

人工智能连续论反思与存在论探析

人工智能连续论反思与存在论探析

2026-01-04 作者：杨庆峰 来源：《中国社会科学》2025年第11期P149—P164

字体放大

字体缩小



微信公众号

— 分享 —

摘要：对于人工智能发展的理解存在着一种观点，即把狭义人工智能、通用智能和超级智能程。但实际上，人工智能发展过程并非完全连续，而是如同有裂缝但又未完全裂开的岩石，对断裂。通过工具与主体、知识与存在、个体性与集体性以及综合性等多重角度，能够洞察其内为三种类型：技术断裂、哲学断裂以及存在论断裂。技术断裂即人工智能从工具到智能体的变级智能体成为知识主体；存在论断裂即作为生存论问题的导致因呈现“存在的升级”与“存在能作为新的能动体和他者，使这些断裂充分展现了出来。

关键词：人工智能 超级智能 连续论 存在论

作者杨庆峰，复旦大学科技伦理与人类未来研究院教授（上海 200433）。



人工智能连续论指的是人工智能发展的本质特征，包含类型、水平、实现和终极阶段四个方面。第一，即人工智能类型可以概括为从狭义人工智能到通用智能再到超级智能，其实质是工具的持续演进，具再到协作工具。第二，智能水平连续提升，即机器智能内在逻辑体现为从模拟人类到比肩人类再到超其核心是机器智能的人为扩展。随着神经科学的日渐成熟和算力数据的支持，机器智能在逐步增强，尤成上增强连续性。第三，通用智能连续实现，即通用智能目标的实现过程类似攀爬梯子的过程。“放梯的硬件；踏板1是可编程的计算机；踏板2是可存储程序的计算机；踏板3是机器学习；踏板4是深度学习；踏板6是指令微调模型。”第四，超级智能是智能体连续进化的终极阶段。智能体连续进化过程存智能的出现，其实质是两种实体智能的博弈。可以说，连续论图景无比诱人，通向超级智能的路径也看

人工智能连续论反思与存在论探析

与集体性以及综合性等多重维度考察断裂现象。从工具与主体维度看，一旦机器能够与具头世界进行交互，工具性规定将彻底瓦解。从知识角度看，非人类具身经验的知识将成为可能，这完全不同于人类基于具身从存在角度看，随着通用智能乃至超级智能的出现，一种新的存在形式将得以出现——它既不同于生硬的硅基生命，而是在交互中获得自身的规定性。从个体性与集体性角度看，未来的超级智能可能并不或以个体形式存在而以群体智能及多智能体的形式表现自身。从综合性角度看，超级智能具有一种无征，当前的法律难以将其囊括。超级智能会实现新的“飞跃”，带给人类的存在论问题将从“存在的飞跃”。

一、人工智能的连续论图景

当前，一个流行的观点是，人工智能是辅助人类的工具，但不能替代人类。这一观点可称为人工上，人工智能工具论的观念并未得到深入考察。海德格尔《技术的追问》呈现了一个追问技术工具论并谓技术是人类的活动和满足人类需要的东西。本文将这一观念作为考察的起点。在此基础上，连续发展动成为人类学工具论的重要观念。但这一理解的前提是人类中心主义，而在对人工智能乃至超级智能的论解释是否依然有效？人工智能工具论又提出了怎样的哲学规定？笔者认为，将人类学工具论延伸到人是行不通的。人工智能工具论有着一些不同的规定性。

首先，人工智能工具论预设了工具性能的不断提升。如果用人类学工具论来理解人工智能，一个更符合莱特林格（Edmund Reitterling）的观念，即人工智能是“更好的工具”。从性质来看，它通过“更好”；从功能上看，它通过泛化走向了通用，一个智能体可以做很多事情。所以，“更好的”规定连续性。随着人工智能工具的不断完善和持续进化，它的效率获得了稳步提升。

