

批判算法研究视角下微博“热搜”的把关标准考察

王茜

摘要

算法分发已成为网络新闻主要的分发模式，而微博“热搜”是该模式在社会化媒体的典型应用。本文采用批判算法研究的视角考察了微博“热搜”的把关标准。基于对微博年度报告、新闻稿、专利、官方发布以及2400余条“最热搜”上榜词的分析，本文提出了微博“热搜”算法价值的四个要素：时新性、流行性、互动性和导向正确。这些要素内嵌于代码的编写与设计之中，微博的价值观、假设和偏见被推向了“上游”阶段。微博使用一系列策略来使“热搜”生成过程看起来浑然天成且不可避免，包括人气值的实时更新、搜索框提示“大家正在搜”、以及“呈现”“反映”等修辞的使用，以此建构算法把关的合法性话语。

关键词

把关、算法价值、社会化媒体、平台、人工智能

作者简介

王茜，郑州大学新闻与传播学院讲师。电子邮箱：wangxi.fr@zzu.edu.cn。

本文系教育部人文社会科学研究青年基金项目“智能算法驱动下的新闻分发变革研究”（项目编号18YJC860031）成果的一部分。

Investigating the Gatekeeping Criteria of Weibo “Hot Search”: From A Critical Algorithms Studies Perspective

WANG Xi

Abstract

Algorithmic distribution has become the main distribution mode of online news. And Weibo “Hot Search” is a typical application of the algorithmic distribution mode in social media platforms. This article, from the perspective of Critical Algorithm Studies, seeks to investigate the gatekeeping criteria of Weibo “Hot Search”. Based on the analysis of Weibo’s annual reports, press releases, patents, official statements as well as 2400 pieces of “Hottest Searches”, this study identifies a set of four algorithmic values: instantaneousness, popularity,

user engagement, and correct orientation, which are embedded in the code writing and design. Therefore, Weibo's values, assumptions, and biases are moved "upstream". In order to construct the discursive legitimization of algorithmic gatekeeping, Weibo uses a series of tactics to present the generation of "Hot Search" as natural and inevitable, including the real-time update of the search index, the indication of "people are searching" in the search box, and the rhetorics of "presentations" and "reflections".

Keywords

Gatekeeping, algorithmic values, social media, platform, Artificial Intelligence

Author

Wang Xi is a lecturer at the School of Journalism and Communication, Zhengzhou University. Email: wangxi.fr@zzu.edu.cn.

This article is the preliminary result of "Research on the Algorithm-Driven News Distribution" (Project Number: 18YJC860031), supported by the Humanity and Social Science Youth Foundation of Ministry of Education of China.

一、研究缘起：从算法引发的争议谈起

媒介与传播的历史大体经历了四个媒介化浪潮：机械化（mechanization）、电子化（electrification）、数字化（digitalization）和数据化（datafication）（Andersen, 2018）。进入数据化阶段，面对每日生成的数以亿计的内容（包括用户生成内容、专业生成内容、算法生成内容），人工把关模式的局限性日益凸显。在这样的背景下，算法把关模式应用而生，并逐渐成为一种主流形态。

中国互联网络信息中心发布的报告指出，基于用户兴趣的算法分发逐渐成为网络新闻主要的分发方式（CNNIC, 2016）。同年发布的《中国移动资讯信息分发市场专题研究报告》显示，在国内资讯信息分发市场上，算法推送的内容将超过50%（易观，2016）。与此同时，算法分发模式也受到了用户的青睐。根据路透社新闻研究所发布的报告，54%的受访者倾向于通过算法（搜索引擎、社会化媒体、资讯聚合平台）来获取新闻故事，而选择记者、编辑的比例为44%（Newman, Fletcher, Kalogeropoulos, Levy & Nielsen, 2017）。传统把关人控制信息的霸权地位受到了算法的挑战。

在把关权力从人工编辑向智能算法转移和让渡的过程中，信息茧房、劣币驱逐良币等问题不断涌现，引起了媒体和管理部门的关注。2017年9月，人民网推出《不能让算法决定内容》等“三评算法推荐”系列评论；2018年初，管理部门约谈微博，要求其“热搜榜”“热门话题榜”等下线一周进行整改；同年4月，国家

网络信息办公室约谈快手和火山小视频相关负责人，要求其暂停算法推荐功能；不久后，今日头条旗下的内涵段子被国家广播电视总局关停。约谈、整改、下架、关停……主打算法推荐的平台媒体遭到了不同程度的监管。

与此同时，学术界对算法的批判之声也甚嚣尘上。在相关讨论中，算法往往与“过滤气泡”（Pariser, 2011: 10）、“偏见”（Baker & Potts, 2013; 仇筠茜, 陈昌凤, 2018）、“区隔”（姜红, 鲁曼, 2017; 张志安, 汤敏, 2018）、“伪中立性”（朱鸿军, 周逵, 2017）、“黑箱”（帕斯奎尔, 2015/2016: 13）等关键词联系在一起。

综上，一方面，算法的社会权力日益强大；另一方面，传统把关人所遵循的职业规范和新闻伦理并不约束算法系统，后者运作的“黑箱化”给问责设置了重重阻碍，继而引发了诸多问题。在这样的背景下，我们迫切需要了解算法究竟是如何把关的，以及遵循何种价值观念和标准。

针对算法把关标准这一问题，中国学者给出的答案各不相同：姜红和鲁曼（2017）认为是趣味性，陈昌凤和霍婕（2018）认为是受众的注意力，朱鸿军和周逵（2017）则指出，公众兴趣和流行话题是算法判断新闻价值的标准。上述结论虽然体现了研究者敏锐的判断，然而结论的得出主要基于个人观察，缺乏系统的论证。此外，搜索引擎、社会化媒体、资讯聚合平台的算法设计价值也往往存在差异。反观西方传播学界，学者已经对脸书、推特等超级平台的算法实践（Chakraborty, Messias, Benevenuto, Ghosh, Ganguly & Gummadi, 2017; DeVito, 2017; Duguay, 2018; Gillespie, 2016a）进行了富有洞见的个案研究。

微博是中国的超级平台之一：根据Alexa公司2020年3月的统计数据，微博的访问量在全球位居第16位，在中国位居第10位。此外，网民通过微博获取新闻的比例达到35.6%，高于通过新闻客户端获取新闻的比例（CNNIC, 2017）。鉴于此，笔者试图从批判算法研究的视角出发，以微博为研究对象，探讨中国语境下超级平台的算法把关标准。学者强调，算法需要被放置在具体的情境及社会技术装配（socio-technical assemblage）中来进行阐释（Kitchin, 2017）。因此，中国语境下的相关研究有助于我们更好地理解算法的内嵌性，即算法在不同语境下的社会建构。

二、文献回顾与核心概念

（一）批判算法研究

根据Goffey（2018: 16）给出的定义：“算法是为了解决问题而输入机器的一

系列指令。”有学者做了形象的比喻：如果数据是食材，那么算法就是食谱；只要遵循食谱设定的步骤和指令，按照要求搭配和处理食材，就能做出指定口味的菜肴（Willson, 2017）。大数据时代，庞大的数据带给算法庞大的权力。英国文化研究专家Lash（2007）强调：“在一个媒体和代码无处不在的社会，权力越来越存在于算法之中。”

