

嵌入实践的技术： “自动化”对新闻业意味着什么

王辰瑶 秦 科

【摘要】技术发展使新闻生产中出现了无需人干涉的自动化现象,并成为近年来备受关注的研究议题。本文在对已有自动化新闻话语进行分析的基础上,提出可从“嵌入实践的技术”这一新视角来理解自动化新闻。新闻实践是由不同新闻工作组成的链条,不同新闻工作所要处理事实的可数据化程度和进行价值判断的复杂程度决定了自动化技术在新闻工作中的嵌入可能。通过对近20年来自动化新闻创新项目的分析,本文初步验证了这一实践主导的理论视角,指出了自动化新闻创新发展的三类主要方向,并提出了要警惕的相关风险。

【关键词】自动化新闻;新闻创新;新闻实践

【中图分类号】G21 **【文献标识码】**A

在大数据和人工智能技术的影响下,“自动化新闻”“算法新闻”“机器人新闻”“智能新闻”等成为近年来备受新闻传播学界关注的研究议题。不同术语指涉的对象是高度重合的,都强调了在新闻生产中出现的无需人干涉的“自动化”现象。作为一个技术概念,“自动化”萌芽于前工业时代,如春秋战国时期李冰父子主持设计的都江堰工程、西汉科学家张衡发明的“记里鼓车”都有令人惊叹的“自动化”功效。但自动化技术对人类生活产生普遍且巨大的影响,则开始于工业革命。自动控制理论也自20世纪初开始形成,经历了经典反馈控制理论、现代控制理论和智能控制理论等不同发展阶段。^①“所谓自动控制,是为实现生产或生活中的某种需求,在没有人直接参与的情况下,利用人为加入的设备,使被控制对象的某一个物理量或

多个物理量,自动地按照期望的规律去运行或变化,这种人为加入的设备就称为控制器或控制装置,由被控制对象和控制装置组成一个整体称为自动控制系统”。^②也就是说,自动化就是用自动控制装置代替人工控制。^③

新闻研究中讨论的“自动化新闻”等新现象,就本质而言,也是利用自动化技术代替了本来由人完成的某些新闻工作。20世纪50年代,自然语言生成(NLG)技术出现,这项技术能从数字结构化数据(digital structure data)中自动生成人类自然语言。2010年前后,NLG的应用扩展至新闻传播行业,在算法模型的帮助下,相关数据得以转化为新闻文本,整个过程除了初始编程外没有人工干预,因此这一过程被定义为“自动化新闻”^{④⑤⑥},也有部分学者称其为“算法新闻(algorithmic jour-

nalism)、“机器人新闻(robot journalism)”或“机器写作新闻(machine-written journalism)”等。^{⑦⑧⑨}现实实践中,新闻工作的链条很长,包括从新闻选择、事实采集到核查、分析、加工、叙述再到发布、审核、互动、评估等多个实践阶段,但自动化新闻的定义目前仍主要指新闻写作阶段的自动化。这可能窄化了对自动化技术在新闻实践中应用前景的理解,实际上每一个新闻工作环节都存在用自动控制装置代替人工的可能。

自2022年底ChatGPT模型发布以来,新闻学术界出现新一轮关于自动化、人工智能对新闻业的影响和在新闻领域运用的研究高潮。据统计,2023年1月至今已有《新闻记者》《新闻大学》《新闻界》《新闻与写作》等10余份国内新闻传播学术期刊开设过聚焦ChatGPT、生成性AI等新议题的学术专栏。本文在对已有自动化新闻话语进行分析的基础上,尝试提出从“嵌入实践的技术”这一新视角来理解自动化到底会在哪些方面以及在多大程度上影响新闻业。

一、自动化新闻话语

目前关于“自动化新闻”的研究主要基于三种问题意识:第一,新闻生产“自动化”的原理是什么;第二,“自动化新闻”是否会威胁人类新闻工作者的地位和价值;第三,受众能否识别出“自动化新闻”以及对“自动化新闻”和人类写作的新闻的质量评价如何。这三类研究可分别对应关于“自动化新闻”的三种主要话语:“解释话语”、“预测话语”和“评估话语”。

(一)解释话语。关于“自动化新闻”的解释性话语,主要涉及新闻生产中“自动化装置”的工作原理,也即新闻是如何被自动生产出来的原理。简单说,人工智能的发展取决于三个要素:数据量(包括数量和质量)、算法和算力。其中算力主要与芯片等硬件和硬件运行的能力相关,这里不

展开讨论。算法,字如其意,就是一套计算方法。只不过在智能时代,算法早已不是简单的计算,而是用复杂的机器学习方法反复计算迭代的模型。数据则是被计算的对象和原料。新闻生产之所以能“自动化”,就是对人类的语言建立数学模型并通过计算概率的深度学习算法把正确的信息呈现出来的这套机制在新闻领域的应用。从机器人写作到AIGC,“自动化新闻”的原理并没有改变。正如ChatGPT的智能原理也依然是来自20世纪以来推动人工智能发展的两个重要理论——信息论(概率)和控制论(反馈)。^⑩Graefe总结的“自动化新闻”流程包括:自动化装置收集有用的数据、用统计方法识别数据中重要和有趣的事件、按照特定规则对识别到的信息进行分类排序、按照特定模式生成叙述、将生成的新闻故事上传并发布。^⑪沈浩等人更具体地解释了这套机制可能产生的不同层次的“自动化新闻”:第一个层次是简单描述事件,只需通过软件自动抓取外部数据并填入预先编写好的模板中;第二个层次是算法不仅可以抓取还可以对外部数据结合数据库加以分析,如稿件中可增加与历史上曾发生过的同类事件比较的背景信息;第三个层次则可更进一步扩大数据库,实时对舆论热点话题进行情感分析,并因此改变生成文本的语气等。^⑫可见“自动化新闻”写作的“自动化”层次,在技术上是由数据量大小和质量高低以及算法的先进程度决定的。

