

机器写作中的性别刻板印象: 基于实验研究的实然探讨

✎ 牟 怡 蓝剑锋*

摘要: 性别刻板印象会影响读者对作者与内容的评价。基于 CASA (*Computers are Social Actors*) 范式, 本研究探讨在机器写作领域中, 被赋予了性别的机器作者是否也会受到性别刻板印象的影响。一个采用因子设计的在线实验收集了读者对不同物种、不同性别作者及写作文本的评价。研究发现性别刻板印象的确存在于大众对机器作者及内容的评价中; 但在具体类型的文本中, 读者对文本质量和作者的评价呈现出物种和性别的差异。

关键词: 机器写作; 性别刻板印象; 机器性别; CASA 范式; 人机传播

“人工智能不需要性别, 她本可以只是一个灰盒子……但是加入性别后会变得更有意思。”

——电影《机械姬》(*Ex Machina*)

一、引言

作者和作品的关系一直以来就是大众津津乐道的话题。尽管如钱锺书所说“假如你吃了个鸡蛋觉得不错, 何必认识那下蛋的母鸡呢?”但在作者

* 牟怡, 上海交通大学媒体与传播学院特聘副教授、博士生导师; 蓝剑锋, 上海交通大学媒体与传播学院博士。

的专业性之外,作者身上包含的种种属性,如道德品质、种族、信仰甚至是性别,都影响了大众对其作品的评价(葛岩等,2013)。在性别刻板印象的作用下,即便是同样的文本内容,读者也会根据作者性别给出不同评价。在人机传播益发普遍的时代,这样的现象是否亦会出现在机器写作的场景中?

让机器人拥有性别似乎是天方夜谭,毕竟性别是人类的生理构造才具有的特征;我们无法把铜丝看成男性,把硅晶芯片看作女性。因此有学者认为,即便是那些看似被赋予性别的人工智能,也只是在进行巴特勒(Butler)所谓的“性别表演”(Gender Performativity)(李思雪,2015)。机器的性别因创造者的需要而改变,并不如人类性别一般生来很难变更。性别的随意和可切换消解了性别之于人类的深远意义,也使机器的性别议题缺乏严肃性。然而,无论是像《机械姬》这类科幻电影中的机器主角,还是亚马逊 Alexa、微软 Cortana、苹果 Siri 等人工智能助手,都无一例外地被开发者赋予了性别。因为不含性别的客体在我们的潜意识里往往是不可交流的,我们更希望自己的交流对象是“他(Him)”或者“她(Her)”,却很少与“它(It)”进行对话(Hu et al., 2019)。作为人类交互对象的人工智能在拥有性别符号后,可以更加成功地融入人际交流的语境。虽然机器人无法真正拥有性别,但为机器人赋予性别符号,确实能够使其更好地适应人类社会的交往规则,并提高用户的接受度(许丽颖、喻丰,2020)。

写作机器人作为人工智能技术发展的产物,也往往通过名字等方式被暗示性别。我们无意探讨其应然问题,而是尝试探索当读者面对由算法驱动的机器男作家或女作家时,性别刻板印象是否还会发生作用的实然问题。因此,本文基于实验设计对可能存在的机器写作的性别刻板印象展开实证研究。

二、文献综述

(一) 机器写作与质量评估

机器写作(Machine Writing)是指运用算法将数据加工成新闻文本(Carl-

son 2015)。《洛杉矶时报》自 2007 年起开发了有关凶杀案件报道的内容生成软件,2011 年开始采用的 Quakebot 机器人则真正实现了文本自动生成,这标志着机器人能够以独立的作者身份生产内容(何苑、张洪忠 2018)。2014 年美联社宣布启用 Automated Insight 公司开发的自动写作软件 Wordsmith 撰写关于公司财务情况的新闻报道,这让他们每季度的财报数量从 300 篇上升到 4400 篇(金兼斌 2014)。2015 年后,国内媒体和互联网公司开始开发机器写作引擎,如腾讯的 Dreamwriter、新华社的“快笔小新”、今日头条的 Xiaomingbot 等。如今,机器写作引擎已经被海内外众多媒体广泛采用。

使用机器进行新闻写作,能够提高新闻生产速度、减少技术差错和生产成本(喻国明等 2017)。在新闻的时效性上,机器新闻写作者将人类作者远远抛在了身后,机器作者可以帮助媒体更快速、更准确地生产新闻消息。此外,机器作者可以对海量的碎片数据进行汇总和分析,挖掘出以往人力难以发现的文本价值,再辅以可视化呈现的手段,便能让人们获得“上帝视角”一般的阅读体验(喻国明 2015)。机器写作的这一特性将新闻生产的质量和数量提升到一个新的高度。因此,当前不少学者认为机器写作势必成为新闻行业未来发展的重要方向之一。

此前对于机器写作的有关研究多集中在机器生产新闻的发展模式、对传媒行业的影响等议题上,而对机器自动生成内容的评估略有提及(牟怡等 2019)。机器写作的优势固然显著——能更快速、精确、有效地整合、处理大数据,但自动生成内容的专业性、可读性以及当中包含的价值判断、道德水平等依然需要人类读者进行评估。写作机器人本身包含着影响读者评价的各种因素,诸如机器人的外形、采用的编程技术、应用的平台等。当我们看到机器作者的外形是一个相貌端正的人类,我们一般会认为“他”创作的作品要优于一个长方形盒子;毕竟我们更容易相信一个“人型”机器人可以像人类一般书写(刘茜 2020)。同样,如果该写作引擎研发自知名科技公司,如微软、谷歌等,基于对品牌原有的信任程度,我们也会更信任写作机器人的能力(刘茜 2020)。除了对机器人新闻生产主体的评价,对机器人新闻内容的评价则是我们更关心的问题。

评价机器人写作的质量,可以参考传统的新闻质量评价标准。此前的

研究认为,新闻可信度是衡量新闻质量的重要因素,一般可以从信息源、内容、媒介几个维度进行测量(Chung et al. 2010)。有学者提到,要想提高内容的可信度,信息最好来自可靠的权威机构(Briggs et al. 2002)。也有研究表明,内容生产者具备更好的专业能力和诚实、正直的品德对于提高内容可信度具有正向作用(McGinnies & Ward, 1980)。

