

数字智能的通用迷思

王天恩

摘要 以 ChatGPT 和 Sora 等为标志,人工智能全面步入数字智能发展阶段。数字智能基于统计模型通过大数据纳入人类语境/世界情境进行自然语言和视频处理,获得了在既定大数据范围内甚至超越人类个体的自然语言和视频处理及超强泛化能力,促成了人工智能通用性的空前发展,引发了生成式人工智能是否步入通用人工智能轨道甚至已经通用化的观点对立,关于数字智能的发展由此陷入通用性和通用化构成的通用迷思。数字智能的通用迷思根源于其通用性极度扩展,而人工智能的“通用性”和“通用化”是两个不能通约的概念,即使数字智能通用性无限叠加,也不会走向人工智能的通用化,其中涉及两种不同范式的根本转换。作为信息编码层次的人工智能,数字智能的意义不在于人工智能的通用化,而在于从自然语言到广义语言、从大数据纳入人类语境到纳入世界情境的通用性发展。

关键词 数字智能; ChatGPT; Sora; 通用性; 通用化; 人工智能

中图分类号 TP18 **文献标识码** A **文章编号** 1672-7320(2025)04-0046-08

基金项目 国家社会科学基金重大项目(24&ZD320)

随着大数据的发展,人工智能研究的联结主义进路如鱼得水,以 ChatGPT 为典型标志,人工智能大模型迎来了爆发式发展。ChatGPT 等大模型的通用性空前扩展,具有越来越强大的功能,以致人们认为 ChatGPT 已经进入人工智能通用化轨道,甚至把是否有理解能力尚且存疑的大型语言模型看作通用人工智能(AGI)。由于人工智能的发展空前凸显了其通用性与通用化的微妙关系,ChatGPT 的发布及之后大模型的迅猛发展,触动了越来越多人敏感的神经,造成了对人工智能大模型理解的重重迷雾。由于认为其中有些环节还难以解释,人们甚至认为这些大模型含有人类尚未把握的原理。在人工智能通用性与通用化之间考察人工智能大模型,其在人工智能发展中的性质就会清晰呈现;而这必须建立在对通用人工智能的理解进一步深化的基础上。在人工智能发展的更深层次探析个中奥秘,无论对避免人工智能研究通用化进路锁定甚至锁死,还是对人机关系理解的深化,都具有重要意义。

一、ChatGPT:通用人工智能的正途还是岔道

在关于 ChatGPT 等大型语言模型的理解中,关键问题是其与通用人工智能的关系。正是对此关键问题的理解,业界存在完全不同甚至相互对立的观点。ChatGPT 面世以来大型语言模型的惊艳表现,使人们联想到通用人工智能的到来。ChatGPT 究竟是不是已经意味着通用人工智能? 正是对这一问题的不同理解和回答,形成了两种对立的观点,而且两方面都涉及人工智能领域的顶尖专家。

由于基于至为关键的变换器(Transformer)架构,在生成内容的意义上,ChatGPT 的确意味着新的范式转换。不少观点由此认为 ChatGPT 已经步入通用人工智能轨道,有的甚至认为 ChatGPT 等已经是通用人工智能。鉴于 ChatGPT 的超强功能,包括 OpenAI 在内,业界有不少专家认为 ChatGPT 已经敲开了通用人工智能的大门。在其最新著作中,库兹韦尔就认为大模型人工智能正迈向通用人工智能,“像

Gemini 和 GPT-4 这样的顶级 AI 系统正在扩大其在许多不同领域的能力——这是走向通用智能的鼓舞人心步骤”^[1](P9)。颇具权威性的微软 ChatGPT 研究报告也认定 ChatGPT 是 AGI 的“早期版本”^[2](P1)。在 2023 年 10 月的一次访谈中,Google 研究院院士兼工程副总裁布莱斯·阿圭拉·阿卡斯和斯坦福大学以人为本人工智能研究所的计算机科学家、杰出教育院士彼得·诺维格就认为,通用人工智能(AGI)对不同人有不同的意味,但它最重要的部分已被当前一代先进的大型语言模型实现,如 ChatGPT、Bard、LLaMA 和 Claude。尽管这些“前沿模型”有诸多缺陷,但几十年后它们将被视为第一个真正的通用人工智能(AGI)例示。清华大学智能产业研究院(AIR)院长、中国工程院院士张亚勤在自己的专访中也认为,ChatGPT 开始有了通用人工智能的雏形:ChatGPT 是第一个可以通过“图灵测试”的智能体,他看到一个通用人工智能的雏形。实际上,ChatGPT 甚至被很多业界人士直接定位为“通用人工智能”。