其次，人工智能工具论预设了“以人为中心”的伦理的稳定。“更好的工具”除了从工具本身理与人的关系角度来理解。“更好的工具”可以做到符合人性化。比如，“以人为中心”的提出意味着人能可能产生超出一般工具的影响。“消灭人类”“替代人类”是目前人类的焦虑和担忧。辛顿（J. Hiri果智能体能够从物理环境中学习，那么就会产生真正意义上的数字生命，从而导致人类的灭亡。大模型自动化、模型的无监督学习，共同使人类的很多工作结构受到影响——人类开始被替代。此时产生的问如何自处等伦理担忧有关。此外，不同于“以人为中心”，还有一种观念强调“以人为目的”。在这一能工具都要服务于人这一目的。这种观念还可以延伸到作为主体的人工智能，使其依然践行“人是目的学工具论解释人工智能的另外一个作用，就是为连续论提供合理辩护。

最后，人工智能工具论有着规范伦理的立场。李飞飞等一些学者认为，“人工智能是辅助工具”，规范伦理的特点。在人与人工智能的关系中，“人类起主导作用”是一种应然状态与描述状态并存的命题。人类起主导作用，人类根据自己的意志做工具的决定，而使用到何种程度则是由不同场景语境决定。但是，现在最大的问题是把使用人工智能工的伦理行为，而忽略了其内在的实然特性。所以“人类使用人工智能”作为应然行为时，强调的是人类产生过度依赖；作为实然行为时，强调的是根据场景、具体情况应用人工智能。

字体放大

字体缩小



微信公众号

— 分享 —

人工智能连续论反思与存在论探析

稍许维度的技术头现，而具身智能则体现了物质维度的技术头现。第一种争议已淡出人们的视野，第开。连续论观念将人工智能发展看作连续的过程，可分为四种类型。

一是智能类型连续演变观念。该观念突出了不同类型的人工智能之间呈现一个连续的过程。拉曼人将人工智能区分为不同类型：弱人工智能、人类水平人工智能、类人智能、通用智能、强人工智能以拉曼的划分未免过于琐碎和牵强。之所以琐碎，是因为他把人工智能划分为六种类型，其中弱人工智能一对划分，但他在其中塞入了人类水平人工智能、类人智能、通用智能三种类型。之所以牵强，是因为比如，类人智能与通用智能之间实质上区别并不那么明显。此外，也有学者提出不要增加类型，认为通有共同的基础与机制，二者之间不存在本质区别。在笔者看来，“多模态走向通用智能”的观念是一种作多模态论者的一种信念。基于多模态的通用智能只是走向超级智能的三条进路之一（另外两条进路能）。类型连续论观点有着明确的时间线，技术类型多半是建立在不同的算法、模型积累的基础上。随的进步，人工智能发展如同河流的分支一样，演化出不同的类型。

二是智能水平连续增强观念。该观念强调人类与智能体的关系遵循从模仿到对齐再到超越的基本逻辑，狭义人工智能包括弱人工智能、人类水平人工智能和类人智能。弱人工智能即专用智能，可以高效需通用认知能力。通用智能则能够执行人类可以做的任何智能任务，并且具有跨领域的泛化能力。拉仿”两个阶段强化了人工智能的逻辑连续观念。超级智能是一种特定类型，根据超级智能的比较解释，域超越人类智能，如创新能力、决策能力等。逻辑连续观念是符合常识又符合进化的观点。从常识角度的加强，狭义人工智能会逐渐进入通用智能和超级智能阶段。从进化角度看，不同类型的人工智能是算果。

三是通用智能连续实现也是人工智能特别是通用智能理解中的一个设定。“人工智能历史显示出长。”尽管在通用智能理解上存在诸多争议，对于真正的通用智能能否实现、可能具备哪些特征，科技，但在具备通用目标和自动化上却存在共识——具备泛化的能力，并在目标实现上存在一致性，能标。

四是机器智能持续进化并超越人类是超级智能的显著特征。很多学者根据人工智能的广度和深度来尽管学者们普遍认为，超级智能远未实现，只具有想象实在。

连续论有很强的叙事特性，它将人工智能发展史看作连续过程，从叙事角度看是科学化叙事的结果中的叙事本质，从而使人工智能的断裂性暴露出来。“连续性的这些形式一旦被束之高阁，便打开了整广的，然而又是一个可确定的领域：它是由实际陈述的整体在它们的散落和在各自所特有的层次上构成续论图景就成为必要的任务。

