能力的推理应用程序。
21. xenova/transformers.js - 用于网络的最先进的机器学习技术允许在没有服务器的浏览器中运行🤗 Transformers模型。
22. netease-youdao/QAnything - 基于任何事物的问答。
23. h2oai/h2ogpt - 与本地GPT的私人聊天，支持文档、图像、视频等各种内容。它是100%私密的，基于Apache 2.0协议，支持oLLaMa、Mixtral、llama.cpp等，在给定链接中有示例。
24. pathwaycom/llm-app - 适用于检索增强生成（RAG）、人工智能管道（AI pipelines）和企业搜索的即用型云模板，可处理实时数据，对Docker友好并能与各种数据源同步。
25. ludwig-ai/ludwig - 用于创建像大型语言模型（LLMs）和神经网络这样的定制人工智能模型的低代码框架。
26. vercel/ai - 使用React、Svelte、Vue和Solid构建人工智能驱动的应用程序。
27. microsoft/promptflow - 通过原型制作、测试、生产部署和监控来构建高质量的大型语言模型（LLM）应用程序。
28. Unstructured-IO/unstructured - 用于创建机器学习中自定义预处理管道（如标记、训练或生产任务）的开源库和API。
29. dataelement/bisheng - 必升（BISHENG）是一个面向企业人工智能应用的开放大语言模型（LLM）运维平台。它具备生成式人工智能（GenAI）工作流、检索增强生成（RAG）等功能。
30. togethercomputer/OpenChatKit -
31. llmware-ai/llmware - 一个使用小型专用模型创建企业级检索增强生成（RAG）管道的统一框架。
32. leptonai/search_with_lepton - 使用Lepton AI快速创建一个基于对话的搜索演示。

33. Deeptrain-Community/chatnio - 下一代人工智能一站式B/C端解决方案，支持多种模型和各类功能。
34. Chainlit/chainlit - 在数分钟内快速构建对话式人工智能。
35. modelscope/modelscope - ModelScope将模型即服务（Model - as - a - Service）的概念变为现实。
36. deeppavlov/DeepPavlov - 一个用于深度学习端到端对话系统和聊天机器人的开源库。
37. langchain-ai/opengpts -
38. TaskingAI/TaskingAI - 一个用于开发原生人工智能应用程序的开源平台。
39. wenda-LLM/wenda - 文达（Wenda）是一个大型语言模型（LLM）调用平台，旨在特定环境中高效生成内容，同时考虑到个人和中小企业计算资源的限制以及知识安全和隐私问题。
40. rustformers/llm - 一个用于处理大型语言模型的未经维护的Rust库生态系统。详见自述文件。
41. josStorer/RWKV-Runner - 一个8MB的全自动RWKV管理和启动工具，带有与OpenAI API兼容的接口。RWKV是一个完全开源且可用于商业用途的大型语言模型。
42. langchain4j/langchain4j - LangChain的Java版本。
43. OpenBMB/ToolBench - 一个用于工具学习的大型语言模型训练、服务和评估的开放平台（ICLR'24焦点论文）。
44. microsoft/FLAML - 一个用于自动机器学习（AutoML）和调参的快速库。还有一个可加入的Discord（一款聊天软件）链接。
45. microsoft/lmops - 用于通过大型语言模型（LLMs）和多模态大型语言模型（MLLMs）实现人工智能能力的通用技术。
46. llm-workflow-engine/llm-workflow-engine - 适用于大型语言模型（核心包）的Power CLI和工作流管理器。
47. timescale/pgai - 一组用于更轻松地开发使用PostgreSQL的检索增强生成（RAG）、语义搜索和其他人工智能应用程序的工具。
48. FreedomIntelligence/LLMZoo - LLM Zoo是一个为大型语言模型提供数据、模型和评估基准的项目。
49. casibase/casibase - AI Cloud是一个类似于LangChain的开源检索增强生成（RAG）知识库。它支持多种模型，并拥有聊天机器人和管理用户界面（UI）演示。
50. getzep/zep - Zep：你的人工智能堆栈的内存基础。
51. leptonai/leptonai - 一个用于简化人工智能服务构建的Python框架。
52. pezzolabs/pezzo - 一个开源的、以开发者为先的LLMOps平台，用于简化提示设计和版本管理等各个方面的工作。
53. cheshire-cat-ai/core - 人工智能代理微服务。

54. aurelio-labs/semantic-router - 用于多模态数据决策和智能处理的超高速人工智能。
55. instill-ai/vdp - Instill Core是一款用于数据、模型和管道编排的全栈人工智能基础设施工具，它简化了构建多种以人工智能为先的应用程序的过程。
56. intel/intel-extension-for-transformers - 使用最先进的压缩技术，在您的设备上快速构建您的聊天机器人，并在英特尔平台上高效运行大型语言模型。
57. griptape-ai/griptape - 一个用于人工智能代理和工作流程的模块化Python框架，具有思维链推理、工具和记忆功能。
58. run-llama/LlamaIndexTS - 用于大型语言模型（LLM）应用的数据框架，重点关注服务器端解决方案。
59. Agenta-AI/agenta - 一个集成了提示词游乐场、提示词管理、大型语言模型（LLM）评估和大型语言模型可观测性的开源LLMOps平台。
60. marella/ctransformers - 通过GGML库为C/C++中的Transformer模型提供的Python绑定。
61. devflowinc/trieve - 一个基于API的集搜索、推荐、检索增强生成（RAG）和分析于一体的基础设施。
62. YangLing0818/RPG-DiffusionMaster - [ICML 2024] 通过多模态大型语言模型（LLMs）的重新字幕、规划和生成来掌握文本到图像扩散的角色扮演游戏（RPG）。
63. trypromptly/LLMStack - 一个无代码的多代理框架，用于使用您的数据构建大型语言模型（LLM）代理、工作流和应用程序。
64. getzep/graphiti - 构建和查询具有时间感知能力的动态知识图谱。
65. KimMeen/Time-LLM - ICLR 2024中《Time - LLM：通过重新编程大型语言模型进行时间序列预测》的官方实现。
66. floneum/floneum - 即时、可控且在本地预训练的Rust语言中的人工智能模型。
67. jina-ai/langchain-serve - 使用Jina和FastAPI进行生产的Langchain应用程序。
68. SqueezeAILab/LLMCompiler - 在2024年国际机器学习会议（ICML）上提出的LLM编译器（LLMCompiler）是一种用于并行函数调用的大型语言模型（LLM）编译器。
69. andreibondarev/langchainrb - 使用Ruby构建由大型语言模型（LLM）提供支持的应用程序。
70. psychic-api/rag-stack - 在虚拟专用云（VPC）中部署一个私有版的ChatGPT替代方案，连接到组织的知识库，并支持开源大型语言模型（LLMs）。
71. DAGWorks-Inc/burr - 构建用于决策的应用程序，如聊天机器人等，并在自己的基础设施上进行管理。
72. IntelLabs/fastRAG - 高效检索增强与生成框架。
73. sobelio/llm-chain - “llm - chain”是一个强大的Rust crate（板条箱，可理解为代码库），用于构建大型语言模型中的链，实现文本摘要和复杂任务的完成。
74. microsoft/windows-ai-studio -

75. vercel/modelfusion - 一个用于创建人工智能应用程序的TypeScript库。
76. axflow/axflow - 一个用于人工智能开发的TypeScript框架。
77. gabrielchua/RAGxplorer - 一个用于可视化你的检索增强生成（RAG）的开源工具。
78. parthsarthi03/raptor - 通过递归抽象处理进行树状组织检索的RAPTOR（猛禽）官方实现。
79. google/generative-ai-swift - 用于Google Gemini API的官方Swift库。
80. pinecone-io/canopy - 由Pinecone驱动的检索增强生成（RAG）框架和上下文引擎。
81. safevideo/autollm - 在数秒内推出基于检索增强生成（RAG）的大型语言模型（LLM）网络应用。

模型

1. openai/whisper - 通过大规模弱监督实现稳健的语音识别。
2. CompVis/stable-diffusion - 一个潜在的文本到图像的扩散模型。
3. facebookresearch/llama - 用于Llama模型的推理代码。
4. xai-org/grok-1 - Grok的公开版本发布。
5. Stability-AI/stablediffusion - 使用潜在扩散模型进行高分辨率图像合成。
6. karpathy/nanoGPT - 用于训练/微调中型GPT的最简单、最快速的库。
7. TencentARC/GFPGAN - GFPGAN专注于创建适用于现实场景中人脸修复的实用算法。
8. lllyasviel/ControlNet -
9. tatsu-lab/stanford_alpaca - 用于训练斯坦福羊驼（Alpaca）模型和生成数据的代码与文档。
10. meta-llama/llama3 - 官方Meta Llama 3 GitHub网站。
11. Stability-AI/generative-models - Stability AI的生成式模型
12. lucidrains/vit-pytorch - 在PyTorch中实现视觉变换器（Vision Transformer），仅使用一个变换器编码器在视觉分类任务中达到最先进水平（SOTA）。
13. apple/ml-stable-diffusion - 在苹果硅芯片（Apple Silicon）上使用Core ML的Stable Diffusion。
14. facebookresearch/codellama - CodeLlama模型的推理代码。
15. QwenLM/Qwen - 通义千问（Qwen）的官方代码库，通义千问是阿里云提出的一个聊天和预训练大型语言模型。
16. AI4Finance-Foundation/FinGPT - FinGPT - 开源金融大型语言模型。在HuggingFace发布的训练模型。
17. state-spaces/mamba - 曼巴SSM架构
18. BlinkDL/RWKV-LM - RWKV是一种在大型语言模型（LLM）方面表现良好的循环神经网络（RNN），可以像GPT变换器那样进行训练。它具有性能优异、线性时间等特点。

19. CompVis/latent-diffusion - 使用潜在扩散模型进行高分辨率图像合成。
20. QwenLM/Qwen1.5 - 通义千问2.5是阿里云通义千问团队开发的大型语言模型系列。
21. lucidrains/DALLE2-pytorch - 在PyTorch中实现OpenAI更新的用于文图合成的神经网络DALL - E 2。
22. NVIDIA/Megatron-LM - 继续进行对Transformer模型大规模训练的研究。
23. guoyww/AnimateDiff - AnimateDiff的官方实现。
24. databrickslabs/dolly - Databricks的Dolly是一个在Databricks机器学习平台上训练的大型语言模型。
25. mlfoundations/open_clip - 一个CLIP（对比语言-图像预训练）的开源实现。
26. THUDM/CogVideo - 文本和图像到视频的生成：CogVideoX（2024年）和CogVideo（2023年国际表征学习会议）
27. AIGC-Audio/AudioGPT - AudioGPT与理解和生成语音、音乐、声音以及说话头像相关。
28. nlpxucan/WizardLM - 大型语言模型（LLMs）建立在Evol Insturct（Evol指令）之上：WizardLM（向导语言模型）、WizardCoder（向导编码器）、WizardMath（向导数学）。
29. lucidrains/denoising-diffusion-pytorch - 在Pytorch中实现去噪扩散概率模型。
30. THUDM/CodeGeeX - CodeGeeX是一个开源的多语言代码生成模型（KDD 2023）。
31. Vaibhavs10/insanely-fast-whisper -
32. 01-ai/Yi - 由01 - ai开发者从头开始开发的一系列大型语言模型。
33. lucidrains/PaLM-rlhf-pytorch - 在PaLM上实施类似于ChatGPT的人类反馈强化学习（RLHF）。
34. HumanAIGC/EMO - 在弱条件下使用音视频扩散模型（Audio2Video Diffusion Model）生成富有表现力的人像视频：生动的表情人像。
35. alembics/disco-diffusion - 没有提供可翻译的描述。
36. openlm-research/open_llama - OpenLLaMA是Meta AI的LLaMA 7B的开源复制品，具有宽松的许可协议，并且在RedPajama数据集上进行训练。
37. OpenBMB/MiniCPM - MiniCPM3 - 4B，一个边缘端的大型语言模型（LLM），性能优于GPT - 3.5 - Turbo。
38. LargeWorldModel/LWM - 用于文本和视频建模的大型世界模型，拥有数百万的大语境。
39. LiheYoung/Depth-Anything - 《深度万物：释放大规模无标注数据的力量》，一个用于2024年计算机视觉与模式识别会议（CVPR）中基于单目图像的深度估计基础模型。
40. openai/point-e - 点云扩散用于3D模型合成。
41. google-research/text-to-text-transfer-transformer - 论文《用统一的文本到文本转换器探索迁移学习的极限》的代码