ChatGPT 的出现和大型语言模型令人眼花缭乱的发展,使一些专家将 ChatGPT 理所当然地视为通用人工智能,以致可以看到“ChatGPT 作为通用型人工智能(AGI)产品”的表述,认为“其区别在于它的智能化具有像人一样的思维水平、心理结构甚至人性特征,有如科幻电影的场景照进现实”^[2](P159-160)。ChatGPT 发布后不久,微信公众号“战略前沿技术”2023 年 4 月 9 日发表了《超级智能的涌现:KK 的失控与 Urban 时刻》一文,其观点甚至认为:ChatGPT 已经具有超级智能出现的征兆,ChatGPT 这种神奇的进化趋势,又一次唤醒人们,机器智能已经进入超级智能产生的初期阶段,人类面对超级智能将何去何从,是不得不要深谋远虑的事情了;更确切地说,2023 年 ChatGPT 的出现,预示着人类开始进入一个超级智能出现的转折加速期,这个加速期的一个重要特征就是人工智能开始向超级智能跃迁,以致由此得出“现在人工智能已经陷入失控的竞赛,开发者也无法理解、预测及可靠地控制”的失控结论。埃隆·马斯克、史蒂夫·沃兹尼亚克和约书亚·本吉奥等于 2023 年 3 月 22 日发表的呼吁暂停人工智能研究至少 6 个月的公开信,在不少人那里似乎成为如临大敌的根据。与此相反,另一种观点则认为,ChatGPT 不仅不是通用人工智能,而且还没有进入通用人工智能研究的进路。作为 Meta 首席 AI 科学家,杨乐昆甚至断言,“在通往更强大人工智能之路上,大型语言模型是一岔道(off-ramp)”^[3]。在 ChatGPT 登场之前,他和杰克布·布朗宁在博古睿研究院的电子版思想杂志 *Noema* 发表了《人工智能和语言的局限》一文,认为目前的根本问题不在于人工智能,而在于语言的局限性。一旦人们放弃关于思想与语言之间联系的旧假设,就会清楚地看到这些系统注定只能拥有肤浅理解(shallow understanding),永远无法接近人类所展现的真正思维。最近,研究人员系统评估了 7 个最先进的模型得出结论:“尽管它们在各种任务中很有用,但当前的人工智能大模型缺乏人类那样的自然语言理解。”^[4]随着生成式人工智能的发展,关于以 ChatGPT 为标志的大型语言模型到底是通用人工智能的正途还是岔道,观点的对立越来越尖锐。

对 ChatGPT 与通用人工智能的关系之所以有完全不同的理解,重要原因在于:随着大数据驱动的人工智能发展到 ChatGPT 等大型语言模型,仅在技术层面,其进一步发展就越来越难以理解和把握。据腾讯网 2023 年 3 月 30 日报道,当谈到千名从业者联名叫停大型 AI 研究时,西蒙·坎波斯就明确表示,AI 系统的发明者不确切知道它们是如何工作的,也不知道它们是什么,因而无法管理系统的风险,发明者有能力做到,但不知道如何限制 AI 的行为;微软首席科学家埃里克·霍维茨则认为,关于已经出现的人工智能,有些事情是已知的,有些是未知的,中间则是为我们观察这个世界留下的大门;索菲·布什维克等甚至认为,没人真正知道 ChatGPT 还能做什么,即使是目前的版本,它是如何做到的仍是一个开放的科学问题。由于对通用人工智能理解的局限,关于 ChatGPT 与通用人工智能的关系,目前的认识实质上还处于未确定状态,关于这一点,来自 OpenAI 的思考最能说明问题。在 OpenAI 博客 2023 年 2 月 24 日发表的最新文章《规划面向通用人工智能(AGI)及其后的未来》中,萨姆·奥特曼一方面认为,随着系统越来越接近 AGI,人们对模型的创建和部署将越来越谨慎;另一方面他同时承认,第一个 AGI 将只是智力连续体上的一个点,进步很可能会从那里继续下去,可能会在很长一段时间内保持过去十年中看到的进步速度。这里的“第一个 AGI”应该类似 ChatGPT 的更高级版本而不是类人智能意义上的真正通用人工智

