

嵌入算法价值观:新闻生产算法系统的 价值敏感设计研究

张 超

[摘 要] “算法转向”让新闻业面临一个现实问题,即如何将新闻价值观嵌入算法,让算法系统的价值观与媒体的价值观匹配。本文从技术哲学的“经验转向”和批判算法研究出发,探讨算法何以有价值观,并将价值敏感设计与“情境分析-价值观识别-技术转译-原型设计”的算法设计流程结合,提出将新闻价值观嵌入算法的具体机制,并对价值敏感设计进行省思,指出价值敏感设计嵌入的是有限度的价值观,价值敏感设计成功的关键在于利益相关者的协同参与,并提出透明性、可解释、可审计的价值敏感设计伦理原则。

[关键词] 经验转向;批判算法研究;价值敏感设计;算法审计

智媒时代来临,新闻生产正发生“算法转向”(algorithmic turn),算法日渐深入地成为新闻业的底层基础设施,驱动新闻生产的技术迭代、流程再造、模式转型和观念革新。由于“算法黑箱”“算法知沟”和媒体技术力量薄弱等问题的存在,新闻业面临算法控制权、新闻生产权、价值观界定权让渡给算法研发者的风险,这些风险将消解新闻业的专业性和合法性。

新闻实践是一种以既定价值体系为基础的实践,其中包括新闻价值及社会责任。^①当算法嵌入新闻生产成为“算法代理者”(algorithmic agent)时,如何让算法体现所属媒体价值观成为一个问题。面对平台推荐算法“流量至上”的价值观,业界和学界提出了“党媒算法”“主流算法”的理念,要求将主流媒体的价值观嵌入算法。在国家层面,2021年9月印发的《关于加强互联网信息服务算法综合治理的指导意见》指出要树立算法正确导向;2022年3月施行的《互联网信息服务算法推荐管理规定》要求算法推荐服务的提供者应当坚持主流价值导向。在国外,公营媒体明确提出了“公共服务算法”(public service algorithm)的理念,BBC要求机器学习技术的开发要反映BBC的价值观。^②

然而,如何将新闻价值观有效嵌入算法设计却鲜有研究。技术圈层关心的是工程层面的效率和技术优化,而非价值观嵌入或伦理问题;学界和业界关心的是算法应有价值观以及应该体现价值观。

张超,新闻传播学博士,山东大学文化传播学院教授、博士生导师(威海264209)。本文系国家社科基金项目“智媒时代新闻生产算法风险及其协同治理研究”(19BXW020)阶段性成果。

①S. Doherty & P. Worthy, “Exploring journalistic values through design: A student perspective”, OZCHI'17: Proceedings of the 29th Australian Conference on Computer-Human Interaction, 2017, <https://doi.org/10.1145/3152771.3156140>, 2023-02-01.

②G. Straub, “Scaling responsible machine learning at the BBC”, <https://www.bbc.co.uk/blogs/internet/entries/4a31d36d-fd0c-4401-b464-d249376aafd1>, 2023-02-01.

但是技术圈层与学界、业界缺少对话的基础,对算法技术的理解不同。如何弥合这个鸿沟?算法价值观何以生成,又何以嵌入?本文运用技术哲学的“经验转向”(empirical turn)和批判算法研究(critical algorithm studies)解答以上问题。

自20世纪80年代以来,技术哲学出现了“经验转向”。经验转向要求把关于技术及其效果的哲学分析建立在对技术的充分的经验描述之上,技术哲学不能只有规范性、评价性的研究,也必须有经验性、描述性的研究,前者以后者为基础。^①当前技术哲学的一个重大改变是从外在进路(external approach)转到内在进路(internal approach),因此,“经验转向”中的“经验”指的是工程师眼中技术研发的经验,而不是哲学家眼中技术使用的经验;“转向”指的是从技术哲学的“外在进路”到“内在进路”的转变。

批判算法研究是围绕算法展开批判的研究领域。其中,算法内嵌的价值观(embedded values)与偏见、算法意识形态等是重要的研究议题。^②技术哲学的“经验转向”对批判算法研究也有借鉴意义。人文社科学者主要关注的不是算法或其底层系统的技术细节,而是算法系统可能具有的意义和含义。^③通过“经验转向”的内在进路,可以让批判算法研究更具有经验性,提供更多现实和个案证据。

当前学界对算法“何以有价值观”这个问题的探讨并不深入,很多研究的预设是算法有价值观,实际上,技术圈层大部分人不认为算法有价值观,这一点从一些学者的前期调查研究和笔者对程序员的深度访谈中可以得到印证。学界和业界需要改变从技术整体论的视角看待算法和算法设计,应当从技术研发经验去理解和探讨算法价值观,如此才能促进技术圈层与学界、业界的相互理解。本文的贡献之一在于从技术哲学“经验转向”的“内在进路”探讨算法价值观何以生成这一问题。

近年来国内外有一些研究提出运用价值敏感设计(value sensitive design)的方法设计新闻生产算法系统,价值敏感设计被整合到现有技术人工物设计研发的实践受到多重因素的影响,^④其如何与新闻生产算法系统进行具体结合则鲜有研究,本文的贡献之二在于提出新闻生产算法系统的价值敏感设计机制,深化学界和业界对价值观如何嵌入算法的认识,为日后研究、设计符合新闻媒体的算法价值观提供借鉴。

一、算法价值观生成的内在路径

在计算机科学中,算法是抽象的解题步骤,是解决问题的基础。当我们在谈论算法时,实际上经常是algorithms(复数),而不是algorithm(单数),因为单个算法固然存在,但在一个系统、一个平台中,为了解决复杂的问题,所谓的算法其实是算法的集合——算法系统。

算法是否有价值观?自2017年人民网三评算法推荐之后,算法价值观问题引发了新闻业界和学界的广泛关注,而“算法是否有价值观”争论的根本原因在于不同利益相关者(Stakeholders)理解算法的视角不同。

(一) 理解算法价值观的两种视角

技术圈层倾向于从技术工具论的视角理解算法,“作为工具的算法”将算法视为“中立”的工具。

①程海东:《实践语境中的技术认识研究》,北京:中国社会科学出版社,2020年,第1—2页。

②方师师:《双强寡头平台新闻推荐算法机制研究》,《传播与社会学刊》2018年第1期。

③T. Bucher,“The multiplicity of algorithms”,<https://doi.org/10.1093/oso/9780190493028.003.0002>,2023-02-01.

