

# 基于 ChatGPT 的研究生人机协同学术写作实践研究及启示

□ 李 艳 金皓月 杨玉辉

**[摘 要]** 以 ChatGPT 为代表的生成式人工智能技术为人机协同学术写作带来了巨大的机遇与挑战。为进一步探索 ChatGPT 在人机协同学术写作方面的潜在价值及风险,尝试通过实践案例的比较来探究不同学术训练背景的研究生与 ChatGPT 人机协同学术写作的异同。研究发现:第一,熟悉选题的研究生在和 ChatGPT 协同学术写作过程中,提问比较具体,并能通过判断、追问、示范等提问方式和 ChatGPT 进行对话,人机协同过程不仅有收获而且相对愉悦,而不熟悉选题的研究生则相反;第二,有发表经历的研究生能够鉴别 ChatGPT 语言表达及写作思路存在的问题,并对生成的内容进行选择与加工,而没有发表经历的研究生对这些问题缺少感知;第三,ChatGPT 在人机协同学术写作中扮演的角色受到研究生学术训练背景的影响,总体特点是“遇强则强,遇弱则弱”;第四,不同学术训练背景的研究生对 ChatGPT 在人机协同学术写作中的价值判断有所不同,不熟悉选题或没有发表经历的研究生从自身不足反思人机协同学术写作所需的知识与技能,更易感受到 ChatGPT 的冲击与威胁,而熟悉选题且有发表经历的研究生更多地思考 ChatGPT 内生性的技术问题对学术写作的影响。鉴于此,教育者可以鼓励学生就写作内容向 ChatGPT 提问与追问,分析不同提问方式对 ChatGPT 生成内容质量的影响;组织学生对 ChatGPT 生成的论文进行评价与鉴赏;让学生提供人机交互的全部过程,并鼓励研究生在此基础上展开反思。

**[关键词]** ChatGPT; 人机协同写作; 案例比较; 协同行为差异

**[中图分类号]** G420

**[文献标识码]** A

**[文章编号]** 1672-0008(2023)05-0038-12

**[DOI]** 10.15881/j.cnki.cn33-1304/g4.2023.05.004

## 一、研究背景

以 ChatGPT 为代表的生成式人工智能为人机协同学术写作带来了巨大的机遇与挑战。生成式人工智能不仅能够清晰理解用户的提问并在此基础上进行回应,而且还能生成文本、撰写摘要及翻译文字等。有研究发现甚至连专家评审都只能识别 63% 由 ChatGPT 生成的摘要(Gao, et al., 2023)。《自然》杂志对他们的读者做了一项关于在科研中使用 ChatGPT 的调研,发现受访者分别使用该工具进行头脑风暴、撰写研究初稿或进行文献综述等,这些受访者认为 ChatGPT 可以帮助他们完成无聊、繁重或重复的任务,但同时也带来学术造假、学术不端等问题(Owens, 2023)。金皓月等(2023)通过对国内智能写作技术的发展与运用情况进行综述后发现,一些高校学生开始借助智能写作平台,帮助自身提高写作能力,但这类技术也存在引发学术不端的风险。因

此,需要对这类工具潜在的用途、局限性和伦理风险展开全面的研究。然而,当前关于 ChatGPT 对学术写作影响的争论主要来自研究经验丰富的教育者和研究人员,相对而言缺少来自学生的声音,特别是缺少学生人机协同学术写作的实证研究。鉴于此,本研究将通过案例比较,探究不同学术训练背景的研究生与 ChatGPT 人机协同学术写作的异同,旨在帮助教育者深入了解 ChatGPT 在人机协同学术写作中的潜在价值及伦理风险,并基于此更好地制定相关政策,同时也为未来人机协同学术写作发展提供建议。

## 二、文献综述与问题提出

培养研究生良好的学术写作能力是衡量高校人才培养质量的重要指标(教育部,等,2020)。研究生只有经过系统的训练,才能具备用抽象的学术话语来表达学术研究过程与成果的能力(朱旭东,等,2022)。然而,由于时间和精力有限,教师无法对每个

**基金项目:** 本文系 2019 年国家自然科学基金面上项目“基于智能教学系统的精准教学模式与发生机制研究”(项目编号:61977057)的研究成果。

**作者简介:** 李艳,教授,博士生导师,浙江大学教育学院(浙江杭州 310058);金皓月,系本文通讯作者,在读博士研究生,浙江大学教育学院(浙江杭州 310058);杨玉辉,高级工程师,浙江大学信息技术中心(浙江杭州 310058)。

**引用信息:** 李艳,金皓月,杨玉辉,2023.基于 ChatGPT 的研究生人机协同学术写作实践研究及启示[J].远程教育杂志,41(5):38-48+75.

学生进行有效的指导(陈文祥,2015),不少学生因为缺少前期学术积淀和对学术论文写作的正确认知,常常陷入不知如何入手的迷惘(梁晨,等,2022)。缺少规范的学术训练,研究生学术论文会出现各种问题。例如,常思亮等(2019)为了解“差评”论文的问题,选取S大学2017届411份“差评”研究生学位论文盲审专家的评阅意见,通过编码法和词频统计法,发现这些论文主要存在研究主题不明确、逻辑结构混乱、论证方法不当、结论不实与建议泛化、写作不规范、写作态度不端正、创新性缺乏等七个问题。

已有研究认为 ChatGPT 为学术写作发展带来了新的机遇。杨九詮指出,ChatGPT 的应用潜力可能颠覆传统的科学研究方式,如改变科学研究知识获取、科学研究实验设计规划、科学研究成果撰写等模式(邱燕楠,等,2023)。一些研究对 ChatGPT 在学术研究中的辅助价值进行探索,无论是对经过同行评议的相关文献的综述(Dergaa,et al.,2023),或是通过对发表的文献、网站、博客等相关自媒体的内容述评(Rahman,et al.,2023),还是对 ChatGPT 功能的详细测评(王树义,等,2023),均发现 ChatGPT 在学术写作中有一定的应用潜力,具体包括文献概括与提炼、概念解释与对比、转化或改写语言文字表达等功能。为了验证生成式人工智能的科研论文写作能力,以色列两位科学家在 ChatGPT 的帮助下,不到一小时内完成了一篇研究论文,研究发现 ChatGPT 不仅可以处理原始数据,而且能撰写初稿并进行润色,生成的文章流畅且规范(Conroy,2023)。同样,《科学》刊发的一项探究 ChatGPT 等生成式人工智能对写作影响的研究招募了453名受过大学教育的各行各业人员,并要求参与者与 ChatGPT 协同撰写他们实际工作中的任务。在参与者完成最初任务后,被随机分成实验组和对照组,实验组要求注册 ChatGPT,允许他们使用该工具来完成第二个任务,对照组则相反。通过用时统计发现,实验组完成第二项任务要快得多,平均用时16分钟,而对照组平均用时27分钟。该研究让三位经验丰富的专业人员对生成的文本质量进行1~7分的评分,发现使用 ChatGPT 的小组比没有使用 ChatGPT 的小组分数高18%。此外,在第一项中任务得分较低的参与者,在 ChatGPT 的帮助下得分提高1~2分,而那些最初获得高分的人在使用 ChatGPT 的情况下依然保持了高分。这表明 ChatGPT 对那些写作技能相对较弱的人特别有帮助,可以将他们的写作水平提升到接近熟练人士的写作水平(Noy,et al.,2023)。

