AI
在全球发展中的
PLAYBOOK

TABLE OF CONTENTS

免责声明

这份文件包含外部资源的链接。这些链接反映了起草时可用的URL信息。美国不对链接内容的准确性做出任何声明，链接也不构成对链接材料的美国官方认可。除非文中明确链接到美国政府资源，否则应视为承认该资源作者的观点，并不代表美国政府采纳了这些观点或立场。

术语表

ADD ： 推进数字民主

AI RMF: 国家标准与技术研究院人工智能风险管理框架

AI ： 人工智能

ANCIR: 非洲调查报告中心网络

API ： 应用程序编程接口

闭路电视

CEDI ： 清洁能源需求倡议

cfA ： 非洲代码

COVID - 19 ： 2019 年冠状病毒病

DOE ： 美国能源部

DPGA ： 数字公共产品联盟

DPG ： 数字公共产品

FAAST: 科学 ， 安全和技术的人工智能前沿

FCDO: 联合王国外交、联邦和发展办公室

FLAIR ： 第一语言 AI 现实

GAIRA ： 全球人工智能研究议程

GIZ: Deutsche Gesellschaft f ü r Internationale Zusammenarbeit (德国国际合作组织)

GPAI ： AI 全球伙伴关系

GPT - 4 ： 生成预训练变压器 4

GPU ： 图形处理单元

惠普 ： 惠普

HPC: 高性能计算

IBM ： 国际商用机器公司

IDRC: 加拿大国际发展研究中心

劳工组织 ： 国际劳工组织

IMDA: 新加坡 Infocomm 媒体发展局

LGBTQI +: 女同性恋、男同性恋、双性恋、变性人、酷儿、双性恋和其他

LLM ： 大型语言模型

低收入和中等收入国家 ： 低收入和中等收入国家

MCV ： Mozilla 普通语音

Mila ： 魁北克人工智能研究所

NASA ： 美国国家航空航天局

NCII ： 非自愿的亲密意象

NIST ： 国家标准与技术研究所

NTIA: 国家电信和信息管理局

开放 RAN: 开放无线接入网

OPG ： 开放政府伙伴关系

Playbook ： AI 在全球发展

研发 ： 研发

RCC ： 负责任的计算挑战

RFI: 索取资料

可持续发展目标 ： 联合国可持续发展目标

中小企业 ： 中小企业

STEM ： 科学 ， 技术 ， 工程和数学

结核病 ： 结核病

美国 ： 美国

联合国 ： 联合国

UNESCO: 联合国教育、科学及文化组织

《气候公约》: 联合国气候变化框架公约秘书处 “ 联合国气候变化 ”

联合国大会 ： 联合国大会

URL ： 统一资源定位符 ， 网址

美国国际开发署 ： 美国国际开发署

WIPO ： 世界知识产权组织

图片来源 ： 美国国际开发署

在我们努力实现共同繁荣的过程中 ， 世界正处于一个不稳定的时刻。全球社会正步入正轨 just 17% 到2030年实现联合国可持续发展目标（SDGs），但在许多领域进展已经停滞或倒退。当有效且负责任地使用时，人工智能（AI）有潜力加速可持续发展目标的进展并缩小数字鸿沟，但也带来了可能进一步阻碍这些目标实现的风险。在适当的促进环境下和由多方利益相关者组成的生态系统中，AI可以在健康、教育、农业、能源、制造业以及提供公共服务等领域提高效率并加速发展成果。美国致力于确保人工智能带来的好处在全球范围内公平共享。

正如美国总统约瑟夫·R·拜登在2023年9月联合国大会发言中所说，美国致力于“确保我们利用人工智能的力量造福社会”。2024年3月，美国在联合国大会上牵头通过了首份专门关于人工智能的决议，“抓住安全、可靠和值得信赖的人工智能系统为可持续发展带来的机遇”，该决议确立了全球共识的两大关键目标：利用人工智能的潜力并减轻其风险。美国政府继续扩大在全球范围内具有重大影响且本地相关的人工智能解决方案的工作，并与私营部门、民间社会和国际伙伴合作，以增强人工智能的生态系统，同时确保采取足够的保障措施以保护人权和公共安全。

基于这些努力以及美国政府在数字发展和促进安全、可靠和可信赖的人工智能方面长期的历史，全球发展 playbook 中的人工智能部分 [1] （蓝皮书）描绘了一套关键行动方案，供美国及其发展伙伴（如其他政府、私营部门和慈善机构）参考使用。

促进全球负责任的AI生态系统，推动可持续发展，创造合作伙伴关系和协作的机会，并应对世界上的某些最大挑战。

这份手册是一条发展能力、生态系统、框架、合作伙伴关系、应用和机构的道路，以利用安全、可靠和可信赖的人工智能推动可持续发展。所有这些特性构成了一个人工智能的良好治理制度，这是手册中的一个贯穿始终的主题。良好的治理措施可以培养公民之间的信任，信任可以促进采纳，而采纳又可以推动创新。

《手册》旨在综合现有研究成果，并为设计、部署和使用安全、安全可靠且可信赖的人工智能以促进可持续发展提供建议——包括美国计划采取的一些步骤以支持全球人工智能生态系统。为了提供实用指导，

图片来源 ： Jack Gordon for USAID

playbook 还包括一系列案例研究。这些案例研究突显了在识别领域中表现出色的工作的现有举措和组织。通过展示这些实际案例，playbook 的目标是激发并引导其他人。