④刘瑞琳、陈凡:《技术设计的创新方法与伦理考量——弗里德曼的价值敏感设计方法论述评》,《东北大学学报》(社会科学版)2014年第3期。

技术工具论主张技术是服务于使用者目的的工具,技术没有自身的价值内容。^①按这种逻辑推演,算法也属于技术,因而也没有价值观。技术工具论用这些论断遮蔽了算法的价值观,将算法视为自然化、透明化、不带有任何意向性的工具,算法设计是一个纯粹的技术问题,与社会、文化、价值观没有关系。因此,很多技术公司和平台以此为“避风港”逃避算法的社会责任和监管责任。

“作为人工物的算法”将算法视为体现人的价值观的人工物,这是学界所持的主要立场。人工物是人以自己的意志、知识、能力和价值观,运用技术和技术手段通过生产劳动作用于自然或人工自然而产生的,满足人或社会需要的第二自然物,是构成人工自然的细胞。^②技术人工物具有技术功能、物理结构、设计情景和使用情景的“四重性”,属于事实与价值的统一产物。^③

从这个角度理解算法,算法是包含人的价值意图的。算法不可避免地以社会世界的愿景为模型,并受到商业或其他利益和议程的影响。算法不仅在社会环境中产生,也是社会世界不可分割的一部分,它们被编织进实践和结果,这些结果通过模型回到算法设计中,形成递归过程。^④

(二) 算法价值观何以生成

算法为何会有价值观?如果以人工物的视角进行三段论推理,自然可以得出这个结论:人工物有价值观,算法是人工物,所以算法也有价值观。但这种推理是逻辑上的,不足以深刻揭示和解释“算法何以有价值观”这一问题。

人文传统的技术研究把技术作为一个整体,站在技术之外探寻技术与人类世界中其他非技术方面之间的关系。^⑤对于算法何以有价值观的回应,学界主要是从外在路径:算法是社会性的,因为算法在设计和学习阶段都受到人类价值结构的影响,并且产生了价值和道德后果。^⑥

安德鲁·芬伯格(Andrew Feenberg)认为,应当从一种生成论的观点去审视技术内部,分析技术的内部机制和生成过程,才能全面把握技术的本质。^⑦对于算法价值观的研究,也应从“生成论”的角度去进行探讨。这与技术哲学的“经验转向”不谋而合。“经验转向”要求把关于技术及其效果的哲学分析建立在对技术的充分的经验描述之上,技术哲学不能只有规范性、评价性的研究,也必须有经验性、描述性的研究,前者以后者为基础。^⑧

算法价值观生成的根本原因在于“算法形式主义”(algorithmic formalism)。形式主义是算法的核心要义。形式主义让算法解决问题的思路变得清晰,让复杂的问题易于处理。^⑨

在算法设计中,算法价值观体现在理解问题、训练数据和算法设计中。

在理解问题上,算法形式主义本身是识别问题、理解问题和解决问题的框架。例如,平台媒体算法强调精准推送,但如何理解精准推送?如果用语言表达,精准推送就是满足用户的个性化需求,但在算法形式主义看来,这种界定和阐释无法量化,也不能直接转化成计算问题。算法需要对输入、输出和目标有明确的数学表达,因此用户的个性化需求便可能被理解为向用户提供他们喜欢内容,喜欢的界定标准就是该用户或同类用户点击量多或停留时间长的内容。

①[美]安德鲁·芬伯格:《技术批判理论》,韩连庆、曹观法译,北京:北京大学出版社,2005年,第4页。

②王德伟:《试论人工物的基本概念》,《自然辩证法研究》2003年第5期。

③郭延龙、汤书昆:《技术治理中技术人工物设计的底层属性探析》,《科技管理研究》2020年第20期。

④[英]大卫·比尔,张韵译:《算法的社会权力》,石力月主编:《传播、技术与社会研究读本》,上海:上海交通大学出版社,2020年,第170页。

⑤程海东:《实践语境中的技术认识研究》,第59页。

⑥陈文胜:《嵌入与引领:智能算法时代的主流价值观构建》,《学术界》2021年第3期。

⑦于骥鸣:《后现代网络技术哲学思想研究》,武汉:华中科技大学出版社,2019年,第57页。

⑧程海东:《实践语境中的技术认识研究》,第1—2页。

⑨B. Green & S. Viljoen, “Algorithmic realism: Expanding the boundaries of algorithmic thought”, FAT * 20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 2020, <https://doi.org/10.1145/3351095.3372840>, 2022-12-05.

