

[英]詹姆斯·菲尼克斯 (James Phoenix)

[美]迈克·泰勒 (Mike Taylor) 著

孙霄逸 译

大模型提示 工程实战 链接资源

【致读者】本书包含的 400 条左右的链接资源，由 AI 辅助整理生成，以电子版 PDF 形式免费提供。由于网络环境不断变化，个别链接可能访问异常；相关描述文字也系 AI 生成，可能存在不精准之处。本资料仅供参考，敬请知悉。

2026 年 2 月

目录

前言 1

第 1 章 3

第 2 章 12

第 3 章 18

第 4 章 23

第 5 章 29

第 6 章 34

第 7 章 38

第 8 章 46

第 9 章 52

第 10 章 66

关于作者 71

前言

1. <https://colab.research.google.com>: Google Colab 的官方网站，提供基于云端的 Jupyter Notebook 环境，允许用户在线编写、运行和共享 Python 代码，适合本书中代码示例的执行和实践。
2. <https://incremen.to>: Incremen.to，一家由 Ellis Crosby 创办的人工智能代理机构。
3. <https://jupyter.org>: Jupyter Notebook 的官方网站，提供一个基于浏览器的交互式计算环境，支持运行和共享包含代码、文本、公式和可视化内容的笔记本，非常适合用于数据分析、教学和科研等编程实践。
4. <https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/requirements.txt>: 本书前言中提到的 requirements.txt 文件，列出了运行相关代码示例所需安装的 Python 依赖包及其版本。用户应在虚拟环境中安装该文件中指定的所有包，以确保代码示例能够正常运行且环境配置正确。
5. <https://llama.meta.com/llama3>: Meta 公司开源发布的 Llama 3 模型，属于大型语言模型（LLM），用于自然语言处理任务，是 Meta 在人工智能领域的重要开源项目之一。
6. <https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/>: 本书的 GitHub 仓库，本书中所有的 Python 代码示例，这些代码设计用于在 Jupyter Notebook 或 Google Colab 环境中运行。该资源主要聚焦于生成式人工智能中的提示工程技术，提供相关的示范和实践代码，方便读者边阅读书籍边动手操作学习。
7. <https://twitter.com/MarioKrenn6240/status/1577102743927652354>: 由用户 Mario Krenn 发布在 X 上的推文，指出 arXiv 上 AI 相关论文数量的指数式增长。
8. https://x.com/kylelf_/status/1623679176246185985: 由用户 KyleLF_ 发布在 X 上的推文，展现了 ChatGPT 用户数量的增长。

9. https://huggingface.co/blog/stable_diffusion: 介绍 Stable Diffusion 模型的博客文章，内容涵盖 Stable Diffusion 的功能、使用方法及其在图像生成领域的应用。
10. <https://www.forbes.com/sites/craigsmith/2023/04/05/mom-dad-i-want-to-be-a-prompt-engineer/>: 福布斯网站的文章，标题为“Mom, Dad, I Want to Be a Prompt Engineer”，强调了提示工程作为提升人工智能模型输出质量的一项关键技能的重要性，前言中李彦宏的引文来自此文。
11. <https://gemini.google.com/>: Google 推出的大模型 Gemini 的官方网站。
12. <https://twitter.com/a16z/status/1592922394275872768?s=20>: 用户 a16z 发布在 X 上的推文，描述了 Stable Diffusion 用户数量的快速增长。
13. <https://stability.ai/stable-diffusion>: Stability AI 官方网站上关于 Stable Diffusion 的页面，介绍了 Stable Diffusion 这款开源的图像生成模型的信息、相关工具和使用方式，包括通过 API 调用或在本地 GPU 设备上运行的选项。
14. <https://learning.oreilly.com/library/view/xsl-fo/0596003552>: O’ Reilly 出版的关于 XSL-FO（可扩展样式表语言格式对象）的书的页面，作者为 Dave Pawson，内容主要介绍 XSL-FO 的使用和相关技术。
15. <https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/requirements.in>: 本书前言中提到的名为 requirements.in 的依赖配置文件，位于本书的 GitHub 仓库中。它用于定义项目所需的 Python 包及其版本，从该文件可以通过命令生成对应特定 Python 版本（如 Python 3.9）的 requirements.txt 文件。
16. <https://platform.openai.com/docs/quickstart?context=python>: OpenAI 官方提供的快速入门指南，介绍如何使用 Python 语言调用 OpenAI API，包括如何设置开发环境、配置 API 密钥以及运行基础示例代码，帮助开发者快速开始使用 OpenAI 平台的功能。
17. https://twitter.com/hammer_mt/status/1657730978125869056?s=20: 用户 Mike Taylor 发布在 X 上的推文，描述了 Midjourney 用户数量的快速增长。

18. <https://platform.stability.ai>: Stability AI 开发者平台主页, 主要涉及 Stable Diffusion 模型的 API 接口和使用支持, 供用户通过云端调用或集成其图像生成技术。
19. <https://www.anthropic.com/news/claude-3-family>: Anthropic 公司的 Claude 3 系列模型的官方介绍页面。
20. <https://greylock.com/greymatter/sam-altman-ai-for-the-next-era/>: 对 OpenAI 联合创始人 Sam Altman 的采访, Sam Altman 在其中指出, 真正有效利用 AI 的关键在于“想法的质量和对你所追求的目标的深刻理解”。
21. <https://learning.oreilly.com/library/view/prompt-engineering-for/9781098153427/>: 本书英文版的主页, 还提供了该书的勘误、示例及附加信息, 方便读者参考和学习。
22. <https://www.midjourney.com>: Midjourney 的官方网站, Midjourney 是一款基于 Discord 平台的图像生成工具, 用户需通过 Discord 账号注册使用。它提供先进的 AI 图像生成服务, 常用于生成高质量的艺术和创意图片。
23. <https://www.scarlettpanda.com>: Scarlett Panda 的官方网站, Scarlett Panda 是由 Ellis Crosby 创立的公司, 专注于人工智能生成儿童读物。
24. <https://www.mayooshin.com>: 早期 LangChain 贡献者 Mayo Oshin 的个人主页。

第 1 章

1. <https://chat.openai.com>: 该资源是 OpenAI 提供的 ChatGPT 在线聊天平台的网址, 用户可以通过该链接访问 ChatGPT, 以文本形式与基于大型语言模型 (LLMs) 的人工智能进行交互, 获取自然语言生成的回答和帮助。

2. <https://discord.com>: 该资源是 Discord 的官方网站, 指向一个提供即时通讯和社区交流服务的平台。用户需要注册一个免费的 Discord 账号, 以便使用 Midjourney Bot 进行图像生成等交互操作。
3. <https://openai.com/pricing>: 该资源为 OpenAI 的官方定价页面, 详细说明了基于使用的“tokens”(提示和响应中的词元数量)计费的具体方案和价格标准, 帮助用户了解使用模型时的费用结构及优化提示以控制成本的方法。
4. <https://oreil.ly/0MZ3->: (<https://www.deeplearning.ai/short-courses/chatgpt-prompt-engineering-for-developers/>) 该资源是由 DeepLearning.AI 提供的一个面向开发者的短期课程, 主题为“ChatGPT 提示工程”(ChatGPT Prompt Engineering)。课程内容聚焦于如何设计和优化针对 ChatGPT 等大型语言模型的提示语, 以提升模型输出的质量和实用性, 涵盖诸如“链式思维”(chain of thought)和“元提示”(meta prompting)等关键技术和方法, 帮助开发者更高效地利用 AI 模型完成复杂任务。
5. <https://oreil.ly/1lqzK>: (<https://arxiv.org/abs/2110.08387>) 该 URL (<https://arxiv.org/abs/2110.08387>) 指向的是一篇由 Liu 等人在 2021 年发表的学术论文, 内容涉及通过“预热”(prewarming)或“内部检索”(internal retrieval)等技术改进 AI 提示词设计的方法。这些技术通过在提示中包含规则或最佳实践作为上下文, 引导 AI 生成更合适的输出, 从而提升生成效果。论文详细探讨了如何利用这种策略让 AI 系统更好地遵循自身建议, 从而优化交互和结果。
6. <https://oreil.ly/2KDV6>: (<https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/requirements.txt>) 该资源是一个名为 requirements.txt 的依赖文件, 通常用于 Python 项目中列出所需安装的第三方库及其版本。根据上下文, 该文件位于 GitHub 仓库“BrightPool/prompt-engineering-for-generative-ai-examples”中, 提供了安装环境所需的依赖包列表, 帮助用户解决 OpenAI Python 包的兼容性问题, 便于用户通过创建虚拟环境并安装该文件中的依赖来保证项目环境的稳定与一致。
7. <https://oreil.ly/3TTZ9>: (<https://openai.com/blog/new-models-and-developer-products-announced-at-devday>) 该 URL 指向 OpenAI 在其官方博客发布的一篇文章, 内容介绍了 OpenAI 在开发者大会(Dev Day)上宣布的

新模型和开发者产品，重点展示了其在大规模语言模型（如 GPT-4）上下文窗口扩展及相关技术进展方面的最新成果和应用。

8. <https://oreil.ly/3bWjz>: (<https://www.brandwatch.com/blog/how-to-name-a-product-our-5-golden-rules>) 该资源是 Brandwatch 网站上的一篇博客文章，标题为《如何为产品命名：我们的 5 条黄金法则》。文章内容介绍了为产品命名时的五个关键原则，旨在帮助读者制定有效且有影响力的产品名称策略。
9. <https://oreil.ly/42rSz>: (<https://mitchellh.com/writing/prompt-engineering-vs-blind-prompting>) 该资源是一篇由 Mitchell Hashimoto 撰写的文章，标题为《Prompt Engineering vs Blind Prompting》。文章探讨了“盲目提示”（blind prompting）与“提示工程”（prompt engineering）之间的区别，重点强调了在使用提示（prompt）时评估和提升响应质量的重要性。文中指出，盲目提示通常指通过反复试验来运行提示并观察结果的方式，这种方式适用于临时或一次性的任务；而对于需多次复用或应用于生产环境的提示，则需要更严谨的质量评估和优化方法。
10. <https://oreil.ly/48lq6>: (<https://github.com/reworkd/AgentGPT>) 该资源是一个开源的 AI 自主代理项目，利用“Reason and Act (ReAct)”框架，通过将多个 AI 调用串联起来，实现规划、观察、行动和评估的闭环，从而使 AI 能够自主完成复杂任务。该项目旨在探索和推动自主智能体的发展，尽管目前尚处于早期阶段且存在一定错误率，但展示了在 AI 系统进化中的潜力。
11. <https://oreil.ly/6Ixcw>: (<https://arxiv.org/pdf/2308.00675.pdf>) 该资源是论文《Hsieh et al., 2023》，讨论了在人工智能生成任务中，与直接提供示例相比，给予明确指导（direction）的方法更有效。这篇论文探讨了通过指导而非示例来提升模型表现的策略，特别是在生成多样性和处理边缘案例方面的优势，适用于优化 AI 系统的输入设计和性能提升。
12. <https://oreil.ly/DZ-br>: (<https://aclanthology.org/2021.naacl-main.208.pdf>) 该资源是 2021 年在 NAACL 主会议上发表的一篇论文，地址为 <https://aclanthology.org/2021.naacl-main.208.pdf>。根据上下文推断，该论文内容可能涉及通过大量示例（数千个）进行模型微调，从而超越传统的提示工程方法，提升自然语言处理系统的性能和智能化水平。

13. <https://oreil.ly/Dwszu>: (<https://arxiv.org/abs/2211.01910>) 该 URL (<https://arxiv.org/abs/2211.01910>) 指向的是一篇关于“链式思维 (chain of thought)”及“元提示 (meta prompting)”技术的学术论文。这篇论文探讨了如何通过让大型语言模型 (LLM) 分步骤思考和生成提示, 从而提升模型在复杂任务中的推理能力和表现。论文内容涉及利用模型自身生成提示以引导后续生成, 体现了语言模型作为“人类级提示工程师”的能力。
14. <https://oreil.ly/E7wua>: (<https://community.openai.com/t/openai-api-guide-using-json-mode/557265>) 该 URL (<https://community.openai.com/t/openai-api-guide-using-json-mode/557265>) 指向的是 OpenAI 社区中的一篇指南, 内容主要介绍如何在使用 OpenAI API 时启用和利用“JSON 模式”输出。该指南详细讲解了如何让模型以结构化的 JSON 格式返回数据, 方便程序化解析和处理, 以及如何通过提示词或系统消息确保输出符合预期的 JSON 格式, 从而提高调用 API 时数据的准确性和可靠性。
15. <https://oreil.ly/Fc8cq>: (<https://platform.openai.com/examples/default-product-name-gen>) 该资源是 OpenAI 平台上一个用于生成产品名称的示例项目, 展示了如何通过提示词设计来自动生成创意产品名称。
16. <https://oreil.ly/IaQjS>: (<https://pypi.org/project/python-dotenv/>) 该资源是 Python 的一个名为 python-dotenv 的库, 通常用于从 .env 文件中加载环境变量, 方便在开发和部署过程中管理配置参数。链接指向其在 PyPI (Python 包索引) 上的项目主页, 用户可以在该页面查看库的详细信息、安装方法和使用说明。
17. <https://oreil.ly/J4Hkr>: (https://unsplash.com/photos/vdXMSiX-n6M?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText) 该资源是一张由摄影师 Mimi Thian 拍摄并发布在图片分享网站 Unsplash 上的照片。
18. <https://oreil.ly/KG5Xl>: (<https://microsoft.github.io/autogen/>) 该资源是微软开源的 AutoGen 项目, 属于基于 Reason and Act (ReAct) 框架的自主 AI 代理工具。它通过串联多次 AI 调用, 实现规划、观察、行动和评估的闭环过程, 旨在帮助构建能够自我推理和执行复杂任务的智能代理系统。

-
19. <https://oreil.ly/KGAqe>: (<https://simonwillison.net/2022/Sep/12/prompt-injection/>) 该资源是一篇关于“prompt injection”（提示注入）攻击的文章，介绍了如何防范这类攻击以避免模型执行非预期或有害的提示内容。具体来说，文章可能讲解了提示注入的原理、风险，以及如何设计提示使其在面对已知的提示注入攻击时依然保持稳健和安全。
 20. <https://oreil.ly/KW5PS>: (<https://arxiv.org/abs/2005.14165>) 该 URL (<https://arxiv.org/abs/2005.14165>) 指向的是著名的学术论文《Language Models are Few-Shot Learners》（语言模型是少样本学习者），该论文介绍了 GPT-3 模型及其在零样本（zero-shot）、一样本（one-shot）和少样本（few-shot）学习任务中的表现，展示了通过在提示中添加少量示例，语言模型的任务准确率能够显著提升。
 21. <https://oreil.ly/NQcFW>: (<https://www.anthropic.com/news/claude-2>) 该资源介绍了 Anthropic 公司发布的 Claude 2 模型，重点突出其支持高达 10 万 token 的超大上下文窗口能力，这使其在处理长文本时具有显著优势。
 22. <https://oreil.ly/NW3WX>: (<https://lmsys.org/blog/2023-03-30-vicuna/>) 该资源是关于 Vicuna-13B 模型发布的介绍，Vicuna-13B 是基于 Meta 开源的 Llama 模型进行微调的一个语言模型。文中提到，研究人员在评估较为主观或复杂的任务时，会利用更先进的模型（如 GPT-4）来评价较简单模型的表现，而 Vicuna-13B 的发布就是一个相关案例。该链接内容大概率详细介绍了 Vicuna 模型的训练方法、性能表现及其在自然语言处理中的应用。
 23. <https://oreil.ly/P2IcU>: (<https://chat.lmsys.org/?arena>) 该资源是由 lmsys.org 提供的“Chatbot Arena”平台，用户可以在该平台上通过对聊天机器人回复进行多级评分（如 3、5 或 10 分制）以及进行对比评价，从而实现基于 Elo 等级体系的不同模型性能的公平比较与排名。
 24. <https://oreil.ly/RYYiV>: (<https://www.saxifrage.xyz/post/prompt-engineering>) 该资源是一篇关于“提示工程”（Prompt Engineering）的博文，内容介绍了提示工程的基本原则和通用策略，这些原则适用于各种生成式文本或图像模型，帮助用户优化与人工智能模型的交互效果。文章强调这些原则是

经过时间考验的通用规范，而非短期技巧，并且与 OpenAI 发布的提示工程指南有较高的一致性，适合任何层级的智能系统使用。

25. <https://oreil.ly/TEiQx>: (<https://github.com/yoheinakajima/babyagi>) 该资源是一个名为 BabyAGI 的开源项目，基于 Reason and Act (ReAct) 框架，旨在通过多次调用人工智能模型，实现自主规划、观察、执行和评估动作，从而构建能够完成复杂任务的自主 AI 代理。
26. <https://oreil.ly/TlIdE>: (https://en.wikipedia.org/wiki/Elo_rating_system) 该 URL 指向的资源是维基百科中关于“Elo 评级系统” (Elo rating system) 的条目页面。该页面介绍了一种用于评估棋类、体育竞技等对抗性比赛中选手或队伍相对实力的数学评级方法，由阿帕德·埃洛 (Arpad Elo) 提出。页面内容通常包括系统原理、计算方法、应用范围及历史背景等信息。
27. <https://oreil.ly/UTxpX>: (<https://docs.midjourney.com/docs/image-prompts>) 该资源是 Midjourney 官方文档中关于“图像提示 (image prompts)”的说明页面，介绍了如何在 Midjourney 中使用图像 URL 作为提示输入，包括上传基础图像到 Discord 并将其 URL 粘贴到提示中，从而利用 Midjourney 基于该图像生成内容的功能。页面可能还包含具体的格式要求和使用示例，帮助用户正确地使用图像提示以获得理想的生成效果。
28. <https://oreil.ly/V40zg>: (<https://www.udemy.com/course/prompt-engineering-for-ai>) 该资源是 Udemy 上的一门热门课程，名为“The Complete Prompt Engineering for AI Bootcamp”，专注于教授文本和图像生成领域的提示工程 (Prompt Engineering) 技巧和原则，旨在帮助学习者掌握如何有效设计和优化用于人工智能生成模型的提示语。课程拥有超过 7 万名学员，提供实用的可下载一页纸总结材料，方便学员作为应用提示工程原则时的参考和检查清单。
29. <https://oreil.ly/VCcgy>: (<https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/images/OnePager-Text.png>) 该资源是一个关于文本生成的“一页概览”图示 (One-Pager)，用以简明扼要地展示文本生成技术或方法的核心内容和要点。

-
30. <https://oreil.ly/Xakgc>: (https://www.w3schools.com/js/js_json_intro.asp) 该资源是 W3Schools 提供的一个关于 JSON (JavaScript 对象表示法) 的入门介绍页面, 适合初学者了解 JSON 的基本概念和用法。
 31. <https://oreil.ly/ZbzZy>: (https://unsplash.com/photos/Oalh2MojUuk?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText) 该资源是一张由摄影师 Jason Goodman 拍摄并发布在图片分享平台 Unsplash 上的照片。
 32. <https://oreil.ly/cyhR4>: (<https://blog.google/technology/ai/google-gemini-next-generation-model-february-2024/>) 该 URL 指向的资源是关于谷歌最新一代人工智能模型“Gemini”的介绍, 发布时间为 2024 年 2 月。文章详细阐述了在大型科技公司(如微软和谷歌)以及开源项目和创业公司的 AI 竞赛背景下, 谷歌推出的 Gemini 模型在上下文窗口大小上的重大突破, 支持高达 100 万 token 的上下文长度, 远超其他主流模型(如 OpenAI 的 GPT-4 和 Anthropic 的 Claude 2)。该模型的能力使得处理整本书级别的大量文本成为可能, 体现了谷歌在 AI 模型规模和性能上的最新进展。
 33. <https://oreil.ly/dF8q->: (<https://platform.openai.com/docs/guides/prompt-engineering>) 该资源是 OpenAI 官方发布的“Prompt Engineering Guide”(提示工程指南), 详细介绍了与生成式 AI 模型交互时设计和优化提示词(prompts)的原则和方法。该指南强调通用且长期有效的提示设计策略, 适用于各种生成文本或图像的模型, 帮助用户提升提示效果, 解决常见问题。
 34. <https://oreil.ly/huVRu>: (<https://stability.ai/blog/stable-diffusion-public-release>) 该 URL (<https://stability.ai/blog/stable-diffusion-public-release>) 指向的是关于“Stable Diffusion”公开发布的官方博客文章或公告。该资源详细介绍了 Stable Diffusion 这一开源图像生成模型的发布情况, 可能包括模型特性、使用方法(如 img2img 技术)、应用示例以及社区相关内容。用户可以通过该链接了解 Stable Diffusion 模型的背景信息、技术细节和最新动态。
 35. <https://oreil.ly/iEGZ9>: (<https://stackoverflow.com/questions/38045290/text-summarization-evaluation-bleu-vs-rouge>) 该资源是 Stack Overflow 上的一个问答页面, 讨论了文本摘要评估中常用的两种指标——BLEU 和 ROUGE

