

论机器人新闻的精确性和可读性

王 勇 王 磊

摘 要：作为人类历史上第一次由机器代替人类全自动生产的新闻，机器人新闻自产生以来就受到社会的广泛关注。目前人们对机器人新闻几乎一边倒地认为它具有精确性优势和可读性差的致命缺陷。然而，从机器人新闻的生产原理和实践过程来看，它既具有精确性的一面，也存在出错的可能性，不能仅仅基于对电脑程序的表面认识和对科技的迷信而忽略了机器人新闻存在的犯错风险；机器人新闻虽然是由电脑程序套用既有新闻模板生产的，但它能够做到多样化、个性化，有立场、有感情，有“温度”、有“人性”，以可读性差来否定机器人新闻难以成立，更不能以主观臆断机器人新闻具有可读性差的致命缺陷而认为人类记者就可以高枕无忧。

关键词：机器人新闻；精确性；可读性；模式化；个性化

作者简介：王勇，男，教授，博士。（昆明理工大学 艺术与传媒学院，云南 昆明，650500）

王磊，男，硕士生。（昆明理工大学 艺术与传媒学院，云南 昆明，650500）

中图分类号：G212

文献标识码：A

文章编号：1008-6552（2018）03-0107-06

机器人新闻（robot journalism），又称机器撰写的新闻（machine-written news）、自动化生产的新闻（automated journalism）、计算机生成的内容（computer-generated content）、算法新闻（algorithmic journalism）等等，是指利用计算机、互联网技术搜集、分析数据信息，发现有价值的新闻事实，并利用自然语言生成技术自动生成的新闻。机器人新闻的产生使人类第一次实现了由机器代替人类进行全自动新闻生产，它是近年来大数据技术、人工智能技术飞速发展的产物。

“机器人新闻的出现改变了传统的新闻生产方式”^[1]，在业界引起了巨大震动，在学界和社会上也引起了广泛反响。“近两年来，国内外新闻业界和学界都在热议‘机器人新闻’。”^[2]梳理国内有关机器人新闻的观点发现，人们普遍认为，相比于人类记者采写的新闻，机器人新闻具有精确性优势，同时又具有可读性差的致命缺陷。

那么，相比人类记者采写的新闻，机器人新闻是否具有精确性优势和可读性差的劣势？本文对此进行分析探讨。

一、机器人新闻的精确性

精确，是人们对新闻稿件报道新闻事实的基本要求和评判新闻报道质量的基本标准，是指新闻稿件在叙述新闻事实时，构成事实成分的时间、地点、人物名称、数字、引语、事件原因、结果等都做到准确无误，没有任何差错。新闻报道是人们获取外界信息，认知外界事物和做出判断、决策的重要参考依据，因此精确和真实一样都是新闻的生命，事关新闻报道、新闻媒体存在的价值和意义。对机

基金项目：本文系2017年国家社会科学基金项目“新媒体时代西南沿边民族地区的政治传播与国家安全研究”（17BXW110），2017年云南省网信办项目“网络传播态势及新生代受众研究”（YNWX2017006），2017年昆明理工大学人文社科研究重点培育项目“新媒体时代边疆民族地区的政治传播与意识形态安全研究”（KKZ2017010），2017年昆明理工大学宣传思想工作研究项目“基于宣传效果的新媒体时代高校媒体优化研究”（KKS2017012），昆明理工大学“当代艺术理论创新研究”创新团队项目的阶段性成果。

机器人新闻这种全自动化生产的新闻,目前许多人认为它相比人类记者采写的新闻具有精确性优势,并因此而对其大加肯定。

(一) 认为机器人新闻具有精确性优势的逻辑

机器人新闻作为一种主要利用电脑程序自动搜集、分析数据信息,并套用既有新闻模板撰写而成的新闻,自产生以来,人们几乎一边倒地肯定它具有精确性优势,认为相比人类记者采写的新闻,它能做到更少出错,更精确。如“机器人新闻相比人类记者具有的优点是错误较少,精确”,“比人类记者出错更少”^[3];机器人新闻写作在“精确性”方面“具有独特优势”^[4];机器人新闻“可以做到精确、迅捷的生成和发布”,“差错率远低于人工写作”^[5]。