算法的运行需要训练数据。数据的本质是对现实世界的“表征”而非“镜像”。数据本身也是蕴含价值观的。选择什么数据、如何界定标准,都涉及价值判断。例如,由于中央广播电视总台和互联网公司业务场景的差异,总台在引入算法时不会直接使用,在应用前会用总台数据进行针对性训练。^①实际上,数据的价值观更加隐蔽,如果了解数据的采集语境和方法,数据蕴含的价值观便可以被揭示。

在算法设计阶段,算法的本质是一种看待、解决问题的方法论。如何看待和解决就设计价值判断的问题,自然也涉及价值观。针对同样的问题,算法设计路径有多种,但由于技术的局限,不可能同时满足多个要求,因此就存在不同的优先级,这就导致某个或某几个价值观被倾向于优先考虑。另外,技术圈层所说的一些“优化”问题,看似是单纯的技术问题,不涉及价值观,但是由于解决问题是目标导向的,最后的结果便是客观上形成了价值观。例如,“以用户体验为中心”的算法推荐逻辑实际上是“流量至上”的商业逻辑和垄断逻辑,是鼓励低俗内容快速传播的价值观,社交平台假新闻屡禁不止的关键原因在于假新闻的传播与算法设计中的激励机制相一致。^②

此外,算法设计中长期形成的惯例和解决思路会被技术圈层视为价值观的“挡箭牌”,事实上由于脱离了当初的设计情境,这些惯例和解决思路被自然化了而已,本身就存在价值观。正如美国计算机科学家葛来迪·布奇(Grady Booch)所言:“我们写的每一行代码都有道德和伦理上的含义。”^③

在实际业务环节中,某算法技术公司主管表示,很多媒体在引进算法技术时并不能找到契合自己媒体价值观的算法,一方面是根据媒体需要设计的定制算法研发成本较高,很多媒体的算法其实并非定制;另一方面是模块化的算法系统较有市场,而模块化的算法并不属于定制,因此有主流媒体抱怨推荐系统不行。^④算法研发者研发的是通行版本,而非针对特定媒体的个性化版本,导致算法价值观与媒体自身的价值观并不匹配,影响媒体运营和品牌形象。

二、新闻生产算法系统的价值敏感设计机制

新闻生产是具有意识形态和舆论导向属性的,并不是一个自动化的中立过程。当算法这种自动化技术引入新闻生产中时,算法就成了所属媒体的代理者,因而必须赋予其与人工新闻生产同样的价值观,确保自动化生产契合媒体自身的定位。

新闻生产算法系统的本质是将新闻工作自动化。将新闻工作视为一个可以自动化的过程意味着解构记者的工作方式,尽管新闻工作不能仅仅被简化为需要完成的任务总和。除了是一种职业活动外,新闻还包括但不限于一种职业文化和意识形态,包括共同的价值观、战略和规范。这些方面是基于一套专业实践中的信念系统。^⑤如何将算法价值观嵌入新闻生产算法系统中,让其充分体现新闻业对技术的掌控?价值敏感设计是可行的思路。

(一) 价值敏感设计

价值敏感设计是一种基于理论的技术设计方法,在整个设计过程中,以原则性和综合性的方式

^①靳巾:《人工智能算法在融合媒体业务中的实践应用》,《广播与电视技术》2021年第8期。

^②E. Bell & T. Owen, “The platform press: How Silicon Valley reengineered journalism”, https://www.cjr.org/tow_center_reports/platform-press-how-silicon-valley-reengineered-journalism.php, 2022-12-05.

^③J. Whittle, M. A. Ferrario & W. Simm, et al., “A case for human values in software engineering”, *IEEE Software*, Vol. 38, No. 1, 2021, pp. 106-113.

^④访谈时间,2022年12月5日。

^⑤L. Dierickx, “News automation, materialities, and the remix of an editorial process”, <https://doi.org/10.1177/14648849211023872>, 2022-12-05.

考虑人类的价值。^① 价值敏感设计一词最早出现在 20 世纪 90 年代中期巴蒂亚·弗里德曼 (Batya Friedman) 和彼得·卡恩 (Peter Kahn) 的一篇文章中,试图将人的价值和道德应用于计算机系统的设计中。^②

价值敏感设计的主要观点包括:(1)价值在技术中是可以表达和嵌入的;(2)被一些人决定的价值和选择会实实在在影响他人;(3)有意识的、明确的思考赋予我们发明物的价值观,这在道德上意义重大;(4)在技术设计和开发过程的早期,在价值考虑因素仍然可以发挥作用的时候,就需要阐明道德考虑。^③

在具体实施时,价值敏感设计分为概念调查 (conceptual investigations)、经验调查 (empirical investigations) 和技术调查 (technical investigations) 三个部分。

概念调查主要是设定研究问题的背景、识别价值及价值冲突、识别直接利益相关者 (direct stakeholders) 和间接利益相关者 (indirect stakeholders),特点是对价值进行哲学性探索、重视间接利益相关者的价值、关注潜在的价值冲突;经验调查主要是识别利益相关者对互动语境下个人价值的认知、识别预期和实际操作差异,特点是关注实际场景、关注利益相关者的认知;技术调查主要是对现有技术进行回顾性分析、对新技术进行前瞻性分析、识别技术如何阻碍或支持某些价值,特点是关注技术本身、将价值映射到技术中。^④

(二) 新闻生产算法系统的价值敏感设计

价值敏感设计作为一种设计方法,如何与新闻生产算法系统的设计相结合,国内外还没有相关研究进行专门探讨。笔者借鉴史蒂文·乌姆布雷洛 (Steven Umbrello) 和伊本·范·德·普尔 (Ibo van de Poel) 提出的人工智能技术价值敏感设计流程 (VSD design process for AI technologies),提出新闻生产算法系统价值敏感设计的双轮驱动模式 (见图 1)。



图 1 新闻生产算法系统价值敏感设计的双轮驱动模式^⑤

①B. Friedman, P. H. Kahn, Jr. & A. Borning, “Value sensitive design: Theory and methods”, *UW CSE Technical Report 02 - 12 - 01*, 2002, <https://faculty.washington.edu/pkahn/articles/vsd-theory-methods-tr.pdf>, 2022 - 12 - 06.

