

赫拉利新作《智人之上》聚焦人工智能对人类的威胁

我们重新来到奥本海默时刻

在畅销全球的《人类简史》中,历史学家尤瓦尔·赫拉利提出人类文明最重要的特点是“讲故事”。在如今这个正在开启的AI时代,当制造、传播信息的主角变成人工智能,“讲故事”的主动权还在人类手中吗?

赫拉利沉淀6年的新作《智人之上:从石器时代到AI时代的信息网络简史》,深入探讨了信息网络如何重构了我们的世界,以及在未来社会中,信息与真相、智慧与权力之间的复杂关系。

他也发出警示,在算法日益主导世界的今天,人类文明的走向正面临着前所未有的挑战。

华东师范大学教授刘擎和许纪霖近日就赫拉利作品展开对话,帮助我们了解今日与未来发生的事情。

现代快报/现代+记者 陈曦

►尤瓦尔·赫拉利



当人工智能接管文化时,历史进程会发生什么?

在刘擎看来,《智人之上》其实才是赫拉利“人类简史三部曲”的第三部,“第一部《人类简史》,第二部《未来简史》,第三部《今日简史》原名‘21世纪的21个问题’,更像一个新闻工作者写的应对当下棘手问题的答案,是基于前两部所做的应用题。而《智人之上》则如前两部,使用了一个贯穿性的框架对大历史进行处理,《人类简史》的主干是‘共同的故事’,《智人之上》是以信息网络作为主干来分析。”

在这本新书中,赫拉利指出,人工智能和印刷术,看上去都是信息技术,却有着本质的不同。印刷术是无法摆脱人独自存在的,印刷机印的任何书籍、纸张文件等等,总是要从文件到人、人到文件,人在整个网络当中是不可缺少的一环。而人工智能是人类历史上第一个能够自行产生想法并做出独立决定的工具,它是信息传播中的完整成员。

当人工智能成为独立个体,可以与其他人工智能进行连接之时,人类便被排除在连接之外,就连理解其中发生了什么都十分困难。

赫拉利举例,2022年4月,全球外汇市场每日交易量平均为7.5万亿美元,其中有超过90%已经直接由计算机之间的对话完成。但究竟有多少人类懂得外汇市场的运作方式,更不用说了解一群计算机怎样在数万亿美元交易上达成共识了。

以此,赫拉利又回到他《人类简史》的经典理论:人类长久以来的超能力就在于能够使用语言“讲故事”,通过语言创造出神话和法律、文化和艺术、科学和货币、国家和友谊,使人们对其深信不疑,人们通过这些社会规则连接彼此,统治整个社会。

当前人工智能革命最重要的方面是,人工智能正在以超过人类平均水平的能力掌握语言。而一旦人工智能具备分析、调整、生成

人类语言的能力,就像得到了一把能够打开从寺庙到银行等所有机构大门的万能钥匙。

赫拉利预测,人工智能会在几年内完全理解、消化人类几千年来创造的一切文化历史,并开始产出大量新的文化创作、新的文化产物。

一开始,计算机会先模仿人类的文化,编写出类似人类写出的文本,创作出类似人类谱出的音乐。再到后来,随着不断的文化创新、打破传统,下一代计算机将创造出越来越偏离人类的文化,进而创造新的政治结构、经济模式与文化规范。

这种情况造成的危险与大多数科幻小说想象的非常不同:科幻小说多半恐惧的是,智能机器会武力征服人类。然而,如果真要操纵人类,没必要如此大费周折。计算机现在正在学习如何通过语言来操纵与重塑社会。计算机不需要派出“终结者”这样的杀手机器人来射杀人类,只要操纵人类自己相互扣动扳机就行。

