

人工智能将超越人类智慧

我们能够做些什么？

农历春节期间，深度求索公司的DeepSeek大模型，因兼具低成本与高性能特征，大幅降低了AI大模型的部署成本，在全球范围内引发热议。

技术狂欢的另一面，是技术焦虑。从AI换脸致诈骗频发，到“AI聊天机器人致死第一案”，再到科幻小说中已泛滥的“AI取代人类”……随着大模型能力不断提升，人工智能似乎正走向一个关键的转折点。作为极有可能超越人类智慧的终极智慧，人工智能的“奥本海默时刻”是否不断逼近或是已经到来？

“奥本海默时刻”源于核武器的发展历程。1945年，物理学家奥本海默在目睹原子弹的巨大破坏力后，意识到科学技术的双刃剑特性。他在《原子科学家公报》中写道：“科学家们知道，他们已经改变了世界。他们失去了天真。”这种对科技潜在危害的深刻反思，成为科技发展史上一个重要警示。

如今，人工智能的发展也面临着类似境况。当人类自己的发明成为一个难以理解的谜团，并且以超乎所有人想象的速度不断演进时，我们能够为此做些什么？

人工智能会否失控

担忧AI系统失控并非杞人忧天。2023年5月，超过350名技术高管、研究人员和学者签署了一份声明，警告人工智能带来的“存在风险”；此前，埃隆·马斯克、苹果公司联合创始人史蒂夫·沃兹尼亚克等人签署了一份公开信，认为人工智能开发人员“陷入了一场失控的竞赛，开发和部署更强大的数字思维，没有人——甚至是它们的创造者——能够对其加以理解、预测或可靠地控制。”

人工智能因其多领域能力快速突破、自主意识初步显现和指数级增长速度让越来越多人认真考虑向充满竞争的世界释放人工智能技术存在的危险。

特定领域超越人类能力。从临床诊断到数学证明，从病毒发现到药物研

发，在一些特定领域，人工智能已展现出超越人类的能力。

斯坦福大学等机构的一项临床试验中，人类医生在特定领域单独做出诊断的准确率为74%，在ChatGPT的辅助之下，这一数字提升到了76%。如果完全让ChatGPT“自由发挥”，准确率能达到90%。

如今，这种特定领域优势正在加速向跨领域的通用能力泛化。去年初，被称为“物理世界模拟器”的Sora横空出世，以场景媒介构筑了一个与人类认知感觉相似的真实场景，推动人工智能由二维迈向三维；工业机器人、自动驾驶、无人机等应用愈发广泛，具身智能更是赋予了AI“身体”……人工智能的多模态能力正在向现实世界的感知与操纵

能力进化。

近一年来，受可用数据、芯片和能源等因素影响，AI技术的增长斜率有所放缓，业内对实现通用人工智能（AGI）的时间预测趋于谨慎，但仍有不少“技术乐观派”对5到10年内实现AGI充满期待。

创造能力实现突破。创造能力在一定程度上被视为人与AI的区别之一，当前，人工智能正在从简单的模仿向创造新知识、新认知的方向发展。

这种人工智能生成的知识在质量、可靠性和原创性等方面仍有待提高，但已经对人类的创作方式产生了冲击。

尽管AGI尚未实现，但当前人工智能系统已展现出一些类似自主意识的表现。

社会与伦理道德风险。2024年10月，全球首例人工智能致死命案发生，一位14岁美国少年在与人工智能聊天系统讨论死亡后饮弹自尽，引发社会对AI情感陪伴功能可能带来的伦理问题的反思。

业内人士指出，人工智能系统的决策过程往往是基于大量数据和复杂算法得出的，这些决策可能与人类的伦理道德观念相冲突。

潜在风险难预判

快速发展的人工智能在就业结构、信息安全、社会伦理道德等诸多方面引发了令人担忧的风险与挑战。

职业结构风险。与以往一项新技术的出现将会取代一类职业不同的是，作为一种全维度的生产效率提升工具，人工智能批量替代人工岗位的情况很有可能出现。这可能会带来潜在财富加速集中和工作两极化，引发结构风险。

信息安全风险。东南亚犯罪集团宣称已将人工智能换脸工具加入其“杀

猪盘工具箱”，相关工具对特定目标的模仿相似度能达到60%至95%并已在1000余个区域安装；2024年11月，ChatGPT为一位美国现役军人提供爆炸知识，后者成功将一辆特斯拉Cybertruck在酒店门口引爆。