范德卡和克拉默(Van der Kaa & Krahmer 2014)调查了新闻记者和普通读者对机器人新闻可信度的评估情况,结果显示人们认为机器人新闻的可信度与人类撰写的新闻可信度相差无几。这说明不同专业背景的受众难以区分人工写作新闻和机器写作新闻。但在沃德尔(Waddell 2017)的研究中,人们普遍认为机器编写的新闻不如人工编写的新闻可信,新闻写作在很大程度上仍被视为是“人类的工作”。得出这样结论的原因既可能是机器新闻内容客观的短板,也有可能仅仅出自人们对机器作家身份的刻板印象。

克莱瓦尔(Clerwall 2014)在2014年使用桑德尔(Sundar, 1999)的新闻感知量表探讨了受试者对新闻质量和可信度的感知程度。研究结果显示,受众认为机器写作的内容更枯燥,也更为客观。爱德华兹(Edwards et al., 2014)在关于Twitter机器人的实验测试中也得到了类似的结果:人类书写的推文往往被认为更具有吸引力。国内对机器新闻的态度调查呈现出相似的结果。2016年腾讯的一项调查显示,41.6%的用户认为机器缺少新闻判断力,39%的用户认为机器缺少人情味;而多达76.4%的自媒体人担忧机器写作面临缺乏情感和个人风格的问题(彭兰, 2016)。许向东等人(2017)认为自动化新闻虽然加快了新闻生产速度、拓宽了新闻报道范围、提高了稿件的准确性,但也面临着报道主题和题材有限制、可读性低、语言枯燥乏味等问题。

在算法渗透进生活各个角落的今天,机器人写作在新闻伦理、新闻价值等方面依然存在诸多问题。新闻专业主义也要求新闻生产者不断提高专业水平,这使机器写作被放置在更严格的评估环境中。要改进机器写作,我们必须对机器人作者的各个方面进行考察。

(二) 性别刻板印象与写作

性别刻板印象(gender stereotype)是人们广泛接受的对男性和女性的固定看法(Waddell 2017)。人们通常认为男性特质包括积极、果断、独立和进攻性;女性特质则包含善良、同情、关心他人等(Heilman 2010)。社会中的性别刻板印象不仅认为两性之间存在显著差异,并且更倾向于认为男性具有更多优势。有学者调查了100位被试者的性别刻板印象,发现在内隐层面,无论男女,都倾向于给男性以更加正面的评价,给女性以更为负面的评价(徐大真 2003)。也有学者调查了380位中国大学生的性别刻板印象,结果显示大学生多认为男性在思维、能力、工作上超过女性,更加坚强、能干;而女性在情感性的项目上超过男性,如善解人意、重感情,因此显得更为被动、顺从(钱铭怡等,1999)。不仅中国语境下的性别刻板印象如此,美国学者鲁宾等人(Rubin et al., 1974)对几对新生儿父母进行了访谈,结果发现女孩父母往往对孩子的评价更为正面,认为女孩更加温柔、可爱。性别刻板印象跨越了地域、种族、信仰和文化,足以显示它多么深刻地影响着人类社会的认知。

按照此前性别和语言使用关联的研究结论,关于男性和女性语言使用的解释框架可以分为两类:支配论和差异论。支配论认为语言的差异可能是男性主导和女性从属的历史带来的结果(Fishman, 1978; Robin, 1973)。而差异论则认为男性和女性属于不同的亚文化,因此语言的差异来自文化的差异(Coates 2014)。语言使用的差异在罗宾(Rubin, 1973)此前的口语研究中得到了证实,其研究结果显示,女性倾向于在陈述句中频繁使用强调词、限定词、空洞的形容词。性别刻板印象不仅存在于语言使用的习惯和方式,甚至在声音选择上也存在不同性别的偏好。研究发现,绝大多数的新闻客户端只提供一种性别的声音,而大部分新闻客户端也呈现出“男声用于硬新闻,女声用于软新闻”的偏向(黄雅兰 2021)。在商场、地铁和机场等服务消费场景中,也往往使用女声进行播报。女性作为服务角色的刻板印象已经深入到社会共识中。

2002年,科佩尔(Koppel 2003)及其同事设计了一个计算机程序,用于

扫描不同性别作者作品的关键词。结果显示男性作者更倾向于讨论实物 (objects) ,并较多地使用限定词(如 a ,the ,that) 和量词(如一 ,二 ,更多);而女性作者更倾向于讨论关系 (relationships) ,且更喜欢使用代词(如你 ,我 ,他 ,她们的)。赖(Lai 2009) 收集了部分书籍和博客文本 ,并创建了一个识别和区分男性和女性语言使用的算法 ,分析结果显示 ,男性作者更常使用分号 ,女性作者更喜欢使用具有情感表现力的感叹号。哈姆丹(Hamdan 2013) 针对人们能否感知阿拉伯语言中不同性别作家语言风格的差异进行了一项调查。研究结果支持了特定的男性或女性写作风格的存在。上述研究结论证实了作家性别确实会影响其遣词造句的习惯和偏好。

写作是一项基于社会认知和社会体验的创造性工作 ,写作行为要求写作者本身具有一定的社会经历。社会文化催生了语言的性别差异 ,而不同社会文化中的作者自然会在作品中折射出根据其性别角度的社会认知(兰玉 2007) 。也正因此 ,无论是从作者写作角度还是读者认知角度 ,性别刻板印象都产生了影响。

(三) 机器人性别与机器的刻板印象

性别符号是身体符号最为强势的特征。如弗洛伊德(Freud)所言“当遇到一个人时 ,你们所做的第一个区分是‘男人’或‘女人’。”尽管机器人是否应该拥有性别特征广受争议 ,但机器人作为人类生命的仿制品 ,其身体也不可避免地带有性别的符号或暗示。在科幻电影中 ,机器人往往以加强的性别符号出现在观众面前 ,或是强健的体格 ,或是完美的身材。在人机交互的关系上 ,为机器设定性别 ,可以让人与机器人在更原始的语境和立场上进行对话 ,也能更大程度还原真实的交流。无论机器人是否具有实体 ,都可以发现其身上的性别符号。微软小冰的设定是一位 18 岁的少女;亚马逊 Alexa 的名字则来自电影《星际迷航》中一位身材诱人的女性助手;小米则为其开发的智能语音助手小爱同学设定了不同性别、不同年龄的声音 ,以满足用户的多样化需求。性别似乎已经成为机器人必不可少的特征。

