



构建“以人为本”的人工智能治理体系

□ 傅宏宇

近年来我国人工智能技术快速发展，技术自主性也不断提高。高度普及的移动网络终端催生了万物互联的数据技术，算力和存储从用户端向云端转移降低了步入数字生活的门槛，深刻改变了人的生产生活方式。目前人工智能已经普遍应用于社会生活的各个方面，从出行路径规划、网络购物选择到知识内容获取、信贷申请审批，人工智能也在改变人类惯常的决策方式，但在提高效率的同时也带来了系统性的问题，比如人工智能技术不够安全、稳定，影响了系统的可靠性；人工智能决策目标单一，难以满足多元的社会需求；人工智能使用的数据和模型复杂不透明，结果难以解释等。

揭开人工智能的面纱，不断解决其发展应用中的诸多挑战和问题，是人类社会的共识。我国近年来不断加强对人工智能治理的要求，结合我国实际情况出台了《新一代人工智能伦理规范》《关于加强科技伦理治理的意见》《互联网信息服务算法推荐管理规定》等一系列政策规范，确立了我国人工智能伦理治理的核心目标和基本原则，创设并完善了人工智能影响评估、备案、分级分类等治理管理工具，引导科技企业评估人工智能技术研发应用风险，健全全生命周期伦理治理体系，为社会提供更为安全、更负责任的人工智能产品服务。

帮助提高人的劳动效率，增加劳动价值。人工智能技术本身处在快速发展的过程之中，从专家系统到深度学习、从有监督学习到无监督的自主学习，机器在不断模拟人类智能的过程中智能化水平不断提高，对人的劳动方式和价值实现产生了影响。一方面，人工智能创造了新的生产生活方式，促进了职业模式的创新，新业态的产生带来了新的就业机会。比如，通过人工智能技术对分散的劳动资源进行线上化整合和实时性调度，催生了网约

车司机、外卖配送员等新职业，通过连接消费者、商户建设起共享性的零工经济，但新的劳动生产方式也带来了劳动保障不足、劳动权益被忽视等问题。另一方面，人工智能降低了知识获取和利用的门槛，对部分传统劳动产生挤出效应和替代效果，例如自然语言学习算法的不断优化让人工智能系统的语音转译、外语翻译能力有了显著提升，目前线上会议软件已经实现了实时记录和翻译功能，提高了对会议内容的记录效率，但也一定程度上降低了对速记、翻译等人员的需求。总体来看，人工智能技术目前在创造性、感知性等方面和人类智能还有较大的差距，完全替代人类的强人工智能尚不具备可能。但人工智能技术会持续影响人类的知识学习和生产工作方式。一是人工智能技术的发展可能将替代部分重复性高、规律性强的人类劳动；二是创造新的、非传统的工作形态，影响现有的劳动价值分配方式和劳动保障制度；三是可能扩大技术赋能与技术失能两大群体间的差距。人工智能治理需要关注其对人的劳动的替代以及对劳动价值的影响，让技术方案更便捷、友好、易用，帮助更多的人提高劳动效率；同时，进一步完善针对新就业形态的劳动保障，让人工智能驱动的劳动者享受到更为充分全面的劳动权益保障。

尊重人的权益，做到有克制、不伤害。随着人工智能的广泛应用，人工智能下的公共决策更加精准，人们生活更为便捷，社会效率也得到提升，但技术渗入社会生活也改变了人们的沟通交流方式，产生了一系列新问题、新挑战。一是个人隐私受到威胁。和物理空间一样，数字社会中人们的隐私也受到保护，但人工智能技术需要大量的数据进行模型训练，不加约束地收集和利用数据侵入了人们在数字社会中的隐私空间，对数字身份带来了安全风险。二是公平公正难以保障。当人工智能被用于调配生产生

活资料、分配社会资源的时候，人们期待其能够做出公平、不带偏见的决定。然而受制于有限的数据分布和数据质量，以及在模型设计和部署实施中的人为错误，人工智能存在难以避免的偏差，而当被用于决定资源分配的时候，技术的偏差就可能造成结果的不公平。三是信任机制存在缺失。人工智能系统的设计、部署和应用比较复杂，机器学习更增加了技术的不透明性，给人们留下了“黑箱”的印象，普通用户很难理解人工智能系统如何做决策，当决策影响到其权益的时候，也很难要求说明、提出申辩、获得救济，往往是被动接受而不是主动选择人工智能的决策结果。