42. Lightning-AI/lit-llama - 基于nanoGPT的LLaMA语言模型实现支持多种特性，如快速注意力机制、量化、微调以及预训练，并且采用Apache 2.0许可协议。
43. OpenGVLab/LLaMA-Adapter - 在2024年国际学习表征会议（ICLR）上，使用120万个参数在1小时内对LLaMA进行微调以遵循指令。
44. NVIDIA/DALI - 一个GPU（图形处理器）加速库拥有高度优化的构建模块和用于数据处理的执行引擎，以加速深度学习训练和推理应用程序。
45. allenai/OLMo - 用于建模、训练、评估和推理的OLMo代码。
46. salesforce/CodeGen - CodeGen是一个用于程序合成的开源模型系列，在TPU - v4上进行训练，可与OpenAI Codex相媲美。
47. lucidrains/x-transformers - 一个专注度高、简洁而完整的变换器，具备来自多篇论文的实验性特征。
48. SCIR-HI/Huatuo-Llama-Med-Chinese - 本曹（原名华佗）的代码库，这是一个用中国医学知识对大型语言模型进行指令微调的模型库。
49. luosiallen/latent-consistency-model - 潜在一致性模型：通过少步推理合成高分辨率图像。
50. microsoft/BioGPT -
51. google-research/simclr - SimCLRv2：大型自监督模型是强大的半监督学习器。
52. llSourcell/Doctor-Dignity - 尊严博士（Doctor Dignity）是一个能够通过美国医师执照考试（USMLE）、可离线使用、跨平台且能保护健康数据隐私的大型语言模型（LLM）。
53. google-research/multinerf - 用于Mip - NeRF 360、Ref - NeRF和RawNeRF的代码发布。
54. jaymody/picoGPT - 一个用NumPy实现的非常小的GPT - 2版本。
55. google-research/albert - ALBERT是一种精简版的BERT（Bidirectional Encoder Representations from Transformers，双向编码器表征转换模型），用于语言表征的自监督学习。
56. project-baize/baize-chatbot - 仅用一个GPU在数小时内使用ChatGPT来训练你的聊天机器人。
57. salesforce/CodeT5 - CodeT5旨在为开放代码的大型语言模型（LLM）提供代码理解和生成方面的支持。
58. facebookresearch/jepa - 通过视觉 - 基于联合嵌入预测架构（V - JEPA）从视频进行自监督学习的PyTorch代码和模型。
59. paperswithcode/galai - GALACTICA的模型应用程序接口。
60. dvlab-research/LongLoRA - LongLoRA和LongAlpaca（ICLR 2024口头报告）的代码和文档。
61. baaivision/Painter - 画家与SegGPT系列：来自北京智源人工智能研究院（BAAI）的视觉基础模型。
62. databricks/dbrx - 用于Databricks的大型语言模型DBRX的代码示例和资源。

63. state-spaces/s4 - 结构化状态空间序列模型。
64. google-research/electra - ELECTRA将文本编码器预训练为鉴别器而不是生成器。
65. EleutherAI/pythia - EleutherAI在可解释性和学习动态方面的研究中心。
66. ise-uiuc/magicoder - Magicoder（ICML'24）通过开源指令（OSS - Instruct）实现代码生成。
67. epfLLM/meditron - Meditron是一套开源的医疗大型语言模型。
68. MetaGLM/FinGLM - FinGLM旨在构建一个开放、公益且持久的金融大模型项目，通过开源推动“AI+金融”发展。
69. deepseek-ai/DeepSeek-LLM - 深度求索（DeepSeek）大语言模型（LLM）：将会有答案。
70. allenai/scispacy - 适用于科学/生物医学文档的完整spaCy管道和模型。
71. apple/ml-4m - 4M：大规模多模态掩码建模。
72. google-research/language - 谷歌人工智能语言团队开源项目的共享库。
73. google/maxtext - 一个简单、高性能且可扩展的Jax大型语言模型（LLM）。
74. netease-youdao/BCEmbedding - 网易有道用于检索增强生成（RAG）产品（嵌入和重排序器）的开源模型。
75. SHI-Labs/OneFormer - CVPR 2023中的OneFormer是一种用于通用图像分割的Transformer。
76. google-research/FLAN -
77. lxtGH/OMG-Seg - OMG - LLaVA和OMG - Seg代码库与CVPR - 24（计算机视觉与模式识别会议 - 2024）和NeurIPS - 24（神经信息处理系统大会 - 2024）相关。
78. SHI-Labs/Versatile-Diffusion - 通用扩散（Versatile Diffusion）：一个融合文本、图像和变体的扩散模型，于2022年发表于预印本平台arXiv，2023年在国际计算机视觉大会（ICCV）上展示。
79. time-series-foundation-models/lag-llama - Lag - Llama：概率性时间序列预测基础模型的方法
80. openai/lm-human-preferences - 人类偏好微调语言模型论文的代码。
81. IBM/Dromedary - 单峰驼（Dromedary）旨在成为有用、合乎道德且可靠的大型语言模型。
82. dauparas/ProteinMPNN - 名为ProteinMPNN的论文代码。
83. SHI-Labs/Neighborhood-Attention-Transformer - 2022年发表于arxiv以及2023年发表于CVPR的邻域注意力变换器。此外，2022年发表于arxiv的空洞邻域注意力变换器。
84. THUDM/SwissArmyTransformer - SwissArmyTransformer是一个灵活且强大的用于开发Transformer变体的库。
85. ctllll/LLM-ToolMaker -
86. Xwin-LM/Xwin-LM - Xwin - LM：一个强大、稳定且可复现的大型语言模型对齐。
87. microsoft/ToRA - ToRA（用于ICLR'24）是一系列集成工具以解决困难数学推理问题的大型语言模型智能体。

88. SalesforceAIResearch/uni2ts - 通用时间序列预测变换器被统一训练。
89. replit/ReplitLM - ReplitLM模型系列的推理代码和配置。
90. HazyResearch/safari - 序列建模背景下的卷积。

AI列表

1. fighting41love/funNLP -
2. linexjlin/GPTs - 与GPT相关的泄露提示。
3. e2b-dev/awesome-ai-agents - 一系列人工智能自主代理。
4. eugeneyan/open-llms - 可供商业使用的开放大语言模型（LLM）列表。
5. Shubhamsaboo/awesome-llm-apps - 一组很棒的带有检索增强生成（RAG）功能的大型语言模型（LLM）应用程序，它们使用OpenAI、Anthropic、Gemini和开源模型。
6. RUCAIBox/LLMSurvey - 大型语言模型综述（A Survey of Large Language Models）这篇调查论文的官方GitHub页面。
7. WooooDyy/LLM-Agent-Paper-List - 席之恒（音译）等人所著的86页论文《基于大型语言模型的智能体的兴起与潜力：综述》的论文列表
8. steven2358/awesome-generative-ai - 当代生成式人工智能项目和服务列表。
9. wgwang/LLMs-In-China - 中国大型模型。
10. lonePatient/awesome-pretrained-chinese-nlp-models - 一系列高质量的中文预训练模型、大型模型、多模态模型和大型语言模型。
11. tensorchord/Awesome-LLMOps - 为开发者精选的优秀LLMOps工具大清单。
12. opendilab/awesome-RLHF - 一份持续更新的基于人类反馈的强化学习资源清单。
13. DSXiangLi/DecryptPrompt - 总结提示（Prompt）与大型语言模型（LLM）相关论文、开源数据与模型，以及人工智能生成内容（AIGC）的应用。
14. FreedomIntelligence/Medical_NLP - 医学自然语言处理竞赛、数据集、大型模型与论文。
15. archinetai/audio-ai-timeline - 一个从2023年开始的最新音频生成人工智能模型的时间表。
16. chiphuyen/aie-book - 人工智能工程师的资源以及《人工智能工程》（奇普·休恩著，2025年）的辅助材料。
17. EgoAlpha/prompt-in-context-learning - 用于情境学习和提示工程的优质资源。掌握具有最新更新的大型语言模型，如ChatGPT、GPT - 3和FlanT5。
18. taranjeet/awesome-gpts - 社区制作的所有GPT（生成式预训练转换器）的集合。
19. cfahlgren1/natural-sql - 一系列高性能的文本到SQL的大型语言模型。
20. yokoffing/ChatGPT-Prompts - ChatGPT和必应AI的提示词管理。

推理优化

1. ggerganov/llama.cpp - C/C++中的大型语言模型（LLM）推理。
2. ggerganov/whisper.cpp - 一个用C/C++编写的OpenAI的Whisper模型的移植版本。
3. karpathy/llm.c - 使用简单的原始C/CUDA训练大型语言模型（LLM）。
4. Mozilla-Ocho/llamafile - 仅使用一个文件来分发和运行大型语言模型（LLMs）。
5. unslothai/unsloth - 在使用少70%内存的同时，将Llama 3.3、Mistral、Phi、Qwen 2.5和Gemma大型语言模型（LLMs）的微调速度提高2 - 5倍。
6. mlc-ai/mlc-llm - 带有机器学习（ML）编译功能的通用大型语言模型（LLM）部署引擎。
7. karpathy/llama2.c - 在单个纯C文件中推断Llama 2。
8. Dao-AILab/flash-attention - 快速且高效的精确注意力机制，既快速又节省内存。
9. openai/triton - 特里同（Triton）语言和编译器的开发库。
10. microsoft/BitNet - 1位大型语言模型（LLMs）的官方推理框架。
11. ggerganov/ggml - 一个用于机器学习的张量库。
12. NVIDIA/TensorRT - NVIDIA TensorRT是一个用于在NVIDIA GPU上进行高性能深度学习推理的软件开发工具包（SDK）。这个代码仓库有其开源组件。
13. bigscience-workshop/petals - 以类似BitTorrent的方式在家运行大型语言模型（LLMs），微调与推理速度比卸载（offloading）快多达10倍。
14. NVIDIA/TensorRT-LLM - TensorRT - LLM提供了一个易于使用的Python API，用于定义大型语言模型（LLMs）并构建优化的TensorRT引擎以实现高效的GPU推理，并且具有用于创建Python和C++运行时来执行这些引擎的组件。
15. intel-analytics/BigDL - 在英特尔XPU上加速本地大型语言模型（LLM）推理和微调，并与各种相关框架集成。
16. intel-analytics/ipex-llm - 在英特尔XPU（英特尔架构的加速处理器）上加速本地大型语言模型（LLM）推理和微调，并与多种工具集成。
17. TimDettmers/bitsandbytes - 通过PyTorch的k位量化实现可访问的大型语言模型。
18. google/gemma.cpp - 一个用于谷歌Gemma模型的轻量级独立C++推理引擎。
19. NVIDIA/cutlass - 用于线性代数子例程的CUDA模板。
20. pytorch-labs/gpt-fast - 用不到1000行Python代码实现用于文本生成的简单高效的原生PyTorch变压器。
21. PanQiWei/AutoGPTQ - 一个基于GPTQ算法、易于使用且具有用户友好型接口的大型语言模型量化包。

22. turboderp/exllamav2 - 一个用于在常见消费级GPU上本地运行大型语言模型（LLM）的快速推理库。
23. OpenNMT/CTranslate2 - 快速变压器模型推理引擎。
24. ztxz16/fastllm - 一个纯C++的全平台大型语言模型（LLM）加速库，支持Python调用。它能使单卡ChatGLM - 6B级别的模型达到每秒超过10,000个词元（token），支持GLM、Llama、Moss基础模型，并且在移动设备上能流畅运行。
25. qwopqwop200/GPTQ-for-LLaMa - 使用GPTQ将LLaMA量化为4位。
26. VainF/Torch-Pruning - [CVPR 2023] DepGraph：面向任意结构剪枝。
27. turboderp/exllama - 用于量化权重的HF Transformers版Llama重写版本，其内存效率更高。
28. lucidrains/vector-quantize-pytorch - PyTorch中的向量（和标量）量化
29. mit-han-lab/llm-awq - AWQ：用于大型语言模型压缩与加速的激活感知权重量化（方法）荣获2024年MLSys最佳论文奖。
30. Jittor/JittorLLMs - Jittor模型推理库具有高性能、低配置要求、良好的中文支持和可移植性等特点。
31. FasterDecoding/Medusa - 美杜莎：一个通过多个解码头加速大型语言模型（LLM）生成的简单框架。
32. intel/neural-compressor - SOTA低比特LLM量化（包括INT8/FP8/INT4/FP4/NF4）和稀疏性是用于TensorFlow、PyTorch和ONNX运行时的领先的模型压缩技术。
33. neuralmagic/sparseml - 便于将稀疏化轻松应用于神经网络的库，从而得到更快且更小的模型。
34. IST-DASLab/gptq - 2023年国际学习表征会议（ICLR）上关于生成式预训练变压器（GPT）精确的训练后量化（post - training quantization）的论文代码，名为“GPTQ”。
35. HazyResearch/ThunderKittens - 用于快速内核的图块基元。
36. uTensor/uTensor - 一个微型机器学习人工智能推理库。
37. pytorch-labs/ao - 用于训练和推理的PyTorch原生量化与稀疏性。
38. saharNooby/rwkv.cpp - 用于RWKV语言模型的CPU上的INT4/INT5/INT8和FP16推理。
39. mit-han-lab/smoothquant - SmoothQuant：大型语言模型准确且高效的训练后量化。
40. Lightning-AI/lightning-thunder - Thunder是一个PyTorch源到源编译器，它可以使模型速度提高达40%，并在多个GPU上使用不同的硬件执行器。
41. pytorch-labs/segment-anything-fast - 一个用于批量离线推理的Segment - Anything版本。
42. Vahe1994/AQLM - 官方PyTorch库，包含两篇关于大型语言模型极限压缩的论文：一篇是通过加法量化（https://arxiv.org/pdf/2401.06118.pdf），另一篇是PV - 微调（https://arxiv.org/abs/2405.14852）。