的比较与应用，帮助用户理解如何使用这些指标来衡量生成文本与参考文本之间的相似度。

36. <https://oreil.ly/iZhMl>: (<https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo>) 该资源是 OpenAI 官方文档中关于 GPT-4 及其变体 GPT-4 Turbo 模型的介绍页面，详细说明了这两种模型的能力、特性以及应用场景，特别包括它们支持的上下文窗口大小（如 128,000-token 的扩展版本）等技术参数。该页面旨在帮助开发者理解和使用 GPT-4 系列模型的最新功能和性能表现。
37. <https://oreil.ly/oGv4j>: (<https://platform.openai.com/signup>) 该资源是 OpenAI 平台的注册页面，用户可以在此创建 OpenAI 账户以便获取 API 使用权限和相关密钥。
38. <https://oreil.ly/oHID1>: (<https://platform.openai.com/account/api-keys>) 该资源为用户在 OpenAI 平台上管理和获取 API 密钥的页面，用户需要登录其 OpenAI 账号后，访问此页面以创建、查看或管理用于调用 OpenAI 各项 API 服务的密钥。
39. <https://oreil.ly/q7wQF>: (<https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/images/OnePager-Images.png>) 该资源是一份关于图像生成技术的“一页概要”（One-Pager），用以简明扼要地介绍生成式人工智能中图像生成的关键概念、方法或示例。具体来说，URL 指向的是一个 PNG 格式的图片文件，可能包含图像生成相关的信息图、流程图或示意图，帮助读者快速理解图像生成的核心内容。
40. <https://oreil.ly/tPPW9>: (<https://til.simonwillison.net/llms/python-react-pattern>) 该资源是一个关于使用 Python 和 React 框架构建 AI 代理（Autonomous Agents）模式的技术分享或教程，内容可能涉及如何通过多次调用 AI 模型实现自我推理（Reason and Act, ReAct）流程，从而让 AI 系统能够规划、观察、行动并评估结果，帮助开发者理解和实现复杂任务的自动化智能代理。
41. <https://oreil.ly/txhSZ>: (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389233) 该 URL (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389233) 指向的是一篇学术论文或研究报告，内容可能涉及使用

先进语言模型（如 GPT-4）进行模型评估的相关研究，特别是在通过更强大的模型对较弱模型的输出进行评价的方法。该资源可能详细探讨了基于 Meta 的 Llama 模型微调版本 Vicuna-13B 的发布及其评价方法，反映了当前自然语言处理领域中评估技术的最新进展。

42. <https://oreil.ly/wolEL>: (<https://github.com/openai/evals>) 该资源是 OpenAI 在 GitHub 上开源的一个评测框架，名为“evals”，用于对大型语言模型（LLMs）的性能进行标准化测试和基准评测。该框架提供了一套预定义的问题及评分标准，帮助开发者和研究者评估不同模型在多种任务上的表现，并支持社区贡献更多的评测模板。
43. <https://oreil.ly/yU27T>: (<https://til.simonwillison.net/llms/llama-cpp-python-grammars>) 该 URL (<https://til.simonwillison.net/llms/llama-cpp-python-grammars>) 指向的资源是一篇关于如何在使用 Llama 模型（特别是基于 llama-cpp 的 Python 接口）时，利用“grammars”（语法规则）来控制 and 规范模型输出格式的技术分享或教程。内容可能涉及通过定义 JSON 等结构化格式的语法，确保模型生成的文本符合预期格式，便于程序化解析和后续处理。
44. <https://s.mj.run/TKAsyhNiKmc>: 该资源是一个通过短链接（s.mj.run）形式分享的外部网页或文件，可能与 MidJourney（mj）相关，通常用于快速访问特定的图像、设计作品或生成内容。具体内容需要打开链接后查看，但大概率是与视觉创作或 AI 生成图像相关的资源。
45. <https://www.langchain.com>: 该资源是 LangChain 的官方网站，LangChain 是一个专为 AI 应用设计的开源工具库，提供多种功能以支持将多个提示模板和查询进行链式组合，从而实现更复杂的任务处理和更结构化的流程管理，广泛应用于如渐进式文本摘要等需要分块处理大文本的场景。
46. <https://www.midjourney.com>: 该资源是 Midjourney 的官方网站，Midjourney 是一种基于扩散模型的 AI 图像生成工具，用户可以通过输入文本提示（prompt）生成高质量的图像。该网站提供 Midjourney 服务的访问入口和账户管理，通常与 Discord 平台结合使用，通过 Midjourney Bot 在 Discord 中输入指令实现图像创作。

第 2 章

1. <https://chat.openai.com>: 该资源是 ChatGPT 的官方网站，提供基于先进大模型语言模型（LLM）的对话式人工智能服务，用户可以通过该平台与 AI 进行自然、流畅且上下文相关的文本交流。
2. <https://claude.ai/login>: 该资源是 Claude 2 的登录页面，用户可以通过该链接访问并登录 Claude 2 这一基于“宪法式 AI”（Constitutional AI）方法开发的先进大型语言模型平台。
3. <https://mistral.ai>: 该资源是法国初创公司 Mistral AI 的官方网站，介绍其开发的开源大型生成式人工智能模型及相关技术。网站内容涵盖其代表性产品如参数量为 73 亿的 Mistral 7B 模型、采用滑动窗口注意力机制的高效架构，以及基于该模型衍生的多种先进 AI 模型和应用，体现了其在生成式 AI 领域的创新与影响力。
4. <https://openai.com>: 该资源是 OpenAI 的官方网站，OpenAI 是一家致力于推动人工通用智能（AGI）发展并确保其惠及全人类的人工智能研究机构，知名于其开创性的 GPT 系列大型语言模型。
5. <https://oreil.ly/1C8qm>: (<https://www.theverge.com/2019/7/22/20703578/microsoft-openai-investment-partnership-1-billion-azure-artificial-general-intelligence-agi>) 该资源是一篇发表于 2019 年 7 月 22 日的 The Verge 报道，介绍了微软向 OpenAI 投资 10 亿美元的战略合作伙伴关系，重点阐述了微软通过其 Azure 云平台支持 OpenAI 开发先进的人工通用智能（AGI）技术。这笔投资推动了 OpenAI 在大规模语言模型（如 GPT-3 及其后续版本）方面的研发和应用。
6. <https://oreil.ly/2Ivq2>: (<https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>) 该资源是一篇路透社（Reuters）发布的科技新闻报道，内容聚焦于 ChatGPT 在 2023 年初创下用户增长纪录，成为互联网历史上增长速度最快的消费级应用，用户数

达到 1 亿。文章还分析了 ChatGPT 在客户服务、虚拟助理等多个领域的广泛应用及其影响。

7. <https://oreil.ly/3fOsM>: (<https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them>) 该资源是关于“什么是 tokens 及如何计数”的说明文章，介绍了在自然语言处理（NLP）和大型语言模型（LLMs）中，tokens 作为文本的基本单位的概念，解释了 token 可以是句子、单词或子词（如字符组合），并提供了如何衡量文本大小的方式（例如 100 个 tokens 大约相当于 75 个单词）。该文对理解和管理不同模型的 token 处理限制具有指导意义。
8. <https://oreil.ly/4fNQX>: (<https://aws.amazon.com/blogs/machine-learning/announcing-new-tools-for-building-with-generative-ai-on-aws/>) 该 URL (<https://aws.amazon.com/blogs/machine-learning/announcing-new-tools-for-building-with-generative-ai-on-aws/>) 指向的是亚马逊 AWS 官方博客上一篇关于发布用于构建生成式人工智能（Generative AI）新工具的公告文章。文章介绍了 AWS 在生成式 AI 领域推出的新服务和工具，如 Amazon Bedrock 和 Titan 大型模型（Titan FMs），旨在帮助开发者更便捷、高效地在 AWS 云平台上构建和部署生成式 AI 应用。
9. <https://oreil.ly/4ttmq>: (<https://openai.com/index/hello-gpt-4o>) 该资源是 OpenAI 官网上关于 GPT-4o 模型的介绍页面，详细说明了 GPT-4o 作为一款先进的多模态模型，能够实时处理和推理文本、音频及视觉输入，具备更强的性能和更高的效率。
10. <https://oreil.ly/6NNbg>: (<https://arxiv.org/abs/1706.03762>) 该资源是指向论文《Attention Is All You Need》（注意力机制即一切），这篇由 Google Brain 团队于 2017 年发表在 arXiv 上的开创性论文，首次提出了 Transformer 架构。该架构通过自注意力机制，极大地改进了自然语言处理（NLP）模型的性能，成为后续包括 GPT 系列在内的生成式预训练变换器模型的理论基础和关键技术。
11. <https://oreil.ly/6ib-Q>: (<https://arxiv.org/pdf/2203.02155.pdf?ref=assemblyai.com>) 该资源是指向一篇发表在 arXiv 上的学术论文，论文编号为 2203.02155。结合上下文内容，这篇论文很可能与 OpenAI 的语言模型特别是

ChatGPT 的技术细节、训练方法或微调策略有关，具体描述了如何通过收集全球反馈对模型进行微调，从而提升模型在对话场景中的表现和人机交互的自然度与有效性。

12. <https://oreil.ly/BWxrn>: (<https://openai.com/blog/new-models-and-developer-products-announced-at-devday>) 该资源是 OpenAI 官网博客上一篇文章，介绍了在开发者大会（DevDay）上发布的新模型和开发者产品，内容涵盖最新的 AI 技术进展及其应用，展示了 OpenAI 在推动人工智能领域创新方面的重要举措。
13. <https://oreil.ly/D7jL5>: (<https://x.com/Andercot/status/1767252833819410758?s=20>) 该资源是 Twitter（现名 X）上用户 Andercot 发布的一条推文，内容可能与 OpenAI 的 GPT-4 模型及其参数规模、训练成本或训练数据量相关，作为对上下文中关于 GPT-4 参数和训练信息的补充说明或引用。
14. <https://oreil.ly/EO42O>: (<https://www.wired.com/story/google-rebrands-ai-chatbot-gemini/>) 该资源是一篇来自 WIRED 的报道，介绍了谷歌将其人工智能聊天机器人 Bard 于 2024 年 2 月升级并重新命名为 Gemini 的相关信息，内容涉及谷歌在大型语言模型领域的最新进展及其与 GPT-4 性能的对比。
15. <https://oreil.ly/EbI8->: (<https://developers.googleblog.com/2023/03/announcing-palm-api-and-makersuite.html>) 该资源（URL: <https://developers.googleblog.com/2023/03/announcing-palm-api-and-makersuite.html>）是谷歌开发者博客上一篇于 2023 年 3 月发布的官方公告，介绍了 Google PaLM API 的发布及 MakerSuite 的相关信息。文章详细说明了 PaLM API 如何使开发者能够通过 Google Cloud Platform 访问和使用 Google 的大规模语言模型 PaLM，助力构建智能应用，并且可能涉及 Google 推出的开发工具套件 MakerSuite，支持开发者更便捷地利用 PaLM 模型进行产品开发和创新。
16. <https://oreil.ly/FJ-7D>: (<https://aws.amazon.com/bedrock/titan>) 该 URL “<https://aws.amazon.com/bedrock/titan>” 所指向的资源是亚马逊云服务（AWS）推出的 Amazon Bedrock 平台中的 Titan 系列基础模型（Foundation Models，简称 FMs）。Titan 模型是 AWS 提供的一类大型预训练语言模型，支

持开发者通过云端服务访问和使用，类似于 Google 的 PaLM API 和 Google Gemini，旨在帮助用户构建和部署基于大规模语言模型的应用。

17. <https://oreil.ly/JQd53>: (<https://www.deeplearning.ai/the-batch/issue-212/>) 该资源是《The Batch》第 212 期，由深度学习领域专家 Andrew Ng 主办，内容聚焦于人工智能特别是自然语言处理（NLP）领域的最新研究进展和背景故事，包括早期 NLP 研究与美国军事情报机构资助之间的联系。
18. <https://oreil.ly/LWIwv>: (<https://ai.google.dev/gemma>) 该 URL 指向的是 Google 发布的 Gemini 项目相关资源，Gemini 是一种先进的大型语言模型（LLM）架构，Google 提供了基于该架构的开源模型，旨在为软件工程师提供除 OpenAI 之外的高质量语言模型选项，方便将最先进的 LLM 集成到应用程序中。
19. https://oreil.ly/Lg6_r: (<https://huggingface.co/HuggingFaceH4/zephyr-7b-beta>) 该资源是“HuggingFaceH4/zephyr-7b-beta”模型的主页链接，指向一个基于 Mistral 7B 大规模开放源代码语言模型的衍生版本“Zephyr 7b beta”。该模型由 HuggingFace 团队或相关开发者在 Hugging Face 平台发布，属于高效且经过微调的 7 亿参数级别的生成式 AI 模型，适用于多种自然语言处理任务，体现了基于 Mistral 7B 模型的性能优化与扩展。
20. <https://oreil.ly/LroPn>: (<https://ai.meta.com/blog/large-language-model-llama-meta-ai/>) 该资源是 Meta 官方博客中的一篇文章，介绍了 Meta 在大型语言模型（Llama 系列）方面的策略和发展。文章详细阐述了 Meta 通过分阶段开源发布 Llama、Llama 2 和 Llama 3 模型，推动更开放、包容和协作的人工智能研发生态系统的理念和实践。
21. <https://oreil.ly/NH0jh>: (<https://www.anthropic.com/news/claude-3-family>) 该资源介绍了 Anthropic 公司 Claude 系列下一代模型的相关信息，特别是包括与 GPT-4 智能水平相当的 Opus 模型，以及价格更具竞争力、速度极快的 Haiku 小型模型。
22. <https://oreil.ly/NeZLw>: (<https://about.fb.com/news/2023/07/llama-2/>) 该资源是 Meta 官方发布的关于 Llama 2 语言模型的新闻或介绍页面，详细说明了 Meta 在语言模型领域的开源策略及 Llama 2 的相关信息，旨在推动更加开放和协作的人工智能发展生态系统。

-
23. <https://oreil.ly/Sj24h>: (<https://www.theverge.com/2023/2/8/23590864/google-ai-chatbot-bard-mistake-error-exoplanet-demo>) 该资源是《The Verge》于 2023 年 2 月 8 日发布的一篇文章，内容主要报道了谷歌 AI 聊天机器人 Bard 在演示中出现错误的事件，特别是涉及其关于系外行星 (exoplanet) 信息的错误，反映了 Bard 在发布初期相较于 ChatGPT 仍存在不足和瑕疵。
 24. <https://oreil.ly/Tim9W>: (<https://www.anthropic.com/index/claude-constitution>) 该资源是 Anthropic 公司关于“Claude 的宪法” (Claude’s Constitution) 的官方说明或介绍页面，详细阐述了其“宪法式人工智能” (Constitutional AI) 方法的原则和规则体系，用以指导和约束 Claude 系列大语言模型的训练与行为，从而提升 AI 的安全性和价值对齐。
 25. <https://oreil.ly/UOEBM>: (<https://openai.com/blog/chatgpt-plus>) 该资源是 OpenAI 官网博客中关于 ChatGPT Plus 订阅服务的介绍页面，详细说明了 ChatGPT Plus 用户可以通过付费订阅获得使用 GPT-4 模型的权限，从而体验更强大的语言理解和生成能力。
 26. <https://oreil.ly/Vwlo->: (<https://llama.meta.com/llama3/>) 该资源是 Meta 公司发布的开源大型语言模型系列中的最新版本“Llama 3”的官方网站或介绍页面，旨在展示 Meta 在语言模型领域的最新进展，推动更加开放和协作的人工智能开发生态系统。
 27. https://oreil.ly/_NAq5: (https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked/#google_vignette) 该资源是一篇关于 OpenAI GPT-4 模型架构、训练数据集、训练成本等详细信息的泄露报道或分析文章。文章内容涉及 GPT-4 的参数规模 (约 1.7 万亿参数)、训练费用 (约 6300 万美元) 以及所用训练数据的庞大体量 (相当于 650 公里长的书架书籍)，为读者提供了对 GPT-4 背后技术细节和资源投入的深入了解。
 28. <https://oreil.ly/bkWXk>: (https://huggingface.co/docs/optimum/concept_guides/quantization) 该资源是 Hugging Face 官方文档中关于“量化” (quantization) 技术的概念指导页面，详细介绍了量化在模型压缩中的作用及其实现方法，帮助开发者通过降低模型参数的数值精度，从而减小模型体积，提高模型在有限硬件资源上的运行效率。

-
29. <https://oreil.ly/evOQE>: (<https://github.com/openai/gpt-2>) 该资源是 OpenAI 在 GitHub 上公开发布的 GPT-2 模型代码库。该代码库包含 GPT-2 的预训练模型及相关实现, 方便开发者和研究人员基于此构建自然语言处理应用, 推动 AI 技术的普及与创新。
 30. <https://oreil.ly/iSO7>: (https://en.wikipedia.org/wiki/Byte_pair_encoding) 该资源介绍了“Byte Pair Encoding (BPE)”这一文本分词技术, 是一种用于自然语言处理 (NLP) 中将文本拆分成子词单元的高效方法。BPE 通过迭代合并最频繁的字节对, 能够在保持词汇表规模可控的同时, 有效处理各种词汇, 是多种 NLP 任务中常用的分词算法之一。
 31. <https://oreil.ly/itsJG>: (<https://mistral.ai/news/mixtral-of-experts/>) 该资源介绍了由法国初创公司 Mistral AI 推出的 Mixtral 8x7b 模型, 这是一种“专家混合”(mixture of experts) 架构的生成式人工智能模型, 类似于 GPT-4 的架构。Mixtral 8x7b 在性能上能够达到与 GPT-3.5-turbo 相当的效果, 体现了该公司在大规模开源语言模型领域的技术实力和创新能力。
 32. <https://oreil.ly/pZvMo>: (https://the-decoder.com/gpt-4-architecture-datasets-costs-and-more-leaked/#google_vignette) 该资源是一篇关于 OpenAI GPT-4 模型的详细解析文章, 内容涵盖 GPT-4 的架构细节、参数规模 (约 1.7 万亿参数)、训练所需的数据规模 (相当于占据 650 公里书架的书籍) 以及训练成本 (约 6300 万美元) 等方面的泄露信息与分析。文章旨在揭示 GPT-4 背后的技术与资源投入, 帮助读者深入了解这一先进大规模语言模型的开发情况。
 33. <https://oreil.ly/ttiji>: (<https://huggingface.co/spaces/lmsys/chatbot-arena-leaderboard>) 该资源是 Hugging Face 平台上一个名为“Chatbot Arena Leaderboard”的排行榜页面, 用于展示开源聊天模型的最新性能比较和详细排名信息。
 34. <https://oreil.ly/v45LZ>: ([https://arxiv.org/abs/2305.14705#:~:text=Sparse%20Mixture%2Dof%2DExperts%20\(,training%20LLMs%20to%20follow%20instructions\)](https://arxiv.org/abs/2305.14705#:~:text=Sparse%20Mixture%2Dof%2DExperts%20(,training%20LLMs%20to%20follow%20instructions))) 该资源介绍了“Sparse Mixture-of-Experts”(稀疏专家混合) 技术, 这是一种在训练大型语言模型 (LLMs) 时采用的架构方法, 通过结合多个专家模型的优势, 实现更高效、更精准的推理能力和更强的指令跟

