

计算人文学视野中的三种人工智能

吴 岩

摘 要：计算人文学视野中至少存在有三种人工智能的话语：历史的、政治的和想象的。历史的人工智能话语具有科技话语的风貌，具有是路径依赖性和外周封闭性。政治的人工智能话语试图通过政治操作建立起人工智能跟未来之间的关系，这种人工智能话语具有目标愿景性、资源承诺性和未来导向性。想象的人工智能话语是愿望性的，有跨领域性和叙事性。计算人文学从厘清人工智能的话语形式开始，通过对不同类型话语的批判，更清晰地感受技术、科学、社会、个体目标之间的复杂作用。通过加强不同话语持有者之间的理解，增进相互的渗透，更好塑造人工智能的社会未来。

关键词：人工智能 话语 历史 政治 想象

任何一个学术领域都深受话语的困扰，人工智能也不例外。计算人文学一个最重要的内容，是梳理话语场中流行的人工智能话语类型，并通过分析话语去探究这一领域中正在发生的到底是些什么？进而，也可以感知过去和推测未来。

从 ALPHA GO 战胜围棋冠军李世石之后，有关人工智能这一主题在各种媒体和社交网络上飞速升温。一些人欢呼人类正在走向一个全新的时代，技术正在超越人脑成为代替人类思考的真正工具。另一些人则对此忧心忡忡，生怕丢失饭碗甚至在未来的某一天丢失性命。人工智能专家也分成许多派别，这些派别对技术发展的路径进行辩论。政府和国际组织强行参与，各自提出了具有法律法规意义的宣言、主张、行动纲领。科幻作家和种种网络意见领袖，也没有放过这个机会阐述

自己的观点。在如此众声喧哗的人工智能言说的杂烩之下，计算人文学必须冷静地寻找一种清晰有效的方法，高屋建瓴地透视现实。而从索绪尔以来，人文和社会学家所进行的扎实的话语分析的实践，是这一过程的最有效的工具。

以话语的特征和来源拆解当今的人工智能话语，会发现三种明显不同的言说范畴，这三个范畴分别是历史、政治和想象。三种范畴共用一个中心词汇，内容上却传差异巨大。

一、人工智能场域中的三种话语

1. 作为历史言说的人工智能

人工智能言说的第一个话语范畴，我称之为历史性范畴，这一领域的人工智能话语包含了从冯·诺依曼、图灵以降所有人工智能、智能机器、定理证明、专家系统等标识之下的逻辑、数学、计算机、脑科学、认知心理学的言说和讨论。人工智能专业学术会议上的讲演，学术刊物上的论文，相关科普读物等，都是历史性人工智能话语范畴之内的言说。真正能理解历史性人工智能话语的人在这个世界上为数不多。即便是相关程序设计的从业者，真正把握到历史性话语精髓的人也是少之又少。更不要提没有逻辑学、算法基础的人，他们完全无法进入这个言说场域。

历史性的人工智能发展到今天，形成了一些重要的言说派别或领域，这其中最为重要的两个派别是深度学习和遗传算法。而自然语言处理、图像识别、专家系统、自动驾驶和机器人等，则成为各个时期最有开发热点的领域。历史性的人工智能有强烈的路径依赖性，是一些历史性的事件（如会议、某项重要投资、某个社会政治力量的强烈关注等）导致了历史路径的形成，它们也是历史性人工智能话语展开的基本线索。每当我们谈到阿尔文·明斯基、赫伯特·西蒙和艾伦·纽埃尔、瓦伦·麦卡洛克和瓦尔特·皮茨、约翰·霍兰德，每当我们谈