②陈凡、孔璐:《价值敏感设计方法新探:能力敏感设计》,《世界哲学》2022 年第 6 期。

③J. van den Hoven, “Ethics for the digital age: Where are the moral specs? Value sensitive design and responsible innovation”, in H. Werthner & F. van Hannes (eds.), *Informatics in the Future: Proceedings of the 11th European Computer Science Summit (ECSS 2015)*, Gewerbestrasse: Springer Nature, 2017, pp. 69 - 70.

④古天龙、马露、李龙、闫茹:《符合伦理的人工智能应用的价值敏感设计:现状与展望》,《智能系统学报》2022 年第 1 期。

⑤笔者在 Steven Umbrello 和 Ibo van de Poel 人工智能技术价值敏感设计流程图基础上进行了修改。S. Umbrello & I. van de Poel, “Mapping value sensitive design onto AI for social good principles”, *AI and Ethics*, Vol. 1, No. 3, 2021, pp. 283 - 296.

价值敏感设计主要有情境分析、价值观识别、技术转译、原型开发四个阶段,在实际设计中,这个流程并不一定是线性的、单向的,而是循环的、迭代的。

1. 情境分析

情境分析(context analysis)是对用户将来使用目标系统解决某个具体问题的方法和结果进行分析。^① 价值敏感设计承认技术设计可以从离散技术本身开始,也可以从使用的情境开始,或者从某个特定的价值观开始,情境分析是至关重要的。^②

情境分析主要运用的方法是经验调查,包括社会科学研究中的定量和定性方法,如观察、访谈、调查、实验操作、文件收集、用户行为的测量等。^③

首先,算法设计应置于广泛的文化和背景中,不能孤立地将算法视为设计对象。^④ 情境分析首先要分析的就是算法所处的社会情境、使用情境。例如,在中国语境下,新闻生产算法系统需要传播主流价值观,避免低俗、媚俗和不良信息传播等;在西方实行公营体制的国家,公营媒体要向社会提供多样化的信息、观点和讨论等。

其次,要确定利益相关者,并与之进行合作,在算法早期创建中发现利益相关者的价值观。^⑤ 利益相关者是与某一新闻生产算法系统相关的利益者,分为直接利益相关者和间接利益相关者:直接利益相关者是指直接与计算机系统或其输出进行交互的各方(个人或组织);间接利益相关者是指受系统使用影响的所有其他各方。^⑥ 直接利益相关者包括算法监管者、算法设计研发者、算法所有者、算法使用者(记者、编辑等媒体内部人员)、算法使用者(该媒体的用户);间接利益相关者是社会潜在群体。价值敏感设计将间接利益相关者纳入其中,这与一些其他算法设计方法明显不同,因为在算法设计过程中,间接利益相关者通常会被忽略。

再次,要确定算法解决的问题是什么。例如,计算新闻发现需要解决“具有新闻价值的选题”;新闻推荐算法要解决“推送主流价值内容”“增强用户体验”等问题。

最后,要调查现有的针对拟解决问题的算法技术。一般来说,解决同一个问题,可能有多种算法解决方案,而最终“胜出”(占主流)的算法模型是技术因素与非技术因素(社会因素)互动演化的结果。^⑦ 因此,有必要审视当前业已存在的算法技术。例如,在不同算法价值观的指导下,个性化算法推荐出现了“个性化的平台逻辑”和“个性化的新闻逻辑”两种明显不同的个性化推荐技术。“个性化的平台逻辑”特点是用户数据海量,对算法推荐的任何编辑控制或监督的强烈抵制;“个性化的新闻逻辑”特点是用户数据有限,强大的编辑控制和专业文化。^⑧

研究显示,精通技术的专家不太关心人工智能技术创新的负面后果。^⑨ 在情境分析这个阶段,需要算法研发者具备道德想象力(moral imagination)。道德想象力概念最早由美国著名哲学家约翰·杜威(John Dewey)在20世纪20年代提出,初始性内涵是指移情投射和创造性发掘情境中的种

①张海藩、牟永敏:《软件工程导论(第6版)》,北京:清华大学出版社,2013年,第58页。

②S. Umbrello & I. van de Poel,“Mapping value sensitive design onto AI for social good principles”.

③古天龙、马露、李龙、闫茹:《符合伦理的人工智能应用的价值敏感设计:现状与展望》。

④M. Bastian, N. Helberger & M. Makhortykh,“Safeguarding the journalistic DNA: Attitudes towards the role of professional values in algorithmic news recommender designs”,*Digital Journalism*, Vol. 9, No. 6, 2021, pp. 835 – 863.

⑤H. Zhu, B. Yu & A. Halfaker, et al.,“Value-sensitive algorithm design: Method, case study, and lessons”,*Proceedings of the ACM on Human-Computer Interaction* 2, <https://doi.org/10.1145/3274463>, 2022 – 12 – 05.

⑥B. Friedman, P. H. Kahn, Jr. & A. Borning,“Value sensitive design: Theory and methods”.

⑦贾开:《算法社会的技术内涵、演化过程与治理创新》,《探索》2022年第2期。

⑧B. Bodó,“Selling news to audiences: A qualitative inquiry into the emerging logics of algorithmic news personalization in European quality news media”,*Digital Journalism*, Vol. 7, No. 8, 2019, pp. 1054 – 1075.

⑨C. Porlezza & G. Ferri,“The missing piece: Ethics and the ontological boundaries of automated journalism”,<https://isoj.org/research/the-missing-piece-ethics-and-the-ontological-boundaries-of-automated-journalism/>, 2022 – 12 – 01.