人工智能治理需要重点关注人工智能对隐私、公平的侵犯，提高人工智能技术的透明度，让包括数据使用、模型设计、影响评估等在内的人工智能决策过程更清晰、可解释，为人们提供沟通反馈的渠道、并对产生的负面效果提供及时充分有效的救济，变人工智能“黑箱”为好用、受信任的“工具箱”。

促进社会包容可持续发展。当人工智能成为人类社会生产生活的基本工具，除了考虑此项技术当前的问题，还需面向未来，探讨其在中长期对人的生存和发展带来的影响。在环境保护方面，人工智能技术可以让能源的生产和使用符合可持续发展目标要求，比如通过控制物联网设备实现实时的能源消耗调节，帮助创建更为智慧和低碳的城市；通过智能电网整合可再生能源，实现更为稳定的替代电能供应；利用图像识别技术实时监控分析卫星图，确定荒漠化和海洋污染情况，并提出针对性的方案。但同时，人工智能技术的发展也推高了能源消耗，比如适用于自然语言处理的预训练大模型，在训练过程中算力消耗较大，带来了过多的碳排放。在人的发展方面，人工智能以及更为广义的数据智能改变了人认识和理解世界的方法，在不同的领

域驱动新的创新和创造。目前在医学领域，通过人工智能设计药物分子式、预测药物的理化性质和生物活性，节约了大量的实验成本，而图像识别技术应用于医学影像分析也会大大提高了诊疗的准确性。在文化领域，人工智能创作音乐、影视作品的艺术表现力不断提高。但上述应用对药物研发和诊疗监管、知识产权等针对创新活动的监管制度也带来了挑战，以人的创新为对象的治理体系难以适配人工智能驱动的创新。因此，从长期来看，人工智能的发展需要充分考虑可持续发展目标、推动创新创造，人工智能的开发使用者应当主动承担社会责任，让技术应用服务更为先进的生产力；另一方面，也需要调整现有制度，以适应人工智能对人的生产发展带来的积极改变。

构建“以人为本”的人工智能治理体系，让人工智能坚持以人为中心的价值目标，需要多方协同、共治。政府应当做好顶层设计，确定人工智能发展的价值目标和治理的总体要求，加强对影响隐私、公平等权益的人工智能的监督，在劳动保障、环境保护、知识创新等方面引导人工智能的发展方向。企业应该强化审慎自律，加强人工智能伦理教育培训，对人工智能进行全生命周期管理，防范人工智能对人带来的各种风险，开发和利用更负责任的人工智能。社会应当增进对人工智能技术的理解，对人工智能治理形成共识，积极反馈意见、提供建议，让每一个人成为人工智能的使用者、受益者，也让每一个人成为人工智能的设计者、治理者。



推动社会治理智能化的“技术密码”

□ 刘 蕾 赵雅琼

当今时代，以互联网、大数据、人工智能等为代表的现代信息技术革命改变了人们的生产方式、生活方式，也重塑着社会的组织模式和治理模式。推动社会治理向数字化不断迈进。在现代信息技术革命浪潮中，理解数据在社会治理中的优势，有效释放公共数据价值，发挥数据要素作用，推动社会治理数字化持续发力，是完善社会治理体系的必要之举，也是提升社会治理能力的应有之义。

充分发挥海量数据优势。数据作为新型生产要素，是网络化、数字化、智能化的基础，在破解社会治理难题方面具有一定潜力。社会治理数字化旨在以数据为媒介，实现社会治理物理空间与信息空间的深度融合。依托数据的收集、筛选、整合、分析能力，为社会治理提供科学依据与遵循，从而将数字之治与社会之治有机统一，以数字新技术的创新应用推动社会治理效能迭代发展。在此过程中，各类数据发挥着以下优势：其一，海量数据的客观性有助于减少决策治理的主观臆断与偏见；其二，数据的全面性有助于深入理解社会问题，

推动科学决策与精细化服务；其三，数据的实时性、关联性能够打破时空壁垒与孤岛效应，提高社会治理预警监测有效性；其四，数据的可共享性有助于化解复杂碎片化问题，促进社会治理的跨部门协作与资源整合。海量数据成为破解社会问题、预测趋势和优化决策，推进社会治理准确化、智能化、科学化的“技术密码”。

