

向总统报告 超级加速研究： 利用人工智能 应对全球挑战 总统办公室
总统科学技术顾问委员会 2024年4月

关于总统科学和技术顾问委员会 总统科学和技术顾问委员会（PCAST）是总统任命的联邦咨询委员会，旨在增加总统从白宫内部和联邦机构获得的科学和技术建议。PCAST由全国28位思想领袖组成，他们因在学术界、政府和私营部门的杰出服务和成就而被选中。PCAST就涉及科学、技术和创新政策的事项向总统提供建议，同时就需要为影响经济、劳工赋权、教育、能源、环境、公共卫生、国家和国土安全、种族平等政策提供科学和技术信息的事项提供建议。

有关PCAST的更多信息，请访问www.whitehouse.gov/pcast。

总统科学技术顾问委员会 总统办公室 华盛顿特区20502 拜登总统 白宫 华盛顿特区
尊敬的总统先生， 您的总统科学技术顾问委员会（PCAST）对您的政府在推动人工智能（AI）的安全有效使用方面采取的前瞻性方法感到兴奋。根据您关于人工智能安全、安全和值得信赖的发展和使用的里程碑性行政命令的要求，我们很高兴在此报告AI在研究中应用时可以实现的重大社会和全球挑战。 AI将从根本上改变我们进行科学研究的方式。许多领域的研究人员已经在利用AI来找到解决长期存在问题的新方案。今天，科学家和工程师正在使用AI来构想、预测性设计和创造新型材料和治疗药物。在不久的将来，AI将通过分析现有数据的新方法以及开发和分析新类型的匿名化和验证数据，在社会科学领域实现前所未有的进步。这些进步将使政府更好地了解政策如何影响美国人民，并改进这些政策以更好地满足社会需求和挑战。AI还将使研究人员能够快速运行数百万次基于计算机的模拟实验，以提供关于运行最重要的现实世界实验的指导。在工业实验室中，丰富的模拟将能够识别设计中的危险或故障，以便科学家和工程师可以创造更安全、可扩展和高效的产品，美国工业和美国消费者可以依赖。总之，AI正在革新研究过程，丰富科学模型，并加速数据生成和分析，其影响将是深远的。除了机遇，我们必须认识到AI可能会带来新的问题和挑战，例如提炼嵌入在偏斜训练数据中的错误和偏见，计算过程所需的巨大——且不断增加的能量量，可能会无意中生成错误科学，以及恶意行为者可能利用新的强大AI技术进行恶意目的。专家人类监督、将保护措施纳入AI算法以及包括适当应用监管框架在内的负责任使用文化，如您的AI权利法案蓝图和国家标准与技术研究所的AI风险管理框架所概述的，将对减轻AI的弱点和危险至关重要。幸运的是，可重现性和验证是1的核心原则。【白宫。

（2022年10月）。AI权利法案蓝图：使自动化系统为美国人民服务。】 2

【行政命令14110，88 FR 75191。

（2023年10月）。《人工智能的安全、安全和值得信赖的发展和使用》。】 3

【白宫管理和预算办公室。

（2024年3月）。推进机构使用人工智能的治理、创新和风险管理。[备忘录M-24-10]】

科学方法。因此，科学界已经在人工智能的可重复性、测试和评估方面进行了充满活力和开创性的研究，以便研究人员和所有人最终能够像使用计算器一样自信地使用人工智能工具。最终，我们的愿景是，对人工智能的负责任使用将使科学家和工程师能够利用这一变革性技术的机遇，同时应对和减轻其弱点。 根据您的行政命令，并在您政府的大胆工作基础上，PCAST提供了行动的发现和建议，这将帮助美国充分利用人工智能的潜力，以公平和负责任的方式推动研究，以应对世界面临的挑战。 此致，
总统科学技术顾问委员会

目录 致白宫的信.....	
..... 2 人工智能工作组.....	
..... 9 执行摘要.....	
..... 10 1. 引言.....	
..... 15 科学的加速.....	
..... 16 人工智能和机器学习.....	
..... 17 生成式人工智能.....	
..... 18 2. 人工智能赋能研发的未来愿景.....	
..... 20 人工智能方法将帮助研究人员优先考虑最有可能的解决方案.....	
..... 21 通过处理例行任务，人工智能将使科学家能够专注于核心研究.....	
..... 21 例行实验室流程将被自动化和改进.....	
..... 22 以前难以解决的模拟将变得可能.....	
..... 22 共享模型和数据将减少工作重复，民主化研究，并降低使用人工智能的总成本.....	
..... 23 多模基础模型将整合多种形式的数据，并创造科学分支之间的新协同效应.....	
..... 24 人工智能将帮助研究人员更好地利用数据.....	
..... 25 新形式的合作将出现.....	
..... 26 负责任的人工智能实践将融入研究工作流程.....	
..... 26 一旦必要的人工智能基础设施到位，新的科学“登月计划”将变得可能.....	
..... 27 3. 人工智能为加速发现和解决全球及社会挑战提供的关键机遇.....	
..... 28 3.1. 材料发现的相变.....	
..... 28 3.2 用于设计先进半导体的人工智能.....	
..... 31 3.3. 理解和应对气候变化和极端天气.....	
..... 33 3.4 揭示宇宙基本物理学.....	
..... 35 3.5 研究人类行为、组织和机构.....	
..... 37 3.6 推进生命科学基础理解.....	

.....	39	3.7 生命科学中人工智能的突破应用.....	
.....	41	4. 发现与建议...	
.....			
.....	46	发现1	
: 重要研究受到无法访问先进模型的阻碍.....			
.....	47		

推荐1：扩大现有努力，广泛和公平地分享基本人工智能资源。

发现2：尖端研究需要高质量数据。 推荐2：扩大对联邦数据集的安全访问，以满足批准的关键研究需求，并提供适当的保护和安全措施。

发现3：人工智能为学术界、工业界和联邦政府各部门之间的合作提供了独特资源。 推荐3：支持涉及学术界、工业界、国家和联邦实验室以及联邦机构合作的人工智能基础和应用研究，如NAIRR任务组发展的NAIRR愿景所述。 发现4：缺乏适当的基准度量、验证程序和负责任的实践，人工智能系统可能会产生难以评估质量的不可靠输出，这可能会对科学领域及其应用造成危害。 推荐4：在科学研究过程的所有阶段采用负责任、透明和值得信赖的人工智能使用原则。 发现5：最佳性能需要人工智能和人类专业知识。

推荐5：鼓励创新方法，将人工智能辅助融入科学工作流程中。 结论 附录A：术语表

附录B：外部专家咨询 致谢

总统科学技术顾问委员会 主席 弗朗西斯·H·阿诺德 阿拉蒂·普拉巴卡
林纳斯·鲍林化学教授 科技办公室主任 生物工程和政策 生物化学 总统助理科学家
加州理工学院 技术办公室主任 白宫 研究副总裁 E·A·格里斯沃尔德
地球物理学教授 麻省理工学院 成员 丹·E·阿尔维祖 威廉·达利 前校长
首席科学家兼高级副总裁 新墨西哥州立大学系统研究 苏·德斯蒙德-赫尔曼 校长
特拉华大学 前首席执行官 比尔及梅琳达·盖茨基金会
执行副总裁，首席技术官及大气科学教授 环境责任 加州大学伯克利分校
弗朗西斯·C·科尔
工程学院院长及高级主任，国际气候应用科学和阿瑟·勒格兰德电气与计算机工程教授
洛杉矶市立大学 布隆伯格卓越健康与医疗公正教授及卫生公正中心主任
约翰·O·霍普金斯大学 佛罗里达州立大学，佛罗里达大学，洛斯阿拉莫斯国家实验室
航空学和机械工程百年教授 玛丽·克拉夫特物理学教授 加州理工学院
佛罗里达州立大学

保拉·哈蒙德詹妮弗·里奇森

麻省理工学院研究所教授，教务副校长，心理学菲利普·R·艾伦教授

和科学传播实验室的科赫研究所成员社会感知与整合癌症研究传播实验室主任

麻省理工学院 首席科学官管理实践教授（已退休） 摩根大通商学院

董事长兼首席执行官总裁兼首席执行官 摩根大通和AMD

菲利普·H·奈特教授兼研究生院院长前宇航员 斯坦福大学 前国家航空航天局管理者

国家海洋和大气管理局 生态与进化生物学系 弗雷德里克·D·佩特里荣誉教授

加利福尼亚大学洛杉矶分校 物理学弗兰克林·W·和卡伦·韦伯·达比

教授兼伯克利数据研究所主任首席信息安全官 加利福尼亚大学伯克利分校 高级科学家

劳伦斯伯克利国家实验室生物复杂性研究所访问杰出教授 弗吉尼亚大学

莱斯利·苏尔吉纳计算机教授食品科学与人类 科学与整合生物学营养学教授

德克萨斯大学奥斯汀分校爱荷华州立大学

PCAST工作人员 拉拉·坎贝尔 执行董事 瑞巴·班迪奥帕迪亚 副执行董事
梅丽莎·爱德华兹 助理副执行董事 碧翠（Twee）辛 转化医学与健康创新助理主任
金伯利·劳伦斯 行政专员 丽娅·达尔 实习生 玛雅·米莱特 前实习生

人工智能工作组

工作组成员参与了本报告的准备。PCCaoS-TL的全体成员审查并批准了该报告。

Terence Tao Laura H. Greene

加州大学洛杉矶分校文理学院詹姆斯和卡罗尔·柯林斯讲座教授

国家高磁场实验室首席科学家

佛罗里达州立大学，佛罗里达大学，洛斯阿拉莫斯国家实验室 Marie Krafft

物理学教授 佛罗里达州立大学成员 John Banovetz Saul Perlmutter

执行副总裁，首席技术官和物理学教授 伯克利数据科学研究所所长

加州大学伯克利分校 高级科学家 首席科学家兼高级副总裁，研究

NVIDIA霍华德·艾奇曼 计算机科学和综合生物学莱斯大学教授 首席科学官

德克萨斯大学奥斯汀分校 MJoinc rLoesvoifnt 总裁兼首席执行官

斯坦福大学商学院研究生院菲利普·H·奈特教授兼院长 AMD高级副总裁

谷歌云首席信息安全官

执行摘要 人工智能（AI）有潜力通过为研究人员提供工具来加速科学发现和技术进步，从而彻底改变我们解决人类最紧迫挑战的能力。生成式AI，可以根据庞大的数据集和广泛的计算创建内容，有望产生特别革命性的影响。生成式AI的例子包括大型语言模型、图像生成模型和生成式科学模型。拜登总统于2023年10月30日发布了关于人工智能的安全、安全和值得信赖的发展和使用的全面行政命令，要求科学技术政策委员会（PCAST）就“AI在研究中的潜在作用……解决重大社会和全球挑战”进行报告。科学技术政策委员会很高兴能够提供这份报告，以履行这一任务。通过设计良好、公平共享和负责任使用的基础设施，AI将使科学家能够解决紧迫挑战，包括改善人类健康和在气候变化时代提高天气预测能力。AI可以帮助探索激发和拓展人类创造力的长期科学奥秘，例如揭示宇宙的起源和演化。AI还将帮助研究人员解决持续存在的国家需求，从加速半导体芯片设计到发现新材料以满足我们的能源需求。此外，AI开始消除使科学研究变得缓慢和昂贵的障碍，例如通过提供快速识别最佳药物候选物的手段（从而减少昂贵的实验室试验数量），帮助优化实验设计，并比手工或传统数据科学方法更有效地揭示数据中的联系。如果基本的AI资源、验证数据以及科学工具和培训得到广泛的普及，AI技术有潜力使科学知识民主化，将互联的技术概念带给更多人，并使各种研究人员能够将他们的专业知识和观点带到社会和全球挑战中。正如任何其他新工具或技术一样，实现AI的潜力将需要解决其局限性。这些问题包括误导性或不正确的结果，模型训练数据中嵌入的模式导致的偏见或不公平以及抽样误差，获取高质量训练数据的限制，保护知识产权和隐私的挑战，训练或部署模型或运行AI算法所需的大量能源，以及坏人或恶意行为者将随时可用的AI工具用于恶意目的的风险。许多公共和私营部门的活动已经在解决这些问题，包括根据2023年10月的AI行政命令任务的政府努力。可重现性和验证是科学诚信和科学方法的关键原则，必须在我们发展负责任AI使用文化和专家人类监督AI应用的过程中继续高度重视。AI有潜力转变每个科学学科以及我们进行科学研究的许多方面。科学家已经在利用AI创造我们目前不知道如何设计的新功能材料；这些包括超导体和热电材料，不仅可以提高我们的能源效率，还可以减少我们的碳足迹。同样，AI模型正在帮助研究人员为制造过程和产品创建新设计，以及科学数据。【来源：自然。Dessimoz, C. and Thomas, P.（2024年3月）。AI与知识民主化。Birhane, A.（2022年10月）。AI算法中看不见的黑人面孔。】

新药物疗法的开发将在未来实现特定癌症和病毒的个性化治疗。人工智能模型还帮助工程师设计半导体芯片，以更少的人力和时间生产更好的设计。在医疗保健领域，人工智能技术正在创造新的方法来分析广泛的医疗数据，用于早期疾病诊断和及时干预以及检测医疗错误。PCAST还预见到广泛可用的人工智能驱动的超个性化医学，定制给特定个体和疾病过程，其中将包括医疗史、基因信息以及健康和 unhealthy 细胞行为等信号的详细信息。人工智能还通过改进我们的科学模型来改变科学。在气候科学领域，人工智能模型开始增强天气预测，以及推进全球模型用于水资源管理、温室气体监测和预测灾害影响。科学家们已经成功使用人工智能来预测蛋白质的结构；新的基础模型将揭示细胞生物学更多秘密，并为细胞内相互作用的计算机模拟提供动力，这可以用于探索新的疗法。人工智能模型承诺帮助我们通过快速模拟来测试众多宇宙学假设，从而帮助我们理解宇宙的起源。这种人工智能驱动的建模甚至可能帮助科学家发现新的物理定律。人工智能将在社会科学领域实现前所未有的进步，通过新的定量技术补充定性方法来分析现有数据，以及开发和分析新类型的数据，例如智能手机上的步数、经过许可从搜索和浏览中提取的匿名数据，或在社交媒体上发布的图像。人工智能可以通过使用大量数据集来加速研究，例如长期由联邦统计机构收集和整理的数据集，理想情况下还应包括私营部门持有的数据，作为设计有效联邦政策的输入。将人工智能应用于这些长期存在的和新的社会科学数据集，可以促进更有效、响应更快、更公平的数据驱动政策制定和服务提供。这些人工智能辅助研究的几个例子说明，通过负责任地使用人工智能技术，人类科学家将能够实现变革性的发现。此外，PCAST预计，负责任地共享基本人工智能资源将有助于民主化科学，并应对重大社会和全球挑战。全球范围内科学技术研究中人工智能的应用正在迅速加速，因此需要我们致力于美国在这一强大新工具应用方面的领导地位。在拜登政府工作的基础上，美国必须大胆而审慎地行动，以保持我们国家在研究、人工智能创新应用以及建立安全和负责任使用人工智能的框架和规范方面的领先地位。在本报告中，PCAST提出了五个具体的发现和行动建议，将有助于美国充分利用人工智能的潜力，以公平和负责任的方式推动科学发现。

推荐摘要 推荐1：扩大现有努力，广泛而公平地分享基本的人工智能资源。广泛支持广泛可访问的共享模型、数据集、基准测试和计算资源对于确保学术研究人员、国家和联邦实验室以及较小的公司和非营利组织能够利用人工智能为国家创造利益至关重要。在美国，这方面最有前途的努力是国家人工智能研究资源（NAIRR），目前是一个试点项目。PCAST建议迅速扩大NAIRR试点项目，使其达到NAIRR任务组设想的规模，并得到充分资助。全面的NAIRR，连同行业合作伙伴和联邦和州级的其他人工智能基础设施努力，可以作为通往国家或国际级人工智能基础设施项目的垫脚石，以促进高影响力研究。

