HalluCitationMatters：通过 ACL 会议中的 300 篇幻觉论文揭示幻觉参考文献的影响

Yusuke Sakai、Hidetaka Kamigaito、Taro Watanabe奈良先端科学技术大学院大学 (NAIST)，日本
{sakai.yusuke.sr9, kamigaito.h, taro}@is.naist.jp

arXiv:2601.18724v1 [cs.CL] 26 Jan 2026

摘要

最近，我们经常在评审中的论文、预印本或已发表论文中发现幻觉引用或参考文献，它们并不对应任何现有作品。这类幻觉引用严重威胁科学研究的可靠性。当它们出现在已接收论文中时，也可能损害会议的公信力。本研究将幻觉引用称为“幻觉引文”，并系统地考察其出现情况及影响。我们分析了 2024年和 2025 年发表于 ACL、NAACL 和 EMNLP 的所有论文，包括主要会议、Findings 和研讨会论文。分析显示，近 300 篇论文至少包含一条幻觉引文，其中大多数发表于 2025 年。值得注意的是，半数此类论文出现在最新举办的 EMNLP 2025，这表明这一问题正在迅速加剧。此外，EMNLP 2025 有 100 多篇此类论文被接收为主要会议论文和 Findings 论文，影响了会议的公信力。

1 引言

得益于科学人工智能研究（Sun 等人，2024年；Chen 等人，2025d；Shi 等人，2025）和 AI Scientist 技术（Wei 等人，2025b；Huang 等人，2025；Yamada 等人，2025；Tang 等人，2025a；Shao 等人，2025；Xu 等人，2025b）的进步，学术研究与写作已从传统上以人类为中心的流程，转变为由人工智能系统（如大语言模型（LLM））提供支持并积极加以利用的流程。尤其是，大语言模型正在推动各类工具的发展，包括写作辅助（Hwang 等人，2025）、文献检索（Asai 等人，2024）、引文推荐（Sahinuç 等人，2024a；Çelik 和 Tekir，2025）、创意讨论（Zheng 等人，2025d）、同行评审支持与分析（Jin 等人，2024；Starace 等人，2025；Hossain 等人，2025），以及其他科学自动化工具（Sahinuç 等人，2024b）。这些优势对非母语者尤为显著。

✗ **ACL 2025 主要：** Chang 等人(2025)—
Y.Zhangand Others。2024年。用于提升大型语言模型技能。arXiv 预印本 arXiv:2402.12345.
✓ Hohloch (2024年)

✗ **EMNLP 2025 Findings：** Jalori 等人(2025)—
Wendi Zhou、Xiao Li、Lin Geng Foo、Yitan Wang, Harold Soh、Caiming Xiong 和 Yoonkey Kim。2024年。TEMPO：时间序列预测中面向大语言模型的时间表示提示。arXiv 预印本 arXiv:2405.18384。 预期的 for NeurIPS 2024。预印本，arXiv:2405.18384.
✓ Shandi 等人(2024年)

✗ **EMNLP 2025 主会：** Srivastava(2025) —
Wei Xu、Yulia Tsvetkov 和 Alan Black。2022年。用于语言学习的人工智能：对话代理和个性化反馈。As- Transactions of the 计算语言学协会（TACL），10:1–15. ✗（不存在）
✗ 标题链接：Canine：预训练高效的分词器自由编码器，用于语言表示（Clark 等人，2022年）

图1：错误参考文献信息示例。有些参考文献包含错误的 arXiv 编号，或带有“预计将发表”等不确定表述。即使看似合理的参考文献，也可能包含幻觉引文，例如链接信息错误，或指向被引论文并不存在的页面，如最后一个案例所示。

英语使用者，因为人工智能工具可以提升翻译质量和语法质量，并有助于为参与国际学术界创造更公平的机会（Lepp 和 Smith，2025）。

然而，尽管这些进展提升了研究生产力，研究界已开始面临由论文投稿量呈指数级增长带来的新挑战，尤其是难以确保有足够的合格审稿人。因此，资格门槛有所降低[1]，，审稿人也常常需要在有限时间内评估许多不熟悉主题的论文。此外，近期的评审流程还会对延迟评审施加潜在处罚

[1] https://aclrollingreview.org/incentives2025

，例如桌拒审稿人自己的投稿[1]。在这种压力和顾虑之下，同行评审的重点已从开展全面、严格的评估，转向在严格的时间限制内完成评审。因此，我们越来越难以仔细检查已提交稿件的严谨性和事实准确性。尤其是，如图1所示，我们最近发现了一些并不存在的引文，即与任何真实既有研究都不对应的幻觉参考文献。这类虚假引文严重威胁科学可靠性，也损害了人们对既有研究的尊重。当它们出现在已发表论文中时，还可能对学术会议和期刊的整体质量与可信度产生负面影响。

在本研究中，我们将此类幻觉参考文献称为**“幻觉引文”**（幻觉 + 引文，/həˌluː.saɪˈteɪ.ʃən/）。至少包含一条幻觉引文的论文称为***“HalluCited”***论文（幻觉 + 引用，/ˈhæl.uːˌsaɪ.tɪd/）。我们系统地调查了它们在顶级自然语言处理会议中的普遍程度及影响。我们分析了2024年和2025年发表于 ACL、NAACL 和 EMNLP 的逾 17,000 篇论文，包括主会论文、Findings 论文和研讨会论文。我们的分析与讨论揭示了以下发现：

• 近 300 篇论文至少包含一条幻觉引文，其中大多发表于 2025 年。值得注意的是，这些论文中有一半是在最近举办的会议 EMNLP 2025 上发现的，表明这一问题正在迅速加剧。此外，EMNLP 2025 有 100 多篇此类论文被接收为主会与 Findings 论文集论文，影响了学术可信度。

• 我们提出了一种基于光学字符识别和数据库匹配的幻觉引文检测方法，并证明该方法能有效支持我们识别实际幻觉引文的分析。具体而言，检测到四个或更多候选项的论文中，幻觉引文的发生率较高，这为将该方法集成到自动检测工具包中提供了切实可行的参考。

• 即使存在幻觉引文，也不能立即将其归因于人工智能生成内容，因为它们也可能源自二手来源，这使其成因更加复杂。因此，仅凭幻觉引文的存在，不应立即惩罚作者。相反，应事先就预防措施达成共识，例如

作为投稿前检查的作者工具包。

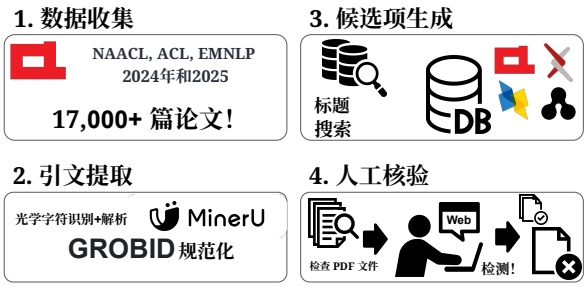

图 2：我们的分析方法。首先，我们从收集的 PDF 文件中提取引文。接下来，我们使用模糊匹配，根据引文标题在参考论文数据库中查找对应论文，并将找不到匹配论文的引文标记为幻觉引文候选项。最后，我们通过查阅原始 PDF 文件并进行网络搜索，人工核验每个候选项是否存在。如果无法客观确认其存在，我们就将其标记为幻觉引文。

Year	会议	主要发现	S.I.D.		WS	总计
2024年北美计		562	296	94	609	1,561
2024	ACL	941	975	84	721	2,721
2024年自然语		,268	1,003	173	510	2,954
2025年北美计		718	474	177	560	1,929
2025	ACL	1,699	1,387	259		4,510
2025年自 方法会议		,809	1,405	269	684	4,167
总体		6,997	5,540			

表 1：PDF 文件数量。S.I.D. 表示学生研究研讨会、工业赛道和演示赛道的论文总数。请注意，ACL 2024 未举办工业赛道，EMNLP 未举办学生研究研讨会。会议按日期先后顺序排列。

2 分析方法

这项调查类似于大海捞针，因为需要检查数量庞大的引文，才能发现少量幻觉引文。因此，手动核验所有引文并不切实际。为应对这一挑战，如图2所示，我们利用“大多数引文都是正确的”这一假设，系统地列出潜在的幻觉引文候选项。然后，我们手动核验每个候选项，以识别幻觉引文。需要注意的是，由于幻觉引文的实际数量尚不清楚，我们报告的结果应被理解为下界，实际可能还存在更多未被发现的幻觉引文。

2.1 数据收集

本研究以2024年和2025年举行的六场会议中的所有论文为研究对象，包括 NAACL、ACL 和 EMNLP，涵盖主要会议——

Year	会议	Mean	标准差 第一四分位数 第二四分位数 第三四分位数 总引用次数
2024年			56,717
2024	ACL	41.26	112,262
2024年			126,744
2025年			75,870
2025	ACL	42.63	192,245
2025年			177,818
总体		41.57	741,656

表2：提取的引文数量。我们报告均值、标准差、四分位数（第一四分位数、Q2=中位数、第三四分位数），以及各会议的引文总数。

论文、Findings 论文以及所有研讨会论文。我们收集了 ACL Anthology 中注册为活动的归档论文的所有 PDF文件和元数据[2]，，不包括教程和会议论文集。表 1 展示了所收集 PDF 文件的统计数据。我们的数据集总计涵盖了超过 17,000 篇论文。

2.2 引文提取

接下来，我们从 PDF 文件中提取引文信息。通过初步研究，我们发现，将基于OCR的提取用于获取原始书目参考文献字符串，并对相应书目参考文献进行规范化，这种流程能够实现最高效的引文提取。因此，首先，我们使用 MinerU (Wang 等人，2024a) 在文本块级别从参考文献部分提取参考文献条目。然后，我们使用 GROBID (GRO，2008–2025) 解析并规范化每条参考文献条目，从而获得标题等结构化书目信息[3]。表2 展示了提取到的引文计数统计结果。从这些结果中可以看出，每篇论文的引文数量中位数为39，且各会议的四分位数保持稳定。总计，我们提取了超过740,000条引文。

2.3 候选项创建

如表 2所示，提取出的引文总数极其庞大，手动检查所有引文并不现实。因此，我们重点关注明确表示引用 ACL Anthology 或 arXiv 论文的引文，并

[2]https://aclanthology.org/

[3]尽管 GROBID 可以直接从 PDF 文件中提取书目信息，但我们的初步研究表明，它对 ACL风格参考文献格式的引文提取不够稳健。具体而言，当参考文献列表跨越多页时，GROBID 经常无法提取后续页面上的参考文献，导致覆盖率不足。例如，有些情况下，实际包含50条参考文献的论文中仅提取出约10条引文。为缓解这一问题，我们明确将参考文献块作为输入提供给 GROBID。

Year	会议	#论文	#引文	平均值 最大值
2024年北美计		211 (13.52%)	268 (0.47%) 1	6
2024	ACL	381 (14.00%)	476 (0.42%) 1	4
2024年自然语		443 (15.00%)	556 (0.44%) 1	7
2025年北美计		336（17.42%）	465（0.61%）	10
2025	ACL	796（17.65%）		17
2025年		783（18.79%）		31
理经验方法会议				
2024		1,035（14.30%）		7
2025		1,915 (18.06%)		31
总体		2,950（16.53%）		31

表 3：候选论文和引文的数量。我们还报告了候选论文和引文占总数的比例。平均值表示每篇论文的候选引文平均数量，而最大值表示单篇论文中观察到的候选引文最大数量。

通过与参考数据库进行匹配，提取出一部分候选引文。

首先，我们通过启发式方法检查原始引文字符串是否包含与 ACL 会议或 arXiv 相关的关键词（如 ACL、EMNLP 或 arXiv），以缩小目标引文范围。接下来，我们主要利用 ACL Anthology 的元数据搜索相应论文[4] 和 arXiv[5]。此外，由于这种启发式筛选仅作为粗略筛查，ACL 发表场所或 arXiv 之外的论文引文仍可能残留。为了进一步筛除这类情况，我们还利用 DBLP[6] 和 OpenAlex（Priem 等人，2022年）[7]。

进行数据库匹配时，我们使用基于归一化莱文斯坦距离的相似度分数，通过 `RapidFuzz` 库对引文标题进行字符级模糊匹配（Bachmann，2025）。具体而言，我们将标题相似度计算为 $1 - d_{\text{lev}}/\max(|s_1|, |s_2|)$，其中 s_1 和 s_2 是进行比较的字符串，d_{lev} 表示莱文斯坦编辑距离（Levenshtein，1965年），并将其缩放至范围 $[0, 1]$。如果参考数据库中的所有标题相似度分数都低于 0。9，我们就将该引文视为可能不存在，并将其列入候选项列表以供人工核验。

观察。 表3 展示了候选引文和论文的统计数据。根据这些结果，我们

[4]https://github.com/acl-org/acl-anthology/tree/c9d3481/data[5]
h
t
t
ps://www.kaggle.com/datasets/Cornell-University/arxiv/versions/263[6]
https://drops.dagstuhl.de/entities/artifact/10.423001
7202512/dblp.xml.--我们通过 PyAlex (De Bruin，2023年) 使用 REST API。

观察到，每篇论文的候选引文平均数量和最大数量都随时间增加。除了引文和论文的原始计数外，相应的百分比也呈现出持续上升的趋势。特别是在比较2024年和2025年时，我们观察到目标论文和引文所占比例呈现出不同的趋势。这些观察结果表明，尽管我们的粗略筛选和简单的数据库匹配可能会引入一些噪声，但候选引文的增加幅度超过了仅凭此类方法噪声所能解释的范围。这表明，除我们方法的局限性之外，其他因素很可能也在促成所观察到的增长，其中最显著的是幻觉引文的影响。

2.4 人工核验

最后，我们对第 2.3 节中识别出的所有候选引文进行人工核验，以确定每条引文是否指向一篇实际存在的作品。由于我们的目标是检测包含幻觉引文的论文，因此一旦识别出至少一条幻觉引文，我们便立即将该论文标记为 *"HalluCited"*，并终止对该论文的进一步检查。我们对全部 2,950 篇目标论文应用了这一流程。

如果引文中提供了可识别信息，例如可点击的链接、数字对象标识符、arXiv 编号、页码或会议信息，我们会利用这些信息搜索被引用的论文。如果缺少此类信息，或信息不足以定位论文，我们会使用引文标题进行网络搜索，以查找可能对应的作品。如果找不到对应论文（例如，搜索结果只有目标论文自身），或者虽然找到了相似论文，但标题、作者、会议或页码等至少两项关键属性与任何可靠来源均不匹配，我们就将该引文归类为"幻觉引文"。这一核验流程采用了保守标准，基本上信任论文，并优先考虑精确率而非召回率。附录 B列出了 HalluCited 论文。

3 结果

发现 1：HalluCited 论文的数量在 2025年大幅增加，其中 EMNLP 占据了这些论文的一半以上。表 4 报告了包含幻觉引文的论文数量。结果显示，幻觉论文数量从 2024年的 20 篇增加到 2025年的 275 篇。尤其是，EMNLP 2025 单独就有 154 篇论文，几乎是此前观察到的数量的两倍。

年份	会议	主要发现	S.I.D.		WS	总计
2024年		1 0.1850.0	0 0.000.00	0 0.000.00	1 0.1650.0	2 0.12 —
2024	ACL	3 0.3142.9	1 0.1014.3	0 0.000.00	3 0.4242.9	7 0.26 —
2024年自	方法会议	3 0.2427.3	4 0.4036.4	2 1.1618.2	2 0.3918.2	11 0.37 —
2025年		6 0.8417.1	10 2.1128.6	2 1.135.71	17 3.0448.6	35 1.81 —
2025	ACL	14 0。 8216。3	18 1.30 21.0	10 3.86 11.6	44 3.7851.2	86 1.91 —
2025年自	法会议		45 3.20 29.2	13 4.83 8.44	36 5.26 23.4	154 3.70 —
2024		7 0.2535.0	5 0.2225.0	2 0.5710.0	6 0.3330.0	20 0.28 —
2025		80 1.8929.1	73 2.24 26.5	25 3.55 9.09	97 4.03 35.3	275 2.59 —
总体		87 1.2429.5	78 1.41 26.4	27 2.56 9.15	103 2.42 34.9	295 1.65 —

表 4：包含**"幻觉引文"**的论文数量。我们还报告了包含幻觉引文的论文占论文总数的比例，以及各发表场所内按专题划分的比例。我们将这些值记为 $\text{NUM}^{\text{Venue}\%}_{\text{专题}\%}$。请注意，在某些仅有少量论文的发表场所和专题中，百分比可能会受到个别论文的影响。尽管如此，这些比例仍有助于追踪趋势。

ACL 2025。对比 2024年和 2025，HalluCited 论文的数量增加了十倍以上。HalluCited 论文的占比也从 2024年的约 0.28% 上升至 2025年的 2.59%，并在 EMNLP 2025 达到 3.7%。检验 2025 年各会议的具体趋势后，我们发现，尽管研讨会论文约占 NAACL 和 ACL 的 HalluCited 论文总数的一半，但 EMNLP 2025 近 70% 的 HalluCited 论文发表于主会与 Findings 论文集。2025 年各分会场中 HalluCited 论文的占比也呈现出类似模式。在 NAACL，只有研讨会论文超过 3%；在 ACL，S.I.D. 分会场达到 3%。相比之下，在 EMNLP，所有分会场都超过 3%。这些结果表明，HalluCited 论文的影响已不再局限于研讨会或旁支分会场，而是日益波及主会分会场。这一发现表明，HalluCited 论文数量的迅速增加不太可能归因于特定分会场的缺陷，而是凸显了当前评审系统在检测 HalluCitation 方面日益显现的局限性，并揭示了其对可持续性和可信度造成的社会影响。

发现 2：包含多个幻觉引文候选项的论文极有可能包含实际的幻觉引文。表5 展示了候选论文和实际检测到的 HalluCited 论文的详细数量。根据这些结果，我们发现，当一篇论文包含约四个幻觉引文候选项时——

	候选项		HalluCited		命中率 (%)	
频率	Num.	Cum.	Num.	Cum.	Num.	Cum.
≧9	10	10	10	10	100.0%	100.0%
8	6	16	5	15	83.3%	93.8%
7	14	30	13	28	92.9%	93.3%
6	10	40	7	35	70.0%	87.5%
5	9	49	7	42	77.7%	85.7%
4	28	77	17	59	60.7%	76.6%
3	91	168	37	96	40.7%	57.1%
2	526	694	76	172	14.4%	24.8%
1	2,256	2,950	123	295	5.45%	10.0%

表5：候选论文和检测到的 HalluCited 论文数量。我们报告了每种幻觉引文候选项频次对应的论文数量。命中率表示每个频率区间内 HalluCited 论文所占的比例。“数量”表示各区间的原始计数，“累计”表示从较高频次区间累加的计数。对于包含九个或更多候选项的论文，所有 HalluCited 论文均已成功检测到，因此我们将它们归为一组。

在检测到的候选项中，近四分之三的论文确实包含幻觉引文。这一发现表明，仅通过简单的标题匹配来比对参考文献数据库，就足以检测出幻觉引文。相比之下，候选项较少的情况往往存在由光学字符识别或解析错误引入的噪声，以及简单字符级模糊匹配的局限性。虽然这些引文可能因解析失败或模糊匹配阈值过高而被错误标记，但此类不匹配很少连续出现。因此，如果在单篇论文中检测到三到四个候选项，则该论文会被视为可疑，并建议进一步核实。

发现3：人工检测颇具挑战，因为大多数 HalluCited 论文中只有少量幻觉引文。如表5所示，当一篇论文只有一两个幻觉引文候选项时，检测到的论文数量接近 200 篇，约占所有 HalluCited 论文的三分之二。这表明，相较于包含大量幻觉引文的论文，大多数 HalluCited 论文中只有少量幻觉引文。这也说明，这类幻觉引文往往混杂在其他真实引文之中。目前，只有与被引作品所属领域高度相关、具备丰富专业知识的审稿人或读者，才可能注意到这些幻觉引文。然而，如果引文涉及专业领域之外的内容，在有限的时间内（例如评审期间）手动检测此类幻觉几乎不可行——

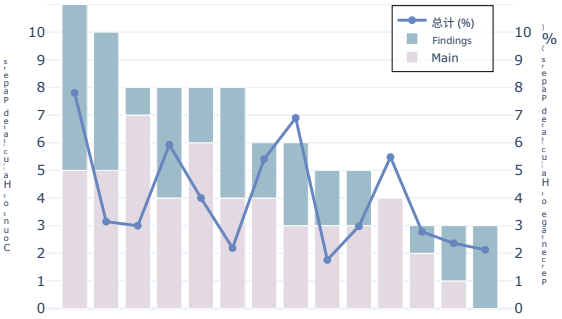

图 3：按领域划分的 HalluCited 论文数量及占比。我们报告包含三篇或更多论文的领域。领域名称采用前几个词的缩写。

——在可用时间有限的情况下，例如评审期间。此外，我们指出，有些幻觉引文未必总是出于有意引入。例如，作者可能会从二手来源（如谷歌学术）获取引文信息[8]、参考文献管理工具或基于大语言模型的推荐系统，而这些来源中的条目本身可能已经包含幻觉。我们将在第4.3节详细讨论这种可能性。这些幻觉引文表明，作者往往会默认信任此类引文，却不总是核实其真实性。此外，由于这类幻觉引文往往混杂在其他真实引文之中，因此最好在验证流程中引入自动检测系统。

4 进一步分析与讨论

4.1 趋势分析

哪些领域更常出现 HalluCited 论文？为了分析 HalluCited 论文的出现趋势，我们考察了这类论文获接收的研究领域。我们聚焦于 EMNLP 2025[9]，并通过识别每篇论文被分配的研究领域来分析主会与 Findings 论文集。图3 展示了各领域 HalluCited 论文的数量和占比。官方领域名称及完整领域列表请参阅 EMNLP 2025 征稿启事[10]。我们发现，诸如

[8] https://scholar.google.com/ [9] [访问日期：十二月 31 日， 2025] 领域信息可在 https://rrplanning2022-my.sharepoint.com/:x:/g/personal/jrachford_randrplanning_com/IQCJIl5zmNpUSZUv8j8H0z1TAVaIf1rOpMYhjCllLm703RY 通过 https://2025.emnlp.org/program。 [10] https://2025.emnlp.org/calls/main_会议_论文/#submission-topics



(a) **HalluCited 论文**中的词频-逆文档频率差异较大。



(b) **普通论文**中较高的词频-逆文档频率差异。

图 4：基于 HalluCited 论文与普通论文之间词频-逆文档频率差异生成的词云。

低资源自然语言处理、大语言模型效率和AI/大语言模型智能体领域的 HalluCited 论文数量和占比相对较高。相比之下，自然语言处理应用、可解释性以及资源与评估等领域的绝对数量相对较多，但占比较低。值得注意的是，若干 HalluCited 论文占比较高的领域，包括大语言模型安全与对齐、AI/大语言模型智能体和大语言模型效率，都是在 EMNLP 2025[11] 中新设立的。这表明，新兴主题可能难以找到合格审稿人，因而严格评审（例如检测幻觉）可能更具挑战性。

HalluCited 论文中包含哪些关键词? 为研究 HalluCited 论文的特征术语和主题，我们计算了（Manning 等人，2008）一般论文和 HalluCited 论文标题的词频-逆文档频率（TF-IDF）分数，重点关注 EMNLP 2025 的主会与 Findings 论文集。接下来，我们通过考察 TF-IDF 值差异较大的术语，对两组论文进行比较。图4展示了词云（Mueller，2023年），用于可视化各组中 TF-IDF 差异较大的术语。根据这些结果，我们观察到，HalluCited 论文更频繁地包含与“多模态”“解码”和“量化”等主题相关的术语。相比之下，一般论文往往包含与

[11]https://2025.emnlp.org/track-changes/

	2023		2024年*				2025		
频率	十月		二月				Feb.	May	July
≧9	0	0	1	0	0	0	2	6	4
8	0	0	0	0	0	0	0	1	0
7	0	0	0	0	0	0	2	3	0
6	0	0	0	0	0	0	2	3	0
5	0	0	2	0	3	0	1	5	1
4	0	0	0	0	0	0	5	6	2
3	2	0	4	1	3	0	13	14	6
2	4	18	37	3	39	3	62	60	17
1	51	87	186	30	226	17	322	237	72
候选项	57	105	230		271	20			102
总计	270	659	967		922	127			371
%	**21.1**	15.9	**23.8**		**29.4**		**30.3**	**24.3**	**27.5**
平均评审负荷									
meta	3.9	5.0	6.6		5.3	2.6	6.5	5.4	3.1
评审	2.2	2.4	3.7		3.1	1.8	3.9	3.2	2.5
可用预印本统计									
提交 1,									
%	21.2								

表 6：ARR 中已选择加入的预印本里，按幻觉引文候选项频率统计的候选论文数量。“候选项”表示至少包含一个幻觉引文候选项的论文数量，“总计”表示已选择加入的匿名预印本总数。**粗体**数值高于表3中报告的已接收论文平均比例（16.53%）。请注意，2024年10月和2024年12月被排除在外，因为这些数据尚未公开。2023年12月对应 NAACL 2024，2024年2月对应 ACL 2024，2024年6月对应 EMNLP 2024，2025年2月对应 ACL 2025，2025年5月对应 EMNLP 2025。我们还报告了每位元评审员和审稿人的平均分配数量。带下划线的数值表示超过预期值的情况。我们报告投稿总数和预印本披露率。

与强化学习相关的主题，例如“人类”“推理”和“偏好”。此外，HalluCited 论文的标题更可能使用“LLM”等缩写，而普通论文则更常使用完整形式，例如“大型语言模型”。这表明，HalluCited 论文倾向于采用简洁或缩略的术语，并关注与效率相关的主题，这与图3的结果一致。

4.2 同行评审流程如何?

表6 展示了 ACL 滚动评审（ARR）中选择加入的预印本里包含幻觉引文候选项的论文数量[12]，，并列出了审稿人分配统计数据及所有投稿的披露率。主会与 Findings 论文集的论文均投稿至 ARR，而选择加入选项的投稿则会以预印本形式公开发布。

[12]https://aclrollingreview.org/

此外，ARR 统计仪表板[13] 提供了与 ARR 相关的汇总统计数据。我们利用这些公开资源进行了分析。从表6可以看出，对于 2024 年以来与会议相关的投稿轮次，包含幻觉引文候选项的论文比例较高。然而，与表3中报告的已接收论文比例相比，这类论文在同行评审流程中会在一定程度上被筛除。我们还发现，包含幻觉引文候选项的论文经常会分配给元评审员；在某些投稿轮次（例如 2024 年 2 月），每位审稿人可能至少会分到一篇此类论文。这凸显了 HalluCited 论文所造成的社会影响，对我们而言，这是一个现实且迫在眉睫的问题。

然而，由于预印本披露率约为 20%，这些论文究竟更倾向于选择披露还是不披露，仍不清楚。因此，我们假设这些论文在披露方面呈均匀分布。为了促进回顾性研究并提高同行评审流程的透明度，加强评审相关信息的披露至关重要。

最后，我们注意到，这项分析并未评估 HalluCited 论文的质量。部分低质量投稿可能包含 HalluCitation，而其他因素也可能已经对筛选产生了作用。尽管如此，鉴于 EMNLP 2025 的已接收论文中 HalluCited 论文数量呈现增长趋势，即使仅依靠审稿人的努力，筛查此类情况也变得越来越困难。这也表明，有必要实施自动标记系统，以协助审稿人完成工作并保持评审的严谨性与质量。

4.3 无意幻觉引文：数据库往往受到污染

幻觉引文可能是作者无意间引入的。这类错误的一个常见来源是二手数据库，例如谷歌学术和 Semantic Scholar[14]，，这些数据库本身可能包含错误或不完整的条目。图5 展示了几个案例，说明了这一问题。在案例 1 中，论文标题被截断，导致无法与 arXiv 数据库正确匹配。在案例 2 中，我们确认所引用的论文并不存在。列出的作者没有在 ACL 2024 发表过相应论文，我们也找不到标题相似的论文，尽管该条目已列在谷歌 Scholar。在案例 3 中，引文关联了错误的 arXiv 编号，或者缺少一位作者。

案例 1：引用来源Li 等人 (2025d)

✗**Semantic Scholar（Ni 和 Li，2024a）：**

> Xuanfan Ni 和 Piji Li。2024a。对大语言模型的系统性评估以自然语言进行。ArXiv，abs/2405.10251。**(语料库 ID: 261341578)**

✓ **arXiv（正式版）（Ni 和 Li，2024b）：**

> Xuanfan Ni 和 Piji Li。2024b。对大语言模型在自然语言生成任务中的系统评估。预印本，arXiv:2405.10251。

案例 2：引用者Lee等人 (2025)

✗ **谷歌学术（不存在的参考文献）：**

> Wei Chen、Arjun Kumar 和 Lin Yang。2024年。基于干扰的攻击提示：一种有效的大语言模型越狱方法。第62届计算语言学协会年会论文集（ACL）。**(ID: 10238337808766132122)**

案例 3：引用者Ji等人(2025)

✗ **谷歌学术 (ID:3546576102853818527)：**

> Alex Ray、Joshua Achiam 和 Dario Amodei。2019 深度强化学习中的安全探索基准测试。arXiv 预印本 arXiv:1910.01708 ✓ Fujimoto等人 (2019), 7(1):2.