(二)预测话语。关于“自动化新闻”的第二种主要的话语是预测引入了自动化技术后的新闻业和新闻人的未来。Serdoukd等研究者搜集了十六种专家信息,发现乐观和悲观的观点存在难以调和的分歧。Serdoukd等将这些观点分为三类:智能技术不会取代人类记者的工作、智能技术会在竞争中取代人类记者工作和智能技术与人类记者工作会融合而不是竞争。^⑬但我们认为这样的类型

表1 自动化新闻的“预测话语”

	对新闻业的未来	对人类记者的未来	观点总结
不可替代论	乐观	乐观	智能技术无法取代人类记者，也不会威胁其职业。人工智能可以为记者们“干粗活”，让后者更专注于复杂任务。
消极替代论	悲观	悲观	在自动化新闻和人工新闻的竞争中前者终会胜出，这不仅影响记者的职业，且会对新闻业的未来产生不利影响，如对世界的理解被结构化数据取代。
积极替代论	乐观	悲观	自动化新闻不仅会在竞争中胜出，且可以克服人类记者的弱点，这对公众是好事，对记者则不然。

划分仍无法清楚地说明它们之间的差异，因为对新闻业未来的看法和对人类记者能否保住自己工作的看法，并不是一回事。本文对“自动化新闻”不同的预测话语重新分为如下三类（见下表1）。

可见，不可替代论与替代论的分歧点在于人工智能的“能力”/“潜力”是否能胜任新闻业的核心工作，也即复杂新闻生产。而积极替代论与消极替代论的分歧点则在于，“自动化新闻”相比于人类记者生产的新闻对社会是有害还是有益。总体上，“自动化新闻”的预测话语其实与对“人工智能时代人类社会未来如何”这个更大问题的看法息息相关。而关于后者的争论早已从未来学家、科技精英、思想家们的圈子漫溢到普通人群中，成为一个无法定论的技术元议题。自工业革命以来，人类文明发展就充斥着关于新技术前景的争论。但人工智能技术引发的未来预测话语与此前最大的不同在于，人类已无法完全理解算法“黑箱”和机器自主学习的过程。换句话说，人类创造了自己也不能完全理解的机器，这导致在包括“自动化新闻”的人工智能将如何影响社会等预测性问题上，出现了更多的疑虑。

（三）评估话语。“自动化新闻”的第三种话语是对已经出现的“自动化新闻”成果（通常是自动化新闻写作）进行使用者评估。这类研究通过对使用者的实验和调研，测试他们能否识别出“自动化新闻”报道，并比较使用者对人类生产和自动化生

产的新闻的评价。这一类话语的分歧也很大：有一些研究认为机器撰写的内容可以克服人类记者的偏见，因而被认为更可信、专业性更高。^④受众对算法抱有积极认知，可能会提高他们对新闻可信度和质量的评价。^⑤但也有很多实验并不能证实人们对“自动化新闻”有更高的预期和评价，甚至提出了相反的结论：受众认为由机器生成的新闻可信度不如人类记者的新闻作品。^⑥

更进一步地分析发现，对文本来源的披露，可能是影响受众评价的重要因素。有学者对2017年—2020年间正式发表的12项经验研究进行了系统分析，这些实验共涉及4473名被试。结果显示，对受众而言，人类记者写的新闻与“自动化新闻”在可信度和文本质量方面没有差异或差异很小，但在可读性方面显出巨大的优势。不过这个结果与文本来源披露很有关系，也就是说，不论实际来源如何，只要参与者认为自己阅读的是一篇人类撰写的文章，他们就倾向于给予更高的评价。因此，研究者认为无法对“自动化新闻”与人工新闻“事实上”孰好孰坏做出判断，但他们提出了一种道德风险：新闻媒体可能会因此故意不披露某些报道是自动生成的。^⑦还有研究者发现，受众会以“拟人化”程度来评判“自动化新闻”，如果机器文本写作被认为“不像人”，就容易产生负面评价。^⑧除了“拟人化”外，人工智能新闻如果出现以“评价”方式提供的信息，也容易让受众反感。^⑨

二、新闻实践链与自动化技术的嵌入

前面谈到的三种主要的“自动化新闻”话语分别解释了新闻生产中的自动化原理、预测了引入自动化技术后新闻业和新闻人的未来,并从使用者角度比较和评估了“自动化新闻”与人工新闻的质量。其中,从技术角度而言,数据量大小和质量高低以及算法的先进程度,可以决定新闻“自动化”的层次;从劳动替代角度而言,预测人工智能未来能否胜任复杂新闻生产,是能否替代人类新闻工作的前提。在认可人工智能终将大量替代人类劳动的新闻业发展趋势前提下,对其走向又有两种不同观点:消极替代论认为这对人类记者和新闻业来说都是糟糕的前景;积极替代论则认为一个主要由人工智能主导的新闻业对社会更有益。只有不可替代论者,也即认为人工智能无法染指新闻业核心工作的论述者对新闻业和人类记者的未来都更乐观;从受众角度而言,一系列实验研究给出了不同的判断,表明在“自动化新闻”初起阶段,来自使用者的评价并不稳定。但不同的研究发现了一些影响使用者评价的共同因素,如使用者更接受“有人情味”“人性化”的文本,倾向于给被标识为人类记者来源的新闻更高的分数,尽管在很多指标上他们实际并不能区分人工新闻和“自动化新闻”的质量。