有效利用公共数据资源。社会治理数字化的基础是具备大量且有效的数据。社会治理数字化的持续发力需要全面化、跨时长、质量好、结构优的数据信息。因此兼顾公共数据的数量与质量，推动公共数据的开放、共享、运营有效性是社会治理数字化建设的源泉性策略。首先，在保证公共数据数量层面，当前公共数据获取有两条渠道，即政府服务平台与公共数据专区，因此推动公共数据开放共享，挖掘更多数据，促进“政府信息公开”向“政府数据开放”转变是必要之举；其次，在提升公共数据质量层面，明确公共数据的供应、授权、运营、利用的责任方，在保障公共数据安全的前提下推动公共数

据的有效运营。这样才能释放公共数据价值，提升公共数据的利用程度，将海量的“数据存量”转变为社会治理的“效能增量”。

发挥数据使用倍效效用。激活数据技术在社会治理中的更多活力需要充分发挥数据的多场景、可重复利用特点，使数据带来的效益成倍增长。一方面，需要同组数据在不同场景重复利用，挖掘出数据的多重适配价值；另一方面，同组数据需要依靠不同分析方法和不同模型处理，得到不同结果，才能满足社会治理多层次、多类别、多维度、多主体的治理要求。在社会治理数字化过程中，通过数据能够更好激活复杂系统中1+1>2的涌现机制，为社会治理带来更多效能。其一，依托数据之间的复杂关系和动态交互，产生出全新的数据结构和特征；其二，以数据更高效便捷的整合与共享特性，构筑复杂信息关系网络，由此产生出超越个体自身的治理效果。将社会治理数字化整体性地纳入系统治理轨道，激活其中的治理动能，是提高社会治理效能与效率的重要途径。

提高数据在社会治理中的精准度。信息技术时代推动了社会治理场景的变化，将传统的社会治理从物理空间扩展到数字空间。这种场景的转变使得信息技术在赋能社会治理的过程中更加精准。在传统社会治理中，受物理空间限制，社会主体往往只有通过“身体在场”才能实现“治理入场”。社会治理数字化所带来的治理场景的变化，能够打破社会治理的空间局限。一方面，持续创新建设公民网上参与平台、研发远程实施互动工具、构建智能辅助决策模型等多元方式，为社会治理构建出更多数字虚拟空间，营造社会治理新型样态。另一方面，以数字信息技术扁平化、多样化、互动化为基础，增强社会治理便捷性，提升公民参与社会治理的意愿。打通数字治理与社会治理的边界，开辟多元主体参与社会治理的新路径。



2024年2月发布的文生视频模型Sora引起关注。这一技术的问世被认为是视频生成领域的重大突破。对比同样引起轰动的ChatGPT，二者在技术路线和产品定位上有所联系，又有所区别。一方面Sora结合了ChatGPT所采用的模型架构，能够结合数个静止画面生成连续的视频，也可以自动修补残缺的视频段。另一方面Sora和ChatGPT都有良好的自然语言理解能力，可以根据用户的描述生成与调整视频内容，也可以对用户提供的素材进行总结归纳和补充。Sora的问世是OpenAI在模型架构、数据管理等方向积累创新的结果，其背后是GPT系列的技术转化和新技术创新的融合。ChatGPT侧重于对文本内容的理解和生成，用户可以与之进行对话问答，也可以指令其撰写文章、编写代码等，其基础功能主要由文字交互实现。最常见的应用场景主要是文本内容的处理，包括对文字的解读、重组、拓展、排序等。Sora则侧重于视频内容的创造，其核心功能是根据文本提示生成视频内容，在理解和模拟物理世界的前提下，构造出虚拟的世界并展现这个世界的交互规律。

文生视频模型都在多个方面表现出强大的能力，但其本身并不完美。类似的生成式模型仍然存在一些问题，如不能完全反映所有的物理规律，主要表现为一些违背生活常识和科学认识的情况仍然会发生。上述文生视频模型缺陷由来已久，并且在短期内难以得到很好的解决，在维持、扩大模型优势的同时克服这些缺陷将成为人工智能产业下一阶段的重点。

首先文生视频模型的首要缺陷是其巨量的算力资源消耗。相较于主流的大语言模型算力，图像参数要占用更多的算力资源。为了追求更好的模型性能，全球人工智能产业的算力需求还会进一步提高，算力匮乏的国家和地区将处于技术不利地位。

其次是模型幻觉现象依然严重。与文本生成模型类似，视频生成模型也受幻觉效应影响。训练数据在压缩等处理环节中失真、用户提示词过于模糊或者根据安全策略不能响应时，模型会被迫填补空缺内容。这种填补有可能使模型陷入幻觉认识，输出与事实不符或者用户未要求的内容。