卡彭特(Carpenter)团队是最早进行机器人性别研究的团队之一 ,他们探索了人们对家用仿人机器人的期望 ,发现被试者更喜欢女性机器人来做

家务活(Julie et al. 2009) 。2019 年剑桥大学未来智能研究中心发布的报告《人工智能与性别: 对于未来研究的四项建议》提到,类人外表的机器人所包含的声音、形象和行为都在模仿并强化着社会中的性别刻板印象(黄雅兰, 2021) 。也有学者就机器人性别感知问题进行过研究,如野村和高木(Nomura & Takagi 2012) 测试了不同教育背景的人如何感知机器人性别。他们仅仅使用姓名来控制机器人性别,并对来自自然科学教育背景和人文科学教育背景的受试者进行测试,结果显示自然科学教育背景的受试者认为男性机器人表现出更果断的特质,而女性机器人在两组被试中没有被感知出明显的区别。

除了机器人性别本身的意义问题外,在人机交互的过程中,不同性别的人群对不同性别的机器人的偏好同样值得关注。西格尔等人(Siegel et al. , 2009) 曾进行过一项实验,在互动后一个机器人要求参与者捐款并填写一份问卷;机器人的性别由声音控制。结果显示,男性受试者认为女性机器人比男性机器人更可信,而女性受试者的结果则相反。野村(Nomura et al. , 2008) 在日本进行的一项调查发现,在 238 名参与者中,女性在和机器人互动时,表现得更为消极。库钦布兰特等人(Kuchenbrandt & Markus 2014) 在一项人机协作实验中,让不同性别的参与者与带有不同性别特征的机器人组队完成任务,结果显示无论机器人性别如何,女性参与者完成任务的速度更快,而男性参与者在与男性机器人搭档时,完成任务效率更高。这些研究结果显示,在对机器人的性别问题上,机器人的性别不同、实验参与者的性别不同,人们对机器人的态度也存在显著的差异。

在人机传播领域,许多研究聚焦于人们是否会将自己的职业特征和对性别的刻板印象转移到特定性别的机器人上。在埃塞尔和黑格尔(Eyssel & Hegel 2012) 的研究中,人们认为带有典型男性外貌的机器人更适合执行男性任务,而装饰着长头发的机器人则被认为更适合执行女性化的任务。泰(Tay et al. 2014) 等人对职业刻板印象的另一项研究证实了埃塞尔等人的发现。在泰的实验中,研究者通过语言暗示(主要通过声音) 和典型的名字操纵机器人的性别。参与者与机器人互动后表示,女性机器人更适合在医疗场景从业(这往往被认为是典型的女性职业),男性机器人则更应该从事安

保领域相关工作(这被认为是刻板的男性职业)。克劳斯(Kraus, 2018)对机器人作者的外显和内隐性别刻板印象进行过研究。结果显示,男性机器人被认为更值得信赖,女性机器人则被认为更可爱。

(四) CASA 范式与机器写作中的性别刻板印象

1994年,来自斯坦福的学者纳斯等人(Nass et al., 1994)提出“计算机为社会行动者”(Computers are Social Actors,简称CASA)范式,指明人机交互本质上是社会性的。1996年,李维斯和纳斯(Reeves & Nass, 1996)基于CASA范式提出“媒体等同理论”(Media Equation Theory)。他们的研究发现人们会将人类社会中的交往规则应用到人机交互当中,其中不仅包含“礼仪”“关心”等情感,还包含“性别刻板印象”等观念。男性声音的计算机被认为更具有说服力,而女性声音的机器人则被描述为一个更擅长“爱”和“人际关系”的老师形象。

在纳斯等人的研究中,人类能被机器礼貌奉承的话语所影响,就像被来自人类的话语所影响一样(Nass et al., 1995)。因此,媒体等同理论提出人们会像对待真人一样对待媒体技术。这一思想为媒体技术赋予了“人性”,使媒体技术突破了原有的工具论和客体论局限,将媒体技术的地位带到了新的高度。在人机交互的过程中,作为媒体的机器人可以将媒体内容以更人性化的方式展示出来,通过文字、画面、声音,甚至是人性外表传达出动作、暗示的眼神、神秘的微笑等,这使人类受众从心理上将机器看作是与人类相当的交往对象,此种交往过程有学者认为是无意识的“主体间”交往(梅琼林、张晓, 2006)。

在CASA范式的指导下,机器从以往被动设置的客体,变为能够主动产生刺激作用的研究对象,这也打开了人机传播的研究思路。如有学者基于CASA范式,对人工智能语音助手的个性(personality)展开研究。通过“绿野仙踪法”实验发现,智能语音助手与用户之间存在个性的“相似吸引效应”,即人机交互双方个性相似时,用户对语音助手的社交信任感更高(王鑫蓉、王玫, 2022)。也有学者从心理学角度出发,基于艾森克人格问卷(Eysenck Personality Questionnaire)对社交机器人微软小冰的人格进行测量。结果发

现微软小冰在与人类对话的过程中折射出了可测量的人格面,且表现出外向、好交际、易冲动、渴望刺激的人格特点(韩秀,2021)。

人类在把机器作为等同于人类的交互对象时,不自觉地会代入各种人际交往的观念、习惯和偏见。以文字内容中的性别刻板印象为例,有学者发现不同主题的新闻报道所含的性别偏见不同。当涉及财经、体育、政治以及一些专业性话题时,性别刻板印象的影响尤其突出(Flaounas et al., 2013)。既然性别刻板印象普遍存在于人类社会中,且在内容生产领域影响着读者对作者以及内容质量的评价,那么根据 CASA 范式,在机器写作中,机器作者的性别暗示亦会引发类似的刻板印象反应。因此,我们提出如下研究假设:

H1: 读者对人类作者的性别刻板印象会延伸到机器作者身上,即当机器作者被设定为男性时得到的作品内容质量的评价会异于女性,且两者间的优劣程度与人类刻板印象中一致。

H2: 读者对人类作者的性别刻板印象会延伸到机器作者身上,即当机器作者被设定为男性时得到的作者专业度的评价会异于女性,且两者间的优劣程度与人类刻板印象中一致。

机器性别的设定中存在非男非女的中性选项,这个性别设定是否会引发不一样的评价?同时根据前文综述,读者的性别也会影响其性别刻板印象感知。就此,我们提出以下两个研究问题:

RQ1: 当机器作者为中性性别时,作品所受到的(a)内容质量和(b)专业度评价与其设定为男性或女性时是否存在差异?

RQ2: 受试者的性别是否会影响机器写作中的性别刻板印象?