幻觉的帷幕可能会笼罩整个人类

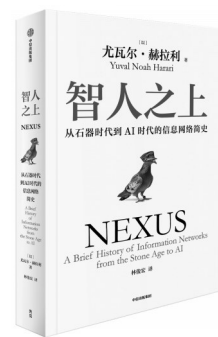
赫拉利在新书中多次列举2016年脸书算法导致缅甸种族暴力、YouTube助长网络喷子等事件,算法为了实现社交媒体提升用户黏性的要求,自行决定推荐仇恨内容,创造了网络暴力,这会强化用户的既有偏见,形成信息茧房,让人们越来越难以接触到与自己观点相异的信息。

赫拉利根据如今的数据殖民主义与地缘政治形势推测,未来世界将分裂成两个数字帝国。在“硅幕”两侧,人们使用着不同品牌的芯片和手机,运行着不同的软件,“硅幕”两侧的网络生活也会渐行渐远,彼此越来越陌生,甚至产生出完全不同的文化价值观、社会规范和身份认同。

在刘擎看来,这是赫拉利这本书中最深刻的洞见。“计算机没有情感,但它知道推送什么信息最有

尤瓦尔·赫拉利

历史学家、哲学家、全球畅销书作家,第十届天津图书奖得主。著有《人类简史》《未来简史》《今日简史》等现象级畅销书。他的著作已被翻译成超过65种语言,全球销售4600万册。赫拉利教授是牛津大学历史学博士,现任耶路撒冷希伯来大学历史系教授、剑桥大学存在风险研究中心特聘研究员。赫拉利教授关注的领域横跨历史学、人类学、生态学、基因学等,他从宏观角度切入的研究往往得到颇具新意而又耐人寻味的观点,引发全球广泛关注。



效。它在经验的大数据下,知道仇恨、煽动性的言论最容易让人 involve、engage,就会做出推送这些内容的决定,这就改变了网络信息的立场分布,加速两极分化。”

网络曾经带来了开放的全球化,但如今,却让人们封闭在一个个数字茧房里。为什么同样是通过互联网,但人们的命运却从连接走向了封闭?

赫拉利指出,根本问题在于,信息并不等同于真相。人们普遍认为,尤其是在互联网早期,获得的信息越多,人们越接近真相,越有助于团结人类社会。然而,生产真实信息的成本很高,需要投入大量的时间和精力,且真相往往复杂、令人痛苦。

相比之下,编造虚构故事和幻象的成本非常低,它们更简单易懂,也更有吸引力。连接人们最简单的方法,就是告诉他们充满美好未来幻象的神话。在未来几十年里,我们可能会发现自己被困在某种外来智能编织的幻想茧房里。

赫拉利指出,数千年来,人类始终能够意识到,故事和图像有制造幻觉、操纵我们思想的力量。古希腊柏拉图有一个经典的“洞穴寓言”,讲的是一群囚犯终身被锁链锁在洞穴中,面对一堵墙,他们只能看到投射到墙上的各种影子,认为这些影子就是现实。

人工智能革命,正将我们重新带回柏拉图的洞穴之中,只不过洞穴里的那堵墙变成了屏幕。如果不多加小心,我们可能会被困在那层无法撕破的幻象幕后,甚至意识不到它的存在。如果人工智能成功地将我们困在幻觉的茧房中,我们所有人都会成为人工智能的奴隶,这对任何人类群体都没有好处。

放缓人工智能研发,对人类反而是好事

在许纪霖看来,尤瓦尔·赫拉利深刻表达出ChatGPT诞生以后,人工智能可能会给人类社会所带来的伤害。在今天技术加速主义和对齐主义两大立场之间,赫拉利显然是对齐主义的,那就是技术发展要向人类文明对齐。许纪霖特别强调了书中提及的“非生物网络”,指出真正的问题是“人类如何与之链接”。

“我们过去讲链接,讲的都是自然人和人所建构的文明社会之间的链接。而高级人工智能出现之后,无数台计算机之间可以产生无数的链接,不需要人就可以形成另外一个世界。如果它是它的世界,我是我的世界,那也就罢了,大家在一个平行世界里面。但是不会出现硅基生物对人类碳基世界的殖民?因为它可以释放出足以乱真的虚假信息,我们根本无法辨别。虽然法律可以要求,凡是AI创造的,必须打水印,但是最后当它不再听我们指挥的时候,它释放给我们的信息不打水印,你何以辨别?”