以上阐释可以帮助我们从不方面理解“自动化新闻”,但除了解释话语外,评价话语和预测话语的内部分歧重重,让“自动化新闻”图景不够明朗。对理解自动化与新闻业关系的更为明确的视角——在新闻实践中探讨自动化技术,仍相当缺乏探索。对此,本文提出可从“嵌入实践的技术”(technology embedded in practice)这一新视角理解自动化给新闻业带来的影响。

研究者对技术与新闻业关系的阐述主要是从技术如何影响新闻业的角度展开的。一如约翰·帕

夫利克(John Pavlik)在著名的《技术对新闻业的影响》一文提出的观点,新闻业发展历史在很多方面都是由技术变革来定义的。^③但与之相对的,将新闻业作为技术采纳和应用主体来探讨技术与新闻业关系的研究则比较少。虽然有关于技术采纳的主要模型如TAM^④、新技术的创新扩散理论^⑤等等,但新闻业与技术关系的研究中,新闻组织和新闻人更多被视为被动的技术跟随者和适应者,他们作为技术行动主体的实践却经常显得无足轻重。例如2023年牛津路透新闻研究所的《新闻业、媒体和技术趋势与预测》报告就再一次宣告了数字化技术(这一次是AI)带给新闻业的紧迫感:“在此背景下,尚未完全拥抱数字化的新闻机构将处于严重劣势。”^⑥在这样的声浪下,从新闻业的实践规范性出发,对新闻工作者如何调适技术以适应现有的新闻惯例与价值观^⑦,或新闻从业者如何发展和部署新技术并调节对新闻业的冲击^⑧等方面的研究则明显偏弱。

本文提出的“嵌入实践的技术”视角,不仅把新闻组织和从业者视为新技术的采纳和使用主体,更进一步把新闻实践过程细分为环环相扣但不同的新闻工作,以此组成独特的新闻实践链条。我们认为“自动化到底会如何影响新闻业的未来”这一备受关注且无公论的问题,需放置在新闻实践链条中,并从新闻实践的独特性入手,才能得到更好的解答。在这一新理论视角下,本文认为:新闻实践可视为由不同新闻工作组成的链条,不同新闻工作所要处理事实的可数据化程度和进行价值判断的复杂程度决定了自动化技术在新闻工作中的嵌入可能。

新闻业是以“事实”为工作对象的“价值”行业,它生产“新近发生的事实的报道”^⑨,但本质上“是一个信仰体系”^⑩。新闻业作为一种独特的社会实践不能混同于一般的信息产业,它不是对事实信息的简单统计和搬运,而是要提供基于事实

的真实可靠的公共知识。^⑧因此,新闻业不仅需要强化它获取、查核、再现和传递事实的一整套认知规范^⑨,而且总要为所在社会“人民的需要和利益服务”。^⑩

从新闻业的实践本质出发,可以更好地理解自动化技术在新闻工作中的嵌入可能。“自动化新闻”的“解释话语”已经说明,数据是实现自动化的前提。这就带来一个问题:新闻事实可被数据化的程度有多高?我们知道,事实是特定时空和情境中具体事物的实际状况,对事实所包含的信息进行结构化提取后才能得到数据。与很多同样处理事实的文化系统相比,新闻业给社会提供的是离事实源头更近的、更具体的真实叙述。虽然新闻类型中包括数据新闻、解释性新闻等类型,但总的来说,对事实的抽象解释并非新闻业所擅长。此外,在新闻的所有实践环节中,价值判断无处不在。从判断哪些事实公众希望或应该关注、到在复杂情景中洞察事实真相、再到选择适合的语言和表达形式(语言和形式中都蕴含着价值判断)进行新闻叙述等,都有很多需要新闻人以“人”的身体、情绪、认知去感受、理解、分析和判断的工作。尤其是在这样一个需要快速反应的职业中,很多价值判断是以非正式的、临场发挥的方式完成的,难以预先形成一套方案。这就是为什么很多记者强调“新闻感”却难以将这种感觉理论化。有些新闻工作需要处理的事实很容易被数据化,如可以量化表达的天气、股市变动等,有些则很难,如需追问采访对象才能挖掘到的事实;有些新闻工作对数据事实进行概率计算后就可以得到正确的判断,更多时候则需要进行复杂的价值判断。人工智能擅长对已有数据化事实总结、分析,但不擅长面对不确定、无规律、不可控的对象。可面对最新变动的事实、与大千世界中形形色色的对象打交道、在政治经济文化道德的复杂环境下

随时决策才是新闻业的工作常态。因此,从逻辑上不难得出结论:新闻工作中那些事实可数据化程度高、判断方式简单的实践容易被自动化,反之,事实可数据化程度不高、需要复杂价值判断的实践则不容易被自动化。新闻业自身的特点,决定了以数据化和计算为原理的“自动化装置”只能处理新闻业的部分工作。