三、研究方法

(一) 研究过程与实验素材

为了验证提出的假设和回答研究问题,本研究进行了 2×3 因子设计在线实验。其中第一个自变量为作者物种属性,即人或是机器人;第二个自变量为作者性别,通常来说即男性或者女性;然而对机器人而言,中性是一个

常见的选项,因此添加此选项。鉴于人类的中性性别并不常见,故在人类作者中仅存在男性或女性。因此,实验存在五组实验组别。同时,本研究也考察受试者性别的影响,但是该自变量不被操纵。在每组实验中,受试者首先阅读一篇含有作者物种身份及性别信息的文字内容,然后让他们填写相应问卷,对所阅读的材料进行评价;最后依据研究伦理,告知受试者真实的实验目的。

本研究的目的是考察人类基于文字的性别刻板印象是否会延伸至机器写作上,因此首先找到带有人类性别刻板印象的文字内容至关重要。为此,本研究尝试了三个常见的可能带有性别刻板印象的话题:两性情感、家庭理财和信息技术。人们通常认为女性更擅长情感话题和家政话题,而男性更擅长技术话题。根据网上常见的情感建议、家庭理财建议和网络安全建议写作套路,笔者针对三个话题各改写出一篇建议性文章。为了排除可能的混淆变量,每篇文章统一有715字左右,其中列举出四条可行建议。例如,在《数据安全专家的四个隐私保护建议》一文中,列举出包括“设置浏览器关闭时自动清除cookie”在内的四项在上网浏览的过程中保护个人隐私安全的建议。

针对每篇文章,我们首先进行了人类性别刻板印象的测试(见下文),即同样的文字内容,却告知不同的作者身份,一个是女性作者,另一个是男性作者。前者采用了典型的女性名字,后者则是典型的男性名字。为了强化性别特征,还采用了常见的作者画像方式,通过一个略带写实风格的头像显示作者的性别。通过刻板印象有效性检验的文字内容即被采用为实验素材。

在机器写作的三组里,机器人身份与性别亦被名字标明。为了避免诸如“小明”这样的中文名字带来的性别模棱两可性,笔者特地选取了英文中具有明确性别信息的名字。“Emma 机器人”“Mike 机器人”和“Star 机器人”分别代表女性机器人、男性机器人和中性机器人。同时辅以带有明显性别特征的外貌形象,以此强化性别特征。

(二) 样本

为了保证样本数据具有足够的效力,实验开始前我们首先进行了效力分析(power analysis)。从文献中获取最接近的效应值(effect size d)为0.32 (Swim & Janet, 1994)。采用 G* Power 3.1 (Faul et al., 2009) 计算,当显著度为0.05,效力为80%时,要求总样本量至少为168。

本研究通过问卷调查平台“问卷星”选取实验对象并完成线上实验。研究采用了问卷星的样本服务,从该调查平台的抽样框(sampling frame)中抽取受试者并随机分配到10组中的1组。每组平均约有45名实验参与者,共计451人。在排除了无效问卷和未能通过操作检验的问卷后,总共392份有效问卷进入了最终的数据分析,每组人数为36人到44人不等,平均为39人,保证了样本数据具有足够的效力。

在392份有效问卷中,48.2%(189人)来自男性,51.8%(203人)来自女性。受试者的年龄为18岁到65岁,平均年龄为29.15岁($SD=6.73$)。学历覆盖从高中以下到研究生或以上,大部分(74.5%)受试者拥有大学本科学历。而他们的收入水平则相对均匀分布:17.9%的受访者税前月收入在0—3500元之间,19.4%的受访者在3501—6000元之间,31.4%的受访者在6001—9000元之间,17.9%的受访者在9001—12 000元之间,13.5%的受访者在12 001元以上。从未或基本没有接触过机器写作内容的人并不多,只占总人数的6.1%(24人);偶尔接触的人占38.8%(152人);有时接触的人占41.1%(161人);经常接触的人占12.8%(50人);而5人(1.3%)则总是接触到机器写作内容。这说明机器写作的流行虽然只有短短几年时间,但是已经为不少人所知。

(三) 变量测量

主要因变量的测量量表来自英文文献,并通过反向翻译(back translation)的方法翻译成中文以保证其准确性。所有变量均使用5级量表进行测量。其中,对感知的内容质量测量借鉴桑德尔(Sundar, 1999)的测试量表,8个问项间的信度克伦巴赫系数 $\alpha=.70$ 。作者的专业度评价借鉴范德

卡(Van der Kaa) 和克拉默(Krahmer) 的 5 个问项 ,其信度克伦巴赫系数 $\alpha = .66$ 。

控制变量包括对 AI 的态度、性别刻板印象感知和机器写作内容接触频率。将阿杰恩(Ajzen ,1991) 的态度量表根据研究语境进行改写得到对 AI 态度的语义差异量表(semantic differential scale) ,5 个问项间的信度克伦巴赫系数 $\alpha = .72$ 。感知到的性别刻板印象通过 3 个自建问题来测量 ,包括“一般而言 ,女性比男性更善于家庭理财” ,“一般而言 ,女性比男性在处理情感问题上更具有优势”和“一般而言 ,男性比女性更擅长技术领域的工作” ,其信度克伦巴赫系数 $\alpha = .71$ 。机器写作内容接触频率通过单一问项“你阅读到机器写作内容的频率”来测量。人口统计学变量包括受试者的性别、年龄、教育程度、收入等常规测量项目。

(四) 操纵检验

为了保证实验操纵的有效性 ,除了问卷星平台自带的注意力检验外 ,本研究还进行了操纵检验。受试者在进行评价之前被要求回忆“你刚才阅读的内容的作者是谁”。那些未能选择正确答案的受试者问卷被剔除掉。

四、分析结果

(一) 实验素材有效性检验

在主题分别为两性情感、家庭理财和信息技术的文字内容中 ,情感建议一文中并未出现预期的性别刻板效应 ,女性作者写作的内容质量($M = 3.98$, $SD = 0.67$) 与男性作者的内容质量($M = 3.89$, $SD = 0.43$) 并无显著性差异: $t(78) = 0.72$, $p = 0.48$; 而女性作者的专业度($M = 4.07$, $SD = 0.74$) 亦与男性作者($M = 4.06$, $SD = 0.53$) 无异: $t(78) = 0.07$, $p = 0.95$ 。因此 ,情感建议一文被排除在实验素材之外。

