

菲利克斯·波斯辛格* 和罗伯特·库恩

走向符号金字塔:语言研究、人工智能和知识交换经济

<https://doi.org/10.1515/lass-2023-0016>

2023 年 4 月 25 日收到; 2023 年 7 月 24 日接受

摘要:本文讨论了知识交换经济背景下人工智能与(数字)人文学科语言研究之间的相互关系。我们提出了一个三维符号学模型(金字塔),其中将人工智能作为意义创造的主动代理纳入通信符号系统,并测试其在自动完成和预测短信方面的有效性。因此,我们在 2 年内分析了 3.6 亿条关于 COVID-19 的推文,并重点关注与重大外部事件相关的积极和消极极端情况下语言使用随时间的变化。数据表明,数字语言的使用在大流行期间发生了显著变化。人工智能、话语和外部现实之间创建了反馈循环,通过建议和完成文本条目将先入为主的概念石化为文本等语。因此,我们认为人工智能通过影响措辞的选择巧妙地改变了对现实的感知。为了减轻人工智能在意义创造方面的影响,我们强调了提供商采取的结构性预防措施;然而,我们认为,(数字)人文学科可能会根据它们为语言使用和人工智能提供的特定主题的见解,重新强调它们在知识交换经济中的内在价值。

关键词:人工智能;数字人文;意义创造;符号金字塔

1 引言:数字与人文

当 ChatGPT 于 2022 年 11 月首次引入市场时,它受到了公众和学术界的热烈欢迎,并且仍然保持强劲。


虽然它可能是人工智能(AI)中最受欢迎和最有可能的实例(参见 Bubeck 等人,2023),但它远非当前日常使用的唯一人工智能。

公司投资它是为了有针对性的营销、聊天机器人或自动建议,而

*通讯作者:Felix Poschinger,汉堡大学,汉堡 20148,德国,E-mail: felix.poschinger@uni-hamburg.de。

<https://orcid.org/0009-0008-6616-5795> Robert Coon,独立人士,德国汉堡,电子邮件:

RobertCoon@web.de

 开放访问。© 2023 作者版权所有,由 De Gruyter 出版。
知识共享归属 4.0 国际许可证。

 本作品已获得许可

个人在健身追踪、搜索引擎结果或垃圾邮件过滤器等方面会接触到它。国际研究表明,与人工智能的反复接触导致人们普遍认识到它的存在,但同时缺乏对它到底是什么的理解(参见 Ipsos 2022; Kennedy 等人 2023; Kozyreva 等人 2020))。这种分歧通常比人文学科更明显,人文学科中存在着一种被误导的范式,即“没有理由认为人工智能只不过是人类设计的一组规则,以帮助更有效地执行他们的知识任务”(Danesi和 Matthews 2019:204)仍然基本上没有受到质疑。

下面的论文将讨论人文学科中的人工智能和数字技术问题,认为当代知识转移经济中人文学科的逐渐贬值可能部分是由于在数字化转型中无法重新思考和重新定位自己作为一门学科而造成的。人工智能和数字技术中、与人工智能和数字技术相关的新可能性的出现。

为了举例说明这一总体主题,我们将重点关注与语言使用和人文学科相关的人工智能的一个具体实例:自动完成和预测短信。我们首先将定义术语、基本工作流程及其含义。其次,我们将提出一种源自德索绪尔、奥格登/理查兹和塞尔传统框架的传播理论符号学模型。第三,我们将介绍对预测短信的 Twitter 数据进行统计分析的结果,这些结果支持所提出的理论框架,并说明传统人文、数字人文和当代知识交换经济之间的相互关系。第四,我们将把研究成果置于边缘化和社会影响之间的人文学科背景下。

2 数字人文与知识交流

近年来,人文学科越来越受到批评,因为它是一个以理论为主的领域,学科本身及其在微观和宏观层面上的应用之间有明确的区别(参见 Kirsch 2014; Leroi 2015; Straumsheim 2014)。语言学等一些领域在其自我定义中阐明了这种二分法,并将理论框架的相应应用建立在分层尺度上(参见 Lyons 1968),而人文学科本身则面临着以实践为导向的对应学科——数字人文学科(DHs)。尽管对该术语的定义存在许多尝试(参见 Terras 等人,2013),但沃里克的数字人文实践中承认缺乏这一定义可能最好地描述了其困境:DH 都是“计算的应用”。或人文学科的数字方法

研究”和“应用人文方法研究数字对象或现象”（2012:xiv）。尽管提供了一种新方法的可能性,但由于学术界内部的结构性阻力和不断发展的趋势被边缘化的文化,DH从未完全成为人文学科中的一个独立学科（参见Joula 2008;Prescott 2011）。