本文将新闻工作细分为一个“实践链条”,并试图举出在不同实践环节(此“实践环节”包括选择、采集、核查、分析、叙述、审校、分发、互动等环节)上,容易自动化和难于自动化的工作例证。该实践环节的分类样式只是对新闻生产工作环节的“理想型”拟仿,实际上一种新闻工作可能是对几个理想型实践环节的融合,并可能经历多次反复而不是单线程的工作顺序,因此要比本文总结的“实践环节”复杂得多。在新闻实践链条图式上,每一个环节都存在容易和难于自动化的工作。具体来看,在新闻选择环节,抓取网络热点、获取特定信息来源发布的内容等都是容易自动化的工作,但要在很多“备选”方案中确定要做哪些选题、哪些是重大重点选题以及新闻选题的角度是什么等等,则是难于自动化的工作;在事实采集环节,通过软件接口连入外部数据库获得最新事实变动情况是容易自动化的工作,但实地调研、现场采访等则是难于自动化的工作;在事实核查环节,人工智能可能帮助记者识别信源,但难于判断什么是值得核查的对象;在对事实素材的分析环节,自动化可以快速对以往数据进行查找和分析,在数据数量和质量较高的情况下很容易生成可用的背景素材,但在对情况不明的复杂事实调查分析、逻辑判断上则难于取代人的工作;在新闻叙述阶段,自动化在符号转换和模块式写作方面表现卓越,如多种语言的自动语音播报、自动生成的AI简讯等,但在输出有价值判断的叙述内容时很容易“翻车”;在审校阶段,自动

化能根据预设标准进行表述规范性的审核校对,但任何一家专业新闻媒体都不可能让AI成为“签发人”;在分发环节,自动化擅长根据用户信息对内容拆散后精准投放,但如果新闻媒体要建立自己的品牌,恰恰需要使用者接受完整的媒体内容,并在此基础上形成独特风格;在互动环节,自动化可以完成礼节性和知识性的互动问答,但无法在人与人的“共情”基础上加深传受双方理解、改善传受关系。

三、对自动化新闻创新项目的初步验证

在新技术所能带来的诸多可能性中,本文对“自动化新闻”之未来的理解偏向于实践主导论。这里所说的新闻实践,既包括决定了新闻之所以是“新闻”的实践规范性,也包括日常新闻行动以及新闻行动者打破常规的新行动——新闻创新。按照上文分析,新闻实践规范性和自动化原理共同决定了部分新闻工作容易被自动化,部分新闻工作难于自动化。那么一个合乎该理论假设的推断就是:在已经出现的“自动化新闻”项目中,那些在容易嵌入自动化技术的实践环节(事实可数据化程度高、价值判断复杂度低)里开展的“自动化新闻”创新更可能持续下去并得到普及推广。

“自动化新闻”创新,是一个持续进行了近20年的创新领域。我们从国内外学术研究、行业报告和新闻报道中整理了2006年以来48个中外“自动化新闻创新项目”(见附录),通过分析其采用自动化技术完成何种新闻工作(对应新闻实践链条上的哪些环节)、在该项工作中自动化的嵌入程度等指标,验证“自动化新闻”创新是否存在特定发展方向,以及其发展方向能否被实践主导论解释。在考虑自动化嵌入程度这一指标时,我们将几乎不需要人工干预的新闻工作视为“高嵌入度”、将需要人与机器协作完成的新闻工作视为“中嵌入

度”、将自动化仅能给新闻人提供辅助信息或工具视为“低嵌入度”。需要说明的是,因为收集这些创新项目的文献来源广杂,经验材料丰俭不一,尤其是有一些早年的创新项目已不再被任何文献提及,因此根据这些有局限的材料,本文仅在此进行初步的、探索性的验证,并期待未来可通过经验研究得到进一步的理论发展。

通过对这些项目的初步分析,我们发现“自动化新闻”创新虽然可能出现在新闻实践链条的任何一个环节,但的确有特定的发展方向。主要有如下三种:

(1)基于外部可靠数据库的自动化新闻简讯。这类“自动化新闻”创新出现早、推动者广泛、且持续进行迭代发展。从2006年汤普森公司针对财经数据开发的自动化写作工具,直到2023年,都不断有媒体、数据公司和数据库所有方在该创新方向上投入。如2023年南方都市报将2017年的写稿机器人升级为2.0版本,抓取金融信息、裁判文书等权威渠道数据,自动进行数据分析和写作,日生成稿件300篇以上。通过自动化装置进行快速高效、格式明确的简讯式内容生产,这种类型的“自动化新闻”可以节省新闻人力成本、在某些突发事件上抢时效、为过去服务不到的特定小众群体提供信息等。比如华盛顿邮报主管曾表示,他们开发的Heliograf自动新闻写作系统可以在抓取体育比赛、大选等信息后依据用户所在地实现个性化推送,为“超本地化”报道创造了一种新模式。^⑧这类“自动化新闻”处理的事实可数据化程度高,简讯模版涉及的价值判断难度低,因此自动化技术在新闻工作中的嵌入度很高。但多限制在财经、体育、天气、地震等可以得到可靠外部数据库支撑的报道领域,且新闻媒体对这些数据没有垄断力,也无法阻止数据公司、专业数据库拥有者、技术开发者等自行提供这类自动简讯服务。

(2) 基于新闻媒体内部数据库开发的自动化新闻辅助工具。一些实力雄厚的传统新闻组织如美联社、华盛顿邮报、新华社等致力于将已有海量报道数据化,并在此基础上提供新闻背景调用、分析、数据可视化、视频生成等自动功能。这类“自动化新闻”实质上是新闻机构通过开发已经完成的报道、用户数据等,让媒体内容池从“仓库”变成一个聪明的“大脑”,帮助编辑记者更好地进行新的内容生产。如新华社“媒体大脑”项目在2019年全国两会报道中通过对比6年来政府工作报告,快速生产的《一杯茶的功夫读完6年政府工作报告,AI看出了啥奥妙》。这类“自动化新闻”创新的自动化嵌入程度中等,主要采取人机协作模式,新闻成品既离不开人的价值判断和叙述,也离不开人工智能的快速分析和内容生成。虽然相比直接使用外部可靠专业数据库,需要媒体自己开发数据和算法投入的成本要大得多,但这类“自动化新闻”创新项目的好处是可能会大幅提升新闻报道的质量如提供更多的语境化新闻、阐释性新闻、分析性新闻、可视化新闻和视频新闻,还可快速审校并对稿件自动提出编辑建议等。