能,这点在文章结尾的一段话中可以看得更清楚,奥特曼认为,人们希望为世界贡献一个能带来如此繁荣的 AGI,在他看来,创建 AGI 似乎需要大量的计算;如果说其中的“似乎”只是在一定程度上表明不确定性,那么,“一个”通用人工智能的表述则肯定无疑地意味着其对通用人工智能理解存在深层次问题。

关于目前对人工智能理解的这种观点对立,本身完全可以理解,这与作为人工智能直接基础的信息研究密切相关。正是信息理解涉及的范式转换,构成了当前人工智能发展理解的天花板。由于信息本身的科学和哲学一体化性质,人工智能发展到大数据驱动的 ChatGPT 等大型语言模型时,关于其理解的科学和哲学一体化要求就提到了更高层次,以致对其核心机制的理解本身都成为问题。进一步深入信息层次理解人工智能当前的发展,个中问题便可以清楚看到。

二、作为数字智能的人工智能大模型

要更到位地理解 ChatGPT 进而人工智能的通用化进路,必须深化人工智能、大数据和信息的哲学与科学的一体化研究。

关于信息的长期曲折探索表明,信息既不是物质也不是能量^[5](P132),而是感受性关系^[6]。对信息的感受性关系理解,使信息编码及其与信息的关系得以清晰呈现:信息编码是感受性关系的物能化和观念化。因此,信息编码相应有两种基本类型:信息的物能编码和信息的观念编码。前者典型如电信号和 DNA 中的基因,它们分别是信息物能编码中的生物编码和物理编码;后者典型如作为自然类(natural kinds)概括产物的概念和符号,比如“苹果”“红色”等概念和自然数等符号。

信息的观念编码是我们既完全陌生又非常熟悉的领域,因为观念的研究不仅具有悠久的历史,而且积累了丰富的成熟理论,但又没有与信息编码关联起来。事实上,作为观念体系的人类知识正是在信息观念编码特别是概念编码基础上建立起来的。信息的概念编码至关重要,正是概念编码构成了人和动物的根本区别。概念研究之所以没有与信息概念编码相关联,一个重要原因是绝大多数概念都不是信息的概念编码,只有在感受性关系建立过程中信宿对信息的概括才是概念编码。比如“红”“绿”“蓝”等是对具体个别事物建立的感受性关系的概括,属于信息的概念编码,而在其基础上进一步抽象概括得到的“颜色”和“色谱”等则只是在信息概念编码基础上推导出的更高层次概念,而不再是信息的概念编码。正是在信息概念编码基础上层次越来越高的概念推导,人类才得以建立知识体系。由此可以看到,作为人类创造的产物,人工智能的发展正是在人类智能的基础上首先发展出人类知识层次的人工智能,专家系统正是其最高层次成果;然后从人类知识进一步深入作为其基础的信息编码,发展出信息编码层次人工智能;由此进路可以想见,如果人工智能再进一步深入作为信息编码基础的信息,就将触达信息层次的人工智能,也就是类人机器智能^[7](P39-48)。在人工智能发展的三个层次中,可以大大深化目前对人工智能发展的理解。

在信息的观念编码中,信息的符号编码是最抽象的形式。信息的符号编码又有一种特殊的方式,那就是信息的数字编码。其之所以特殊,就在于它既可以是作为信息观念编码的符号编码,又可以是作为信息物能编码的物理编码。作为数字编码,0 和 1 是信息观念编码中的符号编码;而电流的开和关又是信息物能编码中的物理编码,因此,数据不是信息本身而是信息编码。大数据正是信息数字编码发展的产物。目前,大数据驱动的人工智能大模型正是信息编码层次的人工智能,由于发展到信息数字编码层次,因而也被理解为数字智能,但其更准确的概括当是数码智能,即信息数字编码人工智能,由此可以在作为通用智能的人类智能进化过程中,深化人工智能通用化的理解。

从作为感受性关系的信息层面可以清楚地看到,作为通用智能,人类智能建立在作为感受性关系的信息基础之上。通用人工智能必须深植于作为感受性关系的信息。只是扎根于信息编码,机器智能处于信息编码的天花板之下,永远不会拥有自己的感性经验和相应语境,不会有自己的理解能力,从而不会有类人意义上的真正通用人工智能。由此反观生物智能的进化,也可以看到重要佐证。

