arXiv:2305.04853v2 [cs.CL] 1 Aug 2023

摘要的现状

Fabian Retkowski

卡尔斯鲁厄理工学院 (KIT)
retkowski@kit.edu

2023年8月2日

摘要

随着文本信息的爆炸式增长，摘要系统变得越来越重要。本研究旨在简明地介绍抽象式文本摘要领域当前最先进的水平。作为其中的一部分，我们概述了当前向预训练编码器-解码器模型和大型自回归语言模型转变的范式。此外，我们进一步探讨了摘要系统评估所面临的挑战，以及指令微调模型在零样本摘要中的潜力。最后，我们简要概述了摘要系统目前如何被整合到商业应用中。

1 引言

摘要是从文本中提取最重要信息并以精简形式呈现的过程。随着信息以前所未有的速度大量涌现，组织和个人都面临着独特的挑战，对高效摘要系统的需求也随之增加。对于许多领域的研究人员而言，要跟上各自领域的最新进展颇具挑战，其中也包括人工智能本身；这一点可以从每年的期刊出版数量间接看出，自D. Zhang 等人，2022年以来，该数量几乎增长了两倍。

一般而言，摘要分为两种不同形式：抽取式和抽象式。在抽取式摘要中，系统的任务是从文档中选取段落并将其纳入摘要。另一方面，抽象式摘要旨在以不同的句法重新表述文档中最重要的内容。随着语言模型的能力日益增强，研究重点正逐渐从抽取式摘要转向抽象式摘要；后者被认为更具挑战性，但也更加流畅、多样且易读。

本文介绍了抽象式文本摘要领域的最新进展，重点关注预训练编码器-解码器模型（第 2 节）、大型自回归语言模型（第 3 节）以及经过指令微调的变体（第4 节）。在力求全面的同时，图 1 概述了本文涵盖的模型。第 5 节结合向大语言模型的范式转变，讨论了当前的评估协议。最后，本文讨论了局限性、潜力（第 6 节）以及当前的商业化工作（第 7 节）。

2 预训练编码器-解码器模型

近年来，预训练编码器-解码器模型广受欢迎，如今已在自然语言处理领域得到广泛应用。这些模型在大型未标注语料库上以自监督方式进行训练。著名的例子包括去噪自编码器 BART（Lewis 等人，2020年）和 T5（Raffel 等人，2020年），它们采用填空式预训练目标进行训练。UL2（Tay 等人，2022年）是一个较新的例子，它对多种去噪预训练目标进行了泛化和整合。通过在特定任务数据集上对这些模型进行微调，它们在包括摘要在内的许多任务上都取得了最先进的结果。一些预训练模型专门针对摘要任务设计，其预训练目标与摘要任务相似。例如，在图 2 中，PEGASUS 的架构（J. Zhang 等

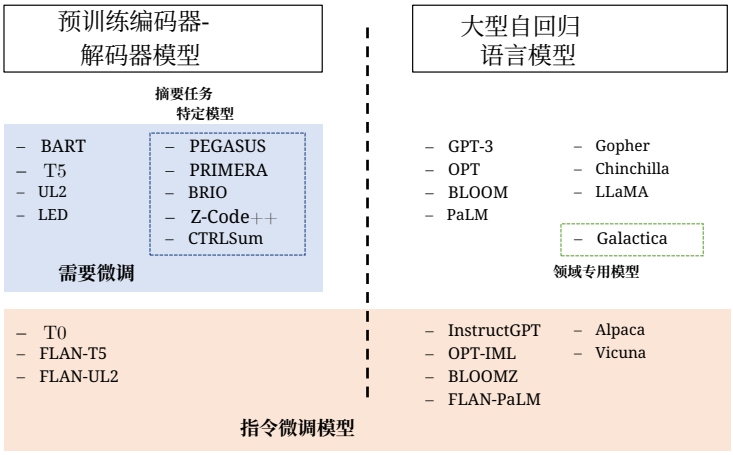


图 1：当前摘要系统大致可分为预训练编码器-解码器模型和大型自回归语言模型。总体而言，在零样本摘要方面，指令微调模型的能力最强。其他编码器-解码器模型通常需要微调，而没有进行指令微调的自回归大语言模型效果较差。一些预训练编码器-解码器模型专门针对摘要任务进行设计。

等，2020年）可以观察到，该模型通过从输入文档中删除重要句子，并要求模型重新生成这些句子来进行训练。在一项针对 23 个摘要任务模型的综合评估中，Fabbri 等人(2021年）得出结论，PEGASUS、BART 和 T5“在大多数维度上的表现始终最佳”，评估内容既包括人工评估，也包括自动指标。近期，一种名为 BRIO（Y. Liu、Liu 等人，2022年）的摘要专用微调机制被提出。该方法引入了对比学习组件，以避免将全部概率分布质量分配给参考摘要，而是同时考虑候选摘要。BRIO 已应用于多个模型，包括 BART 和 PEGASUS。另一个值得关注的模型是 Z-Code++ (P. He 等人，2023年），它在针对特定摘要任务进行微调之前，利用广泛的摘要数据集集合，引入了一个中间的任务自适应微调步骤。研究表明，该方法在低资源环境下尤其有效。

3 大型自回归语言模型

另一项重大的范式转变是大型自回归语言模型（大语言模型，LLMs）的近期兴起。这些仅解码器模型通常具有更多的参数，并采用传统的因果语言建模目标进行训练，即预测序列中的下一个词元。Brown 等人(2020年) 首次证明，这种方法在大规模应用时，能够通过零样本提示完成各种各样的下游任务。无需进行任何梯度更新，只需使用特定于任务的自然语言提示（例如，”Question: 〈问题〉 Answer:”）对模型进行预热，然后通过从模型中采样来生成输出。同样的范式也支持零样本摘要，除了其他方法外，还可以通过在提示末尾添加 ”TL;DR:” 来实现。

这一类别中最受欢迎的模型是 GPT-3（Brown 等人，2020年），参数量达 1750 亿。OPT（S. Zhang 等人，2022年）和 BLOOM（BigScience Workshop，2022年）是旨在复现其成果的两款开源替代模型。Gopher（Rae 等人，2022年）和 PaLM（Chowdhery 等人，2022年）则将这一方法推向极致，把模型规模扩展至多达 5600 亿参数。相比之下，Chinchilla（Hoffmann 等人，2022年）和 LLaMA（Touvron 等人，2023年）更加严格地考量缩放定律和计算预算，因此分别训练出了 700 亿和 650 亿参数的模型，同时仍能媲美或超越更大型的模型。值得一提的还有 Galactica 120B 科学语言模型（Taylor 等人，2022年），它展现了专业化大语言模型的有效性。该模型采用精心设计的数据集，并融入适配领域的分词方法，因此在特定领域的表现优于其他大语言模型。它还通过引入专用任务标记，对引文以及化学式、蛋白质序列等模态进行了特殊处理。

