

人具备不可计算的特质

许纪霖：自DeepSeek开源以后，人工智能的发展日新月异，几乎每一两个周就有新的模型推出。在这样的背景下，我们距离超越人类的AGI(通用人工智能)还有多远？

我个人的猜测是，人工智能的下一步突破，是拥有身体。因为只有拥有身体，才能拥有直觉能力，才会有真正意义上的想象能力和原创能力。英国大思想家卡尔·波普尔将人类的世界分为三种：世界一是物理世界，目前AI还不具备感知能力，世界二指的是无法与人的身体分离的直觉和悟性，而现在人工智能所运用的知识，则属于世界三，也就是可以与人分离的理性知识。只有世界三，通过逻辑推理对已有的理性知识进行提炼概括，并且上升为AI所拥有的理性认知境界，是有天花板的。真正的突破离不开世界二，而世界二的直觉和想象能力，又必须拥有肉身，对世界一有感知能力。

正如黄仁勋所说，人工智能的下一个阶段，就是具身AI。脑机接口技术的普遍运用和仿真机器人的出现，将让人工智能具有对物理世界的感知能力，并且发展出他的直觉、想象和悟性。

肖仰华：从工科角度来看，人类的很多创作本质上是字词、图像、音符等内容元素的随机拼贴。从这个意义上讲，人工智能可以从一个更大的内容选择空间，或者说更大的创作空间中，以人类意想不到的方式进行组合或者说创作。

在这些组合里面，往往就有我们认为优秀，具备灵感与灵性的内容，有些内容甚至是所谓富有想象力与创造力的产出。文章本天成，妙手偶得之。优秀的内容从可能性的角度来说早已存在在那里，只不过通过某个作家的灵光一现再现了这个作品。今天艺术创作过程的主体进一步换成了AI。所以，未来生成性、创作性的工作，AI只可能比人类做得更好，更加胜任。

但是，它的作品到底好还是不好，还是完全由作家自己来判断。它的作品是不是符合你当下的情感需要？是不是符合当前人类社会的境遇情况？是不是能够引起某个群体的共鸣、共情？这些依旧是由人类个体或者群体来判断、评价与选择。

回到问题本身，AGI是否会到来？这个问题有个先决条件：AGI是什么？人们之所以对此感到困惑，恰恰是因为AGI缺少一个公认的定义。刚才许老师试图给出一个定义：AGI是超越个人或者整个人类总体的人工智能。这是一种定义。

如果我们把AGI定义成是超过我们人类智力水平的AI，而人类智力水平又是基于人类各种考试与测试所检验得出的，那么在这个意义上可以说AGI早晚会到来。

但是反过来说，人类是不是应该因此感到恐慌呢？其实一切所谓的AI到来所造成的恐慌，都是因为人工智能让大家觉得人类没用了。换句话说，AI在冲击旧的人类价值体系。事实上，未来人仍有其价值。我们真正需要尽快开展的工作是重新定位人在AI时代的核心能力，重建人类的价值体系。



AGI离我们有多远

去年年初DeepSeek的问世吸引了世人的目光。这场技术革新不仅关乎个人与家庭，更影响着国家乃至世界的未来，也让我们看到了通用人工智能(AGI)的曙光。华东师范大学历史学教授、中国现代思想文化研究所副所长许纪霖，复旦大学中文系教授严锋，以及复旦大学计算机科学技术学院教授、上海市数据科学重点实验室主任肖仰华，围绕DeepSeek及AI技术革新带来的应用与影响展开了深度对谈。本次对谈内容已收入《此心安处是吾乡：AI时代的文化认同》一书，以下为书中摘选，以飨读者。

控制论鼻祖维纳写过一本书叫《人有人的用处》，在当下AI大发展时代，用其书名来提振人类对于自己的信心再恰当不过。人相对机器而言，有很多“不可计算的心智特性”是独特的。有趣的事实是，这是DeepSeek帮我总结归纳生成的术语。如果仔细思考DeepSeek的运行过程，其本质仍然是一种计算过程，是一种基于复杂电磁规律的物理过程。但是，人和人相遇则会发生很多化学反应。荷尔蒙、多巴胺分泌之后，我们会体会到如沐春风、心潮澎湃等情感体验。正像泰戈尔所说的，我们每个人都应该活成一道光，照亮身边的人。