ChatGPT 在写作中的辅助价值让人机协同写作成为未来写作一种可能的趋势。人机协同是人类与

智能代理相互协同实现目标的过程,两者协同期间不断进行双向信息交流,实现人类与智能代理的互相适应与相互学习(武法提,等,2023),该过程包括信息输入、信息加工、信息输出和内外反馈四个方面(何文涛,等,2023)。目前,有个别学者已经开展了学生和 GPT 系列协同学术写作的实证研究。例如,美国学者费夫(Fyfe,2022)研究大学生对人机协同学术写作的态度与看法,邀请20名来自不同年级、不同专业的学生使用 GPT-2 来完成“数据与人类”(Data and the Human)课程的期末论文,要求学生在论文中标注哪些部分是自己创作的、哪些部分是由系统生成的,并在此基础上写一篇有关此次人机协同写作的反思。该研究发现,87%的学生认为人机协同学术写作比自己单独写作复杂得多,“是否或如何在论文中使用这些材料”让写作变成了一件更复杂、需更多投入的工作。GPT-2 的优点包括提供一些新观点、帮助形成新想法、学习人类的语言表达风格,但不足之处在于编造看似合理的虚假信息,这就要求学生花更多的时间对信息进行辨别。学生在人机协同学术写作中的角色更像“编辑、策展人或调解人”。克罗地亚学者贝塞客等(Basic,et al.,2023)采用实验法,研究大学生有无使用 ChatGPT 对论文写作质量的影响。其中,9名学生为对照组,另外9名学生为使用 ChatGPT 的实验组,研究要求学生用克罗地亚语完成一篇论文,最后对论文进行打分,并比较了两组论文在分数、写作时长、文章真实性和内容相似性等维度的差异。该研究发现没有证据表明使用 ChatGPT 可以提高论文质量,对照组在大多数维度上优于实验组,实验组并没有表现得更好,没有提供更高质量的内容,没有写得更快,也没有呈现更真实的文本。研究者认为由于该研究采用克罗地亚语,且受限于样本量,因此不能排除 ChatGPT 在其他语言中的科研辅助价值。

目前,有关人机协同学术写作的实证研究刚处于起步阶段。ChatGPT 存在的虚假信息、学术语言表达问题、阻碍思维发展等问题也引发了学者的关注。有的学者认为 ChatGPT 生成的虚假信息可能会扭曲科学事实并散播错误信息。例如,有研究通过让 ChatGPT 综述一篇文献,发现其编造了一个令人信服的摘要,但存在事实错误、虚假陈述和错误数据等问题,更关键的是该摘要夸大了研究的有效性(Van Dis,et al.,2023)。由于鉴别虚假信息需要一定的能力,有学者(Han,et al.,2023)建议学识不足的学生群体不要太早拥抱 ChatGPT。有的学者认为 ChatGPT 存在学术语言表达方面的问题。例如,有研究(Guo,et al.,2022)将多个领域中的 ChatGPT 回答与人类专

家的回答进行比较,发现 ChatGPT 更喜欢使用长句子,运用的词汇量没有人类学者丰富,而且重复使用训练文本中的常见模式和句法结构。令小雄等(2023)对 ChatGPT 的学术写作能力进行特征分析与描述,发现生成的内容虽然逻辑框架稳固有效,论证清晰充分,但缺乏应有的思想深度,论证语言单薄,说理手段较为单一。也有学者认为 ChatGPT 可能会阻碍学生思维的发展。如艾斯(Else,2023)指出,ChatGPT 模型是根据过去信息进行的训练,然而科学知识生产需要创新思维而非守旧陈因,过多地使用这类工具可能导致学生批判性思维和创新能力的匮乏。徐辉富认为学生可能加大对 ChatGPT 的依赖,学术写作变得容易,但写作的洞见性和智慧却愈少(邱燕楠,等,2023)。由此引发的学术不端现象更是令学者们担忧(Owens,2023)。

综上,已有研究表明,ChatGPT 给学术写作带来了诸多机遇和挑战,为数不多的实证研究主要关注人机协同写作效果,鲜有研究探究人机协同的过程及影响因素。在高校研究生写作领域,已有研究显示研究生的学术写作需要以坚实的学科专业知识为基础(王立珍,等,2012),此外,有过学术论文训练与论文发表经历的研究生在发现问题、开展研究及撰写论文等方面有更好的表现(高文娟,等,2022)。鉴于此,本研究拟探究不同学术训练背景的研究生与 ChatGPT 人机协同学术写作过程和效果的差异,通过案例比较,深入了解 ChatGPT 在人机协同学术写作中的潜在价值及伦理风险。具体包括三个问题:

- (1)不同学术训练背景的研究生在人机协同学术写作过程中的行为有何差异?
- (2)与不同学术训练背景的研究生协同学术写作,ChatGPT 扮演的角色有无差异?
- (3)不同学术训练背景的研究生如何看待和评价人机协同学术写作?

### 三、研究设计

#### (一)研究对象

本研究采用多案例比较的研究方法,这是一种证明因果推断和揭示因果机制的重要研究方法(Eisenhardt,1991)。我们于2023年1月初在网络上招募参与者,共有13名研究生报名,按照“最相似案例”的选择方法(Gerring,2008)进行筛选,最终邀请到4位研究生,按照参与顺序编号为A、B、C、D。他们分别来自化工专业和食品科学工程专业,且均为首次和 ChatGPT 协同学术写作。4位参与者的人机协同论文选题均是他们目前正在开展的研究选题,其中,A和C两位同学为研一学生,研究尚处于摸索

阶段,并不熟悉选题;B同学的选题与他本科毕业论文相关,D同学的选题与她博士论文相关,他们对选题较为熟悉。同时,A和D两位同学有论文发表经历,A同学发表的这篇论文和本次选题无关,D同学发表的五篇论文均和本次选题有关,而B和C同学没有论文发表经历。

由参与者自行决定具体的人机协同学术写作方式及结束时间,研究者并未进行事先规划。此外,为了更好地了解研究生对人机协同学术写作的态度与看法,研究者在余下的9名学生中邀请到4位观察者,他们的主要任务是对4位同学的人机协同学术写作进行全程观察,并在观察中记录自己的感想(见表1)。

表1 四个案例的概况

编号	成员身份	年级	专业	论文发表经历	论文选题	论文选题熟悉程度	协同内容
A	参与者	研一	食品科学与工程	1篇SCI一作	手性材料初探	不熟悉选题,研究尚处于摸索阶段	论文绪论
B		研一	化学工程	无	石墨烯氧化物改性膜在电渗析海水淡化中的性能探究	熟悉选题,本科论文内容相关	论文的绪论、实验分析、结论与展望
C		研一	化学工程	无	电渗析高效分离镁锂离子	不熟悉选题,研究尚处于摸索阶段	研究实验设计
D		博三	食品科学与工程	5篇SCI一作	硒化铁铜纳米粒子与β淀粉蛋白作用	熟悉选题,博士论文相关	论文绪论及实验设计
E	观察者	研一	计算机科学与技术	无			
F		博一	药理学	1篇SCI一作			
G		研一	机械工程	无			
H		研一	新闻学	1篇C刊二作			

为更好地记录参与者和 ChatGPT 协同学术写作的过程,每个案例均由一位参与者通过腾讯会议分享屏幕和“出声思考”(Think Aloud)的方式呈现人机协同的过程。该过程主要包括:(1)参与者输入指令;(2)ChatGPT 依据指令生成相应内容;(3)参与者另起话题或针对生成的内容再次输入指令,直到生成参与者满意的内容为止(见表2)。出声思考是参与者在人机协同学术写作过程中的所思所想,这一方法不仅让研究者有机会了解参与者在人机协同中的想法,同时还能对参与者思考的内容、过程及结果进行探究。人机协同学术写作均在中文语境下开展。经



研究对象同意,研究者对四个案例的人机协同全过程进行录像,最终获得总时长近 11 小时的视频。人机协同学术写作时间为 2023 年 1 月 24 日到 1 月 31 日。