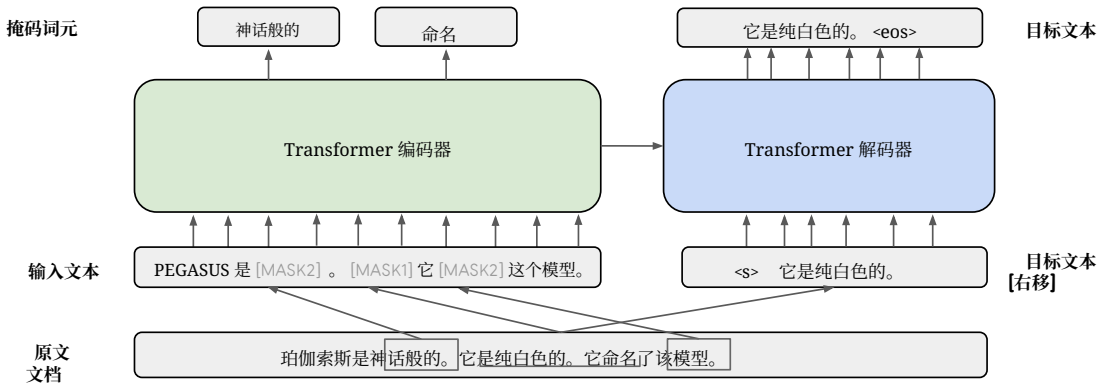


图 2：采用预训练目标的 PEGASUS 架构。该模型结合了掩码语言建模（MLM）和缺句生成（GSG）。在 GSG 中，重要句子会被掩码，并作为解码器的目标。重要性通过一个句子与文档其余部分之间的 ROUGE-1 近似评分。（J. Zhang 等人，2020年）

4 指令微调模型

指令微调是指使用多种数据集对预训练模型进行微调的过程，这些数据集以自然语言任务指令加以描述。这一步使训练过程更贴合模型在推理阶段的实际使用方式，并已被证明能显著提升模型在零样本任务上的表现。它使用户能够直接、可靠地指示模型执行特定任务。例如，现在可以使用“总结这篇文章：〈文章〉”作为摘要任务的提示。更多提示示例见图3。对模型进行指令微调时，最常见的方法是监督微调和基于人类反馈的强化学习（RLHF，Christiano 等人，2017年）。对于预训练编码器-解码器模型，目前已有多种流行的指令微调模型。例如，基于 T5 的 T0（Sanh 等人，2022年）和 FLAN-T5（Chung 等人，2022年）都受到业界从业者的广泛关注。大型自回归语言模型也同样如此，其中大多数都有对应的指令微调版本：InstructGPT（Ouyang 等人，2022年）、OPT-IML（Iyer 等人，2023年）、BLOOMZ（Muennighoff 等人，2023年）和 FLAN-PaLM（Chung 等人，2022年）。Taylor 等人（2022年）通过 Galactica 展示了另一种实现基础指令提示的方法，即提示预训练。这种方法是在预训练阶段加入任务提示，而不是在预训练后微调模型。开源社区近期的一种趋势是，利用现有且能力强大的指令微调大语言模型（如 ChatGPT）生成的对话与指令遵循数据，对大语言模型进行微调。这促成了 Alpaca 和 Vicuna 的开发，两者均基于 LLaMA（Taori 等人，2023年；Vicuna 团队，2023年；Y. Wang 等人，2023年）。大多数自然语言提示数据集都涵盖摘要任务。例如，在 InstructGPT 使用的 API 提示数据集中，4.2% 的指令属于“摘要”应用场景。类似地，T0 使用可用于微调模型的指令模板，对 CNN Daily Mail（Nallapati 等人，2016年）或 SamSum（Gliwa 等人，2019年）等经典摘要数据集进行了扩充。

5 大型语言模型的评估

摘要系统最常见的评估方式是使用自动化指标。ROUGE (Lin, 2004) 在该领域具有悠久的应用历史，尤其用于衡量参考摘要与生成摘要之间的词汇重叠。近年来，BertScore (T.Zhanget al.,2019) 和 BARTScore (Yuan et al., 2021) 等能够更好地捕捉语义等价性的指标，也正逐渐得到广泛认可。然而，随着大语言模型的能力不断增强，并逐渐泛化到各种各样的任务，它们在摘要任务方面接受的专门评估却越来越少，也越来越不充分。相反，它们通常通过侧重于问答的基准测试套件进行评估

CNN 文章： https://www.cnn.com/2022/09/09/politics/judge-throws-out-trumps-rico-lawsuit-against-hillary-clinton-and-democrats/index.html

长度约束提示

用两句话概括上述文章。

GPT-3： 唐纳德·特朗普对希拉里·克林顿、民主党的诉讼。国家委员会以及其他几名个人和实体已被被联邦法官驳回。法官裁定，特朗普的主张是没有任何法律依据支持，而且他的许多具体事实主张断言不合理或缺乏支持。

关键词约束提示

简要概括上述文章，重点关注 ***Alina Habba***。

GPT-3： Alina Habba 是唐纳德·特朗普的律师。她表示，特朗普将“立即对这一裁决提起上诉。”Habba 还表示，法官的意见“充斥着对法律的错误适用”，并且“无视了众多独立的政府调查，这些调查证实了我们的主张，即被告共谋进行虚假牵连我们的客户，并破坏2016年总统选举。”

图3：使用 GPT-3 进行零样本摘要的示例指令。值得注意的是，大语言模型的自然语言指令能够更好地控制任务，例如长度约束摘要。（Goyal 等人，2022年）

以及常识推理，例如 SuperGLUE（A. Wang 等人，2019年）或 MMLU（Hendrycks 等人，2020年），这些基准测试并未明确涉及摘要任务。因此，近年来，多个研究团队独立研究了大语言模型在摘要任务中的能力与局限性（Goyal 等人，2022年；Bhaskar 等人，2023年；Y. Liu 等人，2023年；Qin 等人，2023年；Xiao 等人，2023年；Yang 等人，2023年；T. Zhang 等人，2023年）。根据 Goyal 等人的研究，与经过微调的编码器-解码器模型（T0 和 BRIO）相比，指令微调的 GPT-3 生成的摘要在自动指标上的得分较低。尽管如此，该模型在人工评估中的表现却显著优于这些模型。T. Zhang 等人开展的人工评估表明，这些摘要甚至在质量上超过了参考摘要，并且与专门为此次评估单独收集的高质量摘要不相上下。这些研究对现有评估协议提出了极大的质疑，尤其是在这一范式转变的背景下。其中一些研究指出，自动指标与人类判断之间的相关性较低、参考摘要质量不高、缺乏标注者间一致性，以及摘要风格（在长度、抽象程度和正式程度方面）的差异，都是亟待解决的问题。这与此前研究中提出的问题一致，例如 Fabbri 等人（2021年）指出，摘要评估协议缺乏可比性，无论是对于自动化指标还是人工评估，都是如此。考虑到这些问题，以及摘要系统的表现已可与人类水平相媲美，T. Zhang 等人推测，在评估“单文档新闻摘要”方面已经达到了极限，而 Yang 等人则呼吁“重新思考各种文本摘要任务的未来发展方向”。事实上，“玻璃天花板”现象在自然语言生成领域也得到了更广泛的观察，即使是近期的自动化指标，在提高其与人类判断之间的相关性方面也几乎没有取得进展（Colombo 等人，2022年）。