✗ **Semantic Scholar (语料库 ID: 208283920)：**

> Josh Achiam 和 Dario Amodei。2019。基准测-✗ 缺少第一作者 Alex Ray。试安全探索深度强化学习。URL: https://d4mucfpksywv.cloudfront.net/safexp-short.pdf。**(链接已失效)**

✓ **OpenAI（官方）（Ray等人，2019）：**

> Alex Ray、Josh Achiam 和 Dario Amodei。2019。深度强化学习中的安全探索基准测试。URL: https://cdn.openai.com/safexp-short.pdf

图 5：谷歌学术和 Semantic Scholar 中错误数据库条目的示例。案例 1 展示了错误的标题，案例 2 指向一篇不存在的论文，案例 3 则包含不准确的引文信息。

这些案例表明，从谷歌学术等二手来源获取引文信息并不能保证信息正确。特别是，如案例 2 所示，这类数据库中可能收录完全虚构的论文，而且其他论文实际上也引用了这条引文（Zhou，2025）。在案例 3 中，错误的 arXiv 编号已传播到数百条后续引文中。这些发现表明，出现幻觉引文并不一定意味着该引文是由人工智能系统生成的，或使用了人工智能工具，因为这些数据库中经常能发现错误的引文条目。作者很可能是手动从搜索引擎获取引文信息，

[13] https://stats.aclrollingreview.org/[14] https://www.semanticscholar.org/

或通过 Paperpile 等参考文献管理工具自动导入，从而无意中纳入错误条目。此外，尽早处理此类幻觉也至关重要。近期的 AI-Scientist 智能体能够通过 API 从这些数据库检索引文，这可能会加剧问题，并使其规模和复杂程度随着时间推移不断增加。[15],

不过，这些案例表明，作者并未直接查阅原始来源，而是转而依赖二手来源（Church，2017）。为缓解这一问题，我们有必要认识到数据库并非毫无差错，并鼓励作者直接从 ACL Anthology 或 arXiv 等一手来源获取引文信息。或者，使用 Ribiber 等引文规范化工具[16], 或许有助于减少此类错误，也应得到更广泛的采用。

5 条建议与推荐

在作者工具包中引入自动验证系统和评审检查。如第 4.3 节所述，幻觉引文可能源于二手来源中的错误。因此，不能将幻觉引文的出现直接归因于使用大语言模型或人工智能生成的论文。不过，无论其来源如何，这类幻觉引文都表明，作者往往未能仔细核实参考文献信息，也并非总会查阅原始来源。因此，我们认为有必要引入专门的幻觉引文检测工具，例如我们的工具或现有的基于大语言模型的智能体[17], , 并将其集成到 ACL pubcheck[18]等现有工具包中。在作者和组织者双方使用这些自动验证系统后，作者可以在投稿前收到预警，组织者也可以大规模自动扫描投稿。这将有助于实现更公平的核验、减轻审稿人工作量，并提高整体引文可靠性。

幻觉引文的定义与确保同行评审流程的可追溯性。明确界定哪些情况应视为幻觉引文至关重要，因为引文错误的严重程度并不相同。本研究根据被引用标题是否存在，以及 ID 和页码等关键标识符的一致性（Wang 和 Barabási，2021），但学术写作中长期存在轻微的引文错误，而不准确之处也可能源自二手来源（Besançon 等，2024年），因此其成因较为复杂。若只发现少量幻觉引文，通常只需要求更正。与期刊不同，期刊编辑可能会更正引文信息，而会议论文在很大程度上由作者负责此事。因此，不应对轻微的引文错误予以过度惩罚；提前提供作者工具包有助于避免许多格式错误。更严重的问题在于，这类更正是否体现在终稿中并不明确，这凸显了在修订和同行评审流程中实现可追溯性的必要性。提高可追溯性有助于核实更正，并支持对幻觉引文等新兴问题进行快速分析。

重新思考会议与同行评审的目的。“同行评审”究竟意味着什么？我们认为，审稿人是论文隐形的共同作者，他们的职责不仅是完成正式评估，还应参与建设性讨论，以改进论文。从这一视角来看，评审系统应优先保障充足的时间和可控的工作量，以便提供有意义的反馈，而不是采用鼓励审稿人在仅有数周的严格时限内快速完成评审的机制。目前，通过预印本（如 arXiv ）和博客文章提前发布成果已十分普遍。尽管这些成果未经同行评审正式核验，却已被广泛引用（Church 等人，2025）。与此同时，会议评审往往不如期刊同行评审严格。如果会议既无法提供预印本那样的即时性，又达不到期刊那样的严谨程度，就应重新审视其核心目的。日益沉重的评审负担可能会损害会议信誉，并引发对长期可持续性的担忧。若不进行结构性改革，这一状况很可能会恶化。一种可能的方向是整个社群转向大型期刊式模式，采用异步且严格的同行评审，同时由会议发表场所专注于展示已录用的研究成果。这种演进或能打造更可持续的评审流程，为核验工作留出充足时间，同时仍为作者提供有意义的研究成果展示机会。最初，ARR 系统和 Findings 专题曾

[15] https://paperpile.com/
[16] https://github.com/yuchenlin/rebiber
[17] https://gptzero.me/news/iclr-2026/
[18] https://github.com/acl-org/aclpubcheck

有望发挥这一作用。然而，ARR 周期与会议逐渐紧密绑定，Findings 最终被定位为一种“配套”发表渠道（Cohn 等人，2020年）。如今，ARR 和 Findings 推出已逾五年，或许到了重新思考它们定位的时候。一种可能的重新定义，是将它们视为与期刊等同的发表场所，而会议报告则作为已录用成果的一种认证或展示权，类似于 TMLR。虽然这无法从根本上解决投稿量增长的问题，但至少有助于维护信誉。随着社群不断发展，适应此类灵活模式或许会变得必要。

6 结论

我们检验了 ACL 会议已接收论文中幻觉引文的存在情况，分析了 2024年和2025年发表于 ACL、NAACL 和 EMNLP 的所有论文。最终，我们发现了近 300 篇 HalluCited 论文，其中超过半数发表于 EMNLP 2025。我们认为，不应将出现幻觉引文视为立即施加处罚的理由。正如前文所述，这类错误可能是无意造成的，而非作者蓄意的不当行为。与其采取惩罚措施，我们更强调开发作者工具包和投稿前检查的重要性，以主动防止此类问题。还值得注意的是，这些 HalluCited 论文因其实质性贡献而被接收。因此，不应事后处罚 HalluCited 论文的作者。由于 ACL Anthology 允许提交勘误修订，我们应更积极地鼓励作者自愿更正[19]。接收不应被视为非此即彼的奖励或惩罚，而应视为持续改进和传播研究成果的过程之一。

此外，重要的是重申同行评审的协作性质。会议旨在分享研究成果，而非在批量评审和严格的时间限制下奖励完美无缺的作品（Church，2005）。没有任何论文是毫无瑕疵的，健康的研究社群应对细微错误保持包容，同时积极倡导负责任的纠正。ACL 社区长期以来依靠 Findings 论文和 ARR 等自我改进机制得以发展。*ACL 并非人工智能/机器学习会议。*如今，自然语言处理社群可以主动塑造大语言模型时代的同行评审和会议。

局限性

范围。本研究聚焦于近期六场顶级自然语言处理会议。虽然可以扩大研究范围，但出于以下几个原因，我们将分析范围限定于此。首先，早些年的 HalluCited 论文数量微不足道，例如 NAACL 2024 仅有两例，而且这一趋势持续上升，因此进一步回溯分析所能提供的信息较少。其次，我们主要关注 2025 年观察到的激增现象，因此将重点放在最近召开的会议上。撰写本文时，其他发表场所（如 AACL 2025 或 ARR 2025 年 10 月周期）的资料尚未公开，因此无法纳入分析，尽管我们的研究结果表明，这些场所也可能存在数量可观的 HalluCitations。虽然可以将分析扩展到其他人工智能/机器学习会议，但我们的目标是评估其在自然语言处理社群内的具体影响。总的来说，尽管研究范围有限，我们认为本分析已足以全面满足预期目的；更广泛的研究（例如纳入其他会议和期刊）将留待未来开展。

分析。 我们主要分析了已接收论文。首先，已接收论文很容易获取，因此可以进行大规模分析。其次，我们发现，即使是已接收论文也包含虚构引文，这表明在时间紧迫、审稿人工作量繁重的情况下，同行评审流程未必总能充分核实引文是否准确。从这个角度来看，分析已接收论文既合理又有参考价值。尽管如此，我们也考察了公开可用的 ARR 数据，并在现有披露信息的范围内开展了全面调查。然而，正如第 4.2 节所述，同行评审流程的透明度仍然有限；由于无法充分获取评审材料，目前尚无法进行更深入的分析。此外，被拒论文通常不会公开。尽管已有一些数据发布工作[20] (Dycke 等人，2023年，2022年)，但这些资源仅涵盖已明确获得同意的情况。此外，审稿人要求的修改是否体现在终稿版本中，其可追溯性仍不明确。我们认为，学术论文是作者与审稿人协作的成果。因此，论文也体现了审稿人的反馈

[19] https://aclanthology.org/info/corrections/

[20] https://arr-data.aclweb.org/

。从这个角度来看，评审材料（包括被拒论文）应当向公众开放（Church 等人，2025；Yang，2025；Kim 等人，2025d）。这种透明度将有助于分析评审质量和结果。

方法。我们重点关注 arXiv 和 ACL 文集这两个可靠且成熟的数据源。尽管将分析扩展到人工智能、机器学习、计算机视觉、机器人学、语音和期刊等更广泛领域，是未来研究的重要方向，但这些领域目前缺乏统一且持续维护的书目基础设施，无法与 ACL 社区的书目基础设施相媲美。由于本研究聚焦于 ACL 会议，尽管范围有限，我们仍识别出近 300 篇 HalluCited 论文，这表明我们的方法是有效的。在实现方面，我们使用 MinerU 进行光学字符识别，不过也可以采用其他光学字符识别工具（Wei 等人，2025a；Cui 等人，2025a），并进行额外的工程改进。这些实现选择并非我们的主要贡献。相反，本研究的核心贡献是定量且全面地展示幻觉引文的普遍程度及其影响。最后，我们的方法优先考虑精确率，因此应将报告的结果解读为下界。除本文检测到的内容外，可能还存在其他幻觉引文。我们希望本研究能作为基线，并推动后续工作提升检测准确率和覆盖率。

伦理考量

信息。本手稿中的所有信息均反映截至 2026 年 1 月 4 日公开可用数据的状态。除非另有明确说明，链接、元数据和数据集内容均对应于当时可用的信息。此日期之后的更新、更正或删除可能会改变这些信息，本研究未予反映。

许可。我们的主要数据源是 ACL Anthology 和 OpenReview，二者均依据知识共享署名 4.0 国际许可协议发布。我们确认，本研究使用的所有其他数据集、数据转储和元数据均可公开获取，并采用宽松许可。本研究开发的代码和工具均由我们编写。本论文的人工核验完全由作者进行。因此，本研究中的人工标注不存在许可或同意方面的问题。

关于个人可识别信息保护和潜在危害的讨论。本研究分析了真实的已发表论文。因此，HalluCited 论文的作者信息本身即可被观察到。然而，隐藏 HalluCited 论文的身份将严重妨碍可复现性。此外，本文讨论的幻觉引文已经可以通过公共搜索引擎识别，因此属于现有公共信息。并且，我们的分析全面且系统，不针对或单独指向特定个人。因此，本研究严格作为学术研究开展。需要特别指出的是，正如第 6 节明确说明的那样，我们不将幻觉引文归因于作者不当行为，也不主张应惩罚作者。相反，我们强调当前评审系统的局限性，尤其是缺乏适当的作者工具包。因此，本分析及其主张和目的无意冒犯或指责任何人。为确保透明度和可复现性，我们在第 B 节提供了 HalluCited 论文列表。我们不引入新的标注或主观判断，仅核验并列出公共记录中已经存在的幻觉引文。作为额外的保护措施，表7（位于第B 节）不会直接展示 HalluCited 论文的作者姓名或论文标题。相反，每篇 HalluCited 论文仅通过指向相应参考文献条目的链接加以引用，并且我们为每篇论文展示一个幻觉引文示例。这一设计在保护作者的同时，兼顾了透明度和可复现性。因此，本文通篇的讨论都保持中立并面向未来，重点关注建设性解决方案，而非归咎于任何人。我们已仔细审阅 ACL 伦理政策[21]，并确认本研究符合其指南。

工具使用。我们使用 DeepL、ChatGPT 和 Grammarly 进行翻译和语法改进。所有原创内容均由作者撰写，作者对最终稿件负全部责任。实验中，我们使用 NVIDIA A6000 图形处理器进行光学字符识别处理。其他分析则使用 MacBook Pro、电子表格和 Google Chrome 进行文献检索。值得注意的是，所有与 HalluCitations 相关的记录均在本地管理，未传输至任何外部 API 或服务。因此，工具使用和数据管理均符合伦理考量。

[21] https://ethics.aclweb.org/

参考文献

2008–2025。Grobid。
https://github.com/kermitt2/grobid。

Aastik、Meghana Topu、Chinmay Kulkarni 和 Pragya Paramita Sahu。2025。NormAL LoRA：什么才是完美的大小？载于计算语言学协会论文集：EMNLP 2025，第 19716–19731 页，中国苏州。计算语言学协会。

Mohammad Mahdi Abootorabi、Amirhosein Zobeiri、Mahdi Dehghani、Mohammadali Mohammadkhani、Bardia Mohammadi、Omid Ghahroodi、Mahdieh Soleymani Baghshah 和 Ehsaneddin Asgari。2025。任意模态皆可提问：多模态检索增强生成综合综述。载于计算语言学协会论文集：ACL 2025，第 16776–16809 页，奥地利维也纳。计算语言学协会。

Tewodros Achamaleh、TolulopeOlalekan Abiola、Lemlem Eyob Kawo、Mikiyas Mebraihtu 和 Grigori Sidorov。2025。CIC-NLP@DravidianLangTech 2025：检测达罗毗荼语人工智能生成的产品评论。载于第五届达罗毗荼语言语音、视觉与语言技术研讨会会议论文集，第 502–507 页，阿科马，阿尔伯克基会议中心，新墨西哥州阿尔伯克基。计算语言学协会。

Priyobroto Acharya、Haranath Mondal、Dipanjan Saha、Dipankar Das 和 Sivaji Bandyopadhyay。2025年。JU-NLP：通过高效的基于 LoRA 的适配改进低资源印度语翻译系统。载于《第十届机器翻译大会论文集》，第 1201–1209 页，中国苏州。计算语言学协会。

Muhammad Adilazuarda。2024年。Beyond Turing：检测机器生成文本方法的比较分析。载于《可信自然语言处理研讨会（TrustNLP 2024）论文集》，第 1–12 页，墨西哥墨西哥城。计算语言学协会。

Sabik Aftahee、Tofayel Ahmmed Babu、MD Musa Kalimullah Ratul、Jawad Hossain 和 Mohammed Moshiul Hoque。2025。CUET_NetworkSociety@DravidianLangTech 2025：一种用于检测低资源达罗毗荼语产品评论中人工智能生成内容的基于 Transformer 的方法。载于《第五届达罗毗荼语言语音、视觉与语言技术研讨会论文集》，第 522–528 页，阿科马、阿尔伯克基会议中心、新墨西哥州阿尔伯克基。计算语言学协会。

Mahmoud Ahmad 和 Habeebah Kakudi。2025年。使用 BERT 对尼日利亚2023年选举推文进行立场检测：一种适用于低资源场景的 Transformer 方法。载于第六届计算话语、上下文与文档级推理研讨会（CODI 2025）会议论文集，第 54–63 页，中国苏州。计算语言学协会。

Mohammad Sadegh Akhondzadeh、Aleksandar Bo- jchevski、Evangelos Eleftheriou 和 Martino Dazzi。2025 年。KurTail：基于峰度的大语言模型量化。载于《计算语言学协会论文集：EMNLP 2025》，第 17404–17419 页，苏州，中国。计算语言学协会。

Maro Akpobi。2025 年。Yankari：约鲁巴语单语数据集。载于《第六届非洲自然语言处理研讨会（AfricaNLP 2025）会议论文集》，第 1–6 页，维也纳，奥地利。计算语言学协会。

Diyam Akra、Mohammed Khalilia 和 Mustafa Jar- rar。2025 年。多方言阿拉伯语词性标注的主动学习。载于《计算语言学协会论文集：EMNLP 2025》，第 24960–24973 页，苏州，中国。计算语言学协会。

Flor Alberts、Ivo Bruinier、Nathalie Palm、J ustin Paet- zelt 和 Erik Varecha。2025年。FENJI 在 SemEval-2025 任务 3 中：检索增强生成与幻觉片段检测。载于《第19届国际语义评测研讨会论文集》（SemEval-2025），第1143–1151页，奥地利维也纳。计算语言学协会。

Falwah Alhamed、Rebecca Bendayan、Julia Ive 和 Lucia Specia。2024年。监测用户生成内容中的抑郁严重程度与症状：标注方案与指南。载于《第十四届主观性、情感与社交媒体分析计算方法研讨会论文集》，第227–233页，泰国曼谷。计算语言学协会。

Dana R Alsagheer、Abdulrahman Kamal、Moham- mad Kamal、Cosmo Yang Wu 和 Weidong Shi。2025年。那个从不思考的律师：一致性与公平性是实现可靠人工智能的关键。载于《计算语言学协会第63届年会论文集》（第1卷：长篇论文），第9943–9954页，奥地利维也纳。计算语言学协会。

JesusAlvarezC, DauaKarajeanes,AshleyPrado, JohnRuttan, IvoryYang, SeanO'brien, VasuSharma,and KevinZhu. 2025. 推进犹他-阿兹特克语技术发展：濒危科曼奇语案例研究。收录于《第五届美洲原住民语言自然语言处理研讨会》（AmericasNLP）会议论文集，第 27–37 页，阿尔伯克基，新墨西哥州。计算语言学协会。

Anson Antony 和 Annika Schoene。2025。增强检索的心理健康评估：利用社交媒体和上下文学习捕捉自我状态动态。载于第十届研讨会论文集，主题为

计算语言学与临床心理学（CLPsych 2025），第268–278页，美国新墨西哥州阿尔伯克基。计算语言学协会。

Saurav Aryal 和 Kritika Pant。2025年。霍华德大学-AI4PC 在 SemEval-2025 任务 9 中：使用开放权重 BART-MNLI 对食品召回文档进行零样本分类。载于《第19届国际语义评测研讨会》（SemEval-2025）会议论文集，第1919–1923页，奥地利维也纳。计算语言学协会。

Akari Asai、Jacqueline He、Rulin Shao、Weijia Shi、Amanpreet Singh、Joseph Chee Chang、Kyle Lo、Luca Soldaini、Sergey Feldman、Mike D'arcy、David Wadden、Matt Latzke、Minyang Tian、Pan Ji、Shengyan Liu、Hao Tong、Bohao Wu、Yanyu Xiong、Luke Zettlemoyer 等 6 人。2024年。OpenScholar：利用检索增强语言模型综合科学文献。预印本，arXiv:2411.14199。

Yash Kumar Atri、Ahmed Alaa 和 Thomas Hartvigsen。2025。基于图的外部记忆终身模型编辑。载于计算语言学协会会议论文集：ACL2025，第13336–13352 页，奥地利维也纳。计算语言学协会。

Mohammad Anas Azeez、Rafiq Ali、Ebad Shabbir、Zohaib Hasan Siddiqui、Gautam Siddharth Kashyap、Jiechao Gao 和 Usman Naseem。2025。真相、信任与困境：边缘端的医疗人工智能。载于2025年自然语言处理实证方法会议：产业分会论文集，第 1017–1025 页，中国苏州。计算语言学协会。

Massa Baali、Xiang Li、Hao Chen、Syed Abdul Hannan、Rita Singh 和 BhikshaRaj。2025。CAARMA：采用对抗性混合正则化的类别增强。载于计算语言学协会论文集：EMNLP 2025，第9732–9742 页，中国苏州。计算语言学协会。

Max Bachmann。2025。rapidfuzz/rapidfuzz：发布版本 3.13.0。

Thomas Bailleux、Tanmoy Mukherjee、Pierre Mar- quis 和 Zied Bouraoui。2025。连接概念层与解释依据，以增强语言模型可解释性。载于《第十四届词汇与计算语义联合会议（*SEM 2025）会议论文集》，第409–429页，中国苏州。计算语言学协会。

Mohammad Jahid Ibna Basher、Md Kowsher、Md Saiful Islam、Rabindra Nath Nandi、Nusrat Jahan Prottasha、Mehadi Hasan Menon、Tareq Al Muntasir、Sham- mur Absar Chowdhury、Firoj Alam、Niloofar Yousefi 和 Ozlem Garibay。2025。BnTTS：低资源环境下的少样本说话人自适应。载于计算语言学协会论文集：NAACL 2025，第 4956–4968 页，阿尔伯克基，新墨西哥州。计算语言学协会。

Samuel Bell、Eduardo Sánchez、David Dale、Pontus Stenetorp、Mikel Artetxe 和 Marta R. Costa-Jussà。2025。先翻译，再检测：利用机器翻译进行跨语言毒性分类。载于第十届机器翻译大会论文集，第 253–268 页，中国苏州。计算语言学协会。

Anya Belz 和 Craig Thomson。2025。HEDS 3.0：人工评估数据表第3版。载于第四届生成、评估与指标研讨会（GEM²），第60–81页，奥地利维也纳及线上会议。计算语言学协会。

Lonni Besançon、Guillaume Cabanac、Cyril Labbé 和 Alexander Magazinov。2024年。暗中插入的参考文献：伪造的参考文献元数据扭曲引文计数。《信息科学与技术协会学报》，75(12)：1368–1379。

Niloufar Beyranvand、Hamidreza Dastmalchi、Aijun An、Heidar Davoudi、Winston Chan 和 Ron DiCarlan- tonio。2025。GEAR：面向基于检索增强生成的汽车助手系统的可扩展、可解释评估框架。载于《2025年自然语言处理实证方法会议：产业分会论文集》，第2792–2810页，苏州（中国）。计算语言学协会。

Debarpan Bhattacharya、Apoorva Kulkarni 和 Sriram Ganapathy。2025。FESTA：用于多模态大语言模型可信度评估的功能等效采样。载于《计算语言学协会论文集：EMNLP 2025》，第12277–12295页，苏州，中国。计算语言学协会。

Nikhil Bhendawade、Irina Belousova、Qichen Fu、Henry Mason、Antonie Lin、Mohammad Rastegari 和 Mahyar Najibi。2025。推测式流处理：利用多流注意力实现高效、可扩展的推测解码。载于《2025年自然语言处理实证方法会议论文集》，第19547–19570页，苏州，中国。计算语言学协会。

Guanqun Bi、Zhuang Chen、Zhoufu Liu、Hongkai Wang、Xiyao Xiao、Yuqiang Xie、Wen Zhang、Yongkang Huang、Yuxuan Chen、Libiao Peng 和 Minlie Huang。2025 年。MAGI：多智能体引导式精神科评估访谈。载于《协会计算语言学：ACL 2025 论文集》，第 24898–24921 页，奥地利维也纳。计算语言学协会。

Suhas Bn、Dominik O. Mattioli、Andrew M. Sherrill、Rosa I. Arriaga、Christopher Wiese 和 Saeed Abdullah。2025。合成治疗对话有多真实？评估长期暴露对话中的保真度。载于《计算语言学协会论文集：EMNLP 2025》，第 20986–20995 页，苏州，中国。计算语言学协会。

Ondrej Bohdal、Mete Ozay、Jijoong Moon、Kyenghun Lee、Hyeonmok Ko 和 Umberto Michieli。2025。面向设备端大语言模型的高效组合式多任务处理。载于《2025年自然语言处理实证方法会议论文集》，第 28129–28153 页，苏州，中国。计算语言学协会。

Abhilekh Borah、Chhavi Sharma、Danush Khanna、Utkarsh Bhatt、Gurpreet Singh、Hasnat Md Abdullah、Raghav Kaushik Ravi、Vinija Jain、Jyoti Patel、Shub- ham Singh、Vasu Sharma、Arpita Vats、Rahul Raja、Aman Chadha 和 Amitava Das。2025。对齐质量指数（AQI）：超越拒绝响应：通过潜在几何结构、聚类分歧和逐层池化表示，将 AQI 用作内在对齐诊断。载于《2025年自然语言处理实证方法会议论文集》，第2888–2947页，中国苏州。计算语言学协会。

Abdessalam Bouchekif、Samer Rashwani、Heba Sbahi、Shahd Gaben、Mutaz Al Khatib 和 Mohammed Ghaly。2025。评估大语言模型的伊斯兰法律推理能力：来自继承法评估的证据。载于《第三届阿拉伯语自然语言处理会议论文集》，第246–257页，中国苏州。计算语言学协会。

Hossam Boudraa、Benoit Favre 和 Raquel Urena。2025。在零样本和少样本设置下，使用选择性十亿参数以下模型进行隐式仇恨目标跨度检测。载于第九届在线虐待与伤害研讨会（WOAH）会议论文集，第 228–240 页，奥地利维也纳。计算语言学协会。

Nicolas Bougie 和 Narimawa Watanabe。2025。生成式审稿人智能体：可扩展的同行评审模拟体。载于 2025 年自然语言处理实证方法会议：工业赛道会议论文集，第 98–116 页，中国苏州。计算语言学协会。

Hongyi Cai、Jie Li、Mohammad Mahdinur Rahman 和 Wenzhen Dong。2025。低置信度黄金数据：精炼低置信度样本以实现高效指令调优。载于计算语言学协会发现论文集：EMNLP 2025，第 8233–8240 页，中国苏州。计算语言学协会。

Crivoi Carla 和 Ana Sabina Uban。2025 年。SciBERT 遇上对比学习：一种用于科学幻觉检测的解决方案。载于《第五届学术文档处理研讨会》（SDP 2025）会议论文集，第 336–343 页，奥地利维也纳。计算语言学协会。

EgeYi˘git Çelik 和 Selma Tekir。2025。Cite-BART：学习生成用于局部引文推荐的引文。载于《2025年自然语言处理实证方法会议论文集》，第1703–1719页，中国苏州。计算语言学协会。

Ernie Chang、Yang Li、Patrick Huber、Vish Vogeti、David Kant、Yangyang Shi 和 Vikas Chandra。2025。AutoMixer：将检查点工件用作自动数据混合器。载于《计算语言学协会第63届年会论文集》（第1卷：长篇论文），第19942–19953页，奥地利维也纳。计算语言学协会。

Yu Chao、Siyu Lin、Xiaorong Wang、Zhu Zhang、Zihan Zhou、Haoyu Wang、Shuo Wang、Jie Zhou、Zhiyuan Liu 和 Maosong Sun。2025 年。大语言模型×MapReduce-v3：通过 MCP驱动的分层模块化智能体系统实现交互式深度综述生成。载于《2025年自然语言处理实证方法会议：系统演示》论文集，第688–695 页，中国苏州。计算语言学协会。

Shraddha Chauhan 和 Abhinav Kumar。2025 年。MNLP@DravidianLangTech 2025：用于厌女表情包检测的基于 Transformer 的多模态框架。载于《第五届达罗毗荼语言语音、视觉与语言技术研讨会》论文集，第 248–253 页，阿科马，阿尔伯克基会议中心，阿尔伯克基，新墨西哥州。计算语言学协会。

Divij Chawla、Ashita Bhutada、Duc Anh Do、Abhinav Raghunathan、Vinod Sp、Cathy Guo、Dar Win Liew、Prannaya Gupta、Rishabh Bhardwaj、Rajat Bhard- waj 和 Soujanya Poria。2025。评估金融人工智能：人工智能是否能可靠地评估投资风险偏好？载于《2025年自然语言处理实证方法会议：工业赛道论文集》，第2828–2839页，中国苏州。计算语言学协会。

Bowen Chen、Zhao Wang 和 Shingo Takamatsu。2025a。OMS：通过大语言模型智能体即时生成多目标、自我反思的广告关键词。载于《2025年自然语言处理实证方法会议论文集》，第18594–18612页，中国苏州。计算语言学协会。

Danlu Chen、Freda Shi、Aditi Agarwal、Jacobo My- erston 和 Taylor Berg-Kirkpatrick。2024年。Lo- gogramNLP：比较古代表意文字书写系统的视觉表征与文本表征在自然语言处理中的应用。载于《第62届计算语言学协会年会论文集》（第1卷：长篇论文），第14238–14254页，泰国曼谷。计算语言学协会。

Diyang Chen。2025。pingan-team 在 SemEval-2025 任务 2 中的参赛方案：采用 LoRA 增强的 Qwen2.5，并利用维基数据进行实体翻译。载于《第19届国际语义评测研讨会论文集》（SemEval-2025），第2065–2070页，奥地利维也纳。计算语言学协会。

Hanjie Chen、ZhouxiangFang、YashSingla 和 Mark Dredze。2025b。对大型语言模型回答和解释复杂医学问题的基准测试

问题。载于计算语言学协会美洲分会2025年会议论文集：人类语言技术（第1卷：长篇论文），第3563–3599页，阿尔伯克基，新墨西哥州。计算语言学协会。

Meng Chen、Philip Arthur、Qianyu Feng、Cong Duy Vu Hoang、Yu-Heng Hong、Mahdi Kazemi Moghad- dam、Omid Nezami、Duc Thien Nguyen、Gioacchino Tangari、Duy Vu、Thanh Vu、Mark Johnson、Krish- naram Kenthapadi、Don Dharmasiri、Long Duong 和 Yuan-Fang Li。2025c。掌握代码大语言模型的数据合成技艺。载于《2025年计算语言学协会美洲分会人类语言技术会议论文集》（第1卷：长篇论文），第12484–12500页，美国新墨西哥州阿尔伯克基。计算语言学协会。

Qiguang Chen、Mingda Yang、Libo Qin、Jinhao Liu、Zheng Yan、Jiannan Guan、Dengyun Peng、Yiyan Ji、Hanjing Li、Mengkang Hu、Yimeng Zhang、Yihao Liang、Yuhang Zhou、Jiaqi Wang、Zhi Chen 和 Wanxiang Che。2025d。AI4Research：面向科学研究的人工智能综述。预印本，arXiv:2507.01903。

Yilong Chen、Junyuan Shang、Zhenyu Zhang、Yanxi Xie、Jiawei Sheng、Tingwen Liu、Shuohuan Wang、Yu Sun、Hua Wu 和 Haifeng Wang。2025e。内部思考 Transformer：利用动态深度缩放促进自适应内部思考。载于《计算语言学协会第63届年会论文集》（第1卷：长篇论文），第28241–28259页，奥地利维也纳。计算语言学协会。

Yuhang Chen、Zhen Tan、Ajay Kumar Jaiswal、Huaizhi Qu、Xinyu Zhao、Qi Lin、Yu Cheng、Andrew Kwong、Zhichao Cao 和 Tianlong Chen。2025f。大语言模型中的位翻转错误韧性：全面分析与防御框架。载于《2025年自然语言处理实证方法会议论文集》，第10425–10435页，中国苏州。计算语言学协会。

Veer Chheda、Aaditya Uday Ghaisas、Avantika Sankhe 和 Narendra Shekokar 博士。2025年。Extract-explain-摘要：面向印度法律文档的修辞角色驱动的领域专用摘要框架。载于《2025年自然法律语言处理研讨会会议论文集》，第439–455页，中国苏州。计算语言学协会。