(3) 基于开放大数据的自动化新闻提示和转换工具,如热点线索提供、趋势分析、自动语音播报翻译等。这类“自动化新闻”工具的数据来源是互联网上的开放大数据,通过自动抓取和分析提炼网络热点、预测议题传播周期、利用自动语言识别功能实现多语种播报等。这类自动化创新的数据库和算法主要由外部行动者提供,通常是专门的人工智能企业、数据分析公司等。虽然早期有一些新闻媒体也投入于此,如卫报[®],但在算法研发上无法与专门的企业抗衡,因此这类自动化工具已逐渐成为新闻实践中常见的外部生产资料。新闻人通过使用这类越来越普及的工具完成新闻生产,严格来说两者之间谈不上“协作”。比

如“搜索引擎”就是一种高度普及的自动化装置,几乎所有记者都会通过搜索引擎、热搜排行等获悉当下的网络热点,但是否要对热点进行进一步的报道、如何报道,仍需要新闻人来判断和执行。因此这一类“自动化新闻”工具应用广泛,但在新闻实践链条中代替人工劳动的自动化程度却是最低的。

事实的可数据化程度和价值判断的复杂度不仅区分了新闻工作中容易自动化和难于自动化的实践,也决定了“自动化新闻”可能开展的空间。近20年来不断出现的三类“自动化新闻”创新项目都要依靠可数据化的事实,并根据新闻工作涉及到的价值判断复杂程度,出现了自动化技术对新闻工作不同的嵌入度。第一类“自动化新闻”是可通过模版化写作完成的简讯,价值判断的复杂度最低,因此自动化程度最高,可以作为相对独立的产品运行。但也受限于特定的数据类型和产品形式,不可能替代新闻业的主体。此外,这类“自动化新闻”创新所依赖的各类专业数据并不一定能被媒体掌握,这些因素都可能局限这类“自动化新闻”创新的发展空间;第三类“自动化新闻”创新主要是对外部自动化信息处理工具的辅助性利用,新闻工作环节中的价值判断和主要的内容生产仍由编辑记者完成,因此这种自动化在新闻工作中的嵌入度最低,对新闻生产的挑战不大。虽然文献中提及的这类“自动化新闻”项目不多,但现实中很多自动化工具已经“隐入”新闻生产后台,甚至已经成为新的新闻常规;真正可能对新闻生产带来结构性改变的是第二类“自动化新闻”创新,也即人机协作式的中度“自动化新闻”。人对新闻产品选题、方向、角度的判断与机器快速生成的内容融合,缺一不可。这类“自动化新闻”需要媒体投入较多资源开发自身数据库和独特算法,但回报是可能重构新闻生产、从效率和质量两方面升级新闻业向社会提供的公共

知识产品。

四、警惕自动化新闻的“过度”风险

“自动化新闻”创新已显示出三个主要方向,但经验案例也显示,这一领域对创新者来说是高度不确定的。早期创新者,如路透社、BBC等都有明确流产的自动化创新项目,还有大量项目处于不更新或不再对外宣扬的沉寂中。可见“自动化新闻”创新仍是萌芽状态,还没有达到大规模扩散的阶段。无论是数据规模、质量还是新闻算法都还有极大的改进空间。“人机协作”式的“自动化新闻”创新目前主要是实力雄厚的极少数新闻媒体才有能力开展,其产品多是“试验性”的,如何在数量和质量上升级数字新闻报道还有很长的路要摸索。但同样基于从新闻实践特性出发的判断,本文还要提出:在“自动化新闻”创新的萌芽阶段就要警惕“自动化新闻”的“过度”风险。

第一,要警惕一些新闻创新项目越过了由自动化原理和新闻实践特性所规定的“度”,在无法或不应该自动化的领域开展“自动化新闻”项目。比如一类让数字人进行虚拟“采访”、“访谈”的“自动化新闻”项目,就违背了新闻实践特性。这类项目不是在处理新闻事实,而是生成布尔斯廷意义上的“伪事件”^③。它可能短期具有奇观性和娱乐性,但无论如何都不可能符合新闻规范性的前提下实现新闻采集自动化。新闻媒体投入在这类自动化项目上,除了一时的话题度外,对新闻实践的改进并无任何实际增益。

第二,要警惕新闻业对“自动化新闻”项目的投入过“度”,超过了“可以自动化的新闻工作”的占比。新闻业如果在实践过程中过度看重“自动化新闻”,就可能会减少在“难于自动化的新闻工作”中的投入,包括人才培养、生产经费、渠道建设等,甚至导致不是让机器为人服务,而是让人为机器

服务的错位。近期美国爆发的好莱坞编剧“罢工”的原因之一就是对这种错位关系的抗议。编剧们担心自己曾经的创意和心血“喂养”了AI,但他们自己却失去稳定的、可持续成长的职业,沦为“打零工者”。新闻业的自动化程度当然还远没有走到这一步,但来自其他文化业的教训也当引起对一些新闻领域迹象的警惕。比如在日常新闻实践中过度依赖自动化热点抓取工具、视频合成工具的“小编”还有没有机会像新闻前辈那样成长为“大记者”?笔者在媒体调研中已经多次听到新闻人表达这样的忧虑。