人们认为机器人新闻具有精确性优势,主要基于以下逻辑:

1. 避免了人的主观性和人为干预

一些人认为,机器人新闻生产,它是由电脑按照既定程序自动搜集、分析数据信息,套用既有模板自动撰写新闻,不像人类记者采写新闻那样主观;同时与人类记者采集数据信息、撰写新闻不同,它自动生成新闻,自动发布,避免了人为的干预,更客观。如“相对于人类记者常用的‘质化’信息采集方式,机器人新闻则完全自动采集、加工和发布”,因此具有“准确”的特点。^[2]

2. 电脑对数据的收集处理比人精确

机器人新闻是一种基于数据的发生和变动,以数据为基础和核心内容的新闻。许多人认为,在对数据的收集、计算、分析等方面,特别是对大量的、复杂的数据的收集、计算、分析方面,电脑比人脑更不易犯错,更准确,因此,机器人新闻相比人类记者采写的新闻会更少出错,精确性更高。如新闻机器人“长于处理在人类看来琐碎、庞杂的数据信息,通过既定程序对既定算法的实现,可以保证生成文本在新闻要素、新闻事实、引用数据等方面的完整和准确”;“严格遵守程式设定的数据采集和分析规则,套用标准的新闻格式,使出错概率降至极低”^[6];“面对庞大的数据,它的精确性也无与伦比”^[7]。

正是基于以上两方面的逻辑,很多人认为机器人新闻相比人类记者采写的新闻具有精确性优势。

(二) 机器人新闻存在出错风险的原因

应该说,人们认为机器人新闻具有精确性优势在逻辑上是成立的,但对机器人新闻的生产原理和过程的认识并不全面,以致对机器人新闻的精确性认识存在一定的片面性。

1. 机器人新闻生产同样渗入了人的主观意志和偏见

虽然从表面和机器人新闻的具体生产过程来看,机器人新闻是由电脑程序自动完成的,一般没有人为干预(少数媒体设有最后人工审核把关环节),也不像人类记者采写新闻那样凭人的主观意志,但并不是说机器人新闻生产就没有“人”的参与,没有掺进“人”的主观意志,因为机器人新闻的数据搜集、过滤、分析程序,套用的新闻价值判断模板、新闻报道模板都是事先由人设计、编制的,新闻机器人执行的是设计人的设想和设计,体现的是设计人的意志,甚至不可避免地带有设计者的偏见。德国传媒思想家弗里德里希·基特勒就曾指出,软件产品中已经事先被嵌入了权力结构,设定了“有限性、许可、特权和障碍”^[8]。所以机器人新闻生产,并不是没有人的参与,没有人的意志,因此它就不可能是完全客观、完全中立的;还有,新闻机器人既然是人的设计,也就不可能做到十全十美,就有可能存在不全面、不完善的地方。这些都可能影响机器人新闻报道的精确性。正因为如此,美国哥伦比亚大学 Tow 数字新闻中心曾发布“算法新闻责任报告”,提出新闻机构要“告知公众机器人在实际运行中带有的偏见、所犯的错误”。《纽约时报》也设立专门网页,对其机器人新闻的算法原理进行了通俗易懂的详细解释,并“对该机器人算法存在的偏见直言不讳”^[2]。

2. 新闻机器人对数据质量的判断、真伪的甄别存在缺陷

虽然在数据信息特别是大量的、复杂的数据信息的录入、计算、分析方面，电脑比人脑更不易犯错，精确性更高，但是在对数据信息的质量的判断、真伪的甄别等方面，新闻机器人可能比不上经验丰富的人类记者，特别是在机器人新闻的生产实践中，对数据信息质量的判断、真伪的甄别等方面存在缺陷。