随能力。文章详细阐述了这种方法如何突破单一模型的限制，提升模型的性能和灵活性，类似于 GPT-4 所使用的混合专家机制。

35. <https://oreil.ly/xuovP>: (<https://theaisummer.com/attention/#self-attention-the-key-component-of-the-transformer-architecture>) 该资源是关于“自注意力机制”（self-attention）的详细介绍，作为 Transformer 架构中的核心组成部分，解释了自注意力如何使模型能够在处理句子时让每个词关注句中其他所有词，从而更好地理解词的上下文含义和语义关系，提升自然语言处理的表现。
36. <https://oreil.ly/zORsB>: (<https://arxiv.org/abs/2106.09685>) 该资源 (<https://arxiv.org/abs/2106.09685>) 是一篇关于 LoRA (Low-Rank Adaptation, 低秩适配) 技术的学术论文，介绍了如何通过低秩近似的方法对大型神经网络模型进行高效微调，以显著减少训练和推理时的计算资源需求，同时保持模型性能，从而实现模型在较小硬件环境中的部署和应用。

第 3 章

1. <https://oreil.ly/30rlQ>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_3/hubspot_blog_post.txt) 该资源是一个文本文件，位于 GitHub 仓库“BrightPool/prompt-engineering-for-generative-ai-examples”的第三章内容目录下，文件名为“hubspot_blog_post.txt”。根据文件名推测，该文件可能包含与 HubSpot 相关的博客文章内容，可能用于展示生成式 AI 在内容创作或提示工程中的应用示例。
2. <https://oreil.ly/7N-bs>: (<https://flask.palletsprojects.com/en/3.0.x/>) 该资源 (<https://flask.palletsprojects.com/en/3.0.x/>) 是 Flask 框架的官方文档网

站，提供了 Flask 3.0.x 版本的详细说明、使用指南和 API 参考，帮助开发者了解和使用这一轻量级的 Python Web 应用框架。

3. <https://oreil.ly/8pf40>: (<https://github.com/f/awesome-chatgpt-prompts>) 该资源是一个名为 “Awesome ChatGPT prompts” 的项目，提供了大量可用于 ChatGPT 的角色扮演提示语列表，帮助用户更有效地与 ChatGPT 互动。
4. <https://oreil.ly/A0otS>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_3/one_hierarchical_list_generation.py) 该资源是一个示例代码文件，位于 GitHub 仓库 BrightPool 的 “prompt-engineering-for-generative-ai-examples” 项目中，文件路径为 `content/chapter_3/one_hierarchical_list_generation.py`。该示例演示了如何解析 (Parsing) 一个层级结构的列表 (hierarchical list)，可能用于生成式人工智能的提示工程相关应用。
5. <https://oreil.ly/GKDnc>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_3/sentence_detection_in_spacy.py) 该资源是一个示例代码文件，演示如何使用 spaCy 库进行句子检测 (sentence detection)。具体来说，它展示了利用 spaCy 进行文本的句子边界识别的实现方法，属于生成式人工智能相关的提示工程示例之一。
6. <https://oreil.ly/HA2QD>: (<https://pypi.org/project/tiktoken>) 该资源是 Python 包管理平台 PyPI 上的 tiktoken 项目页面，提供了该库的安装方法及相关信息。tiktoken 是一个用于文本编码和解码为令牌 (tokens) 的工具，常用于自然语言处理任务中的分词操作。用户可以通过该页面获取该库的版本、安装指令、使用示例及相关文档。
7. <https://oreil.ly/K6ZQK>: (<https://platform.openai.com/tokenizer>) 该资源是一个用于方便检查字符串分词结果的工具，通常称为 “OpenAI Tokenizer”，可以帮助用户查看文本是如何被分解成令牌 (tokens) 的。
8. <https://oreil.ly/LcMtv>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_3/

two_hierarchical_list_generation.py) 该资源是一个示例代码文件, 展示了如何将层级结构的列表解析并转换为 Python 字典的实现方法。具体来说, 它演示了从层级列表文本或数据中提取层级关系, 并以字典形式组织这些层级信息, 便于后续处理和使用。

9. <https://oreil.ly/MDOsz>: (<https://pubmed.ncbi.nlm.nih.gov/29949179>) 该 URL (<https://pubmed.ncbi.nlm.nih.gov/29949179>) 指向的是一篇发表在 PubMed 上的医学文献, 标题为《Tumor Microenvironment: Recent Advances in Various Cancer Treatments》。该文献综述了肿瘤微环境在多种癌症治疗中的最新研究进展, 探讨了肿瘤微环境对癌症发展的影响及其作为治疗靶点的潜力。
10. <https://oreil.ly/MoJHn>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_3/three_json_parsing.py) 该资源是一个位于 GitHub 上的 Python 脚本文件, 名为 `three_json_parsing.py`, 属于 “prompt-engineering-for-generative-ai-examples” 项目的第三章内容。该脚本示例演示了如何使用 Python 解析 JSON 格式的数据, 帮助用户理解 and 处理生成式 AI 输出的 JSON 结构。
11. <https://oreil.ly/OHurh>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/tree/main/content/chapter_3) 该资源是一个 GitHub 仓库中的目录, 路径为 `content/chapter_3`, 包含与第三章内容相关的示例文件和代码, 特别是用于正确读取 `hubspot_blog_post.txt` 文件的上下文, 适用于运行相关的 Python 代码或启动 Jupyter Notebook。
12. <https://oreil.ly/QZE8n>: (<https://platform.openai.com/docs/guides/prompt-engineering>) 该资源是 OpenAI 官方文档中关于 “提示工程” (Prompt Engineering) 的指南, 介绍了多种提升 AI 生成结果准确性的方法和原则, 帮助用户设计更有效的提示语以优化 AI 模型的输出效果。
13. <https://oreil.ly/TRK0i>: (<https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/requirements.txt>) 该资源是一个名为 `requirements.txt` 的文件, 位于 GitHub 仓库 `BrightPool/prompt-engineering-for-generative-ai-examples` 中。该文件通常用于列出

一个 Python 项目所依赖的第三方库及其版本信息，方便通过 pip 等包管理工具快速安装所需环境。结合上下文可知，该文件可能包含用于示范生成式 AI 相关示例的 Python 依赖包，但不包括 Flask（Flask 仅作为演示用途，未被包含在该 requirements.txt 里）。

14. <https://oreil.ly/aCkDo>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_3/sliding_window.py) 该资源是一个名为“Sliding window”的代码示例，位于 GitHub 仓库 BrightPool 的项目“prompt-engineering-for-generative-ai-examples”中，具体路径为 content/chapter_3/sliding_window.py。该示例代码可能用于展示如何在生成式 AI 提示工程中应用滑动窗口技术，以处理文本或数据序列，实现分段输入、上下文管理等功能。
15. <https://oreil.ly/c4MvQ>: (<https://spacy.io/api/tokenizer>) 该资源是 spaCy 官方网站中关于其分词器（tokenizer）API 的页面，介绍了 spaCy 中用于将文本拆分成单词、标点等基本单元的分词器功能和使用方法。
16. <https://oreil.ly/l3Ihy>: (<https://github.com/Significant-Gravitas/Auto-GPT>) 该资源是一个名为 Auto-GPT 的多智能体系统项目，托管在 GitHub 上（地址：<https://github.com/Significant-Gravitas/Auto-GPT>）。它利用类似演员-评论家（actor-critic）的机制，实现自动自我评估和上下文检查，帮助大型语言模型（如 GPT-4）在执行任务前判断当前提示信息是否足够，进而动态请求更多上下文或优化提示，从而提高决策的准确性和任务完成度。
17. <https://oreil.ly/n3sNy>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_3/character_chunking_hubspot_blog_post.py) 该资源是一个示例代码文件，演示了“Character chunking”（字符分块）技术的具体实现，代码存放于 GitHub 仓库中，属于“prompt-engineering-for-generative-ai-examples”项目的第三章内容，文件名为 character_chunking_hubspot_blog_post.py，可能基于 HubSpot 博客的相关内容进行说明或应用示例。
18. <https://oreil.ly/nfc3f>: (<https://arxiv.org/pdf/2310.01783.pdf>) 该资源 (<https://arxiv.org/pdf/2310.01783.pdf>) 是一篇学术论文，内容涉及大型语言

模型 (LLMs) 作为评估者的能力研究。论文探讨了 LLMs 在评价任务中表现的有效性和一致性, 评估了如 GPT-4 等先进模型在对较小或较简单模型 (如 GPT-3.5-turbo) 输出进行评分时的可靠性和实用性。

19. <https://oreil.ly/oSpVe>: (<https://pypi.org/project/tiktoken>) 该资源是“Tiktoken”的 Python 包项目页面, 提供了一个快速的字节对编码 (BPE) 分词器, 用于将文本拆分成子词单元, 专为 OpenAI 模型设计, 具有比同类开源分词器更高的性能。
20. <https://oreil.ly/wTmI7>: (<https://www.nltk.org/api/nltk.tokenize.html>) 该资源是 NLTK (Natural Language Toolkit) 库中与文本分词 (tokenization) 相关的 API 文档页面, 介绍了 NLTK 中用于将文本拆分成单词、句子或其他单位的各种分词工具和方法。
21. <https://oreil.ly/ykkzY>: (<https://arxiv.org/abs/2305.17926>) 该资源 (<https://arxiv.org/abs/2305.17926>) 是一篇关于大型语言模型 (LLMs) 作为评估者能力的研究论文。文中探讨了 LLMs 在作为人类水平的评价者方面的有效性, 分析了不同研究中对 LLMs 评价一致性和可靠性的不同结论, 并结合实际经验指出如 GPT-4 在评估较小模型 (如 GPT-3.5-turbo) 生成的回答时表现出较高的有效性和一致性。
22. <https://spacy.io>: 该资源是 spaCy 的官方网站, spaCy 是一个用于自然语言处理 (NLP) 的开源 Python 库, 提供句子切分、词性标注、命名实体识别等功能。
23. <https://www.midjourney.com>: 该资源是一个名为 Midjourney 的图像生成平台, 用户可以通过输入文本提示 (prompt) 生成对应的视觉图像, 常用于将产品描述等文字内容转化为图像表现。
24. <https://www.mturk.com>: 该资源是 Amazon Mechanical Turk (简称 Mechanical Turk 或 MTurk) 的官方网站, 提供一个在线众包平台, 允许用户发布需要人工完成的小任务 (如数据标注、质量评估等), 并由全球的低成本人力工人完成这些任务, 用于辅助人工智能模型的训练和评价。

第 4 章

1. <https://cloud.google.com/vertex-ai>: 该资源是 Google 提供的 Vertex AI 平台，旨在帮助用户构建、部署和管理机器学习模型，支持多种模型的快速实验和集成，提升模型开发效率。
2. <https://llamahub.ai>: 该资源 “<https://llamahub.ai>” 是一个开源库，专注于文档加载器的开发，能够创建适用于 LangChain 框架的特定 Document 对象，方便与大型语言模型（LLMs）相关的应用集成与开发。
3. <https://mistral.ai>: 该资源是一个提供 API 服务的平台，用户需要在该网站注册并订阅以获取 API 密钥，然后通过环境变量进行配置，方便在终端或程序中调用其 API 接口。
4. <https://openai.com>: 该资源为 OpenAI 的官方网站，提供人工智能模型和相关技术的介绍与服务，支持开发者访问和使用先进的 AI 模型。
5. <https://oreil.ly/-koAO>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_4/langchain-evals.ipynb) 该资源是一个名为 “langchain-evals.ipynb” 的 Jupyter Notebook 文件，位于 GitHub 仓库 BrightPool 的 “prompt-engineering-for-generative-ai-examples” 项目中。该笔记本内容可能涉及使用 LangChain 框架进行评估（evaluations）相关的示例和代码，结合上下文中的数据步骤，推测其展示了如何在生成式 AI 应用中利用 LangChain 对交易描述等数据进行处理和评估的具体实践。
6. <https://oreil.ly/0Fn94>: (<https://docs.smith.langchain.com/evaluation/custom-evaluators>) 该资源是 LangChain 官方文档中关于 “自定义评估器”（custom evaluators）的说明页面，介绍如何在 LangChain 框架中创建和使用自定义的评估指标（evals）来衡量和优化 AI 模型的响应性能，辅助开发者进行模型调试、监控和测试。
7. https://oreil.ly/0p_nE: (https://python.langchain.com/docs/guides/evaluation/string/embedding_distance) 该资源介绍了如何使用 LangChain

中的评估器 (evaluator)，特别是基于字符串距离（如 Levenshtein 距离）和嵌入向量距离（embedding distance）的方法来进行答案的相似度评估。它解释了 Levenshtein 距离用于衡量两个文本之间的编辑距离，实现模糊匹配；而嵌入距离则利用向量表示计算语义上的相似性，适用于答案与参考答案不完全一致但语义上接近的场景。

8. <https://oreil.ly/5Vl0W>: (https://js.langchain.com/docs/get_started/introduction) 该资源是 LangChain 的官方入门介绍文档，详细说明了 LangChain 框架的基本概念、功能和使用方法，帮助用户快速了解如何利用该框架构建基于大型语言模型 (LLM) 的应用程序。
9. <https://oreil.ly/9NcTB>: (https://github.com/jerryjliu/llama_index) 该资源是 LlamaIndex 的 GitHub 仓库，LlamaIndex 是一个与 LangChain 框架无缝集成的开源大型语言模型 (LLM) 框架，旨在为文档加载和处理提供支持，方便构建基于语言模型的数据连接和应用。
10. <https://oreil.ly/9O7Mb>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_4/langchain-evals.ipynb) 该 URL 指向的资源是一个 Jupyter Notebook，演示了如何使用 LangChain 库中的评估器 (evaluator) 进行生成式 AI 模型输出的对比评测。具体来说，该笔记本展示了利用 GPT-4 作为评估器，通过 “labeled_pairwise_string” 方法对两个模型（如 Mistral 和 GPT-3.5-turbo）生成的文本输出进行成对比较，并给出选择理由和评分，从而帮助用户分析和判断不同模型或提示之间的表现差异。
11. https://oreil.ly/9xr_6: (https://python.langchain.com/docs/use_cases/summarization#use-case) 该资源是 LangChain 官方文档中关于“文本摘要”用例的说明页面，详细介绍了如何使用 LangChain 框架实现文档内容的自动摘要功能。
12. <https://oreil.ly/A15G3>: (https://en.wikipedia.org/wiki/Levenshtein_distance) 该 URL 指向的资源是维基百科上关于“Levenshtein 距离”的条目，介绍了一种用于衡量两个字符串之间差异的编辑距离算法。Levenshtein 距离通

过计算将一个字符串转换成另一个字符串所需的最少单字符编辑操作（插入、删除或替换）的数量，来评估字符串的相似度，常用于模糊匹配和文本相似性计算。

13. <https://oreil.ly/DqDOF>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_4/langchain-evals.ipynb) 该资源是一个位于 GitHub 上的 Jupyter Notebook 文件，文件名为 “langchain-evals.ipynb”，属于 “prompt-engineering-for-generative-ai-examples” 项目中的第四章内容。该笔记本包含用于生成式人工智能模型评估的示例代码，特别涉及基于 LangChain 框架的评估方法，演示如何使用数据框（dataframe）进行模型评估指标的计算与分析。
14. <https://oreil.ly/Ec578>: (https://huggingface.co/docs/transformers/en/model_doc/mixtral) 该资源是 Hugging Face 官方文档中关于 Mixtral 模型的介绍页面，详细说明了 Mixtral 8x7b（API 中称为 mistral-small）模型的功能、性能及使用方法。文档旨在帮助用户了解该模型在自然语言处理任务中的表现，特别是在与更大模型（如 GPT-4）相比时的准确性和效率优势，以及其开源特性和自托管的可能性。
15. https://oreil.ly/FQUK_: (<https://api.python.langchain.com/en/latest/>) 该 URL (<https://api.python.langchain.com/en/latest/>) 指向的是 LangChain 库的 Python 版本最新英文文档的官方 API 参考页面，提供了关于如何实现各种文档链（document chains）及其功能的详细技术说明和使用指南。
16. <https://oreil.ly/PjV9o>: (https://python.langchain.com/docs/get_started/introduction.html) 该资源是 LangChain 的官方文档中的入门介绍页面，主要用于帮助用户了解如何使用 LangChain Python 库进行数据处理与集成，涵盖基础概念和快速上手指南。
17. <https://oreil.ly/QIMih>: (<https://docs.pydantic.dev/latest>) 该资源是 Pydantic 库的官方文档，提供关于如何使用 Pydantic 进行数据验证和解析的详细说明，介绍了基于 Python 类型注解创建数据模型和自动校验输入数据的功能与用法。

18. <https://oreil.ly/S40bZ>: (<https://platform.openai.com/finetune>) 该资源是 OpenAI 平台上用于模型微调 (fine-tuning) 的页面, 提供用户自定义训练和优化预训练模型的方法, 以提升模型在特定复杂任务上的表现。
19. <https://oreil.ly/WKOma>: (<https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/tree/main>) 该 URL 指向的是一个 GitHub 仓库中的目录, 具体是 “prompt-engineering-for-generative-ai-examples” 项目的主分支下的内容。该资源包含与生成式人工智能相关的提示工程示例, 可能包括代码、示例脚本和相关文件, 用于帮助用户学习和实践生成式 AI 的提示设计与应用。
20. <https://oreil.ly/YPid>: (https://python.langchain.com/docs/get_started/introduction.html) 该资源是 LangChain 官方文档中的入门介绍页面, 详细说明了 LangChain 这一多功能框架的基本概念和用途, 帮助用户了解如何使用 LangChain 创建基于大型语言模型 (LLM) 的应用程序。
21. <https://oreil.ly/a4Hut>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_4/langchain-evals.ipynb) 该资源是一个名为 langchain-evals.ipynb 的 Jupyter Notebook 文件, 位于 GitHub 仓库中, 展示了如何使用 GPT-4 生成的参考答案来进行评估指标的计算。具体来说, 该笔记本示例演示了对金融交易描述列表进行分类, 将每笔交易标注为相应的交易类别 (transaction_category) 和交易类型 (transaction_type), 从而用于模型评估和验证。
22. <https://oreil.ly/bedrock>: (<https://python.langchain.com/docs/integrations/chat/bedrock>) 该资源是 LangChain 官方文档中关于与 Amazon Bedrock 聊天模型集成的说明页面, 介绍了如何在 LangChain 框架下使用 BedrockChat 模型, 实现跨平台的模型调用与快速实验。
23. <https://oreil.ly/cVTyI>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/tree/main/content/chapter_4) 该资源是一个 GitHub 代码库中的目录, 位于路径 content/chapter_4, 包含与第四章相关的示例代码和内容, 主要用于指导用户在该章节中新建 Jupyter

Notebook 或 Python 文件，并运行相关 Python 包安装命令（如 pdf2image、docx2txt、pypdf）以支持示例的执行与学习。

24. <https://oreil.ly/dcCOO>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_4/langchain-evals.ipynb) 该 URL 指向的资源是 GitHub 仓库中一个名为 “langchain-evals.ipynb” 的 Jupyter 笔记本文件，属于 “prompt-engineering-for-generative-ai-examples” 项目的第四章内容。该笔记本很可能演示了如何使用 LangChain 进行评估 (evals)，结合生成式 AI 模型（如 GPT-3.5）对数据（例如交易记录）进行处理和分类的示例代码。具体包括对数据框 (DataFrame) 行的遍历与处理，展示了基于不同模型输出的交易类别和类型的对比分析。
25. <https://oreil.ly/hW8Wr>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_4/langchain-evals.ipynb) 该资源是一个 Jupyter Notebook 文件，位于 GitHub 仓库 “BrightPool/prompt-engineering-for-generative-ai-examples” 的第 4 章目录下，文件名为 “langchain-evals.ipynb”。该笔记本展示了如何使用 LangChain 框架对生成式 AI 模型的输出进行评估，包括示例代码演示如何遍历数据框 (dataframe) 中的每一行，运行评估器 (evaluator)，并将评估结果及其推理过程保存回数据框中。
26. <https://oreil.ly/iahTJ>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_4/langchain-evals.ipynb) 该资源是一个 Jupyter 笔记本，位于 GitHub 仓库 “BrightPool/prompt-engineering-for-generative-ai-examples” 的第 4 章目录下，文件名为 “langchain-evals.ipynb”。该笔记本示例展示了如何使用基于 LangChain 的评估工具，通过成对比较 (pairwise comparisons) 的方法，利用更智能的模型（如 GPT-4）对不同提示词或模型（例如 GPT-3.5-turbo 与 Mixtral 8x7b）的表现进行准确率评估，并提供推理分析以帮助调试和理解模型选择的原因。