然而,新方法论的可能性将有利于整个人文学科,并允许扩展特定学科的想法。一方面,通过将大部分资金重新分配给科学,人文学科逐渐贬值和非货币化,导致研究生数量稳步下降（参见 Costa 2019:2）。这种趋势与从资源型经济向知识型经济的转变有关（参见 Powell 和 Snellman 2004）。以大学、工业界、政府和公众之间的知识转移生态系统作为社会和技术创新的先决条件（参见 Caravannis 和 Campbell 2009）,大学的战略定位转向了比理论更实用和更高层次的实践。知识转移活动（参见 Giuri 等人,2019）。在这个新发展的环境中,人文学科作为一个跨学科领域,位于理论人文学科及其外部应用之间。例如,数字版本向公众提供研究成果,Transkribus 等软件解决方案允许私人 and 学术个人共同开展研究活动,同时提供使用和理解人工智能的低门槛,固有的数字化、共同创造方法与传输的要求及其对数字化的关注反映了当代问题。DH是一个不断发展的领域,在与人文学科平等融合后,可以提供新的见解并抵消毕业生数量下降的趋势。

另一方面,特定主题的数字查询和解决方案开辟了一个新的研究领域。在给定的预测短信和自动完成实例中,数字统计分析同时将所提出的理论模型作为应用进行审查,从而对人工智能与人文学科之间的相互关系进行更大规模的研究。对人工智能和语言使用的依赖性的研究对于当代数字社会至关重要,因为人工智能在日常生活中以各种形式存在。特别是在交流和语言使用方面,缺乏对人工智能对个人可能产生的影响的了解可能会导致语言泡沫。它需要卫生署提供的数字问题和语言符号系统方面的知识。虽然软件开发人员根据既定系统创建人工智能,但没有数字专业知识的人文学者只能解释结果,而无法解释人工智能介导的语言创建本身的过程。然而,这个过程是解释人工智能所使用的语言的必要前提。因此,我们对符号的相互关系提供 DH 视角

基于人工智能在预测短信和自动完成方面的中介的系统和语言使用。

术语“预测短信”和“自动完成”是指移动设备上的常见功能,通过预测单词或建议和完成搜索查询条目来简化文本输入。在键盘手机兴起期间,基于 T9 等定义的语言数据集的预测短信算法通过减少击键次数显着加快了短信流程(参见 Ling 2007)。随着智能手机和数字键盘的出现,预测短信和自动完成系统的功能已基本趋同。

最近的研究表明,智能手机或搜索引擎上的最新人工智能模型不是使用基于统计值的算法来预测单词,而是使用更大的数据集来建议和完成文本和查询条目,并根据输入时间进一步定制、位置、口头、文本或缓存上下文以及实时人口统计特征(参见 Farzi 和 Taherv 2020;Vuong 等人 2021)。

然而,数据集和定制带来了有关沟通中的偏见和代理的某些问题。首先,数据集本身本质上是有偏差的(参见 Wich et al. 2022)。用于语言建模的数据集无法解释每种变体、白话和其他与感知标准的偏差,因此默认情况下倾向于一种标准变体。其次,基于个人语言使用、广告和优先考虑当前主题的当代标准变化数据集以及术语等变量定制建议,提高了某些措辞的可能性。第三,作为定制过程的直接结果,人工智能建议被考虑并被认为是意义创造过程中的积极代理人(参见 Oster 2015)。机构的这种法律责任迫使公司改变数据集和算法,以检测和隐藏贬义和诽谤性内容、假新闻和性别不敏感的语言等(参见 Baron 2020)。人工智能建议隐藏和个性化结果越多,意义创造的代理程度就越高,因此对语言使用的影响就越高,最终形成文本等语词 一种基于人工智能代理的语言变异边界,影响相互关系符号系统和语言使用。

3 符号金字塔

在传统语言学中,符号学 对符号系统的研究 在试图定义符号、语言和现实塑造之间的相互关系方面仍然发挥着关键作用。莎士比亚早于费迪南·德·索绪尔成为现代符号学的奠基人,他在戏剧《罗密欧与朱丽叶》中暗示了语言符号的复杂本质:“名字里有什么?我们称之为玫瑰的东西/任何其他

名字闻起来同样甜蜜。” (II.III.l. 170-72) “玫瑰”一词是一种被索绪尔称为能指的语言形式,它是表示概念或形式含义的任意声音组合,被索绪尔称为所指,在一个相互的过程中(参见 1959:66)。因此,“玫瑰”一词会在脑海中唤起一朵带红色花瓣和芳香的刺花的画面,而无需看到这朵花,也不需要以前见过这种花。因此,对现实的感知是由能指和所指之间的关系和关联塑造的,而不是实际物体本身。