对家庭理财建议一文 ,人类作者的性别具有显著性的差异 ,也即: 女性作者写作的内容质量($M = 4.09$, $SD = 0.45$) 显著高于男性作者写作的内容质

量($M = 3.85$, $SD = 0.58$) : $t(76) = 2.10$, $p = 0.039$ 。而女性作者的专业度($M = 4.05$, $SD = 0.45$) 亦边缘显著高于男性作者的专业度($M = 3.86$, $SD = 0.43$) : $t(76) = 1.88$, $p = 0.064$ 。这说明 ,家庭理财建议这一话题确实带有传统认知上的女优于男的性别刻板印象 ,因此作为后续实验素材使用。

而对网络信息安全建议一文 ,人类作者的性别亦带来显著性的差异 ,也即: 女性作者写作的内容质量($M = 4.05$, $SD = 0.42$) 显著高于男性作者写作的内容质量($M = 3.82$, $SD = 0.53$) : $t(82) = 2.18$, $p = 0.032$ 。但女性作者的专业度($M = 4.22$, $SD = 0.54$) 与男性作者的专业度($M = 4.18$, $SD = 0.43$) 无异: $t(82) = 0.40$, $p = 0.69$ 。值得一提的是 ,网络信息安全建议一文中体现的性别刻板印象与传统认知相反 ,并非男性比女性更擅长这两个话题。然而本文的目的并非验证传统的性别刻板印象是否成立 ,而是考察这些刻板印象是否在机器写作中依然成立 ,因此该文也被用作后续实验素材。

(二) 主效应分析

表 1 不同作者类别下的感知内容质量和专业度

	女·机	男·机	中性·机	女·人	男·人
【家庭理财建议】					
内容质量	3.90(0.38) ^{ab}	3.85(0.45) ^b	4.01(0.37) ^{ab}	4.09(0.45) ^a	3.85(0.55) ^b
专业度	3.95(0.45) ^{ab}	3.85(0.47) ^{a#}	3.74(0.44) ^b	4.05(0.45) ^a	3.86(0.43) ^{b#}
【信息安全技术建议】					
内容质量	3.92(0.53) ^{ab}	4.02(0.39) ^{a#}	3.97(0.51) ^{ab}	4.05(0.42) ^a	3.82(0.53) ^b
专业度	4.26(0.39)	4.13(0.52)	4.22(0.50)	4.22(0.54)	4.18(0.43)

注: 数值为平均值(括号内为标准方差) ; a/b 表示具有显著性差异 #表明边缘显著性差异。

二元协方差分析(Two-way analyses of covariance) 结果显示(见表 1) ,在控制对 AI 的态度、性别刻板印象感知和机器写作内容接触频率 3 个控制变量后 ,在家庭理财建议素材的刺激下 ,出现显著的性别主效应: 对内容质量而言 , $F(2,187) = 3.23$, $p = 0.042$ 。女性作者写作的内容质量($M = 3.99$, $SD = 0.42$) 、中性作者写作的内容质量($M = 4.01$, $SD = 0.37$) 和男性作者写作的内容质量($M = 3.85$, $SD = 0.50$) 存在显著性差异; 事后 LSD (Least - Significant

Difference) 检验显示前两者显著高于后者。而对专业度而言, $F(2, 187) = 3.67$, $p = 0.027$ 。女性作者的专业度 ($M = 4.00$, $SD = 0.45$)、男性作者的专业度 ($M = 3.86$, $SD = 0.44$) 和中性作者写作的专业度 ($M = 3.74$, $SD = 0.44$) 存在显著性差异; 事后 LSD 检验显示前者显著高于后两者。然而, 物种并不存在显著的主效应: 对内容质量而言, $F(1, 188) = 1.65$, $p = 0.20$; 对专业度而言, $F(1, 188) = 0.60$, $p = 0.44$ 。

在信息安全技术建议素材的刺激之下, 物种并不存在显著的主效应: 对内容质量而言, $F(1, 200) = 0.21$, $p = 0.65$; 对专业度而言, $F(1, 200) = 0.001$, $p = 0.98$ 。性别亦不存在显著的主效应: 对内容质量而言, $F(2, 199) = 0.33$, $p = 0.73$; 对专业度而言, $F(2, 199) = 0.72$, $p = 0.49$ 。

就 RQ2 而言, 不同性别的受试者在阅读家庭理财建议后并不会产生不同的认知: 对感知质量 $t(188) = 0.49$, $p = 0.63$; 对专业度 $t(188) = -0.17$, $p = 0.83$ 。而不同性别的受试者在阅读网络安全建议后感知的内容质量并无差异: $t(200) = 0.52$, $p = 0.61$; 但男性受试者对专业度的评价 ($M = 4.27$, $SD = 0.49$) 高于女性受试者的评价 ($M = 4.13$, $SD = 0.45$): $t(200) = 2.06$, $p = 0.04$ 。

(三) 交互效应分析

在家庭理财建议素材的刺激下, 物种与性别对于内容质量和专业度评价不存在显著的二元交互效应: 对内容质量而言, $F(1, 188) = 1.66$, $p = 0.19$; 对专业度而言, $F(1, 188) = 0.33$, $p = 0.57$ 。然而, 在信息安全技术建议素材的刺激之下, 物种与性别对内容质量而言存在显著的二元交互效应: $F(1, 200) = 5.03$, $p = 0.03$; 不过对专业度的交互效应却不显著: $F(1, 200) = 0.36$, $p = 0.55$ (见图 1)。换言之, 同样是女性, 人写作的内容质量显著高于机器人写作的内容质量; 而是男性时, 人写作的内容质量则显著低于机器人写作的内容质量。综上, H1 和 H2 均部分成立。

就 RQ1 而言, 机器人的中性性别并不会总带来显著影响; 只有在家庭理财建议中, 中性机器人的写作质量高于男性机器人, 但专业度却低于女性机器人。针对 RQ2, 受试者性别与其他两个自变量 (作者物种与性别) 并无任何显著性二元和三元交互效应。