推荐2：扩大对经批准的关键研究需求提供联邦数据集的安全访问，同时提供适当的保护和防护措施。通过允许经批准的研究人员有限、安全地访问联邦数据集，以及允许向NAIRR等策划资源中心发布经过仔细匿名化的版本，所带来的好处是巨大的。PCAST强烈鼓励扩大现有的安全数据访问试点项目，并制定包括最新隐私保护技术的联邦数据库管理指南。利用现代人工智能技术自动化这些数据集的策划的潜力巨大。PCAST鼓励利用人工智能改进数据策划，作为联邦数据共享倡议（如data.gov）的长期目标。PCAST支持联邦机构的努力，要求负责任地共享其资助或开展的研究产生的数据集。我们鼓励进一步执行这些要求，包括共享基于联邦资助研究数据训练的人工智能模型，同时提供足够的资源支持所需的行动。

推荐3：支持涉及跨学术界、行业、国家和联邦实验室以及联邦机构合作的人工智能基础研究和应用研究，如NAIRR任务组制定的NAIRR愿景所述。由联邦资助的学术研究和私营部门研究之间的界限模糊。许多研究人员在学术机构、非营利组织和/或私营公司之间流动，并且所有人工智能研究和开发（研发）的相当一部分是由学术机构、非营利组织和/或私营公司的研究人员共同进行的。

目前得到私营公司的支持。为了充分利用人工智能对科学带来的潜在好处，必须支持涉及广泛有前途和富有成效的假设和方法的研究。这可能需要资助机构扩大他们与工业界合作的立场，以及支持哪些研究人员，以促进创新研究工作和不同领域之间的合作。这种合作的例子可能包括从多个来源创建高质量的公共科学数据集，或者创建多模态基础模型。

建议4：在科学研究过程的所有阶段采用负责任、透明和值得信赖的人工智能使用原则。管理人工智能在科学用途中可能导致不准确、有偏见、有害或不可复制的发现的风险应该从研究项目的初期规划，而不是事后进行。PCAST建议，联邦资助机构应考虑更新其研究行为准则，要求研究人员制定负责任人工智能使用计划。这些计划应包括来自机构办公室和委员会的建议最佳实践，以解决潜在的与人工智能相关的风险，并描述任何自动化过程使用的监督程序。为了减少对研究人员的额外行政负担并建立一种责任文化，在列举主要风险后，机构应提供风险缓解的模型流程。与此同时，像国家科学基金会（NSF）和国家标准与技术研究所（NIST）这样的机构应继续支持负责任和值得信赖的人工智能科学基础研究。这项研究应包括开发用于衡量人工智能模型属性的标准基准，如准确性、可重复性、公平性、韧性和可解释性人工智能，以及监测这些属性并在基准不符合定义的规范时进行调整的人工智能算法。这类研究的另一个目标应该是开发工具来评估数据集中的偏见，并区分合成数据和真实世界数据。

建议5：鼓励创新方法将人工智能辅助融入科学工作流程中。科学企业是一个极好的“沙盒”，可以在其中实践、研究和评估人类与人工智能助手之间的新合作范式。目标不应该是最大程度地实现自动化，而是让人类研究人员能够以负责任的方式利用人工智能助手实现高质量的科学。资助机构应该认识到这些新工作流程的出现，并设计灵活的程序、指标、资金模式和挑战问题，以鼓励战略性的创新。

新的人工智能辅助方式来组织和执行科学项目的实验。 这些工作流程的实施还为来自各种学科的研究人员提供了机会，例如人因工程、工业和组织心理学，以推动我们在人机团队合作领域的知识。 更广泛地说，跨资助机构、学术界和学术出版行业的激励结构可能需要更新，以支持更广泛范围的科学贡献，例如策划一个高质量且广泛可用的数据集，这可能不会得到传统研究生产力指标的足够认可。 14

在人工智能（AI）驱动的创新推动下，我们正处于一个非同寻常的时代边缘，跨学科的科学领域有望大大加速美国在科学知识和解决方案方面的长期领导地位。这种以AI为动力的科学工具和方法的范式转变及时而关键，因为人类在能源、气候、医疗保健等重要领域面临着艰巨的挑战。通过负责任、公平、有效地利用AI进行研究，科学家们可以为我们的社会提供更大的韧性，提高我们提供清洁水、充足电力、健康和福祉等各种好处的能力。然而，AI改变科学的最重要回报将是实现以往无法达到甚至无法想象的科学进步和理解的可能性，这些发展有潜力改善所有美国人的生活。我们已经看到利用AI解决困难的健康挑战，包括癌症、自身免疫性疾病、神经退行性疾病和耐药感染。与此同时，研究人员正在利用AI在材料科学领域取得令人惊讶的进展，有望创造下一代电池、超导体和计算机芯片。除了在核心科学和工程学科方面取得进展外，AI方法承诺提供高保真度模型——“数字孪生体”——可以帮助我们穿透不确定性和复杂性，预测、规划和指导政策制定，在当前稀缺数据和模型使评估潜在前进路径困难的情况下。

希望，T. 等人（2023年8月）。科学发现的计算推动。《自然》。

Wang, H. 等人（2023年8月）。人工智能时代的科学发现。Mock, M. 等人（2023年9月）。AI可以帮助加速药物发现——但前提是我们提供正确的数据。美国科学、工程和医学国家学院（2022年5月）。自动化研究工作流程以加速发现：闭环知识发现。

Thierry, A.（2023年1月）。通过DNA片段组学和癌症筛查进行城市化。

Danieli, M. 等人（2024年2月）。自动免疫疾病中的机器学习应用：现状和未来展望。

Cumplido-Mayoral, I. 等人（2023年4月）。利用机器学习对结构性神经影像数据进行生物脑龄预测：多队列验证与阿尔茨海默病和神经退行性疾病生物标志物性别分层。

Wong, F. 等人（2023年12月）。发现一类可解释深度学习的抗生素结构。

请参阅本报告的第3.1节。Chen, C. 等人（2024年1月）。加速计算材料发现与可扩展人工智能和高性能计算：从大规模筛选到实验验证。

Pogue, E. 等人（2023年10月）。闭环超导材料发现。

Liu, Y. 等人（2023年12月）。材料专家-材料发现的人工智能。

例如，Liu, M. 等人（2024年4月）。ChipNeMo：用于芯片设计的领域自适应LLMs。数字孪生体是一个不断使用来自该系统的实时数据更新的物理系统的高分辨率模型。这种孪生体通常依赖于传统模拟来模拟系统的基本过程，但还可以使用AI模型来精细、加速和优化模拟。有关这些模型的当前状态，请参阅美国科学、工程和医学国家学院（2024年）。数字孪生体的基础研究差距和未来方向。

2023年10月的《关于人工智能的安全、安全、可信发展和使用的行政命令》中，PCAST被要求就人工智能对旨在解决重大社会和全球挑战的科学研究潜在影响以及确保这些技术的有效和负责任使用所需的实践进行报告。人工智能对国家安全和关键基础设施、劳动力市场、内容的真实性（如图像或视频）、知识产权、教育和刑事司法系统等重要主题的影响将在行政命令规定的其他报告和行动中得到解决。因此，本报告专注于人工智能在科学研究中的影响，而不是这些更广泛的影响。广义上说，科学进步历来是通过三种范式的结合进行的：经验研究和实验；科学理论和数学分析；数值实验和建模。近年来，第四种范式——数据驱动的发现，已经出现。这四种范式相互补充和支持。然而，所有四种科学模式都遇到了进展的障碍。通过实验验证科学假设、仔细观察或通过临床试验可能会很慢且昂贵。要考虑的候选理论范围可能对人类科学家来说过于庞大和复杂。真正创新的新假设可能只能通过偶然的发现或极具洞察力的研究人员发现。数值模型可能不准确或需要大量的计算资源。数据集可能不完整、存在偏见、异质性或噪音，无法使用传统数据科学方法进行分析。人工智能工具在数据驱动科学中有明显的应用，但长期以来，人们一直希望利用这些技术消除或至少减少其他三种范式中遇到的许多障碍。随着人工智能的当前进展，这一梦想即将成为现实：科学问题的候选解决方案正在迅速被识别，复杂的模拟正在丰富，正在开发出强大的新数据分析方法。通过将人工智能与其他三种研究模式结合起来，科学进步的速度将大大加快，研究人员将能够及时应对紧迫的全球挑战。像大多数技术一样，人工智能是双重用途的：人工智能技术既可以促进有益的应用，也可以导致有害的应用，并且如果不负责任地部署或缺乏专业 and 道德的人类监督，可能会导致意想不到的负面后果。尽管如此，PCAST认为人工智能在加速科学技术为社会和地球造福方面具有巨大潜力。在本报告中，我们提供了一个高层次的愿景，说明如果负责任地使用，人工智能如何可以改变科学的进行方式，拓展人类知识的边界，并使研究人员能够找到解决社会最紧迫问题的方法。我们将通过七个不同科学领域展示这种潜力。行政命令14110号，88 FR

75191。《2023年10月》“人工智能的安全、安全和可信发展和使用”。数据在科学研究中当然已经发挥了重要作用几个世纪了；但在现代时代收集和处理这些数据的能力是前所未有的，参见Hey,

T. 等人（2009年10月）。第四范式-数据密集型科学发现。生物医学计算方法和程序 Smalheiser, N. 和Swanson,

D.（1998年11月），使用ARROWSMITH：一种计算机辅助的方法来制定和评估科学假设。

Wang, H. 等人（2023年8月）。人工智能时代的科学发现。

描述当前领域的最新技术水平，人工智能开始被部署以克服进展障碍的方式，我们认为人工智能可以在中长期内帮助实现的好处，以及必须克服的挑战的例子。最后，我们提供关于实现我们对未来基于人工智能的科学愿景所需的跨领域发现和建议。机器学习人工智能指的是一系列旨在执行或协助以前仅能通过人类智能实现的认知任务的技术。人工智能工具实际上并不复制人类思维的机制，这些机制本身仍然未被现代科学完全理解。相反，大多数人工智能系统目前通过机器学习（ML）运行，这是一系列技术，通过在大型数据集上对人工智能模型进行训练来利用统计推断“学习”。然后可以将这些模型应用于推断相关问题的答案。人工智能系统并不深入理解它们试图解决的任务，缺乏外部验证时，人工智能模型提供的答案不能保证正确。尽管如此，这些模型在许多复杂和不精确指定的任务上表现出色，利用它们的训练过程来识别以前隐藏在数据集中的模式和关系。在过去的二十年里，许多机器学习技术已经发展到可以在消费者和商业产品中得到常规应用的程度。作为经典例子，机器学习驱动垃圾邮件过滤器，并且还用于对电子邮件进行优先排序。在更近期的应用中，一款商用智能手机能够识别语音命令，通过面部识别识别其所有者，或从图像中提取文本然后将其翻译成另一种语言。机器学习技术也应用于汽车：传感器、自动驾驶功能和其他驾驶员辅助系统正在部署在新车辆中，以显著提高道路安全性。如今，许多公司常规地依赖机器学习算法来检测欺诈、改善物流和营销、向消费者提供定向广告，或预测客户的信用价值。尽管这些应用已被证明有效，机器学习和更广泛的人工智能技术仍然存在一些弱点。它们的输出通常通过不透明的过程得出，没有正确性的保证，可能涉及使用受知识产权保护的数据或包含有关个人的敏感私人信息。训练数据集中的偏见，以及训练过程中的系统性偏见，可能导致算法偏见问题。一些人工智能系统还具有额外的“微调”组件，例如通过对其输出的人类反馈调整模型参数的过程，这被称为“带人类反馈的强化学习”。事实上，甚至没有一个共识认为人工智能真正理解“智能”这一概念。根据Joksimovic等人（2023年）的研究，人工智能为支持复杂问题解决提供了机会。例如，Idemia面部识别门禁系统现在已经部署在全球许多机场，以简化安全程序。在这一类别中，流行的软件工具包括适用于视力低下者的Voice Dream Scanner或Google翻译。根据AAA最近的一项研究，这些系统在未来30年内可以防止多达3700万起汽车事故。

37 模型输出。为解决这些问题，拜登政府关于AI权利法案的蓝图提出了在将算法偏见和其他弱点纳入现实世界系统之前如何减轻的指导原则。NIST AI风险管理框架进一步增加了具体的测量和管理方法，以减轻和减少采用AI所带来的不良结果，而管理和预算办公室（OMB）关于推进治理、创新和风险管理以促进机构使用人工智能的备忘录进一步澄清了AI应如何利用AI。最近在机器学习工具类别中发生的一些最显著和引人注目的进展发生在被称为生成AI的机器学习工具中，其中包括流行的大型语言模型，如Chat GPT、Gemini、Claude和LLaMA，图像生成模型，如Midjourney和DALL-E，以及科学生成模型，如AlphaFold、RoseTTAfold和ChemBERTa。通过使用经过训练的大型和复杂的深度神经网络模型，这些模型已经在许多TB的数据上进行了训练，并且通常经过数千小时的人类输入进行了改进，这些生成AI工具分析用户提示并使用具有数十亿甚至数万亿学习关系或参数的模型生成输出。生成算法将输入扩展为可能的序列或结构，使这些模型能够根据许多人类专家的响应来创建文本、图像和其他媒体。来自人类反馈的强化学习可以进一步改进这些生成模型。生成AI仍然是一项非常新颖且快速发展的技术，因此存在需要进一步研究和改进的问题和挑战。例如，除了先前提到的机器学习工具的一般弱点外，一些生成AI的输出也容易出现“幻觉”。由于生成AI模型并不专门编码事实知识，它们可能会自信地做出没有事实依据但令人信服的断言，或者可能生成具有与物理现实不一致的图像。对AI生成的真实性的挑战不仅限于训练数据的准确性：生成的概率性质可能导致AI模型创建看似合理但不准确的生成物。在某些应用中，例如与AI系统合作生成创新性可能性，这种“富有想象力”的过程可能是有用或需要的。然而，在大多数科学应用中，目标是得出真实的推断。继续制定策略和减轻措施以减少错误生成和生成AI输出的其他观察到的问题。正在研究方法来确定科学AI模型何时生成错误，以为这些模型的输出提供良好校准的置信水平，并确保这些输出与物理现实一致。

生物或化学限制。对生成式人工智能技术还存在进一步的担忧。目前最强大的模型需要大量的计算基础设施和能量来进行训练和微调。此外，大多数这些模型由少数几家大型技术公司构建和控制，目前该领域的领先者大多位于美国。这种情况仍在不断发展。例如，已发布了相当能力强大的开源和/或开放权重生成模型，最近的研究表明，小型语言模型可以提供高水平的性能，有时可与最大模型的输出相媲美，但资源成本明显更低。虽然人工智能工具已经取得了相当大的进步，但将生成式人工智能完全取代人类在现实世界的流程和工作流程中的前景仍然主要是遥远的未来，而本报告所设想的未来是一种人工智能工具辅助而非取代人类的未来。 尽管它们并不完美，生成式人工智能技术具有巨大的潜力和前景。根据最近的一项估计，生成式人工智能的应用可能每年为全球经济增加2.6万亿至4.4万亿美元。人工智能技术在科学研究领域尤其具有变革性的机会，通过将人工智能模型与其他科学工具和方法相结合，可以解决许多前述弱点。越来越多的研究表明，外部科学验证方法，如实验室实验、临床试验、数值模拟、形式验证软件以及限制人工智能输入从策划的科学文献中获取，可以用来减轻人工智能模型的幻觉。例如，科学模型可以在广泛公开且更好策划的科学数据集上进行训练，这些数据集的知识产权限制相对较少。 科学界非常重视透明度和可重复性，许多用于研究的人工智能模型和方法是公开可用的。数据和训练资源的共享可以大大减少科学家构建和训练这些模型所需的大量时间、计算资源和能源消耗。资源共享还有助于验证输出、验证结果以及直接在先前工作基础上构建。基础研究也可以成为新人工智能技术实验试验的平台，这可能为如何（或是否）在其他情境中负责任地部署人工智能提供宝贵的见解。 人工智能技术在科学领域的不断发展确实带来了新的潜在风险，特别是涉及恶意应用。增强的能力也给科学界带来了新的挑战 and 压力，比如过分信任模型的风险。数据中心使用的冷却需求也可能导致大量的水资源使用。模型在训练过程结束时披露获得的权重，而“开源”模型披露模型结构和训练过程，但不一定披露最终权重。这一区别对于生成式预训练变压器（GPT）模型尤为重要，因为训练过程极其计算密集。

（以上内容为翻译，非原创新闻稿件）

针对AI工具的便利性以及不检查结果的做法可能会影响科学诚信的既定规范和原则。然而，如果以负责任、人类监督和经过验证的方式部署AI工具，这些技术的科学益处可能会超过风险。事实上，由于我们的全球竞争对手和合作伙伴肯定会开发AI的应用，减轻其风险的最佳途径是在制定规范和最佳实践方面处于领先地位。更大的风险将是错失领先世界发展和理解这些强大工具的机会，并将其应用于我们最紧迫的全球问题。 2.