-
27. <https://oreil.ly/n0hDD>: (<https://unstructured-io.github.io/unstructured/index.html>) 该资源是“Unstructured”项目的官方网站或文档页面，提供了一个用于数据摄取和处理的 Python 包，帮助用户高效地处理非结构化数据。
 28. <https://oreil.ly/u9gMx>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/tree/main/content/chapter_4/data) 该资源是一个代码或示例项目中的数据文件夹，位于 GitHub 仓库中，路径为“content/chapter_4/data”。它包含了除 .docx 文件之外的所有数据文件，供用户导入各种数据加载器时使用，用于汇总和存储多个数据源中的 Document 对象，方便进行统一处理和操作。
 29. <https://oreil.ly/uz05O>: (<https://github.com/openai/tiktoken>) 该资源是一个由 OpenAI 维护的开源工具库，名为「tiktoken」，用于计算 GPT-4 等大型语言模型调用中的令牌（token）数量，同时支持将文本字符串拆分成适当大小的块和文档，方便进行分段处理和模型输入管理。你可以在 GitHub 上访问该项目，获取其源码和使用说明。
 30. <https://oreil.ly/vPUfI>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_4/langchain-evals.ipynb) 该 URL 指向的是一个 GitHub 仓库中的 Jupyter Notebook 文件，文件名为“langchain-evals.ipynb”，位于“prompt-engineering-for-generative-ai-examples”项目的第四章内容目录下。该 Notebook 示例展示了如何使用 LangChain 框架对生成式 AI 模型（如 Mistral）的输出进行评估，特别是通过比较模型预测结果与参考答案的字符串匹配来计算准确率，从而实现对模型性能的基本评估。
 31. <https://oreil.ly/wvuO7>: (<https://pypi.org/project/python-dotenv>) 该资源是 Python 的一个库——python-dotenv，托管在 PyPI（Python 包索引）上，主要用于从 .env 文件加载环境变量，方便在开发中管理和使用环境配置。
 32. <https://wandb.ai/site>: 该资源是一个名为“Weights and Biases”的机器学习平台网站，提供类似于 LangSmith 的平台功能，支持对大型语言模型（LLMs）的调试、监控和测试，具备性能评估和追踪能力。

33. <https://www.anthropic.com>: 该资源为 Anthropic 公司官方网站, Anthropic 是一家专注于开发先进人工智能模型和技术的公司, 提供可与 LangChain 等平台集成的 AI 模型服务, 支持快速实验和高效的模型切换。
34. <https://www.llamaindex.ai>: 该资源是一个名为 LlamaIndex 的工具或库, 属于用于数据摄取 (data ingestion) 的 Python 软件包之一, 可能用于结构化和 管理数据以便后续处理或分析。

第 5 章

1. <https://hf.co/settings/tokens>: 该资源是 Hugging Face 网站上的个人 API 令牌管理页面, 用户可以在此创建、查看和管理自己的 API 访问令牌, 用于通过编程方式访问 Hugging Face 的各种服务和接口。
2. <https://oreil.ly/39rgU>: (https://python.langchain.com/docs/modules/data_connection/retrievers/self_query/) 该资源介绍了 “Self-querying” (自查询) 技术, 阐述了其带来的多项重要优势。
3. <https://oreil.ly/4wR7o>: (https://github.com/jamesaphoenix/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_5/rag.ipynb) 该 URL (https://github.com/jamesaphoenix/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_5/rag.ipynb) 指向一个 Jupyter Notebook 文件, 内容涉及利用 FAISS (Facebook AI Similarity Search) 进行向量数据库的构建与查询, 具体应用于处理和存储文档向量以支持基于向量相似度的检索增强生成 (RAG, Retrieval-Augmented Generation) 技术示例。该资源属于一本关于生成式 AI 提示工程的示例代码库中的第五章, 演示如何将文档分块后生成的向量存储到 FAISS 数据库中, 从而提高检索效率和减少重复计算。

4. <https://oreil.ly/6RzTy>: (https://github.com/jamesaphoenix/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_5/embeddings.ipynb) 该资源是一个 Jupyter 笔记本文件，位于 GitHub 仓库中，内容涉及如何使用 OpenAI 的 text-embedding-ada-002 模型生成高维文本向量 (embeddings)。该笔记本展示了获取和处理文本嵌入的示例代码，帮助用户理解和应用生成式 AI 中的向量表示技术，特别是在语义捕捉和向量空间可视化方面的实践。
5. <https://oreil.ly/8OV3c>: (<https://huggingface.co/sentence-transformers>) 该资源是 Hugging Face 提供的 Sentence Transformers 库的页面，介绍了一个开源的用于生成文本嵌入 (embeddings) 的模型集合，支持高效的句子和文本向量化，常用于自然语言处理任务如语义搜索和文本相似度计算。
6. <https://oreil.ly/9JbTt>: (https://python.langchain.com/docs/modules/data_connection/retrievers/time_weighted_vectorstore) 该资源介绍了“Time-Weighted Vector Store Retriever”这一检索器，它在文档检索过程中引入了时间衰减机制，以增强对近期信息的权重，从而提升检索的时效性和相关性。同时，该资源也指出时间衰减可能导致对较早文档的忽视，存在丢失历史上下文信息的风险。
7. <https://oreil.ly/BBYD4>: (<https://docs.pinecone.io/docs/metadata-filtering>) 该资源 (<https://docs.pinecone.io/docs/metadata-filtering>) 是 Pinecone 官方文档中关于“元数据过滤” (metadata filtering) 的说明页面，介绍了如何通过查询时使用元数据过滤条件（如指定某个批次号 batch）来筛选向量搜索结果，从而只返回符合特定元数据条件的匹配项。
8. <https://oreil.ly/KMNU8>: (<https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo>) 该资源介绍了 OpenAI 的 GPT-4 及其变体 GPT-4 Turbo 模型，重点说明了这些模型在处理大规模上下文时的能力和限制，包括其支持的最大上下文窗口大小（如 GPT-4-1106-preview 的 128,000 个 token）及其成本效益对比。同时，该文档还探讨了在实际应用中如何通过检索相关信息（RAG，检索增强生成）以高效利用模型上下文窗口，从而平衡性能与成本的最佳实践。

9. <https://oreil.ly/Q0rIw>: (https://github.com/jamesaphoenix/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_5/pinecone.ipynb) 该资源是一本书的 GitHub 代码仓库中的 Jupyter 笔记本文件，内容演示了如何使用 Pinecone 这一商业向量数据库进行向量索引的创建与管理，具体包括将文档分块、向量化、索引入库，以及基于向量检索相似文档以用作生成式 AI 提示上下文的完整操作流程。
10. <https://oreil.ly/RmXVR>: (<https://www.tensorflow.org/text/tutorials/word2vec>) 该资源是 TensorFlow 官网提供的一个教程，介绍如何使用 word2vec 方法训练自己的词向量 (embeddings)。通过该教程，用户可以学习如何将词语表示为向量，从而捕捉词语的语义信息，实现自定义的词嵌入训练。
11. <https://oreil.ly/YC-nV>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_5/pinecone.ipynb) 该资源是一个存放在 GitHub 上的 Jupyter Notebook 文件，文件路径为 “content/chapter_5/pinecone.ipynb”。该笔记本内容涉及使用 Pinecone 向量数据库，具体演示了如何将文本分块及其对应的向量批量存储 (upsert) 到 Pinecone 索引中，包含更新和插入操作的示例，适用于生成式 AI 中的向量检索应用。
12. <https://oreil.ly/apikeys>: (<https://platform.openai.com/account/api-keys>) 该资源是用户管理和获取 OpenAI API 密钥的页面，用户可以通过访问该链接创建账户并生成用于调用 OpenAI 接口的 API 密钥，以便在代码中配置环境变量进行身份验证和计费。
13. <https://oreil.ly/dlcmb>: (<https://www.pinecone.io/blog/sparse-dense/>) 该资源是一篇由 Pinecone 发布的博客文章，介绍了稀疏向量 (sparse vectors) 与密集向量 (dense vectors) 在向量表示中的区别及应用。文章详细解释了现代变换器模型生成的上下文相关的密集嵌入向量与传统具有大量零值维度的稀疏向量之间的差异，探讨了两者在语义搜索和关键词搜索中的优势及结合使用的趋势。
14. <https://oreil.ly/gHb3F>: (<https://scikit-learn.org/stable/>) 该资源是 scikit-learn 的官方网站，提供一个开源的科学计算库，主要用于机器学习和数据分析。

用户可以利用该库实现包括 TF-IDF 计算和余弦相似度在内的多种算法，用于文本相似度计算、分类、聚类等任务。

15. <https://oreil.ly/gIcTI>: (<https://github.com/facebookresearch/faiss>) 该资源是 Facebook AI 开发的开源库 FAISS (Facebook AI Similarity Search) 的 GitHub 代码仓库，提供高效的密集向量相似度搜索和聚类功能，广泛用于向量数据库的存储与查询，帮助用户快速检索和管理向量数据。
16. <https://oreil.ly/jIuJh>: (https://python.langchain.com/docs/modules/data_connection/retrievers/ensemble) 该资源介绍了 LangChain 中的 Ensemble Retriever 模块。Ensemble Retriever 是一种基于多重检索器组合结果的混合检索方法，旨在利用不同检索算法的优势以提升检索效果。该方法通过同时调用多个检索器并整合它们的结果来实现更强大的查询能力，但也因此带来了更高的计算负载，可能影响检索速度。
17. <https://oreil.ly/jXSXQ>: (https://python.langchain.com/docs/modules/data_connection/retrievers/parent_document_retriever) 该资源介绍了“Parent Document Retriever”这一检索器模块，主要功能是通过检索原始的文文档（即从中拆分出更小文档块的源文档），来维护丰富的文档背景信息。该方法有助于提升信息的完整性和上下文关联性，但可能会增加计算资源的消耗。
18. <https://oreil.ly/pgvector>: (<https://supabase.com/docs/guides/database/extensions/pgvector>) 该 URL (<https://supabase.com/docs/guides/database/extensions/pgvector>) 指向的资源是 Supabase 官方文档中关于其数据库扩展功能“pgvector”的指南页面。该页面介绍了如何在 Supabase 平台上使用 pgvector 扩展来实现向量搜索功能，支持 AI 相关的向量数据存储与检索。通过该扩展，Supabase 能够提供类似于专业向量数据库（如 Pinecone）的向量搜索能力，方便开发者在其托管的数据库中集成和应用向量检索技术。
19. <https://oreil.ly/qsG7J>: (https://python.langchain.com/docs/modules/data_connection/document_transformers/text_splitters/split_by_token) 该资源介绍了 LangChain 中用于处理文本的“文本分割器” (text splitters) 模块，特别是基于“token”进行切分的方法。它解决了在将长文本拆分成固定 token 数量的片段时，避免在句子或段落中间断开的难题，从而保留文本的语义

完整性和结构信息。该页面详细讲解了如何根据 token 限制合理拆分文本片段，以便在使用如 GPT-4 这类有 token 上限的语言模型时，能够高效且语义连贯地构建模型输入。

20. <https://oreil.ly/uuzpG>: (https://python.langchain.com/docs/modules/data_connection/retrievers/MultiQueryRetriever) 该资源介绍了 MultiQueryRetriever，这是一个用于改进基于向量数据库的检索方法的工具。它通过从单个用户输入查询生成多个不同视角的查询，从而扩大检索到的相关文档集合，提供更全面的见解。但也存在不同查询结果可能冲突或重叠的挑战。
21. <https://oreil.ly/wbx1f>: (<https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo>) 该 URL (<https://platform.openai.com/docs/models/gpt-4-and-gpt-4-turbo>) 指向的是 OpenAI 官方文档中关于 GPT-4 及 GPT-4 Turbo 模型的介绍页面。该资源详细说明了这两种模型的功能特点、使用限制（如 token 限制）、性能差异及应用场景，帮助开发者理解如何高效调用和利用 GPT-4 系列模型进行自然语言处理任务。
22. <https://oreil.ly/wzqVg>: (https://python.langchain.com/docs/modules/data_connection/retrievers/contextual_compression/) 该资源介绍了“Contextual Compression Retriever”这一检索工具，主要用于处理长文档，通过压缩与当前上下文无关的内容来提升检索的相关性。该方法的难点在于需要具备一定的专业知识来判断信息的相关性和重要性，从而实现高效且精准的文档压缩和检索。
23. <https://supabase.com>: 该资源是 Supabase 官网，Supabase 提供数据库托管服务，并通过集成 pgvector 插件支持向量搜索功能，适用于构建和管理支持 AI 用例的数据库应用。
24. <https://weaviate.io>: 该资源是 Weaviate，一个提供托管向量数据库服务的平台，专注于支持人工智能相关的用例和向量搜索功能。
25. <https://www.llamaindex.ai>: 该资源是 LlamaIndex 的官方网站，LlamaIndex 是一个专注于检索增强生成（RAG, Retrieval-Augmented Generation）技术的框架，适用于构建复杂的 AI 应用，特别是在结合外部知识库进行智能信息检索与生成方面具有优势。

26. <https://www.pinecone.io>: 该资源是 Pinecone 官网, Pinecone 是一家知名的商业托管向量数据库服务提供商, 专注于支持人工智能应用中的向量检索和相似度搜索。用户可以在该平台上创建索引、上传向量数据, 并通过 API 进行高效的向量查询, 用于实现基于内容的检索和推荐等功能。
27. <https://www.trychroma.com>: 该资源 Chroma 是一个托管的向量数据库提供商, 专注于支持人工智能相关的应用场景, 提供向量搜索功能, 类似于 Weaviate 和 Pinecone。

第 6 章

1. <https://oreil.ly/1rYDI>: (<https://arxiv.org/abs/2305.10601>) 该 URL (<https://arxiv.org/abs/2305.10601>) 对应的资源是一篇关于 “Tree of Thoughts (ToT)” 框架的学术论文。该论文提出了一种新颖的方法, 利用语言模型进行问题求解时, 突破了传统基于线性、逐词推理的限制, 通过树状思维结构引入更高级的策略前瞻和复杂推理, 从而提升模型在需要关键初始决策和多步推理任务中的表现。
2. <https://oreil.ly/2wu-y>: (<https://github.com/weaviate/weaviate>) 该资源是 Weaviate 的 GitHub 仓库, Weaviate 是一个开源的向量搜索引擎, 用于构建基于机器学习的语义搜索和推荐系统。
3. <https://oreil.ly/5AVfM>: (https://www.ted.com/talks/jim_fan_the_next_grand_challenge_for_ai?language=en) 该 URL 链接指向 TED 官网上的一个演讲视频, 演讲者是 Jim Fan, 主题是 “人工智能的下一项重大挑战” (The Next Grand Challenge for AI)。该演讲可能探讨了人工智能领域中尚未解决的重要问题和未来的发展方向, 尤其涉及智能体 (agent) 在不同任务中可执行动作范围 (action space) 的问题与挑战。

4. <https://oreil.ly/5tAB0>: (<https://python.langchain.com/docs/integrations/tools/filesystem>) 该资源是 LangChain 官方文档中关于“文件系统工具” (File system tools) 的介绍页面, 详细说明了如何在 LangChain 框架中集成和使用文件系统相关的工具和功能。
5. <https://oreil.ly/8EhXl>: (<https://python.langchain.com/docs/modules/callbacks/>) 该资源介绍了 LangChain 框架中的“callbacks”模块, 内容主要讲解如何利用回调函数实现对应用程序的无缝监控和问题定位, 帮助开发者更高效地调试和追踪 AgentExecutor 等链式调用中的执行过程。
6. <https://oreil.ly/8vYF5>: (https://github.com/langchain-ai/langchain/blob/master/cookbook/plan_and_execute_agent.ipynb) 该资源是 LangChain 项目中一个实验性功能的示例笔记本, 名为“plan_and_execute_agent.ipynb”, 演示了如何使用 LangChain 框架实现“计划与执行”类型的智能代理。
7. <https://oreil.ly/9R3pU>: (<https://github.com/chroma-core/chroma>) 该资源是 Chroma 的官方网站或代码仓库链接, Chroma 是一个用于存储和管理上下文数据的开源向量数据库, 常用于增强任务管理系统中的上下文信息检索与存储, 帮助系统更好地对任务进行生成和优先级排序, 从而实现与预定目标的有效对齐。
8. https://oreil.ly/ECD_n: (<https://python.langchain.com/docs/integrations/memory/>) 该资源是 LangChain 官方文档中关于“内存 (memory)”模块的介绍, 详细说明了生成式 AI 模型中对话消息的存储机制, 包括临时的内存列表存储和持久化数据库存储, 并介绍了多种数据库集成方案, 帮助开发者高效管理和利用对话状态和上下文信息。
9. <https://oreil.ly/ECS4r>: (<https://python.langchain.com/docs/integrations/tools/twilio>) 该资源是 Twilio 的相关文档或介绍链接, 通常用于展示 Twilio 服务的功能和使用方法。结合 URL “<https://python.langchain.com/docs/integrations/tools/twilio>”, 可以推断该链接指向的是 LangChain 文档中关于 Twilio 工具的集成说明, 介绍如何在 LangChain 框架中使用 Twilio 进行短信或语音通信的集成与调用。

10. <https://oreil.ly/LXQNY>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_6/react_agent_with_memory.ipynb) 该资源是一个 Jupyter Notebook 文件, 名为 “react_agent_with_memory.ipynb”, 位于 GitHub 仓库 “BrightPool/prompt-engineering-for-generative-ai-examples” 的第六章节目录下。该笔记本展示了一个带有记忆功能的 React 智能体的完整实现示例, 属于生成式人工智能提示工程的相关内容。
11. <https://oreil.ly/M4z8K>: (<https://github.com/Significant-Gravitas/Auto-GPT>) 该资源是一个名为 AutoGPT 的开源项目, 基于 “计划-执行” 框架, 用于实现自动化的人工智能任务处理。它托管在 GitHub 上, 旨在让用户能够构建和运行能够自主规划和执行多步骤任务的智能代理。
12. <https://oreil.ly/TjrnF>: (https://python.langchain.com/docs/integrations/tools/google_search) 该资源是 LangChain 文档中关于 Google Search 工具集成的说明页面, 介绍了如何在 LangChain 框架中使用 Google 搜索功能。
13. https://oreil.ly/_v6dm: (<https://python.langchain.com/docs/integrations/toolkits/>) 该资源是关于 “Agent toolkits” 的文档页面, 介绍了在 LangChain 框架中集成和使用各种工具包 (toolkits) 以增强智能代理 (Agent) 的功能和能力。
14. <https://oreil.ly/fAeLo>: (<https://arxiv.org/abs/2201.11903>) 该资源是指向 arXiv 上的一篇学术论文, 内容很可能与 “Chain of Thought” (CoT) 方法相关, 介绍了通过让大型语言模型 (LLM) 分步骤思考以解决复杂问题的技术, 从而提升模型的推理能力和理解深度。
15. <https://oreil.ly/fub1z>: (https://github.com/langchain-ai/langchain/blob/master/cookbook/tree_of_thought.ipynb) 该资源是 LangChain 项目中的一个 Jupyter 笔记本文件, 展示了 “思路树 (Tree of Thought, ToT)” 方法的实际实现。通过以数独谜题为例, 演示如何在满足数独规则的前提下, 将通配符(*) 替换为数字, 从而体现 ToT 方法在复杂推理任务中的应用。
16. <https://oreil.ly/hYTus>: (<https://openai.com/blog/function-calling-and-other-api-updates>) 该资源介绍了 OpenAI 关于函数调用 (function calling)

功能及其他 API 更新的内容，重点讲解了 OpenAI 发布的更细化调优的大型语言模型（LLMs），这些模型专门优化用于函数调用，以提供除传统 ReAct 模式之外的另一种工具使用方案，同时仍将 LLM 作为推理引擎进行应用。