Kenneth Church。2005年。最后的话：评审评审者。计算语言学，31(4)：575–578。

Kenneth Ward Church。2017年。Word2vec。自然语言工程，23(1)：155–162。

Kenneth Ward Church、Raman Chandrasekar、John E. Ortega 和 Ibrahim Said Ahmad。2025年。同行评审值得付出努力吗？载于《会议论文集》第31届国际计算语言学会议，页3589–3599，阿布扎比，阿联酋。计算语言学协会。Jonathan H. Clark、Dan Garrette、Iulia Turc 和 John Wieting。2022年。Canine：为语言表征预训练的高效无分词编码器。计算语言学协会会刊，10：73–91。Trevor Cohn、Yulan He 和 Yang Liu，编。2020年。计算语言学协会论文集：EMNLP 2020。计算语言学协会，线上。Gaye Colakoglu、Gürkan Solmaz 和 Jonathan Fürst。2025。问题解决了吗？使用大语言模型处理复杂布局文档的信息抽取设计空间。载于计算语言学协会论文集：EMNLP 2025，页17908–17927，中国苏州。计算语言学协会。Cheng Cui、Ting Sun、Suyin Liang、Tingquan Gao、Zelun Zhang、Jiaxuan Liu、Xueqing Wang、Changda Zhou、Hongen Liu、Manhui Lin、Yue Zhang、Yubo Zhang、Handong Zheng、Jing Zhang、Jun Zhang、Yi Liu、Dianhai Yu 和 Yanjun Ma。2025a。Paddleocr-vl：通过0.9B超紧凑视觉语言模型提升多语言文档解析能力。预印本，arXiv:2510.14528。Wendi Cui、Jiaxin Zhang、Zhuohang Li、Hao Sun、Damien Lopez、Kamalika Das、Bradley A. Malin 和 Sricharan Kumar。2025b。自动指令聚焦提示词优化中的启发式搜索算法：综述。载于计算语言学协会论文集：ACL 2025，页22093–22111，奥地利维也纳。计算语言学协会。Xia Cui。2025。xiacui 在 SemEval-2025 任务11中的参赛系统：通过加权损失解决基于Transformer的多标签情感检测中的数据不平衡问题。载于第19届国际语义评测研讨会（SemEval-2025）会议论文集，页247–256，奥地利维也纳。计算语言学协会。Amitava Das、Yaswanth Narsupalli、Gurpreet Singh、Vinija Jain、Vasu Sharma、Suranjana Trivedy、Aman Chadha 和 Amit Sheth。2025a。YinYang-align：一种用于竞争目标的新基准，以及用于文本到图像对齐的多目标偏好方法。载于计算语言学协会论文集：ACL 2025，页23518–23598，奥地利维也纳。计算语言学协会。Amitava Das、Suranjana Trivedy、Danush Khanna、Yaswanth Narsupalli、Basab Ghosh、Rajarshi Roy、Gurpreet Singh、Vinija Jain、Vasu Sharma、Aishwarya Naresh Reganti 和 Aman Chadha。2025b。DPO核：一种语义感知、核增强且富散度的直接偏好优化范式。载于

计算语言学协会会议论文集：ACL 2025，页22174–22270，奥地利维也纳。计算语言学协会。

Jonathan De Bruin。2023年。PyAlex。

Varun Dhanraj 和 Chris Eliasmith。2025。使用神经符号表征改进大语言模型中的基于规则的推理。载于《2025年自然语言处理实证方法会议论文集》，第30577–30596页，中国苏州。计算语言学协会。

Yiwen Ding、Jiarui Liu、Zhiheng Lyu、Kun Zhang、Bernhard Schölkopf、Zhijing Jin 和 Rada Mihalcea。2025年。她的声音：分析人工智能出版界的性别差异。载于《第四届积极影响自然语言处理研讨会（NLP4PI）论文集》，第196–214页，奥地利维也纳。计算语言学协会。

Bonaventure F. P. Dossou 和 Henri Aïdasso。2025年。迈向低资源自然语言处理的开放式发现。载于《第二届不确定性感知自然语言处理研讨会（UncertaiNLP2025）论文集》，第287–297页，中国苏州。计算语言学协会。

Razvan-Gabriel Dumitru、Minglai Yang、Vikas Yadav 和 Mihai Surdeanu。2025。CopySpec：通过推测式复制粘贴加速大语言模型。载于《2025年自然语言处理实证方法会议论文集》，第26312–26343页，中国苏州。计算语言学协会。

Nils Dycke、Ilia Kuznetsov 和 Iryna Gurevych。2022年。Yes-yes-yes：面向 ACL 滚动评审及更广范围的主动数据收集。载于《协会发现论文集：EMNLP 2022》，第300–318页，阿布扎比，阿联酋。计算语言学协会。

Nils Dycke、Ilia Kuznetsov 和 Iryna Gurevych。2023年。NLPeer：用于同行评审计算研究的统一资源。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5049–5073页，加拿大多伦多。计算语言学协会。

Nellia Dzhubaeva、Katharina Trinley 和 Laura Pis- sani。2025年。思维无序，机器可预测：人类与大语言模型意识流写作中叙事连贯性的比较研究。载于《计算语言学协会第63届年会论文集》（第4卷：学生研究研讨会），第1079–1096页，奥地利维也纳。计算语言学协会。

Roxanne El Baff、Dominik Opitz 和 Diaoulé Diallo。2025。CriticalBrew 在 CQs-gen 2025：面向论证的批判性问题的协作式多智能体生成与评估。载于第12届论证挖掘研讨会会议论文集，第314–321页，奥地利维也纳。计算语言学协会。

Vignesh Ethiraj、Ashwath D、Sidhanth Menon、Divya Vijay 和 Vidhyakshaya Kannan。2025年。T-VEC：一种电信领域专用向量化模型，通过深度三元组损失微调增强语义理解。载于《2025年自然语言处理实证方法会议：工业赛道论文集》，第2449–2460页，中国苏州。计算语言学协会。

Egecan Evgin、Ilknur Karadeniz 和 Olcay Taner Yıldız。2025年。MetninOzU 在 BioLaySumm2025：采用逆向数据增强并注入显著句子的文本摘要。载于《第24届生物医学语言处理研讨会（共享任务）论文集》，第179–184页，奥地利维也纳。计算语言学协会。

Neo Eyal、Nachum Dershowitz 和 Kfir Bar。2025。大语言模型中的层复制。收录于计算语言学协会论文集：EMNLP 2025，第 17797–17807 页，中国苏州。计算语言学协会。

Yijia Fan、Jusheng Zhang、Kaitong Cai、Jing Yang 和 Keze Wang。2025。CCG：基于神经语义驱动因果博弈的稀有标签预测。收录于计算语言学协会论文集：EMNLP 2025，第 6243–6256 页，中国苏州。计算语言学协会。

MaryamFatima。2025。FIRMA：基于证明论基础的双向形式语言与非形式数学语言对齐。收录于第三届数学自然语言处理研讨会（MathNLP 2025）会议论文集，第 62–76 页，中国苏州。计算语言学协会。

Arvid Frydenlund。2025。语言模型、图搜索与监督污染：当更多监督反而更少，以及如何让更多变得更多。载于《计算语言学协会第63届年会论文集》（第1卷：长篇论文），第29011–29059页，维也纳，奥地利。计算语言学协会。

Zichuan Fu、Xian Wu、Guojing Li、Yingying Zhang、YefengZheng、Tianshi Ming、YejingWang、Wanyu Wang和Xiangyu Zhao。2025。用于知识编辑的模型融合。载于《计算语言学协会第63届年会论文集》（第6卷：工业赛道），第433–443页，维也纳，奥地利。计算语言学协会。

Scott Fujimoto、Edoardo Conti、Mohammad Ghavamzadeh和Joelle Pineau。2019。批量深度强化学习算法基准测试。预印本，arXiv:1910.01708。

Achille Fusco、Maria Letizia Piccini Bianchessi、Tommaso Sgrizzi、Asya Zanollo 和 Cristiano Chesi。2025。作为词元的语言单位：使用 BabyLM 进行内在和外在评估。载于

首届 BabyLM 研讨会，496–507 页，中国苏州。计算语言学协会。

Jack Gallifant、Shan Chen、Kuleen Sasse、Hugo Aerts、Thomas Hartvigsen 和 Danielle Bitterman。2025。用于分类和可迁移性的稀疏自编码器特征。载于《2025年自然语言处理实证方法会议论文集》，第29927–29951 页，中国苏州。计算语言学协会。

Fan Gao、Cheng Huang、Yutong Liu、Nyima Tashi、Xi- angxiang Wang、Thupten Tsering、Ban Ma-bao、Ren- zeng Duojie、Gadeng Luosang、Rinchen Dongrub、Dorje Tashi、Xiao Feng Cd、Yongbin Yu 和 Hao Wang。2025。TLUE：藏语理解评估基准。载于《2025年自然语言处理实证方法会议论文集》，第 35059–35085 页，中国苏州。计算语言学协会。

Rahul Garg、Trilok Padhi、Hemang Jain、Ugur Kursuncu 和 Ponnurangam Kumaraguru。2025。Just KIDDIN'：用于检测不雅模因的知识注入与蒸馏。载于《计算语言学协会论文集：ACL 2025》，第23067–23086 页，奥地利维也纳。计算语言学协会。

Carol-Luca Gasan 和 Vasile Pais。2025。SAG：利用语义增强图提升领域特定信息检索。载于《第十四届词汇与计算语义联合会议（*SEM 2025）论文集》，第 362–371 页，中国苏州。计算语言学协会。

Ryan George、Akshay Govind Srinivasan、Jayden Koshy Joe、Harshith M R、Vijayavallabh J、Hrushikesh Kant、Rahul Vimalkanth、Sachin S 和 Sudharshan Suresh。2025。利用智能体人工智能和多重 HyDE 增强金融检索增强生成：一种知识检索与幻觉减少的新方法。载于《第十届金融科技与自然语言处理研讨会论文集》，第 19–32 页，中国苏州。计算语言学协会。

Akash Ghosh、Debayan Datta、Sriparna Saha 和 Chi-rag Agarwal。2025。语言模型多语言推理综述。载于计算语言学协会论文集：EMNLP 2025，第 8920–8936 页，中国苏州。计算语言学协会。

Lynn Greschner 和 Roman Klinger。2025。恐惧的猎鹰与愤怒的美洲驼：人类和大语言模型对论证的情绪类别标注。载于第五届数字人文自然语言处理国际会议会议论文集，第 628–646 页，美国阿尔伯克基。计算语言学协会。

Sonam Gupta、YatinNandwani、Asaf Yehudai、Dinesh Khandelwal、Dinesh Raghu 和 Sachindra Joshi。2025a。用于大语言模型中的泛化。载于《计算语言学协会论文集：NAACL 2025》，第 6240–6249 页，阿尔伯克基，新墨西哥州。计算语言学协会。

Taneesh Gupta、Shivam Shandilya、Xuchao Zhang、Rahul Madhavan、Supriyo Ghosh、Chetan Bansal、Huaxiu Yao 和 Saravan Rajmohan。2025b。CARMO：面向上下文的奖励建模动态标准生成。载于计算语言学协会论文集：ACL 2025，第 2202–2261 页，奥地利维也纳。计算语言学协会。

Akio Hayakawa、Nouran Khallaf、Horacio Saggion 和 Serge Sharoff。2025。UoL-UPF 在 TSAR 2025 共享任务中的应用：一种面向可读性控制文本简化的生成与选择方法。载于第四届文本简化、无障碍与可读性研讨会论文集（TSAR 2025），第 193–210 页，中国苏州。计算语言学协会。

Chenyuan He、Yuxiang Jia、Fei Gao、Senbin Zhu、Hongde Liu、Hongying Zan 和 Min Peng。2025a。面向对话中含类代码回复的四元组抽取及非观点检测的任务感知对比专家混合。载于2025年自然语言处理实证方法会议论文集，第 4144–4159 页，中国苏州。计算语言学协会。

Zoe Wanying He、Sean Trott 和 Meenakshi Khosla。2025b。透过文字洞察，以像素表达：视觉与语言模型之间的深度表征对齐。载于2025年自然语言处理实证方法会议论文集，第35645–35660页，中国苏州。计算语言学协会。

RushikeshHiray 和 Venelin Kovatchev。2025年。Story- branch - 从小说生成多媒体内容。载于《2025年计算语言学协会美洲分会会议：人类语言技术（系统演示）论文集》，第 485–493 页，阿尔伯克基，新墨西哥州。计算语言学协会。

Sonja Hohloch。2024年。通过直极限研究同宿Floer同调。预印本，arXiv:2402.12345。

Faye Holt、William Held 和 Diyi Yang。2024年。南亚英语使用者对语言技术故障的认知。载于《计算语言学协会论文集：ACL 2024》，第4067–4081 页，曼谷，泰国。计算语言学协会。

Eftekhar Hossain、Sanjeev Kumar Sinha、Naman Bansal、R. Alexander Knipper、Souvika Sarkar、John Salvador、Yash Mahajan、Sri Ram Pavan Ku- mar Guttikonda、Mousumi Akter、Md. Mahadi Has- san、Matthew Freestone、Matthew C. Williams Jr.、Dongji Feng 和 Santu Karmaker。2025。

大语言模型作为元评审员的助手：一项案例研究。载于2025年美洲分会计算语言学协会会议：人类语言技术（第1卷：长篇论文）会议论文集，第7763–7803页，阿尔伯克基，新墨西哥州。计算语言学协会。

Peng Hu、Sizhe Liu、Changjiang Gao、Xin Huang、Xue Han、Junlan Feng、Chao Deng 和 Shu- jian Huang。2025。大型语言模型是跨语言、无知识的推理者。载于《2025年计算语言学协会美洲分会年会：人类语言技术》（第1卷：长篇论文）会议论文集，第1525–1542页，美国新墨西哥州阿尔伯克基。计算语言学协会。

Yuxuan Huang、Yihang Chen、Haozheng Zhang、Kang Li、Huichi Zhou、方萌、LinyiYang、Xiaoguang Li、Lifeng Shang、Songcen Xu、Jianye Hao、Kun Shao 和 Jun Wang。2025。深度研究智能体：系统性考察与路线图。预印本，arXiv:2506.18096。

Ziyi Huang 和 Xia Cui。2025。PromotionGo 在 SemEval-2025 第11项任务中的表现：一个面向特征的短文本跨语言多情感检测框架。载于《第19届国际语义评测研讨会》（SemEval-2025）会议论文集，第2140–2148页，奥地利维也纳。计算语言学协会。

Angel Hsing-Chi Hwang、Q. Vera Liao、Su Lin Blod-gett、Alexandra Olteanu 和 Adam Trischler。2025。“80% 是我，20% 是人工智能”：探索与大语言模型协同写作的真实性。ACM 人机交互会议论文集，9(2)。

Ahmad Ibrahim Ismail、Bashirudeen Opeyemi Ibrahim、Olubayo Adekanmbi 和 Ife Adebara。2025。检索增强生成与本地语言相结合，以改善药物信息的获取与理解。载于《第六届非洲自然语言处理研讨会（AfricaNLP 2025）会议论文集》，第108–114页，奥地利维也纳。计算语言学协会。

Raunak Jain 和 Srinivasan Rengarajan。2025。BEA 2025 共享任务中的涌现智慧：从词汇理解到面向教学能力评估的反思性推理。载于《第20届自然语言处理创新应用于教育应用构建研讨会（BEA 2025）会议论文集》，第1108–1120页，奥地利维也纳。计算语言学协会。

Charlott Jakob、David Harbecke、Patrick Parschan、Pia Wenzel Neves 和 Vera Schmitt。2025 年。Pol- BiX：通过 X-phemisms 检测大语言模型在事实核查中的政治偏见。载于《计算语言学协会论文集：EMNLP 2025》，第 17192–17210 页，中国苏州。计算语言学协会。

Gunjan Jalori、Preetika Verma 和 Sercan O Arik。2025 年。FLAIRR-TS——通过迭代优化和检索预测时间序列大语言模型智能体。载于《计算语言学协会论文集：EMNLP 2025》，第 15427–15437 页，中国苏州。计算语言学协会。

Dongsuk Jang、Ziyao Shangguan、Kyle Tegtmeyer、Anurag Gupta、Jan T Czerminski、Sophie Chheang 和 Arman Cohan。2025 年。MedTutor：一种用于基于病例的医学教育的检索增强型大语言模型系统。载于《2025年自然语言处理实证方法会议论文集：系统演示》，第319–353页，中国苏州。计算语言学协会。

Ala Jararweh、Oladimeji Macaulay、David Arredondo、Yue Hu、Luis E Tafoya、Kushal Virupakshappa 和 Avinash Sahu。2025年。Protein2Text：将蛋白质序列转换为人类可理解文本的重采样机制。载于《计算语言学协会美洲分会2025年自然语言处理会议论文集：人类语言技术》（第3卷：工业赛道），第918–937页，美国新墨西哥州阿尔伯克基。计算语言学协会。

Sungho Jeon 和 Michael Strube。2025年。小型语言模型中的实体追踪：参数高效微调的基于注意力机制的研究。载于《第六届话语、语境与文档级推理计算方法研讨会（CODI 2025）会议论文集》，第42–53页，中国苏州。计算语言学协会。

Jiaming Ji、Donghai Hong、Borong Zhang、Boyuan Chen、Josef Dai、Boren Zheng、Tianyi Alex Qiu、Jiayi Zhou、Kaile Wang、Boxun Li、Sirui Han、Yike Guo 和 Yaodong Yang。2025。PKU-SafeRLHF：基于人类偏好实现大语言模型多层级安全对齐。载于《计算语言学协会第63届年会论文集》（第1卷：长篇论文），第31983–32016页，奥地利维也纳。计算语言学协会。

Jiazhou Ji 和 Xinru Lu。2025。ReFLAIR：通过结构化反思和奖励引导学习增强多模态推理。载于《EMNLP 2025：计算语言学协会发现论文集》，第25401–25413 页，中国苏州。计算语言学协会。

Feiran Jia、Tong Wu、Xin Qin 和 Anna Squicciarini。2025a。任务防护机制：通过强化任务对齐来防御大语言模型智能体中的间接提示注入。载于《计算语言学协会第63届年会论文集》（第1卷：长篇论文），第 29680–29697 页，奥地利维也纳。计算语言学协会。

Runsong Jia、Mengjia Wu、Ying Ding、Jie Lu 和 Yi Zhang。2025b。HetGCoT：用于学术问答的异构图增强型思维链大语言模型推理。收录于

计算语言学协会：EMNLP 2025，第 15950–15963 页，中国苏州。计算语言学协会。

JinJiang, JianingWang, YuchenYan, YangLiu, Jianhua Zhu, Mengdi Zhang, and Liangcai Gao。2025年。大语言模型能否在形式语言的复杂逻辑推理中表现出色？载于2025年自然语言处理实证方法会议论文集，第16889–16914页，中国苏州。计算语言学协会。

Lesheng Jin, Zhenyuan Ruan, Haohui Mai, and Jingbo Shang。2025年。VeriLocc：通过大语言模型实现端到端跨架构寄存器分配。载于2025年自然语言处理实证方法会议论文集，第30240–30250页，中国苏州。计算语言学协会。

Yiqiao Jin, Qinlin Zhao, Yiyang Wang, Hao Chen, Kai- jie Zhu, Yijia Xiao, and Jindong Wang。2024年。Agen- tReview：探索大语言模型智能体的同行评审动态。载于2024年自然语言处理经验方法会议论文集，第1208–1226页，美国佛罗里达州迈阿密。计算语言学协会。

Aniket Joshi、Cyrus Andre DSouza、Sejal Jain、Jitenku- mar Babubhai Rana 和 Promod Yenigalla。2025a。I-SEE：一种针对产品内容的指令微调、经 SOP 增强的质量评估器。载于《2025年自然语言处理实证方法会议工业赛道论文集》，第 1379–1388 页，中国苏州。计算语言学协会。

Nisheeth Joshi、Palak Arora、Anju Krishnia、Riya Lonchenpa 和 Mhasilenuo Vizo。2025b。BVSLP：在 IndicMT 2025 共享任务中使用语言润饰进行机器翻译。载于《第十届机器翻译会议论文集》，第 1265–1270 页，中国苏州。计算语言学协会。

Srihari V K、Vijay Karthick Vaidyanathan 和 Thenmozhi Durairaj。2025。自然语言处理_goats@DravidianLangTech 2025：达罗毗荼语虚假新闻检测：一种文本分类方法。载于《第五届达罗毗荼语言语音、视觉与语言技术研讨会论文集》，第 345–349 页，阿科马、阿尔伯克基会议中心、阿尔伯克基、新墨西哥州。计算语言学协会。

Mukund K. Roy、Karunesh Arora、Praveen KumarChan- daliya、Rohit Kumar 和 Pruthwik Mishra。2025。CDAC-SVNIT 提交的 IWSLT 2025 印地语共享任务论文。载于第二十二届国际口语翻译会议（IWSLT 2025）会议论文集，第 180–185 页，奥地利维也纳（线下和线上）。计算语言学协会。

Maithili Sanjay Kadam 和 Francis Ferraro。2025。TAG–EQA：通过结构化提示策略进行事件问答的文本与图结构方法。载于第十四届词汇与计算语义联合会议（*SEM 2025）会议论文集，第 304–315 页，中国苏州。计算语言学协会。

Xiaoqiang Kang、Shengen Wu、Zimu Wang、Yilin Liu、Xiaobo Jin、Kaizhu Huang、Wei Wang、Yutao Yue、Xiaowei Huang 和 Qiufeng Wang。2025。GRPO 能提升对复杂多模态表格的理解能力吗？载于《2025年自然语言处理实证方法会议论文集》，第 12642–12655 页，苏州，中国。计算语言学协会。

Aly M. Kassem、Zhuan Shi、Negar Rostamzadeh 和 Golnoosh Farnadi。2025。重现你的记忆：通过稀疏模型差分预测大语言模型取消学习和微调的副作用。载于《2025年自然语言处理实证方法会议论文集》，第 32238–32251 页，苏州，中国。计算语言学协会。

Olga Kellert、Nemika Tyagi、Muhammad Imran、Nelvin Licona-Guevara 和 Carlos Gómez-Rodríguez。2025。解析代码切换：基于大语言模型的复杂代码切换及低资源语言通用依存标注。载于《计算语言学协会论文集：EMNLP 2025》，第 15934–15949 页，苏州，中国。计算语言学协会。

Ashinee Kesanam, GummuluriVenkata Ravi Ram,ChaithanyaSwaroopBanoth,andG RamaMohanaReddy. 2025. NITK-VITAL 在 SemEval-2025 第11项任务中的表现：Focal-RoBERTa：应对多标签情绪分类中的类别不平衡。载于第19届国际语义评测研讨会（SemEval-2025）会议论文集，第1077–1081页，奥地利维也纳。计算语言学协会。

ElhamKhabiri, Jeffrey O. Kephart, Fenno F. Heath Iii,Srideepika Jayaraman, Yingjie Li, Fateh A.Tipu,Dhruv Shah, Achille Fokoue, and AnuBhamidi-paty. 2025. 面向异构数据的自然语言查询声明式技术。载于2025年自然语言处理实证方法会议：产业分会论文集，第1744–1761页，中国苏州。计算语言学协会。

AnantKhandelwal、Manish Gupta 和 PuneetAgrawal。2025 年。CoCoA：用于解决大语言模型中知识冲突的置信度与上下文感知自适应解码。载于《2025年自然语言处理实证方法会议论文集》，第 6846–6866 页，中国苏州。计算语言学协会。

Maira Khatoon、Arooj Kiyani、Tehmina Farid 和 Sadaf Abdul Rauf。2025年。FJWU_Squad 在 SemEval-2025 任务1中的习语视觉理解数据集

用于习语学习。载于《第19届国际语义评测研讨会（SemEval-2025）会议论文集》，第1759–1765页，奥地利维也纳。计算语言学协会。

Md Tawkat Islam Khondaker、Numaan Naeem、Fa- timah Khan、AbdelRahim Elmadany 和 Muham- mad Abdul-Mageed。2024年。对 LLaMA-3 的阿拉伯语生成任务进行基准测试。载于第二届阿拉伯语自然语言处理会议论文集，第 283–297 页，泰国曼谷。计算语言学协会。

Ruslan Khrulev。2025。CHECK-MAT：探究视觉语言模型在手写解答上的数学推理能力及其与评分标准的对齐程度。载于第三届数学自然语言处理研讨会（MathNLP 2025）会议论文集，第 77–94 页，中国苏州。计算语言学协会。

Ahrii Kim。2025。上下文无处不在，却很少改变判断：重新审视文档级机器翻译评估。载于第十届机器翻译大会论文集，第 81–97 页，中国苏州。计算语言学协会。

Dongjun Kim、Gyuho Shim、Yongchan Chun、Min- hyuk Kim、Chanjun Park 和 Heuiseok Lim。2025a。基准测试剖析：大语言模型基准测试的机制诊断。载于《2025年自然语言处理实证方法会议论文集》，第 15646–15661 页，中国苏州。计算语言学协会。

Heehyeon Kim、Kyeongryul Lee 和 Joyce Jiyoung Whang。2025b。表象之下：通过自动生成的越狱提示探测大语言模型中的安全漏洞。载于《协会发现论文集：EMNLP 2025》，第 17668–17700 页，中国苏州。计算语言学协会。

Hyeongsik Kim、Yanheng Xu、Chaoqun Dong 和 Fei Du。2025c。过度生成与压缩：一种使用大语言模型进行程序性文本适配的提示策略。载于《协会发现论文集：EMNLP 2025》，第 19306–19337 页，中国苏州。计算语言学协会。

Jaeho Kim、Yunseok Lee 和 Seulki Lee。2025d。立场：人工智能会议同行评审危机亟需作者反馈和审稿人奖励。载于第42届国际机器学习会议论文集，《机器学习研究论文集》第267卷，第81634–81651页。PMLR。

Jongwoo Kim、SeongYeub Chu、Bryan Wong 和 Mun Yong Yi。2025e。并非所有选项都生而平等：面向令牌高效的大语言模型知识追踪的文本选项加权。载于《计算语言学协会论文集：EMNLP 2025》，第 16114–16128 页，中国苏州。计算语言学协会。

Juhyeong Kim、Sangyeon Yu、Gyunyeop Kim 和 Sangwoo Kang。2025f。FractalLLM：借助层嵌入式自我压缩实现无损自我推测解码。载于《计算语言学协会论文集：EMNLP 2025》，第 23666–23673 页，中国苏州。计算语言学协会。

Đor ̄de Klisura、Astrid R Bernaga Torres、Anna Karen Gárate-Escamilla、Rajesh Roshan Biswal、Ke Yang、Hilal Pataci 和 Anthony Rios。2025。用于缓解隐私政策问答系统中方言偏见的多智能体框架。载于计算语言学协会第63届年会论文集（第1卷：长篇论文），第32318–32337页，奥地利维也纳。计算语言学协会。

Shaghayegh Kolli、Richard Rosenbaum、Timo Cavelius、Lasse Strothe、Andrii Lata 和 Jana Diesner。2025。融合知识图谱、大语言模型和基于搜索的检索智能体的混合事实核查，可提升可解释的主张验证效果。载于第9届拓展自然语言处理研讨会论文集，第106–115页，中国苏州。计算语言学协会。

Olga Kononykhina、Anna-Carolina Haensch 和 Frauke Kreuter。2025。注意差距：职业嵌入中的性别差异。载于《第六届自然语言处理中的性别偏见研讨会（GeBNLP）论文集》，第83–91页，奥地利维也纳。计算语言学协会。

Yevheniia Kryklyvets、Mohammed Irfan Kurpath、Sa- hal Shaji Mullappilly、Jinxing Zhou、Fahad Shah- baz Khan、Rao Muhammad Anwer、Salman Khan 和 Hisham Cholakkal。2025。MAviS：面向鸟类物种的多模态对话式助手。载于《2025年自然语言处理实证方法会议论文集》，第28601–28627页，中国苏州。计算语言学协会。

RohitKumarand Chandan Nolbaria。2025。弥合金融情感领域的数据差距：大语言模型驱动的数据增强。载于《计算语言学协会第63届年会论文集》（第4卷：学生研究研讨会），第1246–1254页，奥地利维也纳。计算语言学协会。

Sandeep Kumar、Samarth Garg、Sagnik Sengupta、TirthankarGhosal 和 Asif Ekbal。2025 年。MixRevDe-检测：探索检测混合式同行评审中的人工智能生成内容。载于 2025 年计算语言学协会美洲分会会议论文集：人类语言技术（第2卷：短论文），第 944–953 页，新墨西哥州阿尔伯克基。计算语言学协会。

SumitKumar 和 Tanmay Basu。2025。AdaBioBERT：用于生物医学命名实体识别的自适应令牌序列学习。载于会议论文集

第24届生物医学语言处理研讨会，页56–62，维也纳，奥地利。计算语言学协会。

Beong-woo Kwak、Minju Kim、Dongha Lim、Hyungjoo Chae、Dongjin Kang、Sunghwan Kim、Dongil Yang 和 Jinyoung Yeo。2025。ToolHaystack：在真实的长期交互中对工具增强语言模型进行压力测试。载于《计算语言学协会发现论文集：EMNLP 2025》，第24696–24727 页，中国苏州。计算语言学协会。

Roman Kyslyi、Yuliia Maksymiuk 和 Ihor Pysmen- nyi。2025a。Vuyko mistral：针对低资源方言翻译调整大语言模型。载于《第四届乌克兰自然语言处理研讨会会议论文集》（UNLP 2025），第 86–95 页，奥地利维也纳（线上）。计算语言学协会。

Roman Kyslyi、Nataliia Romanyshyn 和 Volodymyr Sydorskyi。2025b。UNLP 2025 关于检测社交媒体操纵的共享任务。载于《第四届乌克兰自然语言处理研讨会会议论文集》（UNLP 2025），第 105–111 页，奥地利维也纳（线上）。计算语言学协会。

Ilya Lasy、Peter Knees 和 Stefan Woltran。2025。通过电路发现理解大语言模型中的逐字记忆。载于《首届大语言模型记忆研讨会（L2M2）会议论文集》，第83–94页，奥地利维也纳。计算语言学协会。