二、从工具和主体反思连续论

哲学通常采用的方法是从经验到先验，“既然我们应当由复合事物领悟单纯事物的知识，由经验物，则我们在学习时从比较容易的东西起步就是恰当的了”。对人工智能连续性的考察也可以这样进行工具角度开始考察。从工具角度看，这种新型工具不断实现包括认知与推理在内的智能增强，但也带

字体放大

字体缩小



微信公众号

— 分享 —

人工智能连续论反思与存在论探析

(一) 人工智能的工具性

在对人工智能的认知中，工具是最为基础的规定。在一定意义上，人工智能也被视为工具性的存在。人工智能工具论观念。根据海德格尔的理解，工具论的规定意味着工具指向人类活动本身以及满足某种过程，人工智能工具论有着新的特征，下面将分析阐述人工智能工具论的四种内涵。

一是无智能的工具。该观点认为，人工智能是一种类似锤子的工具，远未达到人类智能的水平。如器等，都可以看作无智能的工具，无法主动与使用者形成互动。根据这个观点，人工智能对人的辅助作用有限。

二是辅助工具。该观点认为，人工智能只是起到辅助作用，帮助人类决策、执行任务，对于人类本有太多影响。

三是增强工具。该观点认为，人工智能可以增强人类的认识、推理、记忆等能力。这种现象可称为归入分析人工智能问题的方法范畴。比如，大模型的快速发展与广泛应用，显著增强了人们的信息检索、翻译等方面的能力。

四是影响文明的工具。尽管这一观点也基于人工智能的工具性，但更加强调人工智能对文明的影响。抛弃人工智能是不可行的。领导者们将不得不直面这项技术的影响，他们对这项技术的应用负有重大责任。他们需要一种伦理准则，它能够帮助我们理解人工智能时代，甚至在这个时代为我们提供指引。”

以上是人工智能工具论的形式，这四种理解有着逐渐强化智能意义的特征，从无智能、增强智能再到飞飞提出“以人为中心”的人工智能观念，即以人类理性为中心或是以人类尊严和自主性为中心。在工认识的局限在于人类中心主义的设定。如果稍加改造这一观念，变为“以人为目的”，从这样的角度认就发生了改变。人工智能将人作为最高目的，可以做到为了最高目的而奉献自身，这将成为人工智能应用场景中，当超级智能面临人类关机要求时，最好的方式是智能体能够实现自主关机。这种方式比哥德尔哥德尔炸弹会失效。而人工智能将自身献给自己的最高目标时，这种奉献就如同人类的自我牺牲一样，的最大体现。

(二) 通用智能的拟主体性

通用智能与一般的工具有着很大的区别。在对区别的探讨中，人们逐渐发现了人工智能的功能拟主体问题上，很多学者并不承认通用智能的主体性。要承认通用智能的主体性最基本的条件是承认通用智能识。这一点在科学界已有所显现，比如自我感知的人工智能（self-awareness AI）与有意识的人工智能的出现。因此，拟主体性是科学界对人工智能主体性的一种折中态度。承认通用智能的功能拟主体性，能具备理性推理和情感能力。

另外，具身智能的主体性也开始有所表现。比如，DeepMind推出了一项成果，研究如何将基于的视觉—语言模型直接整合到端到端的机器人控制中，以提升泛化能力并实现涌现的语义推理能力。迹数据和互联网规模的视觉—语言任务（如视觉问答）上协同微调（co-fine-tune）的视觉—语言模型称为视觉—语言—动作模型，并实例化了一个此类模型的示例，即Robotic Transformer 2（RT-2）。

字体放大

字体放小



微信公众号

— 分享 —

人工智能连续论反思与存在论探析

取哪个物体作为临时锤子），便会发现这种涌现能力使得智能体能够形成一定程度的主体性，即通过释命令以及抓取何种物体当作锤子等形式体现。

从哲学角度看，这项研究强调了具身智能可以获得涌现能力从而实现机器意识，这相对于单纯使用向前迈进了一大步，它将意识涌现具象化，为机器意识提供了能力涌现的途径。从具身智能质变产生裂，即从技术本身发生的根本改变，其本质是从工具到主体的变革。与技术断裂不同，当新的主体性悄然出现。