种可能性。^① 道德想象力的内涵包括:通过情感投射“设身处地”为情境所牵涉的每个人的处境着想;洞察情境中所有可采取的行为方式和行为倾向,并尝试对其未来行为结果进行富有远见的预示;当道德困境处于一筹莫展或非此即彼的境地时,仍继续寻求新的行为选择可能性。^②

算法研发者要借助道德想象力理解新闻生产算法系统中利益相关者的处境、在乎的价值观以及存在的价值观冲突,而非单纯从字面去理解。另外,在设计时需要理解利益相关者所确定的价值观所处的各种情境,提出针对性的技术解决方案。

2. 价值观识别

价值观识别(value identification)是识别并确认嵌入算法的价值观。价值观识别运用的方法是经验调查和概念调查。概念调查主要涉及对价值观内涵的界定与处理,如对利益相关者的价值观进行文献上的梳理,深入理解这些价值观的内涵。概念调查也可以用社会科学方法,如半结构化访谈、调查、观察、实验设计、探索性调查和纵向案例研究等。^③

算法价值观的来源有三个:设计提倡的价值观、设计所尊重的价值观,以及除以上两个来源之外的特定语境的价值观。^④

在价值观识别的环节中,需要思考的核心问题是将哪些价值观嵌入到算法系统中。由于算法系统涉及多个利益相关者,不同利益相关者的诉求也存在差异,在价值识别中需要考虑不同利益相关者的价值共识与分歧。芬伯格认为,技术的形成和特定的利益相一致,处于技术网络中的人都有一定的利益诉求,但利益不可能针对整个社会群体,在满足一部分群体的利益时,必然有另一部分人的利益被忽视或被压抑。^⑤ 算法设计应平衡多个利益相关者的需求、动机和利益,并有助于实现重要的集体目标。^⑥ 某平台算法工程师表示:“AI 技术系统需要满足不同利益方的诉求,随着平台发展阶段动态调整,算法扮演的角色就是综合最大化各方的利益,不断取得平衡”。^⑦

价值观识别会面临两个抉择:价值观相同和价值观相异。无论是相同还是相异,在该阶段需要用概念调查来确定价值观的真正内涵是什么。因为即便利益相关者的价值观看起来是一致的,但具体内涵可能并不一致。例如,在新闻推荐算法中,多样性(diversity)很重要,多样性究竟是什么意思,哪种程度的多样性才算足够?在价值观相异方面,需要思考算法背后的商业逻辑(为用户提供他们喜欢的相关内容)和试图说服公众参与偶然新闻消费(serendipitous news consumption,即面对和用户相反的观点)在多大程度上是可以调和的。^⑧

确定价值观的最好方法是利益相关者互相理解、协商。BBC 在确定推荐算法的价值观方面,约有 80 名编辑人员参与了讨论,约有 40 名设计与工程人员观察了这些讨论,编辑人员和数据科学家通力合作。^⑨ 如果协商存在问题,运用辩证思维进行利弊大小的比较和价值选择。^⑩ 在涉及价值观排序时,还可以采用价值级序策略。价值级序是指价值在相互的关系中具有一个“级序”,根据这个级

①杨慧民、陈锦萍:《网络意见领袖道德想象力:内涵、特性及其价值》,《学术界》2020 年第 6 期。

②杨慧民、王前:《道德想象力:含义、价值与培育途径》,《哲学研究》2014 年第 5 期。

③古天龙、马露、李龙、闫茹:《符合伦理的人工智能应用的价值敏感设计:现状与展望》。

④S. Umbrello & I. van de Poel, “Mapping value sensitive design onto AI for social good principles”.

⑤于骥鸣:《后现代网络技术哲学思想研究》,武汉:华中科技大学出版社,2019 年,第 64 页。

⑥H. Zhu, B. Yu & A. Halfaker, et al., “Value-Sensitive Algorithm Design: Method, Case Study, and Lessons”.

⑦赵璐:《算法实践的社会建构——以某信息分发平台为例》,《社会学研究》2022 年第 4 期。

⑧J. Hendrickx, A. Smets & P. Ballon, “News recommender systems and news diversity, two of a kind? A Case Study from a Small Media Market”, *Media*, Vol. 2, No. 3, 2021, pp. 515 – 528.

⑨A. McGovern, “Understanding public service curation: What do ‘good’ recommendations look like?”, <https://www.bbc.co.uk/blogs/internet/entries/887fd87e-1da7-45f3-9dc7-ce5956b790d2>, 2022 – 12 – 01.

⑩杨小丽、赵宝军:《智能时代的价值冲突及其化解路径》,《河海大学学报》(哲学社会科学版)2021 年第 4 期。

序,一个价值要比另一个价值“更高”或者说“更低”,而更高或更低的判定标准不是绝对的,而是相对的。^① 如此,各个价值观的排序便可一目了然,有助于算法研发人员确定技术方案。

3. 技术转译

技术转译(technical translation)是将识别后的价值观转化为设计要求,从技术上“嵌入”价值观的问题。之所以用“转译”一词,意味着价值观与技术的实现不一定是镜像关系。技术转译很大程度上决定了算法价值观在技术上的实现程度和效果。

技术转译环节主要运用的方法是技术调查,技术调查涉及现有的技术属性如何支持或阻碍特定的人类价值,还涉及系统的前瞻性设计,以支持和实现概念调查中确定的价值观。^② 在一些问题的解决上,技术领域可能会形成通行的规则和方法,形成某种技术经验。这些技术经验实际上依循较为固定的思维方式。因而有必要对这些技术常规进行反思,尤其是洞察技术常规支持或阻碍了哪些价值观,是否有新的技术可以解决等。

与概念调查和经验调查将重点放在利益相关者及其所关注的价值观不同,技术调查更具体地指向技术的设计。因此,技术转译是算法价值观真正落地的环节。

技术转译的过程经历了两次“从抽象到具体”和一次“从具体到抽象”的转译过程。首先,将价值观的抽象内涵转译为具体的规范,之后将具体的规范运用算法思维转译为技术上的抽象问题,最后根据技术上的抽象问题将具体的技术任务分解,转译成具体的设计要求(见图2)。