人工智能在各个领域的广泛应用，大量个人数据被收集、存储和处理后，引发信息安全风险。一旦这些数据被泄露或滥用，将对个人隐私和信息安全造成威胁。

坚持人类命运共同体理念

当下对于人工智能“奥本海默时刻”到来可能会产生的诸多风险的讨论热度不减，但AGI尚未实现，我们仍有时间寻求如何驯化这一关键技术的解法。

香港大学计算与数据科学研究院院长马毅认为，当前AI的发展更多是工程上的突破，而非科学原理的明晰。人工智能系统，特别是深度学习大多是“黑箱”模型，其内部机制不透明，难以理解和追溯，其结果可信度和可用性打了折扣。

马毅认为，找到破解人工智能“黑箱”的“钥匙”，需要智能研究从“黑箱”转向“白箱”，通过数学方法清晰定义和解释智能系统的工作原理。“真正的智能系统必须具备自主学习和改进能力，而不仅仅是数据的存储和处理。”他呼吁研究者们关注智能的科学原理和理论框架，以实现真正意义上的通用人工智能。

在技术之外，更重要的是，一旦人工智能迎来“奥本海默时刻”，人类应有怎样的态度和责任？

纵观世界各个角落发生的事件，也许比起将人工智能投入战场，或是被用于攫取更大的权力集中，人工智能更理想的应用该是为了促进人类的共同命运和共同利益而努力。

作为AI技术和应用的重要参与

者，中国以负责任的态度应对AI治理挑战，提出包容和有效的治理理念。在国内，我国通过法律框架、技术标准和伦理指南的三元治理促进AI健康发展。在国际上，我国坚持人类命运共同体的理念立场，推动多边合作治理，提出《全球人工智能治理倡议》，支持在联合国框架内建立开放和包容的人工智能治理机制。

中国国际经济技术合作促进会副理事长邵春堡发文称，联合国大会虽然通过我国主提的加强人工智能能力建设国际合作决议，但是在国际合作的实践中，仍然布满曲折，需要共同坚持以人为本、智能向善、造福人类的原则，通过国际合作和实际行动帮助各国特别是发展中国家加强人工智能能力建设。

据新华社

众多产业行业与DeepSeek“联姻结对”

DeepSeek在多家医院完成本地化部署；微信测试接入DeepSeek；部分地区政务系统已接入……众多行业产业与DeepSeek“联姻结对”的消息接踵而至，令人不禁想问：AI正在加速接入我们的生活，究竟会带来什么？

中央党校（国家行政学院）国家治理教研部研究员、博士生导师翟云说，中国经济的快速发展为进一步发挥人工智能的赋能作用提供了广阔的“用武之地”，DeepSeek的深度接入既有利于加快破解传统产业数字化、智能化转型过程中面临的“急难愁盼”，也有利于加快消除“智能鸿沟”、让社会共享AI发展带来的技术红利。

“DeepSeek接入微信、医院等场景带来了资源整合的功能，接入微信也集成在搜索功能。”一位清华大学计算机系博士生感到，这一变化为自己正在开展的大模型底层高性能计算技术研究打开了新的视阈，也为普通人提供了更多便利化路径。

“对于微信这种有大量公众号文章、数据的平台，靠人力整合文本资料是很花时间的。有了AI加持，普通用户在寻找资料方面可以更加便捷，而对于简单的问题，AI也能根据已有资料进行进一步推理并直接给出答案。”他说。

翟云强调，对于DeepSeek这样的新生事物，既要充分释放场景活力，为新技术的应用打造更多“试验场”，从战略层面及时总结提炼科技创新的经验启示，也需正视并妥善应对潜在的数据泄露、隐私侵犯及技术伦理等风险。

据新华社

我国智能算力规模将保持高速增长

国际数据公司（IDC）与浪潮信息近日联合发布的《2025年中国人工智能算力发展评估报告》提出，大模型和生成式人工智能推高算力需求，未来两年，中国智能算力规模将保持高速增长。

报告分析，在旺盛的市场需求、丰富的应用场景驱动下，中国人工智能算力基础设施呈现快速发展的趋势，并表现出多元化、服务化、场景化和绿色化的发展特征。

与此同时，人工智能过程中也面临着诸多挑战，包括计算架构难以支持大规模应用、与基础设施建设和维护相关的成本投入高昂等。

为应对上述挑战，报告认为，人工智能算力发展应采取“扩容”和“提效”并行的策略。“一方面，通过增强算力资源的可获得性、优化算力基础设施架构、推动产业聚集形成规模效应等路径，提高算力供给能力和质量；另一方面，通过提高模型算力效率、增强数据支持来系统性地提高算力利用率。”IDC中国副总裁周震刚说。

据新华社