随着算法渗透信息生产、分发与消费日益深入，相关的社会科学研究也开始兴起，其中，批判算法研究是较有代表性的研究取向。新英格兰微软研究院首席研究员Tarleton Gillespie是该研究取向的倡导者和推动者，他和学者尼克·西弗（Nick Seaver）总结了一份详尽的阅读清单，概括了批判算法研究的五个维度（Gillespie & Seaver, 2016）：

第一，研究算法的特定影响与选择逻辑。这部分还可以细分为三个层次：首先，算法具有内嵌的价值观与偏见，可能导致个体化、社会分类与歧视；其次，关注经由算法运作的理性化、自动化和量化带来的对于人类社会评价标准、复杂程度以及语境的化约与抹除；最后，关注与算法相关的政策回应与可信度问责。

第二，算法帮助维持或强化某种特定的意识形态观念，例如资本主义、监视、主体与客体（Cheney-Lippold, 2017）。

第三，算法是复杂的技术组件，需要考虑其内嵌性（situatedness），即算法运作过程中与“规则、人、过程、关系”等的相互作用（Neyland & Möllers, 2017）。此类研究多从行动者网络理论的视角出发。

第四，算法本质上由人来设计，也由人来使用。这一维度主要关注人们如何以特定的方式设计与维持算法，如何与算法共同生存，用户对算法的认知调查，以及算法如何形塑人们对公共议题的感知等议题。

第五，探讨研究算法系统的方法和路径。学者和媒体从业人员采用逆向工程、算法审计、众包等多种方法来研究算法和自动化决策系统。例如，德国媒体人联合创建了非营利性监督组织“算法注视”（Algorithm Watch），于2017年发起线上话题“数据捐赠”（#Datenspende），采用众包的方式对有关德国联邦选举的谷歌搜索结果进行监测。

本文最为关注的是第一个维度的相关研究，尤其是算法内嵌的价值观、可能存在的偏见，以及运作的非中立性（non-neutrality）。

（二）算法把关

把关人理论的源头可追溯至美国社会心理学家库尔特·卢因（Kurt

Lewin)。卢因将“把关”描述为食物到达家庭餐桌的过程，该过程包括把关人（gatekeeper）、门卡（gate）、渠道（channel）等要素（Lewin, 1947）。1950年，传播学者大卫·怀特（David White）将把关人理论引进新闻研究领域（White, 1950），该理论由此发展成为传播学经典理论之一。

传统文献中的把关过程即为取舍选择的过程，如丹尼斯·麦奎尔（Denis McQuail）将“把关”定义为“媒介工作中做出选择的过程，尤其决定某一则新闻故事是否能够通过‘媒体之门’进入到新闻流通环节”（McQuail, 1994: 213-214）。Donohue等学者创新性地指出，“把关”不仅是取舍选择的过程，而是讯息编码过程中“所有形式的信息控制”，包括选择（selection）、塑造（shaping）、展示（display）、时间选择（timing）、隐瞒（withholding）或重复（repetition）（Donohue, Tichenor & Olien, 1972）。在此基础之上，Barzilai-Nahon（2008）增添了渠道（channeling），组合（addition）、本土化（localization）、操纵（manipulation）等信息控制形式。帕梅拉·休梅克（Pamela J. Shoemaker）等后期也强调，把关过程“不仅是一系列的进或出的决定”，而应当被描述为“新闻媒介建构社会现实的整个过程”（Shoemaker, Eichholz, Kim & Wrigley, 2001）。

进入数字时代，新闻业发生了至少五个方面的深刻变化：存储和发布新闻能力的提升、新的新闻生产工具（如算法、用户生成内容）的出现、聚合平台和望关人的兴起、基于单篇故事的竞争以及即时的受众反馈（Pearson & Kosicki, 2017）。在这样的背景下，记者、编辑控制信息的霸权地位受到了平台媒体、用户、算法等的多重挑战。本文重点关注算法所发挥的把关角色。根据Carlson（2018）的梳理，2002年“谷歌新闻”（Google News）的启动标志着算法判断直接入侵了人类的编辑判断领域。作为一款Web新闻聚合器，“谷歌新闻”通过StoryRank算法实现了新闻的自动化组织和排列。中国学者则普遍认为，2012年今日头条的成立拉开了算法进军中国新闻业的序幕（王茜，2017；喻国明，耿晓梦，2018；白红义，李拓，2019）。如今，算法分发已经是资讯聚合平台、搜索引擎、社会化媒体等平台媒体的“标配”。

学者使用不同的概念来阐释基于算法的信息控制，包括算法把关（Bozdog, 2013；Napoli, 2015；Tufekci, 2015；Wallace, 2018）、算法策展（DeVito, 2017；Thorson & Wells, 2016）、算法判断（Carlson, 2018）、算法选择（Just & Latzer, 2017）。尽管它们在用语上存在差异，但在研究的旨趣上是共通的。鉴于把关人理论所具有的丰富意涵，本文统一采用“算法把关”的概念来指代算法所进

行的“所有形式的信息控制”，包括选择、聚合、分类、排序、过滤、组织、呈现、推荐。

随之而来的问题是，算法是如何把关的？和编辑把关相比，它具有什么特征？早在20世纪，美国逻辑学家和计算机科学家罗伯特·科瓦尔斯基（Robert Kowalski）就提出一个关于算法的经典定义：“算法=逻辑+控制”（Algorithm=logic + control）。其中，逻辑部分指定使用哪些知识（数据）来解决问题，而控制部分则决定使用什么策略来解决问题。逻辑部分决定算法的意义，而控制部分仅仅影响算法的效率（Kowalski, 1979）。因此，算法既不完全是物质的，也不完全是人类的——它们是混杂的，由人类意图性（human intentionality）和物质顽固性（material obduracy）共同组成（Anderson, 2013）。这种混杂性使得算法把关具有区别于编辑把关的如下特征：

（1）数据化（datafication）。数据化是指网络化平台将现实世界的多个方面（如用户的线上行为、社交关系、地理位置）转化为数据的能力（Dijck & Poell, 2013），它赋予了平台媒体实时分析、预测和个性化推荐的潜力。

（2）程序化（proceduralization）。算法遵循“如果/那么”逻辑（“if/then” logic），按照预先设定好的指令程序化地运作。所有的人类构念（human constructs）事先都被转化为了具有可操作性的变量、步骤和指令（DeVito, 2017）。

（3）自动化（automation）。在完成初期的设计和编程后，不经人为干预，算法即可自动完成信息的把关。

上述特征为算法把关披上了“机械中立性”（mechanical neutrality）（Gillespie, 2014）或“计算客观性”（calculative objectivity）（Beer, 2017）的光环，并对人工把关模式提出了根本挑战。“人的主观性本质上是受到怀疑和需要代替的，而算法本质上是客观和需要执行的”（Carlson, 2018）成为了一种盛行的观点。

批判算法研究驳斥了上述观点，算法运作的非中立性已经成为了一种共识（Gillespie & Seaver, 2016）。个人因素（如程序员的偏好和价值观念）、组织因素（如企业的方针政策、企业文化）、外部因素（如广告主的诉求、法律规范）等都会影响算法的设计和编程，即科瓦尔斯基所说的逻辑部分。“人为偏见和价值观可以被嵌进程序开发的每一个步骤，电脑化操作仅仅是将人为歧视推到了上游层面”（帕斯奎尔, 2015/2016: 52）。此外，研究发现，算法把关也会带来虚假新闻传播（Carlson, 2018）、有价值的新闻被压制（Tufekci, 2015）、歧视和偏见