17. <https://oreil.ly/jyLab>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_6/custom_tools_memory_agent.ipynb) 该资源是一个位于 GitHub 上的 Jupyter Notebook 文件，文件名为“custom_tools_memory_agent.ipynb”，属于“prompt-engineering-for-generative-ai-examples”项目的第六章内容。该 Notebook 示例深入展示了如何使用 OpenAI 的函数调用机制和内存管理来构建智能代理（agents），演示了多个关键组件的集成与应用，帮助用户理解和实现具备记忆功能的自定义工具代理。
18. <https://oreil.ly/nTBox>: (<https://python.langchain.com/docs/integrations/memory>) 该资源介绍了与“数据库支持的记忆”（database-backed memory）相关的多种集成方案，这些方案能够将记忆的使用从短期、会话特定的上下文，转变为更为稳健的长期存储解决方案。
19. <https://oreil.ly/ssdnL>: (<https://arxiv.org/abs/2210.03629>) 该资源是指向 arXiv 上编号为 2210.03629 的学术论文，内容可能与“ReAct”框架相关。ReAct 框架结合任务分解、思维循环和多种工具来解决问题，该论文很可能详细介绍了该框架的设计原理、实现方法及其在复杂任务中的应用。
20. <https://oreil.ly/vZjm1>: (<https://python.langchain.com/docs/integrations/tools/requests>) 该资源是关于 Python 中 Requests 库的文档或教程，可能介绍如何在 LangChain 框架中集成和使用 Requests 工具进行 HTTP 请求操作。
21. <https://oreil.ly/x6FHN>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/tree/main/content/chapter_6) 该资源是一个 GitHub 代码仓库中的目录，位于路径“content/chapter_6”，包含与第六章相关的示例代码或文件，主要用于演示如何在 Jupyter Notebook 或 Python 文件中使用函数如 `create_csv_agent` 以及类 `ChatOpenAI` 和

`AgentType`，以基于大型语言模型（LLM）和数据集文件路径创建 CSV 代理（agent）。这些内容属于一个关于生成式 AI 提示工程（prompt engineering）的示例项目。

22. <https://oreil.ly/xiejG>: (<https://github.com/yoheinakajima/babyagi>) 该资源是一个名为“BabyAGI”的项目，属于计划与执行（plan-and-execute）框架的实现版本之一，通常用于构建具备自动任务规划和执行能力的智能代理系统。

第 7 章

1. <http://interiorai.com>: 该 URL (<http://interiorai.com>) 指向的是一个基于开源 AI 模型（如 Stable Diffusion）构建的商业应用服务，专注于利用人工智能技术进行室内设计相关的图像生成和创作。通过该平台，用户可以借助 AI 自动生成或优化室内空间的视觉效果，满足设计、展示或创意需求。
2. <https://civitai.com>: 该资源是一个名为“Civitai”的在线平台，主要用于分享和下载各种基于开源 AI 模型（如 Stable Diffusion）的模型文件，特别是包含成人内容（NSFW）的模型。平台支持爱好者和开发者访问、修改和扩展 AI 模型，从而促进模型的快速发展与创新应用。
3. <https://ideogram.ai>: 该资源是一个名为“Ideogram”的新兴图像生成模型项目网站，由前谷歌员工创立，于 2023 年 8 月发布，致力于在图像生成领域提供有竞争力的技术解决方案。
4. <https://oreil.ly/0inw3>: (<https://github.com/AUTOMATIC1111/stable-diffusion-webui>) 该 URL (<https://github.com/AUTOMATIC1111/stable-diffusion-webui>) 指向的是一个名为“AUTOMATIC1111’s web UI”的开源项目仓库。该项目为 Stable Diffusion 模型提供了一个功能丰富且用户友好的网页界面，方便用户访问和使用诸如参数调节（如 Classifier Free Guidance、Denoising、Seed）、高级图像生成功能（如 inpainting、outpainting）、以及

集成了 ControlNet 和 Segment Anything 等扩展功能。该界面极大地推动了 Stable Diffusion 模型的普及和创新，使爱好者和开发者能够更便捷地自定义、扩展和应用该模型。

5. <https://oreil.ly/2mgh5>: (<https://twitter.com/levelsio/status/1680665706235404288?s=20>) 该 URL 指向 Twitter 用户 levelsio 发布的一条推文，内容可能与 Stable Diffusion 模型的发布、版本差异、训练数据集（如 LAION-5B）或模型性能相关的讨论有关，尤其是围绕版本 2.0 中对 NSFW 图像的过滤及其对生成效果的影响。
6. <https://oreil.ly/341QB>: (<https://www.theverge.com/2022/11/2/23434361/google-text-to-image-ai-model-imagen-test-kitchen-app>) 该资源是一篇发表于 2022 年 11 月的 The Verge 文章，介绍了谷歌开发的文本生成图像 AI 模型“Imagen”及其相关测试应用“Test Kitchen”。文章讨论了谷歌在图像生成领域的进展、Imagen 模型的特点以及该技术的潜力和挑战。
7. <https://oreil.ly/3ISZk>: (https://en.wikipedia.org/wiki/Reinforcement_learning_from_human_feedback) 该资源介绍了“从人类反馈中进行强化学习”（Reinforcement Learning from Human Feedback, RLHF）的方法，这是一种通过利用人类的评价和反馈来训练和优化机器学习模型的技术。RLHF 被认为是像 ChatGPT 等大型语言模型取得成功的关键技术之一。具体来说，它结合了人类的偏好和指导，用以提升模型的表现和生成内容的质量。
8. <https://oreil.ly/57szp>: (<https://scholar.harvard.edu/binxuw/classes/machine-learning-scratch/materials/foundation-diffusion-generative-models>) 该资源是由 Binxu Wang 在哈佛大学开设的机器学习课程中提供的关于扩散生成模型（Diffusion Generative Models）数学基础的教学材料，详细介绍了扩散模型的去噪过程及其理论原理。
9. <https://oreil.ly/BT-k5>: (<https://techcrunch.com/2022/10/17/stability-ai-the-startup-behind-stable-diffusion-raises-101m/>) 该资源是一篇 TechCrunch 的新闻报道，介绍了人工智能公司 Stability AI 完成了超过 1 亿美元的融资，以及其在生成式 AI 领域的最新进展，包括推出了改进图像中真实文本生成能力的 DeepFloyd 模型和性能更强大的 Stable Diffusion XL 1.0 模型。

-
10. <https://oreil.ly/CsJgT>: (<https://animatediff.github.io/>) 该资源 (<https://animatediff.github.io/>) 是一个名为 AnimateDiff 的开源项目, 致力于基于 Stable Diffusion 模型扩展, 实现从文本到视频、图像到视频乃至视频到视频的生成技术, 能够逐帧生成连贯一致的动画效果。该项目是推动文本到视频生成领域的重要开源工具之一。
 11. <https://oreil.ly/EqdtP>: (<https://arxiv.org/abs/2204.06125>) 该资源是指向一篇关于 DALL-E 2 模型的论文, 发布于 2022 年 4 月, 介绍了 OpenAI 新一代图像生成模型 DALL-E 2。该论文展示了该模型在图像生成质量上的显著提升, 并引发了广泛关注与使用者排队等待体验的热潮。论文中包含了著名的“宇航员骑马”图像示例, 体现了模型强大的创意生成能力。
 12. <https://oreil.ly/JKZzD>: (<https://discord.com/invite/midjourney>) 该资源是 Midjourney 的官方 Discord 服务器。用户需加入该服务器, 通过在开放频道或私信中提交文本提示来使用 Midjourney 的图像生成工具。服务器内的社区高度活跃, 生成的图像大多公开共享, 促进用户之间的学习和快速迭代。该 Discord 社区用户数量庞大, 体现了 Midjourney 的强大社区支持和互动氛围。
 13. <https://oreil.ly/K5vX2>: (<https://laion.ai/blog/laion-5b/>) 该资源是 LAION-5B 数据集的相关介绍或博客页面。LAION-5B 是一个大规模的图像与文本配对数据集, 主要用于人工智能研究, 尤其是训练图像生成模型如 Stable Diffusion。该页面可能详细说明了数据集的规模、构成、用途及其在训练开源图像生成模型中的重要作用。
 14. <https://oreil.ly/K8oWv>: (<https://imagen.research.google>) 该资源是 Google 推出的“Imagen”图像生成模型的官方网站, 展示了 Google 在基于扩散模型 (diffusion models) 技术的图像生成领域的研究成果和相关信息。Imagen 模型虽未公开发布, 但代表了 Google 在人工智能图像生成技术上的重要探索 and 实力。
 15. <https://oreil.ly/OV46r>: (<https://docs.midjourney.com/docs/plans>) 该资源为 Midjourney 官方文档中关于服务订阅计划的说明页面, 介绍了 Midjourney 提供的不同价格方案及其对应的功能 (如私密模式仅在最高价位方案中开放), 帮助用户了解各类套餐的区别及选择合适的订阅计划。

-
16. <https://oreil.ly/OrAHA>: (<https://scholar.harvard.edu/binxuw/classes/machine-learning-scratch/materials/foundation-diffusion-generative-models>) 该资源是一份关于扩散生成模型 (diffusion generative models) 基础知识的教学材料, 介绍了扩散模型的训练原理和过程, 包括通过多步向图像添加随机噪声, 再通过去噪 (denoising) 逆转扩散过程, 从而生成符合描述的图像。内容结物理中的扩散现象, 讲解了模型如何通过调整神经网络权重, 使其能够根据描述从噪声中生成对应图像。
 17. <https://oreil.ly/SGn9B>: (<https://arxiv.org/abs/2301.13188>) 该资源是一篇由 Carlini 等人在 2023 年发表的学术论文, 探讨了扩散模型 (如 Stable Diffusion) 是否直接复制训练数据中的图像。研究结果表明, 尽管模型经过大量图像训练, 但实际上很难从模型中复现训练集中的绝大多数图像, 仅成功复现了 109 张图像, 表明模型主要是通过学习数据分布生成新图像, 而非简单复制已有作品。该论文对理解扩散模型的版权和生成机制具有重要意义。
 18. <https://oreil.ly/UCQ3I>: (<https://github.com/deep-floyd/IF>) 该资源是 DeepFloyd 项目的 GitHub 代码仓库, DeepFloyd 是由 Stability AI 开发的一种先进的图像生成模型, 专注于在生成图像时更准确地生成真实文本内容, 解决了其他模型在图像文本生成方面存在的问题。
 19. <https://oreil.ly/UuApM>: (<https://stability.ai/news/stable-video-diffusion-open-ai-video-model>) 该资源介绍了 “Stable Video Diffusion” 这一基于 Stable Diffusion 模型的开源文本生成视频技术, 发布于 2023 年 11 月, 能够将文本转化为短视频片段或为已有图像制作动画, 标志着 Stable Diffusion 社区在文本到视频及图像到视频生成领域的重要进展。
 20. <https://oreil.ly/XzQrU>: (<https://gemini.google.com/>) 该资源是谷歌 (Google) 推出的 “Gemini” 系列人工智能模型相关页面, Gemini 1.5 是谷歌在图像生成和 AI 多模态领域与 OpenAI 等竞争对手展开较量的最新技术产品, 展示了谷歌在 AI 生成模型方面的进展和应用。
 21. https://oreil.ly/YrmA_: (<https://uk.pcmag.com/news/146207/ai-image-generator-midjourney-pauses-free-trials-amid-huge-demand-abuse>) 该资源是一篇来自 PCMag 英国站的新闻报道, 标题为 “AI 图像生成器

Midjourney 因巨大需求和滥用暂停免费试用”。文章介绍了 Midjourney 团队规模小（截至 2023 年 3 月仅 11 人），但在没有大量资金支持的情况下，依然开发出比 DALL-E 更强大的功能，如负面提示、加权词汇和图像描述反向生成等。同时，Midjourney 目前仅通过 Discord 平台提供服务，缺乏 API 接口，这可能限制了其更广泛的应用和推广。文章重点报道了 Midjourney 因用户需求激增及滥用问题，暂停了免费试用的决定。

22. <https://oreil.ly/a4Fyp>: (<https://www.theverge.com/2023/12/4/23988403/getty-lawsuit-stability-ai-copyright-infringement>) 该链接指向一篇《The Verge》于 2023 年 12 月 4 日发布的报道，内容涉及 Getty Images 针对 Stability AI 提起的版权侵权诉讼，聚焦于人工智能图像生成模型因使用大量网络图片数据进行训练而引发的版权争议。
23. <https://oreil.ly/cjm8A>: (https://keras.io/examples/generative/random_walks_with_stable_diffusion/) 该资源是一个基于 Stable Diffusion 的生成式模型示例，演示如何通过随机游走（random walks）技术在图像生成过程中，从一个图像逐步过渡到另一个图像，展示了生成图像的中间步骤和变化过程。
24. <https://oreil.ly/dalle1>: (<https://openai.com/research/dall-e>) 该资源是 OpenAI 于 2021 年 1 月发布的文本生成图像模型 DALL-E 的研究介绍页面。DALL-E 基于改进版的 GPT-3 模型，能够根据文本描述生成高质量的图像，展示了生成式人工智能在艺术创作方面的突破性能力。页面详细介绍了 DALL-E 的技术背景、创新点及其在文本到图像生成领域的应用。
25. <https://oreil.ly/dalle2>: (<https://openai.com/dall-e-2>) 该资源是 OpenAI 官方页面，介绍和展示了 DALL-E 2——一种基于扩散模型的先进图像生成技术，能够根据文本描述生成高质量的图像。
26. <https://oreil.ly/dalle3>: (<https://openai.com/dall-e-3>) 该资源介绍了 OpenAI 最新发布的图像生成模型 DALL-E 3，强调其作为扩散模型在文本生成图像领域的先进性。DALL-E 3 已集成入 ChatGPT，实现了文本与图像生成的无缝结合，提升了生成图像的质量和交互体验。该页面可能包含关于 DALL-E 3 的功能特点、应用场景以及与其他同类模型（如 DALL-E 2、Stable Diffusion、Midjourney 等）的比较分析。

-
27. <https://oreil.ly/dqshh>: (https://en.wikipedia.org/wiki/Th%C3%A9%C3%A2tre_d%27Op%C3%A9ra_Spatial) 该 URL (https://en.wikipedia.org/wiki/Th%C3%A9%C3%A2tre_d%27Op%C3%A9ra_Spatial) 指向维基百科中关于“Théâtre d’ Opéra Spatial”的条目页面。根据 URL 中的法语名称推断, 该资源很可能介绍一个名为“Théâtre d’ Opéra Spatial”的空间歌剧剧院, 可能涉及该剧院的历史、建筑特色、演出活动及其文化意义等内容。
 28. <https://oreil.ly/gcT4t>: (<https://stability.ai/stablediffusion>) 该资源是 Stable Diffusion 模型的官方网站页面, 介绍由 Stability AI 开发的先进图像生成模型 Stable Diffusion, 特别是其最新版本 Stable Diffusion XL 1.0。该页面可能包含模型的技术详情、版本更新、下载和使用信息, 展示了该模型在生成高质量图像方面的能力及其参数规模。
 29. https://oreil.ly/gjNJ_: (<https://huggingface.co/docs/diffusers/en/using-diffusers/sdxl>) 该资源是 Hugging Face 官方文档中关于使用 Diffusers 库进行图像生成的教程和说明页面, 具体聚焦于 SDXL (Stable Diffusion XL) 模型的使用方法。文档介绍了如何通过 Diffusers 框架调用和操作先进的扩散模型 (如 SDXL) 来实现从文本到高质量图像的生成, 适合希望深入掌握扩散模型技术及其应用的开发者和研究人员。
 30. <https://oreil.ly/j51L0>: (<https://www.midjourney.com/home>) 该资源是 Midjourney 的官方网站主页, Midjourney 是一个基于扩散模型的图像生成平台, 用户可以通过文本描述生成高质量的艺术图像, 属于近年来在文本到图像生成领域备受欢迎的工具之一。
 31. <https://oreil.ly/jeFYV>: (https://www.theregister.com/2022/08/01/david_holz_midjourney/) 该资源是一篇发表于 2022 年 8 月 1 日的新闻报道, 介绍了 Midjourney 创始人 David Holz 及其团队在人工智能图像生成领域的成就。文章重点讲述了 Midjourney 团队规模小、未依赖大量资金却实现了盈利, 并开发了超越 DALL-E 的功能, 如负面提示、加权词汇和图像反向描述等。同时, 文章提到 Midjourney 目前仅通过 Discord 平台提供服务, 缺乏 API 接口, 影响了其更广泛的应用和普及。

-
32. <https://oreil.ly/17KHB>: (<https://stable-diffusion-art.com/video-to-video/>) 该资源是一个关于“video-to-video”技术及其在 Stable Diffusion 社区中应用的介绍页面，主要内容包括利用扩散模型实现视频帧间一致性的视频生成与转换技术，展示了相关开源项目（如 AnimateDiff）以及 Stable Diffusion 联合创始人 RunwayML 在文本到视频生成领域的最新进展和工具（如 Gen-2 模型、Stable Video Diffusion、Stable Diffusion Turbo 和 Sora）。页面旨在说明视频生成技术的发展趋势及其与图像生成提示技术的关联。
33. <https://oreil.ly/ot4vw>: (<https://www.newscientist.com/article/2329690-ai-art-tool-dall-e-2-adds-black-or-female-to-some-image-prompts/>) 该资源是一篇来自《New Scientist》的文章，报道了 AI 图像生成工具 DALL-E 2 在处理带有人物的图像提示时，因数据集偏见问题而自动向某些提示中添加“black”或“female”等词汇的现象。这一行为被视为该工具试图纠正其训练数据中存在的种族和性别偏见，但方式较为生硬。文章讨论了 DALL-E 2 在伦理和安全方面的考虑，以及其在生成包含人物的图像时所采取的限制措施。
34. <https://oreil.ly/pwPGX>: (<https://twitter.com/justintrimble/status/1598775188778123265?s=20>) 该 URL 指向的一条 Twitter 状态（推文），很可能是由 Justin Trimble 发布，内容与 Stable Diffusion 图像生成模型相关。推文可能讨论了 Stable Diffusion 模型的开源发布、其计算资源支持、性能表现或其 AI 生成图像领域内的影响和流行度，呼应了上下文中提到的 Stable Diffusion 的迅速流行和技术突破。
35. <https://oreil.ly/qOpis>: (https://keras.io/examples/generative/random_walks_with_stable_diffusion/) 该资源是 Keras 官网上的一个示例项目，展示了如何结合 Stable Diffusion 模型进行生成式随机游走（random walks）的实现与应用，旨在通过编码和解码潜在空间中的路径来生成图像，帮助用户理解和实践基于 Stable Diffusion 的生成模型技术。
36. <https://oreil.ly/sFaeW>: (<https://arxiv.org/abs/2205.11487>) 该 URL “<https://arxiv.org/abs/2205.11487>” 指向的是 2022 年 5 月发布的一篇关于 Google 的 Imagen 模型的论文。该论文介绍了一种先进的基于扩散模型的图像生

成技术，展示了在文本驱动图像生成任务中的出色效果。尽管该模型在学术上取得了显著成果，但出于 AI 伦理和安全方面的考虑，Google 并未将其公开发布。

37. <https://oreil.ly/sora>: (<https://openai.com/sora>) 该资源是 OpenAI 于 2024 年 2 月发布的“Sora”，代表 OpenAI 在文本生成视频 (text-to-video) 领域的重要进展，展示了其在视频生成技术上的最新成果和应用潜力。
38. <https://oreil.ly/u-Glg>: (<https://www.theverge.com/2024/2/21/24079371/google-ai-gemini-generative-inaccurate-historical>) 该资源是一篇发表于 2024 年 2 月的 The Verge 文章，报道了谷歌在生成式人工智能领域的新进展，特别是其名为“Google Gemini”的生成模型。文章讨论了 Google Gemini 在图像生成等方面的表现，以及其面临的准确性和历史事实错误等挑战，同时提及谷歌此前未公开发布的 Imagen 模型和前谷歌员工创立的 Ideogram 等相关内容。
39. <https://oreil.ly/uMAkh>: (<https://stability.ai/news/stability-ai-sdxl-turbo>) 该资源介绍了 Stability AI 发布的“Stable Diffusion XL Turbo”（简称 SDXL Turbo），这是一款高性能的图像生成模型，能够实现近乎实时的图像生成。文中提到 SDXL Turbo 在 2023 年 11 月推出，支持从文本生成短视频和对现有图像进行动画处理，体现了 Stable Diffusion 社区在扩展模型能力、推动文本到视频、图像到视频等多模态生成方向的最新进展。该资源可能详细介绍了 SDXL Turbo 的功能特点、应用场景以及其在快速生成高质量视觉内容方面的优势。
40. <https://oreil.ly/vS0mA>: (<https://runwayml.com/ai-magic-tools/gen-2/>) 该资源 (<https://runwayml.com/ai-magic-tools/gen-2/>) 是 RunwayML 推出的“Gen-2”模型，属于领先的文本生成视频 (text-to-video) 技术工具。它能够将文本描述转换成短视频片段，或对现有图像进行动画处理，实现基于文本的高质量视频生成，是当前在稳定扩散 (Stable Diffusion) 视频生成领域的重要创新成果。
41. <https://photoai.com>: 该资源“<https://photoai.com>”指向的是一个基于开源 AI 图像生成模型 Stable Diffusion 构建的在线平台或服务，名为 PhotoAI。该平台利用开源模型的灵活性和可调节参数，为用户提供图像生成和编辑功能，支持创意艺术创作和商业应用，且已实现显著的商业收入。