43. hao-ai-lab/LookaheadDecoding - 使用前瞻解码来打破大型语言模型（LLM）推理中的顺序依赖关系（ICML 2024）。
44. horseee/LLM-Pruner - [NeurIPS 2023] 大型语言模型（如Llama - 3/3.1、Llama - 2、LLaMA等）结构剪枝的LLM - Pruner。
45. kuleshov/minillm - MiniLLM是一种以最小化方式在消费级GPU上运行现代大型语言模型（LLM）的系统。

信息聚合

1. binary-husky/gpt_academic - 为GPT/GLM等大型语言模型（LLM）提供实用的交互界面，尤其优化论文阅读、润色和写作体验。支持多种功能并集成多个模型。
2. imartinez/privateGPT - 使用GPT私下与文档交互，无数据泄露。
3. Mintplex-Labs/anything-llm - 桌面与Docker AI应用是一体化的，内置了检索增强生成（RAG）和AI代理。
4. khoj-ai/khoj - 它是一个可以自托管的人工智能第二大脑，能从各种来源获取答案、构建自定义代理、安排自动化任务以及进行研究，并且能够免费将大型语言模型转化为个人人工智能。
5. PromtEngineer/localGPT - 通过GPT模型私下与本地文档聊天，数据不会离开设备。
6. kaixindelele/ChatPaper - 使用ChatGPT对科研进行全流程加速，包括总结arXiv论文、专业翻译、润色、同行评审和回应同行评审。
7. assafelovic/gpt-researcher - 基于大型语言模型（LLM）的自主代理对任何主题进行本地和网络研究，并创建一份带有引用的综合报告。
8. arc53/DocsGPT - 文档聊天机器人能够与数据聊天，可私人部署，并将知识集成到人工智能工作流程中以进行共享。
9. mayooear/gpt4-pdf-chatbot-langchain - 用于大型PDF文档的GPT4和LangChain聊天机器人。
10. danswer-ai/danswer - Gen - AI Chat for Teams就像ChatGPT一样，但可以获取团队的特殊知识。
11. josStorer/chatGPTBox - 将ChatGPT深度集成到你的浏览器中。你所需的一切都在这里。
12. facebookresearch/nougat - 牛轧糖（Nougat）在学术文档神经光学理解方面的应用。
13. bhaskatripathi/pdfGPT - PDF GPT能够通过GPT功能与PDF内容进行交互，是一种将PDF转变为聊天机器人的有效开源解决方案。
14. whitead/paper-qa - 用于回答基于科学文献且带有引用的问题的高精度检索增强生成（RAG）技术。
15. weaviate/Verba - 一个检索增强生成（RAG）聊天机器人由Weaviate提供动力。
16. run-llama/rags - 使用你的数据构建ChatGPT，全部采用自然语言。
17. MuiseDestiny/zotero-gpt - GPT与Zotero相遇。

18. madawei2699/myGPTReader - 一种使用ChatGPT进行阅读和聊天的、社区驱动的与人工智能机器人交互的方法。
19. swirlai/swirl-search - 人工智能搜索和检索增强生成（AI Search & RAG）能够在确保数据安全和快速部署的同时，从众多应用程序中的公司知识中获取即时答案。
20. dvorka/mindforger - 一个思考笔记和一个Markdown编辑器。
21. kha-white/manga-ocr - 主要用于日漫的日文字符光学字符识别。
22. nlmatics/llmsherpa - 用于加速大型语言模型（LLM）项目的开发者应用程序接口（API）。
23. ucbepic/docetl - 一个由自主语言模型（LLM）驱动的数据处理和ETL（抽取、转换、加载）系统。
24. KnowledgeCanvas/knowledge - 知识是用于网站、文档和文件的各种操作（如保存、搜索等）的工具。
25. rotemweiss57/gpt-newspaper - 一个基于GPT的自主代理程序，能根据用户偏好创建个性化报纸。
26. nlmatics/nlm-ingestor - 此存储库提供用于llmsherpa API连接的服务器端代码以及用于不同文件格式的解析器。
27. kha-white/mokuro - 在浏览器中阅读可选择文字的日本漫画。
28. BruceMacD/chatd - 通过本地人工智能与你的文档进行聊天。
29. akshata29/entaoai - 使用自己的数据进行聊天和提问。快速上传企业数据，以便使用OpenAI服务对上传的数据进行聊天和提问。

代码助手

1. abi/screenshot-to-code - 插入一张屏幕截图并将其转换为简洁的代码（HTML/Tailwind/React/Vue）。
2. gpt-engineer-org/gpt-engineer - 一个用于体验人工智能软件工程师的基于终端的平台，与https://gptengineer.app不同。
3. OpenDevin/OpenDevin - OpenHands：用更少的代码实现更多功能。
4. Pythagora-io/gpt-pilot - 第一位真正成为人工智能领域开发者的人。
5. getcursor/cursor - 人工智能代码编辑器。
6. OpenBMB/ChatDev - 通过具有大型语言模型（LLM）支持的多智能体协作，依据自然语言思路创建定制软件。
7. paul-gauthier/aider - Aider是终端中的人工智能结对编程。
8. TabbyML/tabby - 一个自托管的人工智能编码助手。
9. continuedev/continue - Continue是一个开源的人工智能代码助手。它可以连接到模型和上下文，以便在VS Code和JetBrains中进行自定义自动补全和聊天。

10. stitionai/devika - 迪维卡（Devika）是一个智能体人工智能软件工程师，能够理解人类指令，分解指令，进行研究并编写代码。它的目标是成为认知人工智能（Cognition AI）公司开发的德文（Devin）的开源替代品，并且没有官方网站。
11. emilwallner/Screenshot-to-code - 一个用于将设计模型转换为静态网站的神经网络。
12. fauxpilot/fauxpilot - FauxPilot是GitHub Copilot服务器的一个开源替代品。
13. eosphoros-ai/DB-GPT - 带有AWEL（智能体工作流表达式语言）和智能体的原生人工智能数据应用开发框架。
14. princeton-nlp/SWE-agent - SWE - 代理使用GPT - 4或其他语言模型自动修复GitHub问题，也可用于攻击性网络安全或竞争性编程挑战。[NeurIPS 2024]
15. Sinaptik-AI/pandas-ai - 与各种数据库（SQL、CSV等）进行交互，并使用PandasAI通过大型语言模型（LLMs）和检索增强生成（RAG）进行对话式数据分析。
16. vanna-ai/vanna - 与你的SQL数据库进行交互。使用基于检索的生成（RAG）技术，通过大型语言模型（LLM）生成准确的文本到SQL语句。
17. ShishirPatil/gorilla - 大猩猩：用于函数调用（工具调用）的大型语言模型的训练与评估。
18. codota/TabNine - 人工智能代码补全是指人工智能系统为程序员提供建议或补全代码段的功能。
19. TheR1D/shell_gpt - 一款由GPT - 4等人工智能大语言模型驱动的命令行生产力工具，有助于更快速高效地完成任务。
20. Nutlope/aicommits - 一个利用人工智能为你编写git提交信息的命令行界面（CLI）。
21. GreyDGL/PentestGPT - 一款由GPT赋能的渗透测试工具。
22. joshpxyne/gpt-migrate - 轻松地在框架或语言之间迁移您的代码库。
23. kuafuai/DevOpsGPT - 一个由人工智能驱动的软件开发多智能体系统将大型语言模型（LLM）与DevOps工具相结合，将自然语言需求转化为可运行的软件，支持任何开发语言并扩展现有代码。
24. di-sukharev/opencommit - 用于Git的GPT封装器能够使用大型语言模型（LLM）在1秒内生成提交消息，与Claude 3.5配合良好，并支持本地模型。
25. sqlchat/sqlchat - 一款基于聊天功能的、面向未来十年的SQL客户端与编辑器。
26. Exafunction/codeium.vim - 一个用于Vim和Neovim的免费且超快的Copilot替代方案。
27. varunshenoy/GraphGPT - 使用GPT - 3从非结构化文本推断知识图谱。
28. Nutlope/llamacoder - 克劳德（Claude）制作的开源制品，使用Llama 3.1 405B构建。
29. mckaywrigley/ai-code-translator - 利用人工智能在不同语言之间翻译代码。
30. shobrook/adrenaline - 与代码库进行交互并对其可视化。
31. QwenLM/Qwen2.5-Coder - 通义千问2.5 - Coder是通义千问2.5的代码版本，通义千问2.5是阿里云通义千问团队开发的大型语言模型系列。
32. ricklamers/gpt-code-ui - OpenAI公司的ChatGPT代码解释器的一个开源实现。

33. gofireflyio/aiac - 人工智能基础设施代码生成器。
34. defog-ai/sqlcoder - 用于将自然语言问题转换为SQL查询的最先进的语言模型。
35. gptscript-ai/gptscript - 构建用于与您的系统进行交互的人工智能助手。
36. RootbeerComputer/backend-GPT -
37. mpoon/gpt-repository-loader - 将代码库转换为对大型语言模型（LLM）提示友好的格式，该格式主要由GPT - 4创建。
38. Canner/WrenAI - 一个开源人工智能代理使数据和产品团队能够通过文本到SQL（Text - to - SQL）与数据进行交互聊天，创建图表、电子表格、报告和商业智能（BI）。
39. nus-apr/auto-code-rover - 一位了解项目结构的自主软件工程师致力于自主程序改进。它在两个基准测试中完成了一定比例的任务，且每项任务的成本低于0.7美元。
40. fern-api/fern - 输入OpenAPI，并输出SDK（软件开发工具包）和文档。
41. georgia-tech-db/evadb - 一个由人工智能驱动的应用程序数据库系统。
42. AbanteAI/mentat - Mentat - 人工智能编码助手
43. emcf/engshell - 一个由大型语言模型（LLMs）驱动的英语语言外壳（shell），可在任何操作系统（OS）上使用。
44. AI-Citizen/SolidGPT - 一个用于搜索角色的开发者人工智能。
45. context-labs/autodoc - 一个使用大型语言模型（LLMs）自动生成代码库文档的实验性工具包。
46. knuckleswtf/scribe - 从Laravel代码库为人类生成API文档。
47. jina-ai/dev-gpt - 你的虚拟开发团队可以是一群远程工作的开发人员，通过虚拟协作来开发软件或其他项目。
48. Pythagora-io/pythagora - 使用大型语言模型（LLMs）为Node.js应用程序生成自动化测试，开发人员无需编写任何代码。
49. eli64s/readme-ai - 一个由人工智能驱动的自述文件生成器。
50. mattzcarey/code-review-gpt - 使用大型语言模型（GPT4、Sonnet 3.5）和嵌入（Embeddings）进行代码审查可提高代码质量并在预生产阶段发现错误，并且与Github/GitLab/Azure DevOps持续集成（CI）集成。
51. smallcloudai/refact - 用于微调以及自行托管开源大型编码语言模型的WebUI。
52. eylonmiz/react-agent - 开源的React.js自治大型语言模型（LLM）代理。
53. gorilla-llm/gorilla-cli - 用于命令行界面（CLI）的大型语言模型（LLMs）。
54. huggingface/llm-vscode - 在VSCode中由大型语言模型（LLM）驱动的开发。
55. peterw/Chat-with-Github-Repo - 这个存储库有两个Python脚本，用于通过Streamlit、OpenAI GPT - 3.5 - turbo和Activeloop的Deep Lake创建聊天机器人。