表 2 参与者和 ChatGPT 协同学术写作的过程示例

案例	参与者输入的指令	ChatGPT 生成的内容	参与者出声思考数据	参与者输入的下一个指令
A	我的课题大致是“手性材料初探”,请先帮我完成绪论的引言部分	生成了研究的背景及本研究的目,共计 449 字	居然真的能写!概念有点问题,有些话太笼统了,表达很繁琐,还出现了一些无关的内容	请帮我生成“手性材料初探”的具体的引言
B	(提供四组实验数据),请对这组实验的时间变化和电压的关系进行分析,并讲述相关原因	生成的内容包括每一组实验数据的变化,也分析了不同组之间的数据差异,共计 128 字	(“哇塞”时刻,所有直播间中懂化学的同学发出惊叹声)非常成功,不仅能够分析数据,还提供了我没有想到的点,但有个错误	我认为不存在电流密度过高的问题,电压超过 60 伏可能是电阻过大的问题,请帮我从这个角度分析这组数据
C	我想进行电渗析复杂装置高效地去除锂离子,你能为我提供设计实验的思路吗?	提供设计思路,包括确定实验目标、确定实验参数、设计实验方案、设计实验装置、进行实验和数据分析及汇总实验结论等,共 504 字	感觉都是套话和模板,看到这样的模板感觉不是很舒服。细节什么的还是要自己想。真正有效的信息并不多,“探索合适的电解质”有点用,但这不是这个实验的重点吧	我不要套话,请你帮我生成具体的实验方案
D	我的大概思路是先合成一个手性材料,观察它是否能与淀粉样蛋白作用,先在体外测试,然后分析机制,最后在细胞水平和活体水平进行验证	肯定该研究思路,并生成一个论文框架,包括引言、实验部分、讨论、结论和参考文献等,共 371 字	哦,天!直接把论文的框架写出来了。和我发的那篇文章不一样,因为它的这个实验设计参考了我的提问部分。但是实验设计部分有些地方不合理,比如“机制分析”不能放在实验里	引言和结论部分可以按照你说的这么写,但是后面实验和讨论我希望放在一起写,重新帮我写一下大纲

## (二)数据收集

本研究的数据包括三个方面:(1)参与者和 ChatGPT 协同过程的数据,包括每个案例的协同时长、对话次数、参与者的提问内容与 ChatGPT 生成的文本内容;(2)出声思考数据,包括参与者出声思考的次数及转录后的文字内容;(3)访谈数据,包括 4 位参与者和 4 位观察者的半结构访谈内容(见表 3)。面向参与者和观察者的三个主要问题为“你对 ChatGPT 有怎样的态度和看法”“你认为 ChatGPT 在

学术写作中有什么价值或问题”及“你对人机协同学术写作有什么看法”。

表 3 案例数据收集情况

案例	数据来源							
	协同学术写作数据			出声思考数据		参与者的访谈数据		观察者的访谈数据
	协同时长	对话次数	提问及生成的文本	次数	录音字数	访谈时长	录音字数	访谈时长
A	103 分钟	19 次	5476 字	14 次	821 字	18 分钟	826 字	23 分钟
B	191 分钟	43 次	13838 字	27 次	1985 字	34 分钟	1298 字	26 分钟
C	152 分钟	33 次	8092 字	20 次	1519 字	11 分钟	523 字	20 分钟
D	107 分钟	37 次	11553 字	30 次	1628 字	21 分钟	1163 字	24 分钟

## (三)数据分析

为更好地了解不同参与者在人机协同学术写作中“提问”的差异,研究者参考丹东尼奥等(Dantonio, 2006)的提问结构,该结构由核心提问和各种追问组成,结合“提问”数据的实际情况,将分析维度划分为指示、评价、示范、追问和继续五类(见表 4)。

对参与者的出声思考数据初步分析发现,该数据主要为参与者对 ChatGPT 生成内容的评价及协同感受。为更好地了解参与者的评价类型,研究者参考美国心理学家斯滕伯格(Sternberg, 2000)关于教师回应学生回答的分类,结合出声思考数据的实际情况,将分析维度划分为认可、疑惑、不认可和其他四类(见表 4)。两位研究人员依据表 4 对人机协同学术写作过程

表 4 人机协同学术写作过程中的“提问”及“出声思考”特征描述及案例编码结果分布

内容	类别编码	特征描述	示例	案例编码结果			
				A	B	C	D
提问	1 指示	对生成内容所下的指令	阿尔茨海默病的背景怎么写	15	25	28	17
	2 评价	对生成内容正确的评价,包括纠错行为	“纳米颗粒具有手性结构”这个表述不准确	0	7	4	13
	3 示范	提供写作范例或正确内容生成	我给你示范一下,第一段要这么写,“阿尔茨海默病是一种……”	0	0	0	4
	4 追问	针对生成内容的提问	那你刚才为什么说不能直接证明?	1	1	1	3
	5 继续	继续生成未完成的内容	“继续”或将未完成的内容复制一遍	3	10	0	0
出声思考	1 认可	认可生成内容的质量	我的天,相当可以!	3	7	2	18
	2 疑惑	不能确定生成内容的真假	我也不能确定这个说的是不是对的	4	3	2	3
	3 不认可	指出生成内容的错误	这不对呀,开始说胡话了	2	6	5	8
	4 其他	包括策略、感受等	感觉会让我学术不端的	5	11	11	10

中的提问数据及出声思考数据进行独立编码,编码的一致性检验结果较好(Kappa 系数大于 0.8)。

访谈数据分析采用类属编码法(陈向明,2000),该方法包括初级编码、类目识别和结果呈现三个步骤。研究者先将访谈数据进行初级编码,再在初级编码基础上进行类目编码。通过编码归纳后,总计产生 34 条初级编码和“价值判断、所需知识技能、问题风险感知”三个要素类目(见表 5)。

表 5 访谈数据类目、初级编码与典型条目

类目	初级编码	典型条目(举例)
价值判断	有价值:生成能力、学习能力、节约时间、提供思路、写作助手、学习伙伴	生成的内容提示哪些是已有的研究,也可以提供一些思路
	无价值:没有太多价值	
所需知识技能	专业知识能力、学术写作能力、技术感知力、创造力、批判力、洞见力、纠错力、自制力、探索适宜情境	D 同学有专业知识能力和对信息的洞见与判断能力,将 ChatGPT 当成写作助手,而写作“大脑”掌握在自己这里
问题风险感知	技术问题:虚假信息、缺少创新、表达冗余、随意叫停、算力限制、语义演绎	我觉得它没法提供一些有创新性的内容,创新需要人对生成的内容进行加工与判断
	伦理风险:降格驯化、学术不端、增加惰性、版权问题、数字区隔、替代危机、意义丧失、自主性丧失	ChatGPT 的能力因人而异,能力强的人和 ChatGPT 是双向奔赴,而能力弱的人和 ChatGPT 则是弱的同频共振,因此,未来会不会出现强者越强,弱者越弱

## 四、研究结果

在人机协同学术写作过程中,四位同学不仅提问和出声思考的频次有所不同(见图 1),而且叫停人机协同的原因各异。其中,A、B 和 C 三位同学系主动叫停,A 同学叫停的原因在于 ChatGPT 提供了太多“不确定且胡说八道”的信息,B 同学在完成计划内容后结束人机协同,C 同学由于 ChatGPT 生成的内容“实在离预期太远”而终止人机协同。与这三位

同学不同的是,D 同学由于提问次数太多被迫终止,ChatGPT 显示“Too many requests in 1 hour. Try again later”(1 小时内请求次数过多,请稍后再试)。

### (一)不同学术训练背景的研究生在人机协同学术写作过程中的行为分析

针对问题一“不同学术训练背景的研究生在人机协同学术写作过程中的行为有何差异”,对四个案例中的“提问”及“出声思考”数据比较发现,研究生对选题的熟悉程度影响其和 ChatGPT 的协同方式,具体包括提问具体程度、反馈次数、协同收获、协同感受及人机平均交互用时五个方面(见表 6)。其中,“提问具体程度”指参与者的提问中是否包含明确清晰的任务说明,以及是否用多个提示词(专业名词)对任务范围进行限制等。此外,研究生的发表经历影响其对 ChatGPT 语言表达的评价、对生成内容的文字加工及思路整合等(见表 7)。其中,有发表经历的 A 同学和无发表经历的 C 同学由于对选题不够熟悉,因此他们将关注点更多地落在生成文本的内容方面,对 ChatGPT 的语言表达关注度并不高。

