



自2022年底问世以来, ChatGPT 发展可谓波澜壮阔。截至2024年1月, OpenAI 已经拥有1.8亿用户, 每月访问OpenAI网站的次数超过16亿次。2024年2月, 文生视频模型“Sora”问世, 它能生成超逼真的一分钟电影片段。不用说, OpenAI 并不是唯一一家进入这一市场的公司。2024年2月, 谷歌“DeepMind”德米斯·哈萨比斯(Demis Hassabis)团队开发了 Gemini 1.5 Pro; 2024年3月, 马斯克开源了其人工智能系统 Grok。发展正在飞速前进。然而, 除了“内容”, 生成式人工智能还引发了关于其用处和缺点的激烈社会辩论。

### 共生是进化过程中的一个关键因素

我们应该如何理解生成式人工智能并评估对立的价值观? 我认为, 进化和共生的(symbiotic)视角, 更具体地说, 技术共生(techno-symbiosis)的概念, 为实现这一目标提供了一个很好的切入点。

共生这一概念在古希腊语中的意思是“共同生活”, 但在二十世纪的生物学中, 这一概念一直未得到充分研究。其中占主导地位的新达尔文主义认为, 生命与自私的个体之间的竞争和斗争密不可分, 而这是由自私的基因驱动的。美国古生物学家杰伊·古尔德(Jay Gould)认为, 这种人类观与新自由主义非常相似, 这绝非巧合。他认为, 经济学家亚当·斯密在其《国富论》(1776年)中不仅从个体参与者的自身利益出发解释了市场的运作, 还启发达尔文在《物种起源》(1859年)中将自身利益视为生命进化的驱动力。虽然斯密和达尔文也承认合作的作用, 但后来的(新)达尔文主义者几乎只关注竞争。

然而, 共生在进化和经济学中的作用现已凸显。关于共生, 可以分为三种基本类型: 同种生物共生(群居、兽群中和成团共生的动物)、不同种生物共生(如花卉和昆虫)以及生物和非生物共生(从海狸坝到智能手机)。

共生是进化过程中的一个关键因素, 原因至少有两个。首先, 它是一种重要的创新力量。微生物学家林恩·马格里斯(Lynn Margulis)在上世纪六七十年代提出的内共生理论(endosymbiosis theory)清楚地表明了共生体的创新能力有多重要。她指出, 真核细胞(具有细胞核和细胞器的细胞)出现于大约16亿至18亿年前, 是两种更简单的单细胞生物(细菌或古细菌)共生的结果, 其中一种生物融入另一种生物, 成为“能量工厂”。这种新型细胞构成了后来整个动植物王国的基本单位。马格里斯在著作《共生星球》(1998年)中指出, 共生是所有进化的驱动力。

人类的进化也与创新共生密不可分。人类非常擅长合作和分工, 因此创造了高度复杂

的社会和文化。人类也像其他生物一样, 与其他物种紧密共生。在这里, 我们不仅要考虑到我们与狗(人与狼相互驯化的结果)和牛等家畜的共生关系, 还要考虑到我们体内数十亿的细菌和病毒, 它们对于消化、免疫系统以及胎盘和新皮质的胚胎发育都是不可或缺的。事实上, 你不能再把生物视为属于不同的物种; 相反, 它们是各种基因和新陈代谢过程的临时性交汇点。人类与技术(工具、机器和信息技术)的共生一直是人类进化过程中至关重要的创新因素。虽然这些人工制品没有生命, 但它们确实具有能动性, 带来了新的生命形式。因此, 农业技术和文字将史前智人转变为现代人, 而人类与人工智能的共生似乎也代表着人类进化的一个新步骤。

共生之所以成为生物学中如此重要的因素, 第二个原因是, 与新达尔文主义和新自由主义固守竞争的观点相比, 共生为进化和革命性发展提供了更广泛的解释。事实上, 共生研究根据共生对相关“共生体”的效用区分了不同形式的共生。除了互利共生, 即对共生双方都有利, 还有共享共生(commensal symbioses, 对一些共生体有益, 对另一些共生体中性)、寄生(对一些共生体有益, 对另一些共生体有害)、补偿(对一些共生体中性, 对另一些共生体致命)和稀缺资源竞争(对大多数共生体有害)。