6 局限性与新前沿

如前所述，当前的评估指标和评估方案存在严重局限，因此，寻找新的标准是未来研究的重要方向。Y. Liu 等人（2023年）例如，建议使用原子事实来减少人工评估中的歧义；而机器翻译领域近期的一项研究表明，大语言模型本身就是最先进的评估器，其与人类判断的相关性高于任何其他自动指标（Kocmi 和 Federmann，2023年）。Kadavath 等人（2022年）的研究也支持这一结论，他们发现大语言模型能够进行自我评估。与此同时，众所周知，大语言模型会产生幻觉（Ji 等人，2023年）；随着摘要的抽象程度不断提高，其事实性也受到质疑。诸如Bhaskar 等人（2023年）和Goyal 等人（2022年）的研究表明，摘要事实性仍是大语言模型尚未解决的问题；而

其他研究则公开探讨了究竟该如何衡量事实性（Kryscinski 等人，2020年；Pagnoni 等人，2021年）。

长文档摘要。尽管取得了指数级进展（见图4），许多当前的摘要系统仍受限于语言模型有限的上下文窗口，无法处理更长的文档，而这些文档尤其适合进行摘要，例如长篇新闻文章、科学论文、播客或书籍。克服这一限制有几种常见策略。一种简单的方法是截断输入文本（Y. Zhao 等人，2020年；A. Wang 等人，2022年）。对于新闻文章等某些文档类型，这可能是一种合理的策略，因为这类文章往往在开头传达最重要的信息。事实上，选取前 k 句（Lead-k）常被用作新闻摘要系统的基线摘要（See 等人，2017年；Zhong 等人，2019年）。类似地，在对科学论文进行摘要时，通常只将摘要、引言和结论（AIC）交给摘要器，因为先前研究发现这些部分最为重要（Sharma 等人，2019年；Cachola 等人，2020年）。另一种方法是在两阶段系统中采用抽取式摘要器或检索模块，例如稠密段落检索器（Karpukhin 等人，2020年），先选出重要片段，再将文本交给生成式摘要器（Y. Liu 和 Lapata，2019年b；Ladhak 等人，2020年；A. Wang 等人，2022年）。此外，还有一些 Transformer 架构不受这些限制，例如 LED（Beltagy 等人，2020年）或 LongT5（Guo 等人，2022年），它们以更高效的注意力模式取代了 $O(n^2)$ 注意力模式。最后，也有研究尝试对文本片段进行摘要，并可能经过多轮迭代，最终生成连贯的摘要（Gidiotis 和 Tsoumakas，2020年；Y. Zhao 等人，2020年；Wu 等人，2021年；Y. Zhang 等人，2022年；Yang 等人，2023年）。

多文档摘要。从一组与特定主题相关的文档中生成摘要的过程称为多文档摘要（MDS）。这与长文档摘要面临的挑战类似，因为涉及多篇文档时，上下文长度受限的问题会更加突出。要有效完成这项任务，理解文档之间的关系也至关重要。MDS 的第一种策略是将所有文档简单拼接成一篇长文本，并使用为单文档摘要设计的技术。然而，这要求模型处理非常长的序列。因此，人们通常采用类似于长文档摘要的两阶段流程（P. J. Liu 等人，2018年；Y. Liu 和 Lapata，2019年a）。最先进的方法还采用分层架构或基于图的方法来捕捉文档间关系（Y. Liu 和 Lapata，2019年a；W. Li 等人，2020年；Pasunuru 等人，2021年）。与此同时，MDS 方法日益倾向于使用 BART、T5 或 PEGASUS 等预训练编码器-解码器模型（Goodwin 等人，2020年；Pasunuru 等人，2021年）。这一类别中近期一个值得关注的模型 PRIMERA，专为 MDS 设计，其构建基础是 PEGASUS（Xiao 等人，2022年）。对于 GSG 目标，PRIMERA 会选择能够代表文档簇的句子。它采用文档拼接方法，并在架构上使用 LED 处理长序列。因此，该模型具有广泛的适用性，不依赖特定数据集。尽管目前尚无科学评估，但 LangChain 等实用工具近期的兴起与流行 [1] 和 LlamaIndex[2] 表明，人们正尝试使用大语言模型处理文档集合。例如，LlamaIndex 支持将文档存储在按树状结构组织的索引中，其中每个节点都代表其子节点的摘要。

可控摘要。可控摘要是一项多面向的研究课题，既涉及摘要的形式或风格（如长度、正式程度或抽象程度），也涉及摘要的内容。摘要可以根据特定方面或实体生成，也可以更广泛地根据任意给定的关键词或查询生成。近年来，人们提出了各种各样的方法。其中最全面的系统之一是 CTRLSum（J. He 等人，2022年），这是一种预训练编码器-解码器，通过同时利用关键词和提示来实现通用的可控性。在评估中，作者展示了该方法在长度控制和实体控制方面的有效性，以及在一些更专业的任务（例如专利目的摘要）中的表现。近期，Goyal 等人（2022年）、Xiao 等人（2023年）和Yang 等人（2023年）开展的研究，初步揭示了 GPT-3 和 ChatGPT 等指令微调大语言模型的潜力。这些系统在多种摘要任务中展现出巨大前景

[1] https://github.com/hwchase17/langchain
[2] https://github.com/jerryjliu/llama_index

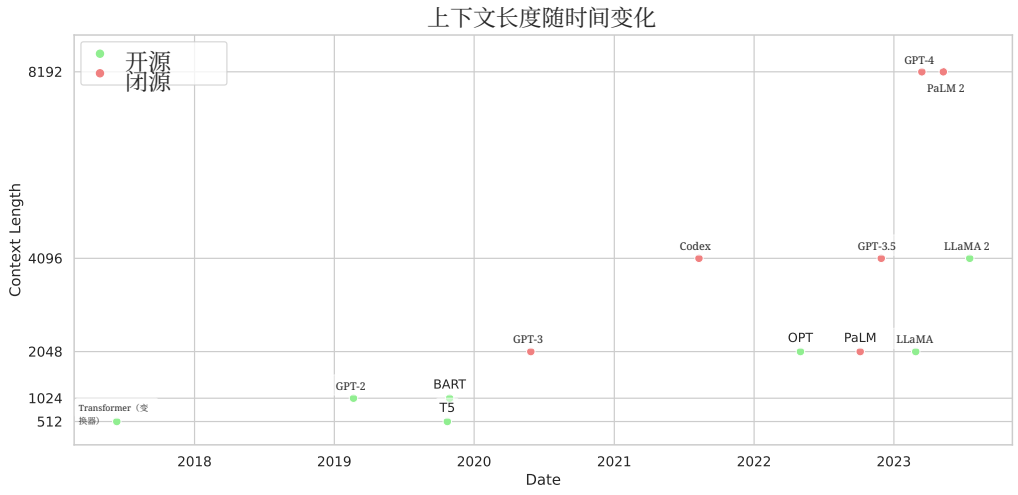


图4：开源语言模型和闭源语言模型的上下文长度都在持续呈指数级增长。此处未考虑 LED 等模型；这类模型专门尝试最大化上下文长度，但会以牺牲其他方面的性能为代价。

，包括基于关键词、方面和查询的摘要。图 3 展示了两个示例，说明零样本提示如何使这类系统实现可控摘要。尽管如此，大语言模型在这项任务中的潜力仍未得到充分探索。Yang 等人（2023年）指出，他们的结果只能作为一个下限，因为模型仅通过简单提示来驱动，未采用提示调优或自我纠正。关于更复杂提示策略潜力的初步展示来自Xiao 等人（2023年），他们建议根据单独训练的模型提供的指令，使用编辑器模型修改生成摘要。与此形成鲜明对比的是，还有大量研究只关注摘要的一种控制维度。例如，仅在长度可控摘要领域，人们提出了多种系统：提前停止解码过程（Kikuchi 等人，2016年）、在将信息交给摘要器之前先进行筛选（LPAS；Saito 等人，2020年），或将长度信息作为输入的一部分（Kikuchi 等人，2016年；Y. Liu 等人，2018年）。最近，Y. Liu、Jia 和 Zhu （2022年）还提出了长度感知注意力机制（LAAM）。