AI-Enabled R&D;未来愿景 鉴于AI技术在科学应用中的实验性和快速发展性（特别是涉及生成式AI），很难预测随着AI技术的应用科学工作流程将如何演变以及何时会发生。

然而，我们设想未来AI-enabled科学将需要在以下三个领域持续关注：

负责任使用人类科学家； 分享AI工具； 以及基本AI资源的共享。正如附图所示，这三个主题相互加强。共享基本的AI基础设施，如高性能计算集群上的时间访问，将使科学家能够使用和理解先进的AI工具，有助于制定负责任使用的共同标准，并通过为所有机构的研究人员提供访问权限，而不仅仅是那些资源最丰富的机构，提高公平性。负责任使用的文化将鼓励安全地分享模型和数据的方式，促进AI辅助研究项目的深思设计，以增强而不是降低其科学产出的质量。最后，多样化的科学 图1.

负责任使用、共享资源和赋权的良性循环

利用AI加速科学研究。本报告探讨了这种相互作用的多个维度。

AI工具广泛赋能的社区将为当前紧迫挑战带来创新解决方案，从而证明对共享资源和负责任使用政策的初期投资是合理的，并创建一个利益相关者社区，以持续发展和建设这些投资。本报告结尾的建议旨在支持和鼓励这些发展。研究人员优先考虑最有可能的解决方案。许多现代科学前沿的问题都是复杂且跨学科的。因此，对于未经辅助的人类专家来说，筛选所有相关科学领域的大量数据和分析，或评估数十万个候选项（例如，用于药物或工程应用的化合物或材料）以确定给定问题的最有前途的解决方案，变得越来越困难。科学家们开始使用机器学习工具，包括生成式人工智能，来填补这一空白。研究人员现在正在使用数据驱动模型来筛选实验室或临床试验中测试的材料、药物和芯片设计的可能候选项，从而通过减少具有负面结果的测试数量并利用有限的实验资源最大化价值，潜在地节省大量时间和费用。未来，这些工具还可以提出对经验观察现象的可能解释，或揭示两个科学领域之间的联系或类比，否则这些联系可能被忽视。从假设上讲，人工智能甚至可以帮助科学家发现新的自然规律，然后通过更传统的理论计算和实验室实验来验证这些规律。重要的是要明确，我们并不预见由人工智能驱动或辅助的推理将取代独特的人类能力和天才，以建立启发式的联系和结论。相反，我们期望传统形式的研究将继续在更广泛的科学社区中发挥重要作用，使科学家能够专注于核心研究。AI所提供的一些最直接的生产力收益将不是直接帮助解决最困难的科学研究挑战，而是来自AI可以提供的对科学家工作时间中占据很大比例的次要任务的支持，例如开发计算机代码、协助论文和报告、进行文献综述以及获取邻近科学领域的专业知识。目前，通用的大型语言模型已经能够为科学文件的撰写提供有用的初稿或实验性变体，尽管现有大型语言模型的输出质量还不足以直接用于科学文件的撰写。然而，我们预计更专业质量的AI写作助手将在不久的将来整合到许多文本编辑平台中，并最终在科学界获得广泛应用，同时更新专业写作标准和指南，考虑到这些AI助手的能力和局限性，例如通过开发类似于最近法律实践中的“LegalBench”等基准数据集的科学写作基准。此外，人类作者应对AI生成的写作输出中的任何错误负责，并应期望对内容保持完全编辑控制。

54, 55, 56, 57, 58 常规使用。然而,我们预计在可预见的未来,实地、非人工智能辅助的研究经验将继续是科学家们的重要部分。许多常规实验室流程非常适合与人工智能合作,使人类能够将时间和精力投入到我们最擅长和最感兴趣的事物上,如设计、分析和协作。一些人工智能技术,包括物体识别、强化学习和生成式人工智能,开始被用于控制机器人系统,使它们能够在不可预测的环境中处理复杂指令并适应感官反馈。这种灵活的问题解决能力将很快使得由人工智能驱动的机器人能够在实验室中执行各种实验,或合成大量材料,而无需为每项新任务仔细重新编程机器人。实验室工作流程的几乎每个方面,从实验设计到数据收集再到数据解释,都有可能通过人工智能辅助而部分或完全自动化,尽管我们认为这种自动化实验室需要专业人员进行监督以应对潜在的问题。人工智能算法成功地被用于大大加速和增强用于模拟复杂系统的计算昂贵的计算机模型,例如地球气候、材料的量子化学以及蛋白质和细胞结构的复杂动态,减少了必须始终返回从第一原理对这些系统进行耗时建模的需求。例如,人工智能模型正在被用于为解决化合物的薛定谔方程提供更有效的近似解,这些解有助于定义化合物的性质。

54物联网和物理系统 Ray, P. (2023年4月)。ChatGPT: 背景、应用、主要挑战、偏见、限制和未来研究的综述。癌症研究杂志。Huang, J. 和Tan, M. (2023年4月)。ChatGPT在科学交流中的作用: 撰写更好的科学评论文章。Boiko, D. 等 (2023年12月)。大型语言模型进行自主化学研究。尽管如此,可能会出现“杰文斯悖论”(或“无纸办公室神话”的版本),即由人工智能工具提供的写作效率提高会转化为(至少在短期内)产生更多科学写作,而不是减少写作任务所花费的时间。

SReelilnefno, Ace. manedn th loearvrnitizn, gE. (2023年11月)。AI自动驾驶的崛起: 从航空界和其他领域汲取的设计经验。arXiv。

是一种通过要求算法进行所需动作的训练方法,以获得所需的结果。Zhou, C. 等 (2021年11月)。智能机器人运动规划算法综述。国家科学、工程和医学院 (2022年5月)。用于加速发现的自动化研究工作流程: 闭环知识发现。

62至67 稳定性和其他属性。目前正在进行多项努力，以开发用于许多应用的AI驱动的“基础模型”和“数字孪生体”。一旦完成，整个研究社区将能够在这些强大而广泛的平台上构建更多定制模型，以快速创建各种科学和工程目的模型。

还将开发大型模型的“轻量级”版本 – 这些版本将足够小，可以在个人计算机上运行和微调，同时仍保留原始模型的大部分功能。这将降低在科学应用中部署AI的成本和环境影响，并应该实现更广泛、更公平地访问这些模型和由此产生的创新。正如先前所指出的，应当注意避免纯粹依赖AI模型。在可能的情况下，AI模拟应当经过验证，并与数值模拟等传统验证方法进行基准测试，以确保数据的准确性和可靠性。为了有效使用AI工具，首先必须收集高质量的数据集，使其可访问，并以可用格式呈现。接下来，模型需要在经过筛选的数据上进行训练，注意限制算法偏见的同时保护可能包含在数据集中的任何个人信息的隐私。重复这些步骤对每个研究项目来说是低效的，浪费了大量的时间和资源。例如，与其目前正在进行的多项努力建立大规模基础模型，将这些努力结合或连接起来，集中资源和人才关注模型规模和质量是可行的。通过诸如构想中的国家人工智能研究资源（NAIRR）之类的中央资源，研究人员将获得标准化模型和筛选数据集，并分享最佳实践。 62《科学报告》 Radu, A. and Duque,

C. (2022年2月)。用于解决任意量子阱中薛定谔方程的神经网络方法。 63

例如, Mukkavilli,

S. 等 (2023年9月)。天气和气候的AI基础模型：应用、设计和实施。arXiv。

例如, Houben, M. (2020年11月)。数字孪生体，植物表型研究的未来 – Rijk

Zwaan的TechnoHouse。 65 例如, Corstian, A. 等 (2023年3月)。利用现有技术的结

构健康监测的数字孪生体：增强管理、运营和维护的新范式。 66《儿科前沿》

例如, van Willegen, B. 等 (2022年9月)。胎儿循环模型的综述研究，用于开发围产期生命支持系统中的数字孪生体。 例如, Geddes,

L. (2023年11月)。数字孪生体如何实现个性化健康治疗。 68 基础模型是经过广泛数据训练的ML模型（通常需要大量计算资源），然后可以相对廉价地对其进行微调以用于更专业的应用。有关更多讨论，请参阅Moosavi Rezvani,

B. (2022年7月)。基础模型的机遇和风险。arXiv。 Javaheripi, M. and Bubeck,

S. (2023年12月)。Phi-2：小型语言模型的惊人力量。 70 IEEE探索。 Dilmaghani,

S. 等 (2019年12月)。AI系统中大数据的隐私和安全性：研究和标准视角。2019年IEEE国际大数据会议。 Metcalf, J. and Crawford,

K. (2016年6月)。大数据研究中的人类主体在哪里？新兴的伦理分歧。 72

国家人工智能研究资源（NAIRR）试点。（2024年4月9日访问）。

73 将推动人工智能基础科学的进展。一个中心资源还将促进各领域研究人员安装、微调和操作上述基础模型，而无需具备专业的人工智能专业知识。 74 一旦正确的基础设施就位，人工智能辅助研究将不仅仅成为高资源公司 and 研究团体的可能，也将成为小型机构、私营部门组织，甚至普通公众的可能，为发现和开发利用人工智能工具的创新方式提供更加公平的机会。然而，对于可能具有潜在有害应用或需要隐私保护的模型和数据，需要一些访问限制。美国政府正在采取措施促进安全可信的人工智能。例如，最近发布的OMB备忘录概述了联邦机构将被要求遵循的一套最低风险实践，以减轻人工智能在涉及权利和安全的情境中的风险。 NAIRR试点项目为正在进行可信机器学习研究的研究人员提供数据、培训和计算资源。此外，最近成立的美国人工智能安全研究所和人工智能安全研究所联盟(AISIC)将汇集来自政府、学术界和公民科学领域的专家，共同合作推进人工智能安全研究。这些资源将整合多种形式的数据库，并在科学领域的各个分支之间创造新的协同效应。 大规模语言模型被称为“基础模型”，因为它们为将通用人工智能模型转化为专门任务和特定领域应用的人工智能系统提供了基础或“基础”。这种改进通常通过使用专门的领域特定数据的学习机制来完成。将基础模型在特定领域转化为高性能模型的过程是一种“迁移学习”。在机器学习研究中，这个术语已经被用了几十年，用来指代涉及将在一个领域上训练的ML模型，例如无机化学，适应其“预训练”以提高模型在其他领域学习能力的方法。 73 国家人工智能研究资源任务组。《加强和民主化美国人工智能创新生态系统：国家人工智能研究资源实施计划》。 74 这样的基础设施不仅包括硬件和软件资源，还包括基准、监管原则以及关于人工智能负责任使用的指导方针（如《人工智能在人类研究中使用的IRB考虑事项》）；最重要的是，具备在公共和私营部门中使用和部署人工智能专家的人力资本。 75 对于某些“双重用途”科学应用，例如病毒功能增强研究，将人工智能辅助科学进步的广泛应用扩展到普通公众并不一定是可取的，这将需要一些额外的监管和监督。然而，我们看到许多科学领域，增加公众参与和互动，不将最强大的人工智能工具限制在少数资源充足的团体或基于其他国家的研究人员手中，将是有益的。 76 白宫管理与预算办公室。《推进治理、创新和风险管理：机构使用人工智能的备忘录》。[备忘录M-24-10]。 例如，Zhuang, F. 等人。《关于迁移学习的综合调查》。IEEE会议论文集。 24

在最近的工作中，研究人员一直在探索构建多尺度和多模态基础模型，这些模型可以利用联合表示学习，并利用迁移学习。多模态学习的目的是结合各种不同类型、不同尺度甚至来自不同科学领域的数据集。除了联合学习包含多种类型数据的数据集外，还正在开发有效的方法，用于将已经独立训练的模型与来自不同科学领域或焦点的数据连接起来。在这个领域的工作包括使用适配器：轻量级连接器模型，经过训练可以将两个或多个现有模型连接起来。虽然多模态模型和相关功能可以以不同方式构建，但我们预计开发跨越多个学科领域以及多个空间和时间分辨率的人工智能工具，将为科学家提供强大的新能力，用于描述或模拟复杂系统。这些能力可能远远超出仅通过特定领域模型可以实现的范围，并将促进跨学科思维和合作的发展。通过人工智能展示的推断背景和利用复杂、嘈杂、非定量的现实世界数据的能力，例如自然语言文本，人工智能算法显示出处理现代数据驱动科学基础的极大潜力，包括自动组织、组合和“清理”极其庞大和异构的数据集，以及识别异常并揭示数据中的重要相关性和模式。人工智能工具还提供了新的方法来实现“超分辨率”，以提高单个图像的质量。此外，在许多领域，人工智能工具已经被常规使用，帮助生成“合成”数据集，当被负责任地使用时，这些数据集可以极大地提高经验性生成数据的质量和预测能力，保护这些数据集中敏感信息的隐私，减少算法偏见的风险，并帮助从基础数据中推断出更广泛领域的结论。然而，这种合成数据将需要被仔细和永久地标记，以区别于数据多尺度建模。联合表示是一种建模策略，同时使用不同尺度的多个模型来描述一个系统。多模态表示是一种机器学习模型从多种类型的数据中学习的手段，例如自然图像与文本结合在一起的方式。Fei, N. 等人（2022年6月）。通过多模态基础模型实现人工通用智能。Li, Z. 等人（2024年2月）。MLIP：利用分歧编码器和知识引导对比学习增强医学视觉表示是一个最近的例子，这是一个从带注释医学图像中学习的多模态模型。Poli, M. 等人（2023年3月）。决策导向学习的理想抽象。arXiv。指的是从数据集中移除或修复部分数据，这些数据不完整、不准确或无关，以提高数据集的质量以供分析或训练使用。例如，Chen, H. 等人（2022年3月）。真实世界单图像超分辨率：简要回顾。MIT新闻。信息与决策系统实验室（2020年10月）。合成数据：真实数据的替代品。人工智能。Savage, N.（2023年4月）。合成数据可能比真实数据更好。Jordon, J. 等人（2022年5月）。合成数据的技术、原因和方法。arXiv。Listgarten J.（2024年1月）。人工智能生成数据的永动机和ChatGPT作为“科学家”的分心。

通过真实世界的观察、感知、调查和实验收集的数据，以避免对“数据污染”问题做出贡献。一项现代科学研究项目通常由少数几位资深科学家领导，指导更大规模的研究生和学生执行更专业化的子任务。人工智能工具将自动化或至少协助执行许多这些子任务，并通过其自然语言处理能力，极大地促进来自不同科学背景、专业水平或母语的研究人员之间的交流方式。由于这一点，新的合作范式将出现，例如增强人工智能专家、人类辅助人工智能系统、结合互补人工智能技术的混合人工智能技术，或者大规模的众包、去中心化和/或高度跨学科的项目，其中个人贡献通过人工智能工具和更严格的评估方法的组合进行验证和整理。除了促进科学家之间的交流，人工智能辅助还可以让普通人参与尖端研究项目。例如，公众可以通过专门的聊天机器人以科学准确、易于理解且真正双向的方式直接并有意义地参与研究，公众的评论可以反馈到假设或输出的制定中。即使在人工智能辅助变得普遍的未来，我们也设想传统科学研究和参与将继续发挥作用。传统研究方法和途径是人工智能辅助科学的必要和补充方法，提供定性背景、通过实验提供重要证据、为培训科学家提供关于工具如人工智能可能带来的错误的关键方法，以及最重要的是，提供研究工作流程中的智力多样性。科学方法本质上通过独立复制和审查机制融入了自我纠正机制。特别是，随着科学的成熟，实验设计不断发展，随着时间推移引入标准程序以减少偏见、最小化对人类或动物实验对象的伤害、提高可复制性并避免浪费。在研究中使用人工智能工具也将发生类似的演变，但这将需要学术机构、科学专业组织和资助机构致力于建立负责任和道德的人工智能实践文化，并帮助研究人员开发、测试和评估这些实践。科学、学术和政府领导者必须鼓励最佳实践，例如科学家确保在其工作和结果中适当引用人工智能模型的使用，并与他们的社区分享有关所使用模型特定版本的详细信息；这些引用对于促进结果的复制和验证至关重要。