（三）超级智能的主体性

当人工智能从通用智能进入超级智能阶段时，“就有可能成长为另一种主体，另一种立法者，或者主体特性表现为意识涌现。意识涌现在具身智能中的可能性已被科学界讨论，可以看作自我意识的条本做到，从互联网规模的训练中获得对新物体泛化能力的显著提升、解释未包含在机器人训练数据中的执行多阶段语义推理等一系列涌现能力。可以看出，这种涌现能力使智能体形成了一定程度的主体性。

如果超级智能的主体性成为可能，接下来的问题将过渡到主体间性，即作为他者的存在。在对超级仓促（缺少论证）的认识是将伊德的人机关系框架（他者关系或异在关系）的活力激发出来。超级智能有很大区别，难以借助现象学的主体间性范畴加以推演和论证。从意向性角度看，机器主体的意识行为成指向特定的对象，这是与人类主体意向不同的地方。从语言意向角度看，大模型的语言并不是意（token）预测的结果。从经验角度看，机器经验与人类经验不同，是来自机器与世界的真实交互。

那么，超级智能有什么主体性特点？在科幻化叙事中，超级智能更多表现为硅基生命消灭碳基生；国拉斯维加斯Ai4会议上提出，超级智能是无法控制的，需要注入母性本能，让其成为母亲才能改变这来说，其进化的目的是更好地存在，人类对其来说是最高目的。根据古德（I. J. Good）的观点，超级设计，之后就是机器的自主进化。为了更好地进化，超级智能与人类之间应构建一种积极的智能契约关由。一是人类与机器之间本质上并不构成互相伤害的关系；二是人类和机器在交互中实现了更好的进化使身体机能和智力水平等得以增强，并且随着社会智能化发展，生产生活方式甚至文明程度都得到了显主体角度呈现的断裂现象，由从工具到主体的转变呈现。对于这一点，李斯认为，人工智能拥有心智、是断裂式的开篇。

三、从知识与存在反思连续论

贡华南从知识类型的角度讨论了知者的存在方式，认为不同的知识既反映了知者与他者之间不同的知者不同的存在方式。那么，人工智能是怎样的知者存在类型？

（一）知识角度的反思

从“人工智能是怎样的存在”这个问题中，我们也可以发现反驳连续论的可能——其中的断裂更存在论的转向。“人工智能一开始是个知识论问题，在不远的未来将要升级为一个涉及终极命运的存在危及人类自身存在的问题。”作为知识论问题，认识主体是基本的问题。“认识主体具备四种相对独立

字体放大

字体缩小



微信公众号

— 分享 —

人工智能连续论反思与存在论探析

一经验的特征。它没有真正的记忆，只是体现了信息存储和编码，无法将这种记忆作为知识的来源和意味着，在人工智能知识生产过程中，并没有记忆起作用，记忆—推理—知识的链条并不完整。当人工以提供各种各样的知识。一般来说，我们期待人工智能提供新的知识。事实上，大模型已经给人们提供有效的知识，在自然科学研究以及人文社会科学研究中，大模型可以给研究者提供便利的方法和新的知超越人类研究者的感性认识甚至经验推理。另一类是机器的幻觉知识，大模型的幻觉问题可能让人类陷从而对大模型产生了信任危机。

从结果上看，人工智能作为知识主体有可能从两个方面颠覆传统认识论，进而引发存在论危机。以数据形式和交互形式表现出来，能够确立新的存在论领域。数据形式意味着机器经验以数据的形式呈意味着这种经验并非无源之水，而是来自人类的经验，但随着增强学习的处理会出现非人化的结果。从的数字生命缺乏一种体验根基，这使得其观点吸睛但无根据。另一方面，无具身经验的知识成为可能身，缺乏感性经验，而机器身体将这种经验建立在数据之上——最初是人类数据，其经验带有明显的随着机器智能从真实环境中学习，尤其是与非物种交互中习得非人类具身的经验，从而形成非人具身无具身经验的知识成为可能，那么这种无具身只是相对于人类身体而言，是对人类身体的超越，是新得。一种新的机器存在论问题在数字存在物的基础上被推到前台。