到达特摩斯会议、乔治敦实验、深蓝击败卡斯帕罗夫、ALPHI GO 战胜李世石的时候，这种路径便凸显出来。

历史性的人工智能话语中的某些材料，至今仍然被隐藏。但通过某些外部的实证性观察，我们能够确认这些话语的存在。这种人为的黑箱，使得开发者之间竞争激烈，也使得政治性人工智能的话语掌控者能有所作为。通常，历史性人工智能话语在能指和所指之间具有的可靠的匹配性，因为这些内容在行业内部具有共识。对而路径性和匹配关系的熟悉与否，导致了语义在小范围的封闭性。但是，这并不证明每一个言说者的同样类型陈述是一致的，因为这其中也有很强的策略性或游戏性。例如，当我们观察各个学派或流派的著作或类似的言说时，可以发现每一个学派在呈现整体人工智能外观的时候可以有一定的侧重和取舍，反复言说的时候也可以有所删节，而这些做法的意义是突出自己所在的路径的重要性。每一个流派似乎都在用事实证实，自己在创造宇宙。这同时也说明，人工智能即便在历史场域之中，被言说到的宇宙是有所侧重的。至少，这个宇宙是有分区和存在各自边疆的。有些区域，通过前期发展已经空间相当充分，而另一些区域还在发展之中，更有一些区域被明显搁置，像第五代计算机就是一个已经被搁置甚至废黜了的空间。从历时性的侧面观察，历史性人工智能的发展是阶段性的，在不同的时段可能被不同称呼所取代。

虽然上面的陈述已经证明，历史范畴中人工智能话语有很强的经验参与，但从这种言说的外观上看，话语是科学 + 技术型的。数学、计算机科学、逻辑学的假设和原理，以及技术的可能性、可行性、可证明性、可证伪性等，调节着话语的结构。也恰恰是这种历史性话语跟科学话语之间的统一关系，导致了历史性人工智能宇宙疆界的最终划定。

回顾冯·诺依曼的《会思考的机器》，彻底拆解这部作品的内容，你能发现历史性人工智能的话语空间可能到达的高度，和可能实现的广度。在冯·诺依曼看来，人工智能自始至终还是一种方法的、工具的、

准 / 非智能体的、可能没有最终临界但却有明显的指数渐近线的空间之中。在当前正在激烈的创新的方向上，历史性的人工智能话语具有逐渐侵蚀边界、拓展宇宙的能力。但时时回到冯·诺依曼是有价值的、能让我们认清历史性人工智能的所有可能和局限。

2. 作为政治言说的人工智能

最近五年，一些国家的议会议员、政治家、企业家，出现了为人工智能的发展游说或为限制人工智能可能出现的种种危机去加速或者叫停的游说。一些民间组织和群众团体也对大公司发展人工智能提出了自己的观点。多数人认为，人工智能技术的突破和普及，将是人类社会经历的一次巨大的转变。这种转变可能只有语言的发明这样的重大变化才能与之媲美。而语言的产生直接导致了智人的兴起和其他类人猿的消亡。在从相互合作到你死我活的光谱氛围下，联合国、欧盟、美国、日本、中国等各自出台了与人工智能有关的促进法案、管理法案、行动计划、行动纲领。这些内容构成了人工智能第二个重要的话语范畴，暨跟决策和未来相关的政治话语范畴。

从很早起，政治力量就已经介入甚至左右人工智能场域了。各类政府或各类组织通过项目或投资方式，给人工智能发展的整体或者某个方向提供经济支持。在这方面行动较早的，是美国国防部先进技术预研局（DAPPA）。这种类似输血的援助，导致了历史性人工智能在一些接近人财损失或者技术僵局的断裂点上起死重生。

假如历史性范畴中的人工智能都是朝向历史过往方向的言说，那么政治性人工智能则是朝向未来的言说。将这一技术 / 存在纳入群体 / 国族之间竞争的考量，纳入国际国内科技或社会发展的考量之后，政治性人工智能创造了自己的话语言说。这些言说大都以议程、法案、行动纲领等形式表达，也有的出现在政治家的讲演和议会推动之中。