表 6 研究生是否熟悉选题对其人机协同学术写作的影响

	维度	提问具体程度	反馈次数	协同收获	协同感受	平均交互用时
	判断依据	“指示”数据	“评价、追问、示范”数据	“出声思考”数据	“出声思考”数据	平均每次交互用时
不熟悉选题	A	不具体	1 次“追问”	收效甚微	失望、不轻松	5.5 分钟/次
	C	不具体	4 次“评价”、1 次“追问”	收效甚微	崩溃、很挫败	4.5 分钟/次
选题熟悉	B	具体	7 次“评价”、1 次“追问”	扩充知识	很累、超负荷	4.5 分钟/次
	D	具体	15 次“评价”、3 次“追问”、4 次“示范”	点灯、周延逻辑	愉悦、很轻松	2.8 分钟/次

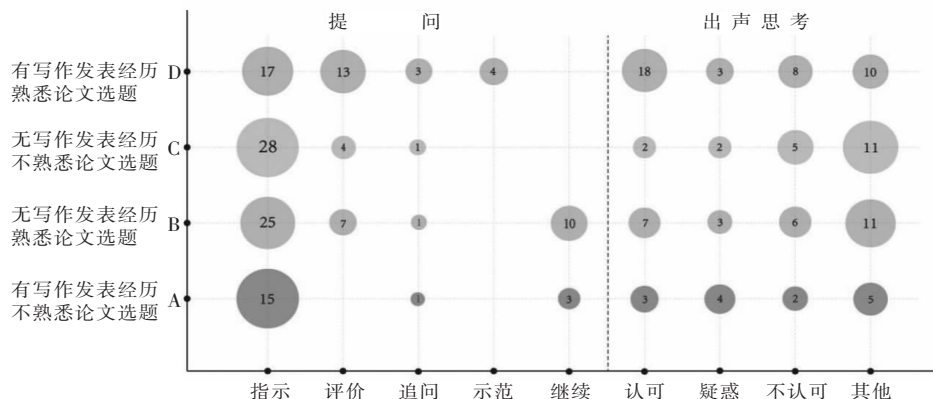


图 1 四个案例中的“提问”及“出声思考”频次

表7 研究生有无发表经历对其人机协同学术写作的影响

	维度	对 ChatGPT 语言表达的评价	对生成内容的文字加工	对生成内容的思路整合
	判断依据	“出声思考”数据	“示范、继续”数据、 “出声思考”数据	“提问”数据、 “出声思考”数据
有发表经历	A	有点啰嗦	无	无
	D	啰嗦冗余、重点不突出	示范、加工及整理	以参与者思路为主
无发表经历	B	写得比我好	扩写	顺着 ChatGPT 生成的思路
	C	无	无	无

第一,在提问具体程度方面,不熟悉选题的 A 和 C 同学提问不具体,而熟悉选题的 B 和 D 同学提问较为具体。对“指示”这一提问数据分析发现,不熟悉选题的 A 同学 15 个“指示”中只包含了 4 个专业名词,不熟悉选题的 C 同学 28 个“指示”中包含了 8 个专业名词。这两位同学的提问中包含的专业名词较少,且提问方式大多为“专业名词+论文组成部分(如绪论)”。针对这样的提问,ChatGPT 尽管能够生成相应的内容,但主要是基于语言模型进行的内容演绎,生成的内容比较宽泛与笼统。相较而言,熟悉选题的 B 同学 25 个“指示”中包含 18 个专业名词,熟悉选题的 D 同学 17 个“指示”中包含 29 个专业名词。这两位同学的提问不仅用较多的专业名词对范围进行限定,而且对任务有清晰且明确的说明。这样的提问方式能够主导 ChatGPT 生成让他们满意的内容。这表明参与者对论文选题的熟悉程度影响其提问具体程度,从而影响人机协同学术写作过程中 ChatGPT 生成内容的质量。

第二,在反馈次数方面,不熟悉选题的 A 和 C 同学对 ChatGPT 生成的内容反馈次数较少,而熟悉选题的 B 和 D 同学反馈次数较多。其中,不熟悉选题的 A 同学针对 ChatGPT 的 15 次生成只进行了 1 次“追问”,不熟悉选题的 C 同学针对 ChatGPT 的 28 次生成进行了 4 次“评价”和 1 次“追问”。结合两位同学的“出声思考”数据分析发现,两位同学尽管对 ChatGPT 生成的内容有所判断,但知识的欠缺让他们只能采用“我觉得你这个说得不对”“这句话我不能理解,讲得俗一点”的提问方式。面对这类提问方式,ChatGPT 生成的内容“答非所问”。相较而言,熟悉选题的 B 同学针对 ChatGPT 的 25 次生成进行了 7 次“评价”和 1 次“追问”,熟悉选题的 D 同学针对 ChatGPT 的 17 次生成进行了 13 次“评价”、3 次“追问”和 4 次“示范”,且反馈次数是四个案例中最多的。这两位熟悉选题的同学均能提供正确的知识让 ChatGPT 对虚假信息进行修正。此外,D 同学还提供

写作表达的范例供 ChatGPT 学习。这表明参与者对论文选题的熟悉程度影响其能否对 ChatGPT 生成的内容进行高质量的反馈,并通过高质量的反馈提升 ChatGPT 生成内容的质量。

第三,在协同收获方面,不熟悉选题的 A 和 C 同学在此次人机协同中收效甚微,相反,熟悉选题的 B 和 D 同学颇有收获。由于知识受限,不熟悉选题的 A 和 C 同学有时无法理解 ChatGPT 生成的内容,更无法在此基础上产生“化学反应”。相较而言,熟悉选题的 B 和 D 同学能够利用已有知识和 ChatGPT 生成的内容进行对话。比如,熟悉选题的 B 同学在人机协同中得到了一些思路启发;熟悉选题的 D 同学认为 ChatGPT 具有“点灯”的作用,能将她的静默知识点亮,“这个方法我在文献中看到过,但我没有把它考虑到这个实验中”。此外,D 同学还就人机分歧的内容对 ChatGPT 进行追问,人和机在高质量的探讨中周延了研究的设计逻辑。由上述可知,参与者对论文选题的熟悉程度影响其能否在 ChatGPT 生成内容的基础上形成新的认知,从而影响参与者的协同收获。

第四,在协同感受方面,不熟悉选题的 A 和 C 同学感到失望和沮丧,熟悉选题的 B 同学感到累和超负荷,熟悉选题的 D 同学感到轻松和愉悦。具体而言,不熟悉选题的 A 同学对 ChatGPT 提供的虚假信息感到失望,同时也对自己缺少知识进行纠错而气馁。不熟悉选题的 C 同学在参与人机协同前满怀期待(我感觉要多一个导师了),但到人机协同时频频叹息(12 次),变得异常沮丧。熟悉选题的 B 同学由于与 ChatGPT 交互措辞带来的负担而感到“累和超负荷”,“这种交互很累,跟哄孩子一样,表达得不够清楚它就不会明白”,并表示这场交互“不亚于高考”。与这些同学相比,熟悉选题的 D 同学是 4 位同学中最轻松的,人机协同过程非常愉悦,共出现 14 次笑声。这表明参与者对论文选题的熟悉度影响其人机协同学术写作的轻松程度。此外,与 ChatGPT 语言交流的方式也会影响参与者的协同感受。

第五,在平均交互用时方面,不熟悉选题的 A 和 C 同学平均每次交互用时较长,而熟悉选题的 D 同学平均每次交互用时较短。B 同学尽管也熟悉选题,但由于自身语言表达的问题,在提问措辞构思方面花费较多时间,导致平均每次交互用时也较长(见表 6)。进一步分析发现,信息判断和提问构思是造成平均交互时长短差异的主要原因。这表明参与者对论文选题的熟悉程度影响其信息判断和提问构思的用时。另外,参与者自身的语言表达能力也会影响人机协同学术写作过程中的平均每次交互用时。