多模态摘要。 迄今为止，大多数研究都集中在文本摘要系统上。然而，播客、电影和会议等媒体与内容十分丰富，它们不仅涉及文本，还涉及图像、视频和音频等其他模态。这些其他模态可能包含纯文本摘要系统会遗漏的关键信息，从而造成语义鸿沟。例如，H. Li 等人 (2017年) 已证明，在多媒体新闻摘要任务中纳入音频和视频信息十分重要；而 M. Li 等人 (2019年) 的研究则表明，在会议摘要中纳入参与者的头部朝向和视线方向很有价值。多模态摘要系统面临的一项关键挑战，是如何融合不同的输入模态。目前，大多数系统采用后期融合方法（参见Jangra 等人， 2023年），例如使用预训练编码器。然而，近年来，研究人员提出了若干前景可观的基于 Transformer 的模型，能够接收多种模态的输入，例如 Perceiver IO（Jaegle 等人， 2021年）和 GATO（Reed 等人， 2022年），但这些模型尚未应用于摘要任务。

7 商业化

随着语言模型的性能超越某一水平，将这些模型创建并集成到产品和工具中已变得日益普遍，从而引发了自然语言处理初创公司的“淘金热”（Butcher，2022年；Toews，2022年）。尤其对于摘要系统而言，模型的上下文长度至关重要。近年来，上下文长度呈指数级增长，如图4所示，已达到足以支持更多任务并具备商业可行性的水平。因此，

过去一年中，许多摘要系统已经产品化，并通过面向消费者的界面提供服务。2022年，谷歌在 Google Docs 中推出了文档摘要功能（Saleh & Kannan，2022年），并在 Google Chat 中推出了对话摘要功能（Saleh & Wang，2022年）；这两项功能均通过微调 PEGASUS 模型实现。然而，数据集中的低质量摘要被认为是一个问题。为了解决这一问题，开发者采用了数据集蒸馏、数据格式化和清理等技术，同时继续收集更多训练数据。通过知识蒸馏，他们将模型蒸馏为更高效的混合架构，由 Transformer 编码器和循环神经网络（RNN）解码器组成。此外，他们还训练了一个额外的模型，用于过滤低质量的生成摘要。最近，微软宣布计划在2023年第二季度于 Microsoft Teams 中推出由 GPT-3.5 驱动的会议摘要功能（Herskowitz，2023年），但尚未提供更多技术细节。社区消息平台 Discord 使用“OpenAI 技术”将消息归类为不同主题，以生成对话摘要（Midha，2023年）。Zoom 最近推出的智能录制功能包含会议摘要和智能章节划分，并含糊地提到使用 GPT-3 来“增强”其自有模型（Parthasarathy，2023年）。Cohere 刚刚推出了专用的文本摘要端点（Hillier & Gall´e，2023年），在很大程度上规避了大语言模型的一些问题，例如需要进行提示工程以及上下文长度有限。此外，该端点还提供了多项设置，以便更好地控制生成摘要：抽取性水平、长度和格式（流畅文本或项目符号）。更广泛地说，访问任何标准大语言模型后，只需指定相应的提示，就可以自然地进行摘要。这适用于 OpenAI 的 GPT-3、AI21 Studio、Antrophic 的 Claude 或 Cohere Generate——这里只列举一些可通过付费 API 使用、并为众多商业应用提供摘要功能的模型。ChatGPT 可能尤其值得注意，因为它还支持以更具交互性的方式进行摘要。特定领域的摘要工具是另一个值得关注的方向。例如，Zoom IQ for Sales（Larkin，2022年）旨在为销售会议提供洞察和摘要；而 BirchAI 是艾伦人工智能研究所的衍生公司，专注于为呼叫中心提供客户通话摘要。与此同时，除了大型科技公司和知名人工智能实验室之外，摘要系统也开始覆盖更多场景，例如浏览器（Opera；Szyndzielorz，2023年）、电子邮件客户端（Shortwave；Wenger，2023年）或笔记应用（Notion；I. Zhao，2023年）。这一趋势表明，摘要并非一种独立的应用，而是一项基础功能，未来可预见的时间内将广泛部署在大多数界面上，并为用户普遍提供访问。

8 结论

文本摘要是一个快速发展的领域，近期经历了两次范式转变。首先，研究方向转向微调预训练编码器-解码器模型；其次，近来又转向对经过指令微调的语言模型进行零样本提示。随着这些进展，单文档摘要似乎已到达一个转折点：继续着力提升自动化指标，收益正日渐减少，甚至可能误导研究界。因此，我们建议将重点转向改进人工评估协议，并探索大语言模型的自我评估。此外，还应针对某些特定方面（如事实性）开展更有针对性的评估，并进一步发掘预训练语言模型和微调摘要模型的能力。然而，若从更广阔的视角审视摘要，多文档摘要和多模态摘要等任务仍面临重大挑战。尽管如此，面向单篇文档的抽象式文本摘要系统已日趋成熟，并正迅速集成到消费产品中。

致谢

本报告所依据的项目由大众汽车基金会资助。

参考文献

Beltagy, I., Peters, M. E., & Cohan, A.（2020年12月）。*Longformer*：长文档 *Transformer*。arXiv。检索于 2023-03-03，来源：http://arxiv.org/abs/2004.05150(arXiv:2004.05150 [cs])