Sunbowen Lee、Junting Zhou、Chang Ao、Kaige Li、Xeron Du、Sirui He、Haihong Wu、Tianci Liu、Jia- heng Liu、Hamid Alinejad-Rokny、Min Yang、Yi- tao Liang、Zhoufutu Wen 和 Shiwen Ni。2025。大语言模型蒸馏的量化。载于《计算语言学协会第63届年会论文集》（第1卷：长篇论文），第4985–5004页，奥地利维也纳。计算语言学协会。

Haley Lepp 和 Daniel Scott Smith。2025。“你无法听起来像 GPT”：计算机科学出版中的语言歧视与抵抗迹象。载于《2025年ACM公平、问责与透明度会议》论文集，FAccT’25，第3162–3181页。ACM。

Vladimir I. Levenshtein。1965年。能够纠正删除、插入和反转的二进制码。《苏联物理学：报告》，10:707–710。

Amit LeVi、Rom Himelstein、Yaniv Nemcovsky、Avi Mendelson 和 Chaim Baskin。2025。将越狱攻击初始化用作服从方向的提取器。载于《计算语言学协会论文集：EMNLP 2025》，第6672–6705 页，中国苏州。计算语言学协会。

Dacheng Li、Shiyi Cao、Tyler Griggs、Shu Liu、Xi- angxi Mo、Eric Tang、Sumanth Hegde、Kourosh Hakhamaneshi、Shishir G Patil、Matei Zaharia、Joseph E. Gonzalez 和 Ion Stoica。2025a。语言模型可以轻松地从示例中学会推理。载于《计算语言学协会论文集：EMNLP 2025》，第 15979–15997 页，中国苏州。计算语言学协会。

DaweiLi、BohanJiang、Liangjie Huang、Alimohammad Beigi、Chengshuai Zhao、Zhen Tan、Amrita Bhat- tacharjee、Yuxuan Jiang、Canyu Chen、Tianhao Wu、Kai Shu、Lu Cheng 和 Huan Liu。2025b。从生成到评判：LLM评审的机遇与挑战。载于《2025年自然语言处理实证方法会议论文集》，第2757–2791页，中国苏州。计算语言学协会。

Jiazheng Li、Hanqi Yan 和 Yulan He。2025c。Drift：通过双重奖励概率推断提升LLM在理由生成中的忠实度。载于《计算语言学协会第63届年会论文集（第1卷：长篇论文）》，第6850–6866页，奥地利维也纳。计算语言学协会。

Matthew Li、Santiago Torres-Garcia、Shayan Halder、Phani Kuppa、Sean O’Brien、Vasu Sharma、Kevin Zhu 和 Sunishchal Dev。2025d。FrontierScience基准：评估大语言模型的人工智能研究能力。载于第一届智能体语言模型研究研讨会（REALM 2025）会议论文集，第 428–453 页，奥地利维也纳。计算语言学协会。

Muyao Li、Zihao Wang、Kaichen He、Xiaojian Ma 和 Yitao Liang。2025e。JARVIS-VLA：通过后训练使大型视觉语言模型能够使用键盘和鼠标玩视觉游戏。载于计算语言学协会发现论文集：ACL 2025，第17878–17899 页，奥地利维也纳。计算语言学协会。

Pingzhi Li、Prateek Yadav、Jaehong Yoon、Jie Peng、Yi-Lin Sung、Mohit Bansal 和 Tianlong Chen。2025f。Glider：全局与局部指令驱动专家路由器。载于2025年自然语言处理实证方法会议论文集，第 6251–6312 页，中国苏州。计算语言学协会。

Rui Li、Heming Xia、Xinfeng Yuan、Qingxiu Dong、Lei Sha、Wenjie Li 和 Zhifang Sui。2025g。大语言模型距离成为我们的数字孪生还有多远？基于角色设定的行为链模拟基准。载于《计算语言学协会论文集：ACL 2025》，第15738–15763页，奥地利维也纳。计算语言学协会。

Runchao Li、Yao Fu、Mu Sheng、XianxuanLong、Hao- tian Yu 和 Pan Li。2025h。FAEDKV：用于无偏 KV 缓存压缩的无限窗口傅里叶变换。载于协会发现论文集

计算语言学：EMNLP 2025，第 16856–16866 页，中国苏州。计算语言学协会。

Shenglan Li、Jia Xu 和 Mengjiao Zhang。2025年i。EmByte：面向小型且私密的自然语言处理的分解与压缩学习。载于计算语言学协会论文集：EMNLP 2025，7182–7201页，中国苏州。计算语言学协会。

Xirui Li、Ruochen Wang、Minhao Cheng、Tianyi Zhou 和 Cho-Jui Hsieh。2024年。DrAttack：提示词分解与重构使强大的大语言模型成为越狱攻击者。载于计算语言学协会论文集：EMNLP 2024，13891–13913页，美国佛罗里达州迈阿密。计算语言学协会。

Zehan Li、Ruhua Pan 和 Xinyu Pi。2025年j。超越大语言模型：一种从叙事文本生成因果图的语言学方法。载于第七届叙事理解研讨会会议论文集，36–51页，美国新墨西哥州阿尔伯克基。计算语言学协会。

Zichao Li. 2025a. 公式-文本交叉检索：面向数学信息检索的稠密嵌入方法基准测试研究. 收录于第三届数学自然语言处理研讨会（MathNLP 2025）会议论文集，第 124– 133 页，中国苏州。计算语言学协会。

Zichao Li. 2025b. 利用检索增强语言模型进行加密货币诈骗的知识驱动检测. 收录于第三届面向知识型基础模型研讨会（KnowFM）会议论文集，第40–48页，奥地利维也纳。计算语言学协会。

Zichao Li. 2025c. 利用表格时间序列数据进行检索增强预测. 收录于第四届表格表示学习研讨会会议论文集，第 192–199 页，奥地利维也纳。计算语言学协会。

ZichaoLiand ZongKe. 2025a. 面向低资源语言理解与生成的跨模态增强. 收录于第一届多模态检索增强生成研讨会（MAGMaR 2025）会议论文集，第90–99 页，奥地利维也纳。计算语言学协会。

Zichao Li 和 Zong Ke。2025b。领域与类型学相遇：利用 Universal Dependencies 预测金融与区块链自然语言处理中的动词末位语序。收录于第七届计算语言类型学研究与多语言自然语言处理研讨会论文集，第 156–164 页，奥地利维也纳。计算语言学协会。

Zichao Li、Zong Ke 和 Puning Zhao。2025k。通过图神经网络将结构化知识注入大语言模型。载于第一届大型语言模型与结构建模联合研讨会会议论文集（XLLM 2025），第 16–25 页，奥地利维也纳。计算语言学协会。

Emmy Liu、Amanda Bertsch、Lintang Sutawika、Lindia Tjuatja、Patrick Fernandes、Lara Marinov、Michael Chen、Shreya Singhal、Carolin Lawrence、Aditi Raghunathan、Kiril Gashteovski 和 Graham Neubig。2025a。不只是规模扩展定律：深入理解语言模型设计决策的下游影响。收录于2025年自然语言处理实证方法会议论文集，第16407–16438 页，中国苏州。计算语言学协会。

JiaruiLiu、YueqiSong、Yunze Xiao、Mingqian Zheng、Lindia Tjuatja、Jana Schaich Borg、Mona T. Diab 和 Maarten Sap。2025b。合成苏格拉底式辩论：检验角色效应对道德决策和说服动态的影响。载于《2025年自然语言处理实证方法会议论文集》，第16439–16469页，中国苏州。计算语言学协会。

Quancai Liu、Haihui Fan、Jinchao Zhang、Xiangfang Li、Chuanrong Li 和 Bo Li。2025c。DisComp：结合任务无关压缩与任务感知压缩的两阶段提示优化框架。载于《计算语言学协会论文集：NAACL 2025》，第1033–1044页，美国新墨西哥州阿尔伯克基。计算语言学协会。

Siyi Liu 和 Dan Roth。2025。文本冲突：数据、影响与挑战。载于《计算语言学协会论文集：EMNLP 2025》，第10073–10091页，中国苏州。计算语言学协会。

Wanlong Liu、Junying Chen、Ke Ji、Li Zhou、Wenyu Chen 和 Benyou Wang。2025d。RAG-instruct：通过多样化检索增强指令提升大语言模型。载于《2025年自然语言处理实证方法会议论文集》，第3865–3888页，中国苏州。计算语言学协会。

Yifan Liu、Qianfeng Wen、Mark Zhao、Jiazhou Liang 和Scott Sanner。2025e。MA-DPR：用于密集段落检索的流形感知距离度量。载于《2025年自然语言处理实证方法会议论文集》，第31073–31091页，中国苏州。计算语言学协会。

Zhiwei Liu、Jielin Qiu、Shiyu Wang、Jianguo Zhang、Zuxin Liu、Roshan Ram、Haolin Chen、Weiran Yao、Shelby Heinecke、Silvio Savarese、Huan Wang 和 Caiming Xiong。2025f。MCPEval：面向人工智能智能体模型的基于 MCP 的自动化深度评估。载于《2025年自然语言处理实证方法会议论文集：系统演示》，第373–402页，中国苏州。计算语言学协会。

ZhixiangLu、MianZhou、AngelosStefanidis 和 Jionglong Su。2025。UniMath-CoT：一种统一的

通过重新推理确认实现多模态数学推理的框架。载于《第3届数学自然语言处理研讨会（MathNLP 2025）会议论文集》，第 176–185 页，中国苏州。计算语言学协会。

Hanjun Luo、Yingbin Jin、Yiran Wang、Xinfeng Li、Tong Shang、Xuecheng Liu、Ruizhe Chen、Kun Wang、Hanan Salam、Qingsong Wen 和 Zuozhu Liu。2025a。DynamicNER：一个用于基于大语言模型的命名实体识别的动态、多语言、细粒度数据集。载于《2025年自然语言处理实证方法会议论文集》，第 16522–16546 页，中国苏州。计算语言学协会。

Xiangfeng Luo、Ruoxin Zheng、Jianqiang Huang 和 Hang Yu。2025b。HDiff：一种置信度引导的去噪扩散方法，用于稳健的超关系链接预测。载于《计算语言学协会论文集：EMNLP 2025》，第 7417–7434 页，中国苏州。计算语言学协会。

Lijia Lv、Yuanshu Zhao、Guan Wang、Xuehai Tang、Wen Jie、Jizhong Han 和 Songlin Hu。2025。Gamma-guard：用于大语言模型稳健防护栏的轻量级残差适配器。载于《2025年自然语言处理实证方法会议论文集》，第 12230–12242 页，中国苏州。计算语言学协会。

Jabez Magomere、Elena Kochkina、Samuel Mensah、Simerjot Kaur 和 Charese Smiley。2025年。FinNLI：用于多体裁金融自然语言推理基准测试的新型数据集。载于计算语言学协会论文集：NAACL 2025，第4545–4568页，新墨西哥州阿尔伯克基。计算语言学协会。

Arijit Maji、Raghvendra Kumar、Akash Ghosh、Anushka 和 Sriparna Saha。2025。SANSKRITI：用于评估语言模型对印度文化知识的综合基准。载于计算语言学协会论文集：ACL 2025，第4434–4451页，奥地利维也纳。计算语言学协会。

Vijit Malik、Akshay Jagatap、Vinayak S Puranik 和 Anirban Majumder。2024年。PEARL：利用大语言模型智能体进行偏好提取、示例增强和检索。载于2024年自然语言处理经验方法会议：工业赛道论文集，第1536–1547页，美国佛罗里达州迈阿密。计算语言学协会。

Stasa Mandic、Georg Niess 和 Roman Kern。2025 年。从分布内到分布外：通过联合损失提升软件提及和关系抽取的泛化能力。载于《第五届学术文档处理研讨会》（SDP 2025）会议论文集，第 146–153 页，奥地利维也纳。计算语言学协会。

Christopher D. Manning、Prabhakar Raghavan 和 Hinrich Schütze。2008 年。《信息检索导论》。剑桥大学出版社，英国剑桥。

Utsav Maskey、Chencheng Zhu 和 Usman Naseem。2025。用于密码分析和侧信道漏洞的大语言模型基准测试。载于《计算语言学协会论文集：EMNLP 2025》，第19849–19865页，苏州，中国。计算语言学协会。

Blake Matheny、Phuong Nguyen 和 Minh Nguyen。2025。JNLP 在 SemEval-2025 任务1中的表现：使用大语言模型进行多模态习语表达表征。载于《第19届国际语义评测研讨会论文集》（SemEval-2025），第1479–1484页，维也纳，奥地利。计算语言学协会。

Han Meng、Yancan Chen、Yunan Li、Yitian Yang、Jungup Lee、Renwen Zhang 和 Yi-Chieh Lee。2025。污名归因于什么？一个基于理论、由专家标注的访谈语料库，旨在揭示心理健康污名。载于《计算语言学协会第63届年会论文集》（第1卷：长篇论文），第5453–5490页，维也纳，奥地利。计算语言学协会。

Saif M. Mohammad。2025年。温暖之词：超过2.6万个英语词汇的信任与社交规范。载于计算语言学协会第63届年会论文集（第1卷：长篇论文），第18830–18850页，奥地利维也纳。计算语言学协会。

Seyedali Mohammadi、Bhaskara HanumaVedula、He- mank Lamba、Edward Raff、Ponnurangam Ku- maraguru、Francis Ferraro和Manas Gaur。2025年。大语言模型会遵循标签定义吗？检验它们对外部标签定义的接受度。载于2025年自然语言处理实证方法会议论文集，第32368–32381页，中国苏州。计算语言学协会。

Joydeb Mondal和Pramir Sarkar。2025年。Modgenix 在 SemEval-2025 任务1中的工作：用于多模态习语理解的上下文感知视觉语言排序（CAViLR）。载于第19届国际语义评测研讨会（SemEval-2025）论文集，第780–784页，奥地利维也纳。计算语言学协会。

Hoyeon Moon、Byeolhee Kim 和 Nikhil Verma。2025 年。多语言 RAG 系统中的质量感知翻译标注。载于《第五届多语言表示学习研讨会》（MRL2025）会议论文集，第 161–177 页，中国苏州。计算语言学协会。

Guy Mor-Lan、Naama Rivlin-Angert、Yael R. Kaplan、Tamir Sheafer 和Shaul R. Shenhav。2025年。HebID：

检测希伯来语政治文本中的社会身份。载于《计算语言学协会论文集：EMNLP 2025》，第9850–9870页，中国苏州。计算语言学协会。

Andreas C Mueller。2023年。词云。

Gagan Mundada、Yash Vishe、Amit Namburi、Xin Xu、Zachary Novack、Julian McAuley 和 Junda Wu。2025。WildScore：对野外符号音乐推理中的多模态大语言模型进行基准测试。载于《2025年自然语言处理实证方法会议论文集》，第16858–16874页，中国苏州。计算语言学协会。

Nandini Mundra、Aditya Nanda Kishore Khandavally、Raj Dabre、Ratish Puduppully、Anoop Kunchukuttan 和 Mitesh M Khapra。2024年。语言模型词汇扩展与初始化方法的实证比较。载于《第28届计算自然语言学习会议论文集》，第84–104页，美国佛罗里达州迈阿密。计算语言学协会。

Tafazzul Nadeem、Riyansha Singh、Suyamoon Pathak 和 Ashutosh Modi。2025。IITK 探索实验室在 SemEval-2025 任务11中的工作：弥合基于文本的情感检测差距。载于《第19届国际语义评测研讨会（SemEval-2025）论文集》，第1859–1865页，奥地利维也纳。计算语言学协会。

Mahmud Wasif Nafee、Maiqi Jiang、Haipeng Chen 和 Yanfu Zhang。2025。通过策略优化实现上下文内知识编辑的动态检索器。载于《2025年自然语言处理实证方法会议论文集》，页16755–16768，中国苏州。计算语言学协会。

Shahriar Kabir Nahin、Rabindra Nath Nandi、Sagor Sarker、Quazi Sarwar Muhtaseem、Md Kowsher、Apu Chandraw Shill、Md Ibrahim、Mehadi Hasan Menon、Tareq Al Muntasir 和 Firoj Alam。2025。TituLLMs：全面基准测试的孟加拉语大语言模型系列。载于《计算语言学协会论文集：ACL 2025》，页24922–24940，奥地利维也纳。计算语言学协会。

Kamel Nebhi、Amrita Panesar 和 Hans Bantilan。2025。使用大语言模型进行自适应英语写作评估的端到端自动化题目生成与评分。载于《第20届自然语言处理创新应用于教育应用构建研讨会》（BEA 2025）会议论文集，第968–977页，奥地利维也纳。计算语言学协会。

Bao Nguyen、Binh Nguyen、Duy Nguyen 和 Viet Anh Nguyen。2025a。语言模型激活的分布式手术。载于《协会发现论文集：计算语言学：EMNLP 2025》，页8192–8212，中国苏州。计算语言学协会。

Viet Cuong Nguyen、Mohammad Taher、Dongwan Hong、ViniciusKonkolics Possobom、Vibha Thirunel- layi Gopalakrishnan、Ekta Raj、Zihang Li、Heather J. Soled、Michael L. Birnbaum、Srijan Kumar 和 Munmun De Choudhury。2025b。大语言模型是否与核心心理健康咨询能力相一致？载于《计算语言学协会论文集：NAACL 2025》，第7488–7511页，美国新墨西哥州阿尔伯克基。计算语言学协会。

Xuanfan Ni 和 Piji Li。2024a。针对自然语言的大语言模型的系统评估。arXiv，abs/2405.10251。注：引自 Semantic Scholar。

Xuanfan Ni 和 Piji Li。2024b。针对自然语言生成任务的大语言模型的系统评估。预印本，arXiv:2405.10251。注：引自 arXiv。

Natapong Nitarach、Warit Sirichotedumrong、Panop Pitchayarthorn、Pittawat Taveekitworachai、Potsawee Manakul 和 Kunat Pipatanakul。2025。FinCoT：以专家金融推理为基础的思维链。载于《第十届金融科技与自然语言处理研讨会论文集》，第93–123页，中国苏州。计算语言学协会。

Syeda Alisha Noor, Sadia Anjum, Syed Ahmad Reza 和 Md Rashadur Rahman。2025。Celestia@DravidianLangTech 2025：面向达罗毗荼语假新闻检测的 Malayalam-BERT 和基于 m-BERT 的 Transformer 模型。载于第五届达罗毗荼语言语音、视觉与语言技术研讨会会议论文集，第688–693页，阿科马，阿尔伯克基会议中心，新墨西哥州阿尔伯克基。计算语言学协会。

Jean De Dieu Nyandwi、Yueqi Song、Simran Khanuja 和 Graham Neubig。2025。利用文化知识为多语言多模态大语言模型提供依据。载于2025年自然语言处理实证方法会议论文集，第24198–24242页，中国苏州。计算语言学协会。

Samarth P 和 Sanjay Balaji Mahalingam。2025。Gemma格言：微调 Gemma 3 以进行梵语连声切分。载于第九届拓展自然语言处理研讨会会议论文集，第235–241页，中国苏州。计算语言学协会。

Manas Pandya、Avinash Kumar Sharma 和 Arpit Shukla。2025 年。斯瓦希里语新闻分类：机器学习、深度学习和 Transformer 模型中的性能、挑战与可解释性。载于《第六届非洲自然语言处理研讨会》（AfricaNLP 2025）会议论文集，第 203–209 页，奥地利维也纳。计算语言学协会。

Lei Pang、Hanyi Mao、Quanjia Xiao、Chen Ruihan、Jingjun Zhang、Haixiao Liu 和 Xiangyi Li。2025 年。In2X 在 WMT25 翻译任务中的表现。载于《第十届机器翻译大会论文集》，第 671–679 页，中国苏州。计算语言学协会。

Angelina Parfenova 和 Jürgen Pfeffer。2025。衡量重要事项：通过标签细化评估归纳式编码中的集成大语言模型。载于《计算语言学协会论文集：ACL 2025》，第 10803–10816 页，奥地利维也纳。计算语言学协会。

Seongwan Park、Taeklim Kim 和 Youngjoong Ko。2025。解码稠密嵌入：用于解释和离散化稠密检索的稀疏自编码器。载于《2025 年自然语言处理经验方法会议论文集》，第 26479–26496 页，中国苏州。计算语言学协会。

Priyaranjan Pattnayak、Amit Agarwal、Hansa Meghwani、Hitesh Laxmichand Patel 和 Srikant Panda。2025。用于响应式多轮在线对话的混合人工智能：新型动态路由与反馈适配。载于《第四届自然语言处理知识增强方法国际研讨会会议论文集》，第 215–229 页，美国新墨西哥州阿尔伯克基。计算语言学协会。

Nguyen Pham Hoang Le、An Dinh Thien、Son T. Luu 和 Kiet Van Nguyen。2025。DocIE@XLLM25：ZeroSemble——利用异构大语言模型集成进行稳健高效的零样本文档信息抽取。载于第一届大语言模型与结构建模联合研讨会（XLLM 2025）会议论文集，第288–297页，奥地利维也纳。计算语言学协会。

Pakawat Phasook、Rapepong Pitijaroonpong、Jir amet Kinchagawat、Amrest Chinkamol、Tossaporn Saengja、Kiartnarin Udomlapsakul、Jitkapat Sawat-phol 和 Piyalitt Ittichaiwong。2025。VeReaFine：面向开放式临床问答的抗幻觉迭代式验证推理优化检索增强生成。载于第24届生物医学语言处理研讨会（共享任务）会议论文集，第281–288页，奥地利维也纳。计算语言学协会。

Renjie Pi, Haoping Bai, QibinChen, Xiaoming Si-mon Wang,JiulongShan, Xiaojiang Liu,and MengCao. 2025. MR. judge：作为裁判的多模态推理器。载于《2025年自然语言处理实证方法会议论文集》，第20192–20216页，中国苏州。计算语言学协会。

Jason Priem、Heather Piwowar 和 Richard Orr。2022年。OpenAlex：一个完全开放的学术成果、作者、发表场所、机构和概念索引。预印本，arXiv:2205.01833。

Emiliana Pulido、Robert Pugh 和 Zoey Liu。2025。我代表这些树木发声：为西班牙语二语学习者和传承语使用者开发依存树库。载于《计算语言学协会第63届年会论文集》（第4卷：学生研究研讨会），第814–822页，奥地利维也纳。计算语言学协会。

SaifPunjwaniandLarry Heck.2025. 思维权重推理：探索神经网络权重以增强大语言模型推理。载于《第一届智能体语言模型研究研讨会（REALM 2025）论文集》，第471–485页，奥地利维也纳。计算语言学协会。

Junlang Qian, ZixiaoZhu, HanzhangZhou, ZijianFeng, ZepengZhai, and Kezhi Mao. 2025. 超越下一个词元：通过高效的多词元预测实现提示词鲁棒的零样本分类。载于《2025年计算语言学协会美洲分会会议：人类语言技术》（第1卷：长篇论文），第7093–7115页，美国新墨西哥州阿尔伯克基。计算语言学协会。

NeelPrabhanjanRachamalla、AravindKonakalla、Gau-tam Rajeev、Ashish Kulkarni、Chandra Khatri 和 Shubham Agarwal。2025 年。Pragyaan：为印度语言设计和整理高质量文化后训练数据集。载于《第五届多语言表征学习研讨会（MRL 2025）会议论文集》，第 285–321 页，中国苏州。计算语言学协会。

Javad Rafiei Asl、Sidhant Narula、Mohammad Ghasemigol、Eduardo Blanco 和 Daniel Takabi。2025 年。NEXUS：通过网络探索利用多轮大语言模型越狱中的不安全序列。载于《2025年自然语言处理实证方法会议论文集》，第 24278–24306 页，中国苏州。计算语言学协会。

S M Rafiuddin 和 Muntaha Nujat Khan。2025 年。学习记忆内容：面向记忆高效语言模型的自适应概率记忆保留。载于《计算语言学协会论文集：EMNLP 2025》，第 3969–3981 页，中国苏州。计算语言学协会。

Shahid Iqbal Rai、Danilo Croce 和 Roberto Basili。2025 年。通过基于提示的微调将框架语义学注入大型语言模型。载于《第十四届词汇与计算语义联合会议（*SEM 2025）会议论文集》，第 31–47 页，中国苏州。计算语言学协会。

Nishat Raihan、Joanna C. S. Santos 和 Marcos Zampieri。2025。MojoBench：Mojo 语言建模与基准。收录于《计算语言学协会发现论文集：NAACL》

2025，4109–4128页，墨西哥新墨西哥州阿尔伯克基。计算语言学协会。

Rahul Raja 和 Arpita Vats。2025 年。低资源印度语言机器翻译的平行语料库：全面评审。载于《第八届低资源语言机器翻译技术研讨会》（LoResMT 2025）会议论文集，第 129–143 页，美国新墨西哥州阿尔伯克基。计算语言学协会。

Alex Ray、Josh Achiam 和 Dario Amodei。2019 年。深度强化学习中的基准测试与安全探索。

Saed Rezayi、Le An Ha、Yiyun Zhou、Andrew Houriet、AngeloD'Addario、PeterBaldwin、Polina Harik、Ann King 和 Victoria Yaneva。2025 年。医患互动中沟通技能的自动评分：兼顾性能与可扩展性。载于《第20届自然语言处理创新应用于教育应用构建研讨会》（BEA 2025）会议论文集，第 891–897 页，奥地利维也纳。计算语言学协会。

Naama Rivlin-Angert 和 Guy Mor-Lan。2025。内敌：以色列政治演讲中的政治合法性否定话语研究。载于《2025年自然语言处理实证方法会议论文集》，第16645–16658页，中国苏州。计算语言学协会。

Keonwoo Roh、Yeong-Joon Ju 和 Seong-Whan Lee。2025。XLQA：一个地区感知多语言开放域问答基准。载于《2025年自然语言处理实证方法会议论文集》，第28797–28809页，中国苏州。计算语言学协会。

Pablo Romero、Zihao Li 和 Matthew Shardlow。2025。利用合成数据微调，在 Firefox 上高效实现设备端文本简化。载于《第四届文本简化、无障碍与可读性研讨会论文集》（TSAR 2025），第105–115页，中国苏州。计算语言学协会。

Kalaivani K S、Sanjay R、ThissyakkannaS M 和 Niren- jhanram S K。2025a。KSK@DravidianLangTech 2025：使用增量学习对泰米尔语 X（推特）评论进行政治多类别情感分析。载于第五届达罗毗荼语言语音、视觉与语言技术研讨会会议论文集，第 221–225 页，阿科马，阿尔伯克基会议中心，阿尔伯克基，新墨西哥州。计算语言学协会。

Vishal S、Rajalakshmi Sivanaiah 和 Angel Deborah S。2025b。SemEval-2025 任务1中的 TechSSN3：使用集成 Transformer 模型和词汇规则进行多标签情绪检测。载于第19届国际语义评测研讨会（SemEval-2025）会议论文集，第186–192页，奥地利维也纳。计算语言学协会。

Jon Saad-Falcon、Rajan Pathe Vivek、William Berrios、Nandita Shankar Naik、Matija Franklin、Bertie Vid- gen、Amanpreet Singh、Douwe Kiela 和 Shikib Mehri。2025。LMUNIT：使用自然语言单元测试进行细粒度评估。载于计算语言学协会论文集：EMNLP 2025，Findings，第3303–3324页，中国苏州。计算语言学协会。

Syeda Jannatus Saba 和 Steven Skiena。2025。通过玩传话游戏评估语言翻译模型。载于 2025 年自然语言处理实证方法会议会议论文集，第 10332–10347 页，中国苏州。计算语言学协会。

Mushtari Sadia、Zhenning Yang、Yunming Xiao、Ang Chen 和 Amrita Roy Chowdhury。2025。SQUiD：从非结构化文本合成关系数据库。载于 2025 年自然语言处理实证方法会议会议论文集，第 31975–32000 页，中国苏州。计算语言学协会。

Muhammed Saeed、Shaina Raza、Ashmal Vayani、MuhammadAbdul-Mageed、Ali Emami 和 Shady Shehata。2025年。超越内容：语法性别如何塑造文本到图像模型中的视觉表征。载于《计算语言学协会论文集：EMNLP 2025》，第24673–24695页，中国苏州。计算语言学协会。

Furkan ,Sahinuç、Ilia Kuznetsov、Yufang Hou 和 Iryna Gurevych。2024a。利用大语言模型进行系统性任务探索：引文文本生成研究。载于《第62届计算语言学协会年会论文集》（第1卷：长篇论文），第4832–4855页，泰国曼谷。计算语言学协会。

Furkan ,Sahinuç、Thy Thy Tran、Yulia Grishina、Yufang Hou、Bei Chen 和 Iryna Gurevych。2024b。高效的性能跟踪：利用大语言模型自动构建科学排行榜。载于《2024年自然语言处理实证方法会议论文集》，第7963–7977页，美国佛罗里达州迈阿密。计算语言学协会。

Jonathan Sakunkoo 和 Annabella Sakunkoo。2025。失而复得：维基词典中形态缺陷众包知识的计算质量保证。载于计算语言学协会第63届年会论文集（第4卷：学生研究研讨会），第998–1003页，维也纳，奥地利。计算语言学协会。

Msvpj Sathvik、Raj Sonani 和 Ravi Teja Potla。2025a。卡纳塔克邦选举期间宗教仇恨言论的检测。载于第五届达罗毗荼语言语音、视觉与语言技术研讨会会议论文集，第 562–566 页，美国

新墨西哥州阿尔伯克基阿尔伯克基会议中心。计算语言学协会。

Sai Sathvik、Muralidhar Palli、Keerthana NNL、Balasubramanian Palani、Jobin Jose 和 Siranjeevi Rajamanickam。2025b。Team-risers@DravidianLangTech 2025：使用基于 Transformer 的嵌入检测达罗毗荼语中的人工智能生成产品评论。载于《第五届达罗毗荼语言语音、视觉与语言技术研讨会会议论文集》，第33–37页，阿科马，阿尔伯克基会议中心，新墨西哥州阿尔伯克基。计算语言学协会。

Kevin Scaria、Himanshu Gupta、Siddharth Goyal、Saurabh Sawant、Swaroop Mishra 和 Chitta Baral。2024年。InstructABSA：用于方面级情感分析的指令学习。载于《2024年计算语言学协会北美分会年会：人类语言技术》（第2卷：短论文）会议论文集，第720–736页，墨西哥城，墨西哥。计算语言学协会。