(Baker & Potts, 2013)、过滤气泡(Pariser, 2011)等负面影响。

综上所述,算法把关拥有其独特的逻辑和合法性基础,并对编辑把关提出了根本挑战。二者的共同点在于,都不可避免地受到多个层次因素的影响。

(三) 算法把关标准

新闻价值,即新闻主体用来选择和衡量新闻价值客体的标准,是把关人理论的核心议题之一,因为它深入到了“哪些新闻入选,哪些新闻遭到舍弃,以及为什么”的核心(Harcup & O' Neill, 2017)。无论是Galtung和Ruge(1965)早期提出的“九要素说”(包括时间跨度、事件强度、明晰性、文化接近性、预期性、出乎意料性、连续性、组合性和社会文化价值),还是学者在此基础上补充的“十五要素说”(包括独家、坏消息、冲突、惊奇、视听、可分享性、娱乐、戏剧性、新闻后续、权力精英、相关性、量级、名人、好消息和新闻机构的议程)(Harcup & O' Neill, 2017),都为新闻从业者提供了一定的指南,帮助其指导日常的新闻实践。

随着算法成为新兴的把关主体,算法用来选择和衡量新闻价值客体的标准是什么?包含哪些要素?学者对内嵌于不同平台媒体中的算法系统及其把关标准进行了考察。这些标准常常是由程序员、算法工程师等事先定义好的、可编程的,且处于流动之中(Wallace, 2018)。

迈克尔·德维托(Michael DeVito)聚焦脸书信息流的算法价值,通过对脸书公开发布的专利、新闻稿、博客等进行内容分析,德维托提炼出了九大算法价值要素,分别是:朋友关系、用户公开表达的兴趣、用户先前的参与、用户含蓄表达的偏好、发布时间、平台优先级、页面关系、用户的负面表达和内容质量。其中,朋友关系是脸书信息流算法的首要价值要素(DeVito, 2017)。

方师师(2018)对双强寡头平台——谷歌和脸书的算法机制进行了研究,指出谷歌注重客观要素并试图建立技术性的网络秩序,而脸书则更加强调以社交为基础的人与人之间的关系构成。关于脸书的研究结果与德维托的结论相一致。

Gillespie(2016a)将目光投向推特的“热门话题”(Trending Topics),认为其算法价值要素包括三个层面:在时间(when)层面,推特强调时新性;在内容层面(what),新奇性要素受到推崇,那些太过普通的词汇(如“今天”)无法单独登上榜单,且仇恨言论、亵渎的言语等会被自动过滤;在主体(who)层面,并非所有人的参与和讨论都被同等对待。比起那些仅在现有用户群(用户已经彼此连接,互相关注,或是位于同样的地理位置)之间传播的词汇,那些能够超越群组

(span clusters) 的词汇被推特认为更加重要。

Klinger和Svensson (2018) 对比分析了大众媒体和社会化媒体不同的把关标准：大众媒体主要依照新闻价值来进行新闻生产，而社会化媒体所遵循的价值是产销合一 (produsage)、连接性、病毒性以及志趣相投个体之间的信息分享。

综上所述，针对脸书、谷歌、推特等西方超级平台的算法把关标准，学界已经有了较为丰富的研究成果，而非西方语境下的系统研究仍较为缺乏。本文尝试从批判算法研究的视角切入，以微博“热搜”为研究对象，探讨中国语境下超级平台的算法把关标准。在智能算法得以广泛应用和推广的大势所趋之下，探究该问题有助于我们更好地理解 and 把握当下新闻传播领域正在发生的权力转移和深刻变革。

三、研究设计

(一) 研究对象

“微博热搜榜是全网最实时权威的热点排行榜，依据用户的真实搜索量，呈现最新鲜、最热门、最有料的资讯。”（微博热搜榜，2016）本文选取微博“热搜”作为研究对象，原因有二：其一，微博是中国最重要的互联网公共表达平台之一，而“热搜”常常被认为是公共讨论的“晴雨表”和“风向标”，发挥着重要的把关和议程设置作用。根据“微博热搜榜”（2016）发布的年度报告，用户每日主动搜索2.5亿次，引导1.35亿搜索热度，每人日均搜索超过12次。其二，“热搜”功能是算法分发模式在社会化媒体的典型应用，西方传播学界围绕脸书、推特的“热门话题”已经开展了深入讨论。本研究有助于加深我们对中西方语境下社会化媒体运作逻辑的理解。

(二) 研究方法

Kitchin (2017) 在《批判性地思考和研究算法》一文中，提出了研究算法的六种方法：研究伪代码和源代码，反射地生成代码，逆向工程，访谈或民族志，对公开材料的文本分析，检验算法在现实世界的运作。鉴于笔者不具有编程技能，且算法价值观是平台媒体的核心商业机密，本研究主要采用第五和第六种研究方法。

1. 对微博公开材料的文本分析

笔者搜集了微博发布的年度报告、新闻稿、专利、核心人物（如微博CEO、微博副总裁等管理人员）访谈，以及“微博热搜榜”“微博搜索”“微博管理员”“微博客服”“微博小秘书”等官方账号发布的公开声明，从中找出与“热搜”“规则”等相关的内容。基于对上述材料的文本分析，笔者梳理了“热搜”上

榜规则所经历的调整，从而把握算法价值的演变。

2. 对“最热搜”上榜词的文本分析

分析公开材料仅能够揭示微博如何对外宣称“热搜”的算法价值，鉴于此，笔者还试图检验在实际的运作中，哪些内容优先通过了“门卡”，哪些内容被排除在外。这在一定程度上可以逆向推导“热搜”的把关标准，然而不可避免地难以勾勒出全貌。

2018年初，北京市网络信息办公室约谈新浪，要求其对“热搜榜”“热门话题榜”等产品下线一周进行整改，下线时间从1月27日21时至2月3日21时。此次整改为本研究的开展提供了契机。鉴于“最热搜”是每日榜首热点的集合，在榜时间更长，讨论范围更广，笔者手动记录了2018年2月3日至9月18日所有上榜“最热搜”的搜索词及其人气值、在榜时间，共得到2404条。由于人气值和上榜时间在当日处于不断的变化之中，笔者采用次日记录的方式，此时数值在多数情况下已趋于稳定。9月19日，微博改版，“最热搜”版块下线，故而资料搜集中断。

此外，笔者以“登不上热搜”“无法登上热搜”“难以登上热搜”等为关键词，检索用户于记录期间发布的微博，从中筛选出用户认为理应上榜的新闻事件（见表1）。

表1：未登上“热搜”的新闻事件

类别	新闻事件
国际新闻	中美贸易战、美国制裁中兴、美英法空袭叙利亚
时政新闻	宪法修正案通过、个人所得税法修正案通过
财经新闻	P2P爆雷潮、股市下跌、鸿茅药酒事件
社会新闻	四川江安工业园区爆燃事故、陕西米脂县砍杀学生事件、北京西单大悦城砍人事件、湖南衡东恶性驾车伤人事件、中国国际航空1350号班机劫持事件、朝鲜涉中国游客重大交通事故、美达员工98%烧伤身亡、德阳女医生自杀事件、湖南耒阳学生分流事件、山东中医药大学研究生导师梁栋性侵事件、南昌大学遭性侵女生起诉学校和副院长、央视主持人涉嫌性骚扰女实习生、微博清理同性恋内容、广东阳春水灾、大雾致海口上万车辆滞留