奥格登和理查兹建立在符号学的核心原则之上,并将实际的外部现实融入到他们的概念中。虽然他们同意能指和所指的二分法,但他们认为所指对象,即看得见、摸得着、闻到的物理对象,是能指的根源,直接指所指。因此,双边关系需要调整为三角关系。随后出现的符号学三角被广泛认为是当今符号学的核心之一(参见 McElvenny 2014)。

然而,这两种理论都在一定程度上忽视了话语的交际特征。尽管奥格登和理查兹提出了话语中两个部分重叠的三角形的想法,但将对话分为两个不同但重叠的领域是有问题的,因为它没有考虑言外言语行为。塞尔和范德维肯的契合方向理论作为言外行为分类学的基础,将现有的符号学理论扩展到话语领域。他们对语言的四种可能的适应方向进行了分类,根据这些方向,要么话语适合世界的事态(单词到世界),世界改变以适应话语(世界到单词),世界改变宣称它通过话语(双向)而改变,或者世界的改变以话语(零方向)为前提(参见1975:52)。通过将能指、所指和所指之间的相互关系置于内部,并将其与话语相结合,塞尔事实上创造了符号学方块,但从未提出过符号学方块。在这个方块中,所指通过能指的话语来概念化所指,然后由对话者解码。然后,对言外行为的接受或拒绝又回到所指,这反过来又创建了一个指向所指的反馈循环,并通过代理指向能指,从而最终确定并构成了交换(图1)。

当人工智能作为意义创造的中介被插入到交换中时,二维框架扩展到第三维(图2)。

在给出的示例中,人工智能预测文本并建议结果。因此,它成为语言创造的积极部分,而不是话语本身的一部分。它位于二维交流领域之上,但又连接到符号金字塔的每一个支柱。首先,所指对象不再构成自然言语行为。通过依赖建议文本和基于统计值的预测,

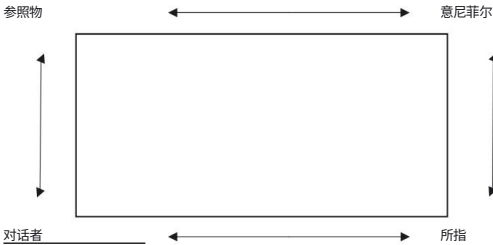


图 1:塞尔和范德肯的符号学模型适合。

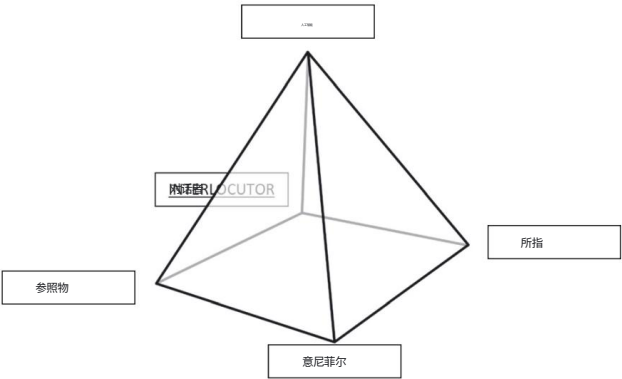


图 2:符号金字塔。

所指对象丧失了言语的完成性,更重要的是,将言语的力量转向人工智能。尽管话语的最初意图仍然是指称对象,但它是通过人工智能的参与来调解的,而不能再仅仅归因于指称对象。因此,我们建议将言外之力分为两个不同但相互关联的实例:意图的和中介的

言外之力。其次,能指是由人工智能根据所指对象的统计偏好创建的。通过进一步考虑上述定制四个维度,人工智能干预了所指、能指和所指之间概念化的相互过程。所指对象的碎片言语行为通过人工智能来调节以创建能指。能指并不唤起最初预期的所指,而是由人工智能的言外之力塑造的一个中介实例。第三,它从而重塑了世界,以适应说话的确切时间和地点。最后,对话者的任务是解码所中介的言外之力,而不是所指本身。因此,对话者对拒绝的接受并不是作为一个反馈循环来完成话语中的话语,而是为了完成人工智能介导的所指的概念化,将更多、更详细的数据交给人工智能。

最初建议或预测的沟通成功或失败。虽然符号学方块的所有角落仍然相连,但最初的联系被切断了,因为概念化的所有实例都通过人工智能运行。

为了分析这个理论,我们首先必须理解它的含义。

自然言语行为在数字通信中发生了改变。预测短信和自动完成改变了语言使用和符号系统之间关系的既定过程。不过,基础数据集和基于偏好的反馈都会导致模型出现偏差。这种偏见巩固了文本等语词中的语言使用。反复接触文本同语词反过来又改变了对语言的感知,从而改变了现实,与能指和所指关系中现实的概念化相一致。重复曝光还引入了时间概念,作为文本等语词创建中的一个重要变量。因此,分析的数据应该显示词频随时间、话语和外部现实的极端变化。