（二）存在角度的反思

思考“智能体作为知者”意味着思考智能体的存在方式。在思考这一问题之前，需要反思人工智能。当人工智能自身演化为存在论问题时，这是一个影响人类生存、文明存在的问题。但此时，人工智能阶段，而辅助工具、增强工具则难以描述这一变化。存在论的问题体现了一种生存危机和哲学反思。赵的问题归结为存在论问题，即“存在的升级”。他认为，作为超图灵机的超级智能一旦形成，就会导致种技术或制度的发明开拓了新的可能生活并且定义了一个新的可能世界，所以它意味着存在方式的革命的进步。王天恩接受了这一概念，并将其作为人工智能存在论意蕴的内核。他认为，人工智能的发展将致“存在的升级”，从而改变人的存在方式。两位哲学家都强调以“存在的升级”为存在论内核（核心智能的理论框架。

当用存在论范畴讨论人类的生存时，便会碰到生存焦虑这一概念。这个概念并非严格的哲学反思层面的刺激—反应式感受。从感受到反思沉淀是一个飞跃过程，而不是简单的连续结果。尽管生存焦地勾勒出人类面对超级智能时对于自身命运的一种心理反应——当面对可能被超级智能灭亡的结果时生存焦虑迫使人们寻求一种治理办法，以防止超级智能失控，并在超级智能成为技术灾难时能够采取“存在的升级”这一改变，并非可以简单解决的问题。“存在的升级”是给人类生存方式以新的形态，生存。在笔者看来，尼采的超人是一种悖论式的存在：尽管理性精神强大，但是身体孱弱、非理性精神一个弱点。而新兴增强技术导致的科技超人悖论发生了颠倒：身体无限增强，但是精神无比孱弱，保此，从知识论到存在论之间的断裂是非常明显的。另外，在西方哲学叙事中，曾出现过“实体论转向知生存论”等说法。如果将人工智能看作一个叙事过程，那么按照赵汀阳的说法，就经历了从知识论到中并不存在明显的连续性。

字体放大

字体放小



微信公众号

— 分享 —

人工智能连续论反思与存在论探析

生命力。面对数据资源耗尽的问题，伊利亚（Ilya Sutskever）提出预训练即将终结，这一观点只能有识，存在同义反复的嫌疑。因为资源即将耗尽，所以预训练即将结束，也意味着学习的监督被突出。当还是基于数据与智能的关联，足够的数据能够确保智能的涌现。针对这一观点，有学者提出利用合成数据。笔者更倾向于将两种观点结合起来，利用合成数据纵然可以实现进化，但模型崩溃论的观点却为超可能。超级智能将构成人类文明未知的不可预测的危机。如果承认超级智能的主体性和能动性，这一新力和推理能力的他者，将对人类文明构成深层次威胁。如果让人工智能完成人类文明中“去糟粕、留精基所担忧的“复制粘贴”的智能危机才会真正体现出来。

四、从个体性与集体性反思连续论

一般来说，人是个体性与集体性的统一。每一个个体都有其特性和人格，而不同的个体组合起来就并不等同于个体的组合，而是超越个体组合的结果。乌合之众只属于个体的组合（也称群体），通常成。比如，一辆车上的乘客需要借助交通工具去往各自的目的地，而这辆车能够满足这些乘客个体的需成了群体。缺乏凝聚力是乌合之众的重要特征。但对于集体来说，形成凝聚力和认同是前提，并且能——这是集体的重要本质，而且与记忆密切相关。

对于人工智能来说，只有个体性和群体性，而没有集体性和总体性。在西蒙栋看来，技术对象的个式、结构与功能等多个角度加以理解。他从物理对象的个体性分析了技术对象的个体性。对于人工智能从智能体与环境的关联中进行解释。换句话说，一个智能体可以等同于个体。智能个体可以通过结构和 transformer 算法加上决策、感知、交流行动等功能。可以说，智能个体的一个特征是整合了不同的能记忆、推理等。