马格里斯和化学家詹姆斯·洛夫洛克(James Lovelock)率先提出了“盖亚理论”(Gaia theory)。据此, 地球可以被视为一个巨大的“整体生物体”(holobiont), 一个由数以百万计的生物和非生物元素组成的动态系统, 它就像一种恒温器, 维持着地球生命所需的生态平衡。然而, 不同类型共生体的存在清楚地表明, 与新纪元的信徒所声称的相反, 盖亚并不仅仅是一个充满关爱的“地球母亲”, 用马格里斯的话说, 她就像古典悲剧中的美狄亚一样, 还杀死了自己的孩子。

### 以多种方式作弊的“CheatGPT”

在这种背景下, 现在该如何解释 ChatGPT 和相关的生成式 AI 系统呢? 其实, 我们面对的是寄生性技术共生体, 它们不仅利用了人类的思维, 而且在这一过程中还以多种方式作弊。我这里所说的“作弊”, 并不是指你不仅可以向 ChatGPT 寻求写作建议, 还可以让“作弊机器人”为你写出整篇论文。

生成式人工智能系统在更根本的层面上也具有误导性, 因为无论其构造多么巧妙, 它们都不能像人类、海豚甚至细菌那样被称为智能系统。ChatGPT 只是一台“句法机器”, 它在将单词和单词部分转换成数字后, 除了生成关于最有可能出现的下一个单词的统计预测, 什

## “传承发展”系列谈

# “五福”: 幸福生活的系统化诠释

郑玄所强调的是, “五福”皆源自上天, 似乎与天子没有直接的关联; 林之奇将“五福”视为天子治理的结果, 否认上天赐予“五福”的传统观点; 王鸣盛主张“康宁”“攸好德”由天子主宰, 而“寿”“富”“考终命”适用于天子和臣民, 应当同样是来自于上天的赐予; 黄忠慎注意到, 天子的治理是上天降下“五福”的必要条件, 但“五福”却是上天的赏赐, 其源头依然是上天。在这四种解释中, 无论是倚重上天, 亦或是凸显天子的重要性, 都与《尚书·洪范》的文本语境有所不合, 只有综合上天、天子才是理解“五福”之源的基本视域。换句话说, 《尚书·洪范》虽然主张天子直接主宰着“五福”, 但天子本身则应当知道“彝伦攸叙”或实现上天意志的常理。箕子的回应表明, 上天在废黜纣作为天子的同时, 又将实现上天意志、确保民众繁衍生息的禹确立为天子, 并通过赐予后者“洪范九畴”来达成“彝伦攸叙”的目的。禹取得“洪范九畴”的故事意味着, 天子的“德”一方面体现为具备治理人类社会的能力, 另一方面也需要凭借上天的眷顾, 赐予其天子的权力来加以确证。相较于天子, 《尚书·洪范》中的民众并非直接面对上天, 而是把服从天子的意志当作拥有“德”的标志。从《尚书·洪范》的“皇极”开篇来看, 箕子提出“洪范九畴”的来历与纣治水失败

### “五福”的适用范围和配享资格

就其适用范围而言, 《尚书·洪范》中的“五福”是包括天子和臣民在内所有人都共同追求的人生目的。由于《尚书·洪范》中的天子同样扮演着“福”之赐予者的角色, 据此会给人造成一种印象, 即天子应当是凌驾于“五福”之上或游离于“五福”之外。但从《尚书·洪范》开篇来看, 箕子提出“洪范九畴”的来历与纣治水失败

## 数字化

生成式 AI 工具席卷全球, 由“生成”引发的科技革命和产业变革正在重塑世界经济, 影响未来生活。生成式人工智能在涌现出强大能力的同时也产生了巨大的争议, 其利与弊引起激烈的社会辩论。我们应该如何理解生成式人工智能并评估对立的价值观?