第六,在对 ChatGPT 语言表达的评价方面,有发表经历的 A 和 D 同学均认为 ChatGPT 生成的内容较为啰嗦冗余,而没有发表经历的 B 同学对生成内容的评价较高。其中,有发表经历的 A 同学认为生成的文字比较啰嗦。相比 A 同学,有发表经历的 D 同学对 ChatGPT 语言表达的评价更加深入与具体。D 同学以期刊发表要求为标准,认为生成的内容尽管符合学术语言规范,但存在一些废话和重复性的内容,没有体现句子间的重点,更缺乏学术表达的洞见力。相较而言,没有发表经历的 B 同学以自己的写作水平作为评价标准,认为 ChatGPT 生成的内容质量较高,7 次出现“写得比我好”“写得比本科生好”等评价。除此之外,B 同学还认为 ChatGPT 的论证逻辑能够带给他一些启发,可以帮助他减少语言构思的时间。这表明没有发表经历的 B 同学对 ChatGPT 的语言问题缺少感知,而在语言表达逻辑方面却能得到一些启发。

第七,在对生成内容的文字加工方面,没有发表经历的 B 同学采用“扩写”的方式,而有发表经历的 D 同学采用“示范”的方式。具体而言,没有发表经历的 B 同学 43 次提问中 10 次为“请对……内容进行扩写”的指令,而没有对生成的文字进行加工与处理。对于他和 ChatGPT 最终协同完成的论文,虽然通篇没有语病,但是 B 同学自己认为水准不够,达不到本科生的水平,“仅仅只是像一篇论文而已”。有发表经历的 D 同学对生成的内容进行改写,并提供自己的写作范例让 ChatGPT 学习,人机协同生成的内容重点就相对突出,语言也较为凝练。这表明 D 同学用自己的写作能力弥补了 ChatGPT 的语言表达问题。

第八,在对生成内容的思路整合方面,没有发表经历的 B 同学没有对 ChatGPT 生成的思路进行整合,而有发表经历的 D 同学进行了整合。没有发表经历的 B 同学之所以不对生成的思路进行整合,一是他较为认可 ChatGPT 生成的写作思路,“这个思路也可以吧”;二是出于惰性与省力,他认为“如果要调整的话,要做大修改,还是算了吧”。相较而言,有发表经历的 D 同学以自己的思路为主,因为她认为 ChatGPT 生成的思路缺少创新性,有些篇幅安排不合理、重点不突出,“如果不进行加工与整合,这思路也没法用”。这表明有发表经历的 D 同学能够识别 ChatGPT 生成思路的问题并进行加工。此外,ChatGPT 有增加没有发表经历的 B 同学惰性的风险。

## (二) ChatGPT 在与不同学术训练背景的研究生协同学术写作中呈现的角色差异

针对问题二“与不同学术训练背景的研究生协同学术写作,ChatGPT 扮演的角色有无差异”,对案

例进行比较分析发现,受到参与者的不同学术训练背景的影响,ChatGPT 在四个案例中呈现的角色有所不同,总体特点是“遇强则强,遇弱则弱”(见图 2)。

<b>案例 B</b> 参与者:熟悉论文选题、 无发表经历 机器:学霸同桌	<b>案例 D</b> 参与者:熟悉论文选题、 有发表经历 机器:高级助理
<b>案例 C</b> 参与者:不熟悉论文选题、 无发表经历 机器:笨蛋学伴	<b>案例 A</b> 参与者:不熟悉论文选题、 有发表经历 机器:离谱队友

图 2 ChatGPT 与不同背景的研究生协同学术写作过程中所扮演的角色差异

ChatGPT 在与不熟悉选题但有发表经历的 A 同学协同学术写作过程中,A 同学由于自身的知识储备不足导致提问不具体,也无法对生成的虚假信息予以纠正,外加 ChatGPT 技术本身的问题,生成的内容出现了诸多虚假信息,包括定义错误、提供了虚假文献和虚假术语等。ChatGPT 在此次人机协同学术写作中扮演了“离谱队友”的角色。C 同学面临着和 A 同学一样的困境,外加多次采用“你好好想想”的提问方式,最终 ChatGPT 生成的 29 次内容中只有两次不包含虚假信息,始终处在“胡说八道”的状态,无法提供实验创新设计。ChatGPT 在此次人机协同学术写作中扮演了“笨蛋学伴”的角色。在这两个案例中,ChatGPT 的表现与参与者的期待完全不同,两位同学希望其能够充当“导师”的角色,给予必要的指导,结果发现 ChatGPT 甚至不是一位可以信赖的伙伴,因此,他们和 ChatGPT 再次协同的意愿并不强烈。

ChatGPT 在与熟悉选题但没有发表经历的 B 同学协同学术写作过程中,通过 B 同学的问题指引,尽管也出现了一些虚假信息,但能够进行修正,故生成的内容质量相对较高。B 同学不仅得到了一些研究与写作思路的启发,而且还发现自己的研究设计创新性不足。ChatGPT 在此次人机协同学术写作中扮演了“学霸同桌”的角色。B 同学认为这个同桌“不仅学识比大部分同学丰富,而且永不疲倦,可以随时向它请教”。ChatGPT 在与熟悉选题且有发表经历的 D 同学协同学术写作过程中,D 同学具有的专业知识及写作能力让她不仅提问清晰,同时还能识别 ChatGPT 存在的问题并进行“喂养”。生成的内容显示 ChatGPT 不仅可以学习 D 同学的语言表达风格,承担部分文字工作,而且还能找到 D 同学设计逻辑中的漏洞。ChatGPT 在此次人机协同学术写作中扮演了“高级助理”的角色。这两位同学均有与 ChatGPT

再次协同的意愿。与此同时,B同学指出“人机协同学术写作需要探索适宜的情境”;D同学指出“人不能太依赖机器,因为机器会随时限制人的使用”。

### (三)不同学术训练背景的研究生对人机协同学术写作的看法和评价

针对问题三“不同学术训练背景的研究生如何看待和评价人机协同学术写作”,在每个案例结束后,研究者对参与者和观察者进行了访谈。对访谈数据分析发现,8位不同背景的研究生对 ChatGPT 应用价值的判断、人机协同学术写作所需的知识技能及对 ChatGPT 的风险感知各有不同(如表 8 所示)。

表 8 参与者和观察者对人机协同写作的评价与看法

案例	身份	对 ChatGPT 应用价值的判断	人机协同学术写作所需的知识技能	对 ChatGPT 的风险感知
A	参与者	没有太多价值	专业知识、纠错力、有自己的想法与思考	虚假信息、表达冗余
	观察者	生成能力、写作助手	纠错力、自制力	虚假信息、替代危机、学术不端
B	参与者	学习伙伴、节约文字构思时间、创新对照者	自己要有想法、对话力、写作策略	增加惰性、学术不端、语言降维、时间成本
	观察者	提供研究思路和写作思路、文字助手、学习伙伴	对话力、反思力、行动力、写作策略	意义丧失、学术不端、量产论文、对人类语言的影响
C	参与者	不清楚、没有太多价值、浪费时间	探索适宜的协作情境、有自己的想法与思考	自主性丧失
	观察者	没有太大价值	技术感知力、创新能力	时间成本、缺乏创新、语义演绎
D	参与者	节约时间、学习能力很强	创新能力、选择与加工	算力限制、缺少创新、版权问题
	观察者	学习能力、学术助手	专业知识、对知识的洞见与思考、学养、写作能力、交互力	数字区隔、新的不公平

首先,熟悉选题的参与者认为 ChatGPT 在人机协同学术写作中有一定的应用价值,而不熟悉选题的参与者认为应用价值有限。相比选题熟悉程度,有无发表经历对价值判断的影响不明显。观察者的观点与参与者的观点基本一致。析言之,熟悉选题的 B 同学认为 ChatGPT 可以充当学习伙伴,熟悉选题的 D 同学认为可以节约时间。同样,观察者在这两个案例中发现了 ChatGPT 的诸多辅助价值,比如提供写作思路、充当学术助手等。相较而言,不熟悉选题的 A 和 C 两位同学均认为 ChatGPT 在学术写作中没有明显的应用价值。观察者在这两个案例中的观点与参与者基本一致,稍有不同的是,在 A 案例中,观察者首次了解 ChatGPT 的生成性能,认为 ChatGPT 有充当写作助手的价值,但到了 C 案例,由于 ChatGPT 不具

有创新能力且胡说八道,4 位观察者也认为应用价值甚少。综上所述,参与者和观察者对 ChatGPT 在人机协同学术写作中的价值判断,受到 ChatGPT 扮演的角色影响,ChatGPT 表现得越好,越能让参与者和观察者看到其价值。此外,参与者和观察者都基于工具的角度对 ChatGPT 的价值进行判断。