94 预计在归因工具和功能的发展方面取得进展，这些工具和功能可以为科学家提供对训练数据来源的理解，这对于生成输出至关重要，以便可以适当引用先前的研究和结果。通过人工智能基础计算机科学的进展，正在开发指标来确定人工智能训练数据集的质量并优化选择，以及标准化方法来补偿这些数据集中的偏见和遗漏。PCAST预计将开发新的人工智能架构，这些架构可以在能耗和环境影响方面提供与现有最大模型相媲美或更优异的性能，但能耗却只是它们的一小部分。 96, 97, 98 开始部署复杂的隐私保护方法，以保护敏感个人数据的隐私，例如可能泄漏到人工智能模型中或从中泄漏出去的医疗信息。人工智能模型中的幻觉和其他不准确之处可以通过改进方法来进行补偿，这些方法可以将人工智能输出与真实世界来源“接地”，并为这些输出附加良好校准的置信水平；我们还可以利用更严格的验证方法，例如形式验证。在可行的情况下，模型、数据集和权重应该负责任地开源，以协助复制和透明度，并倾向于“可解释人工智能”模型，这些模型可以在它们的结论和输入数据之间建立明确的联系。还应建立维持人类监督自动化实验室流程的最佳实践。通过加强自然科学、社会科学和人文科学之间的对话，我们期望更好地理解在研究中部署人工智能工具可能导致的潜在意外后果，并制定评估、预防、最小化和减轻此类后果的标准流程。相反，虽然人们普遍担心科学文献可能面临低质量文章的涌入，这些文章部分或完全由人工智能生成，但我们也期望人工智能工具在维护学术行为标准方面发挥作用，例如通过检测不端行为，促进学术诚信。新的科学“登月计划”将变得可能。当像NAIRR这样的基础资源提供计算能力、安全数据共享服务、开源（和开放权重）人工智能模型以及其他关键基础设施时，将有可能规划雄心勃勃、复杂和大规模的“登月计划”。 94 一样东西可以有的用途或目的；定义其可能的用途或清楚说明它可以或应该如何使用的对象的特性或属性。自然计算科学潜在的未来架构是神经形态计算；参见，例如，Schuman等人（2022年1月）。神经形态计算算法和应用的机会。 96 例如，Papernot, N.和Thakurta, A.（2021年12月）。如何部署自然语言理解：无需深入了解人工智能的网络安全洞见。 Savage, N.（2023年4月）。合成数据可能比真实数据更好。 98 Kaissis, G.等人（2020年6月）。医学影像中的安全、隐私保护和联邦机器学习。 99 Urban, C.和Miné, A.（2021年4月）。应用于机器学习的形式方法综述。arXiv。 100 国家科学院、工程学院和医学院（2018年7月）。设计开放科学：实现21世纪研究愿景。例如，参见Xu, F.等人（2019年9月）。可解释人工智能：关于历史、研究领域、方法和挑战的简要调查。

在计算机模拟中，涉及多个公共和私人合作伙伴的科学项目。此类项目的示例可能包括模拟人类细胞复杂性的基础模型，允许对疾病和实验性治疗进行研究；一个详细的全球模型，利用传统和人工智能模型描述地球系统的组成部分，同时不断更新高度多模态的实时数据；或者一个大型协作努力，通过系统收集、处理和人工智能辅助分析现有数据和文献，结合自动化实验室合成和测试可行候选者，来发现实用的室温超导体。虽然已有个别研究团体或组织正在进行类似的项目，但通过共享人工智能资源基础设施可能实现的更大规模合作将极大受益于规模经济，并同时减少工作重复。在接下来的部分中，我们突出显示了AI工具在科学中应用的七个代表性领域，作为具体案例研究。每个领域都面临着阻碍进展的重大障碍，如果克服这些障碍，可能会带来改善人们生活、减轻全球风险或激励我们作为人类的发现。对于这些领域的每一个，我们描述了该领域的当前状态，确定了进展的障碍，提供了AI当前如何被利用的示例，并描述了我们设想可能实现的未来，以及必须考虑的潜在风险和进展所需的资源。这些示例列表并非旨在全面；还有许多其他未提及的科学领域也可能受到AI的转变。然而，这些小插图展示了跨领域主题、广泛的可能性和AI在各个科学领域提供的关键需求。

材料发现的相变 利用生成式人工智能发现新材料已被证明对推动经济扩张、改善健康结果以及创造对国家防御至关重要的新材料至关重要。历史上，人类状况的重大改善时期是由材料科学的进步推动的：青铜、铁、混凝土、钢铁。如今，我们生活在硅、碳氢化合物和硝酸盐的时代。不久的将来可能是纳米材料、生物聚合物和量子材料的时代。新材料将成为气候友好型能源技术的基础，包括改进的电池和能源存储、碳捕集和氢生产。新的生物制品将改善健康结果并为护理开辟新途径。最后，AI辅助的研发将开启以往仅存在于想象中的可能性，如室温超导体或大规模量子计算机架构。我们的未来将建立在新材料之上；它们将彻底改变我们的社会。 来源：Axios Snyder, A. (2023年12月). 为开发新材料提供人工智能支持。 PCAST (2023年8月). 国家纳米技术倡议第七次评估。 ComPACStic (2022年12月). 生物制造推动生物经济发展。 Stanev, V. 等人 (2021年10月). 用于搜索和发现量子材料的人工智能。

科学家们已经成功地利用深度学习模型进行材料发现。例如，私营公司的跨学科研究团队已经利用人工智能开发了数百万种新型材料设计，其中预测中的近50万种可能稳定到足以在实验室中生长的候选材料。PCAST预计未来AI辅助材料发现的进展将能够缩小类似规模的候选稳定材料特征空间，以分离那些最有可能实现指定目标性质的材料。AI还被用于改进现有材料；例如，AI工具可以帮助科学家优化材料组成，以减少或消除潜在的环境危险材料，同时保持性能。此外，AI可以整合处理参数，不仅优化组成，还可以优化材料制造方法，提高效率，减少浪费，并导致更可持续的新反应和加工路径。美国国家科学基金会（NSF）已经投资7250万美元，推动设计、发现和开发解决重大社会挑战所需的先进材料。设计材料以彻底改变和引领我们未来（DMREF）计划将资助37个新的为期四年的项目，这些项目跨越NSF的许多主任办公室，涉及科学、工程、计算等领域，包括与私营部门的合作；每个项目都将采用深度学习和人工智能。能源部在未来五年内为量子科学中心提供了1.15亿美元的资金，其中包括支持AI辅助的量子材料发现和设计。AI方法在寻找新的超导体材料方面可能特别有价值，这些材料可以高效传导电力，几乎没有能量损失。超导体是MRI机器、粒子加速器、某些实验性量子计算技术以及（在有限的地方）我们电网中无损能量传输的必要组件；但它们目前具有一些不可取的特征。首先，所有已知的实用超导体必须使用液氮冷却至低温（-298华氏度或更低），这是一个涉及昂贵和有限资源的不切实际的程序。其次，与铜等传统导体相比，现有的超导体不具有可塑性，一旦受损就会失去超导性能。第三，它们的使用成本非常昂贵，无论是原材料的成本还是将其工程化为导线所需的努力。更好的超导体——那些可以在更容易实现的温度下工作、更容易工程化到应用中并且更便宜的材料——将是变革性的。它们将使MRI的获取更加民主化，通过减少电网中的电阻损失降低能源成本，并促进经济的进一步电气化。这些材料还将在我们的应用中发挥作用。参考文献： - 陈C.等（2024年1月）。利用人工智能加速计算材料发现：从大规模筛选到实验验证。arXiv。 - 费安提，C.等（2023年12月）。MatterGen：一种用于无机材料设计的生成模型。 - Merchant, A.等（2023年11月）。用于材料发现的深度学习扩展。物理评论快报。 - 美国国家科学基金会（2023年9月）。NSF投资7250万美元设计革命性材料。 - 美国国家科学基金会。设计材料以彻底改变我们的未来（DMREF）。（2024年4月访问）。 - 橡树岭国家实验室（2020年8月）。ORNL，合作伙伴获得1.15亿美元设立量子科学中心。

交通领域，例如实现磁悬浮列车，可以在最小摩擦下行驶，实现更平稳的乘坐体验和更高的效率，真正让科幻变为现实。科学家们还不知道如何设计的另一类材料是热电材料，可以将热量，甚至是来自电力传输或发动机的废热转化为能源，具有许多应用，如冷却和电子设备。研究人员从未能够预测性地设计超导体或热电材料，因为这些量子材料需要独特的物质组成。以往的努力依赖于组合化学方法——涉及创建和筛选大量材料组合的经验性实验，取得了有限的成功。基本上，所有这些至关重要材料的发现都是偶然的，通过实验的试错方法获得的。涉及的变量数量庞大，以及保持这类材料的价格实惠的需求，使得这些材料发现问题变得令人难以应付，几乎不可能通过传统方法解决。人工智能工具将在材料科学领域发挥改变游戏规则的作用的三个领域。首先，人工智能建模的预测能力正在通过连接和利用现有材料、其加工条件和性质的大量数据，实现材料发现的新方法。从这些数据集中，可以确定材料的化学、物理和工程方面的模式，并以独特的方式结合起来，为研究人员提供见解和新方法。其次，人工智能模型可以预测性能（例如，预测量子比特的相干时间、热电材料的效率或超导体的临界温度），从而减少对不可行候选材料的浪费性实验和测试。第三，通过将过程信息与材料组成相结合，可以在材料设计上设定实际边界，加速新材料的规模化生产和商业推广。除了研究“硬”材料如超导体和热电材料外，人工智能还将彻底改变“软”材料如聚合物和流体的开发。软材料的材料发现需要与硬材料相同的大量数据集和预测能力，以及材料科学中复杂的结构-功能关系。不幸的是，迄今为止，将人工智能应用于聚合物发现和加工仍然是一个新兴领域，具有巨大的未开发潜力——很可能是材料科学的下一个前沿。展望未来，人工智能工具可能会导致新型或改进的材料，如冷原子、拓扑绝缘体或基于超导体的量子比特，这些材料将成为量子计算机的构建模块，能够执行某些大规模计算，如果使用传统超级计算机执行将需要不切实际的计算或能源水平。自然物理学

Keimer, B. and Moore, J. (2017年10月). 量子材料的物理。Liu, Y. 等人 (2023年12月). 材料专家-材料发现的人工智能。新视野。例如，Biamonte, J. 等人 (2017年9月). 量子机器学习。自然。例如，Pashilovskiy, D. (2023年6月). IBM量子计算机通过计算里程碑。例如，Breunig, O. 和Ando, Y. (2021年12月). 拓扑绝缘体的机遇。例如，Ballon, A. (2024年3月). 基于超导量子比特的量子计算。麦肯锡公司。量子计算的崛起。(2024年4月9日访问)。

近年来，虽然在实现小型量子计算设备方面取得了重大进展，但这些系统不稳定，容易受到噪音干扰，并且难以扩展到计算解决有用问题所需的规模和复杂性。尽管如此，量子计算的潜在应用引起了国际社会的广泛关注和资金承诺。人工智能的辅助可能是决定这一新兴技术进展速度的关键因素。总的来说，人工智能工具提供了机会，我们今天已经开始意识到，引领着材料发现的新时代，从而加速经济增长和开创新未来。现代电子设备是全球经济和国家安全的基础，运行在被称为“芯片”的小型半导体材料上的集成电路上。随着这些芯片变得更加强大，它们也变得更加复杂，目前最先进的芯片现在包含数百亿个组件。如今，只有最大的公司才能负担得起制造最先进芯片，因为设计芯片需要大量的工程资源和基础设施。重振美国半导体生态系统一直是本届政府的重要优先事项。在未来芯片设计中广泛使用人工智能可以显著提高芯片的质量，减少设计最先进芯片所需的时间和工程师数量。例如，美国半导体行业的一个雄心勃勃的目标是创建平台、方法和工具，使芯片的构建所需的人时减少到今天所需的十分之一，这将大大降低进入半导体市场的门槛，鼓励更多和更多样化的参与者进行创新，并继续在世界半导体设计领域保持领先地位。目前已经存在几种面向芯片设计师的人工智能助手。这些工具可以让初级设计师提出以前需要高级设计师花费时间解决的问题。一些芯片设计人工智能助手还可以总结错误报告和其他设计文档，或者根据简单的英语提示生成其他设计自动化工具的脚本。这些人工智能工具并不取代设计师，而是使设计师能够更加高效，有助于缓解受过训练的芯片设计师总体短缺问题。尽管电路设计是一个成熟的领域，但现有的和正在开发中的人工智能工具承诺提供令人惊讶的设计改进。其中一些工具能够生成比使用传统方法设计的最佳电路更快或更小的电路。

改善电路设计AI助手性能的机制是一种称为强化学习的技术。当AI工具探索可能电路的“状态空间”时，通过生成良好或不良电路的正面“奖励”和负面“惩罚”值来接受强化。它根据这些奖励改变其方法，最终学习哪些策略导致具有理想特征的电路。对于每个新的半导体尺寸，必须重新设计成千上万个“单元”库——现代芯片的标准设计构建模块，以满足工艺的约束。对许多制造商来说，这项任务可能需要大约80人月的工作量。最近开发的工具利用包括用于聚类的生成式AI和用于修复设计规则错误的强化学习等AI方法，能够自动化生成新库，并将工作量减少了一千多倍。类似地，基于机器学习的“布局规划”工具（用于确定芯片上组件的最佳位置、形状和大小）使用强化学习来减少设计时间，并改善放置这些标准单元的芯片布局质量。此外，现场可编程门阵列实现了对最新AI驱动的放置和布线技术的快速迭代，显示出相对效率提高了3倍以上。在创建芯片设计时，需要进行多项分析以验证设计是否符合规格和制造过程的约束。AI已被应用于加速其中许多分析。例如，在执行详细布线之前预测时序，或者预测电路的不良“寄生”特征。使用诸如此类的AI工具，设计师可以快速迭代许多电路想法。过去，例如，必须生成电路的布局才能准确了解寄生特征，这往往会为设计周期的每次迭代增加数天的手动工作。现在，整个设计迭代循环可以在几分钟内完成，以获得符合所需规格的电路。展望未来，我们期望大型语言模型（LLMs）将发展成为能够回答问题、批评和验证设计，并执行常规设计任务的设计助手。我们还期望AI技术将显著提高设计师的生产力，潜在地提高十倍或更多。设计师将在算法和系统级别工作；然后AI助手将在设计的较低级别处理细节。AI综合和分析工具将大大缩短设计周期，使设计能够从高级描述到验证的布局在几小时内完成，而不是今天需要数周的时间。

PCAST预计，通过这些新兴技术整合到芯片制造过程中，美国将在半导体设计领域保持领先地位，同时缓解该领域的关键人才短缺问题。 3.3.