在生物进化过程中,只是建立在机体感受性关系基础上还不会产生生物智能,但机体感受性对智能产生至为重要,没有机体感受性就不可能有感官感受性,也就不可能有人类智能。在人工智能发展过程中,不基于作为感受性关系的信息,只是建立在信息编码的基础上不会有真正意义上的类人智能^[8]。由此可见,仅仅通过大数据的机器学习数字深耕不可能走向人工智能的通用化,因为这只是限于信息编码层次,通用人工智能的发展必须进一步深入作为数据基础的信息。由此也可以看得很清楚,由于没有深入到信息层次,尚未根植于感受性关系,ChatGPT 完全不在通用人工智能轨道上。具有类人理解能力的通用人工智能必须有自己基于整体观照机制的语境基础和理解机制,因此,必须有自己的信息观念编码以及在此基础上建立起来的观念体系——正是由此涉及信息范式的根本转换。

ChatGPT 及进化中的 GPT 只是基于大数据,明显是信息编码层次的人工智能,完全不在通用人工智能发展的正途上,不可能走向通用人工智能。一方面,作为信息数字编码发展的产物,大数据及其利用的发展只能带来信息编码层次的人工智能,不可能进入人工智能通用化。这些只是数据和参数规模的扩展,不可能孕育通用人工智能核心机制。作为统计语言模型,ChatGPT 及其发展只是由参数增加实现的:增加到 1750 亿个参数,ChatGPT 功能爆红,参数再增加几个指数级别,功能还会进一步提升,但模型的参数再怎么增加也只是趋向一个极限,永远不可能走向类人人工智能。另一方面,在所有人工智能的通用化设想中,都是首先创造出单一一个通用人工智能,但真正的通用智能一定是类智能,类智能一定具有社会性,而超级智能不仅同样具有社会性,而且其社会性层次只会更高;不可能出现单独“一个”超级智能,不是在类群中“长大”从而获得类群亲历^[9]的单个通用人工智能不可能出现,这样的人工智能不可能是类人意义上的通用人工智能,而只是泛化能力不断增强、通用性不断扩展的专用人工智能。人工智能大模型的灾难性遗忘,正是其不具有类群亲历性的体现。在神经网络模型中,机器学习在训练一种功能后再去训练另一种功能,会把前面学习的结果给抹除,这正表明通过机器学习并不能使智能体得到作为一个不可分割整体的经历,更不用说形成类群亲历性。这些结论可以在机制层面得到进一步论证。

从机制上看,仅仅在信息编码层次或局限于符号主义和联接主义进路,不实质性涉及行为主义进路,不可能实现通用人工智能,即使将机器人与大型语言模型关联在一起,也并不意味着行为主义进路的机制涉及。仅由此就基本可以断言,ChatGPT 还完全没有接近通用人工智能的核心机制,二者间具有极限趋近性质,但永远不可能趋同。大型语言模型的确意味着人工智能研究的一个重要范式转换,但局限于模型大,正是(也只是)局限于信息编码层次人工智能发展的逻辑,它意味着一种极限。极限不能在所处既定范式中超越,要超越极限困境,只有通过更根本的范式转换。由于这种更根本的范式转换尚未明确,关于通用人工智能的实现,奥特曼心里自然也非真正有谱,在《规划面向通用人工智能(AGI)及其后的未来》的博文中他写道:AGI 可能发生在不久的将来或者很远的将来;从最初的 AGI 到更强大的后续系统的起飞速度可能是缓慢的或者快速的。这种说法反映的通用人工智能理解程度,相当于看上去很不靠谱但具有合乎实际一面的预测:通用人工智能的突破在未来 5-500 年之间。关于通用人工智能的这种预测,正意味着人们对其核心机制还没有基本的理解,不确定性由此而生也就完全可以理解。

事实上,即使处于引发争论的漩涡中心,业界关于 ChatGPT 的发展与实现通用人工智能之间的距离,观点也充满不确定性。据知名微信公众号“图灵人工智能”2023 年 4 月 8 日报道,伊利亚一方面认为,从现在到实现 AGI,这个窗口还将持续很长一段时间,另一方面又说,他认为生成式模型的范式还会走得很远,不会低估它的潜力;但这种确切的范式,也可能不完全是实现 AGI 的唯一方法。在他那里,更根本的实际上还在于,他很难准确地说出下一个范式将是什么,但它可能涉及整合所有过去出现的不同想法。由此可以感觉到,人们对大型语言模型和通用人工智能核心机制之间关系判断的某种不确定甚至不乐观,由此也可见人们对通用人工智能理解的局限。正是由于通用人工智能理解的局限性,人工智能的通用性发展构成了数字智能的通用迷思。随着以 ChatGPT 为标志的大型语言模型和以 Sora 为标志的文生视频模型等的发展,关于数字智能的通用迷思越来越纠缠不清,因为其功能越强大,往往意味着这