机器人新闻，是人类根据某类新闻的通常发生规律，设计编制一套程序去搜集、分析数据信息，并根据数据信息的发生或变化判断它的新闻价值，然后套用既有新闻报道模板生成新闻，它对数据信息的搜集、分析和新闻价值判断是设计者根据一定的模式设定的。例如，谷歌开发的“维基百科实时监测机器人”，它的程序设计逻辑是，在维基百科全球287个不同语言的版本中，如果多人同时对同一个条目的内容进行修改更新，就判断为一条重大新闻正在发生，它就自动进行数据信息抓取和报道。这种设计的好处是，报道全自动化，24小时全天待命，报道快速，“一触即发”，只要有数据信息发生或变动符合程序的设计逻辑，它就自动启动，抓取数据信息报道新闻；不足之处是，它对数据信息的分析和判断是模式化、机械化的，分析和判断的依据就是设计逻辑，对符合设计逻辑的数据信息的真伪、质量不做分辨、甄别，一律采纳。比如“维基百科实时监测机器人”只要发现“多人同时对同一个条目的内容进行修改更新”，它就自动抓取数据信息进行报道，而不会对“多人同时对同一个条目的内容进行修改更新”的原因和真伪进行分析和辨别，可是实际上“多人同时对同一个条目的内容进行修改更新”就存在虚假新闻、谣言甚至人为操纵等多种可能，这样就可能导致维基百科实时监测机器人采集的数据信息及生成的报道不准确。“如果其信源比较权威，机器人新闻出错的可能性较小；但如果其信源来自社交媒体，机器人记者就可能被操纵”；“机器人（算法）既可以快速地采集和发布新闻，也可以快速地犯错。而且，因其高效率和人无法干预，它一旦犯错造成负面影响波及面更广”^[2]。

总之，从对数据的精确计算能力和在具体报道过程中免于人为干预等方面来看，机器人新闻具有精确性优势。相比人类记者，新闻机器人对数据信息的计算处理不易出错，对数据信息的分析报道也更客观；但是从对所搜集的数据信息的质量的判断、真伪的甄别来看，机器人新闻也存在出错的风险。相比社会经验丰富的人类记者，新闻机器人仅根据设计逻辑机械地判断和搜集数据信息，对搜集的数据信息的质量和真伪缺乏深入分析和仔细甄别，因此有可能因源头数据有问题而导致报道犯错。此外，机器人新闻还有可能因程序设计存在片面或偏见而导致新闻报道出现误差。因此，既要认识到机器人新闻具有精确性的一面，也要认识到它存在出错的可能性，而不能笼统地说机器人新闻具有精确性优势。

二、机器人新闻的可读性

可读性是新闻报道适合人们阅读、吸引人们阅读的程度，是影响新闻传播效果的重要方面，在当今信息爆炸、人们注意力有限、“眼球争夺”空前激烈的时代，可读性甚至在一定程度上决定着新闻报道的成败，因此，可读性是评判新闻报道质量的重要指标。对机器人生产新闻这种人类历史上第一次由机器替代人类进行新闻内容生产的新闻生产方式，一些业界、学界专家并不看好，认为相比人类记者采写的新闻，新闻机器人生产的新闻存在可读性差的缺陷，并将其作为否定机器人新闻的重要依据。

（一）认为机器人新闻可读性差的逻辑

自工业革命以来，越来越多的产业工人被机器取代，而新闻记者作为“知识工人”一度高枕无忧。但机器人新闻的产生第一次让从事神圣的内容生产的编辑记者感受到了前所未有的挑战。“机器人新闻的发展激起了新闻内容生产者的恐慌，担心机器的内容生产可能将终结新闻编辑室的工作”^[3]；“一些人在担心这些不知疲倦、不需报酬且努力工作的‘家伙’是否会抢走自己的饭碗”^[9]；“有些人因此预言，机器新闻将会取代新闻记者，成为未来新闻构建的主力军，传统的新闻工作者将面临失业的尴尬

境地”^[10]；“一些媒体已经径直宣布计算机很快将取代记者”^[2]。

然而，一些人认为，相比人类记者采写的新闻，机器人新闻存在一个致命缺陷：可读性差。一些媒体记者甚至在一定程度上将机器人新闻的可读性缺陷看作“救命稻草”。“面对来势汹汹的机器人新闻，记者们往往指出，机器生产内容的质量低下或一般”^[3]；“记者们往往质疑机器人新闻的内容质量，认为它们只能满足读者对清晰和准确信息的最低要求，通常认为机器人新闻缺少精妙的描述，机械化且无聊”^[11]；“‘机器人记者’的报道必然是单调、乏味的，这样的报道已经无法满足受众的品位需求，得不到读者的喜爱。这是‘机器人记者’在文风方面的劣势”^[7]。可读性差也成为当前许多人否定机器人新闻的重要依据。