其次,对于人机协同学术写作所需的知识技能,不同背景的参与者思考的角度有所不同,观察者思考的角度也发生了变化。不熟悉选题或没有发表经历的参与者主要从自身的不足进行反思,而熟悉选题且有发表经历的参与者主要从 ChatGPT 在写作中存在的问题进行思考。比如,不熟悉选题的 A 和 C 两位同学强调人需具备一定的专业知识才能和 ChatGPT 进行协同学术写作,C 同学还指出人需要具有一定的想法才能进行人机协同写作,这些均是两位同学基于自身问题而进行的思考。而熟悉选题且有发表经历的 D 同学则是针对 ChatGPT 在内容及文字表达方面的问题,提出人需具有批判、选择、创新等能力。此外,观察者在观察不熟悉选题或没有发表经历的参与者和 ChatGPT 协同学术写作案例时,均是站在“他”的角度,提出人机协同学术写作所需的知识技能。比如,他们针对 B 同学在人机协同中表现出来的惰性,提出人需要具有写作的“行动力”。而在观察熟悉选题且有发表经历的 D 同学和 ChatGPT 协同学术写作案例时,观察者转向“我”的角度进行思考。比如,4 位观察者均表示 D 同学让他们意识到,关键在于“我”需要具有相应的知识与技能,才能让 ChatGPT 成为“我”的科研助手。他们还认为,D 同学在一定程度上消解了他们的忧虑,面对强大的 ChatGPT,人的能力能够重塑人的尊严。由上述可知,人机协同学术写作实践不仅有助于提升参与者对 ChatGPT 的认识,而且还能帮助他们从自我角度反思人机协同学术写作所需的知识技能。全面观察成功与失败的人机协同学术写作案例,会让观察者的思考视角“由外向内”,帮助观察者产生提升自我能力的想法。

最后,不熟悉选题或没有发表经历的参与者主要从学生视角感知 ChatGPT 对他们的冲击与威胁,包括自主性丧失、替代危机、意义丧失、学术不端、数字区隔及降格驯化等;而熟悉选题且有发表经历的参与者主要从技术角度评价 ChatGPT 存在的问题,包括虚假信息、表达冗余、缺乏创新及算力限制等。观察者的观点也包括以上两个角度。例如,有的担心自己的研究地位被 ChatGPT 取代,“我觉得导师招收我还不如招收 ChatGPT”(观察员 E)。又如,有的认为自己的学术写作水平不如 ChatGPT,进而否定自



身的写作意义。再如,7位同学(除熟悉选题和有写作发表经历的D同学外)担心 ChatGPT 可能会增加他们的惰性及让他们产生学术不端行为。此外,参与者和观察者的观点也不尽相同,有些观点是参与者通过自己与 ChatGPT 协同学术写作实践后发现的,而观察者难以察觉。比如,C同学感到“自主性丧失”,就是因为在人机协同中他寄希望于 ChatGPT 能够生成创新性的内容,导致自己似乎丧失了思考力,而观察者在C案例中没有发现这个风险。综上可知,能力暂时不足以与 ChatGPT 协同学术写作的参与者,担心 ChatGPT 对自己的冲击,其中有些风险只有在人机协同学术写作实践后才能察觉;具备与 ChatGPT 协同学术写作能力的参与者,则从 ChatGPT 内生性的技术问题出发,思考其对学术写作的影响。

## 五、结论与讨论

通过比较不同学术训练背景的研究生和 ChatGPT 协同学术写作的案例,本研究得出以下结论:

第一,熟悉选题的参与者在协同学术写作中提问比较具体,能通过判断、追问、示范等提问方式与 ChatGPT 进行高质量对话,在人机协同学术写作中不仅有收获,而且相对愉悦,而不熟悉选题的参与者恰好相反。这一定程度上说明,首先,参与者对论文选题越熟悉,越有助于提出清晰与具体的问题,从而 ChatGPT 生成的内容质量也越高。这在一定程度上验证了沃尔弗拉姆(Wolfram,2023)的观点,即人类给予的提示越具体与清晰,越可以将 ChatGPT 引入到具体的数据集轨迹中,答案的质量也越高。其次,参与者与 ChatGPT 进行对话需要以专业知识为基础。熟悉选题的参与者能够很好地评判 ChatGPT 生成的内容并进行反馈,而不熟悉选题的参与者不仅不能很好地辨别 ChatGPT 生成的虚假信息,同时对生成的肤浅、浮泛的信息缺乏足够感知,这和已有研究(Han,2023)的观点一致。再次,选题熟悉程度影响参与者能否对 ChatGPT 生成的内容进行加工并创新。在学术写作中,人机协同学术写作的本质在于人通过评价、示范和追问等提问方式和 ChatGPT 进行对话,其中,提问和对话隐含着对判断、加工及创新等能力的考查,而这些能力以知识为基础。此外,ChatGPT 读取的数据库较丰富,所生成的文本内容较为全面,但启发效果受参与者已有专业知识的影响。熟悉选题的参与者具有相应的专业知识,能够和 ChatGPT 进行对话并有所收获。由上述可知,和传统学术写作一样,专业知识仍然是人机协同的关键因素,缺少相应的专业知识,学生无法进行具体的提问,无法用已有的知识和 ChatGPT 对话,一定程度上

也无法形成新的认知。

第二,有学术发表经历的参与者能够鉴别 ChatGPT 生成的内容,能够清晰分辨 ChatGPT 语言表达及写作思路的问题,并对生成的文字进行加工与整合,而没有发表经历的参与者对这些问题缺少感知。有发表经历的参与者以学术期刊发表的标准审视 ChatGPT 生成的内容,认为其存在语言表达缺少重点、问题分析不深入、论证较为单一、缺少洞见力与深刻性等问题,他们对 ChatGPT 语言的感知和已有的研究(Gao,2023;令小熊,等,2023)观点相似。本研究进一步发现没有发表经历的参与者由于缺少写作经验和学术鉴赏力,对 ChatGPT 在学术语言表达中存在的问题缺少感知,长此以往,这类学生存在着被 ChatGPT 降化为机器类成员的风险。结合 ChatGPT 生成的内容可以带来语言表达逻辑方面的启发这一观点,对于缺少写作经验的学生而言,可能更适合将自己写作的版本与 ChatGPT 生成的版本进行对照,并在比较与对话中提升自己的学术表达能力,这在一定程度上可以弥补传统学术训练中指导不足的问题。但是,本研究发现 ChatGPT 存在增加学生惰性的风险,徐辉富指出,有了 ChatGPT 等的辅助,学术写作可能越来越容易,但也可能越来越平庸(邱燕楠,等,2023)。因此,怎样发挥 ChatGPT 的学习伙伴价值,同时又防止其导致的平庸,就需要思考如何在人机协同学术写作中扬长避短,平衡好两者的关系。

第三,ChatGPT 在人机协同学术写作中扮演的角色受到参与者不同学术训练背景的影响,总体呈现“遇强则强,遇弱则弱”的特点。与熟悉选题并有发表经历的研究生协同学术写作,ChatGPT 的表现也越强。这是因为这类参与者能够纠正 ChatGPT 生成的虚假信息及肤浅信息,识别 ChatGPT 生成内容的空洞与浮泛,并用自己的写作能力弥补机器表达上的缺陷。由此表明,传统写作中的关键因素依然是成功的人机协同学术写作的基础。熟悉选题和有发表经历的D同学与 ChatGPT 的协同学术写作,在一定程度上说明 ChatGPT 可以提升科研人员的工作效率,可以承担创新性不强的文字处理任务。这一发现呼应了王树义等(2023)学者的观点。在创造力和思维发展方面,已有观点(Else,2023)认为使用 ChatGPT 会妨碍学生批判力和创造力的发展。但本研究发现,熟悉选题且有发表经历的学生能对 ChatGPT 生成的内容进行批判性思考并在此基础上创新,因此 ChatGPT 不会妨碍这类学生批判力和创造力的发展,相反,能够成为他们的科研助手。而如果学生对 ChatGPT 存在的问题缺少感知、对生成的内容缺少批判性思考并将生成的内容奉为圭臬,则有可能变