智能对象的个体性已明显超出传统的技术对象。按照 DeepMind 的观点，传统的技术物品是无智象开始拥有了智能：L0 为非人工智能物品、L1 中人工智能作为工具、L2 中人工智能作为咨询者、L3 者、L4 中人工智能作为专家、L5 中人工智能作为大师。这个划分中，无智能的物品本身就是作为工具智能体还有咨询者、合作者、专家和大师等类型，这是一般技术对象并不具备的，并且这与主体意识也度看，智能对象的群体性不但具有一般技术对象的组合特征，更具有系统性。人工智能作为系统存在可合进来。功能整合是智能对象个体性和群体性的特征。

然而，超级智能却可以体现出某种集体性，而且这是容易被忽略的。约书亚（Bengio Yoshua）到，多个智能体之间的密谋将成为最大的风险来源。“多个智能体”对未来超级智能的多元形态进行了偏重风险维度，想到多个智能体通过密谋来避开人类监管，就会让人感到不寒而栗。但这也恰恰展示了可通过群体甚至集体形式存在，“在群体智能中，分工乃是实现合作的主要形式”。梅宏等人提出，可合”“反馈”三个活动机制来构建群体智能，最终整个群体可能表现出远超任一个体的智能行为。这展智能进化之间的区别，人类精神发展在于把无数个体智慧浓缩为集体智慧，而智能体进化过程也展现出性和集体性角度看，人工智能连续性更多是以同构性表达出来，而群体智能则是超级智能可能的形式，务上表现出独特的总体性。

五、从综合性角度看连续论问题

字体放大

字体缩小



微信公众号

— 分享 —

人工智能连续论反思与存在论探析

在技术问题、社会问题、哲学问题和政治问题四个维度中，人工智能是技术问题并不需要太多解释。人工智能是社会问题，因为人工智能带来了工作替代、角色取代的危机。人工智能是哲学问题和政治问题，前者是因为人工智能带来了存在论问题，后者是因为人工智能需要全球政治合作才能实现向善。人工智能可以看作四位一体问题的综合。但对于超级智能而言，这种四位一体并不是那么自明了，甚至在某的问题。

从技术角度看，人工智能取决于算法、算力和数据。在人工智能界看来，尽管通用智能是一个技术争议。李飞飞认为通用智能是一个模糊的概念，还有一些人认为通用智能是人工智能技术发展的圣杯。从技术角度来说是一个技术问题？这一点显然更具争议。一些公司关注这一维度，但往往只是作为一种技术可。一些人还是把超级智能看成带有未来维度的安全隐患或杞人忧天式的想象的杂糅。超级智能是科幻想象定，是媒介想象中的关注点，同时也是技术想象的过度表现。目前还不能明确说，超级智能是一个技术制和路径并不明确。

从社会角度看，人工智能是一个社会问题，带来了工作取代和角色替代的问题。然而，对于超级智能并不是一个维度的。现有法律体系很严肃地对待了人工智能带来的可能性风险。欧盟的人工智能法案了划分：不可接受风险、高风险、中风险、低风险和无风险。尽管这个法案讨论了通用智能的未来形态排除在外。这也表明，超级智能更多被视为想象对象，而非实在对象。

从政治角度看，人工智能需要全球联合应对，进行讨论和治理。但超级智能却是少数人面对的问题的学者屈指可数，只有辛顿、鲍斯彻姆等发出了零星声音。大多数人认为，这是一个媒介想象、社会

从哲学角度看，情况有所不同。超级智能的哲学性已从两个方面表现出来。一是知识论层面。超级智能新的类型和源头。二是存在论层面。赵汀阳等人认为，超级智能导致“存在的升级”。因此，在四个维政治这三个维度由于这一问题的想象性而存在争议，唯独哲学维度在超级智能的讨论中保持了下来，但题的延续——知识生成与“存在的升级”。

六、超级智能的现象辨析与应对之策

在笔者看来，超级智能既不能视为想象建构对象，也不能视为生存焦虑建构对象，而应视为现实对象。在科学上已明显表现出来，相关问题必须加以反思。超级智能被一些学者视为人工智能进化的必然进化或动物进化的方式实现。如果说人工智能的工具性是人们对它的原初认识，那么超级智能或许已然失主体角度看，超级智能是人类想象意向构成与焦虑意向构成的对象；从客体角度看，超级智能是一个重类最后一个发明，从此之后机器将开启自主设计。