56. paralleldrive/sudolang-llm-support - Visual Studio Code中的SudoLang大型语言模型（LLM）支持。
57. ricklamers/shell-ai - 一个由LangChain提供支持的用于生成和运行shell命令的命令行界面（CLI）。
58. google/oss-fuzz-gen - 通过OSS - Fuzz由大型语言模型 (LLM) 驱动的模糊测试。
59. kantord/SeaGOAT - 采用本地优先方法的语义代码搜索引擎。
60. OpenAutoCoder/Agentless - 一种用于自动解决软件开发问题的无代理方法。
61. ferrislucas/promptr - Promptr是一个命令行界面（CLI）工具，它能使用通俗易懂的英语指示GPT3或GPT4修改代码库。

AI教程

1. microsoft/generative-ai-for-beginners - 使用生成式人工智能开始构建的21个课程。链接：https://microsoft.github.io/generative-ai-for-beginners/
2. openai/openai-cookbook - OpenAI API使用示例与指南。
3. mlabonne/llm-course - 一个深入大型语言模型（LLMs）的课程，包含路线图和Colab笔记本。
4. rasbt/LLMs-from-scratch - 从头开始逐步在PyTorch中实现一个类似ChatGPT的大型语言模型（LLM）。
5. lutzroeder/netron - 一个用于神经网络、深度学习和机器学习模型的可视化工具。
6. datawhalechina/prompt-engineering-for-developers - 面向开发者的大型语言模型（LLM）入门教程，吴恩达模型系列课程中文版。
7. liguodongiot/llm-action - 本项目旨在分享与大型模型相关的技术原理以及实践经验（大型模型工程和大型模型应用实施）。
8. stas00/ml-engineering - 机器学习工程方面的开放书籍。
9. mikeroyal/Self-Hosting-Guide - 自托管指南：由个人或组织在本地托管和管理软件应用程序，涵盖云、大型语言模型（LLMs）等内容。
10. hua1995116/awesome-ai-painting - 人工智能绘画素材收集，包括国内外平台、教程和新闻，如Stable diffusion（稳定扩散）、AnimateDiff（动画扩散）、Stable Cascade（稳定级联）、Stable SDXL Turbo（稳定SDXL涡轮增压）。
11. Mooler0410/LLMsPracticalGuide - 大型语言模型（LLMs）实用指南资源列表，包括LLMs树、示例和论文。
12. GoogleCloudPlatform/generative-ai - 谷歌云（Google Cloud）上使用Vertex AI中的Gemini进行生成式人工智能（Generative AI）的示例代码和笔记本。
13. kyrolabs/awesome-langchain - 使用LangChain框架的工具和项目的优秀列表。
14. microsoft/DeepSpeedExamples - DeepSpeed示例模型。

15. huggingface/alignment-handbook - 使语言模型符合人类和人工智能偏好的稳健方法。
16. trigaten/Learn_Prompting - 提示工程、生成式人工智能和大型语言模型（LLM）指南，由Learn Prompting提供。加入其Discord（一款聊天软件），获取最大的提示工程学习社区。
17. bbycroft/llm-viz - GPT风格大型语言模型的3D可视化。
18. ray-project/llm-numbers - 每个大型语言模型（LLM）开发者都应该知晓的数字。
19. luban-agi/Awesome-AIGC-Tutorials - 大型语言模型、人工智能绘画等方面的精选教程和资源。
20. georgezouq/awesome-ai-in-finance - 金融市场中一系列精心挑选的优秀大型语言模型（LLMs）、深度学习策略和工具。
21. howl-anderson/unlocking-the-power-of-llms - 使用提示（Prompts）和链（Chains）使ChatGPT成为强大的生产力工具。释放大型语言模型（LLMs）的潜力。
22. ashishpatel26/LLM-Finetuning - 使用PEFT（参数高效微调）对大型语言模型（LLM）进行微调。
23. ray-project/llm-applications - 一份面向生产开发基于检索增强生成（RAG）的大型语言模型（LLM）应用的综合指南。
24. premAI-io/state-of-open-source-ai - 在开源创新这个混乱又快节奏的世界里，需要有清晰的思路。
25. pionxzh/chatgpt-exporter - 导出并分享你的ChatGPT聊天记录。
26. ianand/spreadsheets-are-all-you-need -
27. majacinka/crewai-experiments - 使用本地模型和可通过应用程序接口（API）访问的模型进行实验。
28. thu-vu92/local-llms-analyse-finance -

工作流自动化

1. KillianLucas/open-interpreter - 计算机的自然语言界面。
2. StanGirard/quivr - 用于将生成式人工智能（GenAI）集成到应用中的有主见的检索增强生成（RAG）技术，重点关注产品。可在现有产品中轻松集成并定制，并且在大型语言模型（LLM）、向量存储和文件方面具有多功能性。
3. danielmiessler/fabric - Fabric是一个开源的人工智能增强人类框架，它具有模块化结构，可通过众包人工智能提示来解决问题。
4. openai-translator/openai-translator - 使用ChatGPT API进行翻译的浏览器和桌面应用程序。
5. Skyvern-AI/skyvern - 使用大型语言模型（LLMs）和计算机视觉技术实现基于浏览器的任务自动化。
6. activepieces/activepieces - 您最友好的开源人工智能自动化工具。它是一个具有200多种集成的工作流自动化工具，是企业自动化方面Zapier的替代品。

7. OthersideAI/self-operating-computer - 一种供多模态模型操作计算机的框架。
8. microsoft/UFO - 一个专注于用户界面的Windows操作系统交互代理。
9. yihong0618/bilingual_book_maker - 利用人工智能翻译手段创作双语的epub书籍。
10. lavague-ai/LaVague - 用于开发人工智能网络代理的大型动作模型框架。
11. aisingapore/TagUI - 由新加坡人工智能开发的一款免费的机器人流程自动化（RPA）工具。
12. openchatai/OpenCopilot - 语言到行为引擎
13. KillianLucas/01 - 适用于桌面端、移动端和ESP32芯片的顶级开源语音接口。
14. katanaml/sparrow - 使用机器学习、大型语言模型（LLM）和基于视觉的大型语言模型进行数据处理。
15. xlang-ai/OpenAgents - 2024年的OpenAgents：一个面向野生语言智能体（agents）的开放平台。
16. BAAI-Agents/Cradle -
17. Cormanz/smartgpt - 一个使大型语言模型（LLMs）能够借助插件完成复杂任务的程序。
18. fiatrete/OpenDAN-Personal-AI-OS - OpenDAN是一个开源的个人人工智能操作系统，它整合了各种人工智能模块供个人使用。
19. n4ze3m/page-assist - 使用本地运行的人工智能模型来协助网络浏览。
20. OS-Copilot/FRIDAY -
21. andrewnguonly/Lumos - 一个由本地大型语言模型（LLM）提供支持、用于网络浏览的检索增强生成（RAG）大型语言模型（LLM）副驾驶。
22. Dicklesworthstone/swiss_army_llama - 一个通过预先计算的嵌入、相似性度量以及通过textract支持文件类型来进行语义文本搜索的FastAPI服务。

AI机器人

1. lencx/ChatGPT - 适用于Mac、Windows和Linux系统的ChatGPT桌面应用程序。
2. LAION-AI/Open-Assistant - OpenAssistant是一个基于聊天的助手，能够理解任务、与第三方系统交互并动态检索信息。
3. zhayujie/chatgpt-on-wechat - 一个基于大型模型的聊天机器人，支持多个平台（微信公众号、企业微信应用、飞书、钉钉等）、多个模型（GPT3.5/GPT - 4o/GPT - o1/克劳德/文心一言/讯飞星火/通义千问/双子座/GLM - 4/克劳德/奇米/链爱），能够处理文本、语音和图片，访问操作系统和互联网，并支持基于自有知识库定制企业智能客服。
4. Chanzhaoyu/chatgpt-web - 一个使用Express和Vue3构建的ChatGPT演示网页。
5. janhq/jan - Jan是一个开源的ChatGPT替代品，可在计算机上完全离线运行。
6. Bin-Huang/chatbox - 对人工智能模型/大型语言模型（如GPT、Claude、Gemini、Ollama等）友好的桌面客户端应用程序。

7. joonspk-research/generative_agents - 生成式智能体：人类行为的交互模拟。
8. Unity-Technologies/ml-agents - Unity ML - Agents工具包是一个开源项目，用于通过深度强化学习和模仿学习在游戏和模拟中训练智能体。
9. transitive-bullshit/chatgpt-api - 一个兼容任何大型语言模型（LLM）和TypeScript人工智能软件开发工具包（SDK）的人工智能代理标准库。
10. leon-ai/leon - 利昂是你的开源个人助手。
11. xcanwin/KeepChatGPT - 这是一个增强ChatGPT数据安全性和效率的插件。它提供许多免费的创新功能以提供更好的人工智能体验。
12. lss233/chatgpt-mirai-qq-bot - 一键部署！真正的人工智能聊天机器人，支持多平台和多种功能。
13. getumbrel/llama-gpt - 一个像ChatGPT一样的自托管离线聊天机器人，由Llama 2提供支持，是私有的，没有数据离开设备，现在还支持Code Llama。
14. sfyc23/EverydayWechat - 微信助手：1. 每天定期向朋友（女友）发送定制消息。2. 机器人自动回复朋友。3. 群助手功能（如垃圾分类查询、天气、日历、实时电影票房、快递物流、PM2.5等）。
15. BlinkDL/ChatRWKV - ChatRWKV是一个像ChatGPT一样的开源语言模型，但由RWKV（一种100%的循环神经网络）提供动力。
16. ztjhz/BetterChatGPT - ChatGPT的一个很棒的用户界面，可在网站以及包括Windows、MacOS和Linux在内的多种操作系统上使用。
17. a16z-infra/ai-town - 一个用于构建人工智能小镇（人工智能角色在其中生活、聊天和社交）的、遵循麻省理工学院（MIT）许可的入门工具包可部署且可定制。
18. memochou1993/gpt-ai-assistant - OpenAI、LINE和Vercel结合起来形成了GPT AI助手。
19. miurla/morphic - 一个由人工智能驱动并具有生成式用户界面的搜索引擎。
20. interstellard/chatgpt-advanced - WebChatGPT是一款浏览器扩展程序，可通过网络结果增强ChatGPT提示。
21. linyiLYi/street-fighter-ai - 这是一个针对《街头霸王II冠军版》的人工智能代理。
22. vincelwt/chatgpt-mac - 适用于Mac的ChatGPT驻留在你的菜单栏中。
23. camel-ai/camel - CAMEL：首个也是最佳的多智能体框架，用于发现智能体的扩展定律。(https://www.camel - ai.org)
24. MineDojo/Voyager - 与大型语言模型相关的开放式具身智能体。
25. a16z-infra/companion-app - 具有记忆功能的人工智能伙伴：一个用于创建和托管自己的人工智能伙伴的轻量级堆栈。
26. ConnectAI-E/Feishu-OpenAI - 飞书（结合GPT - 4、GPT - 4V、DALL·E - 3和Whisper）提供了很棒的工作体验，包括语音对话、角色扮演、多话题讨论、图像创作、表格分析和文档导出。
27. simonw/llm - 通过命令行访问大型语言模型。

28. sigoden/aichat - 一款集Shell助手、聊天交互（Chat - REPL）、检索增强生成（RAG）、人工智能工具与代理于一体的大型语言模型（LLM）命令行界面（CLI）工具，可访问OpenAI、Claude等多个平台。
29. lencx/nofwl - 无防火墙（No FireWall，简称NoFWL）桌面应用程序。
30. Kent0n-Li/ChatDoctor -
31. xtekky/chatgpt-clone - 具有改进用户界面的ChatGPT界面。
32. deep-diver/LLM-As-Chatbot - 大型语言模型（LLM）作为一种聊天机器人服务。
33. gragland/chatgpt-chrome-extension - 一个将ChatGPT集成到互联网上每个文本框的ChatGPT Chrome扩展程序。
34. ohmplatform/FreedomGPT - 这个代码库是用于一个带有基于聊天界面的React - Electron（一种将React框架用于构建桌面应用的技术）应用程序，可在Mac和Windows系统本地运行FreedomGPT大型语言模型（LLM）。
35. SoraWebui/SoraWebui - SoraWebui是一个开源的Sora网络客户端，可轻松使用OpenAI的Sora模型从文本创建视频。
36. karthink/gptel - 一个使用大型语言模型的简单Emacs客户端。
37. a16z-infra/llama2-chatbot - LLaMA v2聊天机器人。
38. ItsPi3141/alpaca-electron - 在自己的个人电脑上运行羊驼（Alpaca）和其他基于LLaMA的本地大型语言模型（LLM）的最简单方法。
39. opendilab/DI-star - 一个用于《星际争霸II》的人工智能平台，具备大规模分布式训练和宗师级智能体。
40. jncraton/languagemodels - 使用512MB内存探索大型语言模型。
41. SamurAIGPT/Camel-AutoGPT - 介绍CAMEL，一种针对大型语言模型（LLMs）和自动代理（auto - agents）的角色扮演方法。它使代理能够协作，并在多个领域具有潜力。
42. Syan-Lin/CyberWaifu - 一个由大型语言模型（LLM）+语音合成（TTS）构成的具有真实感的聊天机器人，一个支持表情符号、QQ表情和互联网搜索的QQ机器人。