方面的迷雾越深,因为其功能越强大,具有魔性的硅基智能独特优势越得以凸显。

正是以 ChatGPT 为标志的大型语言模型,空前体现了硅基智能可以利用大数据而具有的独特优势。由于这一优势在语言领域的展开,大型语言模型的通用性正通过其泛化能力不断扩展,而人工智能的通用性和通用化究竟是一种什么样的关系,则是一个颇有些微妙且亟待澄清的重要问题。

三、数字智能通用化的通用性误解

无一例外,认为 ChatGPT 已经步入通用人工智能轨道的基本根据就是 ChatGPT 的通用性大大增强,这也是根据 GPT-4 等更高层次产品的通用性进一步提高,OpenAI 认为自己已经是第一个通用人工智能的原因。ChatGPT 的有些能力超越人类智能毫无疑问,其实机器都具有在某些方面超越人类相应能力的性质,只是当发展到通过大数据纳入人类语境处理自然语言达到了模拟人类理解语言的更高整体层次,给人以意想不到的类人感时,才怂恿人得出通用人工智能降临的结论,但这并非在机制上意味着其自我迭代升级可以直接走向人工智能通用化,通用人工智能不是局限在作为信息编码层次的人工智能可以实现的。最近 GPT-4 Turbo 和 GPT-4o 的上线及 OpenAI o1 和“草莓”(strawberry)模型的推出,当进一步强化对数字智能将走向人工智能通用化判断的支持,但最近的研究亦表明,人工智能大模型的参数量越大、版本越新,其回答的准确度越高,而随着问题难度逐步增加,其回答的准确度有所下降^[10](P61-68)。有研究甚至得到这样的测试结果:将 AI 生成的数据输入一个模型,会导致后续几代模型退化直至崩溃。在一项测试中,关于中世纪建筑的原始文本到了第九代时变成了这样一段字符串:除了是世界上最大的黑@-@尾兔、白@-@尾兔、蓝@-@尾兔、红@-@尾兔和黄@-@尾兔的栖息地之一外……^[11](P755-759)显然,这里涉及人工智能通用性和通用化的深层次原则区别。正是这一区别的根本性,对于确定 ChatGPT 等数字智能的通用性扩展并不导向人工智能通用化至为关键。深化关于人工智能通用性和通用化及其关系的理解,成了更到位地理解当前人工智能发展的重要课题。

在适用范围大幅扩展的意义上,由于在语言领域特有的泛化能力,以 ChatGPT 为标志,数字智能通用性的扩展具有极为重要的意义,但“通用性”和“通用化”是两个密切相关又具有根本不同涵义的概念,其中涉及至为微妙的重要关联。就大型语言模型而言,人工智能通用性的发展具有特殊意义,而在人工智能通用化上却又有特殊局限。因此,在讨论 GPT-3 等大型模型时,自然语言处理专家、斯坦福大学教授曼宁使用的就是“更通用形式的人工智能(a more general form of artificial intelligence)”^[12](P127-138)的表述,用“更通用形式”描述泛化能力空前的大型语言模型,应当更为严谨和确切。将 ChatGPT 的定性理解具体落实到语言处理,诚然比认为 ChatGPT 敲开了通用人工智能大门的判断要有根据得多。

虽然对概念的理解存在不同观点,但作为通用化发展的产物,通用人工智能原则上说是类人智能,具有共识的关键点是可以像人那样从事各种任务。通用人工智能定义的类人含义表达得很清楚:可以在所有或绝大部分有经济价值的任务中达到或超越人类的全部认知能力的机器智能。而人工智能的通用性则意味着只能泛化到更广的特定领域,类似从 AlphaGo 发展到可以在所有棋类游戏中通用的 Alpha Zero——正是在这个意义上,确切地说,包括 ChatGPT 在内的大型语言模型仅仅是 OpenAI 研究人员 2019 年提出的“通用语言智能”(General Linguistic Intelligence)。只是由于就人类而言,作为人机交流甚至人机软融合的人类最佳方式,作为涉及面最广的领域之一,自然语言处理的通用性最容易与人工智能的通用化相联系。任何工具都有一个通用性的问题,但人工智能的通用化却不是通用性发展的结果。通用性的极限扩展并不能趋近通用化,二者具有不可通约性。人工智能通用化和通用性的不可通约性,可以在作为感受性关系的信息与作为其物能化和观念化的信息编码之间找到根据。正是在作为感受性关系的信息层次可以看得更清楚:在人工智能发展中,无论是基于信息的物能编码还是观念编码,都与最终基于作为感受性关系的信息具有根本范式区别。由于信息编码的性质,在人工智能的通用性发展中,大型语言模型有巨大发展空间,从而可以为机软融合进化开辟广阔天地。但大数据驱动的大型语