理解和应对气候变化和极端天气 每年在美国，诸如飓风、野火和洪水等气候灾害造成数百人死亡，数百亿美元的损失，并严重影响那些被迫搬迁或重建的个人和社区的身心健康。虽然罕见，但这些极端天气事件中最严重的那些造成了大部分损失。例如，2023年，美国经历了28起单独的天气和气候灾害，造成929亿美元的损失，包括南部和中西部春秋季节的干旱和热浪（145亿美元）以及三月初南部和东部的恶劣天气（60亿美元）。许多这些极端事件变得更加频繁，并预计随着气候变化将变得更加频繁，预计到本世纪末，数百万美国人可能面临被这些事件迫使离开家园的风险。为了让个人、社区和政府能够为这些事件做好准备并减轻其影响，越来越重要的是获得准确和详细的天气和气候模型，以便实时预测极端天气事件的轨迹，以及未来气候风险的预测。理想情况下，我们希望有一个“水晶球”，可以向我们展示未来十年、二十年或五十年内任何给定位置可能受到的野火、洪水、热浪或飓风的强度。直到最近，最好的“水晶球”是运行巨大天气和气候模型的超级计算机，这些模型并未使用人工智能。这些较老的模型将地球覆盖在一个虚拟网格中，每个网格点模拟了数百万个时间步骤。但这些“水晶球”速度慢、模糊且不确定。例如，美国国家气象局的全球预报系统每天运行四次，以提前16天的天气预报，但空间分辨率约为13公里，并在六小时周期内同化输入数据。对于长期气候建模，一个最先进的例子是由能源部支持的能源级地球系统模型，当完全运行时，可以在几周的超级计算机时间后以3公里分辨率为未来十年产生一个可能的气候结果。尽管这是一个令人印象深刻的改进，但不使用人工智能的超级计算机模型只能在众多可能结果中产生少数结果，仅允许有限探索不确定性，并且无法准确指出在给定网格方格中极端事件最可能发生的位置。

人工智能直接从数据中学习的能力将增强这些“水晶球”，使它们变得更快、更精确。人工智能可以通过俄克拉荷马州龙卷风的数据进行训练，从而学会如何对俄亥俄州的龙卷风进行建模。如果科学家将人工智能训练于当地气候和天气数据，它可以用来将低分辨率气候模型升级为考虑当地地理的高分辨率模型（例如，1公里分辨率或更低），这个过程被称为“降尺度”。一旦人工智能模型被训练，这个过程就会变得极其快速——比传统模拟快上千倍，因此在计算成本和能源使用方面也更便宜。基于人工智能的对粗糙气候模拟的降尺度有潜力产生数百甚至数千种无法被粗糙模型解决的未来过程结果，如热带气旋和风暴潮，从而实现对气候和天气灾害的定量评估。未来，任何公民（或地方政府）都将能够在政府气候门户网站中输入其邮政编码，获取未来十年、二十年或五十年内其地区预期气候变化和潜在天气灾害的多个详细场景。这些数据随后可以与建筑环境数据结合，以估算这些事件的经济影响。这些信息对于设计适当的建筑法规、应急响应准备、保险政策和城市规划将是无价的。就气候变化缓解工作而言，这些快速的人工智能气候模型仿真器还将使科学家能够高效比较许多不同的缓解策略的结果，从而使研究人员和政策制定者能够进行知情的成本效益分析，以实现最大的气候影响，成本和干扰最小化。人工智能天气模拟模型的速度也使其非常适合实时预测极端天气事件，如野火或飓风，尽管在当前技术水平下，这些模型仍必须依赖更传统的预测模型来提供初始数据同化。例如，物理机器学习模型在从传统模型提供的初始条件时，能够在几秒钟内提供全球31公里分辨率的一周天气变量预测，其准确性可与基础模型相媲美，特别适用于预测飓风和大气河等极端事件。欧洲中期天气预报中心（ECMWF）正开始在试验基础上将这些人工智能增强功能纳入其综合预测系统。如果提供足够的历史野火卫星数据，PCAST也预计机器学习模型将能够实时高精度预测野火的轨迹，从而拯救居民和消防员的生命。这些模型不会取代传统的超级计算机模拟，或实时获取天气和气候数据；但它们将极大增强和补充气候科学的这些基本组成部分，传统模拟虽然较慢但更可靠。

用于初始化、验证和基准测试更快但有时不够准确的人工智能模型。即使使用最复杂的模型，气象预测和气候建模中的完美准确性在物理上可能无法实现，但被极端天气事件措手不及将成为过去式。

揭示宇宙基本物理学 也许每个人都会想知道的最基本问题是：“这一切是从哪里来的？宇宙是如何开始的？使其看起来如此并导致我们今天在这里的规则是什么？”这些问题可以让我们团结在对周围世界之美的敬畏中，并在我们对其运作方式的人类理解的每一次进步中感到兴奋。我们聚在一起观看哈勃或韦伯空间望远镜传来的错综复杂、黑暗而明亮的图像，即使在我们的有生之年，我们对膨胀“大爆炸”宇宙的理解也有了重大进展（现在看来它的膨胀速度正在加快！）探索这些问题的宇宙学家和粒子物理学家是一些最早采用和开发人工智能技术的人，因此，先进人工智能的时代也是基础物理学和宇宙学中令人兴奋的发现时代。科学家们不仅有机会利用人工智能进行新的发现，而且我们知道还有等待被发现的发现。我们对宇宙的惊人详细和令人惊讶的经过充分测试的图景揭示了一些引人注目的难题：是什么“暗物质”使星系保持在一起？或者是加速所有星系之间距离扩大的“暗能量”？最近观察到的在形成后不久就显得古老的星系意味着什么？关于我们宇宙的这些基本问题是我们深入了解其运作方式的主要途径。每当我们更多地了解这些基本问题时，我们就能激励和启发下一代学习者。对于我们的实际目标 and 需求至关重要，正是这些新的基本理解也使我们能够跨越技术鸿沟，开辟新的方式来塑造我们的世界，使我们能够茁壮成长。例如，很难想象比广义相对论更抽象、更不切实际的基本理论了；然而，它支撑着全球定位系统（GPS），以一种我们无法预料的方式解决了位置和导航问题，其经济效益估计达数千亿美元。

人工智能为这一挑战带来了什么可能性？物理学家和宇宙学家在他们的研究中都在使用人工智能——它已经迅速成为实验和观测工作的大多数步骤的特征，包括设计、实施和分析。人工智能的一些用途建立在当前方法的基础上，通过计算模拟数据在不同理论下的表现来比较和测试理论。这些模拟可能是一些最具挑战性的超级计算机任务，因为它们计算每一个粒子、恒星或星系的行为步骤。人工智能的一个关键用途是学习这些模拟中隐藏的更大模式，以便科学家可以快速完成这些超级计算机任务，使得在不到一分钟的时间内就能看到一个月超级计算机工作的极好近似。现在，物理学家和宇宙学家正在他们的研究中广泛使用人工智能，这已经成为实验和观测工作的大多数步骤的特征，包括设计、实施和分析。一些人工智能的用途建立在当前方法的基础上，通过计算模拟数据在不同理论下的表现来比较和测试理论。这些模拟可能是一些最具挑战性的超级计算机任务，因为它们计算每一个粒子、恒星或星系的行为步骤。人工智能的一个关键用途是学习这些模拟中隐藏的更大模式，以便科学家可以快速完成这些超级计算机任务，使得在不到一分钟的时间内就能看到一个月超级计算机工作的极好近似。

研究人员可以扫描数百万种可能的理论，每种理论都有不同的宇宙起始图像，以查看哪种理论会导致我们实际观察到的望远镜数据。在这种AI加速领域，我们可能会看到哪些主要结果？到2030年代末，我们将利用这种AI方法分析南希·格雷·罗曼望远镜的十年数据。通过AI辅助分析这些数据，可能会发现令人惊讶的证据，表明我们的宇宙不会以指数扩张的寒冷死亡结束（正如许多宇宙学家目前所断言的那样），而实际上是循环重新启动，不断发生新的大爆炸。物理学中AI的其他用途利用其在极其复杂数据集中找到模式的能力，其中变量的数量超出我们可以轻松跟踪的范围。然后，研究人员可以寻找令人惊讶的事件：这些事件是极其罕见的新发现的指示器，打破了通常的规则，因此它们在这些AI识别的规律中显眼突出。粒子物理学家举办了比赛，寻找搜索这些“异常”的最佳方法，这些异常可以指引我们发现新的物理学发现。获胜者基于AI，并且也推动了AI技术的边界。这些AI方法可能使我们能够在下一代CERN和Fermilab加速器实验（甚至可能在2040年代）中找到一些极其罕见、意想不到的粒子，这些粒子将有助于建立长期寻求的将引力与其他力量结合起来的“万物理论”。基础物理学和宇宙学建立在对数据的统计分析上，以测试理论，因此需要对数据解释中的概率有深入的理解。这一要求推动了能够处理概率严谨性的AI的数学发展。目标不仅是提供最有可能的答案（“那是一张猫的照片”），这可能是对也可能是错的，而是开发能够提供一系列可能答案及每个答案正确概率的AI系统（“69%的概率是一只猫，22%的概率是一只食蚁兽，8%的概率是一只飞艇，1%的概率是一台冰箱”）。对于关键数字的测量，它将提供一系列可能的值，比如68%的概率、95%的概率或99.9%的概率。评估不确定性对于基础物理学至关重要，而概率严谨的AI将对许多其他科学领域产生深远影响，除了在科学之外的应用中也是无价的。理想情况下，我们设想的AI工具还应该知道其“幻觉”极不可能是正确的。在类似的情况下，物理学家和AI研究人员已经开始探索开发能够提供的AI，可以提供国家科学院的论文 Cranmer, K. 等（2020年5月）。基于模拟推断的前沿。

物理学进展报告 Kasieczka,

G. 等（2021年12月）。LHC奥林匹克运动会2020：异常社区挑战 D 1P5re

Ootceecetidoinn gins oHf itghhe ENnaetriogyn aPlh Aycsaicdse.m y of

Sciences. Dai, B. 和Seljak,

U.（2024年2月），“用于稳健和最佳宇宙学分析的多尺度流。

。在许多其他领域也正在追求相同的方法 科学，如上文提到的关于模拟新材料和新蛋白质用于药物发现。这里的区别在于，在先前的例子中，我们正在模拟已知的过程，而在宇宙学和物理学中，我们可以探索未知的过程。

通过已知的物理原理限制的结果。 由于目前可用的生成式人工智能模型甚至能够在不同领域的行话之间进行翻译，因此涉及基础物理学和宇宙学的几个机会应运而生。最明显的是，我们对宇宙的理解的兴奋和丰富性将变得越来越容易接触到更多的人，因为可以提出有关物理学家当前理解的问题，并得到一个适当水平的答案，以带领提问者迈向下一步。显然，这对公众来说是进入这一领域的宝贵途径。但即使对于物理学或宇宙学子领域的专家来说，这些能力也使他们能够访问其他邻近（或更遥远）的领域，提供了极大加速将它们联系在一起的工作的可能性。这种跨学科的努力通常是科学进步最肥沃的途径。预测这种由人工智能支持的跨学科研究将会发现什么，当然是一种挑战想象力。也许在20年后，研究人员可能会看到一种令人兴奋的结果，即在芯片上的量子计算机设备与黑洞之间建立了一种平行关系，从而开辟了一种新颖的台式测试广义相对论的方法，以及一种强大的新型计时技术。科学家们真正会发现什么无疑会大不相同，但也许会更加令人惊讶。 在过去的二十年里，大量详细的数据集使社会科学家能够更好地了解经济流动性、人口统计差异、市场力量以及离岸和自动化对劳动力市场的影响等问题。这些发现引起了公众的关注，并影响了政策辩论。2023年9月，PCAST的一次公开会议展示了这一社会科学的“数据革命”。 人工智能有潜力在社会科学领域引发新一轮发现和进步。已经有社会科学家开始接受许多机器学习工具。这包括用于文本分析的自然语言处理、数据驱动的模式选择，以及将机器学习预测方法与决策协议相结合。实际例子包括设计... **【此处内容过长，请点击网页底部的原文链接查看完整内容】**

在互联网站点上的实验平台，或者用于信用评分或金融证券定价的高维回归分析等领域，机器学习正迅速发展。一项蓬勃发展的研究领域将机器学习与因果推断结合起来——这是社会科学家用来推断因果关系的方法。社会科学中的机器学习承诺提供更详细、通常是个性化的估计，说明医疗治疗、教育计划以及其他干预或公共政策如何影响不同人群。这些新模型已经在实践中得到采用，并在某些情况下得到改进——例如，互联网平台通过突出显示信息或个性化营销提供给个人。关于这些算法的能力和影响的讨论，特别是在社交媒体平台上，突显了在与人类和社会行为互动的方式中应用人工智能工具时引发的一些复杂性。生成式人工智能方法可能使我们重新构想在研究人类行为、组织和机构时什么样的数据才算是“数据”。例如，社会科学家重视人们幸福感的许多方面——比如他们的身份认同、对工作的自豪感、组织和社区的运作方式——这些并不总是可能或实际通过传统调查方法进行大规模测量。然而，这些信念、态度和互动体现在人们说、写和做的事情中，代表了研究人员才刚刚开始探索的一整类数据。新兴的人工智能模型可以组织并理解这些类型的非结构化数据——比如人们在社交媒体帖子中用来描述他们的城市或社区的词语——在大规模上。将人工智能应用于非结构化语言数据的机会为这些研究提供了巨大的机会，将定性数据和研究方法纳入其中，并将其与收入、生产力、教育水平或寿命等更传统的幸福感度量联系起来。人工智能模型还具有显著潜力，可以生成更好的信息，作为政府政策的基础。关于通货膨胀、生产力、就业和国内生产总值等关键变量的数据是向后看的，而且存在众所周知的缺陷。许多这些指标是基于调查的：然而，调查需要时间，而且获得高响应率是具有挑战性的。同时，大量相关数据以实时形式存在，但因为是非结构化的、不完整的或随时间不一致而未被利用。虽然人工智能模型不能神奇地纠正不正确或有偏见的数据，但它们可能极大地提高我们从非结构化和不完整数据中进行预测的能力，为各级政府的政策决策提供更好的实时数据。此外，这些论点同样适用于远远超出联邦政策范围的更广泛范围的其他大型和复杂组织所遇到的问题和所做出的决策。人工智能模型的另一个有前途的用途是改善联邦项目的交付。许多联邦项目涉及联邦雇员和公民的高度手动流程，例如填写表格或接收大量文件。有许多机会利用人工智能方法更快、更有效地提供政府服务。

在提高可访问性选项方面取得了进展，并以更个性化的方式。最近的工作突出显示了自动化不同流程的一个关注点，即算法可能存在偏见，反映了它们基础数据的特点；例如，主要在一个人群上训练的计算机视觉算法在识别另一个人群时可能不够准确。然而，社会科学研究人员也正在探索AI辅助决策可能减少偏见并产生更公平结果的方式，例如在假释决定中。最近的一项研究揭示了如何修改管理IRS审计选择的算法，以减少对黑人纳税人的不成比例审计。这一新兴的研究领域采用算法方法来思考必须数百万次一致应用的政策实施。科学家们在将AI应用于社会科学研究方面仍处于早期阶段，但已经可以看到我们正在进入一个令人兴奋的时期。PCAST目前正在研究联邦政府可能采取的支持未来社会科学进展的步骤。这些步骤可能包括：（1）促进对联邦收集的数据的安全访问，包括行政数据，这是许多社会科学研究的基础；（2）支持更大规模的“团队”社会科学，这正迅速成为分析大规模复杂数据集的标准；以及（3）鼓励联邦机构与社会科学家合作，使政策设计和实施更具有证据基础，更高效，并更加敏感于人类行为的微妙之处。

3.6 推进生命科学的基础理解

人工智能的进步展示了在生命科学领域加速科学进展的激动人心的前景，使我们对生物学的理解取得突破，对医疗保健、农业、能源和材料等领域产生深远影响。在生物科学领域利用机器学习已经带来了前所未有的进展。这些最新进展只是未来可能性的一瞥，以及可以预期的突破。PCAST相信，AI驱动的工具、分析和结果将从根本上改变我们探索和理解生命的基本构建模块以及将这些构建模块组合成具有广泛应用的生物系统的方式。我们预计，基于AI的技术将在生命科学家的日常生活中发挥作用。这些能力包括将AI方法用作计算显微镜，为科学家提供新的能力来观察和理解构成生物系统功能的蛋白质和其他分子的结构和功能，作为模拟生物系统活动的模拟器，以及用于发现和重新定义新候选分子的控制器。解读细胞的复杂运作——生命的基本单位——是一个持续数个世纪的追求。生物学家面临细胞的复杂性和相互关联性所带来的挑战。