一些人断定机器人新闻可读性差主要基于以下逻辑：

一是新闻报道的模式化。机器人新闻依靠自然语言生成技术，套用既有的新闻模板自动生成新闻，因此新闻报道具有模式化的特点。“机器人记者完全依靠人工录入的知识库和模板”，“新闻稿件呈现出模式化、固定化、单一化、同质化的特点”^[12]。

二是缺乏立场和情感。新闻机器人不像人类记者那样有血有肉，有立场有感情，也不可能像人类记者那样带着立场和感情去选择新闻报道的主题，采集新闻报道素材，撰写新闻稿件，因此机器人新闻缺乏人类记者采写的新闻所具有的立场和所包含的感情。“从机器人手中传递出来的新闻虽然是新鲜的却是冰凉的”，“因为机器人本身没有办法判断情绪”^[9]；“新闻是有‘温度’的，机器人做新闻却没有‘温度’，很可能在昙花一现之后，优雅地死掉”^[13]。

正是基于以上逻辑，一些人认为机器人新闻存在可读性差的致命缺陷，无法对传统的新闻生产方式构成挑战和威胁。

（二）可读性差不是机器人新闻的致命缺陷的原因

从上可知，一些人认为，机器人新闻具有报道模式化、缺乏立场和情感，因此具有可读性差这一致命缺陷。但是从机器人新闻的生产原理及实践来看，以上两点并不成立，机器人新闻并不像有些人所断定的那样可读性差，更不是其致命缺陷。

1. 机器人新闻可以多样化、个性化

虽然机器人新闻的生产，不管是新闻报道的主题、角度、结构，还是写作语言都是套用既有的模板，但说“机器人新闻目前主要仍是基于少数几种有限的模板，风格和语调较为生硬”^[2]就不太准确，实际上机器人新闻生产所能套用的新闻报道模板数量非常多，比如被美联社、《纽约时报》、雅虎等美国著名媒体引入使用的新闻机器人 Wordsmith，就有“超过3亿可使用的新闻写作模板”，“且随着机器人记者应用领域的拓展，这一规模还在不断扩大”^[6]。而且从理论上来说，新闻机器人可以将其认为合适的人类已有的所有新闻写作模板都储存起来供其选择使用。而当新闻机器人在新闻生产中所能套用的写作模板足够丰富时，其在新闻生产中实际所选择的写作模板也足够丰富多样时，比如2014年 Wordsmith 就创造出100多种报道类型，新闻机器人所生产的新闻产品就不会给受众以“模式化、固定化、单一化、同质化”的感觉。

而且在机器人新闻生产实践中，有的新闻机器人能够灵活使用基本故事模板，它结合元数据，组合使用，花样繁多，例如根据主语的单复数而配套使用相应的动词形式，或者变换使用同义词，从而使文本读来不那么单调、枯燥；还有的新闻机器人还会在报道文本的生成过程中加入随机因素，并且可以选择多种复杂的叙事语气（如“冷漠的”、“自信的”、“悲观的”和“充满激情的”等等），以让文本显得更加多样。

所以新闻机器人虽然是套用既有的新闻报道模板来生产新闻，但它在实际生产中不断变换模板，灵活使用不同的句子和用词，从而能有效避免新闻产品的模式化、单一化。

特别是随着自然语言生成技术的不断完善和成熟，机器人新闻报道不但可以避免模式化，而且可以实现个性化。它突出表现在以下两方面：

一是针对特定受众提供有针对性的新闻服务，实现新闻服务的“个性化”。如叙述科学公司能依据受众群体特征选择报道角度和内容，比如针对某球队的支持者进行赛事报道时，尽量避免使用负面消息和用词；腾讯公司的新闻机器人 Dreamwriter 在财经报道方面，针对不同目标用户的个性化需求，同步推出“研判版”和“民生版”两个个性化版本；有的新闻机器人甚至根据不同受众群的语言表达习惯提供不同表达方式的新闻作品，比如“90后”群体、女性群体、低文化群体等在语言表达习惯上存在差异，新闻机器人能针对他们的语言表达习惯分别提供不同表达方式的新闻作品，实现同一新闻题材针对不同受众群体提供不同表达方式的个性化新闻服务。