言模型的通用性发展不可能进一步深入作为感受性关系的信息层次,从而超越信息编码天花板走向通用化,直接通过数据处理发展出通用机器智能。由人工智能通用性的叠加永远得不到人工智能通用化,在根本上几乎就像由点的相连永远得不到线,由线的拼接永远得不到面,由面的叠加永远得不到体。

ChatGPT表明的人工智能通用性与通用化间的不可通约性,源于数字智能的根本机制局限于信息编码层次。正是局限于信息编码而未能进一步深入作为感受性关系的信息层次,构成了ChatGPT等大型语言模型发展的信息编码天花板。通用人工智能意味着类人智能,自身没有理解能力的ChatGPT距此不是一个量的距离远近问题,而是天花板上下的质的根本区别。ChatGPT的通用性具有天花板之下的自然语言泛化能力,Sora及数字智能的后续发展甚至可以在人与物理世界间架设数字化桥梁,但不可能突破这一天天花板达至人工智能通用化发展。由于人工智能通用性与通用化有不可通约性,ChatGPT及其代表的人工智能研究进路不可能通过其通用性扩展走向通用化。在人工智能的发展中,大型语言模型的成功在于人类语境的大数据纳入、机器学习深耕和统计语言模型利用。作为人工智能在语言领域通用性向最高层次发展的标志性产物,ChatGPT代表的大型语言模型与通用人工智能有更基本范式的不同:从信息编码层次人工智能发展到信息层次机器智能。不了解人工智能通用性和通用化的根本不同,就可能出现对通用人工智能研究的误解误导,甚至通过将ChatGPT与超级智能相联系的观点,这一点可以看得更清楚。正如将通用性和通用化混为一谈,认为ChatGPT已经是超级智能发展的初期阶段的观点,仍然是把机器智能超越人类智能的单方面能力看作全面超越人类智能的超级机器智能,只是由于所处层次更高,因而更具迷惑性。通用人工智能的核心机制完全不同于数字智能,其通用化不仅不是通用性可以通达,甚至连智商都不足以作为判断人工智能是否具有通用智能核心机制的根据。

据测试,GPT-3.5的智商只有48,GPT-4达到了96,已经接近普通成年人。据估计,2025年AI的智商可达145,2026年AI的智商将达到160以上,人类智商达到140以上就是天才了,这意味着AI已超过爱因斯坦的智商。由此推导人工智能的通用化在即,OpenAI计划2027年实现通用人工智能突破似乎不成问题。事实上,智商只适用于推理等能力的测试,并不足以作为通用人工智能测试标准。由此将智商达到人类水平作为实现人工智能通用化的根据似乎是以偏概全。将智商48的GPT-3.5与6-9岁的人类儿童相比较就一清二楚:GPT-3.5在很多方面能力大大超过人类,却不具有6-9岁人类儿童所具有的常识,更不用说根本的智慧了。GPT-3.5的专用人工智能性质,与人类儿童具有通用智能同样明显。正如通用性怎么发展都不可能走向通用化,人工智能大模型的智商无论有多高,都不构成通用人工智能的质量互变关系。量增导致质变是在两个层次范式间的发展规律,而在同一范式上下,量增并不能带来质变。

总之,由于硅基智能不同于碳基智能的特性,专用人工智能特别是数字智能无不具有单方面超越人类智能的能力,连计算机的计算能力也完全碾压人类;但也正如人工智能通用性的发展并不能构成机器智能的通用化,人工智能单方面超越人类智能能力的叠加并不构成其通用化发展,更不用说走向全面超越人类智能的超级机器智能。由此造成的误导不仅不利于人工智能的正常发展,还可能导致人工智能研究的路径锁定甚至某种程度上的锁死,其引发的进一步严重后果之一,则是可能以错误的方式触发人类新的安身立命焦虑。