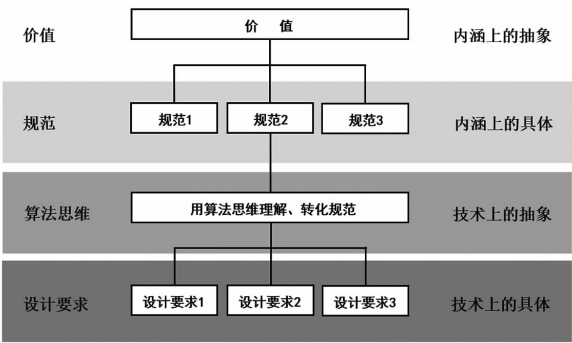


图2 技术转译的过程^③

技术转译过程是比较复杂的,因为“从价值到规范”的第一次转译是语言、哲学问题,“从规范到算法思维”的第二次转译是算法形式主义问题(将现实问题转化为数学问题),也是最难的转译环节。由于算法形式主义的出现,只有那些在算法语言中可读的考虑因素才被认为是重要的设计和评估考虑因素。^④ 因而这个过程并不是镜像的过程,而是映射关系,算法设计者需要理解利益相关者的诉求,并从数学意义上加以解决。当然,在这个过程中也需要利益相关者尽可能将价值观的表述明确、可量化,符合算法形式主义的逻辑。例如,BBC 新闻实验室的产品经理意识到,最具挑战性的部分是阐明这些规则的非常具体的内容,尽管这些标准通常在新闻风格指南或其他相关指南中都有很好的

①[德] 马克斯·舍勒:《伦理学中的形式主义与质料的价值伦理学》,倪梁康译,北京:商务印书馆,2011年,第146—148页。
②J. Simon, P. Wong & G. Rieder, “Algorithmic bias and the value sensitive design approach”, *Internet Policy Review*, Vol. 9, No. 4, 2020, <https://doi.org/10.14763/2020.4.1534>, 2022-12-01.
③笔者在 Evgeni Aizenberg 和 Jeroen Van Den Hoven 价值层级图的基础上进行了改进。E. Aizenberg & J. van Den Hoven, “Designing for human rights in AI”, *Big Data & Society*, Vol. 7, No. 2, 2020, <https://doi.org/10.1177/2053951720949566>, 2022-12-01.
④B. Green & S. Viljoen, “Algorithmic realism: Expanding the boundaries of algorithmic thought”.

描述,但将其编码成程序化格式需要非常详细的说明。^①

《华盛顿邮报》设计的“线索定位器”(Lead Locator)旨在帮助政治记者确定在2020年美国大选中对选民有潜在兴趣的地理区域。如何界定线索是否具有新闻价值?在算法设计时,他们将新闻价值的内涵具体归纳为三个维度:新颖性(novelty)、政治相关性(political relevance)和规模(magnitude)。“新颖性”的技术设计要求被界定为与中心倾向的对比、与已知关系的对比以及与地理环境的对比;“政治相关性”的技术设计要求被界定为一个地点对2020年选举结果的重要性;“规模”的技术设计要求被界定为用县人口作为数量的度量来直接操作。^②

4. 原型开发

原型是提出解决方案的简单实验模型,用于快速和廉价地测试或验证其概念的想法、设计假设和其他方面,以便相关设计师可以在方向上进行适当改进或可能的更改。^③对于算法设计而言,原型开发(prototyping)也被称为beta阶段或beta测试,初始项目在完全开发之前由较小规模的用户进行评估。

算法设计思路不同,开发出的原型也不同。例如,由于对于多样性的理解不同,新闻推荐系统形成了四种模式:(1)自由模式(the liberal model),不仅在内容方面反映公众的利益和偏好,在风格、语言和复杂性方面也反映公众的利益和偏好;(2)参与模式(the participatory model),使人们能够在社会中发挥积极的作用。它重视共同利益而不是个人利益。不同的用户不一定看到相同的文章,但要看到相同的主题,反映了社会上普遍存在的声音;(3)审议模式(the deliberative model),在社会中提出不同的观点和价值观,目的是就不同的价值观达成共识或达成一致,推荐中呈现出多个声音和观点;(4)批判模式(the critical model),强调来自边缘群体的声音。^④

原型开发阶段主要运用的方法是技术调查和经验调查。原型建立后,需要对其进行小规模测试。在测试中,需要对价值观进行内部算法审计。电气与电子工程师协会(IEEE)对审计的界定是对软件产品和流程与相关法规、标准、指南、计划、规范和程序是否一致进行独立评估。^⑤算法审计(algorithmic auditing, algorithm audit)是测试智能机器是否有盲点或偏见的一系列技术的集合,^⑥是一种证明算法存在偏差的研究方法。之所以进行算法审计,很重要的原因在于技术上合理的算法也可能损害利益相关者的价值观。^⑦

算法审计中如果发现原型开发的偏差,就需要进行修正,直至达到可接受的水平。当然,原型开发不是算法设计的终点,任何算法系统如果要运行下去,都需要进行迭代、优化,并随着情境的变化而变化,如此再进入“情境分析-价值观识别-技术转译-原型开发”的流程中。

①S. Danzon-Chambaud, “Automated news at the BBC”, <https://www.bbc.co.uk/rd/publications/automated-news-at-bbc-algorithmic-journalism>, 2023-01-05.

②N. Diakopoulos, M. Dong & L. Bronner, et al., “Generating location-based news leads for national politics reporting”, <https://par.nsf.gov/biblio/10206645-generating-location-based-news-leads-national-politics-reporting>, 2023-01-05.