多模态模型

1. PaddlePaddle/PaddleOCR - 基于飞桨（PaddlePaddle）的超棒多语言光学字符识别（OCR）工具包。它们实用、超轻量，支持80多种语言，可在多种设备上使用。
2. suno-ai/bark - 一种由文本提示的生成式音频模型。
3. openai/CLIP - CLIP（对比语言 - 图像预训练）：为图像预测最相关的文本片段。
4. hpcaitech/Open-Sora - Open - Sora：使每个人都能以民主的方式进行高效的视频制作。
5. haotian-liu/LLaVA - NeurIPS'23 Oral（神经信息处理系统大会2023年口头报告）中的视觉指令调整（LLaVA）旨在获得GPT - 4V级别的能力甚至更强的能力。

6. fishaudio/fish-speech - 最先进的开源语音合成（TTS）技术。
7. borisdayma/dalle-mini - DALL·E Mini根据文本提示生成图像。
8. google-deepmind/alphafold - AlphaFold 2的开源代码。
9. OpenBMB/OmniLMM - MiniCPM - V 2.6是一款在手机上用于单张图像、多张图像和视频的、达到GPT - 4V水平的多模态大语言模型（MLLM）。
10. PKU-YuanGroup/Open-Sora-Plan - 该项目旨在复现Sora（OpenAI的文本到视频模型），并希望开源社区做出贡献。
11. openai/shap-e - 基于文本或图像生成3D对象。
12. facebookresearch/seamless_communication - 用于最先进的语音和文本翻译的基础模型。
13. openai/DALL-E - PyTorch软件包用于DALL·E中的离散变分自编码器（VAE）。
14. google-research/vision_transformer -
15. magic-research/magic-animate - CVPR 2024中的MagicAnimate使用扩散模型实现时序一致的人体图像动画。
16. ashawkey/stable-dreamfusion - 使用神经辐射场（NeRF）+扩散技术进行文本到3D、图像到3D以及网格导出。
17. lucidrains/imagen-pytorch - 在Pytorch中实现谷歌的文本到图像神经网络Imagen。
18. openai/jukebox - 论文《Jukebox：一种音乐生成模型》的代码。
19. deep-floyd/IF -
20. netease-youdao/EmotiVoice - EmotiVoice😊：一款拥有多种音色并且受提示控制的语音合成（TTS）引擎。
21. IDEA-Research/GroundingDINO - 论文《Grounding DINO：将DINO与基础预训练相结合用于开集目标检测》在ECCV 2024中的官方实现。
22. FoundationVision/VAR - NeurIPS 2024口头报告《视觉自回归建模：通过下一尺度预测进行可扩展图像生成》的官方实现。这是一个用于自回归图像生成的极其简单、用户友好的最先进代码库。
23. threestudio-project/threestudio - 一种统一生成3D内容的框架。
24. openai/guided-diffusion -
25. THUDM/CogVLM - 一种最先进的开放式视觉语言模型，一种多模态预训练模型。
26. openai/consistency_models - 一致性模型官方库。
27. levihsu/OOTDiffusion - OOTDiffusion：基于潜扩散的服装融合用于可控虚拟试穿——官方实现。
28. clovaai/donut - ECCV 2022的Donut（无光学字符识别的文档理解变换器）和SynthDoG（合成文档生成器）的官方实现。
29. google/gemma_pytorch - 谷歌Gemma模型的官方PyTorch实现。

30. QwenLM/Qwen-VL - 通义千问 - VL（阿里云的一个聊天和预训练大视觉语言模型）的官方仓库。
31. yl4579/StyleTTS2 - StyleTTS 2旨在通过风格扩散和利用大型语音语言模型进行对抗训练来实现人类水平的文本到语音转换。
32. snakers4/silero-models - Silero模型是用于语音转文本、文本转语音和文本增强的预训练模型，这些模型制作得非常简单。
33. salesforce/BLIP - 用于BLIP（Bootstrapping Language - Image Pre - training，自举语言 - 图像预训练）的PyTorch代码，BLIP用于通过自举语言 - 图像预训练来实现统一的视觉 - 语言理解和生成。
34. google-deepmind/alphageometry -
35. metavoiceio/metavoice-src - 一个类人、富有表现力的文本到语音（TTS）基础模型。
36. Luodian/Otter - 水獭（Otter）是一个基于OpenFlamingo的多模态模型，在MIMIC - IT数据集上进行训练，具有更好的指令遵循和上下文学习能力。
37. NExT-GPT/NExT-GPT - NExT - GPT（一种任意到任意多模态大型语言模型）的代码和模型。
38. openai/improved-diffusion - 发布用于改进的去噪扩散概率模型。
39. X-PLUG/MobileAgent - 移动设备 - 代理：强大的移动设备操作助手家族。
40. dvlab-research/MiniGemini - 《Mini - Gemini：挖掘多模态视觉语言模型的潜力》官方知识库。
41. lucidrains/musiclm-pytorch - 使用注意力网络在PyTorch中实现谷歌最先进的音乐生成模型MusicLM。
42. hustvl/Vim - Vision Mamba（于2024年国际机器学习会议中提出）通过双向状态空间模型实现高效的视觉表征学习。
43. OpenGVLab/Ask-Anything - CVPR2024亮点：VideoChatGPT使ChatGPT能够理解视频。它还支持其他语言模型，如 miniGPT4、StableLM和MOSS。
44. microsoft/lida - 通过大型语言模型自动生成可视化内容和信息图表。
45. google-research/frame-interpolation - ECCV 2022中的大运动帧插值（FILM）
46. InternLM/InternLM-XComposer - InternLM - XComposer2.5 - OmniLive：一个用于长期视频和音频交互的多模态系统。
47. yerfor/GeneFace - GeneFace：广义且高保真的3D说话人脸合成，ICLR 2023，含官方代码。
48. OpenGVLab/InternImage - CVPR 2023中的InternImage：利用可变形卷积探索大规模视觉基础模型。
49. google-deepmind/gemma - 谷歌DeepMind的开放权重大型语言模型。
50. baaivision/EVA - EVA系列：来自北京智源人工智能研究院（BAAI）的视觉表象的幻想。
51. MzeroMiko/VMamba - VMamba：视觉状态空间模型。其代码基于Mamba。
52. deepseek-ai/DeepSeek-VL - DeepSeek - VL旨在对现实世界中的视觉 - 语言进行理解。

53. openai/consistencydecoder - 一致性蒸馏差分变分自编码器。
54. gligen/GLIGEN - 开放 - 基于基础（grounding）的文本到图像生成。
55. dvlab-research/LISA - “LISA：基于大型语言模型的推理分割”项目页面。
56. 3DTopia/LGM - ECCV 2024口头报告中的LGM用于高分辨率3D内容创作。
57. lyuchenyang/Macaw-LLM - 金刚鹦鹉（Macaw）——大型语言模型（LLM）集成了图像、视频、音频和文本，用于多模态语言建模。
58. OpenMotionLab/MotionGPT - MotionGPT在2023年神经信息处理系统大会（NeurIPS 2023）上是一个使用大型语言模型（LLMs）的统一运动 - 语言生成模型。
59. OpenGVLab/InternVideo - ECCV2024（欧洲计算机视觉国际会议2024）中用于多模态理解的视频基础模型和数据。
60. openai/Video-Pre-Training - 视频预训练（VPT）包括通过观察未标记的在线视频来学习行动。
61. THUDM/ImageReward - NeurIPS 2023中的ImageReward：学习和评估人类对文生图的偏好。
62. evo-design/evo - 从分子到基因组规模的生物学基础建模。
63. google-research/tapas - 用于理解表格和文本的端到端神经模型。
64. apple/ml-aim - 该存储库提供用于AIMv1和AIMv2研究项目的代码和模型检查点。
65. showlab/Show-o - Show - o的代码库，一个用于统一多模态理解和生成的单一Transformer。
66. ELLA-Diffusion/ELLA - 为扩散模型配备大型语言模型（LLM）以加强语义对齐。
67. declare-lab/tango - 一种用于文生音的扩散模型家族。
68. OpenBMB/VisCPM - 基于CPM基础模型的汉英双模态大模型系列（聊天与绘画）
69. OpenGVLab/VisionLLM - 视觉大型语言模型（VisionLLM）系列。
70. BAAI-DCAI/Bunny - 轻量级多模态模型家族。
71. Ligo-Biosciences/AlphaFold3 - AlphaFold3的开源实现。
72. Vchitect/SEINE - SEINE：一种用于2024年国际学习表征会议（ICLR）中生成性转换和预测的短视频到长视频扩散模型。
73. google-deepmind/materials_discovery -
74. OpenGVLab/SAM-Med2D - SAM - Med2D的官方实现。
75. OpenMOSS/AnyGPT - 用于“AnyGPT：具有离散序列建模的统一多模态大型语言模型（LLM）”的代码。

多语言模型

1. THUDM/ChatGLM-6B - ChatGLM - 6B是一个开放的双语对话语言模型。
2. ymcui/Chinese-LLaMA-Alpaca - 中国的LLaMA（小羊驼）和Alpaca（羊驼）大型语言模型 + 本地CPU（中央处理器）/GPU（图形处理器）训练与部署

3. UKPLab/sentence-transformers - 最先进的文本嵌入技术。
4. FlagAlpha/Llama2-Chinese - 羊驼（Llama）中文社区已开放Llama3以供在线体验和微调。它更新了Llama3的所有代码，完全开源且可用于商业用途，并且还编写了最新的Llama3学习资料。
5. THUDM/ChatGLM3 - ChatGLM3系列：开源双语聊天大型语言模型。
6. ymcui/Chinese-LLaMA-Alpaca-2 - 中国版LLaMA - 2和Alpaca - 2大型模型二期项目以及64K长文本模型。
7. InternLM/InternLM - InternLM2.5基础模型和聊天模型正式发布，支持100万（1M）的上下文。
8. Facico/Chinese-Vicuna - 中国 - Vicuna：一个遵循指令的基于LLaMA的中文模型 - 一种参考羊驼结构的低资源中文llama + lora解决方案。
9. LC1332/Luotuo-Chinese-LLM - 骆驼（Luotuo）是由华中师范大学的陈启源、商汤科技的李路路和冷子昂开发的开源中文语言模型。
10. wenge-research/YAYI2 - YAYI 2是中科闻歌开发的新一代开源大语言模型，使用超过2万亿个高质量、多语言语料库的标记进行预训练。
11. wenge-research/YaYi - 亚一大型模型由中科文歌算法团队开发，是为客户打造的安全可靠的专属大型模型，它基于LlaMA2和BLOOM系列的大规模中英文多领域指令数据进行训练。
12. TigerResearch/TigerBot - TigerBot是一个支持多种语言和任务的大型语言模型。
13. LinkSoul-AI/Chinese-Llama-2-7b - 开源社区中首个可下载且可运行的中文LLaMA2模型！
14. MiuLab/Taiwan-LLM - 面向台湾的传统普通话语言模型。
15. zjunlp/KnowLM - 一个带有知识的开源大语言模型框架。
16. google-research/multilingual-t5 -
17. SkyworkAI/Skywork - Skywork系列模型在3.2TB的多语言和代码数据上进行了预训练，相关项目已开源。

数据组织

1. photoprism/photoprism - 去中心化网络上由人工智能驱动的照片应用。
2. freedmand/semantra - 一种用于语义搜索的多功能工具。
3. neo4j/NaLLM - NaLLM项目的存储库。

AI服务

1. vllm-project/vllm - 一种用于大型语言模型（LLMs）推理和服务的高通量且内存高效的引擎。
2. guillaumekln/faster-whisper - 使用CTranslate2进行更快速的Whisper转录。
3. bentoml/OpenLLM - 在云端将像Llama和Mistral这样的开源大型语言模型（LLM）作为与OpenAI兼容的API端点来运行。
4. huggingface/text-generation-inference - 大规模语言模型文本生成推理。