42. <https://www.headshotpro.com>: 该资源 Headshot Pro (<https://www.headshotpro.com>) 是一款基于开源 AI 模型 (如 Stable Diffusion) 构建的商业服务, 专注于通过 AI 技术生成或优化专业人像照片 (头像)。它利用先进的图像生成和编辑功能, 为用户提供高质量、个性化的头像制作解决方案, 适用于个人品牌展示、职业形象打造等多种场景。

第 8 章

1. <https://lexica.art>: 该资源 Lexica.art 是一个提供大量 Stable Diffusion 生成图像及其对应提示词的可搜索数据库网站, 用户可以通过关键词或相似图像进行搜索, 以获得创作灵感。
2. https://oreil.ly/0c_en: (<https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Features#outpainting>) 该 URL 指向的资源是 “stable-diffusion-webui” 项目的 Wiki 页面中关于 “outpainting” 功能的介绍。该页面详细说明了如何通过 outpainting 技术扩展现有图像, 尤其是在调整图像宽高比 (如将纵向图片扩展为横向) 时填补图像空白区域的应用方法。同时, 页面还介绍了使用该功能时的操作技巧 (如自定义缩放和分区域提示), 并与其他工具 (如 Midjourney) 做了对比, 说明了其可靠性和使用体验。
3. <https://oreil.ly/0oO4w>: (https://unsplash.com/photos/a-man-and-a-woman-standing-next-to-each-other-on-a-beach-D95hTrZrdg4?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 该 URL 指向的是一张由摄影师 Jessica Hearn 拍摄并发布在 Unsplash 上的照片, 内容是一位男性和一位女性并肩站在海滩上的场景。该图片常被用作 AI 图像生成工具中的示例图像 (base image), 用于通过图像提示 (Img2Img 或类似功能) 帮助模型生成符合特定风格和内容的原创图像。

4. <https://oreil.ly/2qy4E>: (<https://bakztfuture.substack.com/p/dall-e-2-unbundling>) 该资源是一篇由 Bakz T. Future 撰写的文章，发布在 Substack 平台上，标题大致涉及“DALL·E 2 中的风格拆解 (Unbundling)”的内容。文章介绍了如何通过“unbundling”技术——即提取艺术家风格的特征并加以描述——来指导生成模型（如 DALL·E 2 或 ChatGPT-4）创作出既具有该艺术家风格特点又不完全复制其作品的图像，从而实现类似萨尔瓦多·达利 (Dali) 风格但独特的艺术效果。
5. <https://oreil.ly/7DhZE>: (<https://docs.midjourney.com/docs/vary-region>) 该资源介绍了 Midjourney 中名为“Vary Region”的功能，该功能实现了图像的局部修复 (inpainting)，允许用户选择图像的特定区域进行重新生成或修改，从而达到局部调整图像内容的效果。
6. <https://oreil.ly/9A1NL>: (<https://huggingface.co/Gustavosta/MagicPrompt-Stable-Diffusion>) 该资源是一个位于 Hugging Face 平台上的项目，名为“MagicPrompt-Stable-Diffusion”，其主要功能是为 Stable Diffusion 等文本生成图像模型自动生成或优化提示词 (prompt)。该项目致力于通过 AI 模型提升提示词的质量，从而帮助用户获得更高质量的图像生成效果，属于提示词工程 (prompt engineering) 工具的一种实现。
7. <https://oreil.ly/B6E0Y>: (https://unsplash.com/@jessica_hearn?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 该 URL (https://unsplash.com/@jessica_hearn?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 指向的是摄影师 Jessica Hearn 在图片分享网站 Unsplash 上的个人主页。该主页展示了 Jessica Hearn 上传的免费高质量照片，用户可以浏览和下载她的作品，用于各种用途，包括作为 AI 图像生成的示例图片素材。
8. <https://oreil.ly/BQYFP>: (<https://marketingmemetics.com>) 该资源“<https://marketingmemetics.com>”很可能是一个专注于“营销模因学” (marketing memetics) 的平台或网站，介绍如何利用“模因” (memes) 这一文化传播单元，通过分析和拆解艺术风格或文化元素，帮助创作者和营销人员创新表达和推广内容，实现更有效的文化传播和品牌塑造。

-
9. <https://oreil.ly/BaITC>: (<https://x.com/rainisto/status/1709901053053854097?s=20>) 该资源是一条推文，内容展示了利用图像修补 (inpainting) 和扩展 (outpainting) 技术，通过对已有图像的局部修改来保持图像中角色或元素的一致性，尤其是在生成左右并排的两张图像时逐步修补，从而实现角色或元素的连续和一致表现。
 10. <https://oreil.ly/DLqAV>: (<https://marketingmemetics.com/meme-mapping>) 该资源是一篇关于 “meme mapping” (模因映射) 方法的介绍或指南，讲述如何通过系统地识别和分析图像及其相关文字中的反复出现的模式 (模因)，以帮助用户发现和构建有效的提示词 (prompts)。该方法旨在辅助用户从已有的视觉和文本示例中提取灵感，从而更高效地生成或改进创作内容。
 11. <https://oreil.ly/EZmRs>: (<https://app.vexpower.com/sim/what-is-working-in-our-creative/>) 该 URL (<https://app.vexpower.com/sim/what-is-working-in-our-creative/>) 指向的资源很可能是一个关于 “创意内容效果分析” 或 “创意素材表现评估” 的页面，介绍如何通过数据收集、自动化标签及分析技术 (如图像识别、自然语言处理等) 来识别和总结哪些创意元素在营销或内容制作中表现良好，从而辅助优化创意策略。
 12. <https://oreil.ly/EwLNh>: (<https://www.reddit.com/r/dalle2/>) 该资源是 Reddit 上的一个社区版块，专注于与 DALL·E 2 相关的内容交流。用户在该版块中分享、讨论和展示使用 DALL·E 2 生成的图像、提示词 (prompts) 以及相关的创意和技术经验，是一个 AI 图像生成爱好者和创作者的互动平台。
 13. <https://oreil.ly/FvGAi>: (<https://www.adobe.com/uk/products/photoshop/generative-fill.html>) 该资源介绍了 Adobe Photoshop 中的 “生成填充 (Generative Fill)” 功能，这是一种基于 AI 的图像修复和重绘技术，利用类似 Stable Diffusion 的图像修补 (inpainting) 方法，允许用户通过智能生成内容来填补或修改图片的指定区域，实现图像的局部编辑与创作。
 14. <https://oreil.ly/GXfDl>: (<https://app.vexpower.com/sim/what-words-work-best-in-my-ads/>) 该 URL “<https://app.vexpower.com/sim/what-words-work-best-in-my-ads/>” 指向的资源很可能是一个在线工具或应用，帮

助用户分析和识别广告文案中哪些词语效果最佳，从而优化广告内容，提高点击率和转化率。

15. <https://oreil.ly/GdNrt>: ([https://unsplash.com/photos/vdXMSiX-n6M\[Photo\]byhttps://unsplash.com/@mimithian?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText](https://unsplash.com/photos/vdXMSiX-n6M[Photo]byhttps://unsplash.com/@mimithian?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText)) 该 URL 指向的是摄影师 Mimi Thian 在 Unsplash 网站上发布的一张照片。根据上下文，该照片被用作 Midjourney Describe 功能的示例图像，展示了该功能的应用场景或效果。
16. <https://oreil.ly/OirCS>: (<https://twitter.com/dannypostmaa/status/1583020277495046144?s=20>) 该资源是一条推特 (Twitter) 帖子，内容可能与“meta prompting” (元提示) 相关，介绍了在将 AI 系统投入生产时，用户通常不是专业的提示工程师，导致输入的提示质量参差不齐，从而影响系统输出效果。该推文可能进一步讨论了通过由另一个 AI 模型重写提示以提升结果质量的行业常用技巧。
17. <https://oreil.ly/OmZbl>: (<https://artsandculture.google.com/category/art-movement>) 该资源是 Google 艺术与文化平台中关于各种艺术流派的分类页面，提供了丰富的艺术运动 (art movements) 相关信息，帮助用户了解不同艺术流派的特点、历史背景及代表作品，适合对艺术技法和风格感兴趣的人士参考和学习。
18. <https://oreil.ly/VqyG->: (<https://app.vexpower.com/sim/what-attributes-should-we-talk-about/>) 该资源 (URL: <https://app.vexpower.com/sim/what-attributes-should-we-talk-about/>) 很可能是一个关于如何识别和讨论某些核心属性或特征的工具或指南，尤其是在数据整理、标签标注或内容分析的上下文中。结合上下文中的内容，它可能涉及利用自动化或程序化的方法，对收集到的数据 (如图片、文本等) 进行属性识别和模式分析，从而高效地辅助“meme mapping” (梳理和分析网络迷因) 的流程。总体来看，该资源旨在帮助用户确定在分析或讨论某类数据时，应关注哪些关键属性，以便更系统、更大规模地理解和处理信息。
19. <https://oreil.ly/WNsRn>: (<https://lexica.art/prompt/02813962-87d0-49bd-b87f-9034440ccb01>) 该资源是 Lexica 网站上一个具体的图像生成提示

- (prompt) 页面, 地址为 “<https://lexica.art/prompt/02813962-87d0-49bd-b87f-9034440ccb01>”。该页面展示了与特定关键词相关的 AI 生成图像及其对应的文本提示, 用户可以通过参考这些提示来启发和优化自己的图像生成创作过程。
20. <https://oreil.ly/YgL8g>: (<https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Features#inpainting>) 该资源介绍了基于 Stable Diffusion 的图像修复功能——“inpainting” (图像修补/重绘), 详细说明了该技术在 AUTOMATIC1111 的 stable-diffusion-webui 项目中的实现与特点。页面内容涵盖 inpainting 的基本原理、使用方法, 以及与其他平台 (如 Midjourney 的 “Vary Region” 功能、Adobe Photoshop 的 “Generative Fill” 以及 DALL-E 的开创性应用) 之间的比较, 帮助用户理解如何利用该功能在已有图像中选择性地擦除并重新生成部分内容, 从而实现图像局部编辑和创作。
 21. <https://oreil.ly/afGCQ>: (<https://docs.google.com/spreadsheets/d/1-snKDn38-KypoYCK9XLPg799bHcNFSBAVu2HVvFEAkA/edit?usp=sharing>) 该资源是一个由作者之一创建的 “Prompt Engineering Template” (提示工程模板), 用于辅助提升 AI 图像生成中的标签质量和艺术风格选择, 帮助用户更好地设计和优化生成图像的提示词, 从而获得更高质量和符合审美的作品。
 22. <https://oreil.ly/bEEnJ>: (https://unsplash.com/photos/vdXMSiX-n6M?utm_source=unsplash&utm_medium=referral&utm_content=creditCopyText) 该 URL 指向的是 Unsplash 网站上的一张由摄影师 Mimi Thian 拍摄并发布的照片, 图片在文中作为 “Figure 8-6. Midjourney Describe” 的插图素材使用。
 23. <https://oreil.ly/cOcPR>: (<https://platform.openai.com/docs/guides/vision>) 该资源 (URL: <https://platform.openai.com/docs/guides/vision>) 是 OpenAI 官方文档中关于 “视觉 (Vision)” 功能的指南, 介绍如何使用 OpenAI 的多模态视觉模型 (如 GPT-4 Vision) 进行图像识别、实体标注和图像内容分析, 帮助开发者实现结合图像和文本的智能处理与分析。
 24. <https://oreil.ly/fzgno>: (<https://huggingface.co/spaces/pharma/CLIP-Interrogator>) 该 URL 指向的资源是一个名为 “CLIP Interrogator” 的开源项目, 提供图像内容分析和描述生成的工具。该工具能够识别上传图像中的元素及

其内容，并生成相应的文本提示（prompt），以辅助图像生成或理解。不过，相较于 Midjourney，该工具在生成的提示丰富度和对上传图像风格的复现能力方面存在一定差距。

25. <https://oreil.ly/mhujK>: (<https://artsandculture.google.com/category/art-movement>) 该资源（URL: <https://artsandculture.google.com/category/art-movement>）是 Google 艺术与文化平台上关于“艺术流派”（Art Movements）的专题页面，汇集和介绍了各种艺术运动的分类和相关内容，帮助用户了解不同艺术风格及其历史背景，便于学习和识别多样的艺术风格名称。
26. <https://oreil.ly/nnam3>: (<https://www.technologyreview.com/2022/09/16/1059598/this-artist-is-dominating-ai-generated-art-and-hes-not-happy-about-it/>) 该资源是一篇发表于《麻省理工科技评论》（MIT Technology Review）的文章，标题大意为“这位艺术家正主导 AI 生成艺术，但他对此并不满意”。文章聚焦于 AI 图像生成技术如何能够复制流行艺术风格，特别是当代艺术家如波兰数字艺术家 Greg Rutkowski 的风格被广泛模仿的现象。同时，文章探讨了艺术家们对 AI 艺术的态度分歧以及围绕模仿在版权法下是否属于合理使用的法律灰色地带。
27. <https://oreil.ly/upQIh>: (<https://discord.com/servers/midjourney-662267976984297473>) 该 URL 链接指向 Midjourney 的官方 Discord 服务器，这是一个拥有数百万活跃成员的社区，用户在其中分享和交流使用 Midjourney 生成的图像、提示词（prompts）以及相关创作经验。该服务器是 AI 艺术创作者获取灵感、展示作品和互动的重要平台。
28. <https://oreil.ly/wPuMo>: (<https://www.youtube.com/watch?v=nJPERZDfyWc>) 该资源是一段 YouTube 视频，内容可能与数字艺术创作、风格解析或 AI 辅助图像生成相关，特别是涉及如何通过拆解和重组达利（Salvador Dalí）风格中的元素（如超现实主义风景、融化物体、梦境氛围等）来创造更具原创性的图像作品。视频可能演示或讲解利用 AI 工具（如 ChatGPT）和视觉“模因”进行艺术风格重混（remix）的具体方法和技巧。

29. <https://pika.art>: 该资源是一个名为 Pika Labs 的平台, 提供基于图像提示的 AI 视频生成工具, 方便用户通过图像输入来迭代和调整视频风格, 提升视频生成的效率和效果。
30. <https://runwayml.com>: 该资源是 RunwayML, 一个提供 AI 视频生成和创意工具的平台, 用户可以通过输入图像和参数来生成或编辑视频内容, 支持基于图像的提示以便高效迭代和风格调整。

第 9 章

1. <http://127.0.0.1:7860>: 该资源是指向本地主机 (localhost) 上运行的一个网络服务, 监听端口 7860。通常, 这类 URL 用于访问本地部署的网页应用、开发环境或后台管理界面, 可能是某个本地服务器或应用程序的用户界面入口。
2. <https://civitai.com>: 该 URL (<https://civitai.com>) 指向的是一个公开分享和下载各类 Stable Diffusion 及相关 AI 图像生成模型的平台或社区网站。用户可以在该网站上获取包括 DreamBooth 模型在内的多种自定义及预训练模型, 用于文本生成图像、图像修复、放大等多种 AI 绘图任务。
3. <https://github.com/AUTOMATIC1111/stable-diffusion-webui>: 该资源是一个 GitHub 仓库, 地址为 <https://github.com/AUTOMATIC1111/stable-diffusion-webui>, 通常用于通过 git clone 命令克隆。结合仓库名称 “stable-diffusion-webui” 可以推断, 该仓库提供了一个基于 Stable Diffusion 模型的网页用户界面 (Web UI), 方便用户通过浏览器使用和管理 Stable Diffusion 相关的图像生成或处理功能。
4. <https://github.com/Mikubill/sd-webui-controlnet>: 该资源是一个 GitHub 仓库, 名为 “sd-webui-controlnet”, 由用户 Mikubill 维护, 可能与 “Stable Diffusion” 或相关的 Web 用户界面 (web UI) 插件或扩展功能 “ControlNet” 有关。

5. <https://github.com/alemelis/sd-webui-ar>: 该资源是一个托管在 GitHub 上的项目仓库, 地址为 <https://github.com/alemelis/sd-webui-ar>, 可能与 “sd-webui” 相关的某个扩展、插件或工具, 具体内容需访问仓库查看详细说明。
6. <https://github.com/continue-revolution/sd-webui-segment-anything>: 该资源是一个托管在 GitHub 上的开源项目, 名为 “sd-webui-segment-anything”, 很可能是一个基于 Stable Diffusion (简称 sd) 模型的 Web 用户界面 (WebUI) 扩展或工具, 集成了 “Segment Anything” 功能, 用于图像分割或相关的计算机视觉任务。
7. <https://openmodeldb.info>: 该资源是一个提供各种图像处理和生成模型下载的平台, 用户可以在 <https://openmodeldb.info> 上找到多种预训练模型 (如用于图像放大和增强的模型), 以便替换或补充默认模型, 从而提升图像的处理效果。
8. <https://oreil.ly/0DIbv>: (https://huggingface.co/stabilityai/stable-diffusion-xl-refiner-1.0/blob/main/sd_xl_refiner_1.0.safetensors) 该资源是一个名为 “sd_xl_refiner_1.0.safetensors” 的 Refiner 模型文件, 托管在 Hugging Face 的 [stabilityai/stable-diffusion-xl-refiner-1.0](https://huggingface.co/stabilityai/stable-diffusion-xl-refiner-1.0) 仓库中, 通常用于对 Stable Diffusion XL 生成的图像进行细化和提升质量。
9. <https://oreil.ly/2FlyI>: (https://unsplash.com/photos/brown-concrete-palace-surrounded-by-body-of-water-during-daytime-8tS8ywl126Q?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 该资源是由摄影师 Richard Clark 在 Unsplash 网站上发布的一张现代城堡的照片, 图片展示了一座棕色混凝土建筑风格的宫殿, 周围被水体环绕, 拍摄于白天。
10. <https://oreil.ly/2WGxQ>: (https://colab.research.google.com/?utm_source=scs-index) 该资源是指向 Google Colab 平台的链接, Google Colab 是一个基于云端的交互式 Python 编程环境, 允许用户无需本地配置即可运行和共享 Jupyter 笔记本, 适合进行数据分析、机器学习等任务。链接中带有的参数可能用于跟踪访问来源。
11. <https://oreil.ly/2efOO>: (<https://huggingface.co/sd-dreambooth-library>) 该 URL (<https://huggingface.co/sd-dreambooth-library>) 指向的是

Hugging Face 上的一个 DreamBooth 模型库资源，提供多种基于 DreamBooth 技术训练的图像生成模型。用户可以下载这些模型并将其集成到如 AUTOMATIC1111 的 stable-diffusion-webui 中，以便通过特定的触发词或标记生成具有特定风格或主题的图像。