二是模仿特定媒体或人物的写作风格，实现报道的“风格化”。新闻机器人通过学习，可以模仿一些媒体的报道风格，比如新闻机器人 Wordsmith 能模仿美联社的“平实”新闻风格，经过调整后也可以模仿 NPR 的“华丽的”或“更人性化的”新闻风格；有的新闻机器人还可以模拟知名记者或作家的写作风格，实行文风的定制。新闻机器人主要学习和套用以往新闻作品和记者所使用的词汇来组成句子，生成新闻作品，对知名记者或作家的句式和用词习惯的学习和模仿能使其作品呈现出某位知名记者或作家的写作风格，如叙述科学公司的垒球报道，就主要学习和模仿了 20 世纪早期知名体育评论家林·拉德纳的写作用词习惯，呈现出林·拉德纳的报道风格。

总之，机器人新闻可以实现报道的多样化、个性化，避免单调、同质、呆板、枯燥。

2. 机器人新闻也可以有立场和情感

随着智能机器人技术的发展和进步，新闻机器人越来越“聪明”，语言学习能力越来越强，语言表达能力也越来越强。有实验表明，“如果对机器人写作所依托的模板更加精致化，乃至通过输入‘悲伤’‘快乐’‘崇高’等词语，机器人便可以遣词造句，极为迅速地写出符合上述需求的，注入了更为细腻情感的文字。”^[4]一些新闻机器人开发者也通过总结人类记者在新闻报道中表达立场和感情的方式，并让新闻机器人学习，从而使新闻机器人在新闻生产实践中能用合适的句子和语言表达立场和细腻的感情。美联社表示，“机器人写作软件按照程序使用不同语气的能力越来越强”，“如将读者喜好的语气、角度甚至幽默融入一篇报道之中，以机器人报道篮球赛事为例，可能机器人会用同样的数据生成两个版本的新闻：一个版本为获胜球队的粉丝所写，一种欢天喜地的气氛；另一个版本则是为失败的球队准备的，读起来让粉丝沮丧低落”^[9]。2017 年，南方日报的新闻机器人“小南”“写的春运火车票的新闻中采用了不少卖萌的句子”，“小南”开发者北京大学计算机科学技术研究所研究员万小军坦承：“这也是把我们的新闻记者的语言表达做了很好的总结，最后使我们的机器人也能这样表达出来。”^[14]他说他们对新闻机器人的进一步改进目标就是让新闻机器人“在写稿中加入态度和立场，让稿件更加生动和具有人性”^[15]。因此，虽然新闻机器人是没有血肉、没有思想和感情的程序和机器，但它们可以通过模仿和学习人类，从而像人类一样生产出有立场和感情、有“温度”和“人性”的新闻作品。

总之，机器人新闻并不是像一些人想当然认为的那样存在报道模式化、固定化、单一化、同质化，没有感情和“温度”，可读性差等致命缺陷，机器人新闻不但可以做到快速生产、以数据说话，而且可以做到报道多样化、个性化，有立场、有感情。实际上，在国外，经新闻机器人生成的“财经和体育类新闻报道，大多数读者都无法将其与人类记者生产的新闻区别开来”^[2]，在我国，“从今日头条的用户反馈来看，很多人没有区别机器人的稿子和记者的稿子”^[15]。这也从一个侧面反映，在可读性方面，目前机器人生产的新闻与人类记者采写的新闻已经没有非常明显的差别，而且随着新闻机器人的不断进步，机器人新闻的可读性会不断得到提高。