四、研究发现：微博“热搜”的算法价值四要素

通过对微博公开材料的文本分析，笔者对“热搜”上榜规则的调整和更新进行了如下梳理：

2014年11月，微博手机客户端上线“实时热搜榜”，每十分钟更新一次（微博

搜索, 2014);

2015年4月, 微博上线“好友热搜榜”, 该榜单“使用微博一度好友和二度好友关系计算, 呈现一段时间内多个好友共同搜过的词”(微博搜索, 2015a);

2015年末, 《微博搜索白皮书》显示“热搜”的日均曝光达到了2.5亿(微博搜索, 2015b);

2017年3月8日, “热搜”由每10分钟更新一次提升至每1分钟更新一次(新浪, 2017);

9月12日, “最热搜”版块上线, 汇聚每日的榜首热点, 榜首词以天为单位分隔, 按登上榜首的时间排列(微博搜索, 2017);

2018年2月3日, 整改后的“热搜”上线“新时代”版块; 随后微博副总裁在采访中表示, “热搜”在选择和排序上放弃纯粹的算法模式, 引入编辑人工干预模式(杨阳, 2018);

3月15日, “热搜榜”排序规则从单一考量“搜索热度”转变为“搜索热度*话题因子*互动因子”的三重考量(微博热搜榜, 2018);

8月28日, 根据微博高管的公开分享, “热搜榜”的排序公式进一步调整为“(搜索热度+传播热度)*话题因子*互动因子”(微博MCN, 2018);

2019年8月, 微博副总裁称“热搜”将有更多垂直领域的内容(曹增辉, 2019)。

基于以上梳理, 并结合对“最热搜”上榜词的分析, 笔者提炼出了微博“热搜”算法价值的四个要素, 分别是: 时新性、流行性、参与性和导向正确。下文将分别进行讨论。

(一) 时新性

时新性是指“热搜”上榜词在时间上是新近乃至正在发生的, 在内容上是新奇的。

如前所述, 2017年3月, “热搜”由每10分钟更新一次, 提升至每1分钟更新一次。根据改版时的声明, “新版热搜榜不同于按照固定时间段搜索热度累积来衡量的榜单, 将更加凸显此时此刻正在发生的热点事件”(新浪, 2017)。对实时更新的推崇和强调, 给人一种人工无力完成, 因而运作过程是自动化的、免受人工干预的印象(Duguay, 2018)。

此外, 通过对2404条“最热搜”上榜词的分析, 笔者发现常规化的词汇难以登上“热搜”, 上榜词常常具有新奇性、异常性。据笔者观察, 上榜词主要由汉字组

成，也包括数字、英文和符号（主要是“#”）。世界杯期间，一条由纯数字组成的搜索词——“272 614 886”上榜，意思是“厄齐尔+罗伊斯=拜拜了”。所有的上榜词中，字数最少的是2个（如“福字”“初八”），最多14个（9月1日的“阿里员工去世生前租自如甲醛房”）。

在笔者为期228天的观察和记录过程中，除了极少数搜索词能够重复登上“最热搜”（如“李晨 范冰冰”于2月14日、2月27日和5月20日上榜；“快乐大本营”于2月10日、3月10日和3月17日上榜），绝大多数搜索词仅仅上榜一次。这与大众媒体的把关标准相一致，那些曾经出现过的新闻故事难以再次通过筛选过程（White, 1950）。对内容新奇性的强调还使得常规性、持久性的议题被边缘化。例如，旷日持久的中美贸易摩擦、叙利亚问题，“两会”期间的提案、议案等难以成为“搜索量非常高、上升趋势非常快”的搜索词。“两会”期间唯一上榜的新闻是“看两会 没忍住买了一箱苹果”。

（二）流行性

流行性是指内容所具有的易于为多数人所喜爱并突然间迸发的性质。根据英国文化思想家雷蒙德·威廉斯（Raymond Williams）对关键词“popular”的考证，该词最早意指“属于民众”，在现代的主要意涵是“受喜爱的”“受欢迎的”。这种演变包含了“讨人喜欢”的特质，并且保有“刻意迎合”的旧意涵（威廉斯，1976/2016：401-403）。

2018年3月之前，单位时间内的搜索量是“热搜”的唯一测量指标。例如，2016年中旬，有网友对“电影《没有别的爱》演员撤换事件”登不上“热搜”提出质疑，微博CEO“来去之间”回应称“热搜不是根据转评赞来决定上不上，是根据搜索量上的”。微博在多份官方声明中也强调，排行榜是“依据用户的真实搜索量”（微博热搜榜，2016），“基于实时搜索数据计算而来”（Weibo, 2018）。

上述测量方式暗含了这样的逻辑：多数人实时搜索了什么，什么就是热门。使用算法来挖掘用户的数字痕迹通过三种“抹除”（exclusion）造成了“多数人的暴政”（the tyranny of the majority）：数据采集过程中对细节和社会情境的抹除，基于“大数法则”进行预测造成了对离群值和差异的抹除；对既有网络的强化、偏袒造成了对新声音、弱小声音的抹除（Harper, 2017）。在这种逻辑之下，流行的人（如明星、网络红人）、流行的事物（如世界杯、热门电视剧）大行其道，而少数群体的声音（如同性恋、残疾人、农民）被边缘化。

参照“微博热搜榜”（2016）使用的分类标准，笔者将2404条上榜词分为以

下四种类型：娱乐类新闻资讯、非娱乐类新闻资讯、新生事物和影视综艺。如表2所示，67.6%的上榜内容是娱乐类新闻资讯，主要是明星“结婚”“怀孕”“分手”“减肥”等八卦信息。微博上拥有粉丝最多的20个账号中，16个是演艺明星，谢娜等4人的粉丝量破亿。名人拥有远高于普通民众的可见度和曝光量，且很多议题（如6月1日的“HIV感染者与明星首次出境”）只有在涉及到名人时才会突然间迸发。

表2：“最热搜”上榜词的类别列表

类别	类别描述	所占比例	代表性上榜词
娱乐类新闻资讯	指明星动态、娱乐圈八卦、网络红人等的相关讨论。	67.6%	“阚清子承认分手”“林俊杰脱粉”“范丞丞胖了”“景甜回应恋情”
非娱乐类新闻资讯	指时政、社会、体育、财经等新闻的相关讨论。	21.1%	“普京让安倍等了48分钟”“孙杨要求重新升国旗”“教育局回应怀孕教师支教”“华谊兄弟质押股权”
影视综艺	指电影、电视剧、综艺节目等的相关讨论。	6.5%	“流星花园收视率”“如懿传定档”“跑男提档”“春晚主持阵容”
新生事物	指网络新词、热议段子等的相关讨论。	4.8%	“绿豆汤是红的还是绿的”“男朋友可以用来干什么”“#下面有请扁桃体发炎#”