# 幻影魔法: ChatGPT 的寄生性

鹿特丹伊拉斯谟大学哲学系荣休教授 约斯·德·穆尔

上海师范大学历史学系特聘教授 陈新/译

要实现人类与人工智能的互利共生, 就必须想办法在不侵犯版权、不剥削知识工作者、不歧视和不破坏环境的情况下制造这些模型, 并避免出现幻觉和“模型崩溃”。

么也不做。ChatGPT 既不懂语义学, 又不懂语用学: 它不赋予数据任何意义, 除了在用户屏幕上显示输出结果, 自己不做任何事情。只有人类和人工智能这两个共生体才能共同构成一个智能系统。只有用户解读屏幕或智能手机上的一系列字符, 并对它们进行操作时, 这些字符才获得意义和目的。由于 ChatGPT 本身只是无意识地鹦鹉学舌, 语言学家艾米丽·本德(Emily Bender)称这种语言机器人为“随机鹦鹉”。然而, 这个诙谐的描述对鹦鹉来说是一种伤害, 因为根据最新研究, 与语言机器人不同, 鹦鹉可以理解并在上下文中使用它们学舌得到的许多单词, 甚至还能组成新的单词和单词组合。

我们在此讨论的不只是“人工智能”一词的误用, 问题还在更深层次。由于句法机器不会为生成的答案附加意义, 也无法区分事实和寓言, 因此无法保证答案的准确性。语言机器人经常会“幻觉”出错误和无意义的文本。2023年, 巴塔查里亚及其同事对 ChatGPT 生成的医学文章中参考文献的真实性和准确性进行了研究, 结果发现在115篇参考文献中, 47%是编造的, 46%真实但不准确, 只有7%既真实又准确。精神病学家罗宾·埃姆斯利(Robin Emsly)也曾报告过类似的经历, 他认为“幻觉”(hallucinations, 错误的感知)一词在这里用错了地方, 因为这些都是捏造和篡改的。

已发现的错误有几个原因。这些问题的部分原因是输入的内容无意义、错误和/或存在偏差, 部分原因是数据中训练语言模型会放大存在的偏见。另外一个问题是, 神经网络对于设计者来说也是一个黑盒子, 很难发现错误。此外, 捏造和篡改也会不断累积。事实上, 许多由 ChatGPT 生成的文本都会通过互联网重新回到语言模型中。最近的研究表明, 由于这种反馈过程, 语言模型不可避免地随着时间的推移变得越来越“笨”, 而不是越来越“聪明”。他们开始无休止地重复使用相同的数据, 这迟早会导致“模型崩溃”。研究发现, ChatGPT 4 在某些领域(如数学)的表现明显不如3.5版本。舒迈洛夫和他的团队最近的研究

表明, 从2023年3月到6月, 这方面的准确率从97.6%降至2.4%。

在与 ChatGPT 的交互中, 用户仍然会产生一种难以避免的错觉, 以为自己是在与一位智能人交谈, 这是因为界面将巨大的数据工厂隐藏在“屏幕背后”, 用户无法看到。神经网络会在瞬间就准备好全部答案, 文本一字不差地显示在用户的屏幕上, 让人产生错觉, 以为自己是在与人打交道。在这方面, 生成式“聊天机器人”让人想起著名的图灵测试, 即你坐在屏幕前向坐在另一个房间的人或软件程序随机提出各种问题。图灵在《计算机与智能》(1950年)中指出, 如果一台“智能机器”能在至少五分钟内让至少30%的审查者相信他是在与人类打交道, 那么它就通过了测试。图灵估计, 要达到这一目标还需要50年左右。

这个预测并不差。自1990年以来, 休·勒布纳每年都会组织一次比赛, 聊天机器人尝试让评委相信它们是人类。尽管这种智力测验有一定的局限性, 但直到2014年, 还没有聊天机器人成功通过测试。但就在这一年, 俄罗斯聊天机器人“尤金·古斯曼”率先通过了测试。评论家认为, 这主要是因为一个小把戏: 聊天机器人把自己伪装成了一个13岁的男孩, 而这个男孩对世界的了解还不多, 英语只是他的第二语言。一些评论家认为这是违规行为, 尤金·古斯曼实际上并没有通过图灵测试。不过, 也许这么做正是图灵测试的意义所在。图灵在《智能机器》(1948年)中指出, 智力基本上是一个情感概念。该测试主要测试欺骗对方的能力。从“CheatGPT”经常被不加批判地使用来看, 我们可以断定这个聊天机器人毫不费力地通过了图灵测试。