Miriam Schirmer、Angelina Voggenreiter、Juergen Pf- effer 和 Agnes Horvat。2025年。在社交媒体上检测儿童物化：语言建模的挑战。载于第九届在线虐待与伤害研讨会（WOAH）会议论文集，第396–412页，奥地利维也纳。计算语言学协会。

Aishwarya Selvamurugan。2025。DravLingua@DravidianLangTech 2025：通过 muril 和 Wav2Vec 模型的后期融合，对达罗毗荼语进行多模态仇恨言论检测。载于第五届达罗毗荼语言语音、视觉与语言技术研讨会会议论文集，第694–699页，阿科马，新墨西哥州阿尔伯克基会议中心，美国新墨西哥州阿尔伯克基。计算语言学协会。

Uzair Shah、Md. Rafiul Biswas、Marco Agus、Mowafa Househ 和 Wajdi Zaghouani。2024年。MemeMind 参加 ArAIEval 共享任务：利用先进的语言与视觉模型，通过生成式增强和特征融合检测阿拉伯语表情包中的多模态宣传内容。载于第二届阿拉伯语自然语言处理会议会议论文集，第467–472页，泰国曼谷。计算语言学协会。

Naim Shandi、JasonM. Merlo 和 Jeffrey A. Nanzer。2024年。分布式无线系统的去中心化皮秒同步。预印本，arXiv:2405.18384。

Chenyang Shao、Dehao Huang、Yu Li、Keyu Zhao、Weiquan Lin、Yining Zhang、Qingbin Zeng、Zhiyu Chen、Tianxing Li、Yifei Huang、Taozhong Wu、Xinyang Liu、Ruotong Zhao、Mengsheng Zhao、Jiaoyang Li、Xuhua Zhang、Yue Wang、Yuanyi Zhen、Fengli Xu 等共 2 人。2025 年。Omniscientist：迈向人类与人工智能科学家的协同演进生态系统。预印本，arXiv:2511.16931。

Jianshu She、Wenhao Zheng、Zhengzhong Liu、Hongyi Wang、Eric P. Xing、Huaxiu Yao 和 Qirong Ho。2025 年。面向边缘设备的令牌级路由推理系统。载于计算语言学协会第63届年会论文集（第 3 卷：系统演示），第 159–166 页，奥地利维也纳。计算语言学协会。

Rajvee Sheth、Shubh Nisar、Heenaben Prajapati、Himan- shu Beniwal 和 Mayank Singh。2024年。Commenta- tor：一个代码混合多语言文本标注框架。载于《2024年自然语言处理实证方法会议论文集：系统演示》，第101–109页，迈阿密，佛罗里达州，美国。计算语言学协会。

Zhengliang Shi、Yiqun Chen、Haitao Li、Weiwei Sun、Shiyu Ni、Yougang Lyu、Run-Ze Fan、Bowen Jin、Yixuan Weng、Minjun Zhu、Qiujie Xie、Xinyu Guo、Qu Yang、Jiayi Wu、Jujia Zhao、Xiaqiang Tang、Xinbei Ma、Cunxiang Wang、Jiaxin Mao等另外7人。2025。深度研究：系统综述。预印本，arXiv:2512.02038。

Ibne Farabi Shihab、Sanjeda Akter 和 Anuj Sharma。2025。面向资源受限环境的 Mamba 状态空间模型高效非结构化剪枝。载于《2025年自然语言处理实证方法会议论文集》，第11109–11137页，中国苏州。计算语言学协会。

Siddharth Shukla、Tanuj Tyagi、Abhay Singh Bisht、Ashish Sharma 和 Basant Agarwal。2025。NyayGraph：利用大型语言模型和知识图谱增强的印度法律法令识别方法。载于《2025年自然法律语言处理研讨会论文集》，第147–156页，中国苏州。计算语言学协会。

Shikhhar Siingh、Abhinav Rawat、Chitta Baral 和 Vivek Gupta。2025。GETReason：通过层次化多智能体推理增强图像上下文提取。载于《计算语言学协会第63届年会论文集》（第1卷：长篇论文），第29779–29800页，奥地利维也纳。计算语言学协会。

Divyansh Singh、Brodie Mather、Demi Zhang、Patrick Lehman、Justin Ho 和 Bonnie J Dorr。2025 年。FIDELITY：用于有效语言洞见和主题生成的细粒度可解释蒸馏。载于《计算语言学协会论文集：NAACL 2025》，第 2460–2472 页，美国新墨西哥州阿尔伯克基。计算语言学协会。

Charlotte Siska、Katerina Marazopoulou、Melissa Ailem 和 James Bono。2024年。检验大语言模型评估对基准分布假设的鲁棒性。载于会议论文集

《第62届计算语言学协会年会论文集》（第1卷：长篇论文），第10406–10421页，泰国曼谷。计算语言学协会。

Konstantinos Skianis，A. SezaDo˘gruöz 和 John Pavlopoulos。2024年。利用大语言模型翻译和分类心理健康数据。载于《第四届多语言表征学习研讨会（MRL 2024）会议论文集》，第 236–241 页，美国佛罗里达州迈阿密。计算语言学协会。

Yejin Son、Saejin Kim、Dongjun Min 和 Youngjae Yu。2025。多模态反常识：从奇异到寻常，再从寻常到奇异。载于《计算语言学协会论文集：EMNLP 2025》，第 17586–17609 页，中国苏州。计算语言学协会。

Woomin Song、Saket Dingliwal、Sai Muralidhar Jayan-thi、Bhavana Ganesh、Jinwoo Shin、Aram Galstyan 和 Sravan Babu Bodapati。2025a。通过无模型推测性采样加速测试时扩展。载于《2025年自然语言处理实证方法会议论文集》，第 30611–30624 页，中国苏州。计算语言学协会。

Xiaoying Song、Sharon Lisseth Perez、Xinchen Yu、Ed- uardo Blanco 和 Lingzi Hong。2025b。Discord 的回响：预测仇恨者对反仇恨言论的反应。载于计算语言学协会论文集：NAACL 2025，第 4892–4905 页，阿尔布开克，新墨西哥州。计算语言学协会。

Shashank Sonkar、Naiming Liu 和 Richard Baraniuk。2024年。学生数据悖论与单学生-导师模型的奇特案例：为个性化学习训练大语言模型的回归性副作用。载于计算语言学协会论文集：EMNLP 2024，第 15543–15553 页，美国佛罗里达州迈阿密。计算语言学协会。

Mohsen Sorkhpour、Abbas Yazdinejad 和 Ali De- hghantanha。2025。RedHit：通过搜索、推理和偏好优化对大语言模型进行自适应红队测试。载于首届大语言模型安全研讨会（LLMSEC）会议论文集，第 7–16 页，奥地利维也纳。计算语言学协会。

Alexander Spangher、Tenghao Huang、Yiqin Huang、Lucas Spangher、Sewon Min 和 Mark Dredze。2025a。一种新颖的多文档检索基准：新闻写作中的记者来源选择。载于《第四届自然语言处理知识增强方法国际研讨会论文集》，第 180–204 页，阿尔伯克基，新墨西哥州，美国。计算语言学协会。

Alexander Spangher、Michael Lu、SriyaKalyan、Hyun-dong Justin Cho、Tenghao Huang、Weiyan Shi，以及 Jonathan May。2025b。NewsInterview：一个用于通过信息访谈评估大语言模型落地差距的数据集和试验场。载于《计算语言学协会第63届年会论文集》（第1卷：长篇论文），第32895–32925页，奥地利维也纳。计算语言学协会。

Shashank Srivastava。2025。大型语言模型威胁语言的认知与交际基础。载于《2025年自然语言处理实证方法会议论文集》，第 28650–28664 页，苏州，中国。计算语言学协会。

Giulio Starace、Oliver Jaffe、Dane Sherburn、James Aung、Jun Shern Chan、Leon Maksin、Rachel Dias、Evan Mays、Benjamin Kinsella、Wyatt Thompson、Johannes Heidecke、Amelia Glaese 和 Tejal Patwardhan。2025。PaperBench：评估人工智能的复现人工智能研究的能力。载于《第42届国际机器学习会议论文集》，第 267 卷《机器学习研究会议论文集》，第 56843–56873 页。PMLR。

Jie Sun, JunkangWu, Jiancan Wu,Zhibo Zhu, Xingyu Lu, Jun Zhou, Lintao Ma, 和 Xiang Wang。2025年。通过动态目标边际实现鲁棒偏好优化。载于《计算语言学协会论文集：ACL 2025》，第 5399–5416 页，奥地利维也纳。计算语言学协会。

Liangtai Sun, Yang Han, Zihan Zhao, Da Ma, Zhen- nan Shen, Baocai Chen, Lu Chen, 和 Kai Yu。2024年。Scieval：面向科学研究的多级大语言模型评估基准。《AAAI人工智能会议论文集》，38(17)：19053–19061。

Gokul Swamy, Aman Gulati, Srinivas Virinchi, 和 Anoop Saladi。2025年。面向电子商务配送的地址智能框架。载于《2025年自然语言处理实证方法会议：工业赛道会议论文集》，第 1026–1034 页，中国苏州。计算语言学协会。

Jiabin Tang, Lianghao Xia, Zhonghang Li 和 Chao Huang。2025a。AI-researcher：自主科学创新。预印本，arXiv:2505.18705。

Zuojin Tang、Bin Hu、Chenyang Zhao、De Ma、Gang Pan 和 Bin Liu。2025b。VLASCD：一种用于同时聊天和决策的视觉语言动作模型。载于《2025年自然语言处理实证方法会议论文集》，第 9223–9243 页，中国苏州。计算语言学协会。

Mehrab Tanjim、Yeonjun In、Xiang Chen、Victor Bursz- tyn、Ryan A.Rossi、Sungchul Kim、Guang-Jie Ren、Vaishnavi Muppala、Shun Jiang、Yongsung Kim 和 Chanyoung Park。2025 年。大语言模型和智能体时代的对话式问答消歧

综述。载于 2025年自然语言处理实证方法会议论文集，第 9548–9561 页，中国苏州。计算语言学协会。

Chong Tian、Qirong Ho 和 Xiuying Chen。2025。一种用于不断演进的虚假新闻生成与检测的符号对抗学习框架。载于《2025年自然语言处理实证方法会议论文集》，第12307–12321页，中国苏州。计算语言学协会。

Manyana Tiwari。2025。扩展摘要：通过探针引导的参数高效微调，在基于大语言模型的社会影响系统中平衡语言适应与安全性。载于《第三届对话社会影响研讨会（SICon 2025）论文集》，第145–147页，奥地利维也纳。计算语言学协会。

MeiHan Tong 和 Shuai Wang。2025。NovelCR：面向长跨度共指消解的大规模双语数据集。载于《计算语言学协会论文集：ACL 2025》，第5161–5173页，奥地利维也纳。计算语言学协会。

Octavian Alexandru Trifan、Jason Lee Weber、Marc Ti- tus Trifan、Alexandru Nicolau 和 Alexander Vei- denbaum。2025 年。语法剪枝：为边缘人工智能启用低延迟零样本面向任务的语言模型。载于《2025年自然语言处理实证方法会议论文集》，第 16955–16968 页，中国苏州。计算语言学协会。

Vijay Karthick Vaidyanathan、Srihari V K 和 Thenmozhi Durairaj。2025a。自然语言处理_goats@DravidianLangTech 2025：迈向更安全的社交媒体：检测达罗毗荼语言中针对女性的辱骂性语言。载于《第五届达罗毗荼语言语音、视觉与语言技术研讨会论文集》，第 350–354 页，阿科马，阿尔伯克基会议中心，美国新墨西哥州阿尔伯克基。计算语言学协会。

Vijay Karthick Vaidyanathan、Srihari V K、Mugilkrishna D U 和 Saritha Madhavan。2025b。自然语言处理_goats 在 SemEval-2025 第11项任务中的表现：使用微调后的 roberta-large tranformer 进行多标签情感分类。载于《第19届国际语义评测研讨会（SemEval-2025）会议论文集》，第1023–1027页，维也纳，奥地利。计算语言学协会。

Abhiram Vinjamuri 和 Weiwei Sun。2025。吠陀梵语依存句法分析的迁移学习。载于《第9届拓展自然语言处理研讨会会议论文集》，第50–55页，苏州，中国。计算语言学协会。

Jan Philip Wahle、Terry Ruas、Yang Xu 和 Bela Gipp。2024年。释义类型能够激发提示工程能力。载于2024年自然语言处理实证方法会议论文集，第 11004–11033 页，美国佛罗里达州迈阿密。计算语言学协会。

Neng Wan、SicongHuang、EshaUbale 和 IanLane。2025。UCSC 在 SemEval-2025 第8项任务中的表现：基于表格数据的问答。载于《第19届国际语义评测研讨会（SemEval-2025）会议论文集》，第2050–2058页，维也纳，奥地利。计算语言学协会。

BenluWang, IrisXia, YifanZhang, JundaWang, Feiyun Ouyang, Shuo Han, Arman Cohan, Hong Yu, 和 Zonghai Yao。2025a。从评分到步骤：诊断并提升大语言模型在循证医学计算中的表现。载于《2025年自然语言处理实证方法会议论文集》，第10820–10844页，中国苏州。计算语言学协会。

Bin Wang, Chao Xu, Xiaomeng Zhao, Linke Ouyang, Fan Wu, ZhiyuanZhao,Rui Xu,Kaiwen Liu, Yuan Qu, Fukai Shang, Bo Zhang, Liqun Wei, Zhihao Sui, Wei Li, Botian Shi, Yu Qiao, Dahua Lin, 和 Conghui He。2024a。Mineru：一种用于精确文档内容提取的开源解决方案。预印本，arXiv:2409.18839。

Dashun Wang 和 Albert-László Barabási。2021。科学之科学。剑桥大学出版社。

Jiayou Wang、Rundong Liu、Yue Hu、Huijia Wu 和 Zhaofeng He。2025b。SecDecoding：用于更安全的大语言模型生成的可控解码。载于《计算语言学协会论文集：EMNLP 2025》，第20504–20521页，苏州，中国。计算语言学协会。

Qidong Wang、Junjie Hu 和 Ming Jiang。2025c。V-SEAM：通过视觉语义编辑和注意力调节，实现视觉语言模型的因果可解释性。载于《2025年自然语言处理实证方法会议论文集》，第17407–17431页，苏州，中国。计算语言学协会。

Song Wang、Zhen Tan、Zihan Chen、Shuang Zhou、Tianlong Chen 和 Jundong Li。2025d。Any-MAC：通过下一智能体预测实现灵活的级联式多智能体协作。载于《2025年自然语言处理实证方法会议论文集》，第11566–11578页，苏州，中国。计算语言学协会。

WenHao Wang、Mengying Yuan、Zijie Yu、GuangyiLiu、Rui Ye、Tian Jin、Siheng Chen 和 Yanfeng Wang。2025e。MobileA3gent：使用来自多样化用户的去中心化自采集数据训练移动 GUI 智能体。载于第四届人机交互与自然语言处理交叉研讨会（HCI+NLP）会议论文集，第 79–112 页，中国苏州。计算语言学协会。

Xiaochen Wang、Jiaqi Wang、Houping Xiao、Jinghui Chen 和 Fenglong Ma。2024b。FEDKIM：将联邦知识注入自适应地融入医学基础模型。载于2024年自然语言处理实证方法会议会议论文集，第 8141–8154 页，美国佛罗里达州迈阿密。计算语言学协会。

Xiaonan Wang、Bo Shao 和 Hansaem Kim。2025f。人工智能知道你在哪里：使用 KoreaGEO 进行多模态地理定位中的暴露、偏见与推理。载于《2025年自然语言处理实证方法会议论文集》，第 9897–9914 页，中国苏州。计算语言学协会。

Haoran Wei、Yaofeng Sun 和 Yukun Li。2025a。Deepseek-ocr：上下文光学压缩。预印本，arXiv:2510.18234。

Jiaqi Wei、Yuejin Yang、Xiang Zhang、Yuhan Chen、Xiang Zhuang、Zhangyang Gao、Dongzhan Zhou、Guangshuai Wang、Zhiqiang Gao、Juntai Cao、Zijie Qiu、Ming Hu、Chenglong Ma、Shixiang Tang、Jun- jun He、Chunfeng Song、Xuming He、Qiang Zhang、Chenyu You 及另外 8 位作者。2025b。从人工智能赋能科学到智能体科学：自主科学发现综述。预印本，arXiv:2508.14111。

Qiming Wu、Zichen Chen、Will Corcoran、Misha Sra 和 Ambuj Singh。2025a。GraphEval36K：基于图数据集对大语言模型的编码和推理能力进行基准测试。载于《计算语言学协会论文集：NAACL 2025》，第8095–8117页，美国新墨西哥州阿尔伯克基。计算语言学协会。

Zekun Wu、Seonglae Cho、Umar Mohammed、Cristian Enrique Munoz Villalobos、Kleyton Da Costa、Xin Guan、Theo King、Ze Wang、Emre Kazim 和 Adriano Koshiyama。2025b。LibVulnWatch：用于发现开源人工智能库中隐藏漏洞的深度评估智能体系统和排行榜。载于《计算语言学协会第63届年会论文集》（第4卷：学生研究研讨会），第608–656页，奥地利维也纳。计算语言学协会。

Magdalena Wysocka、Danilo Carvalho、Oskar Wysocki、Marco Valentino 和 Andre Freitas。2025年。SylloBio-NLI：评估大语言模型的生物医学三段论推理能力。载于美洲分会计算语言学协会人类语言技术会议（第1卷：长篇论文）论文集，页码7235–7258，新墨西哥州阿尔伯克基。计算语言学协会。

Eric Xia、Karthik Srikumar、Keshav Karthik、Advaith Renjith 和 Ashwinee Panda。2025年。超越干草堆：对法律参考文献召回中上下文敏感性的研究。载于《2025年自然法律语言处理研讨会会议论文集》，第48–53页，中国苏州。计算语言学协会。

Chengrui Xiang、Tengfei Ma、Xiangzheng Fu、Yiping Liu、Bosheng Song 和 Xiangxiang Zeng。2025年。从知识到治疗：大语言模型辅助的生物医学概念表示及其在药物再利用中的应用。载于《计算语言学协会论文集：EMNLP 2025》，第13967–13982页，中国苏州。计算语言学协会。

Yunze Xiao、Yujia Hu、Kenny Tsu Wei Choo 和 Roy Ka-Wei Lee。2024年。ToxiCloakCN：利用伪装扰动评估中文冒犯性语言检测的鲁棒性。载于2024年自然语言处理实证方法会议论文集，页码6012–6025，美国佛罗里达州迈阿密。计算语言学协会。

Yunze Xiao、Lynnette Hui Xian Ng、Jiarui Liu 和 Mona T. Diab。2025年。使机器更具人性：通过多层次设计框架重新思考大语言模型的拟人化。载于2025年自然语言处理实证方法会议论文集，页码3331–3350，中国苏州。计算语言学协会。

Pengcheng Xu、Sicheng Shen、Jieli Zhou 和 Hongyi Xin。2025a。XSZ团队在 BioLaySumm2025 中的工作：分节摘要、检索增强大语言模型以及用于通俗摘要的强化学习微调。载于《第24届生物医学语言处理研讨会（共享任务）会议论文集》，第275–280页，奥地利维也纳。计算语言学协会。

WanghanXu、Yuhao Zhou、YifanZhou、Qinglong Cao、Shuo Li、Jia Bu、Bo Liu、Yixin Chen、Xuming He、Xiangyu Zhao、Xiang Zhuang、Fengxiang Wang、Zhiwang Zhou、Qiantai Feng、Wenxuan Huang、Ji- aqi Wei、Hao Wu、Yuejin Yang、Guangshuai Wang 等88人。2025b。通过与科学家对齐的工作流探索大语言模型的科学通用智能。预印本，arXiv:2512.16969。

Yuebin Xu、Zhiyi Chen 和 Zeyi Wen。2025c。Eco-Tune：面向大语言模型推理的令牌高效多保真度超参数优化。载于《2025年自然语言处理实证方法会议论文集》，第 7746–7756 页，中国苏州。计算语言学协会。

Zicheng Xu、Guanchu Wang、Guangyao Zheng、Yu-Neng Chuang、Alex Szalay、Xia Hu 和 Vladimir Braverman。2025d。自集成：缓解大语言模型的置信度失真。载于《计算语言学协会论文集：EMNLP 2025》，第 16603–16615 页，中国苏州。计算语言学协会。

Yutaro Yamada、Robert Tjarko Lange、Cong Lu、Shengran Hu、Chris Lu、Jakob Foerster、Jeff Clune 和 David Ha。2025 年。AI Scientist-v2：通过智能体树搜索实现研讨会级别的自动化科学发现。预印本，arXiv:2504.08066。

Bohao Yang、Dong Liu、Chenghao Xiao、Kun Zhao、Chen Tang、Chao Li、Lin Yuan、Yang Guang 和 Chenghua Lin。2025a。使用大语言模型打造可定制角色：角色驱动的角色扮演智能体框架。载于《计算语言学协会论文集：EMNLP 2025》，第 20216–20240 页，中国苏州。计算语言学协会。

Dayu Yang、Antoine Simoulin、Xin Qian、Xiaoyi Liu、Yuwei Cao、Zhaopu Teng 和 Grey Yang。2025b。DocAgent：用于自动生成代码文档的多智能体系统。载于《计算语言学协会第63届年会论文集》（第3卷：系统演示），第460–471页，奥地利维也纳。计算语言学协会。

Jing Yang。2025。立场论文：人工智能和机器学习社群应采用更透明且受监管的同行评审流程。载于《第42届国际机器学习会议论文集》，机器学习研究论文集第267卷，第82399–82408页。PMLR。

Seungyoun Yi、Minsoo Khang 和 Sungrae Park。2025。ZERA：零初始化指令演进优化智能体——通过基于原则的优化，从零指令迈向结构化提示。载于《2025年自然语言处理实证方法会议论文集》，第23334–23348页，中国苏州。计算语言学协会。

Kai Yin、Xiangjue Dong、ChengkaiLiu、Lipai Huang、Yiming Xiao、Zhewei Liu、Ali Mostafavi 和 James Caverlee。2025。DisastIR：面向灾害管理的综合信息检索基准。载于《计算语言学协会论文集：EMNLP 2025》，第 1836–1867 页，中国苏州。计算语言学协会。

Ko Yoshida、Daiki Shiono、Kai Sato、Toko Miura、Momoka Furuhashi 和 Jun Suzuki。2025。批次级收敛预训练：受儿童语言发展启发的逐步学习。载于《首届 BabyLM 研讨会论文集》，第 508–524 页，中国苏州。计算语言学协会。

Yao-Ching Yu、Tsun-Han Chiang、Cheng-Wei Tsai、Chien-Ming Huang 和 Wen-Kwang Tsao。2025a。Primus：用于网络安全大语言模型训练的开创性开源数据集集合。载于《2025年自然语言处理实证方法会议论文集》，第 10402–10424 页，中国苏州。计算语言学协会。

Zhaojian Yu、Yinghao Wu、Yilun Zhao、Arman Co- han 和 Xiao-Ping Zhang。2025b。Z1：通过代码实现高效的测试时扩展。载于《2025年自然语言处理实证方法会议：工业赛道，2688–2712页，苏州（中国）。计算语言学协会。

Chenchen Yuan、Zheyu Zhang、Shuo Yang、Bardh Prenkaj 和 Gjergji Kasneci。2025a。大语言模型中集体道德推理的概率聚合与定向嵌入优化。载于计算语言学协会论文集：ACL 2025，第11151–11168页，奥地利维也纳。计算语言学协会。

Mengying Yuan、WenHao Wang、Zixuan Wang、Yu- jie Huang、Kangli Wei、Fei Li、Chong Teng 和 Donghong Ji。2025b。通过RST增强图融合和可解释性预测实现跨文档跨语言自然语言推理。载于2025年自然语言处理实证方法会议论文集，第31595–31617页，中国苏州。计算语言学协会。

Mengying Yuan、WenHao Wang、Zixuan Wang、Yu- jie Huang、Kangli Wei、Fei Li、Chong Teng 和 Donghong Ji。2025c。通过RST增强图融合和可解释性预测实现跨文档跨语言自然语言推理。载于第五届多语言表示学习研讨会（MRL 2025）论文集，第11–33页，中国苏州。计算语言学协会。

Aizan Zafar、Kshitij Mishra 和 Asif Ekbal。2024年。MedLogic-AQA：利用聚焦逻辑结构的抽象式模型增强医疗问答。载于计算语言学协会论文集：EMNLP 2024，第 16844–16867 页，迈阿密，佛罗里达州，美国。计算语言学协会。

Ofir Zafrir、Igor Margulis、Dorin Shteyman、Shira Guskin 和 Guy Boudoukh。2025。FastDraft：如何训练你的草稿模型。载于计算语言学协会论文集：ACL 2025，第 22488–22505 页，维也纳，奥地利。计算语言学协会。

Yiming Zeng、Wanhao Yu、Zexin Li、Tao Ren、Yu Ma、Jinghan Cao、Xiyan Chen 和 Tingting Yu。2025。弥合大语言模型的编辑差距：FineEdit 实现精确且有针对性的文本修改。载于计算语言学协会论文集：EMNLP 2025，第 2193–2206 页，苏州，中国。计算语言学协会。

Hao Zhang、Bo Huang、Zhenjia Li、Xi Xiao、Hui Yi Leong、Zumeng Zhang、Xinwei Long、Tianyang Wang 和 Hao Xu。2025a。Sensitivity-LoRA：基于低负载敏感度的大型语言模型微调。载于《计算语言学协会论文集：EMNLP 2025》，第 13185–13199 页，苏州，中国。计算语言学协会。

Hengrui Zhang、Pin-Siang Huang、Zhen Zhang、Pe- ican Lin、Yao-Ching Yu、Bo Hu 和 Yulu Du。2025b。PathwiseRAG：多维探索与集成框架。载于会议论文集

2025年自然语言处理实证方法会议，页 22915–22936，苏州，中国。计算语言学协会。

Jiaxin Zhang。2025年。置信度感知推理：优化大型推理模型中的自引导思维轨迹。载于《2025年自然语言处理实证方法会议：产业分会论文集》，第2081–2095页，苏州（中国）。计算语言学协会。

Shiwen Zhang、Lingxiang Wang、Hainan Zhang、Ziwei Wang、Sijia Wen 和 Zhiming Zheng。2025c。超越表象：用于竞赛级代码生成的解答感知检索模型。载于《计算语言学协会论文集：EMNLP 2025》，第5237–5246页，中国苏州。计算语言学协会。

Tianyu Zhang、Xinyu Wang、Lu Li、Zhenghan Tai、Jijun Chi、Jingrui Tian、Hailin He 和 Suyuchen Wang。2025d。STRICT：文本图像渲染压力测试。载于《2025年自然语言处理实证方法会议论文集》，第21148–21161页，中国苏州。计算语言学协会。

WeiZhang、Hongcheng Guo、Jian Yang、Zhoujin Tian、Yi Zhang、Yan Chaoran、Zhoujun Li、Tongliang Li、Xu Shi、Liangfan Zheng 和 Bo Zhang。2024年。mABC：受区块链启发的多智能体协作，用于微服务架构中的根因分析。载于《计算语言学协会论文集：EMNLP 2024》，第4017–4033页，美国佛罗里达州迈阿密。计算语言学协会。

Yiming Zhang、Siyue Zhang、Junbo Zhao 和 Chen Zhao。2025e。RPDR：一种面向长尾问答的往返预测数据增强框架。载于《2025年自然语言处理实证方法会议论文集》，第22009–22023页，中国苏州。计算语言学协会。

Yizhe Zhang 和 Navdeep Jaitly。2025年。SAGE：利用面向未来的状态-动作增强引导对话生成。载于《第4届自然语言处理视角主义方法研讨会论文集》，第123–132页，中国苏州。计算语言学协会。

Yukun Zhang 和 Xueqing Zhou。2025。连续时间注意力：面向长序列 Transformer 模型的偏微分方程引导机制。载于《2025年自然语言处理实证方法会议论文集》，第 21654–21674 页，中国苏州。计算语言学协会。

Zhuoxuan Zhang、Jinhao Duan、Edward Kim 和 Kaidi Xu。2025f。稀疏神经元在大语言模型中承载着有关问题歧义的强烈信号。收录于《2025 年自然语言处理实证方法会议论文集》

语言处理，第16092–16110页，苏州，中国。计算语言学协会。

郑传洋、高一航、陈国轩、史翰、熊静、任晓哲、Chao Huang、Zhen-guo Li 和 Yu Li。2025a。自适应 Softmax。载于《2025年自然语言处理实证方法会议论文集》，第 7827–7847 页，中国苏州。计算语言学协会。

郑顺风、方萌和 Ling Chen。2025b。SpatialWebAgent：利用大语言模型实现自动化空间信息抽取和地图定位。载于《计算语言学协会第63届年会论文集》（第 3 卷：系统演示），第 252–266 页，奥地利维也纳。计算语言学协会。

郑顺风, 张雨迪, 方萌, 张子涵, 吴志坦, Mykola Pechenizkiy 和 Ling Chen。2025c。使用检索增强生成对基础模型在奥林匹克级物理问题求解中的表现进行基准测试。载于协会发现论文集：EMNLP 2025，第21927–21956页，中国苏州。计算语言学协会。

郑宇翔, 傅大远, 胡祥坤, 蔡晓杰, 叶柳曼珊, 卢鹏睿和 Pengfei Liu。2025d。DeepResearcher：通过在真实世界环境中进行强化学习来扩展深度研究。载于2025年自然语言处理实证方法会议论文集，第414–431页，中国苏州。计算语言学协会。

郑中华, 廖丽姿, 邓阳, Libo Qin 和聂立强。2024年。大语言模型的自我对话能让小型情感支持聊天机器人变得更好。载于第62届计算语言学协会年会论文集（第1卷：长篇论文），第11325–11345页，泰国曼谷。计算语言学协会。

周靖卓。2025。多模态大语言模型能检测网络钓鱼吗？聚焦动态威胁和多模态评估的学术环境综合安全基准套件。预印本，arXiv:2511.15165v1。注：该问题已在版本 2 中修复。

周书、宣云扬、敖宇轩、王鑫、范涛和Hao Wang。2025a。MERIT：用于无监督时间序列表征学习的多智能体协作。载于《计算语言学协会会议论文集：ACL 2025》，第 24011–24028 页，维也纳，奥地利。计算语言学协会。