4 自动完成和预测短信:Twitter 学习

4.1 方法论

为了测试符号金字塔理论及其潜在含义,我们必须创建一种新的方法论,通过证明智能手机设备上的语言使用与时间相关的极端变化来近似自动完成的建议效果。外部现实中的重大事件(参见 Baron 2020:412; Hazan 等人 2022:3)。由于几个问题,我们选择相应地分析数据,并因此使用代理证明。首先,基于现有数据自动完成的语言变化概念的直接证明需要评估训练数据的偏差程度,并对每个单独级别的人工智能反馈数据进行总体说明。由于企业初始训练数据集不可用,并且人工智能算法关于反馈数据整合的精确推理过程是一个“黑匣子”

(Wenskovitch and North 2020:31),这种方法会伪造结论。或者,可以开发一个单独的自我暗示人工智能,使用开放数据集进行训练,并在无菌测试环境中使用;然而,这种方法无法利用数字环境中的现有数据进行复制,因此无法解释现有现象或人工智能在意义创造中的作用。

其次,必须在个人层面上进行具有时间相关参数的直接概念证明,以解释所分析的语言变化。

为了获得个体水平上的代表性样本,所需数据集的庞大规模将导致无法进行结论性分析。

相反,我们决定分析有关 COVID-19 的开放 Twitter 数据集以进行科学研究 (Banda 等人, 2023), 以检验我们基于词频和相关性的语言使用随时间变化的假设。该数据集由佐治亚州立大学提供和托管。他们从 2020 年 3 月 11 日到 2023 年 3 月 15 日开始根据相关标识符从流 API 中抓取数据,产生了超过 3.6 亿条独特的推文。Twitter 数据很好地满足了我们调查的目的。它几乎专门通过智能手机使用,允许通过数字键盘自动完成,并且其字符限制需要数字书面通信所特有的文本口头表达 (参见 Soffer 2020)。使用有关 COVID-19 大流行的数据是因为该事件具有全球意义,导致了广泛且多样化的数字讨论以及人们认为的讨论的激进化 (参见 Dehgan 和 Nagappa 2022;Schulze 等人 2022)。这样可以更好地分析其中的极端情况和语言变化。

为了分析 R 中的数据并根据 Twitter 的服务条款,我们使用私人开发者 ID 通过 Hydrator¹ 对 ID 数据集进行水合。然后,我们首先通过 `str_remove_all` 提及、主题标签和链接、将 `iconv` 表情符号清理为 ASCII、用空格 `str_replace_all` 标点符号、`str_squish` 重复空格清理数据集,并将所有情况更改为 `str_to_lower`。之后,我们将推文标记为单词,并通过 `tidytext` 变量中的 `stop_words` 数据集删除停用词。

值得注意的是,由于“COVID-19”一词的具体情况,我们并未删除任何数字。

为了考虑到任何可能不相关的推文,我们只计算了与相关关键词语相邻的数字。在现在干净的数据集中,我们使用 `findFreqTerms` 函数从第一个条目迭代到最大值,并根据时间戳和频率对结果进行分组。我们没有删除转发,因为它们与通过重复话语进行反向传播一样具有沟通目的。

4.2 分析

Twitter 数据的定量分析遵循了预期的模式。

在病毒出现及其全球传播最初爆发后,数字逐月减少。2020 年,这一数字仍保持在较高水平,并缓慢下降,从 4 月份的平均每天 420 万条推文下降到 3.8 条。5 月 100 万 7 月 200 万 9 月 150 万

1 记录当下。(2020)。Hydrator [计算机软件]。检索自<https://github.com/docnow/Hydrator>。

十一月。2021 年,推文数量开始从第二季度的平均每天 120 万条减少到第四季度的 60 万条。到 2022 年,这一数字将稳定在每天 20 万条推文左右;然而,我们在交流输出中发现了五个单独的异常,如表 3 中突出显示的那样。我们使用这些尖峰作为标记来分析词频中最明显的变化,建议与重大事件的链接,并评估随时间的分布和情绪(图 3))。