12. <https://oreil.ly/790FZ>: (https://colab.research.google.com/github/ShivamShrirao/diffusers/blob/main/examples/dreambooth/DreamBooth_Stable_Diffusion.ipynb) 该 URL 指向一个 Google Colab 笔记本，内容是基于 Shivam Shrirao 的 GitHub 仓库中使用 HuggingFace 的 diffusers 库实现的 DreamBooth 模型训练示例。该笔记本使用的是 Stable Diffusion 1.5 版本，适合在免费 Google Colab 环境的 Nvidia GPU 上训练，能够在几小时内完成模型微调。
13. <https://oreil.ly/7n02i>: (https://en.wikipedia.org/wiki/Mona_Lisa) 该资源是维基百科中关于《蒙娜丽莎》(Mona Lisa) 这幅著名画作的页面，介绍了这幅由达·芬奇创作的经典肖像画的历史背景、艺术特色及其文化影响。
14. <https://oreil.ly/AJnnL>: (<https://github.com/ShivamShrirao/diffusers/tree/main/examples/dreambooth>) 该 URL (<https://github.com/ShivamShrirao/diffusers/tree/main/examples/dreambooth>) 指向的是 Shivam Shrirao 维护的一个基于 HuggingFace 的 diffusers 库的 DreamBooth 模型训练示例代码仓库。该资源包含用于训练 DreamBooth 模型的示范代码，支持使用较小的模型版本（如 1.5 版）在 Google Colab 环境中进行快速训练，适合于有 Nvidia GPU 的用户进行个性化图像生成模型的开发和实验。
15. <https://oreil.ly/BOKJ7>: (https://unsplash.com/photos/band-practicing-on-studio-By96LAr-34o?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 该 URL 指向的是一张由摄影师 Hans Vivek 拍摄并发布在 Unsplash 上的图片，内容是一支乐队正在录音室排练。该图片用于展示深度图 (depth map) 在图像生成中的应用，帮助理解图中不同物体的空间关系和前后距离。
16. <https://oreil.ly/BuunX>: (<https://github.com/facebookresearch/segment-anything>) 该资源是 Meta（前 Facebook）开源的“Segment Anything

Model” (SAM) 的 GitHub 项目, 提供了一个强大的图像分割模型。该模型基于包含 1100 万张图像和 11 亿个掩码的大规模数据集进行训练, 能够根据用户输入 (如点击图像中的一到三个点) 自动推断图像中的掩码, 或自动分割图像中的所有元素, 生成的掩码可用于图像修复、ControlNet 控制或作为基础图像素材。

17. <https://oreil.ly/DIVUz>: (<https://huggingface.co/docs/diffusers/using-diffusers/sd-xl>) 该资源是 Hugging Face 官方文档中关于使用 Diffusers 库中 Stable Diffusion XL (SDXL) 模型的指南页面。它详细介绍了如何安装和使用基于 SDXL 的扩散模型进行图像生成, 包括依赖安装、模型下载和调用方法等内容。
18. <https://oreil.ly/DsKyU>: (<https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Install-and-Run-on-Nvidia-GPUs>) 该资源是一个指南页面, 介绍如何在搭载 NVidia 显卡的计算机上安装和运行相关软件 (很可能是稳定扩散模型的 Web UI)。页面内容通常包括环境配置、驱动安装、依赖项设置及运行步骤, 帮助用户在 NVidia GPU 硬件上顺利部署和使用该项目。
19. <https://oreil.ly/Dw2rs>: (<https://github.com/Mikubill/sd-webui-controlnet>) 该资源是一个名为 “sd-webui-controlnet” 的开源项目, 托管在 GitHub 上, 提供了 ControlNet 功能的实现和集成, 方便用户在基于 AUTOMATIC1111 的 Stable Diffusion 图形界面中快速使用 ControlNet 模型, 从而实现对文本到图像扩散模型的条件控制。
20. <https://oreil.ly/E-bjw>: (<https://github.com/llyasviel/ControlNet>) 该资源是 ControlNet 的官方 GitHub 仓库, 提供了 “Adding Conditional Control to Text-to-Image Diffusion Models” 论文中提出的一系列开源模型和相关代码, 实现了在文本到图像扩散模型中添加条件控制的功能, 方便用户基于 Stable Diffusion 等模型进行定制化图像生成。
21. <https://oreil.ly/FNx9>: (<https://huggingface.co/runwayml/stable-diffusion-v1-5>) 该资源是 “Stable Diffusion” 模型的一个版本, 具体为 “version 1.5”, 托管在 Hugging Face 平台上的 RunwayML 账号下。它是一个用于生成图像的深度神经网络, 广泛应用于文本到图像的生成任务。与后续版

本相比，1.5 版本因训练数据较为宽松（如包含 NSFW 内容）而被认为更具灵活性，并拥有大量社区定制的模型和资源支持。

22. <https://oreil.ly/GgnJV>: (https://colab.research.google.com/github/huggingface/notebooks/blob/main/diffusers/sd_textual_inversion_training.ipynb) 该 URL (https://colab.research.google.com/github/huggingface/notebooks/blob/main/diffusers/sd_textual_inversion_training.ipynb) 指向的是一个 Google Colab 笔记本，由 Hugging Face 官方维护，内容涉及使用 Diffusers 库进行 Stable Diffusion 模型的 Textual Inversion 训练。该笔记本演示了如何通过文本嵌入的方式训练自定义概念，而无需更新模型的全部权重，是一种较为轻量的模型定制方法。
23. <https://oreil.ly/I3QGY>: (https://unsplash.com/photos/woman-in-red-and-black-plaid-dress-shirt-and-blue-denim-jeans-sitting-on-bed-GxH1DSxzons?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 该 URL 指向的是一张由 Unsplash 提供的图片，内容是一位身穿红黑格子衬衫和蓝色牛仔裤的女性，坐在床上。根据上下文，这张图片用作示例，展示“Normal model”在图像处理中的效果，具体为通过颜色映射估计物体的三维结构。图片来源于 Unsplash，可能由摄影师 Soundtrap 拍摄。
24. <https://oreil.ly/IqrbI>: (<https://github.com/continue-revolution/sd-webui-segment-everything>) 该资源是一个 GitHub 代码仓库，名为“sd-webui-segment-everything”，用于在 Stable Diffusion Web UI 中集成并使用 Segment Anything Model (SAM)。该仓库提供了相关代码和工具，方便用户下载、管理和应用 SAM 模型权重，实现图像的自动分割功能，从而增强图像生成或编辑的效果。
25. <https://oreil.ly/LWVvC>: (<https://stable-diffusion-art.com/comfyui/>) 该资源是 ComfyUI 的官方网站，ComfyUI 是一个面向 Stable Diffusion 的高级用户界面，支持更复杂的工作流程和更高的灵活性（例如支持图像到视频的转换），适合有进阶需求的用户使用。
26. <https://oreil.ly/Lb3xE>: (https://unsplash.com/@bailuk85?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 该 URL

指向的是摄影师 Bailuk85 在图片分享网站 Unsplash 上的个人主页。该主页展示了 Bailuk85 上传的免费高质量摄影作品，供用户浏览和下载使用。

27. <https://oreil.ly/MiSt1>: (<https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Features>) 该 URL (<https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Features>) 指向的是 AUTOMATIC1111 的 Stable Diffusion Web UI 项目的功能介绍页面，详细列出了该 Web UI 提供的各种强大功能和特性，包括图像尺寸调整、高分辨率修复、面部修复、图像平铺、样式保存与复用，以及支持社区扩展插件等内容，帮助用户更好地使用和扩展 Stable Diffusion 生成图像的能力。
28. <https://oreil.ly/NtoiB>: (<https://arxiv.org/abs/2106.09685>) 该 URL “<https://arxiv.org/abs/2106.09685>” 指向的是论文《LoRA: Low-Rank Adaptation of Large Language Models》(LoRA: 大规模语言模型的低秩适配)，该论文提出了一种通过添加并训练模型中的低秩适配层来高效微调大型语言模型的方法，旨在在不显著增加计算资源消耗的情况下，实现对模型的定制化训练和适配。
29. <https://oreil.ly/NzzGm>: (https://github.com/jamesaphoenix/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_9/DreamBooth_Stable_Diffusion.ipynb) 该资源是一本书的 GitHub 仓库中保存的一个 Python Jupyter 笔记本，文件名为 “DreamBooth_Stable_Diffusion.ipynb”。该笔记本演示了如何使用基于 HuggingFace diffusers 库的 DreamBooth 模型进行训练，采用的是较小的 Stable Diffusion 1.5 版本，适合在 Google Colab 的免费环境中利用 Nvidia GPU 进行快速训练。笔记本内容主要用于学习和复现 DreamBooth 模型的训练流程。
30. <https://oreil.ly/OOpas>: (<https://github.com/AUTOMATIC1111/stable-diffusion-webui>) 该资源是一个官方的 GitHub 代码库，名为 “stable-diffusion-webui”，提供了一个基于网页的用户界面，允许用户通过输入文本提示 (prompt) 生成图像。

-
31. <https://oreil.ly/Ob2VK>: (<https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Installation-on-Apple-Silicon>) 该资源是一个关于在 Apple Silicon (苹果自研芯片, 如 M1、M2) 设备上安装和运行 AUTOMATIC1111 版 Stable Diffusion Web UI 的安装指南页面。
 32. <https://oreil.ly/Oc7ix>: (<https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Install-and-Run-on-AMD-GPUs>) 该资源是一个关于如何在 AMD 显卡上安装和运行相关软件 (可能是 Stable Diffusion WebUI) 的指南页面, 详细介绍了在 AMD GPU 环境下的安装步骤和运行方法。
 33. <https://oreil.ly/OmBuR>: (https://colab.research.google.com/github/huggingface/notebooks/blob/main/diffusers/stable_diffusion.ipynb) 该资源是一个托管在 Google Colab 上的 Jupyter Notebook, 链接地址为 “https://colab.research.google.com/github/huggingface/notebooks/blob/main/diffusers/stable_diffusion.ipynb”。该笔记本演示了如何使用 Hugging Face 的 Diffusers 库运行 Stable Diffusion 图像生成模型, 用户可以借助该笔记本在云端免费 (或付费) 获得 GPU 资源, 快速体验和生成基于 Stable Diffusion 的图像。
 34. https://oreil.ly/OtV_O: (https://unsplash.com/@collovhome?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 该 URL “https://unsplash.com/@collovhome?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash” 指向的是 Unsplash 网站上名为 “Collov Home Design” 的摄影或设计作品集页面。该页面主要展示与建筑和室内设计相关的高质量图片资源, 特别是现代公寓和家居设计风格的摄影作品, 适合用于展示直线条和现代建筑元素的视觉效果。
 35. <https://oreil.ly/Pdzb0>: (<https://github.com/git-for-windows/git/releases/download/v2.39.2.windows.1/Git-2.39.2-64-bit.exe>) 该 URL 指向的是 Git for Windows 的安装程序, 具体版本为 Git 2.39.2 的 64 位 Windows 安装包。用户可以通过该链接下载并安装适用于 Windows 系统的 Git 版本。
 36. <https://oreil.ly/RP1Ei>: (https://unsplash.com/@soundtrap?utm_content=creditCopyText&utm_medium=referral&utm_source=

unsplash) 该 URL (https://unsplash.com/@soundtrap?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 指向的是 Unsplash 网站上名为 “Soundtrap” 的摄影作者主页。该页面展示了 Soundtrap 上传的照片资源, 用户可以浏览其作品集, 通常为高质量的免版权图片, 供免费下载和使用。

37. <https://oreil.ly/RjUur>: (https://en.wikipedia.org/wiki/Girl_with_a_Pearl_Earring) 该资源是维基百科中关于《戴珍珠耳环的少女》(Girl with a Pearl Earring) 这幅著名画作的条目页面, 介绍了这幅由荷兰画家约翰内斯·维米尔创作的经典作品的历史背景、艺术特点及其文化影响。
38. <https://oreil.ly/SG9CT>: (https://unsplash.com/@clarky_523?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 该 URL 链接 (https://unsplash.com/@clarky_523?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 指向的是摄影师 Clarky_523 在 Unsplash 平台上的个人主页。该主页展示了这位摄影师发布的高质量免费可商用照片, 可能包括用于书中示例的现代城堡照片。
39. <https://oreil.ly/X3Ilb>: (<https://beta.dreamstudio.ai/account>) 该资源是 Dream-Studio 的用户账户管理页面, 用户可以在此创建和管理自己的账户, 获取用于调用其图像生成 API 的密钥, 并进行账户充值以购买生成图像所需的积分。
40. <https://oreil.ly/YsuBw>: (<https://cloud.google.com/storage/docs/uploading-objects#storage-upload-object-python>) 该资源是 Google Cloud 官方文档中关于如何使用 Python 上传对象到 Google Cloud Storage 的指南, 介绍了在 Python 环境下将文件或数据上传至云端存储的具体方法和示例代码。
41. <https://oreil.ly/ZH-Ow>: (<https://arxiv.org/abs/2302.05543>) 该资源是论文《Adding Conditional Control to Text-to-Image Diffusion Models》(作者: Zhang, Rao, 和 Agrawala, 2023 年), 介绍了一种在文本到图像扩散模型中加入条件控制的方法。论文提出了一系列名为 ControlNet 的开源模型, 用于增强

图像生成过程中的控制能力，实现更精确的图像内容生成。相关模型和技术被广泛应用于如 Stable Diffusion 等文本生成图像的扩散模型中。

42. <https://oreil.ly/ZqdjB>: (<https://arxiv.org/abs/2208.12242>) 该 URL (<https://arxiv.org/abs/2208.12242>) 指向的是一篇名为 “DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation” 的学术论文。该论文介绍了一种名为 DreamBooth 的技术，能够在已有的文本到图像扩散模型（如 Stable Diffusion）基础上进行微调，以便让模型学习并生成之前未见过的新概念。相比从零训练模型，DreamBooth 显著减少了训练时间和资源消耗，允许用户快速高效地定制和扩展图像生成模型。
43. https://oreil.ly/_b7xX: (<https://arxiv.org/abs/2307.01952>) 该 URL (<https://arxiv.org/abs/2307.01952>) 指向的是一篇关于 OpenAI 和 OpenCLIP 模型在图像生成领域应用的学术论文。论文内容可能涉及 Stable Diffusion 最新版本（如 SDXL）中使用的多模态语言模型架构，特别是结合了 OpenCLIP (ViT-G/14) 和 OpenAI 的 CLIP ViT-L 模型，以提升图像生成的语义理解和多样性表现。研究还可能讨论了模型在不同图像尺寸和非标准长宽比上的训练改进，以及用户对新版本模型的偏好和性能提升。
44. https://oreil.ly/_iyxU: (https://unsplash.com/photos/yellow-car-running-on-the-street-between-the-building-during-daytime-WTPp4wgourk?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 该 URL 指向的是一张来自 Unsplash 的照片，内容是一辆黄色汽车在白天穿过建筑物之间的街道行驶。
45. <https://oreil.ly/aGLeX>: (https://github.com/jamesaphoenix/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_9/sd-api.ipynb) 该资源是一个 Jupyter Notebook 文件 (sd-api.ipynb)，位于 GitHub 仓库中，内容示例展示了如何通过 Dream-Studio 的 API 使用 Stable Diffusion 模型进行图像生成。该示例代码方便开发者在无需自行托管模型的情况下，调用 API 批量生成 AI 图像，适合构建 AI 图像应用或自动化脚本。
46. https://oreil.ly/csYK_: (<https://huggingface.co/lillyasviel/ControlNet-v1-1/tree/main>) 该 URL (<https://huggingface.co/lillyasviel/ControlNet-v1-1/>

tree/main) 指向的是 ControlNet 模型的存储库页面, 提供了 ControlNet 系列预训练模型文件的下载和管理。这些模型用于增强图像生成任务中的控制能力, 如基于边缘检测 (Canny edge detection) 等多种控制方式, 帮助用户在使用 Stable Diffusion 等生成模型时实现更精准的图像控制。页面内包含多个大约几 GB 大小的模型文件, 用户可根据需要下载并放置于本地相应文件夹以供程序调用。

47. <https://oreil.ly/dh7jR>: (https://stable-diffusion-art.com/stable-video-diffusion-img2vid/#Use_Stable_Video_Diffusion_with_ComfyUI) 该资源是一个介绍如何使用 ComfyUI 实现 Stable Video Diffusion (基于 Stable Diffusion 的图像到视频转换) 功能的网页, 内容涵盖通过 ComfyUI 进行更高级和灵活的视频生成工作流程的指导和示范, 适合有一定技术基础、希望利用 Stable Diffusion 进行图像到视频转换的高级用户参考。
48. <https://oreil.ly/gfdY3>: (https://github.com/d8ahazard/sd_dreambooth_extension/wiki/ELI5-Training) 该 URL (https://github.com/d8ahazard/sd_dreambooth_extension/wiki/ELI5-Training) 指向的是一个关于使用 sd_dreambooth_extension 扩展进行 Dreambooth 模型训练的入门指南页面。该页面以 “ELI5” (用简单易懂的方式解释) 形式, 帮助用户理解扩展中大量复杂的参数和设置, 适合希望深入学习和实验 Dreambooth 模型训练的高级用户参考。通过该资源, 用户可以掌握在本地环境使用该扩展训练模型的基础知识和操作方法。
49. https://oreil.ly/hs_fS: (<https://github.com/AUTOMATIC1111/stable-diffusion-webui>) 该资源是 AUTOMATIC1111 维护的 Stable Diffusion WebUI 项目的官方代码库, 提供了一个基于浏览器的图形用户界面, 用于运行和管理 Stable Diffusion 模型, 方便用户生成和调整图像。
50. <https://oreil.ly/hwblq>: (<https://huggingface.co/runwayml/stable-diffusion-v1-5/tree/main>) 该资源是 Hugging Face 上 RunwayML 提供的 Stable Diffusion v1.5 模型文件存储库, 用户可以从下载用于图像生成的预训练模型 (如 v1-5-pruned-emaonly.ckpt 文件), 以便在安装了 AUTOMATIC1111 界面的环境中使用。

-
51. <https://oreil.ly/jvCjz>: (https://unsplash.com/photos/3-women-and-man-standing-on-beach-during-daytime-wwtSYc8p9cI?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 该 URL 指向的是一张由摄影师 Luca Baina 拍摄并发布在 Unsplash 上的照片，内容是一群三位女性和一位男性站在白天的海滩上。该照片被用于图像处理示例中，作为在 Img2Img 扩展的 Mask 生成过程中用户选取提示点 (prompt points) 的示例图片。用户可以通过点击照片中的不同位置来辅助模型进行图像分割，从而生成准确的遮罩 (mask)。
 52. <https://oreil.ly/kGiyi>: (<https://www.python.org/ftp/python/3.10.6/python-3.10.6-amd64.exe>) 该资源是 Python 3.10.6 版本的 Windows 64 位安装程序 (可执行文件)，用于在 Windows 操作系统上安装 Python 3.10.6。
 53. <https://oreil.ly/r-2vm>: (<https://github.com/AUTOMATIC1111/stable-diffusion-webui>) 该资源是指向 “AUTOMATIC1111” 项目的 GitHub 仓库，该项目提供了一个功能丰富的 Stable Diffusion 网络用户界面 (web UI)，由 Stable Diffusion 高级用户开发和维护。该界面集成了多种扩展，方便用户利用 Stable Diffusion 强大的开源社区资源，实现图像生成及相关高级功能，适合大多数用户的使用需求。
 54. <https://oreil.ly/rFMJN>: (<https://github.com/continue-revolution/sd-webui-segment-anything>) 该资源是一个名为 “sd-webui-segment-anything” 的扩展，提供在 AUTOMATIC1111 界面中使用 SAM (Segment Anything Model, 图像分割模型) 的功能。通过安装并运行 AUTOMATIC1111 后，可以按照该扩展的说明集成 SAM，实现图像分割相关的操作。
 55. <https://oreil.ly/s739b>: (<https://twitter.com/EMostaque/status/1563870674111832066?s=20>) 该 URL (<https://twitter.com/EMostaque/status/1563870674111832066?s=20>) 指向的是一条由 Emad Mostaque 发布的推文，内容大概率涉及 Stable Diffusion 模型或 DreamBooth 技术的相关讨论或分享，可能介绍了训练大型生成模型的成本、技术细节或使用 DreamBooth 进行模型微调的实践经验。