Bhaskar, A., Fabbri, A., & Durrett, G.（2023年7月）。使用 GPT-3.5 进行提示式观点摘要。载于计算语言学协会论文集：ACL 2023 研究成果（第 9282–9300 页）。加拿大多伦多：计算语言学协会。检索于 2023-07-20，来源：https://aclanthology.org/2023.findings-acl.591 BigScience Workshop。（2022年12月）。BLOOM：一个拥有 1760 亿参数的开放获取多语言语言模型。arXiv。检索于 2023-02-28，来源：http://arxiv.org/abs/2211.05100 (arXiv:2211.05100 [cs]) doi: 10.48550/arXiv.2211.05100 Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., . . . Amodei, D.（2020年）。语言模型是少样本学习者。载于 H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin（编），《神经信息处理系统进展》（第 33 卷，第 1877–1901 页）。Curran Associates, Inc.。检索自 https://proceedings.neurips.cc/paper files/paper/ 2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf Butcher, M.（2022年7月）。自然语言处理初创公司即将迎来淘金热，原因何在。检索于 2023-03-08，来源：https://techcrunch.com/2022/07/28/a-gold-rush-of-nlp-startups-is -a bout-to-arrive-heres-why/ Cachola, I., Lo, K., Cohan, A., & Weld, D.（2020年11月）。TLDR：科学文档的极限摘要。载于计算语言学协会论文集：EMNLP 2020 研究成果（第 4766–4777 页）。线上：计算语言学协会。检索于 2023-07-20，来源：https://aclanthology.org/2020.findings-emnlp.428 doi: 10.18653/v1/2020.findings -emnlp.428 Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., . . . Fiedel, N.（2022年10月）。PaLM：借助 Pathways 扩展语言建模。arXiv。检索于 2023-02-28，来源：http://arxiv.org/abs/2204.02311 (arXiv:2204.02311 [cs]) doi: 10.48550/arXiv.2204.02311 Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D.（2017年）。基于人类偏好的深度强化学习。载于 I. Guyon 等（编），《神经信息处理系统进展》（第 30 卷）。Curran Associates, Inc.。检索自 https://proceedings.neurips.cc/paper files/paper/2017/file/ d5e2c0adad503c91f91df240d0cd4e49-Paper.pdf Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., . . . Wei, J.（2022年12月）。扩展指令微调语言模型。arXiv。检索于 2023-03-01，来源：http:// arxiv.org/abs/2210.11416 (arXiv:2210.11416 [cs]) Colombo, P., Peyrard, M., Noiry, N., West, R., & Piantanida, P.（2022年10月）。自然语言生成自动评估的天花板。arXiv。检索于 2023-03-16，来源：http://arxiv.org/abs/2208.14585 (arXiv:2208.14585 [cs] version: 2) Fabbri, A. R., Kry´sci´nski, W., McCann, B., Xiong, C., Socher, R., & Radev, D.（2021年4月）。SummEval：重新评估摘要评估。《计算语言学协会汇刊》，9，391–409。检索于 2023-07-20，来源：https://doi.org/10.1162/ tacl a 00373 doi: 10.1162/tacl a 00373 Gidiotis, A., & Tsoumakas, G.（2020 年）。一种长文档摘要的分而治之方法。《IEEE/ACM 音频、语音与语言处理汇刊》，28，3029–3040。（会议名称：IEEE/ACM 音频、语音与语言处理汇刊）doi: 10.1109/TASLP.2020.3037401 Gliwa, B., Mochol, I., Biesek, M., & Wawer, A.（2019年11月）。SAMSum 语料库：用于抽象式摘要的人类标注对话数据集。载于《第二届摘要新前沿研讨会论文集》（第 70–79 页）。中国香港：计算语言学协会。检索于 2023-03-01，来源：https://aclanthology.org/D19-5409 doi: 10.18653/v1/D19-5409 Goodwin, T., Savery, M., & Demner-Fushman, D.（2020年12月）。PEGASUS 的崛起？比较 Transformer 在少样本和零样本多文档抽象式摘要中的表现。载于《第 28 届国际计算语言学会议论文集》（第 5640–5646 页）。西班牙巴塞罗那（线上）：国际计算语言学委员会。检索于 2023-05-07，来源：https://aclanthology.org/2020.coling-main.494 doi: 10.18653/v1/2020.coling-main.494 Goyal, T., Li, J. J., & Durrett, G.（2022年9月）。GPT-3 时代的新闻摘要与评估。arXiv。检索于 2023-02-17，来源：http://arxiv.org/abs/2209.12356 (arXiv:2209.12356 [cs])

Guo, M., Ainslie, J., Uthus, D., Ontanon, S., Ni, J., Sung, Y.-H., & Yang, Y.（2022年7月）。LongT5：面向长序列的高效文本到文本转换器。载于《计算语言学协会：NAACL 2022年论文集》（第724–736页）。美国西雅图：计算语言学协会。检索日期：2023-07-20，来源：https://aclanthology.org/ 2022.findings-naacl.55 doi：10.18653/v1/2022.findings-naacl.55 He, J., Kryscinski, W., McCann, B., Rajani, N., & Xiong, C.（2022年12月）。CTRLsum：迈向通用可控文本摘要。载于《2022年自然语言处理实证方法会议论文集》（第5879–5915页）。阿拉伯联合酋长国阿布扎比：计算语言学协会。检索日期：2023-07-21，来源：https://aclanthology . org/2022.emnlp-main.396 He, P., Peng, B., Wang, S., Liu, Y., Xu, R., Hassan, H., . . . Huang, X.（2023年7月）。Z-Code++：一种针对抽象式摘要优化的预训练语言模型。载于《第61届计算语言学协会年会论文集》（第1卷：长篇论文）（第5095–5112页）。加拿大多伦多：计算语言学协会。检索日期：2023-07-20，来源：https://aclanthology.org/2023.acl-long.279 Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D., & Steinhardt, J.（2020年10月）。衡量大规模多任务语言理解。载于《第九届国际学习表征会议，ICLR 2021，线上会议，奥地利，2021年5月3日至7日》。检索日期：2023-07-22，来源：https://openreview.net/forum?id=d7KBjmI3GmQ Herskowitz, N.（2023年2月）。Microsoft Teams Premium：降低成本，提升AI驱动的生产力。检索日期：2023-02-28，来源：https://www.microsoft.com/en-us/microsoft-365/blog/ 2023/02/01/microsoft-teams-premium-cut-costs-and-add-ai-powered-productivity/ Hillier, S., & Gall´e, M.（2023年2月）。推出 Cohere Summarize Beta：文本摘要全新端点。检索日期：2023-03-01，来源：https://txt.cohere.ai/summarize-beta/ Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., . . . Sifre, L.（2022年3月）。训练计算量最优的大语言模型。arXiv。检索日期：2023-02-28，来源：http://arxiv.org/abs/2203.15556（arXiv:2203.15556 [cs]）Iyer, S., Lin, X. V., Pasunuru, R., Mihaylov, T., Simig, D., Yu, P., . . . Stoyanov, V.（2023年1月）。OPT-IML：通过泛化视角扩展语言模型指令元学习。arXiv。检索日期：2023-03-01，来源：http://arxiv.org/abs/2212.12017（arXiv:2212.12017[cs]）doi：10.48550/arXiv.2212.12017 Jaegle, A., Borgeaud, S., Alayrac, J.-B., Doersch, C., Ionescu, C., Ding, D., . . . Carreira, J.（2021年10月）。Perceiver IO：用于结构化输入与输出的通用架构。载于《第十届国际学习表征会议，ICLR 2022，线上会议，2022年4月25日至29日》。检索日期：2023-07-22，来源：https://openreview.net/forum?id=fILj7WpI-g Jangra, A., Mukherjee, S., Jatowt, A., Saha, S., & Hasanuzzaman, M.（2023年2月）。多模态摘要综述。arXiv。检索日期：2023-05-01，来源：http://arxiv.org/abs/ 2109.05199（arXiv:2109.05199 [cs]）Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., . . . Fung, P.（2023年12月）。自然语言生成中的幻觉综述。《ACM计算调查》，55（12），1–38。检索日期：2023-03-06，来源：http://arxiv.org/abs/2202.03629（arXiv:2202.03629 [cs]）doi：10.1145/3571730 Kadavath, S., Conerly, T., Askell, A., Henighan, T., Drain, D., Perez, E., . . . Kaplan, J.（2022年11月）。语言模型（大多）知道自己知道什么。arXiv。检索日期：2023-03-02，来源：http://arxiv.org/abs/2207.05221（arXiv:2207.05221 [cs]）doi：10.48550/arXiv.2207 .05221 Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., . . . Yih, W.-t.（2020年11月）。用于开放领域问答的稠密段落检索。载于《2020年自然语言处理实证方法会议（EMNLP）论文集》（第6769–6781页）。线上：计算语言学协会。检索日期：2023-07-21，来源：https://aclanthology.org/2020.emnlp-main.550 doi：10.18653/v1/2020.emnlp-main.550 Kikuchi, Y., Neubig, G., Sasano, R., Takamura, H., & Okumura, M.（2016年11月）。控制神经编码器-解码器的输出长度。载于《2016年自然语言处理实证方法会议论文集》（第1328–1338页）。得克萨斯州奥斯汀：计算语言学协会。检索日期：2023-07-21，来源：https://aclanthology.org/D16-1140 doi：10.18653/v1/D16-1140