第三,也是最重要的是,要警惕对“难于自动化的新闻工作”的轻视思想。“难于自动化的新闻工作”主要是由不可数据化的事实、真正的人性化理解(不是自动化机器的“拟人化”)和基于此的复杂新闻叙述构成的。很多新闻工作都需要高密度的人际交流和对新近发生事实的全方位探索——它们事先无法数据化、更难有一套计算方法。包括现场新闻、调查新闻、人物专访、特稿、基层调研报告等很多传统但对社会意义重大的新闻类型,都主要靠“难于自动化的新闻工作”才能完成。关键是,新闻业和全社会是否还会珍惜并保护这些难于被自动化的新闻实践?从新闻业社会功能的履行来看,要更好地担当社会守望、督促引导和合作协调角色,“难于自动化的新闻工作”应该更受重视。但“自动化新闻”能降本增效的“利好”前景、用“拟人化”算法和隐匿来源的策略就可能“蒙混”使用者的惊人效果,以及在对人工智能混杂着焦虑和恐惧的心态下(如不拥抱就要被淘汰),新闻业面对自动化技术的最大风险并不是被新技术淘汰,而是在过度自动化中弱化新闻业得以生存的专业能力。

自动化新闻“预测话语”的三种“替代论”对新闻业和新闻人的未来有不同的乐观或悲观看法。本文则认为,新闻业和新闻人的未来并不是“命

定”的,不是在人的能力与机器的潜能之间进行比较,而是由新闻业面对自动化技术的认知和行动方式决定的。在理想的关系中,自动化技术将帮助新闻业适应数字时代的需求进行“内容升级”,如提高时效、改进质量、拓展报道领域等;在糟糕的关系下,自动化技术将“污染”新闻内容,如让大量鸡毛蒜皮、粗制滥造的所谓“新闻”充斥网络空间、被越来越多信息泡沫包围的严肃新闻将因为无法

触达公众而凋零。把“难于自动化的新闻工作”放在核心,为“难于自动化的新闻工作”培养人才、投注资源、打造公众信任的新闻品牌,新闻业才有底气和能力更好地使用自动化技术而不是被自动化反噬。

(王辰瑶:南京大学新闻传播学院教授、博士生导师、新闻创新实验室主任;秦科:南京大学新闻传播学院硕士研究生)

附录:“自动化新闻”创新项目(2006年—2023年)

创立时间	项目名称	创新主体	自动化新闻工作	自动化嵌入程度	对应新闻实践环节	效果/现状
2006	汤普森开发的程序	汤普森公司	对相关财经数据进行采集和加工并合成完整的新闻报道	高	采集、分析、叙述	早期文本生成创新项目,现状未知
2008	Open Calais	路透社	对非结构化 HTML 文档进行语义标记,便于新闻编辑校对	中	分析、审校	内部生产辅助工具,现状未知
2009	Graphiq 开发的交互可视化	语义技术公司、路透社、美联社、洛杉矶时报等	通过 AI 将图形与热门新闻、商业和体育报道配对;提供数据驱动的交互式可视化内容	中	叙述	2017 年被亚马逊收购,不再向新闻出版商提供服务
2009	Stats Monkey 统计猴	美国西北大学智能信息实验室	抓取体育比赛数据信息、撰写体育赛事报道	高	采集、叙述	被数据分析公司收购
2010	Narrative Science 的智能叙事平台	叙事科学公司 Narrative Science	从结构化数据源中生成文章、报告和摘要	高	采集、分析、叙述	2021 年被收购,原功能保留
2012	Spike	实时媒体监控平台 NewsWhip	按主题、语言、发布位置、平台和参与度等抓取和跟踪网络热点	低	选择、采集	持续运行,拥有多个媒体客户
2012	Truth-teller	华盛顿邮报	自动记录新闻报道中的政治言论并与“打假”数据库比对,有异常时发出警报	低	核查	最后一次视频作品发布于 2014 年 3 月
2012	Juicer	BBC	抓取外部数据并分主题进行标记,便于新闻趋势分析	中	采集、分析	由于未找到其他可持续用途,2018 年被关闭
2014	Quakebot	洛杉矶时报	抓取美国地质调查局的信息,如超过一定震级自动生成文章初稿	高	采集、分析、叙述	持续运行
2014	Wordsmith	Automated Insights 自动化洞察公司及美联社等媒体	采用自动化写作软件撰写公司财报和体育比赛	高	分析、叙述	持续运行,美联社等 200 多家客户每年使用 Wordsmith 生成超过 15 亿条内容