5. FMInference/FlexGen - 在以吞吐量为重点的场景下，在单个GPU上运行大型语言模型。
6. triton-inference-server/server - Triton推理服务器为云端和边缘端提供了优化的推理解决方案。
7. dusty-nv/jetson-inference - 《Hello AI World》中使用TensorRT和NVIDIA Jetson部署深度学习推理网络和深度视觉原语指南。
8. openvinotoolkit/openvino - OpenVINO™是一个开源的人工智能推理优化和部署工具包。
9. zilliztech/GPTCache - 用于大型语言模型（LLMs）的语义缓存，与LangChain和llama_index完全集成。
10. Portkey-AI/gateway - 一个速度非常快、集成了防护栏并且可以通过一个API路由到许多大型语言模型（LLMs）和人工智能防护栏的人工智能网关。
11. tensorflow/serving - 一个灵活且高性能的机器学习模型服务系统。
12. xorbitsai/inference - 通过使用Xinference修改一行代码，在你的应用中用另一个大型语言模型（LLM）替换OpenAI GPT，Xinference支持在任何地方运行各种模型的推理。
13. allegroai/clearml - ClearML是一种用于人工智能工作负载（包括实验和数据管理等）的MLOps/LLMOps解决方案。
14. InternLM/lmdeploy - LMDeploy是一个用于压缩、部署和服务大型语言模型的工具包。
15. argmaxinc/WhisperKit - 苹果硅芯片设备端语音识别。
16. kserve/kserve - 一个基于Kubernetes的标准化无服务器机器学习推理平台。
17. neuralmagic/deepsparse - 基于CPU的稀疏感知深度学习推理运行时。
18. huggingface/text-embeddings-inference - 一种用于文本嵌入模型的非常快速的推理解决方案。
19. open-mmlab/mmdeploy - OpenMMLab的模型部署框架。
20. ModelTC/lightllm - LightLLM是一个基于Python的大型语言模型（LLM）推理和服务框架，它轻巧、易于扩展且速度快。
21. predibase/lorax - 一个可扩展至数千个微调大型语言模型（LLM）的多LoRA推理服务器。
22. langchain-ai/langserve - 朗格服务（LangServe）🦜🏓
23. S-LoRA/S-LoRA - S - LoRA：服务大量并发的LoRA适配器。
24. michaelfeil/infinity - Infinity是一种用于文本嵌入、重排序模型、clip、clap和colpali的服务引擎，具有高吞吐量和低延迟的特点。
25. roboflow/inference - 将任何计算机或边缘设备转变为计算机视觉项目的指挥枢纽。
26. ray-project/ray-llm - RayLLM - 基于Ray的大型语言模型。
27. PygmalionAI/aphrodite-engine - 大规模语言模型（LLM）推理引擎。
28. punica-ai/punica - 将多个LoRA微调的大型语言模型作为一个来服务。

29. msoedov/langcorn - 使用FastApi为LLMops自动为LangChain大型语言模型（LLM）应用程序和代理提供服务。
30. mosecorg/mosec - 一种高性能机器学习模型服务框架，具有动态批处理和CPU/GPU管道，可最大限度地提高计算机利用率。

向量数据库

1. facebookresearch/faiss - 一个用于高效稠密向量相似性搜索和聚类的库。
2. milvus-io/milvus - Milvus是一个高性能、云原生的向量数据库，用于可扩展的向量近似最近邻搜索。
3. qdrant/qdrant - Qdrant是一款面向下一代人工智能的高性能、大规模矢量数据库和搜索引擎，也提供云服务。
4. chroma-core/chroma - 人工智能原生的开源嵌入数据库。
5. spotify/annoy - 针对内存使用和磁盘I/O进行优化的C++/Python近似最近邻算法。
6. weaviate/weaviate - Weaviate是一个开源向量数据库。它存储对象和向量，并能够在结构化过滤下进行向量搜索，具备容错性和可扩展性。
7. neuml/txtai - 一个用于语义搜索、大型语言模型（LLM）编排和语言模型工作流的一体化开源嵌入数据库。
8. activeloopai/deeplake - 一个供人工智能使用的数据库，它能够存储诸如向量、图像、文本和视频等各种类型的数据。它可与大型语言模型/语言链（LLMs/LangChain）一起用于存储、查询、版本管理和可视化人工智能数据等操作，并且能够向PyTorch/ TensorFlow实时传输数据。
9. vespa-engine/vespa - 人工智能+数据，可在https://vespa.ai在线获取。
10. lancedb/lancedb - 一种用于人工智能应用的无服务器矢量数据库，对开发者友好，能轻松为大型语言模型（LLM）应用添加长期记忆。
11. marqo-ai/marqo - 统一的嵌入生成和搜索引擎，也可在云端使用 - cloud.marqo.ai。
12. nmslib/hnswlib - 一个仅含头文件的C++/python库，用于快速近似最近邻搜索。
13. unum-cloud/usearch - 适用于多种编程语言的快速开源搜索与聚类引擎。
14. tensorchord/pgvecto.rs - 借助混合功能在Postgres（一种数据库管理系统）中进行可扩展、低延迟的向量搜索。它革新的是向量搜索而非数据库。
15. spotify/voyager - 一个用于近似最近邻搜索的Python和Java库，侧重于易用性、简洁性和可部署性。
16. rapidsai/raft - RAFT拥有用于机器学习和信息检索（IR）的基本算法和原语，这些算法和原语通过CUDA加速以用于高性能应用。

AI教育

1. JushBJJ/Mr.-Ranedeer-AI-Tutor - 一种用于个性化学习体验且可定制的GPT - 4人工智能导师提示词。
2. Nutlope/llamatutor - 一个AI私人导师是基于Llama 3.1构建的。
3. codeacme17/examor - 对于学生、学者、受访者和终身学习者来说，大型语言模型（LLMs）有助于学习。

AI开发工具

1. jina-ai/jina - 使用云原生技术栈构建多模态人工智能应用。
2. iterative/dvc - 数据版本控制与机器学习实验。
3. unifyai/ivy - 在不同框架之间转换机器学习代码。
4. HigherOrderCO/HVM - 用Rust编写的大规模并行最优函数运行时。
5. marimo-team/marimo - Python反应式笔记本可用于可重现性实验、脚本执行、应用程序部署以及使用Git进行版本控制。
6. arogozhnikov/einops - 用于创建可读性和可靠性兼具的代码的灵活且强大的张量操作，适用于PyTorch、Jax、TensorFlow等。
7. replicate/cog - 机器学习中使用的容器。
8. jessevig/bertviz - BertViz：可视化如BERT、GPT2、BART等自然语言处理模型中的注意力。
9. AbdBarho/stable-diffusion-webui-docker - 通过用户友好界面轻松设置用于Stable Diffusion的Docker。
10. huggingface/safetensors - 一种存储和分配张量的简单且安全的方法。
11. wangzhaode/mnn-llm - 基于MNN部署一个大型语言模型（LLM）项目。
12. ajndkr/lanarky - 用于构建大型语言模型（LLM）微服务的网络框架。

模型训练

1. tensorflow/tensorflow - 一个所有人都能使用的开源机器学习框架。
2. huggingface/transformers - Transformer：适用于Pytorch、TensorFlow和JAX的最先进机器学习技术。
3. pytorch/pytorch - 使用强大的GPU加速功能、基于Python的张量和动态神经网络。
4. hpcaitech/ColossalAI - 降低大型人工智能模型的成本、提高其速度并增强其可及性。
5. hiyouga/LLaMA-Factory - 对100多个大型语言模型（LLMs）进行统一高效的微调（ACL 2024）
6. lm-sys/FastChat - 一个用于大型语言模型训练、服务和评估的开放平台，也是Vicuna和Chatbot Arena的发布库。
7. coqui-ai/TTS - 🐸💬是一个用于文本转语音的深度学习工具包，在研究和生产中得到验证。

8. microsoft/DeepSpeed - DeepSpeed是一个深度学习库，用于轻松、高效且有效的分布式训练和推理。
9. ray-project/ray - Ray是一个人工智能计算引擎，它具有核心分布式运行时和人工智能库，用于加速机器学习工作负载。
10. google-research/google-research - 谷歌研究
11. google/jax - Python+NumPy程序可以通过多种方式进行组合转换，例如求导、向量化以及即时编译（JIT）到GPU/TPU等。
12. open-mmlab/mmdetection - OpenMMLab检测工具包与基准测试。
13. tinygrad/tinygrad - 如果你喜欢PyTorch和Micrograd，那么你也会喜欢Tinygrad。
14. huggingface/diffusers - Diffusers：用于生成图像、视频和音频的PyTorch和FLAX最先进的扩散模型。
15. mozilla/DeepSpeech - DeepSpeech是一个开源的语音转文本引擎，可在各种设备上实时使用。
16. modularml/mojo - 莫霍编程语言（给定描述中未提供更多细节）
17. microsoft/unilm - 大规模涵盖任务、语言和模态的自监督预训练。
18. ml-explore/mlx - MLX是一个适用于苹果芯片的数组框架。
19. HigherOrderCO/Bend - 一种具有大规模并行性的高级编程语言。
20. huggingface/peft - PEFT：最佳的参数高效微调。
21. huggingface/candle - 一个用于Rust语言的极简机器学习框架。
22. NVIDIA/NeMo - 一个适用于大型语言模型（LLMs）、多模态和语音人工智能等人工智能领域研究人员和开发人员的框架，该框架具有可扩展性和生成性。
23. PaddlePaddle/PaddleNLP - 一个易于使用且功能强大的自然语言处理（NLP）和大型语言模型（LLM）库，拥有大量优秀的模型，支持从研究到工业应用的各种自然语言处理任务。
24. PaddlePaddle/PaddleSpeech - 一个易用的语音工具包包含多种功能，并获得了NAACL2022最佳演示奖。
25. Lightning-AI/litgpt - 20多个高性能大型语言模型（LLM）以及大规模预训练、微调与部署的相关方案。
26. huggingface/trl - 利用强化学习来训练Transformer语言模型。
27. artidoro/qlora - QLoRA能够实现量化大型语言模型的高效微调。
28. salesforce/LAVIS - LAVIS是一个一站式的语言 - 视觉智能库。
29. nerfstudio-project/nerfstudio - 一个对神经辐射场（NeRFs）协作友好的工作室。
30. mozilla/TTS - 用于语音合成的深度学习（讨论论坛：https://discourse.mozilla.org/c/tts）
31. tracel-ai/burn - Burn是一个全新的动态深度学习框架，由Rust构建，旨在实现灵活性、高效性和可移植性。

32. facebookresearch/pytorch3d - PyTorch3D是FAIR（Facebook人工智能研究）用于3D数据深度学习的库。
33. facebookresearch/xformers - Transformer构建模块可灵活调整且经过优化，支持组合式构建。
34. OptimalScale/LMFlow - 一个用于大型基础模型微调及推理的可扩展工具包，让所有人都能使用大型模型。
35. OpenAccess-AI-Collective/axolotl - 只管去问蝾螈问题就好。
36. FlagOpen/FlagEmbedding - 检索与检索增强型大型语言模型（LLMs）
37. huggingface/accelerate - 一种在各种设备和配置上处理PyTorch模型的简单方法，具有自动混合精度等功能，并支持完全分片数据并行（FSDP）和DeepSpeed。
38. LianjiaTech/BELLE - BELLE是一个开源的面向所有人的中文对话大型语言模型引擎。
39. cloneofsimo/lora - 通过低秩适应快速微调扩散模型。
40. EleutherAI/gpt-neox - 使用Megatron和DeepSpeed库在GPU上实现具有模型并行性的自回归变换器。
41. open-mmlab/mmagic - OpenMMLab是一个多模态工具箱，可用于像人工智能生成内容（AIGC）等各种任务，拥有易用的应用程序接口（API）和模型库。
42. facebookresearch/metaseq - 外部大型工作资料库。
43. Maartengr/BERTopic - 使用BERT和c - TF - IDF生成易于理解的主题。
44. Project-MONAI/MONAI - 一个用于人工智能领域医疗影像（处理）的工具包。
45. yangjianxin1/Firefly - Firefly是一个大型模型训练工具，支持训练多个大型模型，如Qwen2.5、Qwen2等。
46. google-deepmind/graphcast -
47. mosaicml/composer - 增强你的模型训练。
48. cg123/mergekit - 用于组合预训练大型语言模型的工具。
49. CarperAI/trlx - 一个用于使用人类反馈强化学习（RLHF）进行语言模型分布式训练的代码库。
50. pytorch/torchtune - 一个PyTorch原生的训练后库。
51. google-deepmind/open_spiel - OpenSpiel是一组用于通用强化学习以及游戏搜索/规划研究的环境和算法。
52. huggingface/autotrain-advanced - 自动训练进阶版。
53. InternLM/xtuner - 一个高效、灵活且功能齐全的用于对各种大型语言模型（LLMs）进行微调的工具包。
54. mosaicml/llm-foundry - 用于Databricks基础模型的大型语言模型（LLM）训练代码。
55. baidu-research/warp-ctc - 快速并行的连接主义时间分类（CTC）。