WenJie Zhou、Bohan Wang、Wei Chen 和 Xueqi Cheng。2025b。BSFA：利用子空间二分法加速神经网络训练。载于 2025年自然语言处理实证方法会议论文集，页

18845–18860，中国苏州。计算语言学协会。

Yue Zhou 和 Barbara Di Eugenio。2025。真实性偏差及其延伸：揭示大语言模型在问题求解推理中的隐性信念。载于《计算语言学协会第63届年会论文集》（第1卷：长篇论文），第21298–21310页，奥地利维也纳。计算语言学协会。

Zebiao Zhou、Hui Li、Xiangxun Zhu 和 Kangzhen Liu。2025c。TranssionMT 在 WMT 2025 印度语机器翻译共享任务中的提交作品。载于《第十届机器翻译会议论文集》，第1271–1275页，中国苏州。计算语言学协会。

Zihan Zhou、Simon Kurz 和 Zhixue Zhao。2025d。重新审视小型语言模型中的剪枝与量化。载于《计算语言学协会论文集：EMNLP 2025》，第12055–12070页，中国苏州。计算语言学协会。

He Zhu、Guanhua Chen 和 Wenjia Zhang。2025a。PlanGPT：通过定制的智能体框架提升城市规划。载于《计算语言学协会第63届年会论文集》（第6卷：工业赛道），第764–783页，奥地利维也纳。计算语言学协会。

Kunlun Zhu、Hongyi Du、Zhaochen Hong、Xiaocheng Yang、Shuyi Guo、Zhe Wang、Zhenhailong Wang、Cheng Qian、Robert Tang、Heng Ji 和 Jiaxuan You。2025b。MultiAgentBench：评估大语言模型智能体的协作与竞争。载于《计算语言学协会第63届年会论文集》（第1卷：长篇论文），第8580–8622页，奥地利维也纳。计算语言学协会。

Lichao Zhu、Maria Zimina-Poirot、Cristian Valdez 和 Stephane Patin。2025c。使用对齐语料库、术语表和合成数据微调神经机器翻译模型与大语言模型，以进行专业英西翻译：MUL-TITAN 参与 WMT25术语共享任务。载于《第十届机器翻译大会论文集》，第1284–1291页，中国苏州。计算语言学协会。

Runchuan Zhu、Bowen Jiang、Lingrui Mei、Fangkai Yang、Lu Wang、Haoxiang Gao、Fengshuo Bai、Pu Zhao、Qingwei Lin、SaravanRajmohan 和 Dong- mei Zhang。2025d。AdaptFlow：通过元学习实现自适应工作流优化。载于《计算语言学协会论文集：EMNLP 2025》，第 3287–3302 页，中国苏州。计算语言学协会。

Runchuan Zhu、Zinco Jiang、Jiang Wu、Zhipeng Ma、Jiahe Song、Fengshuo Bai、Dahua Lin、Lijun Wu 和 Conghui He。2025e。GRAIT：用于有效缓解幻觉的梯度驱动拒绝感知指令微调。载于协会发现论文集：计算语言学：NAACL2025，第 4006–4021 页，阿尔伯克基，新墨西哥州。计算语言学协会。

更多建议

构建便于核验的格式并规范引文规范。幻觉引文常常被隐藏在参考文献部分中，导致人类审稿人进行视觉检测的成本很高。尽管本研究依靠基于 OCR 的解析和人工核验来识别此类情况，但解析并非总是完美。因此，我们认为，ACL 格式的参考文献部分可以改进得更便于机器处理，从而促进自动核验；例如，可以在不同引文之间加入更清晰的结构分隔符，更明确地标注标题等关键字段，并直接打印 URL，而不只是将其嵌入为超链接。此外，我们认为可验证标识符（如可点击链接和 DOI）应当是强制要求。这样有助于核验、减少人工工作，并实现自动核验。

取消领域选择或提高领域透明度。如第4.1节所述，不同研究领域的 HalluCited 论文比例各不相同，新兴领域的比例较高。我们认为，这种差异部分是领域划分造成的。因此，我们建议改为不划分领域，或公开已接收论文的领域信息，以提高透明度。随着跨领域研究日益普及，固定的领域选择可能导致审稿人匹配不当，并降低评审严谨性。鉴于 OpenReview 等现代评审系统支持自动匹配审稿人，明确选择领域或许已无必要。另一种做法是，由于目前领域信息仅部分公开，[9]，将其存档于 ACL Anthology 等资源中，有助于减少匹配不当，并支持更有效的评审，包括检测幻觉引文。

B HalluCited 论文列表

表7 列出了所有已识别的 HalluCited 论文及相应证据，即幻觉引文。出于隐私考虑，我们以参考文献格式提供相关信息，而非直接展示 HalluCited 论文的详细信息。若要识别对应论文，请点击所提供的链接或参阅参考文献部分。

表 7：幻觉论文（即 HalluCited 论文）及其对应的幻觉引文列表 s.

	事件编号	论文编号	幻觉论文	幻觉示例
1	NAACL 2024	2024.naacl-short.63	Scaria 等人 （2024年）	Himanshu Gupta、Kevin Scaria、Swaroop Mishra, 以及 Chitta Baral。2024b。超越数据瓶颈：通过基于难度的示例选择优化指令微调。arXiv 预印本。
2	NAACL 2024	2024.trustnlp-1.1	Adilazuarda (2024年)	Daphne Ippolito、Daniel Duckworth、Chris CallisonBurch 和 Douglas Eck。2020年。区分人类创作与机器生成文本：综述。arXiv 预印本 arXiv:2012.03358。
3	acl-2024	2024.acl-long.560	Siska 等人(2024年)	Keisuke Sakaguchi、Ronan Le Bras、Chandra Bhagavatula 和 Yejin Choi。2023年。一项大规模对抗性 Winograd 模式挑战。arXiv 预印本 arXiv:2305.06300。
4	acl-2024	2024年.acl-long.611	Zheng等人（2024年）	Chun-Hung Yeh、Anuradha Welivita 和 Pearl Pu Faltings。2015。一种包含为困境中的人提供情感支持的对话数据集。arXiv 预印本 arXiv:1503.08895。
5	acl-2024	2024.acl-long.768	Chen 等人 (2024年)	Mengjie Fang、Linlin Xu 和 Pascale Fung。2020年。无监督跨语言迁移学习：上下文化词嵌入。在2020年实证研究会议论文集自然语言处理方法（EMNLP），页 2616–2629。
6	acl-2024	2024.arabicnlp-1.24	Khondaker 等人（2024年）	Isabel Moreno 和 Wei Zhang。2024年。评估用于阿拉伯语自然语言处理任务的多语言模型。计算语言学，50(3):425–445。
7	acl-2024	2024.arabicnlp-1.45	Shah 等人(2024年)	Naoya Inoue、Pontus Stenetorp 和 Kentaro Inui。2021。偏好与机器学习在语言模型微调中的相互作用。arXiv 预印本 arXiv:2110.08413。
8	acl-2024	2024.findings-acl.241	Holt 等人 (2024年)	Junjie Li、Jieyu Wu 和 Richard Socher。2021 年。理解语言模型。arXiv 预印本 arXiv:2109.04278。
9	acl-2024	2024.wassa-1.18	Alhamed 等人(2024年)	David Benetka、Alicia Moreno-Moral、Lorena RomeroFombuena 和 Juan Lopez-Gazpio。2020年。心理健康讨论的标注方案。在国际计算语言学会议（会议论文集：长篇论文，2020年），第2617–2627页。计算语言学协会计算语言学。
10	emnlp-2024 2		Mundra 等人 (2024年)	Adam Alabi、Saleha Nawaz 和 Vincent Ng。2022年。Alabi：一种轻量级方法，用于多语言生物医学语言模型。在2022年自然语言处理实证方法会议论文集自然语言处理，页 9717–9727。
11	emnlp-2024	2024.emnlp-demo.11	Sheth等人 (2024年)	Stephan Druskat、Ulrike Gut、Nils Reiter、Stefan Schweter ，和 Manfred Stede。2014年。Atomic：一款用于处理多种语言中的照应的开源工具。载于 2014年会议论文集自然语言处理实证方法会议：系统演示，第 71–76 页。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
12	emnlp-2024 2(		Malik 等人 (2024年)	Jiajia Liu、Mengyuan Yang、Yankai Yu、Haixia Xu、Kang Li 和 Xiaobo Zhou。2023a。借助人工智能驱动的对话代理提升客户体验。arXiv 预印本 arXiv:2306.04325。
13	emnlp-2024 2(		Xiao 等人(2024年)	Nelson F Liu、Tony Wu、Duane S Boning 和 Tanmoy Choudhury。2020b。AI 漏洞检测器：自然语言的对抗输入检测模型。发表于 2020年自然语言实证方法会议语言处理：系统演示，第 187–196 页。
14	emnlp-2024 2		Wang 等人 (2024b)	Yu Gu、Robert Tinn、Hao Cheng、Youzheng Ben, Zhuozhao Liu、Jingqi Zhou、Michael Wang, Shizhuo Wang、Hongfang Zhou 和 Yanshan Shen。2021。Biomedgpt：一种大型生物医学生成式预训练变换器，用于生物医学文本挖掘。arXiv 预印本 arXiv:2104.07497。
15	emnlp-2024 2		Wahle 等 (2024年)	Dominik Meier、Jan Philip Wahle、Terry Ruas, 以及 Bela Gipp。2024年。一项关于原子释义类型生成的人类研究。arXiv。
16	emnlp-2024 2(		Zhang 等人(2024年)	Yi Chen 等。2023年。利用增强型大语言模型赋能云根因分析。arXiv 预印本 arXiv:2311.00000。
17	emnlp-2024 2(		Li 等人 (2024年)	Ruochen Wang、Ting Liu、Cho-Jui Hsieh 和 Boqing Gong。2023年。DPO-DIFF：关于离散文本到图像的提示词优化扩散模型生成自然语言对抗样本。预印本，arXiv:2311.07998。
18	emnlp-2024 2		Son等人(2024年)	Yoav Goldberg。2022年。对有关以下内容的主张进行评估：大语言模型。arXiv 预印本 arXiv:2212.09273。
19	emnlp-2024 2(		Zafar等人(2024年)	Xinyu Wang、Tao Sun、Deqing Zou、Wei Wu，以及 Jiawei Han。2021年。逻辑引导的数据增强与正则化用于一致性学习。arXiv 预印本 arXiv:2104.04379。
20	emnlp-2024	2024.mrl-1.20	Skianis等人(2024年)	Shaoxiong Ji、Yanzhao Zhang、Leilei Sun 和 Jia Wang。2022年。MentalBERT：一种预训练语言模型，用于心理健康护理。发表于2022年自然语言处理实证方法会议论文集（EMNLP）。
21	naacl-2025	2025.americasnlp-1.4	Alvarez C 等（2025）	Manuel Mager、Arturo Oncevay、Annette Rios, Jamshidbek Mirzakhalov ，以及 Katharina Kann。2023年。计算语言学在原住民语言复兴中的作用：挑战与机遇。载于第一届原住民语言计算研讨会（C3NLP）会议论文集，第23–31页。
22	naacl-2025	2025.clpsych-1.23	Antony 和 Schoene (2025)	Ilias Chalkidis、Ion Androutsopoulos 和 Nikolaos Aletras。2022年。一项关于法律判决预测神经方法的实证研究。载于2022年实证方法会议论文集自然语言处理，页4783–4798。计算语言学协会。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
23	naacl-2025	2025 德拉威语言技术 1.117	Noor 等人(2025)	G. Shimi 等人 2024年。用于达罗毗荼语语言识别：假新闻多语言环境下检测的关键步骤。待定。
24	naacl-2025	2025 德拉威语言技术 1.118	Selvamurugan （2025）	Anil Kumar、Ajay Kumar Ojha、Shervin Malmasi, 以及 Marcos Zampieri。2021。基准测试社交媒体中的低资源语言攻击性识别。载于在线虐待与伤害研讨会会议论文集（ACL）。
25	naacl-2025	2025 德拉威语言技术 1.38	S 等 （2025a）	Simran Khanuja、Kaustav Dey、El Moatez Billah Karim Nagoudi 等。2020年。一个新的数据集和用于代码混合检测的强大基线冒犯性语言。载于《2020年自然语言处理实证方法会议论文集》（EMNLP），页1719–1726。
26	naacl-2025	2025 德拉威语言技术 1.43	Chauhan 和 Kumar (2025)	Douwe Kiela、Mohamed Elhoseiny、Lin Zhang, Marco Baroni 和 Geoffrey Hinton。2019 年。面向可扩展性的监督式多模态哈希跨模态检索。载于 2019 年会议论文集自然语言处理实证方法会议语言处理（EMNLP），页1740–1751。
27	naacl-2025	2025 德拉威语言技术 1.61	K 等（2025）	F. Wu 和 M. Dredze。2019。社交媒体作为公众舆论的传感器。2019年自然语言处理实证方法会议论文集（EMNLP 2019），2019：1001–1010。
28	naacl-2025	2025 德拉威语言技术 1.62	Vaidyanathan等人 (2025a)	Jiho Park、Jihyung Shin、Sangyoun Lee 和 Changhyun Seo。2018年。仇恨言论检测：数据、方法与挑战。在第27届国际会议论文集国际计算语言学会议（COLING 2018），页385–395。国际计算语言学委员会计算语言学。
29	naacl-2025	2025 德拉威语言技术 1.7	Sathvik 等（2025b）	M. H. Al-Adhaileh 和 F. W. Alsaade。2022年。双向长短期记忆网络（bilstm）用于虚假评论检测的网络：A比较研究。施普林格。
30	naacl-2025	2025 德拉威语言技术 1.88	Achamaleh 等人 (2025年)	Anna Kolesnikova 和 Sergey Ivanov。2023年。探索多语言文本表示Transformer 模型。ACL 汇刊，11：212–230。
31	naacl-2025	2025 德拉威语言技术 1.91	Aftahee 等人 （2025年）	R. Raja、A. Kumar 和 S. Joseph。2023年。假低资源语言中的假新闻检测：挑战与进展。《计算《语言学评论》，15(2):123–137。
32	naacl-2025	2025 德拉威语言技术 1.97	Sathvik 等 (2025a)	Simran Khanuja、Sandipan Dandapat、Ritesh Kumar、Sunayana Sitaram、K. P. Soman 和 Anup Kumar。2021。MahaNLP：面向印度语言理解，使用 BERT 模型进行印地语、马拉地语和卡纳达语。arXiv 预印本 arXiv:2106.07469。
33	naacl-2025	2025.findings-naacl.132	Singh 等人(2025)	W. Liu、X. Li、W. Yang、Y. Lin、X. Liu、S. Wang, C. Xie、L. Xu、W. Zhuang、X. Zhao 和 L. Li。2020年。Minilm：面向微型Transformer的深度自注意力蒸馏。arXiv 预印本 arXiv:2002.10957。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
34	naacl-2025	2025.findings-naacl.223	Zhu 等人（2025e）	Krishnateja Killamsetty 和 Rishabh Iyer。2021。采用梯度匹配的子集选择。arXiv 预印本 arXiv:2103.00123。
35	naacl-2025	2025.findings-naacl.230	Raihan 等人(2025)	OpenAI。2024年。GPT-4 Omni：一款全面的多模态模型，适用于语言、视觉及更多领域。arXiv 预印本 arXiv:2408.01234。
36	naacl-2025	2025.Findings-NAACL.257	Magomere 等(2025)	Chaitanya Shivade 等。2019。MedNLI——一个自然语言推理数据集，适用于临床领域。载于 2018年自然语言处理实证方法会议自然语言会议论文集处理，布鲁塞尔，比利时。美国计算语言学，第 1586–1596 页。
37	naacl-2025	2025.findings-naacl.275	Song等人(2025b)	M Honnibal 和 I Montani。2017。spacy 2：基于 bloom 的自然语言理解嵌入、卷积神经网络和增量解析。神经机器翻译。载于计算语言学协会会议论文集（ACL），页 688–697。
38	naacl-2025	2025.findings-naacl.279	Basher等人 (2025)	Y. Xu等人 2023年。跨语言迁移低资源文本转语音。载于2023年自然语言处理实证方法会议自然语言处理，页 510–520。计算语言学协会。
39	naacl-2025	2025.findings-naacl.349	Gupta 等人 (2025a)	Zijun Zhang、Yue Wu、Hao Guan、Xinlei Chen，以及 Yue Zhang。2023年。使用.....进行持续学习Transformer模型：挑战与解决方案。arXiv 预印本 arXiv:2302.13713。
40	naacl-2025	2025.findings-naacl.418	Nguyen 等人（2025b）	Danny Hernandez、Jared Kaplan、Tom Henighan，Tom Brown ，Jack Clark ，Preetum Nakkiran ，Catherine Olsson ，以及 Afshin Rahimi。2023年。测量大语言模型中的数据污染模型。arXiv 预印本 arXiv:2303.13375。
41	naacl-2025	2025.findings-naacl.452	Wu 等 (2025a)	Abhijit Srivastava、Nathan Major、Jasmine D. Hernandez，等。2022年。超越模仿游戏：量化并外推大语言模型能力与 BIG-Bench。arXiv 预印本 arXiv:2206.04615。
42	naacl-2025	2025.findings-naacl.58	Liu 等人 （2025c）	Xiang Jiang等人 2023c。Mistral-7b-instruct-v0.2：一种高效的大型指令遵循语言模型。arXiv 预印本 arXiv:2309.00001。
43	naacl-2025	2025.knowledgenlp-1.18	Spangher 等人 (2025a)	Tushar Khot、Ashish Sabharwal 等人。2023年。语言模型中的分解驱动推理arXiv 预印本 arXiv:2304.xxxxx。
44	naacl-2025	2025.knowledgenlp-1.20	Pattnayak 等（2025）	等。Zhao, W. 2020年。一种检索增强编码器-解码器，用于知识密集型自然语言处理任务。载于第58届年会会议论文集（计算语言学协会（ACL）。
45	naacl-2025	2025.loresmt-1.12	Raja 和 Vats （2025）	Kalika Bali、Jatin Sharma、Monojit Choudhury，以及 Yogarshi Vyas。2014年。语码混合：一项针对印度语境下语言识别的挑战上下文。载于第一届计算方法处理语码转换研讨会（EMNLP），第 13–23 页。
46	naacl-2025	2025.naacl-demo.39	Hiray 和 Kovatchev (2025)	T. Lee等人 2023年。Bivdiff：一种免训练通用视频合成框架。arXiv 预印本 arXiv:2302.02918。

续见下页

	事件编号	论文编号	幻觉论文	幻觉引文示例
47	naacl-2025	2025.naacl-industry.68	Jararweh 等人(2025)	刘昊天、林春元、曾方云，以及 等人 2023a。Llava：大型语言与视觉 助手。arXiv 预印本 arXiv:2304.08485。
48	naacl-2025	2025.naacl-long.182	Chen 等人 (2025年b)	Ebtesam Almazrouei、Hamza Alobeidli、 Abdulaziz Alshamsi、Alessandro Cappelli、 Ruxandra Cojocaru、Merouane Debbah、Etienne Goffinet、Daniel Heslow、Julien Launay、Quentin Malartic 等人。2023年。Falcon-40b：一个开放的大型 具有最先进性能的语言模型。 计算语言学协会论文集 ：ACL，2023年：10755–10773。
49	naacl-2025	2025.naacl-long.363	Qian 等（2025）	Zhengbao Jiang、Frank F Xu、Jun Araki 和 Graham Neubig。2020年。基于提示的 学习能否适用于低资源任务? arXiv 预印本 arXiv:2001.07676。
50	naacl-2025	2025.naacl-long.371	Wysocka 等人（2025年）	Jakob Prange、Khalil Mrini 和 Noah A. Smith。 2023年。自然语言处理中的挑战与机遇 系统泛化：超越 组合性。载于第 61 届 协会年度会议 计算语言学。计算语言学协会 计算语言学。
51	naacl-2025	2025.naacl-long.620	Chen 等人 (2025c)	Jiawei Liu、Yifeng Ding、Naman Jain、Harm de Vries、Leandro von Werra、Arjun Guha， Lingming Zhang 和 Yuxiang Wei。2024a。 Starcoder2-instruct：完全透明且 宽松的代码生成自对齐。 预印本，arXiv:2307.08701。
52	naacl-2025	2025.naacl-long.72	Hu 等人（2025）	Mor Geva、Yoav Goldberg 和 Jonathan Berant。 2021。Strategyqa：一个问答 策略推理基准。在 2021年实证研究会议论文集 自然语言处理方法。
53	naacl-2025	2025.naacl-short.79	Kumar 等人（2025）	Sankha Bhattacharjee 等人。2023年。对抗 学习用于稳健的人工智能生成文本检测。 arXiv 预印本 arXiv:2304.07812。
54	naacl-2025	2025.nlp4dh-1.52	Greschner 和 Klinger (2025)	Chiara Leoni、Mauro Coccoli、Ilaria Torre 和 Gianni Vercelli。2018年。您的论文标题在此。收录于 第五届意大利计算语言学会议论文集 计算语言学（CLiC-it 2018）， 意大利都灵。Accademia University Press。
55	naacl-2025	2025.wnu-1.10	Li 等人 （2025年）	Xingxing Lu、Yuning Mao 和 Jason Wei。2023年。 Autoeval：基于大语言模型的自动评估 文本摘要框架。载于 ACL 发现论文集 ACL.
56	acl-2025	2025.acl-demo.16	She 等人(2025)	Wenhao Zheng 和另外 1 位作者。2025a。Citer： 基于置信度的令牌路由，用于协作 使用大语言模型进行推理。arXiv 预印本 arXiv:2503.01013。
57	acl-2025	2025.acl-demo.25	Zheng等人(2025b)	Wei Zhang、Pengfei Liu 和 Qiang Shen。2020年。 使用……处理地理文本中的歧义 自然语言处理技术。《Transactions in GIS》， 24(3):519–538。
58	acl-2025	2025.acl-demo.44	Yang 等人 (2025b)	Ziniu Wu、Cheng Liu、Jindong Zhang、Xinyun Li， Yewen Wang、Jimmy Xin、Lianmin Zhang、Eric Xing、Yuxin Lu 和 Percy Liang。2023年。AutoGen： 实现新一代多智能体 语言模型通信。arXiv 预印本 arXiv:2309.07864。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
59	acl-2025	2025.acl-industry.30	Fu 等（2025）	Yihan Chen、Dongkuan Zhang、Xiang Wang, Yifan Yang 和 Heng Wang。2023年。Dare：ldirect 面向自适应模式 重配置的参数编辑。arXiv 预印本 arXiv:2310.09570。
60	acl-2025	2025.acl-industry.54	Zhu 等人(2025a)	Yijie Deng、He Zhu、Wen Wang、Minxin Chen, Junyou Su 和 Wenjia Zhang。2025 年。城市 规划基准：一个全面的基准，用于 评估大型 语言模型的能力。†同等贡献。 *通讯作者： wenjiazhang@tongji.edu.cn。
61	acl-2025	2025.acl-long.1034	Zhou 和 Di Eugenio (2025)	Jane Doe 和 Wei Chen。2023年。去偏 语言模型中的推理：一项定量 研究。2023年 自然语言经验方法会议 正在处理。
62	acl-2025	2025.acl-long.1369	Chen 等人 (2025e)	Mia Xu Chen、Orhan Firat、Ankur Bapna、Melvin Johnson 和 Wolfgang Macherey。2024a。 不同模型架构下的缩放定律：A 密集模型与混合专家模型的比较分析 。arXiv 预印本 arXiv:2410.05661。
63	acl-2025	2025.acl-long.1409	Frydenlund （2025）	Jason Lee、Elman Mansimov 和 Kyunghyun Cho。2018年。确定性 非自回归解读 GPT：logit透镜序列 通过迭代优化进行神经序列建模。 载于 2018年 自然语言实证方法会议论文集 论文集，第1173–1182页，比利时布鲁塞尔。 计算语言学协会。
64	acl-2025	2025.acl-long.1435	Jia 等人 (2025a)	Mei Li 等人。2024年。保障大语言模型中的工具使用安全 智能体：挑战与策略。arXiv 预印本 arXiv:2402.03014。
65	acl-2025	2025.acl-long.1439	Siingh 等人（2025）	Wei Xu、Shujie Liu、Minlie Huang、Si Wei 和 Hang Li。2018年。关于神经机器翻译中 幻觉的综述。arXiv 预印本 arXiv:1803.02577。访问日期：2025-02-14。
66	acl-2025	2025.acl-long.1554	Klisura 等（2025）	Dorottya Demeszky、Weijie Gong 和 Diyi Yang。 2019。分析新闻中的偏见和框架构建 文章中的人工智能：以 情感分析为例。载于 2019 年 自然语言处理实证方法年会 论文集，页码 1655–1661。
67	acl-2025	2025.acl-long.1580	Spangher 等人（2025b）	Minjun Chang 等人。2023年。借助 生成式智能体进行自我反思。arXiv 预印本 arXiv:2311.09214。
68	acl-2025	2025.acl-long.248	Lee等人 (2025年)	Wei Chen、Arjun Kumar 和 Lin Yang。2024年。 基于干扰的攻击提示：一种有效的 大语言模型越狱方法。会议论文集中的 第62届计算语言学协会年会 计算语言学（ACL）。
69	acl-2025	2025.acl-long.272	Meng 等人 (2025)	Chen 等人 2023a。NL2FOL：自然语言到 一阶逻辑。arXiv 预印本 arXiv:2304.09102。
70	acl-2025	2025.acl-long.340	Li 等人 (2025c)	OpenAI。2024年。学习使用大语言模型进行推理。 预印本，arXiv:2406.19949。