人工智能正在为科学家提供强大的工具，帮助揭示病毒、细胞和生物体的奥秘。从生物学的基础开始，人工智能已经为蛋白质提供了新的视角，这些复杂的分子机器驱动着各种细胞过程。几十年来，确定蛋白质的结构一直是一项昂贵而繁琐的任务。例如，一些基于人工智能的蛋白质折叠预测系统利用机器学习来预测数百万个蛋白质的结构。这些系统从已知蛋白质和结构的数据中学习，以及从基础化学知识（如原子之间距离的物理约束）中学习。高效识别结构的工具正在推动生物科学的进一步发展。超越结构，研究人员最近利用人工智能的力量来解读蛋白质的功能，包括蛋白质如何相互作用，为细胞信号传导、新陈代谢和基因调控提供洞察。人工智能工具还被用于设计蛋白质，以实现与受体和其他靶标特定交互的目标。人工智能驱动的蛋白质设计已经在创建疫苗和新型治疗药物方面取得成功。一些设计方法利用了人工智能驱动的图像生成系统中首次展示的“扩散建模”和“填充和擦除”概念。人工智能工具现在正朝着更远的地平线前进，使生物科学家能够构建和研究定义细胞行为的活动网络的丰富表示。在一种方法中，研究人员使用一种称为图神经网络的人工智能方法直接学习蛋白质和药物之间相互作用的网状结构，并迈出了识别与生理反应相关联的转变性的下一步。通过揭示控制细胞功能细节的复杂机制，研究人员可以更深入地了解疾病的分子基础、治疗药物的作用、药物副作用以及其他基于计算的干预措施，为开发更有效和有针对性的疗法铺平道路。此外，模拟和模拟更大的细胞系统的能力“ ”为构建生物科学的丰富实验工具打开了新的可能性。为生物科学构建丰富的多模态、多尺度基础模型，旨在进行整体细胞建模，是一种有前途的方法。人工智能方法使科学家能够构建多模态生物学表示或“嵌入”，包括蛋白质序列和结构、DNA、RNA表达数据、表型数据（如临床观察）、成像数据以及来自电子健康记录的数据。这些丰富的生物科学模型可以在多个尺度上提供洞察，从分子相互作用到细胞过程再到整个生物体的健康或疾病。构建细胞模拟可能看起来遥不可及，但我们正在看到这方面的初步进展。自然 Jumper, J. et al. (2021年6月). 使用AlphaFold进行高度准确的蛋白质结构预测。Baek, M. et al. (2021年8月). 使用三元序列跟踪神经网络进行准确的蛋白质结构和相互作用预测。HumSciPhenreys, I. et al. (2021年11月). 计算核心真核蛋白质复合物的结构。Krishna, R. et al. (2024年3月). 使用RNN和Atom进行广义生物分子建模和设计。Bao Winfu, K. et al. (2024年2月). 通过折叠扩散进行蛋白质结构生成。Zitnik, M. et al. (2018年6月). 使用图卷积网络模拟多药副作用。

在生物科学领域，基础模型Evo的发展方向是整合大型数据集，结合DNA、RNA和蛋白质数据，以阐明协调细胞整体功能的相互作用。这种多模态和多尺度模型可以提供结果预测和分子行为生成，涵盖从原子到生理学尺度的范围。生物科学基础模型有望成为常见工具，使研究人员能够通过扰动系统的各个方面并注意药物和其他干预措施的效果来进行虚拟实验。我们预计这些人工智能系统将促进计算领域内的探索性工作，例如研究基因工程以使一个或多个特定基因失活或移除、药物治疗或环境应激。这些新工具将使科学家能够更快地探究健康和疾病的基础，例如通过建立癌症模型并探索导致癌症的细胞相互作用和网络如何在模拟中被破坏或“治愈”。这些基于计算机的研究将通过实现虚拟筛选和优化潜在治疗化合物，指导药物的识别或开发，从而在进行昂贵和耗时的实验验证之前提供科学家指导。该系统将能够为科学家提供关于理想实验的指导，根据预期价值的预期结果和目标。我们预计通过使用人工智能驱动的方法在生命科学领域取得重大进展。这些创新技术有望帮助生命科学家揭开无数生物学之谜，为调节细胞机制的能力铺平道路。然而，必须谨慎行事并承担责任，以防止潜在的滥用。特别是在涉及人工智能驱动的蛋白质设计等领域，必须采取措施防止生物威胁的产生，无论这种创造是无意的还是出于恶意目的。建立和遵守包括筛选、监测和及时响应机制在内的最佳实践将对减轻人工智能的无意或恶意使用至关重要，这可能导致新毒素的产生或增强传染病的毒力。在生命科学领域，我们预计将在多个领域取得重大突破。我们预见人工智能工具将在农业、能源、可持续性和医疗保健领域发挥重要作用。新的基于人工智能的工具使我们能够理解、调节和工程化的过程，这将对农业产生重大影响，从分子水平到生态系统水平。我们预见人工智能技术将为农民提供新的种植和保护作物的方法，并在面对气候变化和资源枯竭时建立韧性。一些有前途的例子包括，努力将植物模型系统的基因组学和蛋白质组学技术应用于食品生产系统，以及利用人工智能技术增强农业生产系统的韧性。

在农业领域，利用人工智能技术进行种子研究和开发的过程通常需要十年甚至更长时间。然而，我们预计基于人工智能的方法将把培育具有更高抗性和适应性的种子周期缩短至几年甚至更短。在可持续发展的追求中，一个潜力巨大的创新领域涉及减少牲畜，尤其是牛只产生的甲烷排放。在美国，仅牛只产生的甲烷排放就占到总农业甲烷排放的约45%。这一令人担忧的统计数据凸显了追求开发新颖解决方案以减少这类排放的重要性。生成式人工智能方法为制定旨在优化动物饲料和完善粪便存储和处理政策的创新策略提供了有前途的途径。通过利用人工智能驱动的洞见，存在着一个切实的机会，即开发适合每个农场的定制干预措施，可以显著减少牲畜产生的甲烷排放，从而推动农业部门的可持续发展目标。另一方面，人工智能驱动的发现管道的整合有望通过提供超越依赖化学方法的替代途径，从而彻底改变传统农业实践。这些管道促进了对潜在增强农业系统可持续性和生产力的新生物替代品的探索 and 开发。通过利用人工智能驱动的洞见，研究人员可以开发创新解决方案，不仅可以减轻环境影响，还可以增强农业的韧性。然而，人工智能工具帮助开发可能与欧洲转基因生物一样不符合社区或消费者价值观的方法的能力，再次凸显了保持公众参与和人工智能使用透明度的重要性。人工智能方法将在未来几十年内被认为是推动医疗保健领域革命的动力。超越疾病，人工智能方法可以被用来促进健康、活力和长寿。首先，在将当今广为人知的人工智能解决方案投入使用方面有着巨大的潜力。然而，要实现这一目标，需要加强跨学科合作，以确保人工智能技术的有效应用，从而为农业和医疗保健领域带来更多创新和可持续发展的机会。

。

在生物科学领域利用人工智能应用取得了令人瞩目的突破，发现和设计新型治疗药物，利用人工智能技术早期检测和应对疾病初期迹象，并实现“超个性化”医疗。几十年来，通过使用传统机器学习和推理系统，已经取得了一些进展，这些系统可以向临床医生提供建议，提供专家级别的诊断和预测支持。例如，已经开发了用于预测败血症发作、获得医院相关感染风险、检测可避免的患者护理错误、预测意外患者恶化以及出院后再入院风险的系统。到目前为止，这些人工智能系统的采用还比较有限。然而，未来将这些人工智能技术整合到实践中具有显著潜力，可以提高医疗质量的同时降低医疗成本。我们认为应该积极推动这些传统机器学习技术的发展，重点解决与工作流程整合和迄今为止这些系统在实际服务中遇到的采用挑战，例如确保完全遵守患者隐私法规。与此同时，最近发展的深度学习和生成式人工智能基础模型的应用也有望大大减少行政任务的繁琐，如报告、撰写和总结信息。然而，PCAST认为，在生物科学领域的进步利用将带来最大的收益。在接下来的子章节中，PCAST预计人工智能分子发现和设计将带来创新的浪潮，有望彻底改变我们开发治疗药物的方式。生成式人工智能正成为一种强大的工具，既可以发现又可以设计与特定细胞靶点相互作用的新型分子结构，为以新方式干扰传染性细菌和病毒功能、阻断癌症和自身免疫疾病等疾病途径，或增强免疫反应提供巨大潜力。在分子发现方面，通过将人工智能整合到药物发现流程中，研究人员可以快速筛选数百万个潜在化合物，优先考虑那些具有最有前途的化合物。PCAST预计，基于人工智能的分子发现和设计将在生物科学领域带来创新的浪潮，有望彻底改变我们开发治疗药物的方式。生成式人工智能正成为一种强大的工具，既可以发现又可以设计与特定细胞靶点相互作用的新型分子结构，为以新方式干扰传染性细菌和病毒功能、阻断癌症和自身免疫疾病等疾病途径，或增强免疫反应提供巨大潜力。

通过AI驱动的计算筛选过程，最近发现了一种强大的新抗生素化合物。通过利用AI算法来遍历潜在分子结构的广阔化学空间，研究人员能够发现一种对抗多种病原体具有强效抗菌活性的新化合物，包括那些对现有抗生素产生抗药性的病原体。这种基于发现的方法显示出为提供有关分子结构和功能的新见解，例如导致全新类别的抗生素的巨大潜力。除了发现，生物学中生成式AI模型最有前景的用途之一是设计针对特定治疗靶点的新型分子结构，例如定制分子以阻断病毒的结合区域，阻止其与人体组织相互作用，从而制止病毒感染。现在有几类AI算法可生成新的分子结构以实现特定的目标结构和功能，并具有所需的特性，如结合亲和力、特异性和药代动力学特性。这些生成模型是AI驱动的药物和蛋白质设计兴起领域的基础。除了帮助科学家提出新设计外，AI在药物再利用领域也证明是无价的。通过利用机器学习技术和扩展对人类细胞相互作用网络的了解，研究人员正在通过揭示药物与多样细胞网络和生物途径的相互作用，识别现有已批准药物的新用途。整合多种数据源的AI模型，如基因表达谱、蛋白质相互作用和药物靶标关联，已经使得识别大量药物再利用变得可能，丰富了治疗武器库。PCAST预计AI方法将彻底改变我们提供有关疾病特殊风险和倾向以及早期检测和干预的能力，以阻止疾病的发作或在最早期阶段开始应对和治愈疾病。除了帮助提高像放射学这样基于图像的诊断的准确性和及时性外，AI算法还将在解释有关蛋白质、DNA和RNA存在的复杂分子信号方面发挥越来越关键的作用，实现更早和更准确的疾病检测。通过分析包括基因组、转录组、蛋白质组和代谢组数据的大型数据集，AI算法可以揭示可能被传统分析方法忽视的微妙模式和关联。令人兴奋的AI工作已经使医生能够检测到表明早期癌症的微小分子标志物。这一突破代表了我们诊断先前难以在早期阶段检测到的癌症的能力的重大飞跃。

AI技术的发展轨迹具有深远的意义：早期检测有望显著提高癌症的存活率，通过实现更早的干预和更具针对性的治疗策略。除了癌症检测，AI驱动的数据分析承诺彻底改变其他复杂疾病的诊断和管理。例如，研究人员正在利用AI算法解读大量基因组数据，实现罕见遗传变异的识别及其与疾病易感性的关联，并通过制定早期干预措施来发展治疗策略。AI方法将实现医疗的超个性化，当需要精准和细致的监测和调整时。方向包括利用蛋白质组和基因组信息，以及关于RNA和蛋白质在不同细胞类型中表达模式的数据，进行个性化风险评估和量身定制的治疗计划。通过分析个体的遗传基因组、生活方式因素和环境暴露之间错综复杂的相互作用，AI算法有潜力为患者和医生提供关于疾病易感性、疾病进展和潜在治疗反应的见解。将电子健康记录和表型数据整合到多模态嵌入中为个性化医学铺平了道路，实现了基于患者分子特征优化治疗策略的预测模型的开发。例如，用于研究人类基因组变异的AI工具可以增强我们理解基因变异及其对细胞功能的影响的能力。通过准确识别和解释测序数据中的基因变异，可以确定复杂疾病的分子基础，为个性化疗法和有针对性的干预的发展提供宝贵见解。超个性化、适应性治疗最有前景的应用之一是在癌症领域。癌症是一种复杂疾病，其特征是患者之间的大量遗传和分子异常可能存在显著差异。传统的“一刀切”癌症治疗方法往往效果不佳，因为它们未能考虑到这种固有的多样性。此外，癌症继续演变，可能对治疗产生抵抗，需要更新方法以及对潜在分子网络进行多点攻击，以突破许多癌症表现出的韧性。AI算法可以将多组学数据集与临床和环境因素整合，实现构建个性化疾病模型，捕捉推动个体癌症进展的细胞过程错综复杂网络。然后可以利用这些模型来模拟和预测对各种治疗干预的可能反应，为量身定制的治疗策略铺平道路，优化每位患者独特的分子特征。我们看到利用AI驱动的基因组、转录组和蛋白质组数据分析的巨大机遇，可以实现更早的干预和更具针对性的治疗策略，从而提高癌症患者的生存率。参考文献：Nature Frazer, J. et al. (2021年10月). Disease variant prediction with deep generative models of evolutionary nature. Communications Qi, H. et al. (2021年1月). MVP predicts the pathogenicity of missense variants by deep learning. Medical Oncology. Quazi, S. (2022年6月). Artificial intelligence and machine learning in precision and genomic medicine. The Economist. (2024年3月). The AI doctor will see you now... eventually. Esteban-Gil, A. et al. (2019年10月). ColPortal, an integrative multiomic platform for analysing epigenetic interactions in colorectal cancer.

为了获得对每位患者癌症独特分子特征的前所未有洞察，并以精准治疗靶向做出回应。AI和细胞生物学的整合不仅在癌症领域具有巨大潜力，还在包括心血管疾病、神经系统疾病和自身免疫疾病在内的广泛疾病范围中承诺个性化护理。通过模拟这些疾病中涉及的复杂细胞网络和信号通路，AI系统可以识别潜在的治疗靶点，并根据个体独特的分子和环境因素预测各种干预措施的疗效。从治愈到维持健康和活力，PCAST设想了一个未来的个性化预防策略，基于AI驱动的对细胞过程和个体遗传易感性的洞察。量身定制的生活方式建议、膳食干预和靶向疗法可以主动实施，以减轻疾病风险并促进整体健康，引领向真正预防和主动的医疗保健的范式转变。在这个未来，超个性化医疗保健成为一个广泛可用的现实，得益于AI在细胞生物学和将分子生物学与健康和健康联系起来的创新模型方面的进展，并在涵盖代表性人口范围的医疗数据集上进行训练。这一愿景根植于我们对定义我们作为个体的相似性和差异性的日益加深的理解。AI系统有望利用捕捉我们共享的分子和细胞结构以及我们独特特征的数据，从构成我们蛋白质、DNA和RNA的分子，到驱动我们细胞活动的网络，再延伸到我们整体生理结构。将这一追求提升到更高水平的健康和活力，我们预期AI在超个性化医学中的作用将超越分子和生理方面。未来的AI系统可以帮助我们包含关于我们环境、活动和个人的信息，以及关于过上快乐、富有成效和健康生活的建议。负责任地部署的AI技术可以被利用来极大加速科学进步，并在科学的所有子领域中彻底改变其实践，帮助研究人员找到解决当今全球挑战的新颖有效选项。已经有相当多的兴趣、投资和动力来自私营部门、公共部门和学术界，致力于通过应用AI实现科学的潜在重大进展。然而，使用AI工具所需的资源高度集中在少数私营部门实体中。AI被个人、行业和机构采用的速度有时超过了负责任实践和保障措施的实施。此外，现有的激励结构可能不足以吸引人才从事重要的基础AI研究，这些研究在短期内没有明确的商业化目的，阻碍了主要是公共利益而不是商业利益的AI研究进展。在这一部分，我们提出了关于这些问题的发现和推动负责任和公平部署和应用AI以加速科学发现的建议。我们的发现建立在先前报告的基础上，特别是2021年国家安全委员会关于人工智能的最终报告和2023年NAIRR任务组报告。

根据之前报告的建议，已经部分或完全实施，包括AI权利法案蓝图（2022年）、CHIPS和科学法案（2022年）、NIST AI风险管理框架（2023年）、2021年和2023年NSF AI研究所数量的扩大、2023年关于人工智能发展和使用的广泛行政命令、关于推进机构使用人工智能的治理、创新和风险管理的OMB备忘录、国家AI咨询委员会的建议，以及NAIRR试点项目（于2024年1月启动）。尽管在过去三年里采取了大量和迅速的行动，但如果我们要实现AI在解决紧迫社会和全球挑战方面的全部潜力，还需要做更多工作。

发现1：重要研究受到获取先进模型的限制。AI当前的前沿进展往往与获取大量计算能力和数据密切相关。如今，这种访问往往仅限于资源充裕的组织。这种巨大且不断增长的资源差距有可能限制并对我们的AI研究生态系统产生不利影响... 一个广泛可访问的AI研究网络基础设施，汇集计算资源、数据、测试平台、算法、软件、服务、网络和专业知识，将有助于使美国的AI研究和开发领域民主化，造福所有人。 -

NAIRR工作组报告（2023年） 科学研究是一项协作努力。为了让一组研究人员的发现能够被其他人复制、验证并加以利用，重要的是使用的数据和模型应该向研究社区广泛开放。目前，最大和最强大的AI模型往往是专有的，仅供少数研究人员使用，这使得独立验证或调整使用这些模型的研究变得困难。如果这种情况持续下去，我们有可能在AI领域产生新的“数字鸿沟”，以及潜在的“AI单一文化”，即AI应用依赖极少数提供商或架构。研究和科学方法在不同的假设和方法面对同一问题时蓬勃发展，因此“AI单一文化”将通过错过潜在更好的答案或留下整个未被探索的途径来阻碍科学进步。目前，接触尖端AI模型的成本以及领先技术公司与学术或政府机构之间极端的薪酬差距已经导致... 【以上为部分内容，如需继续翻译，请指示】

215 在这些后续领域中，人力资本的损失是非常宝贵的。此外，要学习如何负责地部署这些强大的人工智能模型，并检测和对抗这些工具不负责任使用的影响，对于广泛研究AI系统的各种弱点和偏见，并量化不同方式准确减轻所有缺陷的有效性，这是一个至关重要的公共利益问题。由于AI模型的性能在很大程度上取决于它们的规模，因此AI基础研究人员能够接触到小型和大型模型至关重要。公共和私营部门在使用AI方面有不同的激励和目标，但它们可以宝贵地共同努力，创建负责任使用AI的文化。 框1.