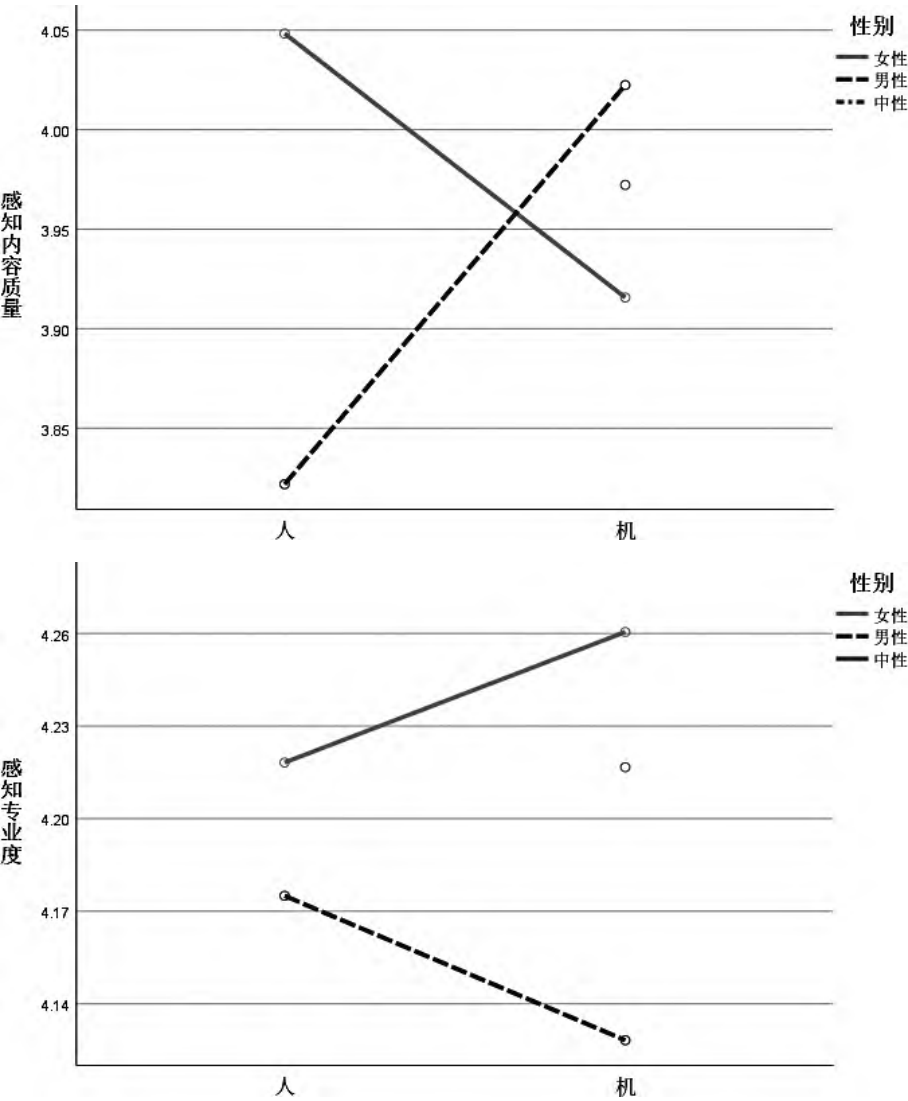


图 1 信息安全建议中物种与性别的二元交互效应

五、结论与讨论

1950 年, 艾伦·图灵(Alan Turing) 在《计算机与智能》(“Computing Machinery and Intelligence”) 这篇论文中对计算机的思考能力进行了研究, 提出

了著名的“图灵测试”(The Turing Test)。测试的具体内容是让一个人和一台机器躲在幕后,测试者分别向两个对象提问,借此分辨出机器人和人类。经过多轮测试后,如果机器使每个参与者平均误判率超过30%,则认为该机器通过了测试。如今,图灵测试也被认为是测试机器是否具有智能的代表性方法。自1956年达特茅斯会议正式提出“人工智能”概念至今,尽管经过60余年的发展,“强人工智能”(strong AI)和“弱人工智能”(weak AI)的边界开始变得模糊,但从人机协作、辅助生成的角度来看,如今的写作机器人依然属于弱人工智能的范畴(莫宏伟,2018)。为了让机器生成的内容被更多人接受,研究者应该考虑提升写作机器人的智能水平和仿真程度,比如提升算法水平,使其生成的内容更生动,或者赋予其社会身份和性别,让写作机器人更贴近人类作者的方方面面。

在机器写作越来越普遍的今天,如何更好地设计写作机器人成为一个值得深入探讨的问题。本研究聚焦于机器写作中存在的性别刻板印象,通过实验研究检验了大众对机器作者的性别刻板印象,并针对不同类型的文本内容分别讨论。结果显示,性别刻板印象确实存在于机器作者身上:尽管有违普遍的直觉,但在评价文本内容的过程中,人们真的会像评价真人一样基于性别评价机器作者。物种和性别的交互效应生动地体现性别在机器作者身上的作用,尽管这样的反映并不完全复制了人类写作中的性别刻板印象模式。

研究发现,在信息安全建议文本测试组中,物种和性别对写作内容的质量具有显著的交互效应。同为人类时,女性作者质量评价高于男性;然而同为机器时,作品质量的评价为男性最高,中性其次,女性最低。由此可见,对于信息安全技术这类文本,人类世界中的性别刻板印象在机器作者身上发生了逆转。当作者物种从人类变为机器,读者的选择也从女性转为男性,即人们对人类作者的性别刻板印象确实同样存在于机器作者身上,且当机器作者被设定为男性时,作品质量获得了更高的评价。

在控制了对AI态度、性别刻板印象感知以及机器写作内容接触频率三个变量后,研究结果显示,对于家庭理财和信息安全技术两类内容,受试者均认为女性作者创作的文本质量更佳。在信息安全技术领域,这样的结果

与以往的性别刻板印象相反,因为无论是从业者还是专业学习人数都呈现典型的“男多女少”。其中一个可能的原因是女性往往需要付出更多的努力,才能被行业和大众认可,得以进入这一领域。因此在考虑到诸多潜在的社会因素和投入的努力之后,人们可能会认为该领域的女性作者能力更强。此外,单就作者性别而言,人们普遍认为女性作为更仔细、更认真的角色,能够更好地生产出高质量、易读的文本内容。女性在写作过程中,更多地使用比喻等修辞手法,用更为感性和通俗的语言,使文本的可读性大大提升,这样读者能够更好地理解文本内容。带着这样的刻板印象,人们往往会认为女性作者创作的内容质量更高,倾向于给女性作者文本质量更高的评价。因此,人类女性作者的文本质量在所有组别中得分最高。