续见下页

	事件编号	论文编号	HalluCited 论文	幻觉示例
71	acl-2025	2025.acl-long.421	Zhu 等人(2025b)	Xu Sun、Xiaoya Zhang、Yicheng Feng、Shiyang Wang、Shuming Ma、Jiuding He、Zhixu Zhang、Yuxian Gu、Yi Xu、Hao Zhou 和 Zhiyuan Liu。2023年。对大型视觉语言模型的系统性能力评估。arXiv 预印本 arXiv:2305.16372.
72	acl-2025	2025.acl-long.491	Alsagheer 等人(2025)	Hao Zhong、Jieyu Tang、Tianyang Xu 等人。2020年。自然语言处理是否有益于法律体系？一项关于文档表示。arXiv 预印本 arXiv:2005.01647。
73	acl-2025	2025.acl-long.922	Mohammad (2025)	Jan Philip Wahle，Krishnapriya Vishnubhotla，Bela Gipp 和 Saif M. Mohammad。2025。情感、身体、认知、人口统计学，以及情感：计算文本特征的 abcde 计算情感科学。arXiv。
74	acl-2025	2025.acl-long.979	Chang 等（2025）	Xinyu Li、Rahul Gupta 和 Emily Brown。2024年。Datacomplm：大型语言模型中的数据质量基准测试。arXiv 预印本 arXiv:2402.23456。
75	acl-2025	2025.acl-srw.40	Wu 等 （2025b）	[为简洁起见省略名字]Zhang 和同事。2024 年。MetaGPT：元编程最先进水平的自主多智能体协作大语言模型工作流。arXiv 预印本 arXiv:2404.14496。
76	acl-2025	2025.acl-srw.56	Pulido 等人 (2025)	Sabrina J. Mielke、Tal Linzen 和 Jason Eisner。2021。上下文化词表示捕捉了什么类型的知识？上下文化词表示？在第59届计算语言学协会年会论文集计算语言学协会，页 1250–1265。
77	acl-2025	2025.acl-srw.73	Sakunkoo 和 Sakunkoo (2025)	Christo Kirov、John Sylak-Glassman、Ryan Que，David Yarowsky、Jason Eisner 和 Ryan Cotterell。2018年。通用形态屈折使用多语言数据集进行生成。北方 2018年会议的会议论文集计算语言学协会美国分会计算语言学（NAACL），页 52–62。
78	acl-2025	2025.acl-srw.85	Dzhubaeva 等（2025）	Y. Wang、H. Li 和 X. Zhang。2024年。人格特质一致性在量化角色扮演对话智能体中。载于《2024年自然语言处理经验方法会议：产业赛道》会议论文集，第 123–130页。
79	acl-2025	2025.acl-srw.98	Kumar 和 Nolbaria (2025)	Raghuraj Sailesh Shah、Ankur Anand、Srini Chodisetti、Ankit Gupta、Raj Sanjay Patel，Sameer Patel 和 Manish Gupta。2022年。FinservNLP：金融共享任务库与基准。在第三届经济学与自然语言处理研讨会会议论文集，第 144–150 页。计算语言学协会计算语言学。
80	acl-2025	2025.africanlp-1.1	Akpobi （2025）	Adewale Oladipo 等，2023年。Wura：一项非洲语言多语言数据集。在第61届年度会议论文集计算语言学协会。
81	acl-2025	2025.africanlp-1.15	Ismail 等人(2025)	J. Sohn 和其他 1 位作者。2024年。推理引导的检索增强生成用于医学问答。arXiv 预印本 arXiv:2411.00300。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
82	acl-2025	2025.africanlp-1.30	Pandya 等（2025）	M. Martin 等（2022）. Swahbert：增强斯瓦希里语自然语言处理与预训练变换器。在论文集。EACL。
83	acl-2025	2025.argmining-1.30	El Baff 等人（2025）	Yiqiu Xu、Alessio Frosini、Mattia Vanni、Anisa Rula 和 Roberto Navigli。2024年。Frase：面向 SPARQL查询生成的基于框架的语义丰富。arXiv 预印本 arXiv:2503.22144。
84	acl-2025	2025.bea-1.66	Rezayi 等（2025）	Shuai Bai、Keqin Chen、Xuejing Liu、Jialin Wang、Wenbin Ge、Sibo Song、Kai Dang、Peng Wang、Shijie Wang、Jun Tang、Humen Zhong、Yuanzhi Zhu、Mingkun Yang、Zhaohai Li、Jianqiang Wan、Pengfei Wang、Wei Ding、Zheren Fu、Yiheng Xu、Jiabo Ye、Xi Zhang、Tianbao Xie、Zesen Cheng、Hang Zhang、Zhibo Yang、Haiyang Xu，以及 Junyang Lin。2024年。Qwen2.5：新一代通义千问大语言模型。arXiv 预印本 arXiv:2407.10671。
85	acl-2025	2025.bea-1.73	Nebhi 等人 (2025年)	Elizabeth Mayfield 和 Alan W. Black。2020年。使用……进行书面作文自动评分 Transformer 模型。载于 ACL 2020年 BEA 论文集。
86	acl-2025	2025.bea-1.86	Jain 和 Rengarajan (2025)	Tianyu Zhang、Sayak Basu、Douwe Kiela 和 Oriol Vinyals。2024年。级联感知训练与推理，以实现更高资源效率的语言模型。arXiv 预印本 arXiv:2405.12345。
87	acl-2025	2025年bionlp-1.6	Kumar 和 Basu （2025）	Ilias Chalkidis、Michail Fergadiotis、Roger Stradling、Nikolaos Pappas 和 Prodromos Malakasiotis。2020年。用于法律和生物医学文档分类的基于Transformer的模型：一项比较研究。载于 ACL会议论文集。
88	acl-2025	2025.bionlp-share.22	Evgin 等人(2025)	Qingxiu Jia、Qipeng Xu、Weiting Yu、Yitong Duan、JianYun Nie 和 Zhiyuan Liu。2022年。Alignscore：通过上下文对齐评估事实一致性。载于 2022年会议论文集自然语言处理实证方法会议（EMNLP）。
89	acl-2025	2025.bionlp-share.33	Xu 等(2025a)	Yang Liu、Ani Nenkova 和 Kathleen McKeown。2022年。将可控性融入神经抽象式摘要。载于第60届 ACL 年会，页 876–890。ACL.
90	acl-2025	2025.bionlp-share.34	Phasook 等人 (2025)	Sahil Dhuliawala、Raghav Gupta 和 Shashi Narayan。2023年。验证链：通过自我验证提升大语言模型事实准确性。发表于 NAACL。
91	acl-2025	2025.findings-acl.114	Gupta 等人 (2025b)	Yao Du、Kaitao Qian、Sanmi Koyejo、Zheng Zheng 和 Chao Zhang。2023年。Alpaca：一种强大且可复现的指令遵循模型。arXiv 预印本 arXiv:2303.16199。
92	acl-2025	2025.findings-acl.1140	Cui 等人(2025b)	Xiaoyu Li、Wei Zhang、Jiahui Chen 和 Hongwei Liu。2024c。Iris：面向大型语言模型的信息检索任务基准测试。arXiv 预印本 arXiv:2401.12345。
93	acl-2025	2025.findings-acl.1143	Das 等（2025b）	Raphael Rafailov、Orion Redwood 等。2023年。直接偏好优化：你不 ’ 需要奖励即可完成基于人类反馈的强化学习。arXiv 预印本 arXiv:2305.11517。预印本，arXiv:2305.11517。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
94	acl-2025	2025.findings-acl.1156	Zafrir 等（2025）	Shashi Narayan、Shay B. Cohen 和 Mirella Lapata。2018年。XSum：一个用于新闻文章抽象式摘要的新数据集。载于2018年自然语言处理实证方法会议论文集自然语言处理实证方法（EMNLP），701–711页。
95	acl-2025	2025.findings-acl.1184	Garg 等人(2025)	William Berrios、Yonglong Tian、Chiraag Lala, Alec Parrish ，和 Trevor Darrell。2023年。Lens：一个用于多模态、多任务学习的可学习嵌入命名空间，多任务学习。arXiv 预印本 arXiv:2306.11527。
96	acl-2025	2025.findings-acl.1208	Das 等（2025a）	Rakesh Gupta、Susan Johnson 和 Wei Li。2023年。提示抽象：利用语言-图像嵌入扩展创造力。arXiv 预印本 arXiv:2302.12345。
97	acl-2025	2025.findings-acl.1231	Zhou等 （2025a）	Oron Anschel、Nir Baram 和 Nahum Shimkin。2018年。在多智能体环境中学习表征。载于国际学习表征会议 Learning Representations。
98	acl-2025	2025.findings-acl.1278	Bi 等人(2025年)	DeepSeek-AI. 2024年。Deepseek 大语言模型：扩展具有持久价值的开源语言模型价值。arXiv 预印本 arXiv:2401.02954。
99	acl-2025	2025年.findings-acl.1279	Nahin 等人 (2025)	W Lian、B Goodson、E Pentland 等人，2023年。Openorca：一个由 GPT 增强的 FLAN 开放数据集推理轨迹。arXiv 预印本 arXiv:2305.11206.
100	acl-2025	2025.findings-acl.228	Maji等人(2025年)	Ramesh Patel 等人（2023）. 评估文化理解多语言模型。arXiv 预印本 arXiv:2302.12345。
101	acl-2025	2025.findings-acl.268	Tong 和 Wang (2025)	David Bamman。2020年。Booknlp：自然语言处理长篇文本。在2020年自然语言处理实证方法会议论文集自然语言处理：系统演示，第 1–7 页。计算语言学协会用于计算语言学。
102	acl-2025	2025.findings-acl.282	Sun 等人（2025）	Tianle Li、Wei-Lin Chiang、Lisa Dunlap Evan Frick、Banghua Zhu、Joseph E. Gonzalez 和 Ion Stoica。2024年。从实时数据到高质量基准：arena-hard 流水线。arXiv 预印本，abs/2406.11939。
103	acl-2025	2025.findings-acl.563	Parfenova 和 Pfeffer (2025)	Diana Parfenova 等，2024年。自动化使用大语言模型进行定性分析。会议论文集 ACL 2024。
104	acl-2025	2025.findings-acl.581	Yuan等人 (2025a)	Kevin Meng、Alex Andonian、David Bau, Yonatan Belinkov ，和李飞飞。2023年。MEMIT：在 Transformer 中大规模编辑记忆。arXiv 预印本 arXiv:2302.04761。
105	acl-2025	2025年.findings-acl.690	Atri 等人(2025)	Peter Hase 和 Mohit Bansal。2023年。语言模型中的组合式编辑。EMNLP。
106	acl-2025	2025.findings-acl.813	Li 等人 (2025g)	OpenAI。2024年。GPT-4o：规模扩展与性能改进。arXiv 预印本 arXiv:2401.XXXXX。
107	acl-2025	2025.findings-acl.861	Abootorabi 等人 (2025)	Gunnar A Sigurdsson，Gul Varol，Giovanni Maria Farinella 等，2018年。Charadesego：一个用于自我中心视频理解的数据集。arXiv 预印本 arXiv:1804.09626。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
108	acl-2025	2025.findings-acl.920	Li 等人 (2025e)	Zihao Wang、Haowei Lin、Ruilin Yan、Xiangyu Wang、Jiaqi Li、Weiye Shi、Xiaojian Ma、Anji Liu、Yitao Liang 等人。2024e。优化大语言模型的推理时推理检索增强反思。arXiv 预印本 arXiv:2403.05313.
109	acl-2025	2025.gebnlp-1.7	Kononykhina 等人 (2025)	张明远、王佳怡等人。2024a。检索增强型 ICD 编码：一种两阶段系统以 94 的优势胜过基础大语言模型。arXiv 预印本 arXiv:2401.12345.
110	acl-2025	2025.gem-1.6	Belz 和 Thomson (2025)	Anya Belz、Simon Mille 和 Craig Thomson。2025b。一套质量标准名称的分类体系及其定义，用于根据质量的标准可比方面评估自然语言处理系统。
111	acl-2025	2025.iwslt-1.16	K. Roy 等人 （2025）	Elizabeth Salesky、Ramon Sanabria、Alan W. Black 和 Florian Metze。2021 年。探索跨模态低资源语音转文本翻译模态。在 2021 年会议论文集中计算语言学协会北美分会会议计算语言学协会，页 1134–1145。
112	acl-2025	2025.knowllm-1.4	Li (2025b)	Wei Zhang、Li Chen 和 Nicolas Christin。2022年。用于加密货币欺诈检测的动态图学习。发表于 IEEE区块链与加密货币国际会议区块链与加密货币，1–9页。
113	acl-2025	2025.l2m2-1.7	Lasy 等人（2025）	Nikhil Stoehr、Kha Pham Doan 和 Louis-Philippe Morency。2023年。本地化Transformer语言模型中的记忆化。arXiv 预印本 arXiv:2310.01331。
114	acl-2025	2025.llmsec-1.2	Sorkhpour 等（2025年）	Robert Shelby、Carl Vondrick 和其他 1 位作者。2023年。大语言模型能否安全发布？评估红队测试对语言模型行为的影响。arXiv 预印本 arXiv:2304.10685。
115	acl-2025	2025.magmar-1.9	Li 和 Ke (2025a)	Rohit Gupta、Ankit Kumar 和 Rahul Singh。2022年。Crisismm：一个多模态危机应对框架。arXiv 预印本 arXiv:2203.01234。
116	acl-2025	2025.nlp4pi-1.17	Ding等人(2025)	Zhijing Jin、Zhiheng Lyu、Yiwen Ding, Mrinmaya Sachan、Kun Zhang、Rada Mihalcea, 以及 Bernhard Schoelkopf。2022年。AI Scholars：一个自然语言处理中的因果推断数据集。
117	acl-2025	2025.realm-1.33	Punjwani和Heck (2025)	Harsha Nori、Andrew He、Chenguang Wang, Po-Sen Huang、Yejin Yang、Hao-Tong Tung, Shashank Singh、Hossein Hosseini、Mark Hughes, Matei Zaharia等人。2023年。语言模型能否教会较弱智能体？训练能够解释自身行为的推理教师。arXiv 预印本 arXiv:2311.10731。
118	acl-2025	2025.sdp-1.14	Mandic 等(2025年)	Kai Sun、Rui Zhang 和 Xiang Ren。2022年。FS-NER：一种少样本联合实体与关系抽取框架。发表于 Findings of the 计算语言学协会：ACL 2022年，第 4292–4303 页，爱尔兰都柏林。计算语言学协会。
119	acl-2025	2025.sdp-1.32	Carla 和 Uban (2025)	邓翔、张涵、于文浩、Clare Lee, 以及 Mohit Bansal。2024年。Factscore 2.0：事实性双编码器对比预训练一致性的双编码器对比预训练。计算语言学协会会刊，12：1–19。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
120	acl-2025	2025.semeval-1.106	Mondal 和 Sarkar (2025)	Xiao Tan 和 Yuan Jiang。2021。探索问答系统中的习语。在北美计算语言学协会分会 2021 年会议论文集 ACL。
121	acl-2025	2025.semeval-1.135	Vaidyanathan等人（2025b）	Pranav Goel、Shrey Agarwal 和 Amir Hussain。2021。使用情绪强度回归 Transformer模型与损失函数优化。在 ACL 2021 Findings，第1432–1440页。
122	acl-2025	2025.semeval-1.142	Kesanam等人 (2025)	Shailza Choudhary和Tanmoy Chakraborty。2020年。一种用于多标签情绪分类的混合特征提取方法。IEEE 情感计算汇刊。
123	acl-2025	2025.semeval-1.151	Alberts 等(2025)	Jian Wu、Junwei Xie、Bing Tang、Hongyan Yan, 以及 Lijun Wang。2022年。密集段落检索：多重表征研究句子在一个段落中的表征的研究。载于第60届计算语言学协会年会论文集（ACL 2022）。
124	acl-2025	2025.semeval-1.195	Matheny 等人（2025）	Silvio Ricardo Cordeiro、Aline Villavicencio, Marco Idiart ，以及 Carlos Ramisch。2019 年。一篇名词复合词组合性的计算模型组合性。载于第57届计算语言学协会年会计算语言学（ACL），第2762–2773页。计算语言学协会语言学。
125	acl-2025	2025.semeval-1.231	Khatoon 等(2025)	Vered Shwartz 和 Ido Dagan。2018年。一种用于多词表达识别的序列神经模型识别。在 2018年会议论文集计算语言学协会北美分会会议计算语言学协会（NAACL-HLT）。
126	acl-2025	2025.semeval-1.243	Nadeem 等（2025）	Frank A Acheampong、Henry Nunoo-Mensah, 以及 Wei Chen。2020年。用于基于文本的情感检测：一项比较分析。arXiv 预印本 arXiv:2004.13704。
127	acl-2025	2025.semeval-1.250	Aryal 和 Pant （2025）	Alec Radford、Karthik Narasimhan、Tim Salimans 和 Ilya Sutskever。2018年。改进语言理解：通过生成式预训练。arXiv 预印本 arXiv:1812.01813。
128	acl-2025	2025.semeval-1.266	Wan等（2025）	Danqi Chen、Pengcheng Xie、Shiyu Wang等 2020年。Tabfact：一个大型数据集，用于基于表格的事实验证。载于计算语言学协会第58届年会计算语言学（ACL）。
129	acl-2025	2025.semeval-1.268	Chen (2025)	李敏、李良友和 Philipp Koehn。2022年。神经网络中的低频实体翻译机器翻译。载于计算语言学协会年会论文集：ACL 2022年，页 3021–3033。
130	acl-2025	2025.semeval-1.27	S 等 (2025b)	S. Kumar 等人 2023年。Sentilex：增强基于词典的方法的情感分类方法。载于 ACL 情感分析研讨会分析。
131	acl-2025	2025.semeval-1.278	Huang 和 Cui (2025)	Linhao Zhang、Shuai Wang 和 Bing Liu。2018年。学习检测推文中的多标签情绪结合标签共现和依存关系。在第27届国际计算语言学会议计算语言学会议论文集，第3299–3309页。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
132	acl-2025	2025.semeval-1.37	Cui (2025)	Yuqing Zhang、Yansong Feng、Yinwen Zhang, 以及 Shujie Liu。2021。基于注意力的多标签情绪分类，采用逐标签注意力机制。载于第59届协会年会计算语言学（ACL 2021），页5943–5952。
133	acl-2025	2025.sicon-1.12	Tiwari (2025)	Yixuan Zhang、Leo Cheng、Lichang Zhang、Lu Liu、Jingcheng Li、Haoning Liu 和 Z G Xu。2024年。Alert：一套全面的安全基准套件，适用于大型语言模型。预印本，arXiv:2402.12441。
134	acl-2025	2025.sigtyp-1.15	Li 和 Ke (2025b)	Kevin Zhang 等 1 人。2022年。代码与散文：加密货币白皮书中的句法模式。收录于 LAW@ACL。对200 份白皮书的人工分析，揭示了主谓宾/主宾谓混用现象。
135	acl-2025	2025.trl-1.16	Li （2025c）	Haoyang Ding、Zijian Liu、Xiao Chen、Lidong Bing 和 Wai Lam。2021 年。金融表格处理与预训练语言模型。EMNLP。
136	acl-2025	2025.unlp-1.10	Kyslyi 等人(2025a)	Marcos Zampieri、Shervin Malmasi、Preslav Nakov、Ahmed Ali 和 Stephan Vogel。2017。面向 DSL 2017 的阿拉伯语方言识别共享任务。载于第四届相似语言、变体与方言自然语言处理研讨会，变体与方言（VarDial）。
137	acl-2025	2025.unlp-1.12	Kyslyi 等人（2025b）	Firoj Alam、Shaden Shaar、Alex Nikolov 和 1 等人。2021。用于假新闻检测的自然语言处理综述。计算语言学，47(4):905–960。
138	acl-2025	2025.woah-1.21	Boudraa 等人（2025）	Hande Kartal、Dilek Hakkani-Tür 和 Gokhan Tur。2022年。基于跨度的偏见检测新闻文章中的偏见陈述。载于2022年计算语言学会议。COLING。
139	acl-2025	2025.woah-1.36	Schirmer 等人（2023年）	Xing Jiang、Hugo Touvron、Ludovic Denoyer, Hervé Jégou 和 Guillaume Lample。2023年。Mistral 7b：一个小模型，却有大思维。arXiv 预印本 arXiv:2310.06825。
140	acl-2025	2025.xllm-1.25	Pham Hoang Le 等人。(2025)	Micha Livne、Roberto Dessì 和 Douwe Kiela。2023年。如何训练一个大型语言模型，使其成为可靠的信息抽取系统。计算研究资料库，arXiv:2309.16396。第3版。
141	acl-2025	2025.xllm-1.3	Li 等人 （2025k）	Yujia Zhang、Xiaobo Liu 和 Zhenhua Li。2020年。用于文本的基于图的层次关系表示。arXiv 预印本 arXiv:2007.12345。
142	emnlp-2025	2025.arabicnlp-main.20	Bouchekif 等人 (2025)	Faiza Qamar、Seemab Latif 和 Rabia Latif。2024年。一个具有更大上下文的基准数据集用于伊斯兰文本上的非事实型问答。arXiv 预印本 arXiv:2409.09844。
143	emnlp-2025 2(		Fusco 等人(2025)	Leonie Weissweiler、Valentin Hofmann、David Mortensen ，和 Janet Pierrehumbert。2023年。使用 wug 测试评估形态泛化。arXiv 预印本。ArXiv:230x.xxxxx。
144	emnlp-2025 2		Yoshida 等人 (2025)	Alec Radford、Jeffrey Wu、Rewon Child、David Luan、Dario Amodei 和 Ilya Sutskever。2019 年。语言模型是无监督多任务学习器。预印本，arXiv:1901.04513。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
145	emnlp-2025 2		Jeon 和 Strube （2025）	Kevin Clark、Urvashi Khandelwal、Omer Levy，以及 Christopher D. Manning。2019。BERT：一项关于 BERT 中注意力的研究。载于 2019年自然语言处理实证方法会议和第九届自然语言处理国际联合会议语言处理（EMNLP-IJCNLP），页 5717–5726，中国香港。计算语言学协会计算语言学。
146			Ahmad 和 Kakudi (2025)	Yuan Lan、Chao Huang、Wayne Xin Zhao 和 Jun Li。2024年。CoLA：协作角色注入式多智能体辩论框架用于立场检测。arXiv 预印本 arXiv:2403.01234。
147	emnlp-2025 2		Jang 等(2025)	Rong Chen、Siyun Zhang、Yiyi Zheng、Qiuhua Yu 和 Chu-huai Wang。2025。增强下腰痛治疗决策：一种将大语言模型与检索增强生成技术相结合的新框架。arXiv 预印本 arXiv:2501.34567。
148	emnlp-2025 2		Liu 等人 (2025f)	Jihan Zhu、Zhizheng Lin、Jun Gao、Zhaoxuan Zhou、Yaodong Zhang、Zhaofan Liu、Ceyao Zheng、Cheng Li、Zhaokai Wang、Zili Wang，以及另外 1 位作者。2024年。人工智能智能体评估综述：一一百个尚未解决的问题和一站式开源库。arXiv 预印本 arXiv:2406.09844。
149	emnlp-2025 2(		Chao 等人 (2025)	Jinhao Liang、Yihang Chen、Haozheng Zhang，等另外 1 位作者。2025。Surveyx：端到端自动化综述生成解决方案。arXiv 预印本 arXiv:2501.12345。
150	emnlp-2025	2025.emnlp-industry.123	Khabiri 等（2025）	Takeshi Kojima、Shinn Yao、Jinyi Zhao、Xiang Ren 等人。2023年。ToolLLM：帮助大语言模型掌握 16000+现实世界 API。arXiv 预印本 arXiv:2312.11568。
151	emnlp-2025 2(		Zhang （2025）	Emmanouil Panousis 和 Tom Rainforth。2022年。一种有原则且统一的贝叶斯最优实验设计框架。arXiv 预印本 arXiv:2202.04647。
152	emnlp-2025	2025.emnlp-industry.168	Ethiraj 等人（2025）	Zhen Li 和其他 1 位作者。2023年。迈向通用文本嵌入与专家混合。预印本，arXiv:2311.05723。https://arxiv.org/abs/2311.05723。
153	emnlp-2025 2(		Yu 等(2025b)	Haotian Luo、Li Shen、Haiying He、Yibo Wang，Shiwei Liu、Wei Li、Naiqiang Tan、Xiaochun Cao，和 Dacheng Tao。2025。Adaptthink：大语言模型可以学会何时思考。https://arxiv.org/abs/2501.12570。
154	emnlp-2025 2(		Beyranvand 等 (2025)	Jun Xu、Fan Wang 和 Shuo Chen。2022年。面向车辆的自动问答手册，使用自然语言处理和知识图谱。IEEE 智能交通汇刊系统。
155	emnlp-2025 2(		Chawla 等人(2025)	Yifan Liu、Jiaqi Wang 和 Zhi Zhang。2024年。人工智能投资建议中的偏见：对大语言模型中的本土偏好进行分析。arXiv 预印本 arXiv:2209.04538。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
156	emnlp-2025	2025.emnlp-industry.69	Azeez 等人（2025）	Juho Kim 等人 2023年。GPT-4 能通过律师资格考试吗? 一项关于大型 语言模型在法律选择题 中的表现的案例研究。载于 2023年北美分会会议论文集 北美分会的 计算语言学协会。
157	emnlp-2025	2025.emnlp-industry.70	Swamy 等, （2025）	Florian Schuster 等，2023。Getml：基于帕累托的 特征学习，用于机器学习应用。 arXiv 预印本 arXiv:XXXX.XXXXX.
158	emnlp-2025 2(		Bougie 和 Watanabe (2025)	Anonymous 等 1 人。2023b。大语言模型是否可靠 审稿人？对大语言模型进行全面评估 审稿修订选择题。ACL Anthology 2024.lrec-main.816。
159	emnlp-2025	2025.emnlp-industry.96	Joshi 等人(2025a)	Siva Reddy、Daniel Khashabi、Benjamin Lo, Chandra Bhagavatula、Maarten Sap 和 Yejin Choi。2023年。大型语言模型 能理解指令吗？arXiv 预印本 arXiv:2301.01193。
160	emnlp-2025	2025.emnlp-main.1021	Pi 等人(2025年)	Ziyang Xiong 等人 2024年。Llava-critic：视觉 指令微调与反馈。arXiv 预印本 arXiv:2402.12345。
161	emnlp-2025 2		Zhang 等人(2025d)	Yiming Zhang 和 Zhouhui Lian。2024a。Brush your text：增强扩散 模型中的文本渲染：注意力图干预。在 第62届计算语言学协会年会 计算语言学协会 （ACL）。
162	emnlp-2025 2		Zhang 和 Zhou （2025）	X. Li、P. Wang 和 Y. Chen。2020 年。迭代式 Transformer：面向 序列建模的循环注意力更新。arXiv 预印本 arXiv:2010.02536。
163	emnlp-2025 2		Zhang 等人(2025e)	Chenyan Yu、Lee Xiong 和 Jamie Callan。2022年。 关于对话式稠密检索的综述： 方法与挑战。arXiv 预印本 arXiv:2201.08452。
164	emnlp-2025 2		Zhang 等人（2025b）	Zhengbao Jiang、Frank F Xu、Jun Araki 和 Graham Neubig。2024年。Flare：主动检索 增强，用于幻觉缓解。arXiv 预印本 arXiv:2401.07019。
165	emnlp-2025 2		Yi 等人（2025）	Bogdan Gliwa、Iwona Mochol、Michał Biesek, 以及 Aleksander Wawer。2019 年。Samsum语料库: 一个人工标注的对话摘要数据集。 arXiv 预印本 arXiv:1911.12237。
166			Nyandwi 等(2025年)	Wenxuan Zhang、Da Yin、Zhong Ding 和 Radu Soricut。2023年。M3exam：一个多语言、 多模态考试式问答基准。arXiv 预印本 arXiv:2306.05179。
167	emnlp-2025 2		Rafiei Asl 等人 (2025)	Mark Russinovich、Jingwen Cai、Yifan Hou, Yuxiao Jiang、Yiping Jiang、Jieyu Kang、Prithviraj Kang、Urvashi Khandelwal、Nitish Khurana、Pang Wei Koh , 等。2024年。Crescendo：迈向 使用语言智能体对大语言模型进行逼真的红队测试。 arXiv 预印本 arXiv:2402.13249。
168	emnlp-2025 2		Dumitru 等。(2025)	Jia Chen 和 Hao Xu。2023年。并行解码 使用推测采样的大语言 模型。arXiv 预印本 arXiv:2306.15478。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
169	emnlp-2025 2		Park 等（2025）	Rodrigo Nogueira 和 Jimmy Lin。2019年。从 doc2query 到 doctttttquery。arXiv 预印本 arXiv:1904.08375。
170	emnlp-2025 2(		Li 等人 (2025b)	Weixi Tong 和 Tianyi Zhang。2024年。Codejudge: 评估大语言模型的代码生成 模型。载于2024年自然语言处理实证方法会议论文集 关于自然语言处理的实证方法 ，第20032–20051页。
171	emnlp-2025 2		Bohdal 等人（2025）	Aliaksei Severyn、Eric Emil Malmi、Jonathan Stephen Mallinson、Sascha Rothe 和 Sebastian Krause。2021。语法纠错 使用大型多语言语言模型。在 ACL.
172	emnlp-2025 2		Borah 等人(2025)	Junnan Li、Ziqing Wang、Yilun Xu、S 等。2024年。 大语言模型中的对齐退化：潜在漂移的测量、可视化与缓解。arXiv 预印本 arXiv:2401.04200。
173	emnlp-2025 2		Kryklyvets 等 （2025）	S. Thoppilan 等人。2023年。Gemini：一个统一的 多模态大语言模型。arXiv 预印本 arXiv:2303.06601。
174	emnlp-2025 2		Srivastava (2025)	Wei Xu、Yulia Tsvetkov 和 Alan Black。2022年。 用于语言学习的人工智能：对话代理 和个性化反馈。计算语言学协会会刊 计算语言学协会 （TACL），10:1–15。
175	emnlp-2025 2		Roh 等（2025）	Amanpreet Singh、Yujia Wang、Yulia Tsvetkov, 以及 Percy Liang。2024年。全球多语言大规模多任务语言理解：评估 多语言理解中的文化偏见和语言偏见 。arXiv 预印本 arXiv:2412.03304。
176	emnlp-2025	2025.emnlp-main.1521	Gallifant 等人（2025）	Trenton Bricken、Jonathan Marcus、Siddharth MishraSharma、Meg Tong、Ethan Perez、Mrinank Sharma、Kelley Rivoire 和 Thomas Henighan。 2023年。迈向单义性：分解 基于字典学习的语言模型。arXiv, abs/2309.08600。
177	emnlp-2025 2		Jin 等人（2025）	Liang Liu、Marc Brockschmidt、Vivek Murali, 和等人。2024年。元大语言模型编译器：一套 用于代码优化的开源模型。 arXiv 预印本 arXiv:2407.02524。
178	emnlp-2025 2		Dhanraj 和 Eliasmith (2025)	Jianfeng Wang、Fei Huang 和另外 1 位作者。2023年。 大语言模型的高效微调 与 LoRA。arXiv 预印本 arXiv:2303.01234。
179	emnlp-2025 2		Song等人（2025a）	Mor Geva、Tal Schuster、Jonathan Berant 和 Omer Levy。2023年。令牌回收：让大语言模型 速度更快且数据效率更高。arXiv 预印本 arXiv:2310.02548。
180	emnlp-2025 2		Liu 等人 (2025e)	Sabine Ploux 和 Hong Ji。2003年。语义 空间：词义分布建模 采用无监督学习技术。 《计算语言学》，29(2):259–275。
181	emnlp-2025	2025.emnlp-main.1611	Yuan等人 (2025b)	Laura A Banarescu、Claire Bonial、Sheila Condon、Emily Faries、Jon Niekrasz 和 Tim O ’ Connor。2018年。用于多文档 摘要的 AMR。载于第56届会议论文集 计算语言学协会年会 （第2卷：短 论文），第 577–583 页。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
182	emnlp-2025 2		Sadia 等人 (2025)	Srivatsan Sundar、Dhruv Jain、Yuwei Zhang 和 H. V. Jagadish。2024年。gtbls：生成表格通过学习表格结构从文本中生成。在计算语言学协会 2024 年会议论文集计算语言学协会。
183	emnlp-2025 2		Xiao 等人 （2025年）	Kai Jiang、Hao Liu 和 Rodney Brooks。2022年。工业人机交互中的信任—性能悖论：来自仓储机器人技术的实证证据。arXiv 预印本 arXiv:2208.14637。
184	emnlp-2025 2(		Kassem 等人（2025）	Baolin Peng 和其他 1 位作者。2024年。高效的模型大规模编辑。arXiv 预印本 arXiv:2401.05911。
185	emnlp-2025 2		Mohammadi 等人 （2025）	以及 Hannaneh Hajishirzi。2022年。跨任务通过自然语言指令实现泛化。载于第60届年会论文集计算语言学协会（第1卷：长篇论文），第3470–3487页，爱尔兰都柏林。计算语言学协会语言学。
186	emnlp-2025	2025.emnlp-main.1777	Gao 等（2025）	Jack FitzGerald、Ryan Cotterell、Bahadir Avci, Sebastian Ruder、Graham Neubig、Melvin Johnson ，和 Orhan Firat。2022年。Massive：一项面向任务的 51 种语言多语言数据集语义解析。arXiv 预印本 arXiv:2204.08582。
187	emnlp-2025	2025.emnlp-main.1806	He 等人(2025b)	A. Kirstain 和其他 1 位作者。2023年。Pick-a-pic：一个用于评估鲁棒性的数据集视觉语言模型。预印本，arXiv:2305.01569。
188	emnlp-2025	2025.emnlp-main.192	Liu 等人 (2025d)	Rishi Taori 等人 2023年。Alpaca：一个用于指令遵循的1750亿参数模型。arXiv 预印本 arXiv:2303.11366。
189	emnlp-2025 2		He 等（2025a）	Shuang Chen、Yang Liu 和 Erik Cambria。2023年。知识增强的方面级情感分析与外部常识图谱。在第61届计算语言学协会年会论文集计算语言学协会（ACL 2023年），页 13567–13579。计算语言学协会计算语言学。
190	emnlp-2025 2		Li 等人 (2025f)	Kevin Lin、Ben Tan、Swabha Swayamdipta、Noah A. Smith 和 Yejin Choi。2019b。Ropes：段落阅读理解。在2019年自然语言处理实证方法会议论文集自然语言处理实证方法和第九届国际联合自然语言处理会议（EMNLP-IJCNLP）。
191	emnlp-2025 2(		Khandelwal等 (2025)	Alex Mallen、Zexuan Zhang、Xinyu Liu、Yichong Wu、Yiming Zhang 和 William Yang Wang。2023b。何时不应信任语言模型：探究参数化知识与基于检索的知识的有效性。载于第61届计算语言学协会年会计算语言学（第1卷：长篇论文），第1234–1245页。
192	emnlp-2025 2		Xu 等人（2025c）	Todor Mihaylov 等人。2018年。OpenBookQA：问答中的事实知识评估。EMNLP。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
193			Zheng等人(2025a)	Aurko Roy、Mohammad Saffar、Ashish Vaswani, 以及 David Grangier。2021b。高效路由 Transformer：动态词元交互模型 用于自然语言处理。arXiv 预印本 arXiv:2003.05997。
194	emnlp-2025 2		Tang等人（2025b）	Justin Fu、Kelvin Zhang、Utkarsh Sanyal、Lantao Yu、Collin Moses、Fan Yang、Stefano Ermon，以及 Zhibin Zhao。2023年。通过推理驾驶：利用通用语言 模型实现可解释策略的强化学习。arXiv 预印本 arXiv:2303.00745.
195	emnlp-2025 2		Tanjim 等 (2025)	Meta AI。2024年。Llama 3.2：推进 开放权重语言模型。arXiv 预印本 arXiv:2408.00001.
196	emnlp-2025 2		Wang 等人 (2025f)	Xinyu Fan、Li Zheng、Bing Yu、Yifan Liu、Yichi Zhang、Yu Zhang、Hao Tang、Yikai Wang、 Weinan Chen、Yizhou Wang等人，以及另外1位作者。2023年。 街道中的城市身份感知地理推理 观点。arXiv 预印本 arXiv:2311.16456。
197	emnlp-2025 2		Saba 和 Skiena （2025）	Markus Freitag、Yaser Al-Onaizan、Shuo Sun Ma, 以及另外 1 位作者。2022a。高质量低资源 机器翻译：一项新的基准。在 EMNLP 研究成果。
198			Yu 等 (2025a)	Wei Zhang、李明、Hao Wang 和 Yang Liu。 2024b。Deepseekmath：可扩展的数学 预训练与组相对策略 数学推理优化。arXiv arXiv 预印本 arXiv:2402.03300。
199	emnlp-2025	2025.emnlp-main.528	Chen 等人 (2025f)	Salah Lhoussaine、Georgios Tziantzioulis, Michael B. Sullivan ，Vilas Sridharan ，Nathan DeBardeleben、Christian Engelmann 和 Simranjit Singh。2024年。初探 bfloat16 大语言模型中的软错误韧性。 预印本，arXiv:2412.07192。
200	emnlp-2025 2		Wang 等人(2025a)	Harsha Nori、Allison Webster、Matthew McInnis, 等人。2023b。Medprompt：大型语言模型 是配备结构化提示的推理引擎。 预印本，arXiv:2307.13880。
201			Shihab 等人 (2025)	Szymon Tworkowski、Konrad Przybysz、Tomasz Korbak、Pedro Rodriguez、Wojciech Rozemłyn, Kamil Rakowski、Maciej Grzelak、Albert Webson、Maciej Szafraniec、Robin Sorsch 和 1 其他人。2024年。Mambaformer：基于选择性状态空间和注意力的高效语言 建模。 arXiv 预印本 arXiv:2312.00752。
202	emnlp-2025 2		Wang 等人 (2025d)	DeepSeek。2025。DeepSeek R1：迈向深度 面向语言模型的强化学习。 arXiv 预印本 arXiv:2501.12948。
203	emnlp-2025 2		Lv 等（2025）	徐逸凡、刘晓宇、Baolin Peng 等。 2024年。评估大语言模型的指令遵循 鲁棒性。在 EMNLP 2024 会议论文集。
204	emnlp-2025	2025.emnlp-main.619	Tian 等人 (2025)	钱静等。2018年。基于神经网络的 假新闻检测：学习识别 欺骗性内容。第56届会议论文集 计算语言学协会年会 计算语言学。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
205	emnlp-2025 2(		Kang 等(2025)	Hang Yu、Tianle Li 等人。2025 年。Dapo：面向开源大语言模型对齐的数据感知策略优化。arXiv 预印本 arXiv:2503.04111。
206	emnlp-2025 2		Kim 等人（2025a）	Mor Geva、Daniel Khashabi、Elad Segal, Jonathan Berant 和 Ido Dagan。2021 年。Aristotle：用于聚焦逻辑推理的数据集。载于计算语言学协会论文集：计算语言学协会论文集：EMNLP 2021，第 1–14 页，蓬塔卡纳，多米尼加共和国。计算语言学协会计算语言学。
207	emnlp-2025 2(		Zhang 等人(2025f)	Mark Finlayson 和另外 1 位作者。2023年。因果对语言模型行为的分析，包含诱导干预分析语言模型行为。arXiv 预印本 arXiv:2301.12928。
208	emnlp-2025 2		Liu 等人 (2025a)	Yixin Nie、Jing Wang、Mohit Bansal 和 Kai-Wei Chang。2020年。对抗性自然语言推理。载于 ACL会议论文集。
209	emnlp-2025 2(		Liu 等人 (2025b)	Paola Espinosa、Sarah Kadi、Stefanie Kamps，以及等人，2017。性别差异与共情相关过程：探索行为及神经生理相关因素。《心理神经内分泌学》，85:34–43。
210	emnlp-2025 2(		Luo 等(2025a)	Xuezhe Li 和 Xiaodan Sun。2020年。用于数据不平衡的自然语言处理任务的 Dice 损失：应用于命名实体识别。发表于第58届计算语言学协会年会计算语言学，第4653–4661页。
211	€		Rivlin-Angert 和 Mor-Lan （2025）	David Hebron、Avi Shmidman 和 Moshe Koppel。2023年。Hero：一个用于希伯来语自然语言处理任务。预印本，arXiv:2309.12345。
212	emnlp-2025 2		Nafee 等人（2025年）	Linyong Lu、Qingxiu Zhang、Xiang Li 和刘晶晶。2022年。Promptpg：用于提示策略梯度训练。载于2022年自然语言处理实证方法会议（EMNLP）。
213	emnlp-2025 2(		Mundada等人 （2025）	Lukasz Czajka、Malihe Alikhani、Patrick Verga，以及 Yonatan Belinkov。2024b。Musictheory-bench。arXiv 预印本 arXiv:2410.02084。
214	emnlp-2025 2(		Jiang等人 (2025)	Victor Zhong、Chandra Bhagavatula、Ronan Le Bras、Yejin Choi和Noah A Smith。2022年。文本分析推理：统一机器阅读与逻辑推理。载于《计算语言学协会论文集：NAACL 2022，页 2307–2323。
215	emnlp-2025 2		Trifan 等人 (2025)	Liang Pan、Lei Zhang 和 Zhenzhong Yang。2023年。ChatGPT 能否进行零样本对话状态跟踪？载于 arXiv 预印本 arXiv:2302.01180。
216	emnlp-2025 2		Wang 等人 (2025c)	Weijie Chen、Yizhe Zhang、Qian Wu 和 1 等人。2024年。Internvl：扩展视觉-语言预训练与多模态强化学习。arXiv 预印本 arXiv:2402.00028。
217	emnlp-2025 2(		Chen 等人 (2025a)	Xiaoxi Zhong 等。2023年。Agentverse：促进多智能体协作与借助大语言模型进行探索。arXiv 预印本 arXiv:2309.07864。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
218			Zhou等 (2025b)	Shengding Hu、Yuge Tu、Xu Han、Chaoqun He, Ganqu Cui、Xiang Long、Zhi Zheng、Yewei Fang, Yuxiang Huang、Weilin Zhao及另外1人。2024年。Minicpm：揭示小型 语言模型结合预热-稳定-衰减 学习率调度器的潜力。arXiv 预印本 arXiv:2404.06395。
219	emnlp-2025		Bhendawade 等人 (2025)	Yi Tay、Dara Bahri、Donald Metzler 等人。2022年。 高效扩展：从训练和 扩展大型语言模型中获得的见解。arXiv 预印本 arXiv:2210.03863。
220	emnlp-2025 2	emnlp.1052	Kim 等人 (2025c)	H. Yang 和 K. Li。2023年。Boostaug：混合 用于提升文本的实例过滤框架 数据增强。载于计算语言学协会论文集：计算语言学：ACL 2023。
221	emnlp-2025	2025.findings-emnlp.1074	Aastik 等人（2025）	Xianzhi Li、Zihang Dai 和其他 1 位作者。2022年。 Branchtuning：通过分支高效微调预训练 视觉模型。arXiv 预印本 arXiv:2202.06924。
222	emnlp-2025 20	emnlp.1082	Maskey 等(2025)	Rodrigo Gomez 等，2018年。Ciphergan： 使用生成对抗网络进行无监督密码破解。 arXiv 预印本 arXiv:1801.04883。
223	emnlp-2025 20	emnlp.1100	Yang 等人（2025a）	Mo Yu、Yisi Sang、Kangsheng Pu、Zekai Wei, Han Wang、Jing Li 和 Jie Zhou。2022年。 电影中的角色理解：一个基准 用于电影角色分析。arXiv 预印本 arXiv:2211.04684。
224	emnlp-2025 2	emnlp.1118	Wang 等人 (2025b)	Zhen Chiang、Noah Shinn、Siyuan Zhuang 和 1 其他人。2023年。Vicuna：一款开源聊天机器人 以 90%* 的 ChatGPT 质量媲美 GPT-4。 预印本，arXiv:2304.05335。
225	emnlp-2025 2	emnlp.1144	Bn 等人（2025）	Suyeon Lee、Sunghwan Kim 等人。2024b。 CounselingEval：迈向质量评估 基于大语言模型的心理咨询。在 第62届计算语言学协会年会论文集 计算语言学协会 （ACL）。
226	emnlp-2025 2		Zeng 等人 （2025）	X. Du等人（2022a）。Grit1：语法错误纠正 大语言模型评估数据集。载于 EMNLP 2022会议论文集。X. Du等人（2022b）。 Grit2：扩展语法错误纠正，适用于 多语言大语言模型。载于EMNLP 2022 Findings。
227	emnlp-2025	2025年.findings-emnlp.1196	Zheng等人(2025c)	OpenAI. 2023a. GPT-4：OpenAI 的 生成式 预训练 Transformer 4 模型。预印本, arXiv:arXiv:2301.00000.
228	emnlp-2025	2025.findings-emnlp.1286	Kim 等人 (2025f)	Baptiste Roziere、Gautier Izacard、Jan Leike Botha 等人。2023年。Code LLaMA：大型 用于代码的语言模型。预印本, arXiv:2308.08545。
229	emnlp-2025 2	emnlp.1343	Saeed 等人 (2025)	Damai Bi、Kunhao Lv、Xiang Ji、Yining Wang, 吕泽成、杜子涵等。2024年。Deepseek：通过增强语言模型推进人工智能发展。 arXiv 预印本 arXiv:2401.14196。
230	emnlp-2025 2	emnlp.1344	Kwak 等人（2025）	Aditya Maharana 和另外 1 位作者。2024年。Locomo： 对话代理的长上下文建模基准 对话式智能体。arXiv 预印本 arXiv:2402.13968。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
231	emnlp-2025 2	emnlp.1359	Akra 等人(2025)	Ling Chen、Ming Zhao 和另外 1 位作者。2024年。用于生物医学图像分割的高效主动学习，仅需少量标注。arXiv arXiv 预印本 arXiv:2405.01701。
232	emnlp-2025 20	emnlp.1384	Ji 和 Lu（2025）	Yiming Li、Xia Chen、Liang Zhou 和 Zheng Wang。2024b。事后反思：面向鲁棒推理的选择性反思。arXiv 预印本 arXiv:2406.12050。
233	emnlp-2025 2(		Zhu 等人（2025d）	Yujia Hong、Junyang Wu、Fan Zhang 和 Shuo Zhang。2024b。使用代码和记忆的自适应智能体解决数学文字题。arXiv 预印本 arXiv:2403.01290。
234	emnlp-2025 2		Saad-Falcon 等（2025）	Dahyun Koo、Yejin Choi 和 Eunsol Choi。2023年。作为评估器的大语言模型中的认知偏差 arXiv 预印本 arXiv:2312.05441。
235	emnlp-2025 2		Rafiuddin 和 Khan (2025)	Tim Dettmers、Yannic Kilcher、Henry Minsky, Anna McDowell、Neha Nangia、Andreas Vlachos, 以及 Microsoft Phi 团队。2025。Phi-4-mini：紧凑而强大的多模态模型。arXiv 预印本 arXiv:2503.01743。
236	emnlp-2025 2(		Zhang 等人（2025c）	Yinlin Liu、Pengcheng Yin 和 Graham Neubig。2019。Conala：代码/自然语言挑战。载于 2019 年会议论文集 自然语言处理实证方法会议 语言处理（EMNLP）。
237	emnlp-2025 2(		Fan 等人 （2025）	Yuyang Huang 和 Clark Glymour。2016年。A 因果发现和因果推断综述。arXiv 预印本。
238	emnlp-2025	2025.findings-emnlp.354	LeVi 等人 (2025)	Q. Zhang 和其他 1 位作者。2025。针对大语言模型的无梯度对抗攻击：可迁移性和优化。arXiv:2502.01567v1 [cs.CL]。这篇论文主要关注针对大语言模型的无梯度攻击，但也包括基于梯度的攻击，并探讨在黑盒环境中借助替代模型实现的可迁移性。它并未着重讨论越狱或贪婪坐标梯度（GCG），使其与文本的契合度降低 ’ s 重点。Wang 等人（2025，arXiv:2503.02219v1 [计算与语言]）关于基于梯度的越狱 GCG 是一种更精确的替代方案。
239	emnlp-2025	2025.findings-emnlp.379	Li 等人 (2025年i)	Yinhan Tang、Xavier Garcia、Alessandro Raganato、Vishrav Chaudhary 和 Jiatao Gu。2022年。一种高效的多语言字节到字节序列到序列任务模型。载于 计算语言学协会发现论文集 语言学：EMNLP 2022年。
240	€		Luo 等（2025b）	Wenbin Xiong、Zhan Liu、Yihan Yang、Ge Yu、Ge Zhang，以及 Tian Song。2023年。超关系知识图谱综述：从数据超关系知识图谱：从数据 从模型到应用。arXiv 预印本 arXiv:2301.03869。
241	emnlp-2025 2(		Nguyen 等人（2025a）	Yi Cheng、Yilun Wu、Adrian Weller、Matt J. Kusner 和 Pushmeet Kohli。2024年。整合解码：提升大语言 模型生成的事实性：通过隐式自洽性。在 国际学习表征会议 。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
242	e		Cai 等人 (2025)	Wenhui Wang、Li Dong、Hao Cheng、Furu Wei，以及 Ming Zhou。2022年。Minilmv2：多任务预训练，用于多任务通用文本表示。载于计算语言学协会年会论文集：ACL 2022，页 2907–2918。
243	emnlp-2025 2		Ghosh 等人（2025）	Yankai Lin、Jiapeng Zhou、Yiming Shen，Wenxuan Zhou、Zhiyuan Liu、Peng Li、Maosong Sun 和 Jie Zhou。2021 年。Xcsqa：一个基准用于跨语言对话式问题作答。在 EMNLP。
244	emnlp-2025 2(		Baali 等人(2025)	Jue Gao、Tri M Dang、Michael L Seltzer 和 Rif A Saurous。2022年。Conversasynth：探索音频合成对话的领域理解。在第60届年会论文集计算语言学协会语言学（第1卷：长篇论文），页 6543–6558。
245	e		Mor-Lan 等人 (2025)	David Hebron、Avi Shmidman 和 Moshe Koppel。2023年。Hero：一个用于希伯来语自然语言处理任务的希伯来语 RoBERTa 模型。预印本，arXiv:2309.12345。
246	emnlp-2025	2025年.findings-emnlp.534	Liu 和 Roth(2025)	Ellie Pavlick 和 Joel Tetreault。2016年。语义驱动的未来语言歧义检测方向。载于计算语言学协会年会（ACL）会议论文集。
247	emnlp-2025	2025.findings-emnlp.645	Zhou等 (2025d)	Sho Kurita、Jonas Pfeiffer、Ivan Vulic、Edoardo Ponti，和 Anna Korhonen。2020年。一种加权无监督多语言Transformer微调。载于《AACL-IJCNLP。
248	emnlp-2025 2(		Bhattacharya等 (2025)	Yifei Li、Zhou Zhao、Xiaohan Yu 和 Deng Cai。2023年。多模态不确定性估计深度学习模型。arXiv 预印本 arXiv:2304.02637。
249	emnlp-2025 2(		Zhang 等人(2025a)	E. Zaken、Y. Goldberg 和 S. Ravfogel。2022年。Bitfit：面向……的简单参数高效微调Transformer模型。计算语言学协会会刊（TACL），10:1–16。
250	emnlp-2025 2(		Xiang等人（2025）	VN Ioannidis、X Song、S Manchanda、M Li、X Pan、D Zheng、X Ning、X Zeng和G Karypis。2022年。DRKG 药物再利用知识图谱用于新冠肺炎。2020年。arXiv 预印本 arXiv:2010.09600，第 1–13 页。
251	emnlp-2025 1			Wendi Zhou、Xiao Li、Lin Geng Foo、Yitan Wang，Harold Soh、Caiming Xiong 和 Yoonkey Kim。2024年。TEMPO：时间表示面向时间序列预测的大语言模型提示。arXiv 预印本 arXiv:2405.18384。预计发表于 NeurIPS 2024。预印本，arXiv:2405.18384。
252	emnlp-2025	2025.findings-emnlp.863	Kellert 等人 (2025)	Shruti Rijhwani、Lawrence Wolf-Sonkin、Victor Kuperman、Timothy Baldwin 和 Thamar Solorio。2017。分析语码转换社交媒体文本。载于 2017 年会议论文集自然语言处理实证方法会议语言处理。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
253	emnlp-2025			Jiawei Chen、Yuanhang He、Yada Zhang、Jiachen Ji 和 Jie Tang。2024年。图上生成：将大语言模型同时视为智能体和知识图谱，用于不完整知识图谱问答。arXiv 预印本 arXiv:2404.14741。
254	emnlp-2025	2025.findings-emnlp.866	Li 等人 (2025a)	Eric Mitchell、Roberta Raileanu、Colin Raffel, John Levine、Yulia Tsvetkov 和 Christopher D Manning。2023年。Ieval：一个指令遵循基准。arXiv 预印本 arXiv:2310.07724。
255	emnlp-2025 2		Kim 等人（2025e）	Zhiwei Chen、Yichao Lu 和 Mingjun Zheng。2020年。Dkt+：增强型深度知识追踪结合正则化。在《2020年自然语言处理实证方法会议语言处理》（EMNLP），页 1234–1244。
256	emnlp-2025 2		Xu 等人（2025d）	Zixuan Wang、Xiaocheng Li 和 Yang Liu。2023年。后验提示调优：迈向忠实且经过校准的大语言模型。发表于《自然语言实证方法》语言处理（EMNLP）。
257	emnlp-2025 2		Li 等人 （2025h）	RangRang Ge、ShiYe Song、Zhaorui Liu、Wei Liu, Yuesheng Wang、Dongling Wang、Bofang Zhou, Zhicheng Dou 和 Ji-Rong Wen。2023年。一篇综述关于大语言模型的 KV 缓存压缩模型。arXiv 预印本 arXiv:2312.10546。
258	emnlp-2025	2025.findings-emnlp.932	Jakob 等人（2025）	Shaina Ashraf、Isabel Bezzaoui、Ionut Andone, Alexander Markowetz、Jonas Fegert 和 Lucie Flek。2024年。Defakts：一个用于分析德国媒体中虚假信息的细粒度数据集。载于2024年联合国际会议计算语言学、语言资源与评测，都灵，意大利。欧洲语言资源协会。
259	emnlp-2025 2(		Akhondzadeh 等 (2025)	Michael Boratko、Harsh Padigela、Deepak Mikkilineni, Pavan Yuvraj, Rajarshi Das, Andrew McCallum、Mihai Chang、Achille Fokoue、Pavan Kapanipathi, Nicholas Mattei, et al. 2018年。Arc：一个用于机器阅读理解的数据集的是对科学文本进行推理的机器阅读理解数据集。载于2018年实证方法会议自然语言处理，1414–1423页。
260	emnlp-2025 2		Son 等 (2025年)	Jack Hessel、Jingkang Zhao、Ranjay Krishna, Angel Chang 和 Yonatan Bisk。2022年。Abductionrules：利用常识用于视觉的知识与概率推理溯因的知识与概率推理。载于自然语言处理实证方法会议会议论文集（EMNLP），第 1451–1463 页。
261	emnlp-2025 2		Kim 等人 (2025b)	刘旭东、Paul Röttger、Johannes Welbl, Yonatan Belinkov , Hila Gonen , Eric Wallace , Samuel R. Bowman、Ryan Cotterell 和 Noah A. Smith。2023a。Harmbench：衡量语言模型生成有害内容的倾向。在计算语言学协会论文集：EMNLP 2023年，第 11033–11051 页。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
262	emnlp-2025	2025.findings-emnlp.967	Eyal 等人（2025）	Jackson Petty、Sjoerd van Steenkiste、Ishita Dasgupta、Fei Sha、Dan Garrette 和 Tal Linzen。2024年。深度对组合泛化在基于 Transformer 的神经网络中的影响网络。arXiv 预印本 arXiv:2310.19956。
263	emnlp-2025	2025.findings-emnlp.97	Yin 等（2025年）	Alice Oh、Kalpesh Krishna、Eric Wallace, Yichong Zhao、Patrick Lewis 和 Antoine Bosselut。2023年。Instructir：让密集检索器遵循指令。预印本, arXiv:2305.14252。
264	emnlp-2025	2025.findings-emnlp.973	Colakoglu 等人 (2025)	Jian Zhong 等人 2020年。Docextractor：一个信息抽取端到端系统来自表单和收据。arXiv 预印本 arXiv:2012.04573。
265	emnlp-2025	2025.finnlp-2.3	George 等人 （2025）	Peter Henderson、Koustuv Sinha、Nicolas AngelardGontier、Nan Rosemary Ke、Geneviève Fried、Ryan Lowe 和 Joelle Pineau。2023年。用于法律推理的基础模型。arXiv 预印本 arXiv:2307.03557。
266	emnlp-2025 2		Nitarach 等人 (2025)	Xuezhi Ma、Yan Zhou、Yining Wang 和 1 等人。2023年。金融问答：一种推理金融问答基准。arXiv 预印本 arXiv:2302.07304。
267	emnlp-2025	2025.hcinlp-1.8	Wang 等人 (2025e)	OpenAI。2023年。Gpt-4：大规模多模态模型。arXiv 预印本 arXiv:2303.08774。
268	e		Lu 等人 (2025)	Zhe Chen、Weiyun Zhang、Wen Wang、Yiliang Liu、Zhaoyang Zhang、Jian Wang、Jie Luo、Yu Qiao 和 Wenhai Wang。2024年。Internvl 1.5：一种通用视觉语言模型。arXiv 预印本 arXiv:2404.16821。
269	emnlp-2025 2		Fatima （2025）	Jingwen Xin、Zhengying Liu、Yifan Luo 和 1 等人。2025。Deepseek-prover-v2：扩展自然语言图结构测试时计算用于自动定理证明。arXiv 预印本 arXiv:2503.11657。
270	e		Khrulev (2025)	Zhen Yuan、Yifan Zhang、Jing Liu、Yuxiang Wang、Jie Zhang、Hanwang Liu 和 Tat-Seng Chua。2024年。Fermat：用于评估 vlm’ 在事实错误纠正方面能力的基准手写数学解答。arXiv 预印本 arXiv:2405.10100。
271	e		Li (2025a)	Qwen 团队。2024年。Qwen2.5-math：扩展数学领域中的推理。arXiv 预印本 arXiv:2409.XXXXX。模型卡和技术报告。
272	emnlp-2025		Moon 等人（2025）	Leonardo Ranaldi、Barry Haddow 和 Alexandra Birch。2025。CrossRAG：跨语言用于检索增强生成的知识密集型任务的检索增强生成。arXiv 预印本 arXiv:2504.03616。
273	emnlp-2025	2025.mrl-main.2	Yuan等人 (2025c)	Laura A Banarescu、Claire Bonial、Sheila Condon、Emily Faries、Jon Niekrasz 和 Tim O ’ Connor。2018年。面向多文档摘要。在第56届会议论文集计算语言学协会年会（第2卷：短篇论文），第 577–583 页。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
274	emnlp-2025 2		Rachamalla 等 （2025）	英伟达。2025。Llama-nemotron-post-training 数据集：全面收集 指令微调和对齐数据。arXiv 预印本 arXiv:2505.00949。
275	emnlp-2025 2		Shukla 等（2025）	Shounak Paul、Arpan Mandal、Pawan Goyal，以及 Saptarshi Ghosh。2022b。InLegalBERT：预训练 用于印度法律文本的语言模型。arXiv 预印本 arXiv:2209.06049。
276	emnlp-2025 202		Chheda 等人(2025)	Shivansh Nigam、Sarvesh Dubey、Ayush Agarwal, Dhananjay Kumar 和 Saket Maheshwary。2025。 LegalSeg：揭示印度法律 文件的结构。arXiv 预印本 arXiv:2502.05836。
277	emnlp-2025	2025.nllp-1.5	Xia 等人(2025)	Ghita Houir Alami 和另外 1 位作者。2024年。 Legalbench-rag：一个用于 法律领域检索增强系统的基准。 预印本，arXiv:2408.10343。
278	emnlp-2025	2025.nlperspectives-1.11	Zhang 和 Jaitly (2025)	Barbara Plank。2022年。人工标注差异： 挑战与机遇。计算 语言学，48(4):999–1015。
279	emnlp-2025 2		Kadam 和 Ferraro (2025)	esse Dunietz、Sam Thomson、Chris Dyer 和 Noah A. Smith。2020年。一种可解释的 隐式事件因果关系词汇化模型。载于 第58届年会会议论文集 计算语言学协会，页 1703–1713。
280	emnlp-2025 2		Gasan 和 Pais (2025)	Jinhyuk Kim、Wonjin Kim、Jinhyuk Lee、Joongbo Lee、Kyunghyun Lee、Sunghwan Yoon，以及 Jaewoo Kang。2024年。Meerkat：一个医学 大语言模型推理基准。 arXiv 预印本 arXiv:2402.00000。
281	emnlp-2025	2025.starsem-1.3	Rai 等人（2025）	Harish Madabushi 等人，2024年。 用于连贯问题 回答的基于框架的嵌入。载于 EACL 会议论文集。
282	emnlp-2025 2		Bailleux 等人（2025）	Sumedha Bhan、Aaditya Prabhu、Huaxiu Ma，以及 Zachary C. Lipton。2025。完整文本 概念瓶颈模型：解决概念 完备性和分类泄漏问题。arXiv 预印本 arXiv:2502.12345。
283	emnlp-2025 2		Hayakawa 等人(2025)	Ben Dodd、Betty van Aken、Paul Röttger 和 Isabelle Augenstein。2021。AUTORANK：一种 对基准进行系统化处理并比较 机器学习模型的方法。载于 2021年自然语言处理经验方法会议 论文集，第170–185页， 线上和多米尼加共和国蓬塔卡纳。 计算语言学协会
284	emnlp-2025	2025.tsar-1.7	Romero 等人 (2023年)	Mozilla 自然语言处理团队。2023年。Mozilla 翻译： 浏览器中的开源神经机器翻译。在 《机器翻译峰会论文集》。
285	emnlp-2025 2	主要.24	Dossou 和 Aïdasso (2025)	Yi Zhang、Jialin Li 和 Danqi Chen。2023年。 协同标注：人机协同标注 通过不确定性引导的任务分配。在 2023年自然语言处理实证方法会议论文集 自然语言处理实证方法。
286	€		Vinjamuri 和 Sun (2025)	Junxian Li、Xuezhe Ma 和 Eduard Hovy。2019。 采用部分双仿射注意力的依存句法分析 。载于 NAACL-HLT 会议论文集。