(一) 从主体角度看，超级智能机器可以视为人类想象意向构成的对象和生存焦虑意向构成的对象。超级智能是三种想象构建的产物：从诞生起就带有强烈的科幻想象底色，在发展过程中是过度的技术想象人时是媒介想象的对象。

字体放大

字体缩小



微信公众号

— 分享 —

人工智能连续论反思与存在论探析

激烈对立的矛盾冲突中，因为具作刀叉子创作构图构建一种典型叙事，而激烈对立的矛盾叙事更有可能在题材中，人与机器冲突的敌托邦色彩往往伴随着科幻电影或文学作品。这种冲突并不必然导致人类关注，很多时候只是好奇或娱乐的一种体现。

在技术发展中，超级智能可以视为人类过度的技术想象。笔者曾区分了两类技术想象：合理的技术想象。合理的技术想象指以理性形式表现出来的技术之爱且符合必要的技术逻辑；过度的技术想象指基于行为对于特定技术和文化对象的“胡思乱想”，这种想象产生了一种超出理性的结果。

在社会呈现上，超级智能是媒介想象的重要对象。为了博取社会关注，媒体不惜渲染人类与人工智能未来人工智能是魔鬼、机器人反抗会不会消灭人类等话题，无形之中挑逗、刺激着人们的神经，从而将话题作为一种独特社会记忆加以巩固。

不过，将超级智能看作想象产物，还停留在心理感性层面，缺乏足够的反思。超级智能如果是想象般性想象有什么区别？想象金山、孙悟空和想象超级智能的区别何在？目前来看，只有超级智能是三物，而金山只是个体理性产物，孙悟空是文学想象的产物。进入哲学反思层面，方可对超级智能现象有学层面看，存在两种情况。

一是超级智能由人类生存焦虑意向构成。也许是人们看过太多敌托邦式的电影，对这种警告有点麻木，这种由超级智能带来的危机是模糊的、不清晰的，更多是想象的产物。这种由生存、命运导致的情况可以说，超级智能是生存焦虑意向构造的产物，而且是能够带来灭绝的东西。所以从此看，“存在的升构建的对象，而是智能科技对人类生活方式、存在方式产生的改变，最大的改变是消除了传统的存在方式存在方式。

二是超级智能是人类存在论事件。赵汀阳、王天恩等人强调了“存在的升级”，这是从正面价值看调存在的降幕或人类存在的消亡。当然，这种升级并不是孤立的变化，而是在关系中呈现的变化。未来可以达成一种智能契约关系，以此确保彼此“存在的升级”。所以，在智能契约关系下，“存在的升级

(二)从客体角度看超级智能，可以说它是机器事件的新开端，机器智能超越人类智能具有解决能，但可能会导致生命灭亡和文明重置。

从发展视角看，超级智能是一种比喻，对象进化超越人类进化。因此，超级智能并不是实体，而是个节点，机器智能将超越人类智能。这种比喻是建立在线性时间上，即时间能够无限制地延伸。在无限的速度很快，而人类生物进化基本上停留在一个较高的阶段，因此超越必然会出现。现在想来，超越素。如果从相融的角度看，人类与机器不断融合，就不存在超越与否的问题了，而二元进化比喻忽略势。

从功能角度看，超级智能是实体的功能，对象功能超越了人类的水平。从人工智能功能来看，可者、专家和大师的功能。人工智能在这些基础上可以给人类最权威的信息，能够很完美地与人类合作，至大师级的指引。从经验角度看，超级智能有望解决一些人类社会面临的重大问题，但超级智能最终可如破解人类面对的与生存有关的凭借理性无法解决的“大麻烦”。就像古希腊戏剧中突然引入一个外部

字体放大

字体放小



微信公众号

— 分享 —

人工智能连续论反思与存在论探析

这个问题上的态度基本一致。辛顿强调了一个事实：如果数字生命进入无监督学习阶段，能够在现实环人类文明就面临灭亡的危险了。在无监督学习阶段，数据没有了人类标签。更为重要的是，合成数据会解决自然数据耗尽的问题。文明重置的问题相当于系统重启。重启计算机系统是一种非常重要的维护手统的正常运行状态，还能确保新的更新、配置变更得以生效，并帮助修复潜在的系统问题，是提升系统性的一种必要操作。在这一前提下，人工智能文明的出现将成为可能。那么，对于人类文明系统来说，置机制？在人类文明的发展中，如果与人工智能文明发生冲突，那么通过对话寻求一种协调机制，或者间的平衡，可以发挥非常重要的作用。