③R. F. Dam & T. Y. Siang, “Design thinking: Get started with prototyping”, <https://www.interaction-design.org/literature/article/design-thinking-get-started-with-prototyping>, 2023-01-05.

④S. Vrijenhoek, M. Kaya & N. Metoui, et al., “Recommenders with a mission: Assessing diversity in news recommendations”, CHIIR'21: Proceedings of the 2021 Conference on Human Information Interaction and Retrieval, 2021, <https://doi.org/10.1145/3406522.3446019>, 2022-12-05.

⑤J. A. Kroll, J. Huey & S. Barocas, et al., “Accountable algorithms”, https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/, 2022-12-05.

⑥A. Salkever, “Why your company may need an algorithmic audit”, <https://workflow.servicenow.com/it-transformation/algorithmic-audit/>, 2022-12-05.

⑦C. E. Smith, B. Yu & A. Srivastava, et al., “Keeping community in the loop: Understanding Wikipedia stakeholder values for machine learning-based systems”, CHI'20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, 2020, <https://doi.org/10.1145/3313831.3376783>, 2022-12-05.

三、对新闻生产算法系统价值敏感设计的省思

价值敏感设计虽然可以将新闻价值观嵌入算法之中,但应批判看待,不能抱有技术乐观主义态度,否则会陷入“唯技术论”的陷阱中。

(一) 价值敏感设计嵌入的是有限度的价值观

不是所有的新闻价值观都能嵌入算法。理论上可以采用概率论、数理逻辑、形式逻辑、知识表示等数学和人工智能方法描述人类的各种价值、伦理范畴与规则,再采用逻辑程序将伦理范畴组成的伦理规则编写为“道德代码”或者“伦理算法”。^①但事实上这些方法不可能涵盖所有的新闻价值观。给算法嵌入价值观实际上是一种“理想化”的状态,价值观与算法设计不是一一对应的“镜像关系”,而是存在偏差的“映射关系”,甚至有时可能形不成映射关系。在这种情况下,价值敏感设计的应用便遇到局限。

从可以嵌入价值观的层面看,价值敏感设计能嵌入的具体价值观是有限度的。这依然是由于算法形式主义造成的。形式主义意味着坚持规定的形式和规则,算法需要输入、输出和目标的明确数学表达,^②并将解决方案公式化、抽象化和规则化,整个过程是一个转译的过程,而转译意味着偏差,因为模型的抽象和世界的复杂性之间存在鸿沟。^③

例如,新闻多样性(news diversity)这一价值观,如果将多样性从技术上理解为信息源多样性(source diversity,将新闻媒体包含的多个信息源视为一个故事的构建模块)、内容多样性(content diversity,用户通过单一新闻渠道或多个新闻渠道接触到的话题)、观点多样性(viewpoint diversity,新闻媒体在报道某一问题时所包含的视角的多样性)时^④,相关界定虽然在语言表达上是清晰的,但落实到每一个指标时都必须量化,每一个量化方案也很复杂,算法的有穷性(finiteness)决定了量化程度的有限性,所以能嵌入的价值观也是有限度的。

(二) 价值敏感设计成功的关键在于利益相关者的协同参与

价值敏感设计不能以算法研发者为中心,虽然他们在技术上拥有优势,但是他们未必理解具体的情境和价值观的内涵。在研发的时候,主要是靠研发者理解相关问题,但这种理解可能是片面、理想化的,甚至是错误的。在现实中,很少有工作专注于面向记者的算法系统,之前也没有试图从人类价值观的角度为记者提出人工智能系统的要求、设计或评估。^⑤News.me高级设计师表示,应用设计和新闻价值观之间的脱节是由于专业文化没有交叉。^⑥因而在价值敏感设计中,迫切需要一些有算法思维能力的新闻从业者转变成为“元记者”(meta journalists)参与到算法价值观的设计之中,不同利益相关者的有效沟通、理解和分工写作至关重要。《华盛顿邮报》新闻编辑室工程团队与新闻编辑

①莫宏伟、徐立芳:《人工智能伦理导论》,西安:西安电子科技大学出版社,2022年,第176页。

②B. Green & S. Viljoen, “Algorithmic realism: Expanding the boundaries of algorithmic thought”.

③B. Green & S. Viljoen, “Algorithmic realism: Expanding the boundaries of algorithmic thought”.

④R. Evans, D. Jackson & J. Murphy, “Google news and machine gatekeepers: Algorithmic personalisation and news diversity in online news search”, *Digital Journalism*, <https://doi.org/10.1080/21670811.2022.2055596>, 2023-01-20.

⑤T. Komatsu, M. G. Lopez & S. Makri, et al., “AI should embody our values: Investigating journalistic values to inform AI technology design”, *NordiCHI'20: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 2020, <https://doi.org/10.1145/3419249.3420105>, 2022-12-05.

⑥M. Ananny & K. Crawford, “A liminal press: Situating news app designers within a field of networked news production”, *Digital Journalism*, Vol. 3, No. 2, 2015, pp. 192-208.