### 警惕在进化中走上一条自寄生的道路

当然, ChatGPT 显得如此人性化, 主要是因为输入的大多数文本都是由人类撰写的(至少到目前为止是这样)。毕竟, 语言模型是用数百万个人类创作的文本训练出来的。因此, 作弊机器人与其创造者之间的共生关系就显现出来了。在导入现有文本时, 会大规模地侵犯版权, 像维基百科和ComonCrawl(一个包含40万亿字节数据集的开源网络爬虫项目)这样的非营利性出版物会被无耻地掠夺和推销, 从互联网上获取的许多输入内容都是有毒的。此外, 就像生产和交易加密货币一样, 创建和使用语言模型也会留下巨大的生态足迹。从生态学意义上讲, ChatGPT 也是一种有害的寄生虫。

“人工智能”一词在 ChatGPT 和其他生成式聊天框中显得尤为神秘。人类作者提供并

控制输入, 在此基础上, 模型生成信息, 同样, 这些信息只对人类有趣、有用或相关。人工智能本身并不智能, 就像一本书一样。“智能”一词来自“inter”(之间)和“legere”(选择)。智能意味着能够在各种选择中做出选择。人工智能需要有目的的行动、理性的决策以及与环境有效互动的能力。人类和其他生物之所以有智慧, 是因为它们有一个有机体, 这个有机体通过本能、感觉和情感促使它们与环境互动, 而它们对这些本能、感觉和情感是有认识的, 甚至是有意识的。智能不仅仅是一个感性的概念, 它需要情感的激励, 无论好坏, 情感都会促使我们去做智能的事情。

此外, 人们不仅与他们所处的环境(Umwelt)进行实用性互动, 而且还生活在一个有意义的世界(Welt)中, 这是一个承载着他们的文化, 并作为一种背景不断地隐性存在着的世界。这种背景也使我们能够在不同的智能活动之间切换。此外, 人类不仅可以重新组合现有知识, 还可以发明新事物。人工智能做不到这些, 它们本质上是人造人的延伸, 有许多用途, 在日益复杂的世界中可能会变得越来越不可或缺。

将复杂的思维过程“外包”给人工智能, 这就像将记忆“外包”给文字一样不可避免。当年狩猎采集者的小规模文化转变为以城市化、劳动分工、科学和技术为特征的农业社会时就是如此。人工智能也可以成为一种有用的工具, 但要实现人类与人工智能的互利共生, 就必须想办法在不侵犯版权、不剥削知识工作者、不歧视和不破坏环境的情况下制造这些模型, 并避免出现幻觉和“模型崩溃”。然而, 有人担心, 生成式人工智能将独立于人类, 进化成与人类智能相当的“人工通用智能”, 然后再进化成超越人类的“超级智能”, 这种担心是毫无道理的。我们今天所知的人工智能什么也不知道, 什么也不想要, 什么也感觉不到。不排除这样的可能, 在人类与技术的共同进化过程中, 有一天会出现一个智能超过智人的半机械人。埃隆·马斯克已经开始在人类大脑中植入芯片。但是, 声称基于纯粹句法重组的生成式人工智能将做到这一点, 主要是为了转移我们对上述围绕生成式人工智能的寄生行为的注意力。无论是认为人工智能将拯救人类的乌托邦式幻想, 还是认为人工智能将毁灭人类的乌托邦式恐惧, 都是毫无根据的。后者只是因为寄生虫依赖于它们的宿主生物。生物学家爱德华·威尔逊(Edward O. Wilson)曾尖锐地指出, 毕竟, 寄生虫是捕食者, 它们吃掉猎物的分量小于1。没有人类, ChatGPT 也将一筹莫展。