-
56. https://oreil.ly/s_trl: (<https://huggingface.co/runwayml/stable-diffusion-inpainting/tree/main>) 该资源是 RunwayML 提供的 Stable Diffusion Inpainting 模型的官方代码仓库，包含用于图像修复 (inpainting) 和图像扩展 (outpainting) 的相关模型文件和示例，支持用户通过掩膜工具对图像特定区域进行编辑和生成，适用于基于稳定扩散技术的图像处理任务。
57. <https://oreil.ly/spsy3>: (<https://huggingface.co/Envvi/Inkpunk-Diffusion>) 该资源是一个基于 DreamBooth 技术训练的 Stable Diffusion 模型，名称为 “Inkpunk Diffusion”，托管在 Hugging Face 平台上（链接：<https://huggingface.co/Envvi/Inkpunk-Diffusion>）。使用时需要在生成提示词中加入特定触发词 “nvinkpunk”，模型的基础为 Stable Diffusion v1.5，建议生成图像尺寸为 512×512 像素。
58. <https://oreil.ly/suOJz>: (<https://arxiv.org/pdf/2302.05543.pdf>) 该资源是指向 ControlNet 论文的 PDF 文件，详细介绍了一种通过边缘检测、姿态估计、深度信息等多种技术来增强图像生成控制能力的方法。论文中提出的预训练模型能够处理输入图像与文本提示，生成风格匹配且结构相似的图像，提升了图像生成的精确度和多样性。
59. <https://oreil.ly/t5T7p>: (<https://huggingface.co/models?other=stable-diffusion>) 该资源为 Hugging Face 平台上与 Stable Diffusion 相关的模型集合页面，用户可以在此浏览、下载和使用多种基于 Stable Diffusion 的文本生成图像模型及其变体，支持包括文本到图像、图像到图像、图像放大和修补等多种功能，方便用户探索和应用开源的图像生成模型。
60. <https://oreil.ly/tlCrf>: (https://unsplash.com/@oneshotespresso?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 该 URL “https://unsplash.com/@oneshotespresso?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash” 指向的是摄影师或内容创作者 “oneshotespresso” 在 Unsplash 平台上的个人主页。Unsplash 是一个提供高质量免版权照片的网站，该主页展示了 “oneshotespresso” 拍摄并分享的照片作品合集，用户可以浏览、下载并用于各种用途。

-
61. <https://oreil.ly/uauNn>: (https://github.com/jamesaphoenix/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_9/stable_diffusion.ipynb) 该资源是本书配套的 GitHub 仓库中的一个 Python 笔记本文件, 文件名为“stable_diffusion.ipynb”, 位于“content/chapter_9”目录下。该笔记本示例演示了如何使用 Stable Diffusion 进行生成式 AI 相关的任务, 建议读者将其上传至 Google Drive 并在 Google Colab 环境中运行, 以避免本地环境配置问题。
 62. <https://oreil.ly/vBOVI>: (<https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Dependencies>) 该资源是 AUTOMATIC1111 的 Stable Diffusion Web UI 项目在 GitHub 上的“依赖项 (Dependencies)”页面, 介绍了运行该项目所需的软件环境和依赖包, 包括必备的 Git、Python 版本及其他相关库和工具的安装说明, 帮助用户正确配置环境以顺利安装和使用 Stable Diffusion 的自动化界面。
 63. <https://oreil.ly/vc1zS>: (<https://arxiv.org/abs/2307.01952>) 该 URL (<https://arxiv.org/abs/2307.01952>) 指向一篇关于“SDXL (Stable Diffusion XL)”模型的学术论文或预印本, 详细介绍了该模型的架构及性能提升。SDXL v1.0 是一个拥有 66 亿参数的扩展版本, 相较于之前的 v1.5 版本 (9.8 亿参数), 在生成图像质量和细节表现上有显著提升。论文中可能阐述了基模型与精炼器 (refiner) 模型的协作机制, 前者负责整体构图, 后者用于细节优化, 从而实现更高质量的图像生成效果。
 64. <https://oreil.ly/wEPLB>: (https://unsplash.com/@robertbye?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 该 URL 指向的是摄影师 Robert Bye 在 Unsplash 网站上的个人主页, 展示了他上传的免费高质量摄影作品。
 65. <https://oreil.ly/wtHRj>: (https://huggingface.co/stabilityai/stable-diffusion-xl-base-1.0/blob/main/sd_xl_base_1.0.safetensors) 该资源是一个名为“sd_xl_base_1.0.safetensors”的基础模型文件, 托管在 Hugging Face 的 Stability AI 账号下, 属于 Stable Diffusion XL 系列的一个版本, 用于图像生成或相关的深度学习任务。

-
66. <https://oreil.ly/x0BMn>: (<https://github.com/AUTOMATIC1111/stable-diffusion-webui>) 该资源是指向 GitHub 上的 AUTOMATIC1111 项目仓库，提供了一个基于 Stable Diffusion 的网页用户界面 (web UI)。该项目方便用户在本地安装和运行 Stable Diffusion 模型，支持通过网页界面进行图像生成和相关操作，通常需要安装 Git、Python 及其它依赖。它广受社区欢迎，支持多版本的 Stable Diffusion 模型及丰富的自定义扩展。
67. <https://oreil.ly/xbt2d>: (https://github.com/d8ahazard/sd_dreambooth_extension) 该资源是一个基于 Automatic1111 平台的 Dreambooth 模型训练扩展工具，托管在 GitHub 上 (https://github.com/d8ahazard/sd_dreambooth_extension)。该扩展实现了 Shivam Shrirao 的方法，允许用户在本地计算机上进行 Dreambooth 模型的训练，支持丰富的高级功能和设置，适合有机器学习背景的用户进行深入实验。相比使用 Google Colab，该扩展的优势在于训练过程可本地持续运行，无需担心超时或进度丢失。
68. <https://oreil.ly/z38do>: (https://unsplash.com/photos/living-room-with-brown-sofa-and-glass-window-4_jQL4JCS98?utm_content=creditCopyText&utm_medium=referral&utm_source=unsplash) 该 URL 指向的资源是一张展示现代公寓室内设计的照片，画面中包含带有棕色沙发和玻璃窗的客厅场景，符合建筑和室内设计中常用的直线检测技术 (M-LSD) 应用示例。这张图片来自 Unsplash，常用于展示现代家居环境的直线结构与设计美学。
69. <https://oreil.ly/z9XC6>: (https://huggingface.co/llyasviel/ControlNet-v1-1/blob/main/control_v11p_sd15_canny.pth) 该资源是 ControlNet 项目中用于 Stable Diffusion 模型的一个预训练权重文件，具体是 “Canny edge detection” (Canny 边缘检测) 控制模型的权重，文件名为 `control_v11p_sd15_canny.pth`。用户可以从该链接下载该模型文件，并将其放置在 ControlNet 的模型文件夹中，用于增强图像生成时对边缘信息的控制，提高生成图像的细节和结构表现能力。
70. <https://oreil.ly/zFH0->: (<https://blog.novelai.net/novelai-improvements-on-stable-diffusion-e10d38db82ac>) 该 URL (<https://blog.novelai.net/novelai-improvements-on-stable-diffusion-e10d38db82ac>) 指向 NovelAI

博客上一篇关于“Stable Diffusion 改进”的文章，内容介绍了 Stable Diffusion 模型的优化方法和技术进展，包括利用 Hypernetworks 等新兴实验性技术来提升生成模型的性能和效率。文章探讨了这些改进相比传统训练方法（如 Dreambooth）的优势与不足，旨在分享 NovelAI 团队在稳定扩散模型上的最新研究成果和实践经验。

71. <https://pytorch.org>: 该资源是 PyTorch 官网，提供开源的机器学习库，支持计算机视觉和自然语言处理等应用。

第 10 章

1. <https://nextjs.org>: 该资源是 Next.js 的官方网站，Next.js 是一个基于 React 的流行前端框架，广泛用于构建高性能的服务器端渲染和静态网站生成的 JavaScript 应用程序。
2. <https://oreil.ly/0IFyI>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_10/article_generation.py) 该资源是一个名为 `article_generation.py` 的 Python 脚本文件，位于 GitHub 仓库 `BrightPool/prompt-engineering-for-generative-ai-examples` 的 `chapter_10` 目录下。根据文件名和所在目录，可以推测该脚本与第 10 章的内容相关，主要功能可能是用于生成文章或自动写作，属于生成式人工智能的示例代码之一。
3. <https://oreil.ly/0d5Hi>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_10/expert_interview_chain.py) 该资源是一个名为 `expert_interview_chain.py` 的 Python 脚本文件，托管在 GitHub 上的 BrightPool 组织的 `prompt-engineering-for-generative-ai-examples` 仓库中，位于 `content/chapter_10` 目录下。根据文件名推测，该脚本可能与“专家

访谈链”（expert interview chain）相关，属于生成式人工智能领域的示例代码，演示如何通过链式调用或流程设计来实现专家访谈场景中的提示工程（prompt engineering）技术。

4. <https://oreil.ly/3UHQs>: (<https://webflow.com/blog/corporate-memphis>) 该资源是一篇由 Webflow 发布的博客文章，介绍了“Corporate Memphis”这种流行于企业对企业（B2B）领域的视觉设计风格。文章详细阐述了该风格的特点，包括鲜艳的色彩、粗犷且不对称的形状，以及有机与几何形态的结合，探讨了其作为一种“昂贵信号”技术的背景和意义，并可能提供如何利用 AI 技术（如 Stable Diffusion）来生成类似风格插图的方法和示例。
5. <https://oreil.ly/400gJ>: (https://python.langchain.com/docs/guides/evaluation/string/embedding_distance) 该资源是 LangChain 官方文档中的一篇指南，介绍如何通过计算文本嵌入（embedding）之间的距离来进行字符串级别的评估。具体内容包括利用向量表示（如 OpenAI 的 text-embedding-ada-002）生成文本嵌入，并通过余弦相似度（cosine similarity）等方法比较生成文本与参考文本的相似度，从而量化文本输出的质量或效果。该指南适用于评估生成式模型输出的准确性和一致性。
6. <https://oreil.ly/GlwJT>: (<https://github.com/AUTOMATIC1111/stable-diffusion-webui>) 该资源是“AUTOMATIC1111”开发的基于 Stable Diffusion 的 Web 用户界面项目，托管在 GitHub 上，提供了一个开源的图形界面，方便用户通过网页直接运行和交互使用 Stable Diffusion 模型进行图像生成。
7. <https://oreil.ly/H3VtJ>: (<https://www.saxifrage.xyz/post/ai-writing-style-prompt-experiment>) 该资源是一篇关于 AI 写作风格提示词优化实验的文章，重点探讨如何通过调整提示词来提升基于 GPT-4 模型的写作质量，从而减少用户后期大量手动编辑的工作量。文章详细介绍了写作过程中提示词设计的重要性及其优化方法。
8. <https://oreil.ly/HNqVX>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_10/gradio_code_example.py) 该资源是一个位于 GitHub 上的 Python 代码文件，

示例代码展示了如何使用 Gradio 库构建界面，以实现自动化流程的演示和操作。具体来说，该文件（`gradio_code_example.py`）可能包含利用 Gradio 创建交互式界面的示例，帮助用户通过简洁的步骤快速搭建和运行生成式 AI 相关的应用。

9. <https://oreil.ly/KyKjS>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_10/custom_summarize_chain.py) 该资源是一个名为 `custom_summarize_chain.py` 的 Python 脚本文件，位于 GitHub 仓库 `BrightPool/prompt-engineering-for-generative-ai-examples` 的第十章内容目录中。该文件实现了与函数 `create_all_summaries` 相关的自定义摘要链逻辑，适合希望深入理解和扩展文本摘要功能的开发者参考学习。
10. <https://oreil.ly/O6RdB>: (https://github.com/hammer-ml/prompt_optimization_blog) 该资源是一个公开的 GitHub 代码仓库，包含用于提示词优化（prompt optimization）的示例代码和参考文本，旨在帮助用户测试和改进提示词以提升生成文本的质量和匹配特定写作风格的效果。用户也可以贡献改进方案，或者尝试通过微调模型来获得更好的结果。
11. <https://oreil.ly/OMMKi>: (<https://platform.openai.com/docs/guides/fine-tuning>) 该资源是 OpenAI 官方文档中关于“微调（fine-tuning）”的指南，介绍了如何通过微调模型以提升其在特定任务或写作风格上的表现。文档中包含微调所需的样本数量建议（如至少 50 个样本）、操作步骤和最佳实践，帮助用户利用已有模型进行定制化训练，从而获得更符合需求的输出结果。
12. https://oreil.ly/XD_jQ: (<https://platform.stability.ai/docs/api-reference#tag/v1generation/operation/textToImage>) 该资源是 Stability AI 官方提供的 API 文档页面，具体介绍了通过 API 使用 Stable Diffusion 模型进行文本生成图像（text-to-image）功能的接口说明和调用方法。用户可以参考该文档了解如何构造请求、传入文本提示词，从而生成对应的图像。
13. <https://oreil.ly/cWpXH>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/blob/main/content/chapter_10/image_generation_chain.py) 该资源是一个名为

“image_generation_chain.py” 的 Python 脚本文件，托管在 GitHub 仓库 “BrightPool/prompt-engineering-for-generative-ai-examples” 的第 10 章节内容目录下。根据文件名推测，该脚本可能实现了一个用于图像生成的链式处理流程，涉及生成式人工智能相关的图像生成技术示例或演示。

14. <https://oreil.ly/chapter10>: (https://github.com/BrightPool/prompt-engineering-for-generative-ai-examples/tree/main/content/chapter_10) 该资源是一个 GitHub 仓库中名为 “chapter_10” 的文件夹，包含用于运行 Gradio 应用的示例代码（如 `gradio_code_example.py`）。该示例演示了如何通过脚本使用 SERPAPI 和 STABILITY 的 API 密钥，展示生成式 AI 相关的实用案例和代码实现。
15. <https://oreil.ly/dspy>: (<https://github.com/stanfordnlp/dspy>) 该资源是一个名为 DSPy 的提示优化框架，旨在通过定义评估指标，自动测试和选择提示中指令与少量示例的最佳组合，从而提升提示工程的效果和质量。
16. <https://oreil.ly/haoTZ>: (<https://every.to/divinations/dall-e-2-and-the-origin-of-vibe-shifts>) 该资源是一篇文章，探讨了由 OpenAI 的图像生成模型 DALL·E 2 引发的视觉风格变迁（vibe shifts）的起源及其影响，特别是对商业插画风格如 “Corporate Memphis” 的演变和应用的影响。文章分析了 AI 生成图像技术如何改变设计趋势及商业品牌的视觉表达方式。
17. <https://oreil.ly/nxEzu>: (<https://tools.saxifrage.xyz/prompt>) 该资源是一个 “视觉提示生成器” 工具，帮助用户通过选择或输入各种艺术风格名称，构建用于指导图像生成模型（如 Stable Diffusion）的提示语，从而生成符合特定风格的图像。
18. <https://oreil.ly/pSrP3>: (<https://www.gradio.app/guides/using-hugging-face-integrations>) 该资源是 Gradio 官方提供的关于如何使用 Hugging Face 集成的指南页面，详细介绍了如何将 Gradio 演示项目上传并托管到 Hugging Face Spaces 平台，帮助开发者快速免费地发布和分享他们的交互式演示。
19. <https://oreil.ly/r7sXi>: (<https://mitchellh.com/writing/prompt-engineering-vs-blind-prompting>) 该资源是一篇关于 “盲目提示”（blind prompting）与提示工程（prompt engineering）区别的文章，探讨在构建提

示语时未经测试直接添加指令可能导致提示质量下降或资源浪费的问题。文章强调通过评估提示效果来优化提示设计的重要性。

20. <https://oreil.ly/saxifrage>: (<https://www.saxifrage.xyz/post/ai-writer>) 该资源是 Saxifrage 博客上的一篇文章，介绍了如何应用五条提示原则，构建一个基于 AI 的端到端内容写作 workflow。该系统通过用户对采访问题的回答，自动生成符合用户写作风格的博客文章。
21. <https://oreil.ly/uMfZa>: (<https://www.contentbot.ai/blog/news/top-10-chatgpt-prompts-for-blog-posts>) 该资源是一篇关于“博客文章写作提示”的内容，可能详细介绍了如何通过精心设计的 ChatGPT 提示语来指导博客文章的写作，包括写作风格、结构安排和关键词使用等方面的建议，旨在帮助用户提升博客内容的质量和效率。
22. <https://oreil.ly/vRRYO>: (<https://www.saxifrage.xyz/post/ai-writing-style-prompt-experiment>) 该资源是一篇关于 AI 写作风格提示词实验的文章，介绍了通过设计不同提示词 (prompts) 来生成文本，并与人工重写的参考文本进行对比，使用 OpenAI 的文本嵌入模型 (如 text-embedding-ada-002) 和余弦相似度方法评估生成文本与参考文本之间的相似度，从而分析和优化 AI 生成内容的风格和质量。
23. <https://oreil.ly/vercel>: (<https://dspy-docs.vercel.app/docs/intro>) 该资源是 DSPy 的官方文档，提供了关于如何使用 DSPy 进行提示词优化的详细介绍和示例，帮助用户定义评估指标并自动测试不同提示词组合，从而提升提示词的质量和效果。
24. <https://oreil.ly/vu0IU>: (<https://twitter.com/8teAPi/status/1752011432034713963?s=20>) 该 URL 指向的 Twitter 帖子，内容可能讨论了 AI 产品开发中关于提示词 (prompting) 策略测试的重要性，强调即使测试过程不够系统化或严谨，也能帮助开发者更有信心地选择合适的写作风格或提示方法，从而推动产品成功。帖子可能以 GitHub Copilot 的开发经验为例，说明实用且灵活的评估方法在 AI 工具设计中的价值。

25. <https://streamlit.io>: 该资源是 Streamlit 的官方网站, Streamlit 是一个开源框架, 旨在帮助开发者快速构建用于展示和交互的用户界面, 特别适合于 AI 原型和数据应用的快速开发与部署。
26. <https://tailwindcss.com>: 该资源是 Tailwind CSS 的官方网站, Tailwind CSS 是一个用于快速构建用户界面设计元素的实用型 CSS 框架, 常与 React 和 NextJS 等前端技术配合使用, 方便进行快速原型设计和样式开发。
27. <https://www.gradio.app>: 该资源是 Gradio 的官方网站, Gradio 是一个开源工具, 方便用户快速构建基于 AI 的交互式网页界面, 广泛用于展示和运行 AI 模型的原型, 支持本地运行和在线分享, 现由 HuggingFace 运营。

关于作者

1. <https://ladder.io>: 该资源是一个名为 Ladder 的增长营销机构的网站, Ladder 由 Mike Taylor 共同创立, 拥有 50 名员工, 并在美国、英国和欧洲设有办事处, 专注于提供增长营销服务。
2. <https://oreil.ly/Nvkdm>: (<https://www.udemy.com/course/prompt-engineering-for-ai/>) 该资源是一门由 James Phoenix 和 Mike 合作开发的 Udemy 在线课程, 主题为“AI 提示工程” (Prompt Engineering for AI), 旨在教授如何设计和优化针对人工智能模型的提示语, 以提升 AI 应用的效果和效率。
3. <https://oreil.ly/QDWh0>: (<https://www.udemy.com/course/prompt-engineering-for-ai/>) 该资源是 Mike Taylor 在 Udemy 平台上开设的关于“AI 提示工程” (Prompt Engineering for AI) 的在线课程, 旨在帮助学员掌握利用提示优化人工智能模型效果的相关技能。
4. <https://oreil.ly/mtaylor>: (<https://www.linkedin.com/learning/search?keywords=Michael+Taylor>) 该资源为 LinkedIn Learning 平台上关于

Michael Taylor 的课程和教学内容的搜索结果页面，展示了由他主讲的市场营销和人工智能相关课程，供用户学习和参考。

5. <https://vexpower.com>: 该资源是一个在线教育平台，提供由 Mike Taylor 和 James Phoenix 等专家讲授的营销和数据科学相关课程。
6. <https://www.brightpool.dev>: 该资源是一个名为 Brightpool 的网站，涉及与提示工程（prompt engineering）相关的项目和服务，可能是一个专注于 AI 提示工程技术应用和开发的平台或团队展示。
7. <https://www.linkedin.com/in/jamesphoenix>: 该资源是 James Phoenix 的 LinkedIn 个人主页，介绍了他作为数据工程师为营销团队构建可靠的数据管道、自动化重复任务的专业背景，以及他在数据科学培训和在线课程开发方面的经验。
8. <https://www.linkedin.com/in/mjt145>: 该资源是 Mike Taylor 的 LinkedIn 个人主页，介绍了他作为一位联合创始人的职业背景及其在增长营销领域的专业经验。