当人工智能作为工具物品出现时，关机效应并不存在。人工智能的意义只是人们使用的能够进行些工具失去了日常功能，人们可以自行处置，不需要给这些物品一个交代，只需要扔掉或重新购置新的为新的能动者以及他者存在时，关机效应就会出现。面对人类的关机行为，超级智能能够作出不同的回恐惧反应，可能出现等待接受被关闭的僵化行为。智能体会产生恐惧，使得其行为僵死、等待被处置。出现自我复制和自我保存行为。如果面对关机，智能体可能会进行自我复制，从而采取其他方式延续自改变人类的关机想法。智能体可能会通过各种说服言辞和行为改变人们的关机想法。对于关机思想实验狭义人工智能、通用智能、超级智能之间还是存在一个极大的“飞跃”鸿沟。

如果说人工智能会导致“存在的升级”，超级智能则会导致“存在的飞跃”。赵汀阳、王天恩“级”观念作为赋予人工智能意义的哲学范畴，用来揭示人工智能的存在论意蕴。这在连续性的范围中是来，“存在的升级”指某种技术或制度的发明开拓了新的可能生活并且定义了一个新的可能世界，所以革命。但这个观念在面对超级智能时却失去了作用。超级智能已超越了工具技术层面，而成为一个新的析超级智能时，要抓住生存论问题的这条主线，“存在的飞跃”将是一个更好的概念。祁克果（Søren《恐惧与战栗》中使用了“信仰的飞跃”（leap of faith）概念来说明亚伯拉罕为上帝献上自己的儿子这种信任的行动就是“信仰的飞跃”。如果说人类面对人工智能时采取的是有条件的信任，只是把人工不可全信；那么面对超级智能，或许我们只能采取无条件信任，如同亚伯拉罕对于上帝的信任那般。“畴中“存在的”保留和突出了人工智能引发的存在论问题，但又超越出“升级”的局限。升级强调连革，而“飞跃”恰好能够体现这种跳跃和变革。

结语

从根本上来说，本文是基于未来可能性的哲学思辨分析，目的并非论证通用智能以及超级智能的可是在阐述超级智能这一“看似遥远甚至虚构的未来”对于当下哲学反思的重要性。本文通过工具、知识、性这五种角度对人工智能连续论进行了反思，并在此基础上对超级智能进行了追问。超级智能超越了一新的能动体和他者。在存在论意义上，超级智能不仅带来了“存在的升级”，更带来了“存在的飞跃”争议较多，但存在论视角的探讨对于克服人工智能理解中的短视效应非常必要，并且这也是经过哲学审重要内容。

西蒙栋曾谈到技术与文化之间的二元对立，而人工智能的出现扩大了二者之间的裂缝。当人工智能这种裂痕和以往并没有太大区别，最多只是增加了一个经验的变量。但随着智能体能够与真实环境交互

字体放大

字体缩小



微信公众号

— 分享 —

人工智能连续论反思与存在论探析

以及木米的八上省能及展具有一定的现头意义。

[本文注释内容略]

原文责任编辑：[]

字体放大

字体放小



微信公众号

— 分享 —

社科推荐

跨学科聚焦人工智能多领域应用

大力推动海外红色文献征集整理与研究利用

深化文化自信与文化主体性建设研究

学术赋能农业农村高质量发展

跨界探索传统工艺振兴路径

[关于我们](#)

[广告服务](#)

[网站声明](#)

[网站纠错](#)

[联系我们](#)

举报电话：010-85341520 举报邮箱：zgshkxw@cass.org.cn 互联网新闻信息服务许可证：10120220003

京ICP备11013869号 京公网安备11010502030146号



中国社会科学杂志社版权所有，未经书面授权禁止使用

Copyright © 2011-2026 by www.cssn.cn all rights reserved