56. JohnSnowLabs/spark-nlp - 自然语言处理的最先进技术。
57. FlagAI-Open/FlagAI - FlagAI是一个用于大规模模型的工具包，它快速、易用且可扩展。
58. mlfoundations/open_flamingo - 一个用于训练大型多模态模型的开源框架。
59. OpenLLMAI/OpenRLHF - 一个易用、可扩展且高性能的人类反馈强化学习（RLHF）框架，具有700亿以上参数的近端策略优化（PPO）完全微调、迭代直接偏好优化（DPO）、低秩自适应（LoRA）、环形注意力（RingAttention）和递归微调（RFT）等功能。
60. google-deepmind/acme - 一个用于强化学习的组件和代理库。
61. open-mmlab/mmpretrain - OpenMMLab的预训练工具箱和基准测试。
62. shibing624/MedicalGPT - MedicalGPT使用ChatGPT训练管道来训练医学GPT模型，实现增量预训练、有监督微调、人类反馈强化学习（RLHF）、直接偏好优化（DPO）和基于排序的偏好优化（ORPO）。
63. iryna-kondr/scikit-llm - 毫无问题地将大型语言模型（LLMs）集成到scikit - learn中。
64. google-research/scenic - Scenic：一个用于计算机视觉研究及其他更多用途的Jax库。
65. facebookresearch/fairscale - 用于高性能和大规模训练的PyTorch扩展。
66. alpa-projects/alpa - 通过自动并行化训练和服务大规模神经网络。
67. microsoft/torchscale - 大中型语言模型的基础架构。
68. google-deepmind/dm-haiku - 基于JAX的神经网络库。
69. eureka-research/Eureka - ICLR 2024会议上发表的论文《Eureka：通过对大型语言模型进行编码实现人类水平的奖励设计》的官方存储库。
70. Alpha-VLLM/LLaMA2-Accessory - 一个用于开发大型语言模型的开源工具包。
71. google-research/t5x -
72. google-deepmind/alphatensor -
73. PhoebusSi/Alpaca-CoT -
74. huggingface/optimum - 使用易用的硬件优化工具来加速Transformer、Diffuser、TIMM和Sentence Transformer的推理和训练。
75. stochasticai/xTuring - 通过xTuring可以轻松地从数据预处理到微调构建、定制和控制您自己的大型语言模型（LLMs），并加入其Discord社区。
76. adapter-hub/adapters - 一个用于参数高效和模块化迁移学习的统一库。
77. openai/weak-to-strong -
78. OpenPipe/OpenPipe - 将昂贵的提示转换为价格实惠的微调模型。
79. lamini-ai/lamini - Lamini API的官方Python客户端。

80. google-research/big_vision - 用于开发视觉变换器（Vision Transformer）、SigLIP、多层感知机混合器（MLP - Mixer）、LiT等的官方代码库。
81. young-geng/EasyLM - EasyLM（基于JAX/Flax）为预训练、微调、评估和服务等各种大型语言模型（LLM）操作提供一站式解决方案。
82. pyro-ppl/numpyro - 使用NumPy进行概率编程，并利用JAX进行自动求导以及将即时编译（JIT编译）到GPU/TPU/CPU。
83. eric-mitchell/direct-preference-optimization - 直接偏好优化（DPO）的参考实现。
84. huggingface/setfit - 使用句向量转换器（Sentence Transformers）进行高效的小样本学习。
85. allenai/open-instruct -
86. allenai/RL4LMs - 一个用于根据人类偏好微调语言模型的模块化强化学习（RL）库。
87. lxe/simple-llm-finetuner - 用于微调大型语言模型（LLM）的简单用户界面。
88. THUDM/P-tuning-v2 - 一种优化的深度提示微调策略在不同规模和任务中的效果与微调一样好。
89. tensorflow/privacy - 一个用于训练机器学习模型且对训练数据有隐私保护的库。
90. xlang-ai/instructor-embedding - 适用于任何任务的单一嵌入器：指令 - 微调文本嵌入（ACL 2023）
91. unslothai/hyperlearn - 机器学习算法的速度提高了2 - 2000倍，内存使用量减少50%，并且可在所有硬件上运行。
92. salesforce/ctrl - 一种用于可控生成的条件转换语言模型。
93. google-deepmind/optax - Optax是一个用于梯度处理和优化的JAX库。
94. google-deepmind/penzai - 一个用于构建、修改和可视化神经网络的JAX工具包。
95. microsoft/i-Code -
96. kubeflow/training-operator - 在Kubernetes上进行分布式机器学习训练和微调。
97. AetherCortex/Llama-X - 开展使LLaMA达到最先进大型语言模型的开放学术研究。
98. salesforce/ALBEF - 一种新的视觉 - 语言预训练方法ALBEF的代码。
99. kubeflow/katib - Kubernetes环境下的自动化机器学习。
100. facebookresearch/multimodal - TorchMultimodal是一个PyTorch库，用于对最先进的多模态多任务模型进行大规模训练。
101. jina-ai/finetuner - 用于BERT、CLIP等的面向任务的嵌入调整。
102. salesforce/CodeTF - CodeTF：用于最新代码大型语言模型（LLM）的一站式Transformer库。
103. AnswerDotAI/fsdp_qlora - 使用QLoRA和全分片数据并行（FSDP）训练大型语言模型（LLM）。
104. nerfstudio-project/nerfacc - 一个基于PyTorch的通用神经辐射场（NeRF）加速工具箱。

105. jquesnelle/yarn - YaRN：大型语言模型的高效上下文窗口扩展。
106. PKU-Alignment/safe-rlhf - 安全的人类反馈强化学习（Safe RLHF）利用基于人类反馈的安全强化学习来实现受限的值对齐。
107. lucidrains/self-rewarding-lm-pytorch - 实现MetaAI在《自奖励语言模型》中提出的训练框架。
108. OpenLMLab/MOSS-RLHF - 这是关于大型语言模型中人类反馈强化学习（RLHF）的秘密，特别是关于近端策略优化算法（PPO）的第一部分。
109. JonasGeiping/cramming - 在有限的计算资源内压缩BERT类型语言模型的训练。
110. AlibabaResearch/DAMO-ConvAI - 达摩 - 对话式人工智能（ConvAI）是拥有阿里巴巴达摩院对话式人工智能（DAMO Conversational AI）代码库的官方资源库。
111. databricks/megablocks -
112. AGI-Edgerunners/LLM-Adapters - EMNLP 2023论文《LLM - 适配器：用于大型语言模型参数高效微调的适配器家族》的代码。
113. KhoomeiK/LlamaGym - 通过在线强化学习对大型语言模型（LLM）智能体进行微调。
114. thunlp/OpenDelta - 一个即插即用的参数高效调整（Delta Tuning）库。
115. Liuhong99/Sophia - 论文《Sophia：一种用于语言模型预训练的可扩展随机二阶优化器》的官方实现。
116. yuchenlin/LLM-Blender - LLM - Blender是[ACL2023]中的一个集成框架，它通过排序来消除弱点，并通过生成融合优势以提升大型语言模型（LLMs）的能力。
117. google-deepmind/xmanager - 一个用于处理机器学习实验的平台。
118. google-deepmind/chex -

AI图像生成

1. AUTOMATIC1111/stable-diffusion-webui - 稳定扩散Web用户界面。
2. lllyasviel/Fooocus - 描述内容是关于专注于提示和生成。
3. upscayl/upscayl - Upscayl是适用于Linux、MacOS和Windows系统的排名第一的免费开源人工智能图像放大器。
4. s0md3v/roop - 一键换脸意味着只需点击一下就可以完成换脸。
5. invoke-ai/InvokeAI - Invoke是一个用于Stable Diffusion模型的创意引擎，它提供了一个WebUI（用户界面）并作为商业产品的基础。
6. facefusion/facefusion - 一个行业领先的面部处理平台。
7. Sanster/lama-cleaner - 由最先进的人工智能模型驱动的图像修复工具可以去除图片中不需要的元素或替换其中的事物。
8. Mikubill/sd-webui-controlnet - 用于ControlNet的WebUI扩展。

9. camenduru/stable-diffusion-webui-colab - 在Colab（谷歌协作平台）上的Stable diffusion（稳定扩散）网页用户界面。
10. divamgupta/diffusionbee-stable-diffusion-ui - Diffusion Bee是在M1 Mac本地运行Stable Diffusion的最简单方法。它有一个一键安装程序，不需要依赖项或技术知识。
11. Baiyuetribe/paper2gui - 将人工智能论文转换为图形用户界面（GUI），以便每个人都能轻松使用人工智能技术。
12. easydiffusion/easydiffusion - 在PC端利用人工智能创作精美艺术作品最简单的方法是一键操作。只需输入文本提示词，就能通过浏览器用户界面生成图像，无需技术知识。
13. Stability-AI/StableStudio - 生成式人工智能的社区界面。
14. carson-katri/dream-textures - Stable Diffusion集成到Blender中。
15. TheLastBen/fast-stable-diffusion - 快速 - 稳定 - 扩散和DreamBooth。
16. godly-devotion/MochiDiffusion - 在Mac上本地运行Stable Diffusion。
17. HumanAIGC/OutfitAnyone - 超高质量的适用于所有人和衣物的虚拟试穿。
18. sensity-ai/dot - 深度伪造攻击工具包。
19. leap-ai/headshots-starter -
20. Nutlope/restorePhotos - 使用人工智能修复模糊的旧人脸照片。
21. jina-ai/discoart - 一行创建迪斯科扩散艺术作品。
22. mlc-ai/web-stable-diffusion - 将稳定扩散模型引入网络浏览器；无需服务器支持，一切都在浏览器中运行。
23. all-in-aigc/aicover - 一个用于生成人工智能封面的工具。

数据集

1. huggingface/datasets - 最大的机器学习模型数据集中心拥有快速、易用且高效的数据操作工具。
2. BuilderIO/gpt-crawler - 从一个网址抓取网站内容以生成创建自定义GPT所需的知识文件。
3. joke2k/faker - Faker是一个用于生成虚假数据的Python包。
4. DS4SD/docling - 为生成式人工智能准备好你的文档。
5. openai/tiktoken - Tiktoken是适用于OpenAI模型的一种快速的BPE分词器。
6. cleanlab/cleanlab - 标准的人工智能包专注于用于质量和机器学习的数据，处理杂乱的现实世界数据和标签。
7. karpathy/minbpe - 用于大型语言模型（LLM）标记化中常用的字节对编码（BPE）算法的简洁代码。
8. huggingface/tokenizers - 快速、先进的分词器，为研究和生产进行了优化。
9. arsenetar/dupeguru - 查找重复文件。

10. QuivrHQ/MegaParse - 优化的文件解析器，用于无损的大型语言模型（LLM）摄入，能够将PDF、Docx、PPTx文件解析成适合大型语言模型（LLM）的理想格式。
11. togethercomputer/RedPajama-Data - RedPajama - 数据存储库中有用于为大型语言模型训练准备大型数据集的代码。
12. lk-geimfari/mimesis - Mimesis是一个Python数据生成器，能够创建多种语言的各类虚假数据。
13. Instruction-Tuning-with-GPT-4/GPT-4-LLM - 使用GPT - 4进行指令微调。
14. yizhongw/self-instruct - 将预训练语言模型与其自身生成的指令数据进行对齐。
15. dedupeio/dedupe - 一个用于精确和可扩展的模糊匹配、记录去重和实体解析的Python库。
16. argilla-io/argilla - Argilla是一种可供人工智能工程师和领域专家共同创建高质量数据集的工具。
17. mshumer/gpt-llm-trainer -
18. life4/textdistance - 使用30多种纯Python算法、通用接口以及可选择使用外部库来计算序列之间的距离。
19. Docta-ai/docta - 一位医生来照管你的数据。
20. alibaba/data-juicer - 为基础模型提供高质量、丰富且易于处理的数据。
21. towhee-io/towhee - Towhee是一个使神经数据处理管道变得简单快速的框架。
22. QData/TextAttack - TextAttack是一个用于自然语言处理（NLP）任务（如对抗性攻击、数据增强和模型训练）的Python框架。
23. seatgeek/thefuzz - Python中的模糊字符串匹配。
24. ekzhu/datasketch - 最小哈希、局部敏感哈希、局部敏感哈希森林、加权最小哈希、超对数、超对数++、局部敏感哈希集成和分层可导航小世界图。
25. thunlp/UltraChat - 大规模、信息丰富且多样的多轮聊天数据和模型。
26. modAL-python/modAL - 一个用Python编写的模块化主动学习框架。
27. chiphuyen/lazynlp - 一个用于爬取和清理网页以生成大型数据集的库。
28. huggingface/datatrove - 提供一组与平台无关的可定制管道处理模块，使数据处理不再依赖脚本编写。
29. refuel-ai/autolabel - 使用大型语言模型对文本数据集进行标记、清理和扩充。
30. google-deepmind/code_contests -
31. Tencent/MedicalNet - 许多研究表明，训练数据量会显著影响深度学习的性能。MedicalNet项目提供3D - ResNet预训练模型和代码。
32. argilla-io/distilabel - Distilabel是一个合成数据和人工智能反馈框架，适用于需要基于经过验证的研究论文构建快速、可靠且可扩展管道的工程师。
33. google-deepmind/mathematics_dataset - 该数据集代码从各种题型中创建学校难度级别的数学问答对。