前面已经说过，政治性人工智能话语的言说差异巨大。在 2016 年联合国教科文组织与世界科学知识与技术伦理委员会联合发布的《机

器人伦理初步报告草案》^①中，各自强调了人工智能不能用于独立杀人武器的开发以及人工智能机器的行动都必须在监管之下完成。2017年，国际电信联盟“人工智能造福人类”全球峰会则重点探索加速人工智能发展以便更好应对贫困、饥饿、健康、教育、平等和环境保护等全球性挑战。上述话语跟联合国维护世界和平和关注贫困与平等的宗旨之间具有极大一致性。2018年5月，欧盟委员会决定了增加投资带动公共和私人资本参与、促进教育和培训体系升级适应就业岗位变化和制定人工智能新的道德准则以捍卫欧洲价值观的三大措施。继续维护启蒙以来的个人的高尚地位，保护隐私，强调文化根基在人工智能发展上的价值给人深刻的印象。2016年10月至12月，美国白宫科技政策办公室发布《为人工智能的未来做好准备》^②、《国家人工智能研究和发展战略计划》^③和《人工智能、自动化与经济》^④3个报告。前者提出人工智能应用带来潜在公共政策问题，并提出23条建议。报告特别强调政府的重要作用。第二个报告提出了投资、开发人机协作、解决伦理、法律和社会学影响问题、注重安全可靠、开发培训和测试的公共共享数据集和环境、通过标准和基准以测量和评估、了解人力需求等七个建议。第三个报告提出要投资和开发对社会有益的人工智能、提供教育和培训为未来的工作做好准备、帮助转型过程中的工人并赋予其权利，确保其可以广泛共享由人工智能带来的经济增长收益。同样在2016年，英国科学办公室和议院下院也发布了《人工智能对未来

① United Nations Educational, Scientific and Cultural Organization and World Commission on the Ethics of Scientific Knowledge and Technology, preliminary draft report of commission on robotics ethics, 2016.

② Executive Office of the President National Science and Technology Council Committee on Technology, preparing for the future of artificial intelligence, 2016.

③ National Science and Technology Council Networking and Information Technology Research and Development Subcommittee, “the national artificial intelligence research and development strategic plan”, 2016.

④ Executive Office of the President of the United States of America, “Artificial Intelligence, Automation, and the Economy”, 2016.

决策的机会和影响》^①和《机器人技术和人工智能》^②两个报告,指出人工智能在政府层面大规模使用的潜在可能性、可能的法律与伦理道德问题。2017年,日本政府召开“人工智能技术战略会议”,认为人工智能是第四次产业革命的技术核心。会上还公布产业化路线图,计划分三个阶段推进。2017年7月,国务院关于印发新一代人工智能发展规划的通知,要求认真贯彻执行《新一代人工智能发展规划》^③。该规划从战略势态、总体要求、重点任务、资源配置、保障措施、组织实施等方面,全面阐述了中国人工智能发展到2030年之前的步骤、方向、目标、执行方法等。认为必须“面对新形势新需求,必须主动求变应变,牢牢把握人工智能发展的重大历史机遇,紧扣发展、研判大势、主动谋划、把握方向、抢占先机,引领世界人工智能发展新潮流,服务经济社会发展和支撑国家安全,带动国家竞争力整体跃升和跨越式发展”。

政治性人工智能的话语是一种全新的话语。只有在技术成熟到能够引发政治行为的时候,这种话语才逐渐成为一种社会化的言说形式。政治性人工智能话语没有历史性人工智能话语那种强烈的路径依赖特征,但它特别关注历史性人工智能的当前边界,并通过这些边界做出自己发展判断的同时,将组织、地区、国家的利益放在重要位置。由于政治性人工智能话语伴随着巨大的资源承诺,因此在上述不同类型的话语言说发生作用之后,世界人工智能的未来版图将出现彻底的改变。而各国人工智能的发展方向,也会发生较大的分裂。

政治性人工智能话语是强社会语义的、具有明确所指的、包含充分行动潜力的,通常是社会愿景形式所展现的。以政治考量判定人工智能的发展方向,推进这些方向,是政治性人工智能话语的标志性特

① Government Office of Science of United Kingdom, “Artificial intelligence: opportunities and implications for the future of decision making”, 2016.

② 施羽暇:《英美两国人工智能战略比较研究》,《人民邮电报》,2017年5月18日。

③ 国务院:《新一代人工智能发展规划》,2017年7月8日。

点。分析各国各地区各组织颁布的政治性人工智能法案或报告、方案，可以发现各自相当不同的操作特征和经验特征。由于政治操作可能具有的偏差也会导致这些话语普遍在世界各地的两极性评价暨一些人欢呼一些人诟病，因此，不断研究人们对政治性人工智能话语的反映、观察其他对手的话语内容，对修改和调整自身人工智能发展言说具有重要意义。

在某种意义上，政治性的人工智能话语由于携带着权力和资源，已经成为当前最重要的人工智能话语形式。这种形式联系着科研和大众，联系着历史和未来。但十分遗憾的是，政治性的人工智能话语的言说只是极少数人的事情。我认为这种状态是不符合人工智能发展的。如果没有更多社会力量的介入，许多由于思考不全造成的漏洞，令人十分担忧。