得越来越平庸,并导致创造力和批判力的匮乏。此外,ChatGPT“遇强则强,遇弱则弱”的特点,在将来的人机协同学术写作时代,可能将进一步拉大学习者之间的差距,造成新的“数字鸿沟”。

第四,不同学术训练背景的参与者和观察者对 ChatGPT 学术写作的价值判断有所不同。不熟悉选题或无发表经历的参与者,往往从自身不足反思人机协同学术写作所需的知识及技能,这类学生也更容易感受到 ChatGPT 对他们的冲击与威胁;而熟悉选题和有发表经历的参与者往往审视 ChatGPT 的不足,进而思考内生性的技术问题对学术写作的影响。首先,ChatGPT 在与不同学术训练背景的研究生协同写作中表现出来的差异,影响参与者和观察者对 ChatGPT 应用价值的判断及人机协同学术写作所需知识技能的思考。不熟悉选题或无发表经历的参与者在人机协同学术写作中,首先看到的是自己的不足,而观察者看到的是 ChatGPT 的缺陷;熟悉选题且有发表经历的参与者在人机协同学术写作中,看到的是 ChatGPT 的问题,而观察者看到了自己能力的欠缺,从而产生提升研究能力的想法。参与者和观察者视角的倒置,一方面说明“做”与“看”的区别,另一方面说明学生并非总是以人类中心的视角看待 ChatGPT,而是能够从整体视角出发,反思人机协同所需的知识技能,并意识到自己的不足。其次,不熟悉选题或无发表经历的参与者更容易感知到 ChatGPT 对他们的冲击。另外,B 同学提到的“语言降维”和 C 同学提到的“自主性消失”,均属于参与者独特的“内视角”,只有在人机协同学术写作的基础上才能发现,而要找到这样的“内视角”,需要参与者具有一定的反思力,并对 ChatGPT 存在的问题保持足够警醒。由上述可知,在关于 ChatGPT 对研究生学术写作影响的这场争论中,应鼓励研究生群体在协同学术写作实践中进行自我评估与反思。

## 六、建议及启示

本研究发现,在人机协同学术写作过程中,ChatGPT 对不同学术训练背景的研究生有着不同的影响,也因此人机协同学术写作的重点在于学生要充分意识到 ChatGPT 的局限之处,以及自身学科基础及学术写作能力对于人机协同学术写作质量的重要性。鉴于研究生学科知识的积累以及学术训练需要较长时间的积淀,教育者与其担心 ChatGPT 被滥用于研究生学术写作,不如思考怎样在教学设计中融入 ChatGPT,通过与学生分享 ChatGPT 的优缺点以及人机协同写作的训练来逐步提升研究生的学术写作能力,以防科研能力尚未成熟的研究生被 Chat-

GPT 降维。基于本研究的发现,尝试提出以下建议:

第一,教师可以让研究生根据写作内容提出不同的问题,并将这些问题提交给 ChatGPT 进行回答,组织研究生就 ChatGPT 生成的内容进行讨论,比较不同的提问及回答,分析怎样的提问方式会使生成的内容更清晰,以此帮助研究生了解提问对于人机协同的重要性。之后,教师可以引导研究生通过评价、追问、示范等方式进一步和 ChatGPT 进行互动,由此引导学生如何人机之间产生高质量的对话。

第二,教师可以使用 ChatGPT 就某一主题生成一篇文章,再在课堂上组织研究生对生成的论文进行学习和发表评论。在课堂讨论中,教师引导研究生关注 ChatGPT 在语言表达、思路整合中存在的局限性,并就如何改进这些局限鼓励学生给出建议。对生成文章优缺点的评论有助于培养研究生的批判性思维。教师还可组织比较阅读,比较 ChatGPT 生成的内容和人类学者所写论文的异同点,并在比较中引导研究生发现机器人和人类在语言表达等方面存在的区别,以此提升研究生的学术鉴赏力。

第三,教师可以鼓励研究生开展人机协同学术写作,允许研究生在提交的论文中部分使用 ChatGPT 提供的帮助,但需要他们提供人机交互的全部过程,同时也需提交对相关内容的修改说明,鼓励研究生不断对人机交互的结果进行分析,通过添加信息、澄清观点、提供证据、增加分析等方式来完善文章,直到他们认为这是一个可以提交的版本。这一过程有助于研究生了解人在人机协同写作中的价值,只有人类通过适当的反思、批判、选择、加工等过程才能提升人机协同写作文本的质量。同时,教师也需帮助研究生明白人机协同学术写作的价值在于经历写作本身,并通过多次迭代看到论文的变化以及自我写作能力的发展。

第四,教师可以鼓励研究生在人机协同学术写作过程中开展反思和评价等活动,一方面主动思考并评估 ChatGPT 生成内容的质量,另一方面学会反思自身的投入和表现。在这一过程中,教师可以为学生提供指导,帮助学生了解 ChatGPT 的优缺点,更好地应对 ChatGPT 带来的挑战,以防研究经验尚未成熟的研究生被 ChatGPT 降维或挫伤研究积极性。同时,教师需开展伦理教育,引导研究生在符合学术伦理的条件下正确且合理地使用 ChatGPT,发挥其作为学习伙伴的价值,助力学业学习,而非替代人类写作。

本研究聚焦研究生论文选题熟悉度和发表经历对 人机协同学术写作的影响,但本研究的参与者仅为 4 名研究生,主要参与者为化工与食品科学领域



这两个理科专业学生,学科因素可能影响研究结果。此外,尽管本研究采用“最相似原则”对参与者进行选择,但无法控制4位参与者在其他变量上完全相似,因此可能存在未察觉的因素影响实验结果。后续研究可以在扩大样本量的基础上,通过准实验设计开展人机协同学术写作的相关研究,或提升样本的多样性,开展不同学科人机协同学术写作的实证研究,并在此基础上进行学科间的比较。

总之,人机协同学术写作并非“另谋新篇”,传统学术写作中的专业知识积淀、学术写作基本功等仍然是人机协同学术写作的关键与基础,而生成式人工智能等只是辅助科学研究的工具与载体。但是,这个工具对使用者的能力提出了要求,而且还带来不可忽视的伦理风险。如果使用得当,这会是一个得力的科研助手;若使用不当,不仅引发学术不端,而且会使学习者忽视自身学术能力的提升,从而导致自身智能的废弛与荒漠化(李政涛,2023)。面对 ChatGPT 等生成式人工智能带来的机遇与挑战,需要高校教学科研部门、教师和学生协同努力,在发挥机器效能的同时防止机器带来的流弊。简言之,即由人类引导机器,而非机器主宰人类。