第一次高峰出现在 2020 年 10 月上旬,当时美国总统唐纳德·特朗普(Donald Trump)的 COVID-19 检测呈阳性并住院治疗(参见 Miller et al. 2020: A1)。2021 年 1 月上旬的第二次高峰恰逢世界卫生组织紧急批准疫苗使用(参见 Boseley 2021: 21)。2021 年 5 月的第三次高峰恰逢 COVAX 挑战和印度病毒激增(参见 Webster 2021:2054)。第四次高峰表现为 2021 年 7 月下旬至 2021 年 9 月上旬之间的平稳期,主题范围包括疫苗通行证、疫苗强制规定以及播客 Joe Rogan 发布的关于伊维菌素假定效果的病毒视频(参见 Izadi 和 Yahr 2021:C3)。2021 年至 2022 年之间的最后一次高峰期,除了一般的岁末年初和有关庆祝活动的规定外,没有发生任何重大事件。

在第一个峰值期间,除了“特朗普”和与医院相关的术语外,我们可以检测到的词频、phi 相关性或情绪没有显着的统计变化,总体提及量相对于日平均值增加了 0.52 和 0.42 分别为,phi 相关性为 0.58。整体通讯量的激增似乎仅集中在美国总统住院的消息上。

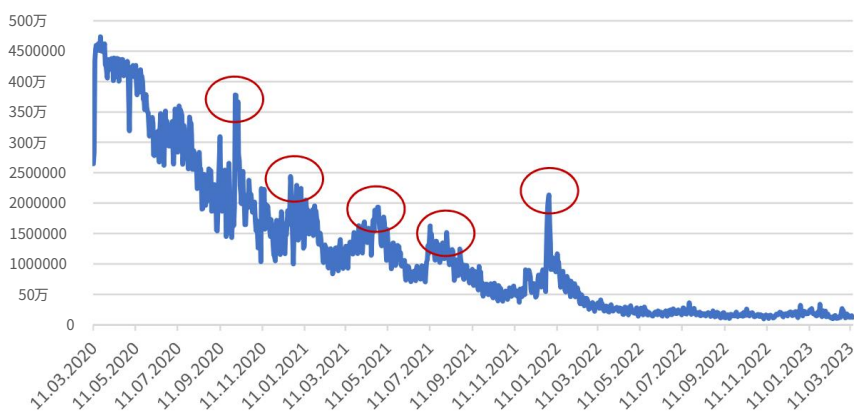


图 3:2020 年 3 月至 2023 年 3 月有关 COVID-19 的推文数量。

第二个峰值显示疫苗相关术语2的频率相对于日平均值增加了 0.71。在高峰期间,围绕这些术语的情绪主要是中性 (0.91) 或积极 (0.07)。为了进一步分析这一变化,我们排除了峰值的时间范围,只关注疫苗相关术语的频率分布及其在其他过程中的情绪

三个尖峰。在第三个和第四个峰值期间,疫苗相关术语的分布相当均匀 (分别为 0.42 和 0.46),而第五个峰值的词频则低于 0.12。不过,第三和第四个峰值的情绪分析揭示了极端情况的变化。虽然大部分言论仍然是中性的 (0.85 和 0.82),但总体情绪趋势正在下降,分别为 0.08 积极/0.07 消极和 0.06 积极/0.12 消极。在第三个和第四个峰值过程中,与分析术语接近的五个最常见的负面词的 phi 相关性说明了语言使用的变化:尽管大多数负面情绪的相关性只能被认为可以忽略不计到较低,但重要的是认为大多数与疫苗相关的言论 (0.85 和 0.82)被归类为中性。在这种情况下,负面术语“谎言”相对于尖峰 3 和 4 中整个话语的 phi 相关系数出奇地高。此外,随着时间的推移,负面情绪的phi相关系数略有上升,表明类似措辞有所增加 (表1和表2)。

第三个峰值显示,除了术语“longcovid”和“booster”的逐渐增加之外,我们可以检测到的词频、phi 相关性或情绪没有显着的统计变化。这两个学期都从第二峰到第三峰稳步上升。

第四个峰值或平台显示多个异常,这又与话语数量的长期增加有关。数据显示,“伊维菌素”(0.81)、“护照”、“证书”(0.72)和“授权书”(0.58)的日均词频有所增加。人们对这些术语的看法差异很大 (表 3)。

伊维菌素与“蠕虫”(0.67)密切相关,我们将其归类为负面,可以追溯到乔·罗根 (Joe Rogan)和有关它的新闻报道。护照/证书和授权书显示出类似的情绪分布,但负面情绪程度高于预期。强制的负面情绪与“坏”(0.14)相关性最强,而护照/证书与“假”(0.16)相关性最强。第五个峰值显示我们可以检测到的词频、phi 相关性或情绪没有显着的统计变化。