Kocmi, T., & Federmann, C.（2023年2月）。大语言模型是最先进的翻译质量评估器。arXiv。检索日期：2023-03-02，来源：http://arxiv.org/abs/2302.14520（arXiv:2302.14520 [cs]）doi：10.48550/arXiv.2302.14520 Kryscinski, W., McCann, B., Xiong, C., & Socher, R.（2020年11月）。评估抽象式文本摘要的事实一致性。载于《2020年自然语言处理实证方法会议（EMNLP）论文集》（第9332–9346页）。线上：计算语言学协会。检索日期：2023-07-20，来源：https://aclanthology.org/ 2020.emnlp-main.750 doi：10.18653/v1/2020.emnlp-main.750 Ladhak, F., Li, B., Al-Onaizan, Y., & McKeown, K.（2020年7月）。探索小说章节摘要中的内容选择。载于《第58届计算语言学协会年会论文集》（第5043–5054页）。线上：计算语言学协会。检索日期：2023-07-20，来源：https://aclanthology.org/2020.acl-main.453 doi：10.18653/v1/2020.acl-main.453 Larkin, T.（2022年4月）。Zoom IQ for Sales：为销售人员提供对话智能。检索日期：2023-07-26，来源：https://blog.zoom.us/zoom-iq-for-sales/ Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., . . . Zettlemoyer, L.（2020年7月）。BART：用于自然语言生成、翻译和理解的去噪序列到序列预训练。载于《第58届计算语言学协会年会论文集》（第7871–7880页）。线上：计算语言学协会。检索日期：2023-07-21，来源：https://aclanthology.org/2020.acl-main.703 doi：10.18653/v1/2020.acl-main.703 Li, H., Zhu, J., Ma, C., Zhang, J., & Zong, C.（2017年9月）。面向异步收集文本、图像、音频和视频的多模态摘要。载于《2017年自然语言处理实证方法会议论文集》（第1092–1102页）。丹麦哥本哈根：计算语言学协会。检索日期：2023-05-06，来源：https:// aclanthology.org/D17-1114 doi：10.18653/v1/D17-1114 Li, M., Zhang, L., Ji, H., & Radke, R. J.（2019年7月）。让会议摘要紧扣主题：抽象式多模态会议摘要。载于《第57届计算语言学协会年会论文集》（第2190–2196页）。意大利佛罗伦萨：计算语言学协会。检索日期：2023-05-01，来源：https://aclanthology.org/P19-1210 doi：10.18653/v1/P19-1210 Li, W., Xiao, X., Liu, J., Wu, H., Wang, H., & Du, J.（2020年7月）。利用图结构改进抽象式多文档摘要。载于《第58届计算语言学协会年会论文集》（第6232–6243页）。线上：计算语言学协会。检索日期：2023-05-06，来源：https://aclanthology.org/2020.acl-main.555 doi：10.18653/v1/2020.acl-main.555 Lin, C.-Y.（2004年7月）。ROUGE：摘要自动评估工具包。载于《文本摘要分支拓展》（第74–81页）。西班牙巴塞罗那：计算语言学协会。检索日期：2023-07-27，来源：https://aclanthology.org/W04-1013 Liu, P. J., Saleh, M., Pot, E., Goodrich, B., Sepassi, R., Kaiser, L., & Shazeer, N.（2018年2月）。通过总结长序列生成维基百科。载于《第六届国际学习表征会议，ICLR 2018，温哥华，不列颠哥伦比亚省，加拿大，2018年4月30日至5月3日，会议论文集》。检索日期：2023-07-22，来源：https://openreview.net/ forum?id=Hyg0vbWC- Liu, Y., Fabbri, A., Liu, P., Zhao, Y., Nan, L., Han, R., . . . Radev, D.（2023年7月）。重审黄金标准：以稳健的人工评估为摘要评估奠定基础。载于《第61届计算语言学协会年会论文集》（第1卷：长篇论文）（第4140–4170页）。加拿大多伦多：计算语言学协会。检索日期：2023-07-20，来源：https://aclanthology.org/2023.acl-long.228 Liu, Y., Jia, Q., & Zhu, K.（2022年5月）。通过预训练信息选择实现抽象式摘要的长度控制。载于《第60届计算语言学协会年会论文集》（第1卷：长篇论文）（第6885–6895页）。爱尔兰都柏林：计算语言学协会。检索日期：2023-03-07，来源：https://aclanthology.org/ 2022.acl-long.474 doi：10.18653/v1/2022.acl-long.474 Liu, Y., & Lapata, M.（2019年a年7月）。用于多文档摘要的层次化Transformer。载于《第57届计算语言学协会年会论文集》（第