四、人工智能通用化与人类新的安身立命问题

正是由于认为ChatGPT意味着通用人工智能的出现,具有内容生成能力的数字智能发展带来了当前的现实忧虑。认为ChatGPT已经进入通用人工智能发展轨道的直接原因,事实上与对人工智能当前发展的理解渐失把握密切相关。对数字智能发展的忧虑,最根本的原因还是缺乏对人工智能发展的更高层次整体理解。为人工智能的健康有序发展提供一个更高层次人工智能观整体观照,必须对其发展有一个更高整体层次的把握,这不仅关乎通用人工智能发展正确进路的选择,而且涉及人类新的安身立命问题及其深入理解和合理应对。出于对人工智能大模型通用化发展的忧虑,OpenAI上层由于不同观

点而发生了震惊人工智能界的“宫斗”，这与数字智能的通用迷思密切相关。人工智能发展到当下，哲学上关于“加速主义”的研究发展到人工智能有效加速主义与有效利他主义的分歧和冲突。在 OpenAI 的“宫斗”中，更倾向于有效加速主义的奥特曼主张快速推动人工智能的商业化和社会普及，而作为有效利他主义的代表性人物，伊利亚则强调在推进人工智能发展中更充分考虑其社会和伦理影响，因此，首先要进行与人类价值观的超级对齐(Superalignment)。

“有效加速主义”和“有效利他主义”的冲突绝不仅仅是关于人工智能对其研究的态度不同，客观上涉及对数字智能发展前景的判断。撇开涉及双方的其他复杂因素，冲突的发生与关于数字智能通用化的判断不无关系，对避免不当触发人类新的安身立命焦虑具有启示意义。由伊利亚和其老师在这方面的反应，既可以看到对人工智能研究的负责任态度，也可以发现值得进一步讨论的问题。

出于对人工智能目前发展的担忧，有人工智能教父之称的杰弗里·辛顿离开谷歌，以利于发表他的一些重要观点，从而成为人工智能最新的“末日预言者”。这既与其关于人工智能发展的长期判断，也与其特殊处境和人类责任感密切相关，这兴许也是将诺贝尔物理学奖颁发给这位机器学习大神的一方面的考量。由于人工智能通用性和通用化的各种方式混同，也由于人工智能运行与人类智能完全不同的速度，身临其境的人工智能龙头企业研究人员更容易出现担忧情绪。这既是人类责任心的突出表现，也是促进人工智能通用化发展进程研究的重要动力。在辛顿的学生伊利亚那里，可以看到人类新的安身立命问题与当下人工智能发展更切近的关联。

ChatGPT 通用性的空前扩展，虽然使 OpenAI 顶尖专家也认为大型语言模型研究已经进入通用人工智能轨道，但即使是作为有效利他主义代表性人物的伊利亚，同时对此也仍然有清醒的认识。作为 ChatGPT 诞生时期的首席科学家，伊利亚虽然对通用人工智能到来的威胁极为敏感，但他仍然不仅认为离实现通用人工智能的窗口期还很长，而且认为 OpenAI 目前追求的是大型语言模型的泛化能力，以致把追求更强的泛化能力看作通用人工智能的本质，因此，他认为大型语言模型的泛化能力越强，其智能水平越高。模型智能程度(泛化能力)=模型泛化效率×训练数据规模，AGI 基础模型的本质是实现最大有效数据集的最大程度无损压缩，由此足见伊利亚的有效利他主义举动与其对人工智能当下发展的判断密切相关，可见关于人工智能通用化发展的理解之于人类新的安身立命问题的重要性。

如果将 ChatGPT 的“泛化能力”理解为其通用性的扩展，就意味着其不仅并未步入通用人工智能轨道，而且还不在于走向通用人工智能研究的正确进路上：从专用人工智能到通用人工智能的发展绝不是通用性量的累积的结果，而是具有颠覆性的范式转换的产物。在信息层次可以清楚地看到，数字智能还没有也不可能涉及人工智能通用化研究范式的根本转换，只是人类必须及早做好必要的准备。对人工智能的发展特别是其通用化来说，时刻保持警觉的确关乎人类未来命运，但准确把握人工智能的发展却对人类现实生活越来越重要。在人工智能发展过程中，深入把握其通用性发展与通用化的关系，对人机关系理解的深化甚至人工智能观的合理化，都具有重要意义。