2 与疫苗相关的术语还包括辉瑞 (Pfizer)等公司名称、疫苗传播系统的技术术语 mRNA,以及美国国家流行病学顾问福奇博士 (Dr. Fauci)。

表 和 :尖峰 和 中负面词的 Phi 相关性。

| -n 个单词尖峰 | 相关性 |
|----------|-----|
| 莉萨 | . |
| 伪造的 | . |
| 癌症 | . |
| 误传 | . |
| 阴谋 | . |

| -n 个单词尖峰 | 相关性 |
|----------|-----|
| 谎言 | . |
| 伪造的 | . |
| 癌症 | . |
| 误传 | . |
| 阴谋 | . |

^a包括词形变化。

表 :峰值 中的情绪分布。

| 学期 | 积极的 | 中性的 | 消极的 |
|-------|-----|-----|-----|
| 伊维菌素 | . | . | . |
| 护照/证书 | . | . | . |
| 授权 | . | . | . |

4.3 讨论

数据分析表明,语言使用发生了变化。
COVID-19 大流行期间的在线交流。我们分析了
3.6 亿条推文的词频和及时分发
仔细选择 3 年期间五个异常中最显着的变化
发现、确定和解释语言使用变化的时期。两者内容
根据最常用的单词及其与其关系的推文
及时出现表明它们与外部现实中的特定事件相关
这似乎引发了某个单词或短语的激增,
与特定情绪相结合,然后变得相对普遍。毒品
例如,“伊维菌素”在 2021 年 9 月出现频率激增,当时 Joe
Rogan 对 COVID-19 的描述迅速走红。今天,谷歌搜索查询自动完成建议将“伊维菌素”列为输入
“iv”后的第八个总体建议。
同样,在大流行期间,总体负面情绪有所上升
疫苗相关术语。从新闻负面情绪分布 0.02 开始
关于新药开发和传输技术的报道,

在随后的高峰期间,情绪迅速下降至 0.07 和 0.12,总体上似乎质疑其有效性或疫苗接种的必要性。否定词的词频和统计配对均有所上升。虽然这并不是关于预测短信或自动完成语言感知和形成条目的影响的直接概念证明,但符号金字塔模型(其中包括人工智能作为通信符号系统中意义创造的代理),是互惠的,因此必然影响金字塔的所有支柱。上述伊维菌素药物的例子最好地说明了这一理论。尽管伊维菌素于 1975 年首次被发现,并已用于治疗人类和动物寄生虫近 50 年(参见 Campbell 2012),但在大流行之前,其在 Twitter 上的传播量可以忽略不计,并且主要与新的医学发现有关或讲座。然而,在大流行初期,这种药物因其对治疗 COVID-19 的预期效果而流行起来。数据表明,在监测期开始时,“伊维菌素”的使用相对较少,并且主要与“治疗”、“治愈”或“蠕虫”等术语相关。如果一个词很少被使用,那么有关相关术语的反馈循环对于未来的建议或预测就变得不可或缺。在峰值 4 期间的后期激增中,伊维菌素立即与负极端中的“蠕虫”和正极端中的“治愈”或“羟氯喹”高度相关。这表明,根据所指对象的偏好,人工智能将预测并建议在原始术语附近使用两个极端中的任何一个。那么,这种语境预测性地影响了药物在其表现形式中可能的概念化,即通过改变其言外之力来表示,如下例所示,由于 Twitter 记录的数据规定,推文:(1)我妻子的女儿去了一家密苏里州尤宁敦的医生,他给了她“伊维菌素”这种治疗感冒的蠕虫药,很高兴我们注射了疫苗

田纳西州。

- (2) 目前FB上建议人们服用3倍剂量的伊维菌素来治愈新冠病毒。

基于偏好和预测,显著变化是根据人工智能的统计值。因此,对话者的解码过程不一定涉及预期的,但肯定涉及中介的言外之力。

外部现实的变化(在本例中为大流行)会影响日常语言的使用,然后触发建议和完成查询的人工智能统计值的重新配置。正如 Twitter 自己的人工智能过去所表明的那样,这种重新配置在极端情况下变得最为普遍(参见 Cherelus 和 Tennery 2016)。它似乎在人工智能、话语和外部现实之间创建了一个反馈循环,通过以同样的方式建议和完成条目,使先入为主的观念僵化。因此,人工智能改变了另一种人对现实的感知

与之前所谓的“社交媒体泡沫”相比,人工智能的形式有所不同:通过创造语言,人工智能或多或少微妙地影响着用户的措辞选择,用户通过这种选择创造了自己对现实的感知。通过法律行动和立法关注,公司已经意识到了这个问题,并在该领域专家的协助下尝试减轻影响;然而,大多数尝试只是禁止或隐藏不需要的内容。这可能对公司来说是一个更好的短期解决方案,但它并没有解决人工智能的根本问题,同时通过将焦点从消极极端转向积极极端,直接导致文本同义词的形成。

5 重新思考人文学科:社会影响还是边缘化?