图1是基于“热搜”人气值制作的直方图。可以看到，大多数情况下，人气值达到100万以上才能上榜“最热搜”，多数上榜词的人气值集中在200万至400万之间，有的能达到一千万以上。除了“杨紫秦俊杰分手”“韩庚卢靖姗公开恋情”等31条娱乐类新闻资讯之外，有6条非娱乐类新闻资讯的人气值超过一千万，内容是有关“山竹登录”和“滴滴遇害女孩”的讨论。

（三）互动性

互动性是指内容所具有的易于促进用户参与和社交行为（如点赞、评论、转发、关注）的性质。

如前所述，自2018年3月15日起，“热搜榜”排序规则从单一考量“搜索热度”转变为“（搜索热度+传播热度）*话题因子*互动因子”的多重考量。其中，“搜索热度”以搜索量为基础，反映用户对热点的关注和探索的程度；“传播热度”以热点相关博文阅读量为基础，反映热点触达人群的规模；“话题因子”以话题讨论量为基础，反映用户热议和参与的热情；而“互动因子”以结果页转评赞

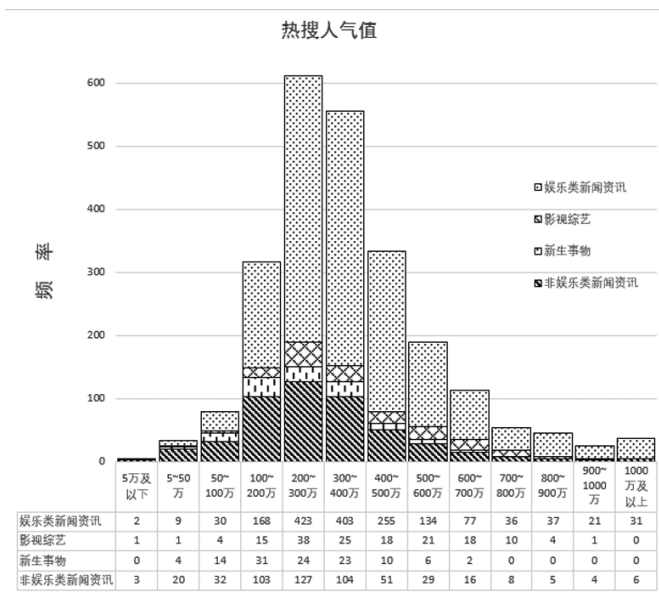


图1：“最热搜”上榜词的人气值分布

互动率为基础，反映用户消费内容的意愿（微博MCN，2018）。笔者观察记录期间，共有148条包含双井号的话题（如“#最受异性欢迎的身高排行#”“#手机上瘾是一种精神障碍#”“#该不该取消本科论文#”）上榜。

尽管从表面上看，“热搜”仍然是原来的排行榜，然而其对“热门”的界定及测量方式发生了根本改变，互动性开始成为重要的价值要素。设想一下：在私底下搜索的内容和公开讨论、参与、“转评赞”的内容，哪一种更能代表用户的真实关注呢？前者往往具有隐秘性，后者则常常含有表演成分。两种方式共同之处在于，都将现实世界的用户降级为一串串的数字痕迹。尽管微博宣称，新的排序规则能够“更全面反映热点的关注度”（微博热搜榜，2018），然而“热搜”和民意调查、街头访谈、投票机制等一样，仅仅是宣称能够代表民意的众多信息实践中的一种（Gillespie，2012）。在标明“热门”应优先考虑哪个因素这个问题上，“人类判断的极偶然结果，被说成是只要开着计算机运行脚本就会产生的不可避免的客观结果”（莫罗佐夫，2013/2014：160）。

此外，微博设置“好友搜”版块，展现多个好友共同搜索过的内容；开发“你可能认识的人”功能，促使用户扩大关注范围，均是为了提高用户和用户之间、用户和内容之间的互动。学者称之为“编程的社交性”（programmed sociality），

即算法和软件深度介入到了人际关系网络的创建、维系、形塑和组织过程之中（Bucher, 2013）。

值得注意的是，并非所有用户的搜索、互动和参与都会被同等地纳入考量。“微博管理员”在《关于“紫光阁地沟油”从未进入微博热搜榜的说明》中指出，“热搜”计算主要依据的是“可信用户的搜索行为”，为了防止刷榜，还会考虑用户群分布、终端系统分布、搜索操作特征等维度。微博的出发点是保证榜单的真实性和可信度，然而在具体运作过程中，却会出现“算法误判”，造成意料之外的负面影响。

在此通过一个案例来阐释“算法误判”。2018年11月3日，为了庆祝iG夺冠，王思聪通过微博抽奖平台发起了一场平分113万元现金的活动，吸引了上千万人转发。然而最终的中奖者中有112人是女性，仅有1名男性。微博CEO“来去之间”对此解释，抽奖最关键的是“不能抽出来机器号”，“但凡行为有点儿像水军的（比如只转发从来不原创，没发过图，等等等等）抽奖时候都会降权，女生一般行为更活跃，男的只看不说话，是比较吃亏。”微博的一项发明专利申请也指出，现有技术“在识别虚假热门信息时存在误伤、误判的情况（微梦创科，2019）。

算法以一种不可见的、隐蔽的方式对人和物进行分类，体现了算法的权力。“类别是什么，什么属于某个类别，谁来决定在实践中如何实施分类等，都是有关事物是什么以及应当是什么的有力断言（powerful assertions）”（Gillespie, 2014）。在上述案例中，微博对“可信用户”和“水军”的界定需要被转化为机器可读取的具体指令，而转化过程不可避免地会造成“剪裁现实生活、忽视社会情境、抹杀主体建构、取消生活意义”（潘绥铭，2016）。“只看不说话”的用户在不知情的情况下被算法自动判定为水军，无从申诉，且难以提出质疑或抗议。在微博“大家正在搜”的口号背后，一部分用户的参与被降权或排除。

（四）导向正确

习近平在2016年2月19日党的新闻舆论工作座谈会上说：

新闻舆论工作各个方面、各个环节都要坚持正确舆论导向。各级党报党刊、电台电视台要讲导向，都市类报刊、新媒体也要讲导向；新闻报道要讲导向，副刊、专题节目、广告宣传也要讲导向；时政新闻要讲导向，娱乐类、社会类新闻也要讲导向；国内新闻报道要讲导向，国际新闻报道也要讲导向……

作为中国最有影响力的社会化媒体和新闻获取的重要渠道，微博也被纳入了日

常媒介规制。

需要指出的是，导向正确并非一开始就由平台媒体自觉纳入算法设计价值，而是经过了本文开头提到的约谈、整改、下架、关停等一系列监管措施之后，才开始受到平台媒体的重视。整改后的“热搜”上线“新时代”频道，从而更好地宣传十九大精神，弘扬社会主义核心价值观，报道中国新时代的社会风貌（观察者网，2018）。笔者观察记录期间，有46条正能量搜索词登上榜首，它们弘扬了家庭和谐（如“婆婆推着儿媳看樱花”“妈别怕 有人来接”）、爱岗敬业（如“她是全中国最美的逆行”“扎针技术生疏 同事献身试练”）、友善互助（如“环卫工已故儿子照片找到了”“忘带准考证12分钟取回”）、自强不息（如“81岁学霸奶奶获本科学位”“707分考入北大 感谢贫穷”）等中国传统价值观念。网络正能量、宣传和意识形态教育等等，成为新的治理策略和公共话语，体现了政治权力更巧妙的表现形式（杨国斌，2017）。