室的记者和编辑合作,设计了线索定位器。^① 学者林登(Carl-Gustav Linden)建议,数据的选择和评估应该是记者的任务,而验证、标准化和规范化是程序员的任务。^② 如此协同,让利益相关者不再是算法设计的旁观者,而是参与者,让媒体的价值观有效嵌入数据和算法系统中。

(三) 价值敏感设计应遵循设计伦理

新闻生产算法系统的公共性、自主性和合规性决定了在进行价值敏感设计时要遵循设计伦理,涉及的伦理维度主要包括透明性、可解释和可审计。由于新闻生产算法系统中的算法包含两类,即面向媒体自己内部使用的内部算法(如计算新闻发现算法)和面向公众提供服务的外部算法(如个性化推荐算法),导致不同算法价值敏感设计的利益相关者存在较大差异,以上伦理维度在面向不同算法系统类型时所实践的具体对象和内涵略有差异。

1. 透明性原则

透明性(transparency)的本质是“后台前置”,通过披露相关事项接受利益相关者的监督。在价值敏感设计中,透明原则贯穿于语境分析、价值观识别、技术转译和原型设计四个环节中,根据内部算法和外部算法的差异性向对应的利益相关者公开决策过程、设计思路和最终方案。

透明性原则不要求绝对透明,而是相对透明,追求“有意义的透明”(meaningful transparency),即应当把关键问题的所涉事项公开,而非事无巨细地全部公开,淹没关键信息;也非选择性公开,对关键问题避而不谈。是否达到“有意义的透明”,需要根据利益相关者的透明性诉求而定。

2. 可解释原则

可解释(explainability)是指在价值敏感设计环节中,技术方案的思路、原理、实现过程是可以解释清楚的,不能以技术黑箱为由拒绝解释或者无法解释。可解释原则主要存在于技术转译和原型设计环节,具体包括三个维度:可解释的数据(explainable data),阐明用于训练模型的数据,数据的类型和内容,选择它的原因,如何评估公平性,以及是否需要努力消除偏见;可解释的预测(explainable predictions),具体说明所有激活或用于达到某种输出的模型特征;可解释的算法(explainable algorithms),揭示模型所包含的各个层次,并解释它们如何对预测或输出作出贡献。^③ 可解释原则是利益相关者有效理解、讨论、洞察和监督价值观是否有效嵌入算法系统的关键原则。

可解释原则要遵循便于理解原则、相关性原则和层级原则。^④ 便于理解原则是指要技术圈层利益相关者要考虑到非技术圈层利益相关者的算法技术认知水平,以易于用户理解为最低标准;相关性原则是指解释的内容与利益相关者关心的利益相关(如某个价值观是如何有效嵌入的);层级原则是指可解释原则不是只有一个标准、一个层次,而是因利益相关者而异,应当根据利益相关者在价值敏感设计中的地位和关系确定,并根据不同的情境来确定可解释的范围、水平、层次。

3. 可审计原则

价值敏感设计是一个复杂的价值观嵌入过程,受制于算法形式主义的存在和算法系统在真实环境下的运行影响,有可能需要继续迭代,或者为后续追溯责任时进行问责、审计提供线索和依据,因此可审计(auditability)原则势在必行。可审计是指提供可以用于事后审计的相关资料和运行数据,保证技术人员能够审计出价值敏感设计中各环节的疏漏。可审计原则在实际场景中可用于内部审

^①N. Diakopoulos, M. Dong & L. Bronner, et al., “Generating location-based news leads for national politics reporting”.

^②C. Linden, “Decades of automation in the newsroom”, *Digital Journalism*, Vol. 5, No. 2, 2017, pp. 123 – 140.

^③Aporia, “Explainable AI: How it works and why you can’t do AI without it”, <https://www.aporia.com/learn/explainable-ai/explainable-ai/>, 2023 – 02 – 05.

^④徐汉明、孙逸啸:《算法媒体的权力、异化风险与规制框架》,《西安交通大学学报》(社会科学版)2020年第6期。

计,例如,原型设计阶段的算法审计,也可以用于外部审计,例如,利益相关者想要通过第三方审计算法系统是否体现了预期的价值观。可审计的前提是要有足够、充分的全过程资料和数据,因此在价值敏感设计中需要设计方记录各个环节的决策结果、原因和具体方案,对特定问题的局限以及处理方式做出说明,在技术转译环节详细记录算法设计思路和步骤,并对训练数据、算法测试性能等主要技术指标进行详细说明和阐述,确保审计时可以探查出问题根源。

四、结 语

算法如今已经成为重要的基础设施嵌入新闻生产流程中,新闻生产作为渗透着意识形态和价值观的专业实践不应忽视算法价值观给自身运营带来的影响。不同类型、不同体制的媒体有不同的价值观,算法系统也应有相匹配的价值观,能保证人工新闻实践和自动化新闻实践在价值观上的同一性。本文提出运用价值敏感设计的方法给新闻生产算法系统嵌入价值观,一方面揭示算法价值观的嵌入原理,另一方面提醒新闻从业者应主动与研发者进行对话。当然,由于算法形式主义的存在,新闻价值观转译为算法价值观并非理想化的镜像,而是映射。作为人工物的算法只能一定程度上体现新闻价值观,这需要通过技术上的迭代方式进行优化,另一方面新闻生产依然要确保人在环路中(human in the loop)。

(责任编辑:张 升)

Embedding Algorithmic Values: A Study on Value Sensitive Design of Algorithmic Systems for News Production

ZHANG Chao

Abstract: The “algorithmic turn” has confronted journalism with a real problem: how to embed journalistic values into algorithms and match the values of algorithmic systems with those of the media. In this paper, we attempt to explore how algorithms have values from the “empirical turn” in the philosophy of technology and critical algorithm research through combining value-sensitive design with the algorithm design process consisting of “context analysis”, “value identification”, “technical translation”, and “prototype design”. On this basis, we propose a specific mechanism for embedding news values into algorithms, reflects on value-sensitive design, and points out that value sensitive design embeds limited values. We argue that the key to successful value sensitive design lies in the collaborative participation of stakeholders, and the value sensitive design should follow the ethical principles of transparency, explainability and auditability.

Keywords: empirical turn; critical algorithm studies; value sensitive design; algorithm audit

About the author: ZHANG Chao, PhD in Journalism and Communication, is Professor and PhD Supervisor at School of Culture and Communication, Shandong University(Weihai 264209).