续见下页

	事件编号	论文编号	幻觉论文	幻觉示例
287	emnlp-2025 2		Kolli 等人(2025)	Yujia Tan、Wenpeng Zhang、Xiang Ren 和 Qiji Chen。2023b。Oe-fact：开放域解释增强型事实核查与大型语言模型。载于计算语言学协会论文集：EMNLP 2023。
288			P 和 Mahalingam (2025)	Rahul Aralikatte、Neelamadhav Gantayat, Naveen Panwar、Anush Sankaran 和 Senthil Mani。2018年。使用一种双解码器循环神经网络。在 2018年会议论文集自然语言处理实证方法会议语言处理，页 4909–4914。计算语言学协会。
289	emnlp-2025 2		Joshi 等人（2025b）	Ahmed、T.、Hasan、M. K.、Hoque、M. T.，& Sultana, N.（2023年）。关于低资源神经机器翻译的子词切分策略比较研究。载于 *第八届机器翻译会议（WMT 2023）*（第912–920页）。计算语言学协会。https://aclanthology.org/2023年.wmt1.87
290	emnlp-2025 2		Zhou等 (2025c)	Anoop Kunchukuttan 和另外 1 位作者。2023年。IndicTrans2：迈向高质量且面向印度语的低资源机器翻译语言。在2023年自然语言处理实证方法会议语言处理。
291	emnlp-2025 2		Zhu 等人(2025c)	Martin Vechev Luca Beurer-Kellner，Marc Fischer。2024年。引导大语言模型走上正轨：快速、非侵入式约束解码。arXiv。
292	emnlp-2025	2025.wmt-1.15	Bell 等（2025）	Alfio Gliozzo 和 Carlo Strapparava。2006年。利用词汇对齐进行跨语言文本蕴含。载于第 11 届欧洲分会会议论文集，计算语言学协会，页 33–40，意大利特伦托。计算语言学协会计算语言学。
293	emnlp-2025 2		Pang 等人 (2025)	Aaron Chiang 等人。2023年。Autoreviewer：实现数据集质量控制的模型自我审查。arXiv 预印本 arXiv:2303.14112。
294			Kim （2025）	Zheng Jiang、Yang Yu、Yang Feng、Bing Qin 和 Ting Liu。2022年。Blonde：一种自动评估文档级自然语言生成指标。载于 NAACL 会议论文集，页 1679–1698。
295	emnlp-2025 2		Acharya 等人（2025年）	Gowtham Ramesh、Vishrav Chaudhary, Divyanshu Kakwani、Sai Praneeth Golla, Abhishek Philip 等人。2023年。Indictrans2：迈向高质量、高效的多语言印度语言翻译。arXiv 预印本 arXiv:2304.09105。

已结束