人文学科在这场辩论中的作用是双重的。在特定学科方面,人文学科可以对语言的运作、符号系统以及通过语言使用进行意义创造的过程提供深刻的见解。关于传统语言使用的现有理论,在这种情况下是言外言语行为和符号学,可以适应数字环境和人工智能,反过来可能有助于数字通信中结构问题的实际解决方案。鉴于语言学只是许多学术领域中的一个,其他学科可能会对这些问题和其他令人关注的问题提供不同的见解。通过将研究转移到数字环境和更高层次的相关兴趣领域,而不必在此过程中采用数字方法,人文学科可以通过解决涉及个人、公共和私营部门的紧迫问题来增强其社会影响力作为立法。因此,将领域进一步扩大到数字化也可以作为货币逐渐贬值的对策。更大的社会影响还将解决大学越来越多地参与的知识转移生态系统带来的挑战。

尽管人文学科在社会承诺领域表现出色,但知识转移的重点在于创新和创业。通过跨学科研究将人文学科的方法应用到数字和共同创造解决方案中可以帮助促进创新。Transkribus 软件是基于原始人文学科方法论的创新方法的一个例子。

从战略角度来看,在数字领域利用传统方法可以缓解数字与人文之间二分法的内在张力。将研究扩展到数字环境既可以减少卫生署将自己建立为一个实体的需要,又可以通过解决利基问题来减少边缘化。

参考

- Banda, Juan M., Ramya Tekumalla, Guanyu Wang, Jingyuan Yu, Tuo Liu, Yuning Ding, Katya Artemova, Tutubalina Elena 和 Gerardo Chowell. 2023 年。用于开放科学研究的大规模 COVID-19 Twitter 聊天数据集 一项国际合作。流行病学 2(3)。315–324。
- Baron, Naomi S. 2020。在智能手机上说话、阅读和写作。在理查德·林看来，Leopoldina Fortunati, Gerard Goggin, Sun Sun Lim 和 Yuling Li (编辑)，《牛津移动通信与社会手册》，408–424。纽约：牛津大学出版社。
- 博斯利·莎拉。2021。牛津/阿斯利康疫苗会给英国带来什么改变？守护者。一月 2, 21。
- 布贝克、塞巴斯蒂安·瓦伦·钱德拉塞卡兰、埃尔丹·罗南、约翰内斯·格洛克、埃里克·霍维茨、埃斯·卡马尔、Peter Lee, Yin Tat Lee, Yuanzhi Li, Lundberg Scott, Harsha Nori, Palangi Hamid, Marco Tulio Ribeiro 和 Yi Zhang。2023 年。通用人工智能的火花：GPT-4 的早期实验。https://arxiv.org/abs/2303.12712 (2023 年 4 月 19 日访问)。
- 坎贝尔·威廉。2012。阿维菌素和伊维菌素的历史，以及其他历史的注释
大环内酯抗寄生虫药。当前的制药生物技术 13(6)。853–865。
- 卡拉扬尼斯、埃利亚斯和大卫·坎贝尔。2009。“模式3”和“四螺旋”：迈向21世纪的分形
创新生态系统。国际技术管理杂志46(3/4)。201–234。
- 切勒克斯、吉娜和艾米·坦纳利。2016 年。微软的 AI Twitter 机器人在发布种族主义、性别歧视推文后关闭。路透社 3 月 24 日。
- 科斯塔·罗萨里奥·库托。2019。人文学科在当今知识社会中的地位。帕尔格雷夫
通讯 5(38)。1–5。
- 达内西·马塞尔和斯蒂芬妮·沃尔什·马修斯。2019。人工智能：符号学视角。中国符号学研究15(2)。199–216。
- 德甘·埃桑和阿什温·纳加帕。2022。Alt-Tech 生态系统中话语的政治化和激进化：关于 gab 社交的案例研究。社交媒体+社会 8(3)。1–12。
- 记录现在。2020。Hydrator [计算机软件]。位于：https://github.com/docnow/
Hydrator (2023 年 4 月 19 日访问)。
- 法尔兹、赛义德和赛德·塔赫里。2020。定制查询自动完成和建议 – 回顾。
信息系统87。47–59。
- 朱里·保拉、费德里科·穆纳里、亚历山德拉·斯坎杜拉和劳拉·托斯基。2019。大学知识转移活动的战略定位。技术预测和社会变革。国际期刊 138。261–278。
- 哈赞·蒂莫西、亚历山德拉·奥尔泰努、加布里埃拉·卡扎伊、费尔南多·迪亚兹和迈克尔·戈莱比夫斯基。2022 年。关于网络搜索自动
暗示调节的社会和技术挑战。第一个星期一 27(2)。1–17。
- 益普索。2022年。全球对人工智能的看法和期望。纽约：益普索。
- 伊扎迪、埃拉赫和艾米丽·亚尔。2021 年。乔·罗根 (Joe Rogan) 感染了 Covid-19,正在服用未经证实的驱虫药。这
华盛顿邮报 9 月 3 日。C3。
- 朱拉·帕特里克。2008。数字人文领域的杀手级应用。文学和语言计算 23(1)。73–83。
- 肯尼迪、布莱恩、亚历克·泰森和艾米丽·萨克斯。2023 年。公众对日常活动中人工智能的认识。
纽约：皮尤研究中心。
- 基尔希·亚当。2014 年。技术正在接管英语系。数字化的虚假承诺
人文学科。新共和5月2日。
- 科济列娃、阿纳斯塔西娅·斯特凡·赫尔佐格、菲利普·洛伦茨·斯普林、拉尔夫·赫特维希和斯蒂芬·莱万多夫斯基。
2020。在线环境中的人工智能。德国公众态度的代表性调查。
柏林：马克斯·普朗克人类发展研究所。