最后是生成视频细节上存在常识性错误。文生视频模型对物理法则的认识还停留在初级阶段，它能够正确反映人和物体宏观上的互动关系，但还不能准确把握涉及物体形状改变的物理规律。生成式人工智能既可能因为缺乏认知而生成错误内容，也可能因为被不当灌输认知而生成错误内容。比如，当人咬下一口饼干后，饼干却依旧保持完整。虽然这类常识性错误通常被控制在很小的规模，但依旧说明模型对真实物理规律的理解还停留在较浅层次。

Sora的技术路线和性能表现依托于大语言模型基座提供的强大的自然语言理解能力。未来用户与生成式人工智能的交互会越来越轻松，模型也会因为被大量用户使用而得到进一步的反馈训练。研发能够理解、再现甚至模拟物理交互的生成式人工智能将成为新的产业发展方向。

一是媒体行业能够借助此类工具提高内容生产效率，包括电视剧、电影、自媒体等行业在内。随着生成时长、场景准确度、提示词遵循度等性能指标的不断提升，生成式人工智能将有效降低媒体行业的制作成本和从业门槛，改变媒体行业的内容生态。融合各种模型架构的生成式人工智能还将在未来胜任不同内容的具体任务，例如可以同时参与电影的本脚本编写、选角协助、镜头规划和剪辑辅助等。在完整意义上的通用人工智能出现以前，可以并行承担某个领域或行业内所有工作的准通用人工智能可能会先一步进入社会生产当中。

二是创意产业生态因生成式人工智能的不断发展而变化。文生视频模型所生成的虚拟视频具备想象力和设计感，能根据关键词、图片或视频生成相关内容，创作者可以将自己的设计、思路 and 半成品交给人工智能，让其生成完整的创意作品；或者找寻已有作品中的可改进之处。现阶段的生成式人工智能多搭载连接多个不同媒体形式的功能，可以融合文字、声音、图像、视频等各种形式素材，创造极其丰富的内容。生成式人工智能产业会不断强化模型对人思想的呈现能力，大幅降低内容创作者的门槛。普通人也将有机会描绘自己心中的艺术世界，创意作品的内容和形式将变得更加丰富，创意产业有望迎来新的发展。

三是游戏与仿真产业将借助生成式人工智能获得新的发展方向。新一代生成式人工智能展现出的数字模拟能力无疑会进一步降低游戏的制作门槛，使小团队也能独立完成大制作的开发；这项

生成式人工智能的发展趋势与应用场景

□ 张凌寒

突破还给数字仿真带来了新的技术路线——如果生成式人工智能能够正确且精准地认识物理规律，那么利用模型演算、预测复杂事件走向将成为可能。未来，生成式人工智能会越来越接近一个完整的虚拟世界引擎。

四是生成式人工智能有望成为元宇宙世界的基点，它们的表现在一定程度上融合了虚拟与现实，一旦与物联网、脑机接口等前端技术结合，将会给社会带来全新的信息交互方式。在大规模的训练后，不仅可以通图像理解视觉世界，还可以对现实世界进行模拟。虽然相关技术还有很大的进步空间，但是最先进的生成式人工智能已经能够模拟部分的物理交互。文生视频只是新一代生成式人工智能的表现形式。物理模拟模型的本质作用是进一步融合虚拟与现实，从虚拟中创造出无限接近现实的内容。由此可见，生成式人工智能有望成为元宇宙建构的另一个基点。

正如婴儿无数次看到母亲从一个地方消失又出现，他才能渐渐理解这个世界的物理法则，生成式人工智能已经开始通过观察动态的视频学习3D一致性、物体连贯性等物理常识。算法从理解这个世界的真实物理关系到模拟出逼真的物理世界，或许只需要短短的数年时间。在这期间，先进的生成式人工智能可以赋能专用工业软件，拓展其功能边界、提升解决问题的效率。更多以模拟世界为目标的模型将会问世，精准地模拟过去难以实现的场景，在自动驾驶研发、产品设计、电影制作等业务中发挥作用。在人工智能的辅助下，更多的人能够在更短的学习周期内就掌握完成多数工作任务的能力，社会劳动力将会得到进一步的解放，出现在我们眼前的生成式人工智能绝不只是单纯的视频生成模型，而是AI与真实世界互动的开始。