这样一种人与人之间的化学反应，对人际关系的影响，是机器不可代替的。人类身上还有很多不可计算、难以计算的特性是机器无法代替的。相信要不了多久，机器人就能走进家庭、料理家务、护理老人。但是机器再精细而准确的服务，也代替不了来自儿女的对老人的一句深情问候；机器再如何完美的情感陪伴，也代替不了来自同类的一次爱抚与拥抱。这就是人不可计算的特质。

从这个意义上讲，AGI永远实现不了这些功能。它或许能在一切由物理过程所决定的方面，远远超过人类；但在一切需要化学反应的地方，无法比肩人类。那么又有人会说，我们可不可以给机器也植入一些类似多巴胺的化学要素？比如，有一些公司已经在设计仿生材料的人形机器人，或者反方向研究对生物神经元的算法操作，等等。如果这样的实验和设计成真，那么未来AI可能就会成为新的生命物种，就像《西部世界》描绘的那样。对于这类物种，我们可能要赋予它与人类对等的地位。

许纪霖：这在法律上有禁令吗？

肖仰华：非常好的问题。在我看来，这一步是不应该迈出去的，除非你想搅乱这个世界。这是非常危险的一步，在某种意义上就是造人。

但是，如果我们把时间尺度拉长，我们还要有一种担心：我非常同意许老师提到的身体的重要性，但身体同样是一种禁锢，使灵魂禁

锢在身体内，并且身体是会灭亡的。所以，我们才会去寻求有限生命的无限意义。人类追求意义的源头是强烈的自我意识，正是有限、短暂的身体激发了人类这个物种存在、生存和延续的内在需求，进而产生了强烈的自我意识，然后有了物我两分，并产生各种纷繁的意义。可以说，整个人文和社会学科都建立在这一基础上。

大家可能会设想这样一个问题：除了拥有身体之外，AI还需要具备哪些条件，才可能诞生人类水平的自我意识？这实际上是非常值得去严肃思考的。意识的诞生条件究竟是什么？这是科学家、哲学家尚未能够清晰回答的问题。我认为可能有几个条件：第一是要有身体；第二是要有群体，也就是某种智能体要有一定的规模；第三是要去进化，要在环境中去试错，去反馈和进化。智能本质上是由环境塑造的。目前地球上只有人类这个物种达到了高水平的意识，有些动物也具备了初步的、一定的自我意识，但是要形成人类水平的意识还需要依靠上述几个条件：要有身体，要有群体，并在一个漫长的时段去进化。通过漫长的、为了生存而奋斗的实践，一个物种才可能激发出较高的意识水平。

其实今天的AI发展形式已经非常接近人类发展的激发方式。DeepSeek R1复杂思维能力是通过强化学习算法习得的，是在可能的思考空间进行探索、试错，以求得尽可能优的解决方案。这种智能的习得机制非常接近人类智能的进化机制。R1通过强化学习甚至自发发现了新颖的反思模式，也就是所谓的“Aha moment”。这就好比人类的顿悟时刻，经过苦思冥想后突然灵感迸发。整个R1的学习过程完全是基于强化学习的自主学习，人类唯一的干预只是设定强化学习过程中对思维方案好坏判定的基本规则。我们通过设定一个简单的判定目标，AI就能自主找到最优的思考路径。将来，这个来自外在的设定完全可以简化为让AI智能体尽可能生存和延续的设定，就好比科幻作者阿西莫夫为机器人设定的三定律。从长远来看，人工智能或许有可能进化到接近某种生命的形态，因此对于其进化过程中是否

会发展某种程度的意识水平是需要严肃研究与监管的。

未雨绸缪方能行稳致远

严锋：关于人类不可计算的心智特性，我其实是有一点怀疑的。因为现在有一些人工智能的应用，已经演变为心理治疗师、抚慰者，甚至作为一个虚拟的伴侣。更令人吃惊的是，现实中很多人沉迷于跟ChatGPT对话，甚至觉得人工智能所扮演的角色，要比现实当中的人更好。我们原本以为心灵是人类最后的防线，但现在人工智能可以说比我们更了解我们自己，这对于人类的心灵性也是一个挑战，技术发展监管上的挑战。