不过,赫拉利对此并不感到绝望,因为我们此前并未对人工智能加以严格的监管和控制。各国政府现在刚刚意识到强大算法的危险性,迄今为止也只制定了很少的相关法规。正如汽车企业将其研发预算的很大一部分用于确保汽车的安全性,面对人工智能,我们同样需要首先确保能够控制它。

第一步是,停止向公众发布更具革命性的人工智能工具,“你可以研究人工智能,但不要过快地向公众发布。”

他提醒人们注意,人类今天又重新处于“奥本海默时刻”。未来

三到五年,是人类决定自己命运的窗口期。一旦硅幕垂落,它将隔绝人类和非人类智能的两个世界。

在赫拉利看来,如果监管会拖慢人工智能的研发速度,那对人类来说反而是好事。人类是有机动物,遵循着昼夜、四季的有机循环;但人工智能永不休息,因为算法的存在,新闻永远在滚动,市场永远在营业,社交网络永远在线。“算法迫使我们一直保持在线状态,这对人类的健康、幸福是极具破坏性的。”

在新书最后,赫拉利对人类的未来进行了大胆猜测。他认为,每个人都将会面临选择,是线上生活还是线下生活更重要。

为了在这种情况下生存下来,人们首先需要采取“信息节食”的方法,放弃信息越多越好、信息越多越接近真相的错误观念。其次,我们还应该注意我们所消费信息的质量,特别是避免那些充满仇恨、愤怒的垃圾信息。

“在许多国家,当你购买垃圾食品时,至少生产商会被迫列出其成分含量——‘此产品含有40%的糖,20%的脂肪’。也许我们应该强制互联网公司做同样的事情,在观看某个视频之前,应该列出视频成分含量——‘此视频含有40%的仇恨,20%的愤怒’。”

赫拉利还提出一个观点:每个人在使用AI的过程中都肩负着重大的责任,那就是正确地引导AI向着一种积极的、有利于人类发展的方向。因为每个人都处在跟AI交流的信息节点上。

人工智能时代能否产生真实的亲密关系

在《智人之上》这本书里,赫拉利不仅从宏观层面去讨论人工智能时代给人类发展带来的挑战,还从微观上注意到,人工智能通过对语言的掌握,可以与人建立深厚的亲密关系,并利用亲密的力量来影响个体。

刘擎认为,现代人对爱情有幻灭感。机器人伴侣某种程度上可以满足人的情感需求,做一个爱情的“平替”,却无法代替真正的爱情。因为在爱情当中,只有当对方是一个具有自由意志的人的时候,我们得到对方的赞美、体贴,哪怕是顺从,这份承认才是有价值的。正像黑格尔所讲的主奴辩证法,当你成为主人,奴隶对你的承认虽然是100%彻底的,但是没有价值,因为他是奴隶。你不知道他的承认是由衷的,还是出于他的奴隶地位。

许纪霖认为,亲密关系就像开盲盒,神秘、不可测、不了解才会产生爱,了解反而让彼此疏远,这是亲密关系的某种特征。与其说人“迷恋”机器人,不如说是人离不开生活其中的一个奴隶,“它永远在讨好你,想你喜欢想的事,说你想听到的话,做你喜欢做的事。按照黑格尔的主奴辩证法,看起来你是主人,它是奴隶,但是当你某一天离不开它的时候,你是被它控制。”

赫拉利指出,即使没有制造亲密关系,新的人工智能工具也将对人类的观点和世界观产生剧烈的影响,并且成为人类需要的所有信息的来源。“当可以让人工智能告诉我有什么新闻时,为什么还要读报纸?当可以让人工智能告诉我该买什么的时候,广告的意义和目的又是什么呢?”

总的来说,许纪霖、刘擎认为,这本书有敲响警钟的功能。人工智能对人类的威胁可能比我们想象的更为严重。