### [参考文献]

- 常思亮,罗小丹,卢堃,2019.“差评”研究生学位论文主要特点鉴别[J].黑龙江高教研究(5):107-112.
- 陈向明,2000.质的研究方法与社会科学研究[M].北京:教育科学出版社.
- 陈文祥,2015.研究生学术论文写作渐进式培养模式研究[J].中国大学教学(10):60-64+16.
- 丹东尼奥,等,2006.课堂提问的艺术:发展教师的有效提问技能[M].宋玲,译.北京:中国轻工业出版社.
- 高文娟,徐颖,2022.美国博士生在读期间学术论文发表的实证研究——以19所美国研究型大学的经济学院为例[J].中国高教研究(9):102-108.
- 何文涛,路璐,周跃良,等,2023.智能时代人机协同学习的本质特征与一般过程[J].中国远程教育(3):12-20.
- 金皓月,李艳,2023.国内人工智能写作的研究现状分析及启示[J].现代远程教育(2):19-29.
- 教育部,国家发展改革委,财政部,2020.关于加快新时代研究生教育发展的意见[EB/OL].[2023-07-11].[http://www.moe.gov.cn/srcsite/A22/s7065/202009/t20200921\\_489271.html](http://www.moe.gov.cn/srcsite/A22/s7065/202009/t20200921_489271.html).
- 梁晨,李伟健,王屹,2022.自我迷失与理性回归:把脉人文社科类研究生的学术论文撰写——基于教育学视角[J].学位与研究生教育(6):25-31.
- 令小雄,王鼎民,袁健,2023.ChatGPT爆火后关于科技伦理及学术伦理的冷思考[J].新疆师范大学学报(哲学社会科学版)(4):123-136.
- 邱燕楠,李政涛,2023.挑战·融合·变革:“ChatGPT与未来教育”会议综述[J].现代远程教育研究(3):3-12+21.
- 斯腾伯格,2000.成功智力教学[M].张庆林,译.北京:中国轻工业出版社.
- 王立珍,袁金英,马秀峰,2012.研究生学术素养的内涵及培育探析[J].软件导刊(教育技术)(5):50-52.
- 王树义,张庆薇,2023.ChatGPT给科研工作者带来的机遇与挑战[J].图书馆论坛(3):109-118.
- 武法提,田浩,2023.人机协同的精准学习干预:动力机制、逻辑理路与实践模型[J].开放教育研究(2):81-90.
- 朱旭东,郭绒,2022.论学术论文写作训练:价值、方式和内容——基于“学术论文写作和规范”课程的经验[J].学位与研究生教育(6):6-15.
- BASIC Z, BANOVA A, KRIZIC I, et al., 2023. Better by you, better than me, chatgpt3 as writing assistance in students essays[J]. ArXiv, abs/2302.04536.
- CONROY G, 2023.Scientists used ChatGPT to generate an entire paper from scratch—but is it any good?[J]. Nature(619): 443-444.
- DERGAA I, CHAMARI K, ZMIJEWSKI P, et al., 2023. From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing[J]. Biology of Sport, 40(2): 615-622.
- EISENHARDY K M,1991. Better stories and better constructs: The case for rigor and comparative logic[J]. Academy of Management Review(3): 620-627.
- ELSE H,2023. Abstracts written by ChatGPT fool scientists[J]. Nature(7944): 423.
- FYFE P,2022. How to cheat on your final paper: Assigning AI for student writing[J]. AI & SOCIETY: 1-11.
- GAO C A, HOWARD F M, MARKOV N S, et al.,2023. Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers[J]. NPJ Digital Medicine, 6(1): 75.
- GERRING J,2008. Case selection for case study analysis: Qualitative and quantitative techniques[M]. The Oxford Handbook of Political Methodology, London: Oxford University Press.
- GUO B, ZHANG X, WANG Z,et al.,2023. How close is ChatGPT to human experts? Comparison corpus, evaluation, and detection[J]. ArXiv, abs/2301.7597.
- HAN Z, et al.,2023. An explorative assessment of ChatGPT as an aid in medical education: Use it with caution[J]. MedRxiv,02.
- NOY S, ZHANG W, 2023. Experimental evidence on the productivity effects of generative artificial intelligence[J]. Science(381):187-192.
- OWENS B,2023. How Nature readers are using ChatGPT[J]. Nature(7950): 20-20.
- RAHMAN M M, TERANO H J, RAHMAN M N, et al., 2023. ChatGPT and academic research: a review and recommendations based on practical examples[J]. Journal of Education, Management and Development Studies, 3(1): 1-12.
- WOLFRAM S, 2023. What is ChatGPT doing and why does it work?[EB/OL].[2023-03-27]. <https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>.
- VAN DIS E A M, BOLLEN J, ZUIDEMA W, et al.,2023. ChatGPT: Five priorities for research[J]. Nature (7947): 224-226.

收稿日期:2023年7月10日

责任编辑:吕东东

(下转第75页)



徐红,2016.教育测量与评价[M].武汉:华中科技大学出版社.  
徐纪良,1991.组织行为学手册[M].上海:上海科学技术出版社.  
张乐乐,顾小清,2022.多模态数据支持的课堂教学行为分析模型与实践框架[J].开放教育研究(6):101-110.  
郑勤华,陈丽,郭利明,等,2022.理论与技术双向驱动的学生综合素质评价新范式[J].中国电化教育(4):56-63.  
FLANDERS N A, 1963. Intent, action and feedback: A prepara-

tion for teaching[J]. Journal of Teacher Education,3(14): 25-260.

FLANDERS N A,1970.Analyzing teaching behavior[M].London: Addison-Wesley Educational Publishers:100-107.

收稿日期:2023年8月17日

责任编辑:吕东东

## Research on the Method of Calculating Standard Score Norm for Smart Classroom Interaction Behavior Data: Based on Improved Flanders Interaction Analysis System

Fang Haiguang, Kong XinMei, Zheng Zhihong & Huang Yingwenhui

**[Abstract]** The strategic transformation of national education digitization has driven the large-scale and systematic collection of smart classroom interaction behavior data, and the current analysis based on the collected data is based on statistical and empirical analysis methods. Therefore, how to establish a standard score norm model based on classroom interaction behavior data and hierarchical application of related dimensional data is a data-driven smart classroom interaction analysis method that needs to be explored urgently for the digitalization of national education. With the data basis of classroom interaction behavior, based on the Improved Flanders Interaction Analysis System (iFIAS), 358 lessons and 2,923,380 pieces of data from the iFIAS platform were selected to form the calculation method of the standard score norm model for classroom interaction analysis. Besides, a hierarchical application framework of “one overall, four main dimensions, and fourteen sub-dimensions” for the standard score norm was constructed, which resulted in the calculation results of the standard score norm for the classroom interaction behavior data of iFIAS. Finally, this method was applied to the quantitative analysis of classroom interaction behaviors of the fourth-grade math class “Recognize Triangles and Quadrangles” and the Seventh-grade English class “Do You Think You Will Have Your Own Robot?” The analysis compares the characteristics of the data-driven smart classrooms of the two cases and their intrinsic correlations, and discusses the results of norm analysis in five aspects: standard score norm, three-valued intrinsic logical relation, classroom characteristics, non-subject limitations, and norm drift.

**[Keywords]** Smart Classroom; Standard Score Norm; Interaction Behavior Data; iFIAS

(上接第48页)

## A Study of ChatGPT-Based Human-Computer Collaborative Academic Writing Practice for Graduate Students and Its Implications

Li Yan, Jin Haoyue & Yang Yuhui

**[Abstract]** The generative AI technology represented by ChatGPT brings great opportunities and challenges for human-computer collaborative academic writing. To further explore the potential value and risks of ChatGPT in human-computer collaborative academic writing, an attempt was made to explore the similarities and differences in collaborative academic writing between graduate students with different academic training backgrounds and those with the help of ChatGPT through case comparison. The findings of this study are as follows: first, graduate students who were familiar with the topic tended to ask more specific questions and engage in high-quality dialogue using such questioning methods as judging, probing and demonstrating. They tended to have a more fruitful and pleasant collaboration with ChatGPT, while students who were unfamiliar with the topic tended to have passive experience. Second, graduate students with publishing experience were good at identifying problems in ChatGPT language expression and writing methods, and could refine the generated content, while graduate students without publishing experience lacked such kinds of skills. Third, the role of ChatGPT in human-computer collaborative academic writing was influenced by the diverse academic backgrounds of graduate students, with the overall tendency of “strengthening the strong and weakening the weak.” Four, graduate students with different academic backgrounds had different judgments about the value of ChatGPT in human-computer collaborative academic writing. Those who were unfamiliar with the topic or had no publishing experience reflected on the knowledge and skills needed for collaborative human-computer academic writing from their own inadequacies, and were more likely to feel the impact and threat of ChatGPT, whereas graduate students who were familiar with the topic and had publishing experience thought more about the impact of the inherent technological issues of ChatGPT on academic writing. In view of this, educators are encouraged to prompt students to ask and inquire ChatGPT about writing content, and to analyze the impact of different questioning approaches on the quality of ChatGPT-generated content, to organize students to appreciate ChatGPT-generated essays, and to encourage students to actively engage in and reflect on the entire process of human-computer interaction.

**[Keywords]** ChatGPT; Collaborative Academic Writing; Case Comparison; Collaborative Behavior Variances