诚然，人工智能的通用性有不同程度，从只能下围棋到可以下其他棋类是其通用性发展的典型表现；由于涉及与人类具有特殊关系的语言，在广义语言领域，大型语言模型的多模态发展更是人工智能通用性扩展的重要维度。但大模型在任何单一具体领域的通用性扩展，都不构成人工智能的通用化。作为类人意义上的通用化，通用人工智能绝不只是能做更多不同的事。人工智能通用性程度之所以无论怎么提高都不可能走向通用化，个中关键在于：由于不具有类人理解能力，人工智能通用性程度的提高有上限，只能局限于工具意义上的通用性。这里涉及对通用人工智能研究有重要意义的关键点，从中不仅可以对大型语言模型基础上人工智能通用化发展更深层次进路有更到位的理解，而且对数字智能和人类发展的关系有更清晰的认识。只要是人工智能通用性的发展，就属于人类高级工具的发展，都不涉及人类新的安身立命的问题，只涉及人工智能与人类价值观的对齐；当人工智能进入通用化发展，对人机关系的把握就必须进入更深层次。无论在哪些种情况下，包括对人类新的安身立命问题在内，人工

智能的通用化与通用性发展间的关系,都是人工智能发展同时也是人类发展越来越重要的研究领域。

参考文献

- [1] R. Kurzweil. *The Singularity Is Nearer: When We Merge with AI*. London: The Bodley Head, 2024.
- [2] 黄楚新,陈伊高. ChatGPT:开启通用型人工智能的数字交往. 中国传媒科技, 2023, (2)
- [3] Percy Liang. Generative AI: Large Language Models' Ability to Generate Text Also Lets Them Plan and Reason. *The Economist*, 2023, (22).
- [4] Vittoria Dentella1, FritzGünther, Elliot Murphy et al. Testing AI on Language Comprehension Tasks Reveals Insensitivity to Underlying Meaning. *Scientific Reports*, 2024, 14(28083).
- [5] Norbert Wiener. *Cybernetic, or Control and Communication in the Animal and the Machine*. Cambridge : the MIT Press, 1985.
- [6] 王天恩. 信息及其基本特性的当代开显. 中国社会科学, 2022, (1).
- [7] W. Tianen, W. Xi. *Information as Receptive Relation*. London, New York: Routledge, 2024.
- [8] 王天恩. 人工智能通用化及其实现途径. 中国社会科学, 2024, (3).
- [9] 王天恩. 人机交会:人工智能进化的类群亲历性. 上海师范大学学报(哲学社会科学版), 2023, (1).
- [10] Lexin Zhou, Wout Schellaert, Fernando Martínez-Plumed et al. Larger and More Instructable Language Models Become Less Reliable. *Nature*, 2024, 634(8032).
- [11] Ilia Shumailov, Zakhar Shumaylov, Yiren Zhao et al. AI Models Collapse When Trained on Recursively Generated Data. *Nature*, 2024, 631(8022).
- [12] Christopher D. Manning. Human Language Understanding & Reasoning. *Daedalus*, 2022, 151(2).

The General Myth of Digital Intelligence

Wang Tianen (Shanghai University)

Abstract Marked by advancements such as ChatGPT and Sora, artificial intelligence (AI) has fully entered the stage of digital intelligence development. Digital intelligence, by processing natural language and video upon statistical models through integrating vast amounts of data into human contexts and world scenarios, has acquired the processing capabilities that, within the scope of the given big data, can even surpass those of individual humans, along with an extraordinary capacity for generalization. This has led to unprecedented advancements in the generality of AI and sparked debates about whether generative AI is on the path to artificial general intelligence(AGI), or even has already achieved generalization, thus creating a generalization myth surrounding the development of digital intelligence. The myth stems from the extreme expansion of its generality, while the concepts of AI generality and generalization are fundamentally incommensurable. Even if the generality of digital intelligence continues to grow, it will not lead to the generalization of AI, which will involve a fundamental shift between two different paradigms. The significance of digital intelligence, as an AI operating on the level of information encoding, lies not in the generalization of AI, but in its development of its generality, which spans from natural language to language in broader sense and from incorporating big data into human contexts to encompassing world scenarios.

Key words digital intelligence; ChatGPT; Sora; generality; generalization; artificial intelligence

-
- 作者简介 王天恩,上海大学智能时代的马克思主义研究中心、智能哲学与文化研究院教授,上海 200444。
 - 责任编辑 涂文迁