5070–5081）。意大利佛罗伦萨：计算语言学协会。检索于2023年-05-06，网址：https://aclanthology.org/P19-1500 doi: 10.18653/v1/P19-1500 Liu, Y., & Lapata, M.（2019年b年11月）。使用预训练编码器进行文本摘要。载于《2019年自然语言处理实证方法会议暨第9届国际自然语言处理联合会议论文集》（EMNLP-IJCNLP）（第3730–3740页）。中国香港：计算语言学协会。检索于2023年-07- 20，网址：https://aclanthology.org/D19-1387 doi: 10.18653/v1/D19-1387 Liu, Y., Liu, P., Radev, D., & Neubig, G.（2022年5月）。BRIO：为抽象式摘要理顺顺序。载于《第60届计算语言学协会年会论文集》（第1卷：长篇论文）（第2890–2903页）。爱尔兰都柏林：计算语言学协会。检索于2023年-07-21，网址：https://aclanthology.org/2022.acl-long.207 doi: 10.18653/v1/2022.acl-long.207 Liu, Y., Luo, Z., & Zhu, K.（2018年10月）。使用卷积神经网络控制抽象式摘要的长度。载于《2018年自然语言处理实证方法会议论文集》（第4110–4119页）。比利时布鲁塞尔：计算语言学协会。检索于2023年-03-21，网址：https://aclanthology.org/D18-1444 doi: 10.18653/v1/D18-1444 Midha, A.（2023年3月）。Discord 是与朋友一起体验 AI 的好去处。检索于2023年-03-09，网址：https://discord.com/blog/ai-on-discord-your-place-for-ai-with-friends Muennighoff, N., Wang, T., Sutawika, L., Roberts, A., Biderman, S., Le Scao, T., . . . Raffel, C.（2023年7月）。通过多任务微调实现跨语言泛化。载于《第61届计算语言学协会年会论文集》（第1卷：长篇论文）（第15991–16111页）。加拿大多伦多：计算语言学协会。检索于2023年-07-21，网址：https://aclanthology.org/2023.acl-long.891 Nallapati, R., Zhou, B., santos, C. N. d., Gulcehre, C., & Xiang, B.（2016年8月）。使用序列到序列 RNN 及其他方法进行抽象式文本摘要。arXiv。检索于2023年-03-01，网址：http://arxiv.org/abs/1602.06023 (arXiv:1602.06023 [cs] 版本：5) Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., . . . Lowe, R.（2022年）。训练语言模型以遵循人类反馈中的指令。载于 S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh（编），《神经信息处理系统进展》（第35卷，第27730–27744页）。Curran Associates, Inc. 检索网址：https://proceedings.neurips.cc/paper files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf Pagnoni, A., Balachandran, V., & Tsvetkov, Y.（2021年6月）。借助 FRANK 基准理解抽象式摘要中的事实性：事实性指标基准。载于《2021年计算语言学协会北美分会会议：人类语言技术论文集》（第4812–4829页）。线上：计算语言学协会。检索于2023年-07-20，网址：https://aclanthology.org/2021.naacl-main.383 doi: 10.18653/v1/2021.naacl-main.383 Parthasarathy, V.（2023年2月）。Zoom 的 AI 创新赋能人们。检索于2023年-03-16，网址：https://blog.zoom.us/ai-driven-innovations/ Pasunuru, R., Liu, M., Bansal, M., Ravi, S., & Dreyer, M.（2021年6月）。高效总结多文档簇的文本与图编码。载于《2021年计算语言学协会北美分会会议：人类语言技术论文集》（第4768–4779页）。线上：计算语言学协会。检索于2023年-05-06，网址：https://aclanthology.org/2021.naacl-main.380 doi: 10.18653/v1/2021 .naacl-main.380 Qin, C., Zhang, A., Zhang, Z., Chen, J., Yasunaga, M., & Yang, D.（2023年2月）。ChatGPT 是通用自然语言处理任务求解器吗？arXiv。检索于2023年-03-20，网址：http://arxiv.org/abs/2302.06476 (arXiv:2302.06476 [cs]) Rae, J. W., Borgeaud, S., Cai, T., Millican, K., Hoffmann, J., Song, F., . . . Irving, G.（2022年1月）。扩展语言模型：Gopher 训练方法、分析与见解。arXiv。检索于2023年-02-28，网址：http://arxiv.org/abs/2112.11446 (arXiv:2112.11446 [cs]) Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., . . . Liu, P. J.（2020年）。探索统一文本到文本转换器迁移学习的极限。《机器学习研究期刊》，21（140），1–67。检索于2023年-07-20，网址：http://jmlr.org/papers/v21/

20-074.html
Reed, S., Zolna, K., Parisotto, E., Colmenarejo, S. G., Novikov, A., Barth-maron, G., . . . Freitas, N. d.（2022年8月）。通才智能体。《机器学习研究汇刊》。检索于2023年-07-22，网址：https://openreview.net/forum?id=1ikK0kHjvj Saito, I., Nishida, K., Nishida, K., Otsuka, A., Asano, H., Tomita, J., . . . Matsumoto, Y.（2020年1月）。通过摘要原型引导实现长度可控的抽象式摘要。arXiv。检索于2023年-03-21，网址：http://arxiv.org/abs/2001.07331 (arXiv:2001.07331 [cs]) doi: 10.48550/arXiv.2001.07331 Saleh, M., & Kannan, A.（2022年3月）。Google Docs 中的自动生成摘要。检索于2023年-02-28，网址：https://ai.googleblog.com/2022/03/auto-generated-summaries -in-google-docs.html Saleh, M., & Wang, Y.（2022年11月）。Google Chat 中的对话摘要。检索于2023年-02- 28，网址：https://ai.googleblog.com/2022/11/conversation-summaries-in-google-chat .html Sanh, V., Webson, A., Raffel, C., Bach, S. H., Sutawika, L., Alyafeai, Z., . . . Rush, A. M.（2022年3月）。多任务提示训练实现零样本任务泛化。arXiv。检索于2023年-03-01，网址：http://arxiv.org/abs/2110.08207 (arXiv:2110.08207 [cs]) See, A., Liu, P. J., & Manning, C. D.（2017年7月）。直奔主题：使用指针生成网络进行摘要。载于《第55届计算语言学协会年会论文集》（第1卷：长篇论文）（第1073–1083页）。加拿大温哥华：计算语言学协会。检索于2023年-07-20，网址：https://aclanthology.org/P17-1099 doi: 10.18653/v1/P17-1099 Sharma, E., Li, C., & Wang, L.（2019年7月）。BIGPATENT：用于抽象式且连贯摘要的大规模数据集。载于《第57届计算语言学协会年会论文集》（第2204–2213页）。意大利佛罗伦萨：计算语言学协会。检索于2023年-03-19，网址：https://aclanthology.org/P19-1212 doi: 10.18653/ v1/P19-1212 Szyndzielorz, J.（2023年2月）。Opera 通过浏览器和内容应用中的新功能进军生成式 AI 领域。检索于2023年-03-01，网址：https://blogs.opera.com/news/2023/02/opera -aigc-integration/ Taori, R., Gulrajani, I., Zhang, T., Dubois, Y., Guestrin, C., Liang, P., & Hashimoto, T. B.（2023年3月）。Alpaca：强大且可复现的指令跟随模型。检索于2023年-05-07，网址：https://crfm.stanford.edu/2023/03/13/alpaca.html Tay, Y., Dehghani, M., Tran, V. Q., Garcia, X., Wei, J., Wang, X., . . . Metzler, D.（2022年9月）。UL2：统一语言学习范式。载于《第十一届国际学习表征会议》（ICLR 2023），卢旺达基加利，2023年5月1–5日。检索于2023年-07-22，网址：https://openreview.net/forum?id=6ruVLB727MC Taylor, R., Kardas, M., Cucurull, G., Scialom, T., Hartshorn, A., Saravia, E., . . . Stojnic, R.（2022年11月）。Galactica：面向科学的大型语言模型。arXiv。检索于2023年-02-28，网址：http://arxiv.org/abs/2211.09085 (arXiv:2211.09085 [cs, stat]) doi: 10.48550/arXiv.2211 .09085 Vicuna 团队。（2023年3月）。Vicuna：一款开源聊天机器人，以90%*的 ChatGPT 水平媲美 GPT-4 | LMSYS Org。检索于2023年-05-07，网址：https://lmsys.org/blog/2023 -03-30-vicuna Toews, R.（2022年5月）。一波价值十亿美元的语言 AI 初创公司即将到来。检索于2023年-03-08，网址：https://www.forbes.com/sites/robtoews/2022/03/27/a-wave-of -billion-dollar-language-ai-startups-is-coming/（章节：AI）Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., . . . Lample, G.（2023年）。LLaMA：开放且高效的基础语言模型。Wang, A., Pang, R. Y., Chen, A., Phang, J., & Bowman, S. R.（2022年12月）。SQuALITY：艰难构建长文档摘要数据集。载于《2022年自然语言处理实证方法会议论文集》（第1139–1156页）。阿布扎比，阿拉伯联合酋长国：计算语言学协会。检索于2023年-07-20，网址：https://aclanthology.org/2022.emnlp-main.75 Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., . . . Bowman, S.（2019年）。SuperGLUE：一个更具挑战性的通用语言理解系统基准——