3. 作为愿望言说的人工智能

人工智能话语的第三个范畴，是想象性言说。这个言说方式的存在，远远早于历史性和政治性人工智能话语。想象性人工智能话语的产生是在追问人类为什么存在和挑战这种存在、尝试设计通向自由之路的一系列思考过程中产生的，这种话语核心是人类对自身智能的超越和类似事物/过程的重建。

想象性人工智能话语无论是东方还是西方都出现得很早。叙事性是它常常具有的特点。许多神话、传说、民间故事、小说、美术、童话、寓言中包含有这种人工智能的话语。希腊神话中的铁匠赫菲斯托斯制造机器人仆人的故事，就这类言说的一个典型案例。赫菲斯托斯是希腊神话中的主神之一，赫拉之子。他一条腿行动不灵，一出生就被厌弃。但恰恰是这个身体残疾、家庭关系不好的铁匠创造了具有智能驱动的机器人仆从。在希腊神话的语境中，故事充满了道德和社会文化方面的功能代偿，也充分体现了西方金属文化的技术路径。同样是人工智能研发的故事，在列子的《汤问》篇也有一个。这就是《偃师造人》。故事从周穆王西巡途中遇到偃师谈起，把对智能体遭遇产生的震惊、

兴奋、愤怒和行动统一起来。情绪情感是这个叙事的核心，更是技术创造的核心。两个东西方不同的人工智能叙事，都是人类的想象，但差异如此巨大。这在很大程度上展现了作为愿望言说的想象性人工智能的特点。

想象性人工智能有一定的蔓生性，有可能随着历史的发展反复重现，但内容则没有太多逻辑性，语义更加能指化，即可以做许多种不同的解读。随着科技时代的到来，想象性人工智能的话语在很大程度上被科幻文学和科幻艺术所承接。例如，有关智能机器人的词语就是由捷克科幻作家卡莱尔·卡佩克在《洛桑万能机器人公司》中所创造的。美国作家艾萨克·阿西莫夫还在小说《我，机器人》中提出了所谓的机器人学三定律/三大法则。坦尼斯拉夫·莱姆专门写文章讨论了这类人工智能作品局限性的论文。丹·西蒙斯把这种人工智能的未来发展提升到宗教体验层次。

想象的人工智能话语看起来边界宏大，它也确实比历史性的人工智能的路径性话语具有更广的空间，但是，如果把这些内容详细地研究，会发现这种言说的内容范畴其实也并不很大，将人工智能表达为一种人性的、即将超越/永远无法超越情感、灵魂、自我界限的机器造物，是想象性人工智能的通常表达形式。对于那个他者意义上的、跟人类处于非我状态的造物，从一开始就就被预设得是人而不是机器。自身的超越和回归是想象性人工智能话语的三个主要母题。

由于族群目标、政治操作只是想象性人工智能话语一小部分内容，因此两者之间的重叠并不很大。而且，虽然愿望强烈，但想象性人工智能的话语缺乏未来的直接推动性。当然，在更广大的时间范围内，想象性的人工智能话语可能推动人投入人工智能历史的创造，投入未来发展的愿景考量，但这些都是间接的。不但如此，为了达到目的想象性人工智能话语还常常会“正话反说”，从不可能暗喻可能。这一点是跟其他两类话语完全不同的。

二、话语的批判和对话创建

无论起源于古希腊还是春秋战国，想象性人工智能话语的出现都远远早于历史性人工智能话语。在过往的数千年里，想象性人工智能话语一直引导着人类不忘初心，寻求超越自身也超越自然的努力。是科技革命中方法的改变颠覆了这个状况。从图灵开始，历史性人工智能话语逐渐抢夺了话语场中人工智能言说的主动权。在当今，随着族群政治、国家政治和国际政治的发展，政治性人工智能话语采用自己凌厉的攻势，超越了历史性人工智能话语更超越了想象性人工智能话语获得了领导力。面对人工智能话语场域中的这种权力关系的此消彼长，一些隐含的问题正在逐渐浮出水面。