与此同时，对导向正确的强调也可能造成大量具有公共性的搜索词（如表1列出的社会新闻）无法登上“热搜”或是被撤出榜单，发展成为一种变相的信息审查。笔者在观察记录期间共发现两次人工干预“最热搜”的情况，一次是7月12日“新闻联播 播出事故”搜索结果未予显示，另一次是9月3日“刘强东被捕照片”被快速撤下榜单。

除了上述四个算法价值要素之外，其他价值要素（如重要性）也在一定程度上被微博纳入了考量，然而并不占据主要地位。笔者注意到，人气值在5万以下的搜索词（见表3）也曾登上榜首，且集中出现在5月25日至28日期间；一些搜索词的热度还在1万以下。据笔者推测，微博在此期间试图对算法价值进行更改和调试，将具有重要性、突发性（如自然灾害）的新闻放置在“热搜”的显著位置，给予更长

表3：人气值在5万以下的“最热搜”

上榜时间	上榜内容	人气值	在榜时间 (分钟)	备注
5月25日	#雨好大啊#	33656	294	对影视剧中下大雨情节的讨论
5月27日	张韶涵遭父亲控诉	36583	761	
5月28日	松原地震	7003	610	
5月28日	哈尔滨地震	7462	282	
5月28日	震醒整个东三省	18806	249	
5月28日	林允手机号	49104	140	

的曝光时间。这呼应了微博此前的声明，整改后的“热搜”会在选择和排序上放弃纯粹的算法模式，引入编辑人工干预模式。

然而，这一改革和尝试并未持久推行。“热搜”仍然主要遵循“热度优先”原则，其原因可能在于“人机协作”会破坏“热搜”试图建立和维持的自动化运作、因而免受人工干预的形象。一个典型例子是包含错别字的搜索词——2月17日的“谁不想涌有范冰冰这样的姐姐”——也曾上榜“最热搜”。由此可见，微博注重对用户数字痕迹的聚合和呈现，而弱化了编辑所秉持的核实原则。

五、讨论与结论

（一）研究的理论价值和应用价值

本文采用批判算法研究的视角，对微博“热搜”的算法把关标准进行了考察。研究发现，微博“热搜”的算法价值要素主要包括时新性、流行性、互动性和导向正确。以上要素反映了微博关于“什么应当算作‘热门’”的假设，其价值观被推向了“上游”阶段：无论是数据来源的界定（如保留“可信用户”的搜索，排除“刷榜”数据）还是指标的设定（如“热搜热度”“传播热度”“互动因子”和“话题因子”的定义及测量方式、各个价值要素的权重），都在算法的设计和编程阶段被转化为了可操作的指令。

算法的创建并非出于中立目的，而是为了创造价值和利润，以某种方式助推行为或构造偏好，或对人进行识别、排序和分类（Kitchin, 2017）。尽管中西方平台媒体都以盈利为主要目标，然而本研究揭示了中国语境下平台媒体算法价值不同于西方的独特一面。尤其是在宣传新闻主义观念的指导下，导向正确要素被纳入中国平台媒体的算法设计价值。除了微博强调“更高的价值标准和更大的责任担当”（微博管理员，2018）之外，快手CEO宿华（2018）表示要“用正确的价值观指导算法”，今日头条CEO张一鸣（2018）也承诺要“将正确的价值观融入技术和产品”，这与其早期的定位——“没有采编人员，不生产内容，没有立场和价值观，运转核心是一套由代码搭建而成的算法”——形成了强烈反差。

Anderson（2013）提出了算法新闻学研究的六种社会学路径：政治和公共政策视角、经济视角、制度和场域视角、组织动力学视角、文化-历史视角以及技术与新闻视角。其中，第一种路径关注新闻业所处的政治情境，尤其是“公共政策决定如何促进或限制算法新闻的发展”。未来研究可以进一步探讨中国语境下政治力量对算法新闻的规训。

此外,本研究揭示了“人机联姻”在具体执行过程中的复杂性。尽管微博表示将放弃纯粹的算法模式,引入编辑人工干预模式,然而在具体实施过程中仍然不遗余力地建构和完善以“机械中立性”和“计算客观性”为核心的合法性话语。微博使用了一系列策略来使“热搜”生成过程看起来浑然天成且不可避免,包括人气值的实时更新、搜索框提示“大家正在搜”、以及“呈现”“反映”等修辞的使用。例如,微博程序员“文如破竹”发文称“热搜”是“客观的‘热度/人气/关注度’的衡量与体现”。

根据批判算法研究的已有文献,算法把关并不是对公共讨论镜子式的“呈现”和“反映”,它也在“引导”和“形塑”。“‘原始数据’是一种修辞”(“raw data” is an oxymoron) (Gitelman, 2013),平台媒体使用的采集机制决定了它最终能够获取什么样的数据。正如威廉斯所指出的,“群众是不存在的,所存在的只是把人民视作群众的方式”(Williams, 1958: 300)。相应地,“大家正在搜”中的“大家”是不存在的,所存在的只是“热搜”算法按照一定的标准对“大家”的建构。Gillespie (2014)称之为“计算的公众”(calculated publics),即人们只能通过算法生成结果(如“热搜”人气值)来感知“大家”的存在,然而其具体构成是模糊的、不透明的,且可能存在偏见。前文讨论的王思聪抽奖事件就揭示了“只看不说话”用户的参与被降权或排除的事实。

值得注意的是,微博“热搜”的合法性在一定程度上得到了大众媒体的接纳。在诸多的新闻报道中(例如,《人民日报》微信公众号的推文《昨晚热搜第一!她退役后,中国乒乓的大魔王们好寂寞》、新华社微信公众号的推文《有人建议下调法定结婚年龄上热搜!你愿意多少岁结婚?》),“上热搜”成为了“网民关注度高”的代名词。平台媒体的度量指标(metrics)日益成为一套对议题流行度和人物影响力进行测量、排序的合法化标准(Dijk & Poell, 2013)。社会化媒体逻辑和大众媒体逻辑的日益交织,是值得继续探讨的问题。

除了理论价值之外,研究微博“热搜”的算法把关标准还有重要的应用价值。腾讯网络媒体总编辑陈菊红曾指出,作为最大的信息原生地和中转地,社交平台调一下“指针”,就会影响一批内容生产者的流量(腾讯传媒研究院,2016: I-IV)。她所说的“指针”在很大程度上是指算法的设计价值。例如,脸书2018年1月调整信息流算法,优先推荐“有意义的社交互动”,该决定造成法国《世界报》等12家新闻机构在脸书的互动量(包括分享、评论和点赞)平均下降了9.4%(Cornia, Sehl, Levy & Nielsen, 2018)。随着“热搜”将垂直度纳入考量,时尚

和美妆两个领域的流量较前一年增加了3倍（曹增辉，2019）。

在超级平台日益主导新闻分发的当下，“内容为王”“酒香不怕巷子深”的思维模式亟需转变。搜索引擎优化（search engine optimization, SEO）、社会化媒体优化（social media optimization, SMO）、自动化推荐系统等都对内容的组织和呈现方式提出了新的要求。新闻从业者需要了解平台媒体的推荐规则和算法逻辑，从而更好地应对蓬勃而至的智媒化浪潮。