国家人工智能研究资源（NAIRR） 国家人工智能研究资源（NAIRR）被设想为一个基础设施，帮助民主化获取进行负责任人工智能研究和开发所需的资源，包括计算资源、数据集以及具有用户支持的研究环境，适用于开放和安全项目。美国国家人工智能安全委员会[a]、美国国会[b]、NAIRR任务组[c]和总统办公室[d]都提出了NAIRR作为美国人工智能创新生态系统的关键元素的想法。作为《国家人工智能倡议法》[e]的一部分成立的NAIRR任务组发布了一份报告，提出了建立NAIRR的实施计划，以刺激创新、增加人才多样性、提高研究能力并推进可信赖的人工智能。NAIRR任务组估计，需要联邦每年约5亿美元的投资，连同实物和其他贡献，才能建立NAIRR[c]。NAIRR试点项目于2024年1月24日启动[e]；然而，截至撰写本文时，尚未拨款用于建立和维持NAIRR。 [a]

国家人工智能安全委员会。《2021年最终报告》。 [b]

S. 2714。《2023年7月》。《2023年创造人工智能法案》。 [c] 国家人工智能研究资源任务组。《2023年1月》。《加强和民主化美国人工智能创新生态系统：国家人工智能研究资源实施计划》。 [d] 行政命令14110，88 FR

75191。《2023年11月》。《人工智能的安全、安全和可信赖的开发和使用》。 [e] 国家科学基金会。国家人工智能研究资源试点项目。《2024年4月11日访问》；能源部。

《2024年1月》。《通过国家人工智能研究资源试点项目推进安全和安全的人工智能研究基础设施》；国家人工智能研究资源（NAIRR）试点项目。《2024年4月11日访问》。建议1：扩大现有努力，广泛而公平地共享基本人工智能资源。为了支持这些科学需求以开放和经济有效的方式，鼓励实验和创新，并民主化AI研究的性能和收益，至关重要的是对共享的AI模型、数据集、基准测试和计算资源提供广泛支持。215 Nix,

N. 等人（2024年3月）。《硅谷正在让学者无法承受AI研究》。《华盛顿邮报》。48

不仅使最大的私人人工智能公司受益，还使学术研究人员、国家和联邦实验室、小型公司和非营利组织受益。在美国，这方面最有前途的努力是国家人工智能研究资源（NAIRR），这是一个旨在实现所有这些目标的计划。NAIRR目前正在作为一个小规模为期两年的试点项目运作（见框1）。然而，在目前的规模下，NAIRR试点项目并不能提供与私营部门领先努力规模相当的计算资源或模型（其中大多数模型是封闭源和专有的），也不能提供与其他国家相媲美的努力（见框2）的访问。进一步扩大这些共享资源，特别是全面和及时实施NAIRR，将是为广泛研究人员提供一种成本效益手段，以充分受益于最先进的人工智能模型而避免重复努力的重要第一步；这也将使美国保持其在人工智能能力方面的技术领先地位。PCAST认为NAIRR在促进较小规模AI模型的开发方面具有特别有前景的用例。进一步扩大这些共享资源，特别是那些资源消耗较少但仍足够强大以满足许多研究应用需求的模型是必要的。正如后续建议中所讨论的，一个功能齐全且资源充足的NAIRR（以及联邦和州级AI基础设施努力）未来可以作为走向国家或国际更雄心勃勃的AI基础设施项目的垫脚石。一个全面的NAIRR和随后的AI基础设施对于诸如构建下一代多模态基础模型或开发进一步的大规模数据处理能力（包括基于云的计算以及对大数据源的高效访问）等项目至关重要，这些项目是为了学术研究目的。在2016年至2021年期间，至少有30个国家公开宣布了国家人工智能战略和倡议，包括明确投资于人工智能的研究和开发。例如，印度、新加坡和芬兰分别宣布了全国计划，通过对公共和私人资源的投资，将其国家提升为先进人工智能研究和开发的领导者。中国宣布了新的立法和国家地区机构，以激励人工智能的研究和开发。几个国家的人工智能投资，包括英国、澳大利亚、德国、土耳其、韩国、新西兰、南非和法国，涉及数亿至数十亿美元。例如，英国推出了人工智能研究资源计划，旨在为研究人员提供AI研发的计算能力，成本为3亿英镑。澳大利亚的国家人工智能中心包括4400万澳元用于四个人工智能和数字能力中心，这些中心整合了计算基础设施、数据和现有专家合作，作为其1.24亿澳元人工智能行动计划的一部分，用于开发和采用负责任的人工智能。德国政府承诺在5年内为人工智能和相关努力投入超过50亿欧元。

新闻发布 2024年4月11日，AI Singapore发布了一份关于人工智能领域的最新报告。根据该报告，美国投资者已向中国的人工智能行业投入了数十亿美元。这表明人工智能领域的投资持续增长，展现出巨大的潜力。 另一方面，英国研究与创新部门于2023年11月宣布将投入3亿英镑启动新的人工智能研究资源的第一阶段。这一举措将进一步推动英国在人工智能领域的发展。 此外，澳大利亚政府于2022年3月宣布投入4400万美元用于建设人工智能和数字能力中心，以推动澳大利亚在人工智能领域的发展。而德国联邦教育与研究部门于2023年11月发布了人工智能行动计划，旨在推动人工智能技术的创新与发展。 最新研究表明，尖端研究需要高质量的数据支持。与自然语言处理和计算机视觉等人工智能领域相比，生物学、化学和物理等科学领域相对缺乏适用的公开数据。获取适当的科学数据通常需要高度训练有素的人员和昂贵设备，这使得整个过程相对昂贵且缓慢。 未来将更加依赖数据驱动。特别是机器学习技术，对大规模、高质量的数据集至关重要。例如，成功的蛋白质折叠模型已经在现有的数十万个蛋白质结构数据库上进行了训练。人工智能气候模型依赖于包括历史记录、超级计算机模拟、卫星数据等在内的多样化现有数据。 这些数据对于科学应用和重要挑战的政策决策都具有无可估量的价值。政府拥有许多其他数据集，这些数据对于许多科学应用以及解决我们面临的重要挑战都将是宝贵的资源。

国家。数据.gov门户网站上已经有大量有用的原始数据可用，但正如我们过去的PCAST报告一再发现的那样，这些数据仍然可以释放出更多的价值。例如，利用机器学习来预测和建模野火的努力将受益于获得来自国防部和其他联邦来源的解密卫星数据。同样，多个机构保存有关极端天气事件的物理和经济影响的记录，这些记录可以作为灾难建模的机器学习方法的输入。解决影响许多美国人的另一个重要问题，研究人员对匿名化医疗数据的访问将对评估和改进患者安全措施至关重要。私营部门创造的数据也有改善公共福祉的潜力。例如，正如本节开头引用的内容所指出的，有关化合物和材料的化学和物理性质的数据更常由私营部门持有。许多联邦数据集包含敏感个人信息，因此无法公开发布；然而，仍然可以使用这些数据进行科学研究，同时保护个人的隐私，例如，通过限制对聚合和匿名化信息的访问，通过引入受控数量的随机噪声，或通过使用从基础数据派生的合成数据集。现在正在进行一些试点联邦计划，允许这种受保护的访问，例如国家安全数据服务示范项目和联邦统计研究数据中心。建议2：扩大联邦数据集的安全访问，以满足批准的关键研究需求，并提供适当的保护和防护。PCAST认识到，联邦数据安全访问的问题既敏感又技术复杂，保护个人信息以及在某些情况下国家安全问题必须始终是首要考虑因素。然而，通过批准的研究人员有限访问这些数据的好处，以及向策展资源中心（如未来的NAIRR或国家安全数据服务）发布精心匿名化版本的这些数据集，是巨大的，因此我们强烈鼓励扩大现有的受保护数据访问试点计划。PCAST还鼓励进一步发展关于联邦数据库管理的现有指南，这些指南结合了隐私保护技术的最新发展，例如差分隐私、同态加密、联邦学习以及使用合成数据。这些技术并非实现绝对隐私保护的“灵丹妙药”，这是

PCAST（2023年2月）。现代化野火灭火以保护我们的消防员。

PCAST（2023年4月）。气候变化中的极端天气风险：增强预测和保护社区。

PCAST（2023年9月）。关于患者安全的变革性努力。

国家科学与工程统计中心。国家安全数据服务示范项目。（2024年4月11日访问）。

美国人口普查局。（2023年7月）。联邦统计研究数据中心。

国家科学与技术委员会开放科学分委员会。

（2022年5月）。联邦资助研究数据存储库的理想特征。

Papernot, N. 和 Thakurta, A.（2021年12月）。如何使用差分隐私部署机器学习。

NIST医学关键词词典。

Torkzadehmahani, R. 等（2022年6月）。医学图像中的隐私保护人工智能技术。

Kaissis, G. 等（2020年6月）。医学成像中的安全、隐私保护和联邦学习。

Savage, N.（2023年）。合成数据可能比真实数据更好。

一项不可行的目标，但它们可以显著降低私人信息泄漏的风险。尤其是当原始数据集，特别是标记有相关“元数据”的数据集，已经对传统数据科学和最近基于人工智能的数据分析具有重要的研究价值时，更加精心策划和维护的数据库，专门面向研究用途，将在多个科学领域产生更为深远的影响。目前，高质量的数据策划是一项昂贵且困难的任务，需要大量专业人员的关注；然而，PCAST认为利用现代人工智能技术部分自动化策划过程具有巨大潜力。未来，我们预计可以利用人工智能工具将许多现有的联邦数据集升级为经过精心策划的形式，我们建议将这作为联邦数据共享倡议（如data.gov）的长期目标。除了联邦政府直接持有的科学数据集之外，还有更多具有公共价值的科学数据集值得收集和共享，以一种尊重隐私和知识产权的方式。PCAST支持联邦机构（如国家卫生研究院（NIH）和国家科学基金会（NSF））努力推动从其资助或开展的研究中产生的数据集的负责共享。我们鼓励进一步发展和执行这些要求，并提供足够的支持资源来满足这些要求。我们还应考虑在适当时鼓励共享基于这些数据训练的人工智能模型。从长远来看，我们设想启动更多雄心勃勃的项目，涉及多个公共和私营合作伙伴，创建安全和高质量的科学数据库，作为国家人工智能和数据基础设施的一部分。我们支持类似NSF

FAIROS研究协调网络的倡议，以促进学术界、工业界和联邦政府各部门之间的合作。

“这项试点项目的合作性质，汇集了学术界、工业界、非营利组织和政府部门，旨在促进跨部门合作伙伴关系。工业界的合作可以促进商业可行的人工智能应用和解决方案的开发，通过创造新的市场和收入来源促进经济增长。” - NAIRR报告（2024） 美国工业在基础和应用人工智能研究上投资数十亿美元。NSF最近宣布投资1.4亿美元，设立七个新的国家人工智能研究所。科学数据对于科学研究具有一些理想的特性，请参阅Wilkinson等人（2016年3月）的《科学数据管理和监护的FAIR指导原则》。对于极大规模（PB级）的数据集，下载原始数据的带宽限制也是一个问题；在这种情况下，能够在数据源附近进行分析和模型训练是可取的。

国家卫生研究院。最终的NIH数据管理和共享政策。（2024年4月11日访问）。

国家科学基金会。NSF公共访问倡议。（2024年4月11日访问）。 国家科学基金会。可查找、可访问、可互操作、可重复使用的开放科学研究协调网络（FAIROS RCN）。

（2024年4月11日访问）。

我们的世界数据。人工智能年度私人投资。（2024年4月11日访问）。

国家科学基金会。（2023年5月）。NSF宣布设立7个新的国家人工智能研究所。

涉及NSF内的每个总局，包括新成立的技术、创新和合作伙伴关系（TIP）总局。NAIRR试点计划已经包含了“NAIRR开放”区域，提供开放获取以民主化研究，以及由NIH和DOE共同领导的“NAIRR安全”区域，支持需要隐私和安全的人工智能研究。这些资源提供的人工智能基础设施，以及人工智能本身的跨学科本质，将促进高度跨学科研究，汇集国家和联邦实验室、用户设施、私营行业以及基础和应用科学领域的学者，形成新颖且潜在变革性的合作形式。这种合作的一个例子是NSF材料创新平台，是材料基因组计划的一部分。这些平台正在发展数据共享基础设施，同时利用人工智能工具作为社区共识的一部分，支持跨学术界、工业界、国家和联邦实验室以及联邦机构之间的合作，正如NAIRR任务组制定的NAIRR愿景所概述的那样。联邦资助的学术研究和私营部门研究之间的界限模糊，许多研究人员在学术机构、非营利组织和/或私营公司之间流动，目前有相当大比例的研究与开发得到私营公司的支持。为了充分利用人工智能的潜在研究益处，我们需要支持涉及广泛有前途和富有成效的假设和方法的研究。这可能需要资助机构拓宽他们与工业界合作的立场，以及支持哪些研究人员，以促进创新研究努力和不同领域之间的合作。私营人工智能领域可以为合作带来重要且高度有益的计算资源、专业知识和资金，但需要注意确保行业合作伙伴的市场激励与项目的公共和科学目标一致，同时明确知识产权和许可计划。跨部门合作的例子可能包括从多个来源创建高质量策划的公共科学数据集，创建多模态基础模型，或开发新的量子计算机量子位架构等下一代技术。随着人工智能领域的快速发展，联邦机构和私营部门需要探索和重新评估新兴国家人工智能基础设施的发展，例如提议的NAIRR平台，以保持美国在领导竞争力和创新方面的地位。

发现4：缺乏适当的基准度量标准、验证程序和负责任的实践，人工智能系统可能会产生难以评估质量的不可靠输出，这可能对科学领域及其应用造成危害。就像在事后解决问题的成本通常远高于在系统初始设计阶段解决问题的成本一样，在研究中，如果在最初考虑这些问题，那么更容易解决疏忽、意想不到的后果和意外结果... 如果他们不与相关利益相关者合作，计算研究人员就有可能构建可能适用于自己（或他们的朋友和同事）但可能对其他人群（包括开发者最初未针对的人群）效果不佳的系统。