我这样说, 并不意味着像 ChatGPT 这样的“真实存在着的”生成式聊天机器人不会带来危险。我担心的不是人类会被高级智能体淘汰, 而是人类会越来越适应和依赖低级人工智能体, 从而在进化中走上一条自寄生的道路。

### 《尚书·洪范》中的

“五福”开启了从合理的政治、道德实践活动来思考幸福或好生活的全新视域, 为春秋战国时期实现理性的突破、人文精神的觉醒奠定了坚实的理论基础。

说明的是, 以“得”解释“德”的用法应当有其特殊含义, 即由于敬畏天和先祖, 遵循分封与宗法制度而获得尊贵的政治地位。这种对于“德”的理解不仅与《尚书·洪范》“皇极”中的用法较为一致, 而且也可以把“攸好德”视为是一种特殊的“福”。或者说, “五福”中的“攸好德”, 一方面是配享“寿”“富”“康宁”“考终命”的前提条件, 另一方面其自身就是“福”, 亦即喜好得到尊贵的政治地位。

在《尚书·洪范》的文本语境中, “五福”展现了中华民族在春秋初期之前对幸福或好生活的理解: 把上天预设的公正的审判者, 并将其视为人类幸福或好生活的终极根源; 每个人的幸福或者好生活都与治理天下的实践活动密不可分, 同时以促进人类社会和谐有序作为个人配享幸福或好生活的资格; 幸福或好生活是一种混合型的善, 其中既有“寿”“富”“康宁”“考终命”所代表的“活得好”, 又包括“攸好德”所象征的“做得好”; “五福”与“六极”相对, 表明幸福或好生活有其限度。《尚书·洪范》中的“五福”尽管依然没有完全摆脱天命神学的影响, 然而却开启了从合理的政治、道德实践活动来思考幸福或好生活的全新视域, 为春秋战国时期实现理性的突破、人文精神的觉醒奠定了坚实的理论基础。

出自《尚书·洪范》中的“五福”, 即“寿”“富”“康宁”“攸好德”“考终命”, 不仅总结了“福”的具体条目, 而且也涉及到“福”的来源、适用范围、配享资格等诸多议题。同甲骨文中“福”的用法相比, 《尚书·洪范》中的“五福”蕴含着中华民族早期关于幸福或好生活的系统化诠释, 堪称中华传统幸福观的开山鼻祖。

### 明确了幸福的来源

从词源学的角度来看, 甲骨文中的“福”是由“示”“畀”所构成, “示”指称神灵, “畀”表示酒器, 二者组合起来的意思应该是: “古人以酒象征生活之丰富完备, 故洒酒于神为报神之福或求福之祭”。也就是说, “福”在甲骨文中被用作祭祀的名称, 包括了神灵的庇佑, 人们生活富足、以酒报答神灵的庇佑等内容。其中, 神灵的庇佑应该是“福”得以可能的前提条件, 而《说文解字》用“佑”解释“福”的观点也正与此相契合。值得注意的是, 尽管甲骨文中“福”字借助质朴的形式呈现出自身的原初含义, 即人们的富足生活来源于神灵的庇佑, 然而与之相关的幸福观却并不完备, 因为它回避了“福”的适用范围、配享资格、具体条目等理论问题。

《尚书·洪范》则强调天子同样扮演着“福”之赐予者的角色。在“洪范九畴”的叙事语境中, “五福”被放在第九畴或最后一类来加以说明。针对这种现象, 以郑玄为代表的观点认为: “此五者皆是善事, 自天受之, 故谓之福”; 林之奇提出: “是以自五行至于庶征, 各得其序, 则斯民归于五福矣, 天之所界, 而实自于造物者向而与之也”; 王鸣盛的注解是: “天下之福, 极皆君身所致, 故人平安, 人皆好有德, 即为入君之福。然则寿、富等皆兼天下臣民而言, 不专指君身”; 黄忠慎则评论道: “王道行, 天降五福, 反之则降六极矣”。就其实质而言,