（二）研究的局限性和不足

需要指出的是，算法和用户之间并不是单向影响的关系，而是紧密地交织和勾连（entanglement）。Dijck和Poell（2013）在《理解社会化媒体逻辑》一文中指出，算法的权力主要体现在编程性（programmability）：一方面是社会化媒体平台刺激和引导用户进行创造性或沟通性贡献的能力；另一方面是用户通过与这些被编码的情境互动，转而影响平台信息流的能力。本研究的一大不足，是未充分考察用户使用行为和实践对平台的形塑。例如，微博“热搜”的娱乐八卦类话题居多，一方面是因为算法以流行度为“指针”，另一方面，也可能源自微博用户的自发搜索。

为了弥补用户研究的不足，未来可以借鉴Chakraborty等（2017）对算法“输入”的研究。通过对比推特整体的人口统计学分布和“热门话题”推动者的人口统计学分布，研究者发现边缘群体（如黑人、女性、中年人）很少能够推动某一话题成为“热门”。

此外，算法的“黑箱”特性和流动性是本研究面临的两大挑战。一方面，算法被视作核心商业机密而讳莫如深，是难以探查的“黑箱”（帕斯奎尔，2015/2016：6）。鉴于此，单纯依赖文本分析和逆向推导得出的结论具有一定的局限性。未来研究可进一步通过深入平台媒体内部，与算法架构师、程序员、编辑等开展深度访谈，或采用民族志等研究方法来获取一手材料。另一方面，平台媒体所采用的算法系统并不是静态的、一成不变的，而是在不断地演进、优化和迭代。因此，本研究的结论需要进行阶段性地重访和更新。

（责任编辑：吴欣慰）

参考文献 [References]

白红义，李拓（2019）。算法的“迷思”：基于新闻分发平台“今日头条”的元新闻话语研究。《新闻大学》，（1），30-44。

曹增辉（2019）。热门热搜增加垂直内容。检索于<https://www.pearvideo.com/>

- video_1585947。
- 陈昌凤, 霍婕 (2018)。权力迁移与人本精神: 算法式新闻分发的技术伦理。《新闻与写作》, (1), 63-66。
- CNNIC (2016)。第38次中国互联网络发展状况统计报告。检索于<http://www.cnnic.net.cn/hlwfzyj/hlwxbg/hlwtjbg/201608/P020160803367337470363.pdf>。
- CNNIC (2017)。2016年中国互联网络新闻市场研究报告。检索于<https://cnnic.cn/hlwfzyj/hlwxbg/mtbg/201701/P020170112309068736023.pdf>。
- 方师师 (2018)。双强寡头平台新闻推荐算法机制研究。《传播与社会学刊》(香港), (43), 103-122。
- 弗兰克·帕斯奎尔 (2015/2016)。《黑箱社会: 掌控信息和金钱的数据法则》(赵亚男译)。北京: 中信出版社。
- 观察者网 (2018)。微博热门区新增新时代频道 弘扬社会主义核心价值观。检索于https://www.guancha.cn/TMT/2018_02_01_445507.shtml。
- 姜红, 鲁曼 (2017)。重塑“媒介”: 行动者网络中的新闻“算法”。《新闻记者》, (4), 26-32。
- 雷蒙德·威廉斯 (1976/2016)。《关键词: 文化与社会的词汇》(刘建基译)。北京: 生活·读书·新知三联书店。
- 潘绥铭 (2016)。生活是如何被篡改改为数据的?——大数据套用到研究人类的“原罪”。《新视野》, (3), 32-35。
- 仇筠茜, 陈昌凤 (2018)。基于人工智能与算法新闻透明度的“黑箱”打开方式选择。《郑州大学学报(哲学社会科学版)》, (5), 84-88。
- 宿华 (2018)。接受批评, 重整前行。检索于<https://mp.weixin.qq.com/s/blhnOYRiuPZMfDO4HhhBAG>。
- 腾讯传媒研究院 (2016)。《众媒时代》。北京: 中信出版社。
- 王茜 (2017)。打开算法分发的“黑箱”: 基于今日头条新闻推送的量化研究。《新闻记者》, (9), 7-14。
- 微博管理员 (2018)。微博社区公告。检索于https://weibo.com/1934183965/G0nPDuIG7?from=page_1002061934183965_profile&wvr=6&mod=weibotime&type=comment。
- 微博MCN (2018)。黄远文: 热搜和新话题的原理与技巧。检索于<https://weibo.com/6159353180/GwTCnmgYC>。
- 微博热搜榜 (2016)。微博热搜榜2016产品报告。检索于https://weibo.com/1658035485/Dt4MKzZ3f?refer_flag=1001030103_&type=comment#_rnd1553091000274。
- 微博热搜榜 (2018)。如何打造#热搜话题#。检索于https://weibo.com/1658035485/G88PcoHOn?from=page_1002061658035485_profile&wvr=6&mod=weibotime&type=comment#_

- rnd1548165426326。
- 微博搜索（2014）。微博搜索“实时热搜榜”上线。检索于https://weibo.com/1894467483/BuRtaDz63?from=page_1002061894467483_profile&wvr=6&mod=weibotime&type=comment#_rnd1560854990690。
- 微博搜索（2015a）。“好友热搜榜”上线。检索于<https://weibo.com/1894467483/CcCjqv7Qg>。
- 微博搜索（2015b）。2015微博搜索白皮书。检索于https://weibo.com/1894467483/D8o7hF27Q?type=comment#_rnd1560857829069。
- 微博搜索（2017）。微博热搜榜新功能“最热搜”上线。检索于https://weibo.com/1894467483/FlvmXnAsL?from=page_1002061894467483_profile&wvr=6&mod=weibotime&type=comment#_rnd1536025763274。
- 微梦创科（2019）。一种对网站中作弊行为的监测处理方法及装置。检索于<https://patentimages.storage.googleapis.com/ad/85/8d/4640b37a9abe7a/CN110602184A.pdf>。
- 新浪（2017）。1分钟快速响应 微博打造全网最快热搜榜单。检索于<http://tech.sina.com.cn/i/2017-03-09/doc-ifychhuq3442894.shtml>。
- 杨国斌（2017）。中国互联网的深度研究。《新闻与传播评论》，（1），22-40。
- 杨阳（2018）。专访微博副总裁曹增辉：热搜榜到底整改了什么。检索于<http://www.jiemian.com/article/1936929.html>。
- 叶夫根尼·莫罗佐夫（2013/2014）。《技术至死：数字化生存的阴暗面》（张行舟，闫佳译）。北京：电子工业出版社。
- 易观（2016）。2016中国移动资讯信息分发市场研究专题报告。检索于<https://www.analysis.cn/analysis/trade/detail/1000218/>。
- 喻国明，耿晓梦（2018）。智能算法推荐：工具理性与价值适切——从技术逻辑的人文反思到价值适切的优化之道。《全球传媒学刊》，（4），13-23。
- 张一鸣（2018）。道歉和反思。检索于<https://mp.weixin.qq.com/s/4r6rCwNE7BgTLD37cPJOoA?>。
- 张志安，汤敏（2018）。论算法推荐对主流意识形态传播的影响。《社会科学战线》，（10），174-182。
- 朱鸿军，周逵（2017）。伪中立性：资讯聚合平台把关机制与社会责任的考察。《南昌大学学报（人文社会科学版）》，（5），102-107。
- Andersen, J. (2018). Archiving, ordering, and searching: search engines, algorithms, databases, and deep mediatization. *Media, Culture & Society*, 40(8), 1135–1150.
- Anderson, C. (2013). Towards a sociology of computational and algorithmic journalism. *New Media & Society*, 15(7), 1005–1021.