tem 等。载于 H. Wallach、H. Larochelle、A. Beygelzimer、F. d. Alch´e-Buc、E. Fox 和 R. Garnett（编），神经信息处理系统进展（第 32 卷）。Curran Associates, Inc. 检索自 https://proceedings.neurips.cc/paper files/paper/2019/file/4496bf24afe7fab6f046bf4923da8de6-Paper.pdf Wang, Y., Kordi, Y., Mishra, S., Liu, A., Smith, N. A., Khashabi, D., & Hajishirzi, H.（2023 年 7 月）。自我指令：使用自生成指令对齐语言模型。载于计算语言学协会第 61 届年会论文集（第 1 卷：长篇论文）（第 13484–13508 页）。加拿大多伦多：计算语言学协会。检索于 2023-07-20，来自 https://aclanthology.org/2023.acl-long.754 Wenger, J.（2023）。AI 邮件摘要：几秒钟内读完邮件。检索于 2023-03-01，来自 https://www.shortwave.com/blog/ai-email-summaries/ Wu, J., Ouyang, L., Ziegler, D. M., Stiennon, N., Lowe, R., Leike, J., & Christiano, P.（2021 年 9 月）。基于人类反馈递归摘要书籍。arXiv。检索于 2023-03-07，来自 http://arxiv.org/abs/2109.10862 (arXiv:2109.10862 [cs]) doi: 10.48550/arXiv .2109.10862 Xiao, W., Beltagy, I., Carenini, G., & Cohan, A.（2022 年 5 月）。PRIMERA：用于多文档摘要的基于金字塔的掩码句子预训练。载于计算语言学协会第 60 届年会论文集（第 1 卷：长篇论文）（第 5245–5263 页）。爱尔兰都柏林：计算语言学协会。检索于 2023-07-20，来自 https://aclanthology.org/2022.acl-long.360 doi: 10.18653/v1/2022.acl-long.360 Xiao, W., Xie, Y., Carenini, G., & He, P.（2023 年 5 月）。用于定制抽象式摘要的 ChatGPT 引导的编辑指导工具。arXiv。检索于 2023-05-08，来自 http://arxiv .org/abs/2305.02483 (arXiv:2305.02483 [cs]) Yang, X., Li, Y., Zhang, X., Chen, H., & Cheng, W.（2023 年 2 月）。探索 ChatGPT 在查询式或基于方面的文本摘要中的能力极限。arXiv。检索于 2023-02-17，来自 http://arxiv.org/abs/2302.08081 (arXiv:2302.08081 [cs]) doi: 10.48550/arXiv.2302.08081 Yuan, W., Neubig, G., & Liu, P.（2021）。BARTScore：将生成文本作为文本生成进行评估。载于 M. Ranzato、A. Beygelzimer、Y. Dauphin、P. S. Liang 和 J. W. Vaughan（编），神经信息处理系统进展（第 34 卷，第 27263–27277 页）。Curran Associates, Inc. 检索自 https://proceedings.neurips.cc/paper files/paper/2021/file/e4d2b6e6fdeca3e60e0f1a62fee3d9dd-Paper.pdf Zhang, D., Maslej, N., Brynjolfsson, E., Etchemendy, J., Lyons, T., Manyika, J., . . . Perrault, R.（2022 年 5 月）。《2022 年人工智能指数年度报告》。arXiv。检索于 2023-02-28，来自 http:// arxiv.org/abs/2205.03468 (arXiv:2205.03468 [cs]) doi: 10.48550/arXiv.2205.03468 Zhang, J., Zhao, Y., Saleh, M., & Liu, P.（2020 年 11 月）。PEGASUS：通过抽取间隔句进行抽象式摘要预训练。载于第 37 届国际机器学习大会论文集（第 11328–11339 页）。PMLR。检索于 2023-07-20，来自 https://proceedings.mlr.press/v119/zhang20ae.html (ISSN: 2640-3498) Zhang, S., Roller, S., Goyal, N., Artetxe, M., Chen, M., Chen, S., . . . Zettlemoyer, L.（2022 年 6 月）。OPT：开放式预训练变换器语言模型。arXiv。检索于 2023-02-28，来自 http://arxiv.org/abs/2205.01068 (arXiv:2205.01068 [cs]) doi: 10.48550/arXiv.2205.01068 Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y.（2019 年 9 月）。BERTScore：使用 BERT 评估文本生成。检索于 2023-07-27，来自 https://openreview .net/forum?id=SkeHuCVFDr Zhang, T., Ladhak, F., Durmus, E., Liang, P., McKeown, K., & Hashimoto, T. B.（2023 年 1 月）。用于新闻摘要的大语言模型基准测试。arXiv。检索于 2023-02-17，来自 http://arxiv.org/abs/2301.13848 (arXiv:2301.13848 [cs] version: 1) Zhang, Y., Ni, A., Mao, Z., Wu, C. H., Zhu, C., Deb, B., . . . Zhang, R.（2022 年 5 月）。Summˆ N：用于长输入对话和文档的多阶段摘要框架。载于计算语言学协会第 60 届年会论文集（第 1 卷：长篇论文）（第 1592–1604 页）。爱尔兰都柏林：计算语言学协会。检索于 2023-07-20，来自 https://aclanthology.org/2022.acl-long.112 doi: 10.18653/v1/2022.acl -long.112 Zhao, I.（2023 年 2 月）。Notion AI 现已面向所有人开放。检索于 2023-03-01，来自 https:// www.notion.so/blog/notion-ai-is-here-for-everyone Zhao, Y., Saleh, M., & Liu, P. J.（2020 年 6 月）。SEAL：分段式抽取与生成长文本摘要。

文本摘要。 arXiv。检索于2023年3月20日，来自 http://arxiv.org/abs/2006.10213(arXiv:2006.10213 [cs]) Zhong, M., Liu, P., Wang, D., Qiu, X., & Huang, X.（2019年7月）。探索有效的神经抽取式摘要：哪些方法有效，未来的发展方向是什么？载于 第*57*届计算语言学协会年会论文集 （第1049–1058页）。意大利佛罗伦萨：计算语言学协会。检索于2023年7月20日，来自 https://aclanthology.org/P19-1100doi: 10.18653/v1/P19-1100