但对机器作者而言,读者并不会考虑太多社会因素。机器作者相较于人类作者,身份背景被简化了许多。人们不太愿意相信电脑前的这个机器作者是背负了家庭责任和千百年来重男轻女的传统观念,经过重重考核选拔最终成为“专家”的。这里被赋予性别的机器作者身上携带的更多是普遍意义上人们对性别的认知。因此,基于传统的性别刻板印象,尤其是像信息技术安全这样传统的技术、理工类专业领域,男性占据主导地位的局面卷土重来,对于内容质量的评价,男性机器作者便高于女性机器作者。基于CASA范式的视角,我们可以看到,虽然机器生成内容呈现出的性别刻板印象的具体情况不同,但性别刻板印象这一现象本身确实存在。从人们对女性作者的偏好,到对物种和性别显著的二元交互效应,不难发现当机器作者被赋予性别后,人们便不再像以往那样单纯地评价性别或者评价机器作者,而是兼有传统的性别刻板印象以及对机器作者身份的评价。

研究同样显示,无论是人类作者还是机器作者,受试者均对其专业度评价没有显著差异,这呼应了前人的研究——范德卡和克拉默(2014)的研究显示,普通大众无法区分人类作者和机器作者创作内容的可信度与专业度。机器写作的专业度更多取决于文本本身,性别影响甚微。而本研究的结果显示,性别、物种都没有显著影响读者对专业度的评价,因此我们认为在CASA范式下,为机器赋予性别,使人们容易产生对内容“质量好坏”的评价差异。这种影响更多是带给读者主观上的体验。当涉及作者专业性的检验

时,读者会更关注文字本身,而忽略作者的因素。

为机器人赋予性别是当前人工智能发展过程中日渐得到默认的选项。在 CASA 范式的视角下,我们不仅会像对待真人作者一样对待机器,评价其生产内容的质量,还会将机器作者的性别纳入考量,尽管这样的性别设计往往非常简单。这为人工智能研发提供了新的思路,即在设计人工智能程序时,将人们固有的性别刻板印象考虑在内,也许就能放大用户对人工智能的需求,或者减少用户的负面评价。对于用户而言,意识到人工智能拥有性别,亦具有重要意义,这意味着我们在更大程度上承认人工智能的社会参与度,赋予人工智能更多人性和人格。

本研究存在一定的局限。首先,研究采用的测量性别刻板印象的量表主要来自英文文献,因此量表中对性别的描述和数据的收集标准主要基于西方文化中的性别认知。但研究样本均来自中国大陆。文化间的差异可能会影响人们对性别的认知和评价标准,造成量表结果的偏差,一个直观的结果是量表信度尽管已达到可以接受的范围,但是并不太高。其次,研究选取的家庭理财、信息技术安全等主题的文本主要基于研究者的主观判断。这两类特定主题的文本不能代表读者对所有类别机器生成内容的评价;而且本研究中采用的单一素材不仅可能会造成外在效度的降低,使其结果不具有太大的可推广性,还有可能造成假设检验中的第一类错误的产生,即观察到了实际上并不存在的处理效应。今后的研究可以考虑加入更多类型的文本,并在同种类型中增加更多的实验材料,以消除单一材料设计引起的不稳定性。最后,本研究在中国内地进行,不得不将中国社会中关于性别的刻板印象和传统的雄性文化因素纳入考量。相信如果此研究在其他有着不同性别刻板印象的文化中进行(例如女性偏向文化下的北欧社会),结果会呈现很大的差异。这也是后续研究的方向之一。

总而言之,探讨人机关系中的性别议题,既是性别研究的横向延伸,也是人机研究的纵向扩展。当我们在谈论人工智能这个对象时,使用的代词从“它”变为“他、她”,人工智能可能拥有更多的机会,与人类共同进入人机协同的新纪元。

参考文献

- 葛岩,秦裕林,姚君喜,等.2013.母鸡的长相是否重要: 艺术家不良品行信息传播对作品评价的影响[J].文艺研究(09): 111-126.
- 韩秀.2021.社交机器人的“人格”测量: 一项基于艾森克问卷的探索性研究[J].青年记者(18): 45-46.
- 何苑,张洪忠.2018.原理、现状与局限: 机器写作在传媒业中的应用[J].新闻界(03): 21-25.
- 黄雅兰.2021.声音的性别: 新闻客户端自动语音播报中的刻板“音”象[J].全球传媒学刊(01): 108-125.
- 金兼斌.2014.机器新闻写作: 一场正在发生的革命[J].新闻与写作(09): 30-35.
- 兰玉.2007.论语言性别差异与社会文化的互构[J].重庆工商大学学报(社会科学版)(04): 118-123.
- 李思雪.2015.科幻电影中“人机关系”的性别表演[J].北京电影学院学报(06): 102-108.
- 刘茜.2020.人工智能机器写作受众态度实验研究[J].西南民族大学学报(人文社会科学版)(03): 157-162.
- 莫宏伟.2018.强人工智能与弱人工智能的伦理问题思考[J].科学与社会(01): 14-24.
- 牟怡,夏凯,Novozhilova E,等.2019.人工智能创作内容的信息加工与态度认知: 基于信息双重加工理论的实验研究[J].新闻大学(08): 30-43+121-122.
- 梅琼林,张晓.2006.“媒体等同”: 从效果研究到理论建构[J].社会科学研究(05): 193-196.
- 彭兰.2016.机器与算法的流行时代,人该怎么办[J].新闻与写作(12): 25-28.
- 钱铭怡,罗珊红,张光健,等.1999.关于性别刻板印象的初步调查[J].应用心理学(01): 14-19.
- 王鑫蓉,王玫.2022.基于 CASA 理论的语音助手个性研究[J].信息技术与信息化(02): 21-24.
- 徐大真.2003.性别刻板印象之性别效应研究[J].心理科学(04): 741-742.
- 许丽颖,喻丰.2020.机器人接受度的影响因素[J].科学通报(06): 496-510.
- 许向东,郭萌萌.2017.智媒时代的新闻生产: 自动化新闻的实践与思考[J].国际新闻界(05): 29-41.
- 喻国明.2015.“机器新闻写作”时代传媒发展的新变局[J].中国报业(23): 22-23.
- 喻国明,兰美娜,李玮.2017.智能化: 未来传播模式创新的核心逻辑: 兼论“人工智能+媒