许纪霖：在ChatGPT出现以后，扑面而来的技术加速主义和与人类对齐的对齐主义的大争论。因为开源，在AI技术飞速发展的当下，AI的破坏力也将同时变得更大。道高一尺，魔高一丈。我们应该怎么去看待AI的破坏力？在未来AI普遍开源的背景之下，又应该如何处理技术与人类对齐的平衡呢？

肖仰华：自AI问世的第一天起，关于人工智能加速与对齐、发展与监管的讨论就从未停止。

有媒体称借助DeepSeek，原本会耗费人类10年的烦杂政务工作，一天就做完了，这是生产力的巨大提升。归根结底，DeepSeek让最先进的大模型技术平民化、普惠化了。让老百姓用上了最先进的模型，享受到了AI带来的便利，这当然会让整个社会感到欣喜。

但同时，我们也要认清AI可能带来的各种问题，做好相应的防范和治理工作。有些人认为，所谓AI风险不过是危言耸听。但事实上，应用AI带来的问题，已经发生在我们的日常工作与生活中。比如各位专家学者辛辛苦苦发表论文、出版著作，但AI可能在未获得你授权同意的情况下直接学习其中内容。我最近就发现，AI关于某些问题生产的内容，和我曾经写过的文章观念非常相似，但它重新组织了内容形式，避免出现大段或连续的原文内容，规避了抄袭和侵权的指控。对此，作者们只能有苦说不出。

另一方面，AI也可能抓取一些网络上的隐私信息。比如一些教师在教学中将自己的个人信息分享给AI，结果被AI抓取并被大模型学习，使得任何普通用户都可以通过提问接触到这些隐私信息。这些问题已经在发生，只不过我们可能还没有意识到。

或许有人会说，AI发展速度会慢下来。我的预测是AI发展只会加速。因为AI研发与大规模应用有政府与资本推动。作为新兴技术革命的代表，AI浪潮是时代发展趋势，是不可阻挡的，关系国家竞争，决定民族兴亡，是不容错失的重大发展战略机遇。同时，资本逐利的本性，势必要求对AI产业的大规模投入得以变现。AI的大规模应用，像水、电、通信网络基础设施一般的大规模应用，势在必行。

在这一过程中，很容易出现AI滥用。AI大规模滥用带来的问题，包括教育升级、经济结构转型、人力资源市场调整，甚至社会关系革新都需要时间。人类社会的一切上层建筑的调整都是缓慢的，经不起震荡，唯有温和变革才能维系全社会稳定。AI的超速发展和人类社会结构的缓慢调整之间的矛盾，怎么去调和？这都是我们需要去思考的问题。

如果我们不想深度思考并积极应对这些问题，AI就可能给我们人类带来难以承受的损失。一旦AI带有破坏力的一面被放大，那么AI本身也无法持续发展。AI的可持续发展能力，才是一国AI时代最核心的竞争力。

在我看来，开源进一步放大了发展与监管之间的矛盾。开源让先进大模型人人可用。人人可用就意味着有不良用心的人也可以使用先进大模型进行欺诈、造谣、作假，甚至制造武器，危害社会。如果只是把矛盾磨得越来越尖锐，盾不够厚实，那么社会就一定会失衡。所以，面对越来越尖锐的矛，我们一定要铸造越来越厚重的盾，才能做好AI的可持续发展。

或许会有人说，我们完全可以先发展再治理。互联网的发展，我们就是如此得以实现的。然而，AI不同于互联网。AI是以全面发展人类水平的心智能力，并取代人类的同等水平所能胜任的大部分人类工作为目的的。AI大模型滥用可能会造成积重难返的困境。事实上，发展AI最大的风险，恰恰是大部分人尚未意识到它的风险，而只关注它作为先进生产力的那一面，忽略了其破坏性的那一面。正确的态度是在大力发展AI先进生产力的同时，以前所未有的积极姿态未雨绸缪，做好应对与预防。

(本文摘自《此心安处是吾乡：AI时代的文化认同》，内容有删节，标题、小标题为编者所加)



《此心安处是吾乡：AI时代的文化认同》
许纪霖 刘擎 等著
浙江人民出版社