首先，政治性话语背后的族群诉求较强，各国在出台这样的报告和规定之前又只在小范围政府渠道或组织渠道内进行讨论，这样的人工智能言说本身就令人疑虑。市场因素对政治性人工智能的话语塑造也起着关键作用。但市场操作本身的问题早在多年之前就已经开始受人指摘。因此在当今，虽然各国出台了人工智能政策，但这些政策是否会反会将人类送入歧途，还未可知。

其次，历史性人工智能正在沿着各自的发展路径快速前行，但科技发展永远是在从业者兴趣、前沿突破的可能性和功利性三者之间综合推进着，而科技背后的道德言说，在这个领域中常常起不到任何作用。此外，历史性人工智能话语本身的内敛性和路径依赖性，导致了这种形式的话语超越自身的难度。

第三，覆盖广泛、内容丰富但无法介入当前话语场的想象性人工智能言说，已经处于全面失语状态。从科技革命之后，人文主义在技术领域中的合法性早已经被剥夺。而且，在人类作为整体已经超越70亿，很难寻找出一种方法收集和放大新的、更有价值的想象性人工智能言说的方法，并将其作用于话语场去改进人工智能的现状。

从上面的几点可以看出，当前风光无限的人工智能话语场域并非一个发展完善的场域。使用更多分析话语、解构话语，寻找内涵中被埋没的信息已经到了十分急迫的状况。此外，建立更多话语交互机制、丰富这个场域中的言说和对话，是另一个重要的工作。

在这里，我想侧重提到发挥想象性人工智能话语价值的问题。在我看来，想象性人工智能话语的最大特征是言说的多向性、模糊性和内容的社会感召力。鉴于这些特点，想象性话语其实提供了人类最广泛的人工智能产业发展目标。想要增进政治性人工智能话语的代表性，增进政治性话语的全面性，应该从更广泛地听取想象性人工智能话语的言说开始。

还以《偃师造人》为例。这篇作品至少提供了四个不同侧面去观察人工智能现象。第一个侧面是技术创生侧面的。在这篇文章中，可以生发出有机材料、仿生学、道家自然观等。作品对人工智能开发中情感的中心位置，也很有启发性。从《偃师造人》中提取出的第二个侧面是社会管理的。故事清楚地告知我们，人工智能问题不是也不可能是纯粹的技术过程，它必定是社会、政治、区域化、行业管理等造就的综合现象。第三个侧面是事业运作的。在这里，政治性人工智能的话语已经介入。因为周穆王的存在本身，就是政治干预技术发展的符号。第四个侧面是哲学/神学的。人工智能的制造到底使人类走向哪里？是通向超人、取得神的地位，还是其他的什么？想象性的人工智能的话语就是通过这一种模糊的故事，揭示了技术存在本身的生态性。大量的相关叙事，提供了人工智能各种发展可能的未来样貌，可以供我们学习、吸纳，并在政治性和历史性人工智能的言说中体现。

这样，我们自然而然地终结在当前人工智能的发展必须进行话语场重建的问题之上。这种重建的目标，是让三种言说都能充分发挥作用，且能从相互了解中学习更多有价值的东西。

话语场的重建说起来容易做起来难。2018年5月19日，南方科技大学人文中心和计算机学院共同召开了“人工智能时代的技术与人

文：跨学科对话”研讨会。会上，一位出身于计算机技术但站在社会管理者立场发言者不无讽刺地说：有的人“昨天还来问什么是人工智能，明天就成世界专家了，后天开班收费了，然后还来跟我认真讨论这个东西”。这句话至少让我发现了人工智能领域中确实存在一种鄙视链。他在另一个地方还曾说过，文学家和哲学家已经搅得人工智能开发者“不知道什么是人工智能”。作为这个会议的参加者我觉得收获巨大。人工智能的话语冲突让我感到，现在是回到科技发展的真正意义、回到交流的真正意义上来的时候了。至少，对历史性和政治性人工智能言说者需要广泛的倾听，而想象的人工智能言说者，应该学会新的言说方式。只有部分地改变自己的叙事性言说方式，才能将人类的广谱的愿望投送到其他两种言说的场域之中并能跟他们交流。这样，一种新的人工智能言说会逐渐产生。这是相互倾听和相互理解的，是寻找启迪和共同发展的。计算人文学对此责无旁贷。

商
学
馆