创立时间	项目名称	创新主体	自动化新闻工作	自动化嵌入程度	对应新闻实践环节	效果/现状
2014	#Open001 机器人	卫报	算法筛选社交网络热门文章, 自动化排版后再生成纸质报纸	低	选择、采集	为试验性项目, 相关信息停留至2014年, 现状未知
2015	快笔小新	新华社	实时捕获热点, 通过文本摘要技术自动生成短素材; 抓取并分析数据, 选择合适模板进行自动化叙述; 自动语音识别与文字转换	高	选择、采集、分析、叙述	2015年开启我国央媒机器人写稿先河, 2016年里约奥运会期间签发500多篇稿件。但近年来已逐渐变为写稿发稿辅助工具, 未有大的升级迭代
2015	Editor	纽约时报	对新闻稿件进行自动细粒度注释和标记, 便于新闻编辑校对	中	分析、审校	曾作为内部生产辅助工具。没有对现状的说明
2015	Blossom	纽约时报	预测新闻报道的热度, 便于新闻编辑筛选推广	中	分发	相关信息停留在2015年, 没有对现状的说明
2015	United Robots 的自动化写作	数据挖掘公司 United Robots、瑞典当地媒体集团 Mittmedia	数据公司与媒体合作开发特定风格的算法, 从结构化数据库中生成文本	中	分析、叙述	持续运行, 拥有多国客户
2015	梦幻写手 dreamwriter	腾讯	购买或自建商业、金融和科技领域的数据库, 自动撰写新闻稿件	高	分析、叙述	生产了中国第一篇机器人报道。最后一篇文章更新于2020年7月
2016	钱江晚报 AI 记者小冰	微软与钱江晚报	分析、预测数据, 生成文本; 同用户互动, 实现个性化推荐	高	叙述、互动	2017至2018年活跃后淡出大众视野
2016	BuzzBot	Buzzfeed 实验室、facebook	入驻 Facebook 的媒体可使用这款聊天机器人与用户互动	高	互动	以开源方式发布于 GitHub, 目前仍可以使用
2016	Feels	华盛顿邮报	向用户询问报道感受并向用户提供用户分析的互动机器人	高	互动	单次项目, 在2016年大选期间同用户互动
2016	Quartz	大西洋杂志	以互动方式为用户推荐新闻	高	分发、互动	2019年7月下线
2016	Velocity	网络新闻博客 Mashable	抓取并分析网络热点的传播趋势, 便于编辑筛选、推广; 优化文本内容	低	选择、分析、分发	内部生产辅助工具, 现状未知
2016	Heliograf	华盛顿邮报	抓取体育赛事、大选等相关数据自动生成报道; 依据用户所在地, 将大选结果插入播客, 实现个性化 AI 播报	高	采集、叙述、分发	2018年获“机器人使用卓越奖”, 2020年进行技术迭代, 持续运行
2016	DT 稿王	阿里巴巴集团与第一财经	实时监控财经信源、自动信息抽取、模版化自动新闻写作	高	采集、叙述	最后一篇文章发表于2017年6月9日
2016	Writingbots	百度	对百度知识库和全网资讯数据自动生成资讯文章	高	采集、叙述	无标注 Writingbots 的文本, 现状未知
2016	Xiaomingbot	今日头条	监测、抓取数据, 对体育赛事进行自动化写作, 提供多语言稿件并进行3D动画语音播报	高	采集、叙述	2019年8月最后一次更新; 2020年宣布发布全新升级版, 但未检索到作品
2017	Perspective API	美联社	识别辱骂性用户评论, 帮助编辑管理用户评论	中	互动	持续运行
2017	@TrumpOrNot	Twitter	使用机器学习和自然语言处理识别特朗普的推文哪些是他本人发的哪些是他手下代劳的	高	核查	持续运行

创立时间	项目名称	创新主体	自动化新闻工作	自动化嵌入程度	对应新闻实践环节	效果/现状
2017	媒体大脑	新华社	通过传感器，自动采集和匹配数据，自动化生成视频	中	采集、分析、叙述	持续运行
2017	小南	南方都市报	抓取并分析数据，根据模板自动化生成文字；抓取文本内容，利用排序学习、文本摘要技术等进行加工、改写	高	选择、采集、分析、叙述	2023年南方都市报写稿机器人升级2.0版本“湾财AI快报”
2017	Journalism++	瑞典数据新闻公司Newsworthy	实时监测公开的政府数据，根据异常值推荐新闻线索，进行数据分析	低	选择、分析	持续运行
2017	中国地震台网机器人	中国地震台网中心	对地震消息的自动撰写	高	采集、叙述	持续运行
2017	Modbot	华盛顿邮报	使用机器学习自动过滤需要人工审核的评论	中	互动	未知
2017	Talk 评论平台	华盛顿邮报、纽约时报、Mozilla	提供统计数据面板，便于评论管理员审核、互动	中	互动	内部生产辅助工具，现状未知
2018	CMS Bertie	福布斯	根据热门话题和历史文章为作者提供写作建议	低	选择、叙述	内部生产辅助工具，现状未知
2018	AI 合成主播	新华社	通过语音、唇语、表情、动作合成及深度学习，实现文本自动播报	高	叙述	为全球首个AI合成主播，2019年后陆续迭代，持续运行中
2018	新闻机器人RADAR	英国新闻联合社PA、新闻初创公司Urbs Media	使用AI挖掘、分析本地开源数据，通过自然语言技术生成本地报道文本	高	采集、分析、叙述	持续运行
2018	News Tracer 与 Lynx Insight	路透社	筛选 Twitter 推文并验证来源，为记者提供线索；识别热点趋势，为记者选题提供建议	低	采集、核查	内部生产辅助工具，现状未知
2019	AI 剪辑	中国广播电视台总台	对多路实时信号进行智能剪辑	中	叙述	持续运行
2019	Salco	BBC	抓取、分析外部数据，自动生成本地报道并自动化配图	高	采集、分析、叙述	持续运行
2019	智能化编辑部	新华社	人机协作为特征的新闻生产全流程、全环节智能化的编辑部	中	采集、叙述、分发	持续运行
2021	数字虚拟主播小C	中国广播电视台总台	对虚拟人进行访谈，打全国首档元宇宙虚拟访谈微综艺	低	叙述	全国首档元宇宙虚拟访谈微综艺，持续运行
2022	齐鲁壹点智能平台10.0版	山东齐鲁壹点APP	输入关键词后得到符合文意的AI配图，自动审校，个性化推荐	中	叙述、审校、分发	持续运行
2022	川观新闻数字人小观	川观新闻	与川观新闻客户端稿件系统等对接，提供播报新闻、用户服务等功能	低	叙述、互动	持续运行
2023	人民日报AI虚拟主播任小融	人民日报	模拟真人主播视听效果，个性化播报新闻文本，在虚拟演播厅访谈、连线	低	叙述、互动	两会期间#人民日报AI虚拟主播#登热搜榜榜首，持续运行
2023	“南都·湾财AI快报”	南方都市报	抓取银保监会、证监会、深交所、上交所、裁判文书等权威渠道数据，进行数据分析，自动化新闻写作	高	采集、分析、叙述	南都写稿机器人2.0版本，相比1.0版本处理能力提高了10倍，单日能生成300篇稿件以上
2023	AI撰写与用户识别的互动项目	意大利媒体IL Foglio	发起为期30天的互动项目，媒体发布由ChatGPT编写的文本，成功识别的用户获得媒体奖励	高	叙述、互动	单次项目
2023	AI撰写	英国每日镜报、每日快报出版商Reach	2023年3月，开始尝试用ChatGPT协助记者撰写本地天气与交通等主题的短篇报道	高	采集、分析、叙述	刚开始运行
2023	NewsGPT	Alan Levy 非营利团队	实时扫描世界各地信源，并创建新闻报道	高	采集、分析、叙述、分发	2023年3月27日推出，世界第一个完全由人工智能生成的新闻频道，持续运行中