34. openai/prm800k - 80万条针对大型语言模型（LLM）解决数学问题的答案的步骤级正确性标签。
35. salesforce/WikiSQL - 一个用于带有语义分析注释的自然语言界面开发的大型语料库。
36. anthropics/hh-rlhf - 用于训练通过人类反馈强化学习的助手（使其有用且无害）的人类偏好数据。
37. moj-analytical-services/splink - 快速、可扩展且精确的概率性数据链接，支持多种SQL后端。
38. dleemiller/WordLlama - 大型语言模型（LLM）的标记嵌入可完成的事情。
39. AI4Finance-Foundation/FinRL-Meta - FinRL - Meta为FinRL提供动态数据集和市场环境。
40. tensorflow/text - 将文本作为TensorFlow中的首要元素。
41. google-research/deduplicate-text-datasets -
42. allenai/dolma - 用于创建和检查OLMo预训练数据的数据和工具。
43. lilacai/lilac - 改进大型语言模型的数据管理。
44. 1e0ng/simhash - 一种Simhash算法的Python实现。
45. J535D165/recordlinkage - 一个用于记录链接和重复数据检测的Python工具包，功能强大且模块化。
46. google-deepmind/tree - Tree是一个用于处理嵌套数据结构的库。
47. xtreme1-io/xtreme1 - Xtreme1是一个多模态数据训练一体化平台，支持3D激光雷达点云、图像和大型语言模型（LLM）进行数据标记和注释。
48. datadreamer-dev/DataDreamer - DataDreamer：提示、生成合成数据、训练和校准模型。
49. HazyResearch/meerkat - 所有数据集的创意性和交互性视图。

模型评估

1. openai/evals - Evals是一个用于评估大型语言模型（LLMs）及其系统的框架，也是一个开源的基准测试注册中心。
2. explodinggradients/ragas - 为你对大型语言模型（LLM）应用的评估注入强大动力。
3. EleutherAI/lm-evaluation-harness - 一种使用少量样本评估语言模型的框架。
4. erikbern/ann-benchmarks - Python中近似最近邻库的基准测试。
5. Trusted-AI/adversarial-robustness-toolbox - 对抗性鲁棒性工具包（ART）是一个用于机器学习安全的Python库，涵盖了红蓝两队（攻防双方）的规避、投毒、提取和推理（攻击）。
6. open-compass/opencompass - OpenCompass是一个大型语言模型（LLM）评估平台，支持100多个数据集以及诸如Llama3、Mistral等各种各样的模型。
7. Arize-ai/phoenix - 人工智能可观察性与评估。
8. NVIDIA/NeMo-Guardrails - NeMo Guardrails是一个开源工具包，可轻松为基于大型语言模型（LLM）的对话系统添加可编程的防护栏。

9. confident-ai/deepeval - 大语言模型（LLM）评估框架。
10. Giskard-AI/giskard - 人工智能与大型语言模型（LLM）系统的开源评估与测试。
11. fchollet/ARC - 抽象推理语料库是一种资源，但没有更多细节的情况下，很难给出更具体的描述。它可能可用于各个领域中与抽象和推理相关的任务。
12. llm-attacks/llm-attacks - 对齐语言模型的通用可迁移攻击。
13. leondz/garak - 大型语言模型（LLM）漏洞扫描器。
14. jeinlee1991/chinese-llm-benchmark - 中国大模型能力评估清单包含134个模型，包括商业和开源模型，并提供能力评分和原始输出。
15. google/BIG-bench - 超越模仿游戏：一个用于衡量和推断语言模型能力的协作性基准。
16. meta-llama/PurpleLlama - 一组用于评估和增强大型语言模型（LLM）安全性的工具。
17. openai/human-eval - 论文《评估基于代码训练的大型语言模型》的代码
18. salesforce/decaNLP - 自然语言十项全能（竞赛）是自然语言处理领域的一项多任务挑战。
19. THUDM/AgentBench - 一个用于评估作为智能体的大型语言模型（LLMs）的综合基准（国际学习表征会议ICLR'24）。
20. truera/trulens - 大型语言模型（LLM）实验中的评估与跟踪。
21. princeton-nlp/SWE-bench - [国际学习表征会议（ICLR）2024] SWE - bench：语言模型能否解决现实世界中的GitHub（代码托管平台）问题。
22. Lightning-AI/torchmetrics - 适用于分布式且可扩展的PyTorch应用程序的机器学习指标。
23. openai/simple-evals -
24. huggingface/evaluate - 一个名为Evaluate的库，用于轻松评估机器学习模型和数据集。
25. embeddings-benchmark/mteb - MTEB是大规模文本嵌入基准测试。
26. Azure/PyRIT - PyRIT是一个开源框架，可帮助安全专业人员和工程师主动识别生成式人工智能系统中的风险。
27. stanford-crfm/helm - HELM是一个用于提高语言模型透明度的框架，它还用于评估其他模型，如HEIM中的文本到图像模型和VHELM中的视觉 - 语言模型。
28. TransformerLensOrg/TransformerLens - 一个用于类GPT语言模型的机械可解释性的库。
29. beir-cellar/beir - 一个异构信息检索（IR）基准，用于轻松评估15个以上不同数据集的模型。
30. tatsu-lab/alpaca_eval - 一种指令遵循型语言模型的自动评估器。它经过人工验证，质量高、成本低且速度快。
31. microsoft/CodeXGLUE - CodeXGLUE是一个项目或者实体，但没有更多的上下文信息的话，很难说得更具体。
32. google-deepmind/bsuite - Bsuite是一组精心设计的实验，用于探索强化学习智能体的核心能力。

33. CalculatedContent/WeightWatcher - WeightWatcher工具用于预测深度神经网络的准确性。
34. facebookresearch/LAMA - 语言模型分析。
35. evalplus/evalplus - 在NeurIPS 2023和COLM 2024对大型语言模型（LLM）合成的代码进行严格评估。
36. vectara/hallucination-leaderboard - 用于比较大型语言模型（LLM）在总结短文时产生幻觉表现的排行榜。
37. hendrycks/test - 在2021年国际学习表征会议（ICLR）上衡量大规模多任务语言理解能力。
38. mlcommons/inference - MLPerf™推理基准的参考实现。
39. openai/grade-school-math -
40. rlancemartin/auto-evaluator - 大型语言模型问答链的评估工具。
41. openai/automated-interpretability -
42. allenai/natural-instructions - 该描述是关于扩展自然指令的。
43. WeOpenML/PandaLM -
44. thu-coai/Safety-Prompts - 用于评估和提升大型语言模型安全性的中国安全提示。
45. salesforce/OmniXAI - OmniXAI是一个可解释人工智能（XAI）库。
46. bigcode-project/bigcode-evaluation-harness - 用于评估自回归代码生成语言模型的框架。
47. hsiehjackson/RULER - 这个存储库包含RULER（关于长上下文语言模型的真实上下文大小）的源代码。

计算管理

1. kubeflow/kubeflow - 一个专为Kubernetes设计的机器学习工具包。
2. Netflix/metaflow - 一个人工智能与机器学习的开源平台。
3. skypilot-org/skypilot - SkyPilot能够在任何基础设施（Kubernetes或12种以上的云）上运行AI和批处理作业，通过一个简单的接口提供统一的执行、成本节约和高GPU可用性。
4. gpuweb/gpuweb - 这是网络中GPU（图形处理器）工作的地方。
5. zenml-io/zenml - ZenML：机器学习与操作之间的联系。
6. higgsfield-ai/higgsfield - 容错且高度可扩展的GPU编排，以及用于训练大规模模型的机器学习框架。
7. Haidra-Org/AI-Horde - 一个用于生成人工智能艺术和文本的众包分布式集群。

AI写作

1. steven-tey/novel - 一个所见即所得且具有人工智能驱动的自动补全功能的类似Notion的编辑器。
2. reorproject/reor - 一款面向高熵人群的、私密且本地化的人工智能个人知识管理应用。

3. shibing624/pycorrector - Pycorrector是一个文本纠错工具包，有多种用于纠错的模型应用且易于使用。
4. BlinkDL/AI-Writer - 人工智能创作诸如奇幻和浪漫网络小说之类的小说。它是一个类似于GPT - 2、使用RWKV模型的中文预训练生成模型。
5. mshumer/gpt-author -
6. Nutlope/twitterbio - 使用人工智能创建你的推特简介。
7. nhaouari/obsidian-textgenerator-plugin - Text Generator是一个Obsidian插件，可用于与OpenAI、Anthropic、Google等各种人工智能供应商以及本地模型一起生成文本。
8. google-deepmind/dramatron - Dramatron利用大型语言模型来生成连贯的脚本和剧本。

智能体监控

1. nebuly-ai/nebuly - 一组用于优化人工智能模型性能的库。
2. langfuse/langfuse - 具有多种集成（用于大型语言模型可观测性、指标等）的开源大型语言模型工程平台。它来自于 YC W23。
3. evidentlyai/evidently - Evidently是一个用于机器学习（ML）和大型语言模型（LLM）可观测性的开源框架，可用于评估、测试和监控与人工智能相关的系统或数据管道，它拥有100多项指标。
4. traceloop/openllmetry - 依靠OpenTelemetry（开放遥测）为您的大型语言模型（LLM）应用提供开源可观测性。
5. Helicone/helicone - 一个开源的大型语言模型（LLM）可观测性平台。一行代码即可用于监控、评估和实验。
6. whylabs/whylogs - 一个用于记录机器学习模型和管道中数据的开源库。它提供数据质量和模型性能的可见性，以及受隐私保护的数据收集。
7. uptrain-ai/uptrain - UpTrain是一个用于评估和改进生成式人工智能应用的开源平台。它为预配置的检查提供评分，分析故障并给出解决方案。
8. labmlai/labml - 通过手机监控深度学习模型训练和硬件使用情况。
9. lmnr-ai/lmnr - Laminar是一个用于构建人工智能产品的开源一体化平台。它通过跟踪、评估、数据集和标签为人工智能应用创建数据飞轮（YC S24）。
10. llmonitor/llmonitor - 大语言模型（LLMs）的生产工具包涉及可观测性、提示管理和评估。
11. lunary-ai/lunary - 大型语言模型（LLMs）的生产工具包包括可观测性、提示管理和评估。
12. dillionverma/llm.report - llm.report是一个针对OpenAI的开源平台，用于记录API请求、成本分析和提示优化。
13. whylabs/langkit - LangKit是一个用于大型语言模型（LLM）监控的开源工具包。它具有用于LLM可观测性的文本质量和情感分析等功能。

视频生成

1. RayVentura/ShortGPT - ShortGPT——一个用于自动化运营YouTube Shorts（短视频）/TikTok（抖音国际版）频道的实验性AI框架。
2. all-in-aigc/sorafm - 由Sora.FM提供的Sora人工智能视频生成器。

数据管理

1. ibis-project/ibis - 可移植的Python数据框库。
2. SuperDuperDB/superduperdb - Superduper能够在现有数据基础设施和首选工具上构建端到端的人工智能应用程序和代理工作流，无需进行数据迁移。
3. run-llama/llama-hub - 由社区制作的数据加载器库，用于LlamaIndex和/或LangChain的大型语言模型（LLM）。
4. webdataset/webdataset - 一个适用于各种规模深度学习问题的基于Python的输入输出（I/O）系统，强力支持PyTorch。
5. NVIDIA/aistore - AIStore：可扩展的AI应用程序存储。
6. mosaicml/streaming - 一种用于高效神经网络训练的数据流库。