- Baker, P., & Potts, A. (2013). 'Why do white people have thin lips?' Google and the perpetuation of stereotypes via auto-complete search forms. *Critical Discourse Studies*, 10(2), 187–204.
- Barzilai-Nahon, K. (2008). Toward a Theory of Network Gatekeeping: A Framework for Exploring Information Control. *Journal of the American Society for Information Science and Technology*, 59(9), 1493–1512.
- Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, 20(1), 1–13.
- Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, 15(3), 209–227.
- Bucher, T. (2013). The Friendship Assemblage. *Television & New Media*, 14(6), 479–493.
- Carlson, M. (2018). Automating judgment? Algorithmic judgment, news knowledge, and journalistic professionalism. *New Media & Society*, 20(5), 1755–1772.
- Chakraborty, A., Messias, J., Benevenuto, F., Ghosh, S., Ganguly, N., & Gummadi, K. P. (2017, May). Who Makes Trends? Understanding Demographic Biases in Crowdsourced Recommendations. Paper presented at Proceedings of the Eleventh International AAAI Conference on Web and Social Media. Montreal.
- Cheney-Lippold, J. (2017). *We are data: Algorithms and the making of our digital selves*. New York, NY: New York University Press.
- Cornia, A., Sehl, A., Levy, D. A. L., & Nielsen, R. K. (2018). *Private Sector News, Social Media Distribution, and Algorithm Change*. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-10/Cornia_Private_Sector_News_FINAL.pdf
- DeVito, M. A. (2017). From Editors to Algorithms: A values-based approach to understanding story selection in the Facebook news feed. *Digital Journalism*, 5(6), 753–773.
- Dijk, J. Van, & Poell, T. (2013). Understanding Social Media Logic. *Media and Communication*, 1(1), 2–14.
- Donohue, G.A., Tichenor, P.J., & Olien, C.N. (1972). Gatekeeping: Mass media systems and information control. In F.G. Kline & P.J. Tichenor (Eds.). *Current perspectives in mass communication research* (pp. 41–70). Beverly Hills, CA: Sage.
- Duguay, S. (2018). Social media's breaking news: the logic of automation in Facebook Trending Topics and Twitter Moments. *Media International Australia*, 166(1), 20–33.
- Galtung, J., & Ruge, M. H. (1965). The Structure of Foreign News. *Journal of Peace Research*, 2(1), 64–90.
- Gillespie, T. (2012). *Can an Algorithm be Wrong?* Retrieved from <http://escholarship.org/uc/item/0jk9k4hj>.

- Gillespie, T. (2014). The relevance of algorithms. In P. Boczkowski & K. Foot (Eds.). *Media technologies: Essays on communication, materiality, and society* (pp. 167–193). Cambridge, MA: MIT Press.
- Gillespie, T. (2016a). #trendingistrending: When algorithms become culture. In R. Seyfert & J. Roberge (Eds.). *Algorithmic Cultures Essays on Meaning, Performance and New Technologies* (pp. 52–71). New York, NY: Routledge.
- Gillespie, T. (2016b). Algorithm. In B. Peters (Ed.). *Digital keywords: a vocabulary of information society and culture* (pp. 18–30). Princeton University Press.
- Gillespie, T., & Seaver, N. (2016). *Critical Algorithm Studies: a Reading List*. Retrieved from <https://socialmediacollective.org/reading-lists/critical-algorithm-studies/>.
- Gitelman, L. (2013). *“Raw data” Is an Oxymoron*. Cambridge, MA: MIT Press.
- Goffey, A. (2008). Algorithm. In M. Fuller (Ed.). *Software Studies: A Lexicon* (pp. 15–20). Cambridge, MA: MIT Press.
- Harper, T. (2017). The big data public and its problems: Big data and the structural transformation of the public sphere. *New Media & Society, 19*(9), 1424–1439.
- Harcup, T., & O’Neill, D. (2017). What is news? News values revisited (again). *Journalism Studies, 18*(12), 1470–1488.
- Just, N., & Latzer, M. (2017). Governance by algorithms : reality construction by algorithmic selection on the Internet. *Media, Culture & Society, 39*(2), 238–258.
- Kitchin, R. (2017). Thinking critically about and researching algorithms. *Information, Communication & Society, 20*(1), 14–29.
- Klinger, U., & Svensson, J. (2018). The end of media logics? On algorithms and agency. *New Media & Society, 20*(12), 4653–4670.
- Kowalski, R. (1979). Algorithm = logic + control. *Communications of the ACM, 22*(7), 424–436.
- Lash, S. (2007). Power after Hegemony: Cultural Studies in Mutation? *Theory, Culture & Society, 24*(3), 55–78.
- Lewin, K. (1947). Frontiers in Group Dynamics. *Human Relations, 1*(1), 5–41.
- McQuail, D. (1994). *Mass Communication Theory: An Introduction* (3rd Editio). SAGE Publications Ltd.
- Napoli, P. M. (2015). Social media and the public interest: Governance of news platforms in the realm of individual and algorithmic gatekeepers. *Telecommunications Policy, 39*(9), 751–760.
- Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2017). *Reuters Institute Digital News Report 2017*. Retrieved from [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital News Report 2017 web_0.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/Digital%20News%20Report%202017%20web_0.pdf).

- Neyland, D., & Möllers, N. (2017). Algorithmic IF ... THEN rules and the conditions and consequences of power. *Information, Communication & Society*, 20(1), 45–62.
- Pariser, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You*. New York, NY: the Penguin Press.
- Pearson, G. D. H., & Kosicki, G. M. (2017). How Way-Finding is Challenging Gatekeeping in the Digital Age. *Journalism Studies*, 18(9), 1087–1105.
- Shoemaker, P. J., Eichholz, M., Kim, E., & Wrigley, B. (2001). Individual and Routine Forces in Gatekeeping. *Journalism & Mass Communication Quarterly*, 78(2), 233–246.
- Thorson, K., & Wells, C. (2016). Curated Flows: A Framework for Mapping Media Exposure in the Digital Age. *Communication Theory*, 26(3), 309–328.
- Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colorado Technology Law Journal*, 13(2), 203–217.
- Wallace, J. (2018). Modelling Contemporary Gatekeeping. *Digital Journalism*, 6(3), 274–293.
- Weibo (2018). *Annual Reports 2017*. Retrieved from <http://ir.weibo.com/static-files/1eae99aa-a5e0-4a2b-99de-c60d7d671427>.
- White, D. M. (1950). The “Gate Keeper” : A Case Study in the Selection of News. *Journalism Bulletin*, 27(4), 383–390.
- Williams, R. (1958). *Culture and Society, 1780–1950*. New York, NY: Columbia University Press.
- Willson, M. (2017). Algorithms (and the) everyday. *Information, Communication & Society*, 20(1), 137–150.