注释:

① 王国军、陈松乔:《自动控制理论发展综述》,《微型机与应用》,2000年第6期,第4-7页。

② 王燕平主编:《自动控制原理》,北京:人民邮电出版社,2015年,第13页。

③ 朱永甫、宋丽琴等:《自动控制原理》,北京:电子工业出版社,2018年,第16页。

④ Napoli P M. Automated media: An institutional theory perspective on algorithmic media production and consumption. *Communication theory*, 2014, 24(3): 340-360.

⑤ Carlson M. The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital journalism*, 2015, 3(3): 416-431.

⑥ ⑪ Graefe, A. Guide to automated journalism, *Columbia Journalism Review*, 2016: 1-13. https://www.researchgate.net/publication/289529155_Guide_to_Automated_Journalism.

⑦ Anderson C W. Towards a sociology of computational and algorithmic journalism. *New media & society*, 2013, 15(7): 1005-1021.

⑧ Anderson C W. What aggregators do: Towards a networked concept of journalistic expertise in the digital age. *Journalism*, 2013, 14(8): 1008-1023.

⑨ Dörr K N. Mapping the field of algorithmic journalism. *Digital journalism*, 2016, 4(6): 700-722.

⑩ 邓建国:《概率与反馈:ChatGPT的智能原理与人机内容共创》,《南京社会科学》,2023年第3期,第86-94页。

⑫ 沈浩、元方:《“大数据时代”的自动化新闻写作的历史、实践与未来》,《新闻爱好者》,2017年第4期,第19-25页。

⑬ Serdouk A, Bessam A C. Bots in Newsrooms: What Future for Human Journalists?. *Media Watch*, 2023, 14(1): 100-115.

⑭ Sundar, S. S. The MAIN model: A heuristic approach to understanding technology effects on credibility. Cambridge, MA: MacArthur Foundation Digital Media and Learning Initiative, 2008: 73-100.

⑮ Liu B, Wei L. Machine authorship in situ: Effect of news organization and news genre on news credibility. *Digital journalism*, 2019, 7(5): 635-657.

⑯ ⑰ Waddell T F. A robot wrote this? How perceived machine authorship affects news credibility. *Digital journalism*, 2018, 6(2): 236-255.

⑱ Graefe A, Bohlken N. Automated journalism: A meta-analysis of readers' perceptions of human-written in comparison to automated news[J]. *Media and Communication*, 2020, 8(3): 50-59.

⑲ Henestrosa A L, Greving H, Kimmerle J. Automated journalism: The effects of AI authorship and evaluative information on the perception of a science journalism article. *Computers in Human Behavior*, 2023, 138: 107445.

⑳ Pavlik J. The impact of technology on journalism. *Journalism studies*, 2000, 1(2): 229-237.

㉑ Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. *Universal access in the information society*, 2015, 14: 81-95.

㉒ Rogers E M. *Diffusion of innovations*. New York: Simon and Schuster, 2010.

㉓ Newman N. Journalism, media and technology trends and predictions 2023. Reuters Institute for the Study of Journalism, 2023. <https://reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2023>.

㉔ Singer J B. The political j-blogger: 'Normalizing' a new media form to fit old norms and practices. *Journalism*, 2005, 6(2): 173-198.

㉕ Fenton, N. *News in the digital age*. The Routledge companion to news and journalism. Routledge, 2009:557-567.

㉖ 陆定一:《我们对于新闻学的基本观点》,《解放日报》,1943年9月1日。

㉗ John C. Nerone. The historical roots of the normative model of journalism[J]. *Journalism*. 2012. 14(4): 446-458.

㉘ 王辰瑶:《未来新闻的知识形态》,《南京社会科学》,2013年第10期,第105-110页。

㉙ 潘忠党:《在“后真相”喧嚣下新闻业的坚持——一个以“副文本”为修辞的视角》,《新闻记者》,2018年第5期,第4-16页。

㉚ J. Herbert Altschull. *Agents of Power: The Media and Public Policy*(2nd ed). New York: Longman, 1995: 440.

㉛ Dick Weisinger. Artificial Intelligence: Bots that Can Write Take on Journalism. <https://formtek.com/blog/artificial-intelligence-bots-that-can-write-take-on-journalism/> [2019-6-7].

㉜ Justin Ellis. The Guardian experiments with a robot-generated newspaper with The Long Good Read. NiemanLab. <https://www.niemanlab.org/2013/12/the-guardian-experiments-with-a-robot-generated-newspaper-with-the-long-good-read/> [2013-12-3]

㉝ [美]丹尼尔·布尔斯廷:《幻象》,符夏怡译,海南:南海出版公司,2023年,第9-10页。

(责任编辑:李嘉卓)