- 《促进负责任的计算研究》，国家科学院（2022年） 科学研究是一个相互关联的生态系统——一个研究小组产生的发现、模型和数据将影响其他基础和应用领域的研究人员的工作。如果不负责任地使用人工智能工具，可能会对这一生态系统产生负面影响，导致受算法偏见影响的发现，无法复制，缺乏对其不确定性的量化，过度拟合训练数据以牺牲对真实世界数据的准确性，或无法通过更透明的技术（如数值模拟和实验室实验）进行验证。此外，个人数据的隐私或数据创建者的知识产权必须得到保护和尊重。从长远来看，发布不正确标记的合成数据可能导致“数据污染”，这将对进一步的科学工作产生负面影响。如果没有专家监督，完全自动化的人工智能工具部署可能会带来这些风险。学术界必须制定并完善依赖于人工智能衍生输入的研究和数据发布标准。高等教育机构将需要更新他们对初级科学家的培训，使他们能够充分利用新的人工智能工具，能够独立验证和基准化这些工具的输出，并能够在道德和负责任的情况下使用它们。科学进步的加速还将需要研究人员和政府科学机构之间更高水平的双向公众参与。特别是当人工智能开始影响政府和私营部门的政策时，将受到影响的公众的优先事项和价值观必须成为开发输入甚至问题的一部分。一些双重用途研究应用，如生物病原体的功能增强研究，将需要更高水平的负责任、透明和值得信赖的人工智能使用，贯穿科学研究过程的所有阶段。NSF和NIST等机构应继续支持负责任和值得信赖人工智能科学基础研究的发展，包括制定衡量标准基准的标准。法律分析杂志 241 Ben-Shahar, O. (2019年9月)，数据污染。 242 例如，PCAST (2023年8月)。促进科学公众参与。 243 目前通过NSF资助的国家研究人工智能研究所是实现这一目标的一种机制，这些研究所已经资助了一些关于人工智能可信度的研究，以及NSF的安全学习启用系统计划和负责任设计、开发和部署新兴技术计划。

AI模型的属性，如准确性、可复制性、公平性、韧性和可解释性，以及开发工具来评估数据集中的偏见，并区分合成数据和真实世界数据。机构应征求来自这些领域的研究人员以及社会和行为科学相关学科的意见，以帮助制定负责任的AI使用政策，这些政策需要随着我们对这些技术的理论和实践的理解不断更新。总的来说，PCAST强烈鼓励开发者和使用者、政策制定者以及人文、法律和社会科学专家之间就AI治理、使用和影响等更广泛主题进行持续对话。广泛的对话和多样化的意见特别重要，因为在行业围绕治理工具上出现越来越多的整合，少数技术和咨询组织控制着市场的大部分份额。同时，管理科学上使用AI产生的不准确、有偏见、有害或不可复制的发现风险不应该仅仅是科学研究项目的附带思考，而需要从项目的初期规划。PCAST建议联邦资助机构考虑更新其研究责任行为准则，要求使用这些技术的研究人员提交负责任的AI使用计划。这些计划应包括对潜在AI相关风险的评估，如算法偏见、敏感信息披露、结果的透明度或可复制性不足，或潜在的有害应用，并描述采取的措施来减轻这些风险，特别是描述任何自动化过程的监督程序。这些要求可以参照（或与之结合）现有的数据管理计划要求，并参照AI权利法案蓝图以及NIST AI风险管理框架。为了最大程度地减少研究人员的额外行政负担并建立责任文化，机构可以帮助列举主要风险并提供潜在的风险缓解流程。PCAST还建议更新数据管理计划要求，要求披露和记录任何将生成或处理拟议研究项目中使用的数据的AI工具。

发现5：最佳表现需要AI和人类专业知识的结合。政府需要AI系统来增强和补充人类的理解和决策能力，以便充分利用人类和AI的互补优势作为最佳团队。实现这一点仍然是一个挑战。科幻小说经常将人工智能描绘为完全自治的实体，独立运行而无需人类监督，个别AI在复杂任务中作为个别人类的完全替代品。支持者经常将这种自动化水平作为AI技术的长期目标，但取代人类创造力并非是PCAST的预期结果。

本报告中展望。 PCAST认为，AI最有效和有价值的部署，以帮助科学和研究解决紧迫问题，将是以辅助工具的形式。此外，在它们当前的技术成熟水平上，生成式AI资源远远不够可靠，无法实现完全自动化的持续运行，独立于人类监督或干预，特别是在科学领域，准确性、可解释性和可复制性至关重要的领域。此外，人类智能和人工智能的优势在很大程度上是互补的。 AI可以不知疲倦地从庞大数据集中找到模式并执行重复任务，而人类可以从相对较小的数据量中推断和得出结论，进行系统性和战略性推理，并与其他人类和AI助手协调行动。当前的AI工具在展示真正的创造力、分析或高级战略思维方面仍然相当薄弱。即使在一个AI辅助普遍的未来世界中，除了当前最先进的ML模型（如大型语言模型）之外，还会出现其他AI技术，我们预计科学研究将（并且应该）继续由人类科学家指导。科学家将利用独特的人类能力，从相对较小的数据量中得出高层次的结论，以补充AI从大数据集中生成建议和联系的能力，并自动化研究的更常规和乏味的方面。简而言之，PCAST预计，在近期和中期，AI辅助科学的范式往往会优于传统的集成AI辅助方法。 科学企业是一个极好的“沙盒”，可以在其中实践、研究和评估人类与AI助手之间的新合作范式。这种范式的例子可能包括科学顾问AI代理，它们将作为操作复杂软件或实验室设备的复杂自然语言界面；将生成式AI算法与专家人类反馈和形式验证方法（或其他类型的AI系统）相结合，以提高生成式AI的准确性；或使用AI工具来实现跨学科、去中心化或众包研究项目，其规模否则是不可行的。目标不应该是最大程度地实现自动化，而是要利用互补性，让人类研究人员负责任地利用AI辅助进行高质量的科学研究。 资助机构应该认识到新工作流程的出现，并设计灵活的程序、度量标准、资金模型和挑战问题，以鼓励（但不是强制）对组织科学项目的新AI辅助方式进行战略性实验。更广泛地说，激励结构（不仅在资助机构，还包括学术界和科学出版行业）可能需要更新，以支持不同类型的科学贡献，例如用专家人类反馈训练AI系统、开发专门针对科学应用的新AI软件工具，或策划一个高质量且广泛可用的数据集，这些贡献可能不会得到传统研究生产力指标的应有认可。 自然

例如，Boiko, D. 等（2023年12月）。大型语言模型的自主化化学研究。

例如，国家科学基金会。人工智能、形式方法和数学推理。（2024年4月25日访问）。

Donoho, D. （2023年10月）。在奇点处的数据科学。arXiv。 例如，更好的工具用于从难以“标记化”的未标记数据中学习（例如，由于数据的高维度），或用于执行不确定性量化。

这些新的指标将补充传统指标；这些指标的结合将更全面地理解并给予各种类型的智力贡献以认可，这些贡献对科学的推动至关重要。如今，我们的世界迫切需要科学进步来解决各种全球和社会挑战。本报告提供了这些进步的“快照”，这些进步可能在人工智能工具的帮助下实现：新材料以实现向能源高效、低碳经济的过渡；准确的气候模型以预测气候变化和极端天气事件的影响；更有效和数据驱动的基本公共服务交付；等等。当得到人类专家的适当和负责的监督，基于高质量数据训练，并使用可靠的科学技术进行验证时，人工智能工具可以成为创新引擎，可以极大地增强科学家和决策者应对这些挑战的能力。为实现这些目标，先进的模型、数据集和基准需要广泛提供给科学界，负责任的人工智能实践需要编织到科学工作流程中，同时需要在人工智能的基础和各种科学应用中进行持续投资。为了为科学建立这样的人工智能基础设施和文化所需的资源和努力是巨大的，但将带来巨大的回报：一个广泛和民主化的研究和开发生态系统，既开放又安全，在这个生态系统中，科学思想可以迅速转化为实验、原型和成功解决紧迫问题的解决方案。人工智能工具可以补充传统的研究实践，依赖于人工智能辅助但由人类指导的合作和研究范式，仍符合最高的科学验证、可复制性和客观性标准。科学是人工智能技术可以做出巨大积极贡献的领域，只要我们警惕地识别和管理其弱点和风险。虽然人工智能将是这一转变的手段，但最终将是我们人类——科学家和普通公众——通过增强合作能力来应对重要挑战，并更深入地理解我们周围世界的复杂性。我们应该热情地拥抱这一机遇，同时充分意识到如果这些工具未得到负责任部署可能带来的潜在风险，并做出深思熟虑的投资，将这种通过人工智能增强但仍然非常人类导向的科学来应对全球挑战的愿景变为现实。

附录A：术语表

- 功能性：物品可以具有的用途或目的；定义物体可能用途或清楚说明其如何可以或应该使用的特质或属性。
- 人工智能（AI）：指能够执行历史上只有人类才能完成的复杂任务的计算机系统，如推理、做决策或解决问题。
- 芯片：主要指应用特定集成电路（ASIC）。一种专门的计算机处理芯片，可提高特定任务的速度和效率，如神经网络计算。
- 数据清洗：指删除或修复数据集中重复、不完整、不准确或无关部分的过程，以提高数据集的质量，以供分析或训练使用。
- 传统机器学习：一系列已经被探索了几十年的监督式机器学习方法，涉及对诊断和预测系统进行训练。
- 数据污染：这个术语含糊不清，其含义常取决于上下文。本报告中最相关的是：(a) 将由LLMs生成的数据包含到用作或分类为真实用户输入数据的数据流中，以及 (b) 数据未完全和彻底地代表AI工作所涉及的主题。然而，还有许多其他考虑数据污染的方式也是有效的。
- 数据集：相关数据的集合；然而，什么构成数据集并不清晰。例如，人们可以认为与研究项目相关的所有数据，无论数据类型如何，都是一个数据集。
- 深度神经网络模型：受人类大脑内神经元结构启发而设计的一类非常成功的机器学习算法。神经网络由相互连接的层组成，每层包含一个更简单的单元或“节点”，可以执行部分计算。
- 数字孪生体：一个物理系统的高分辨率模型，不断使用该系统的实时数据进行更新。这种孪生体通常依赖传统模拟来模拟系统的基本过程，但也可以使用AI模型来完善、加速或分析这些模拟。
- 现场可编程门阵列：一种集成电路（IC）类型，可为快速原型设计和最终系统设计开发自定义逻辑。FPGA与其他定制或半定制IC不同，因为其固有的灵活性允许通过软件下载进行编程和重新编程，以适应其设计所在的更大系统不断发展的需求。FPGA非常适用于当今增长最快的应用，如边缘计算、人工智能（AI）、系统安全、5G、工厂自动化和机器人技术。

- **基础模型：** 一个机器学习模型，经过广泛的数据训练（通常需要大量计算资源），然后可以相对廉价地进行微调，以适用于更专业的应用。
- **生成式人工智能：** 一种人工智能，能够生成新的内容，如代码、图像、音乐、文本、模拟、3D物体、视频等。被认为是人工智能研究和发展的的重要组成部分，因为它有潜力彻底改变许多行业，包括娱乐、艺术和设计。
- **人工智能幻觉：**
由人工智能生成的内容，可能是荒谬的或与提供的源内容不符[; ...] 主要有两种幻觉类型，即内在幻觉和外在幻觉。[内在幻觉是]生成的输出与源内容相矛盾；[外在幻觉是]生成的输出无法从源内容中验证（即输出既不能被支持也不能被反驳）。
- **图像生成模型：** AI图像生成器是使用深度学习算法从头开始生成数字图像（通常是文本）或修改现有图像（通常是图像）的计算机程序。这些生成器可以创建高度逼真和复杂的图像，包括风景、人脸、物体等。它们在各种领域具有实际应用，如艺术、设计、广告和游戏。
- **联合表示：**
机器学习模型从多种数据或特征学习的一种方式，例如结合图像和文本。
- **大型语言模型（LLMs）：** 一类使用深度学习算法的语言模型，训练在极其庞大的文本数据集上，这些数据集可能达到多个TB的大小。LLMs可以分为两种类型：生成式或判别式。生成式LLMs是输出文本的模型，例如回答问题或甚至撰写特定主题的文章。它们通常是无监督或半监督学习模型，用于预测给定任务的响应。判别式LLMs是监督学习模型，通常专注于对文本进行分类，例如确定文本是由人类还是人工智能生成的。
- **机器学习（ML）：**
一系列利用统计推断“学习”的技术，通过对大型数据集上的AI模型进行训练。
- **多模态模型：** 一个扩展了生成式人工智能的机器学习模型，能够处理来自不同模态的信息，包括图像、视频和文本。
- **多尺度建模：**
一种建模策略，同时使用不同尺度的多个模型来描述一个系统。
- **开源：**
披露模型结构和训练过程的模型，但不一定包括最终权重。
- **开放权重：**
披露在训练过程结束时获得的权重的模型。
- **个性化医学：**
一种利用个人基因或蛋白质信息来预防、诊断或治疗疾病的医学形式。

资料来源：
纽约大学图书馆。《机器与社会》。2024年4月。 Google
Cloud。多模态人工智能。2024年4月25日访问。
国家癌症研究所。个性化医学。2024年4月25日访问。

量子计算机：量子计算的一般术语，指进行量子计算的设备（无论是理论上的还是实际实现的）。量子计算机可以是模拟的或门控的，通用的或非通用的，以及嘈杂的或容错的。强化学习：

一种训练算法以执行期望动作的方法，要求该过程最大化给定的函数或结果。

研究责任行为准则：共享价值观，包括研究行为的“规则”，具有科学诚信，无论是在规划研究、进行研究还是报告和审查研究时。路由：

在任何互连网络中选择最佳路径的过程，连接节点之间的路径选择。超导体：

能够高效传导电流，几乎没有能量损失的材料。超导性目前需要低温。

合成数据生成：

使用种子数据创建人工数据的过程，这些人工数据具有种子数据的一些统计特征。

系统性偏见：由特定机构的程序和实践导致的结果，这些机构以某种方式使某些社会群体受益或受到偏爱，而使其他社会群体处于不利或被贬低的地位。这不一定是任何有意识的偏见或歧视的结果，而是多数人遵循现有规则或规范的结果。热电材料：

将热量转化为电能或通过施加电压产生冷的固态半导体材料。附加词汇资源： •

IEEE道德对齐设计词汇表 • NIST的计算机科学词汇表 •

NIST的值得信赖的人工智能词汇表 国家科学、工程和医学院出版社

国家科学、工程和医学院。《量子计算：进展与前景》（2019年）。Steneck, N.（2007年8月）。《研究责任行为导论》。研究诚信办公室。

国家标准与技术研究所。合成数据生成。（2024年4月25日访问）。Shevelkov, A.（2010年）。《热电材料：简介》。

附录B：外部专家咨询 PCAST寻求了一组额外专家和利益相关者的意见。PCAST对以下列出的分享专业知识的人士表示感谢。他们没有审查报告草稿，他们愿意就特定问题与PCAST进行交流并不意味着对本文所表达的观点的认可。对本报告中的观点、发现和建议的责任，以及任何事实或解释错误，均由PCAST承担。

研究与发展主任 物理学教授
布莱恩·巴赫兰 克雷恩大学 核技术与战略公共政策助理主任
信息系统海因茨学院院长 科学与技术政策办公室 查理斯·麦洛姆 亚利桑那大学
白宫全球人工智能伦理学家 AI特别顾问 达克塔·莫布达 白宫国家安全委员会
白宫技术部主任助理 科学与技术政策办公室 约翰·杰弗斯 法律研究主席
德克萨斯大学奥斯汀分校 成员 国家人工智能咨询委员会 成员
国家人工智能咨询委员会 科学计算与成像主任 犹他州立大学 教授 伯克利信息学院
内容真实性高级主任 倡议成员 沃德·沙恩 国家人工智能倡议办公室主任
首席项目经理 科学与技术政策办公室 迈克尔·乔丹 白宫
电气工程、计算机科学与统计学系教授 民主与公民参与总统特别助理
加州大学伯克利分校 白宫 61

致谢词 PCAST成员感谢科学技术政策办公室（OSTP）工作人员在本报告准备过程中的各种贡献。我们感谢国防分析研究所（IDA）科学技术政策研究所（STPI）团队，特别是Emily Grumbling、Sara Jordan、Matthew Ishimaru和Kate Ross在研究和分析方面的帮助。我们还感谢许多同事在附录B中列出的正式咨询之外提供的见解，包括： Anima Anandkumar Tanya Klowden Tapio Schneider Ivan K Schuller Vivian Zapf 62