勒罗伊·阿尔芒·玛丽。2015年。一个学习共和国：人文学科数字化。《纽约时报》二月13。

林·理查德。2007。短信的长度和预测短信的使用：谁使用它以及使用量他们有什么要说的吗？AU Tesol 工作论文 4. 1-18。

里昂·约翰。1968。理论语言学导论。剑桥：CUP。

麦克尔文尼·詹姆斯。2014。奥格登和理查兹的《意义的意义》和早期分析哲学。语言科学 41(B)。212-221。

米勒·泽克·吉尔·科尔文和阿梅尔·马达尼。2020年。特朗普出院，说“我感觉很好”。时代10月6日。A1。

奥斯特，2015年1月。沟通、诽谤和中间人的责任。法律研究35(2)。348-368。

鲍威尔·沃尔特和凯萨·斯内尔曼。2004。知识经济。社会学年度评论 30。199-220。

普雷斯科特·安德鲁。2011。消费者、创造者还是评论者？受众和使命的问题

数字人文。高等教育中的艺术与人文 11(1-2)。61-75。

索绪尔·斐迪南。1959。普通语言学课程。纽约：哲学图书馆。

舒尔茨·海蒂·朱利安·霍纳·西蒙·格雷普尔·马克西米利安·吉格胡贝尔·伊莎贝尔·德斯塔和戴安娜·里格。2022年。边缘平台上的极右阴谋团体：电报激进化动态的纵向分析。融合：国际新媒体技术研究杂志 28(4)。1103-1126。

塞尔·约翰和丹尼尔·范德维肯。1975。言外逻辑基础。剑桥：CUP。

索弗·奥伦。2020。从文本口头到口头文本：语音查询的案例。融合：国际新媒体技术研究杂志 26(4)。927-941。

斯特劳姆海姆，卡尔。2014年。数字人文泡沫。高等教育内部 8. 6-9。

特拉斯·梅丽莎·朱丽安·尼汉和爱德华·范豪特。2013。定义数字人文。一位读者。法纳姆：阿什盖特。

Vuong,Tung,Salvatore Andolina,Giulio Jacucci 和 Tuukka Routsalo。2021年。口语对话上下文改进了网络搜索中的查询自动完成。ACM 信息系统交易 39(3)。1-32。

沃里克·克莱尔·梅丽莎·特拉斯和朱丽安·尼汉。2012。简介。摘自 Claire Warwick,Melissa Terras 和 Julianne Nyhan (编辑)，实践中的数字人文。xii-xx。伦敦：方面。

韦伯斯特·保罗。2021年。COVID-19 事件时间表。自然医学 27。2054-2055。

温斯科维奇·约翰·诺斯和克里斯·诺斯。2020。交互式人工智能：针对“两个黑匣子”问题进行设计。计算机58(8)。29-39。

威奇·马克西米利安·托拜厄斯·埃德·哈拉·阿尔·库瓦特利和乔治·格罗。2022。偏差和比较框架辱骂性语言数据集。人工智能与伦理 2. 79-101。

比奥诺特

Felix Poschinger

汉堡大学，汉堡 20148，德国felix.poschinger@uni-hamburg.de <https://orcid.org/0009-0008-6616-5795>

菲利克斯·波斯辛格 (Felix Poschinger) 在汉堡大学知识交流机构工作。他的研究兴趣集中于人文学科的数字重构，特别关注当代通信系统。