三、结 语

机器人新闻是当今大数据技术和人工智能技术发展进步的产物,是人类不断引进和采用最新科学技术改进新闻生产、提高生产效率、降低生产成本的结果。作为人类历史上第一次由机器代替人类全自动生产的新闻,机器人新闻自产生以来就在新闻业界引起极大的震动,也受到了学界和全社会的广泛关注。对电脑程序自动生产新闻这种对新闻行业产生巨大冲击的革命性新闻生产方式,人们最关注的是它相比传统新闻生产方式具有哪些优势和不足。然而就像对所有新技术、新事物的认识都有一个由模糊到清晰、由肤浅到深入、由片面到全面、由偏差到准确的过程一样,目前对机器人新闻也存在一些不够深入、不够全面的认识。

一些人基于电脑程序的精确计算能力和在具体报道过程中免于人为干预表象,认定机器人新闻具有精确性优势,但是,由于对搜集的数据的质量和真伪缺乏深入分析和仔细甄别,以及程序设计可能存在片面或偏见,机器人新闻也存在出错的风险。我们既要认识到机器人新闻具有精确性的一面,也要认识到它存在出错的可能性,不能仅仅基于对电脑程序的表面认识和对科技的迷信而忽略了机器人新闻存在的犯错风险。从某种意义上看,正由于机器人新闻出错的风险具有隐蔽性,容易被人忽略,加之机器人新闻生产速度快、效率高、生产量大、人为干预少,一旦犯错影响面大,因此尤其需要警惕,切不可盲目迷信其精确性。

另外一些人则基于机器人没有“人性”,报道新闻主要套用既有的新闻模板而断定机器人新闻具有模式化,缺乏立场和感情,因而可读性差的致命缺陷。但实际上,新闻机器人在新闻生产中能够灵活使用丰富多样的模板、句子和用词,从而使报道呈现出多样化;而且随着新闻机器人的语言学习能力、表达能力越来越强,新闻机器人还可以实现报道的“个性化”、“风格化”,甚至可以像人类一样生产出有立场、有感情,有“温度”、有“人性”的新闻作品。因此,以可读性差来否定机器人新闻难以成立,更不能以主观臆断机器人新闻具有可读性差的致命缺陷而认为人类记者就可以高枕无忧。

参考文献:

- [1] 陈小晰. 机器人新闻与记者稿件的对比 [J]. 新闻记者, 2016 (9).
- [2] 邓建国. 机器人新闻: 原理、风险和影响 [J]. 新闻记者, 2016 (9).
- [3] 叶韦明. 机器人新闻: 变革历程与社会影响 [J]. 中国出版, 2016 (10).
- [4] 喻国明, 刘瑞一, 武丛伟. 新闻人的价值位移与人机协同的未来趋势 [J]. 新闻知识, 2017 (2).
- [5] 喻国明. “机器新闻写作”时代传媒发展的新变局 [J]. 中国报业, 2015 (12 上).
- [6] 黄可. 机器人记者: 本质、模式与意义 [J]. 中国记者, 2015 (5).
- [7] 熊国荣, 李贤秀. “机器人记者”对新闻记者就业的冲击及应对 [J]. 编辑之友, 2016 (11).
- [8] [美] 盖恩, [英] 比尔. 新媒介: 关键概念 [M]. 刘君, 周竞男译. 上海: 复旦大学出版社, 2015: 102.
- [9] 孙瑛. 机器人新闻: 一种基于大数据的新闻生产模式 [J]. 编辑之友, 2016 (3).
- [10] 付松聚. 我国首创机器新闻与人工新闻写作之实证性研究 [J]. 传媒观察, 2015 (11).
- [11] Christer Clerwall. (2014). Enter the robot journalist. *Journalism Practice*, (5): 519-531.
- [12] 孙振虎, 张馨亚. 机器人新闻的发展与反思 [J]. 电视研究, 2016 (6).
- [13] Van Dalen, A. (2012). The algorithms behind the headlines: How machine-written news redefines the core skills of human journalists. *Journalism Practice*, (6): 648-658.
- [14] 李诗. 还有什么新闻是机器人不会写的? 北大万小军详解写稿机器人背后的秘密 [EB/OL]. <https://www.leiphone.com/news/201707/4a3BM1YK3fVSwiCL.html>, 2017-07-10.
- [15] 李诗. 专访北大计算机所万小军: 写稿机器人是新媒体时代的产物 [EB/OL]. <https://www.leiphone.com/news/201707/Tg7Fan6T6qVeXfB7.html>, 2017-07-09.