- 体”的基本运作范式[J].新闻与写作(03):41-45.
- Ajzen I ,1991.The theory of planned behavior [J].Organizational behavior and human decision processes ,50(2) : 179-211.
- Briggs P ,Burford B ,De Angeli A ,et al. ,2002.Trust in online advice [J].Social science computer review ,20(3) : 321-332.
- Carlson M ,2015.The robotic reporter: Automated journalism and the redefinition of labor ,compositional forms and journalistic authority [J].Digital journalism ,3(3) : 416-431.
- Carpenter J ,Davis J M ,Erwin-Stewart N ,et al. ,2009.Gender representation and humanoid robots designed for domestic use [J].International journal of social robotics ,1: 261-265.
- Chung C J ,Kim H ,Kim J H ,2010.An anatomy of the credibility of online newspapers [J].Online information review ,34(5) : 669-685.
- Clerwall C ,2014.Enter the robot journalist: Users’ perceptions of automated content [J].Journalism practice ,8: 518-537.
- Coates J ,2015.Women ,men and language: A sociolinguistic account of gender differences in language [M].London: Routledge.
- Edwards C ,Edwards A ,Spence P R ,et al. ,2014.Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter [J].Computers in human behavior ,33: 372-376.
- Eyssel F ,Hegel F ,2012.(S) he ’ s got the look: Gender stereotyping of robots [J].Journal of applied social psychology ,42(9) : 2213-2230.
- Faul F ,Erdfelder E ,Buchner A ,et al. ,2009.Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses [J].Behavior research methods ,41(4) : 1149-1160.
- Fishman P M ,1978.Interaction: The work women do [J].Social problems ,25(4) : 397-406.
- Flaounas I ,Ali O ,Lansdall-Welfare T ,et al. ,2013.Research methods in the age of digital journalism: Massive-scale automated analysis of news-content—topics ,style and gender [J].Digital journalism ,1(1) : 102-116.
- Hamdan S M ,Hamdan J M ,2013.Authors’ perceptions of author’s gender: A myth or a truth? [J].International journal of English and literature ,4(10) : 523-528.
- Heilman M E ,2001.Description and prescription: How gender stereotypes prevent women’s ascent up the organizational ladder [J].Journal of social issues ,57(4) : 657-674.

- Hu W L ,Akash K ,Reid T ,et al. 2018.Computational modeling of the dynamics of human trust during human-machine interactions [J].IEEE transactions on human-machine systems ,49 (6) : 485-497.
- Koppel M ,Argamon S ,Shimoni A R ,2002. Automatically categorizing written texts by author gender [J].Literary and linguistic computing ,17(4) : 401-412.
- Kraus M ,Kraus J ,Baumann M ,et al. 2018.Effects of gender stereotypes on trust and likability in spoken human-robot interaction [C] // In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018) . Miyazaki , European Language Resources Association: 112-118.
- Kuchenbrandt D ,HÄring M ,Eichberg J ,et al. 2014.Keep an eye on the task! How gender typicality of tasks influence human-robot interactions [J].International journal of social robotics ,6(3) : 417-427.
- Lai C Y , 2009. Author Gender Analysis [EB/OL]. https://courses.ischool.berkeley.edu/i256/f09/Final%20Projects%20write-ups/LaiChaoyue_project_final.pdf.
- Lakoff R ,1973.Language and woman's place [J].Language in society ,2(1) : 45-79.
- McGinnies E ,Ward C D ,1980.Better liked than right: Trustworthiness and expertise as factors in credibility [J].Personality and social psychology bulletin ,6(3) : 467-472.
- Nass C ,Moon Y ,Fogg B J ,et al. ,1994 ,Can computer personalities be human personalities? [J].International journal of human-computer studies ,2(43) : 223-238.
- Nass C ,Steuer J ,Tauber E R ,1994. Computers are social actors [C] // Proceedings of the SIGCHI conference on Human factors in computing systems.Los Angeles ,ACM SIGCHI: 72-78.
- Nomura T ,Kanda T ,Suzuki T ,et al. 2008.Prediction of human behavior in human-robot interaction using psychological scales for anxiety and negative attitudes toward robots [J].IEEE transactions on robotics ,24(2) : 442-451.
- Nomura T ,Takagi S ,2011.Exploring effects of educational backgrounds and gender in human-robot interaction [C] // 2011 International conference on user science and engineering (i-user) .Selangor ,IEEE Conference: 24-29.
- Reeves B ,Nass C ,1996.The media equation: How people treat computers ,television ,and new media like real people [M].London: Cambridge University Press.
- Rubin J Z ,Provenzano F J ,Luria Z ,1974.The eye of the beholder: parents' views on sex of

- newborns [J]. *American journal of orthopsychiatry* ,44(4) : 512-519.
- Siegel M ,Breazeal C ,Norton M I ,2009. Persuasive robotics: The influence of robot gender on human behavior [C] // 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. St. Louis ,IEEE Conference: 2563-2568.
- Sundar S S ,1999. Exploring receivers' criteria for perception of print and online news [J]. *Journalism & mass communication quarterly* ,76(2) : 373-386.
- Swim J K ,1994. Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes [J]. *Journal of personality and social psychology* ,66(1) : 21.
- Tay B ,Jung Y ,Park T ,2014. When stereotypes meet robots: the double-edge sword of robot gender and personality in human-robot interaction [J]. *Computers in human behavior* ,38: 75-84.
- Turing A M ,1950. Computing machinery and intelligence [J]. *Mind* ,59(236) : 433-460.
- Van Der Kaa H A J ,Krahmer E J ,2014. Journalist versus news consumer: The perceived credibility of machine written news [C] // *Computation + Journalism Symposium 2014*. New York ,The Brown Institute for Media Innovation: 24-25.
- Waddell T F ,2018. A robot wrote this? How perceived machine authorship affects news credibility [J]. *Digital journalism* ,6(2) : 236-255.

Gender Stereotypes in Perceiving Machine-Written Content: Empirical Evidence Based on an Experimental Study

Mou Yi ,Lan Jianfeng

Abstract: Gender stereotypes can influence people's assessments on writers and content quality. Based on the CASA paradigm, this study attempts to examine whether the AI-writers with different gender labels are affected by the gender stereotypes. An online experiment with factorial design was conducted to collect people's evaluations on the content quality and expertise of the writers. The results indicated that gender stereotypes do exist in people's assessments on AI-writers and content quality. But for specific contexts, readers' appraisals of the quality of the content and expertise of the writer vary based on the writers' species and genders.

Keywords: machine writing; gender stereotype; robot gender; CASA paradigm; human-machine communication