该手册特别适用于希望为低收入和中等收入国家（LMICs）的可持续发展作出贡献的发展实践者、政策制定者、发展和慈善组织以及私营部门参与者。随着人工智能技术的不断进步，这些利益相关方群体都应了解其带来的益处与风险，并认识到各自在构建保障措施和支持负责任的人工智能生态系统方面的重要角色，以推动全球发展。

人工智能技术的模糊和快速变化的性质意味着单一利益相关方群体无法独自成功地实现人工智能带来的益处或减轻其风险。政府、私营部门、民间社会以及学术界（国内外和国与国之间）之间的合作是我们在这份文件中方法的核心要素。

1 《全球发展 playbook 中的人工智能》（Playbook）根据美国关于安全、可靠和值得信赖的人工智能开发和使用的行政命令第 11(c)(i) 条发布。该行政命令要求美国国务卿和美国国际开发署署长，与国家标准与技术研究院（NIST）协调，发布一份人工智能在全球发展领域的 playbook，该 playbook 将 NIST 的人工智能风险管理框架（AI RMF）的原则、准则和推荐实践纳入社会、技术、经济、治理、人权和安全条件中，并借鉴全球发展领域中人工智能项目经验教训。

图片来源 ： Angela Rucker

The Playbook 概述了与全球发展和人道主义援助相关的关键主题下的机会与挑战。识别并应对人工智能的风险是实现其潜在利益的第一步。本手册与现有框架保持一致，并在此基础上进一步扩展。 AI 风险管理框架 从NIST识别出如何在设计、部署和使用人工智能的过程中减轻风险（如对个人、社区、组织、生态系统或社会的伤害）。

The Playbook 的建议——提炼自与数百名政府官员、非政府组织、科技公司和初创企业以及来自世界各地的个人进行咨询的内容——围绕几个核心领域构建而成：

- **增强能力 ， 在所有部门和级别推广与 AI 相关的技能 ， 并保护劳动力。** 通过为更广泛的人群提供人工智能技能，以填补人工智能劳动力市场的缺口，各国可以抓住经济机遇、推动本地创新并创造有助于可持续发展的就业机会。同时，人工智能带来的劳动市场变化要求建立 robust 的社会安全网、采取行动预防和解决任何新的工人权利风险，并包括工人群体和工会在内的社会对话过程。

- **建立可信和可持续的数字基础设施。** 此外，广泛的互联网连接和可靠的能源资源在更广泛的发展中发挥关键作用，可以促进可持续发展的人工智能。通过合作提升数字基础设施不仅能够改善对人工智能技术的访问，还能刺激经济增长，并使社区能够利用人工智能应对当地挑战并提高生活质量。应优先考虑人工智能系统的能效，并仔细考虑气候变化和环境影响，以避免延续现有问题。

- **扩大对数据存储和计算资源的访问。** AI创新者需要访问大规模计算和数据存储资源以支持模型推理、训练和部署。通过扩展应用程序编程接口（APIs）、可信赖的云计算服务及其他资源的访问权限，使计算更加经济实惠且易于获取，可以加速开发满足当地需求的AI应用。

- **创建具有代表性的本地相关数据集和保存文化遗产。** 本地化、语言上和文化上相关的数据集，能够反映低收入和中等收入国家（LMICs）的种族、 Ethnicity多样性及其当地背景和现实情况，可以促进符合社区需求的AI模型的发展和应用，使其更适合当地环境，并推动可持续发展。通过构建具有代表性的数据集，利益相关方可以共同努力，使AI解决方案更加准确、公正且具有影响力，从而最终促进包容性增长和创新。

- **制定战略 ，在实践中实现人工智能的承诺。** 严格的AI开发环境下的测试和基准测试以及广泛公开分享研究发现是确保AI及AI驱动干预措施基于有效证据的基础。这些证据对于指导AI解决方案的扩大应用至关重要，有助于确保它们能够提供广泛的公共价值并解决已有的发展挑战。系统性评估AI项目可以突出成功案例，提供最佳实践，并引导投资，从而在不同领域和地区复制并扩大有效的解决方案。

- **推进良好治理框架 ，以开发和使用安全和尊重权利的人工智能系统。** 人工智能有可能被恶意行为者以危害个人和社会的方式滥用，例如通过非法或任意的监视、促进网络威胁、信息操纵、政治操控或深度合成——包括合成的非自愿亲密图像和儿童色情材料。必须采取积极措施推进良好的治理框架，以保护民主程序；增强透明度；保护网络安全、知识产权和隐私；确保与适用法律框架的一致性；确保商品和服务的公平获取；并促进人权，这有助于在社会中培养信任和韧性。

- **通过开放性、透明度和可解释性培养对人工智能的信任。** 提高AI模型、开发过程、组织实践以及AI政策制定的透明度可以增强可靠性、公平性、知识产权保护和尊重。拥抱AI系统设计、部署和使用中的开放性和可解释性可以促进采用，增强透明度，并提升对AI系统的信任。

- **可持续部署 AI 和气候行动。** 抓住人工智能在各行业节能方面可以作出贡献的机会，在应对气候变化方面将发挥紧急且重要的作用。强调可持续性将减少人工智能技术的净环境足迹，同时利用人工智能提高各行各业的能源效率，从而为全球应对气候变化的努力做出贡献。

美国坚定支持可持续发展，并在全球范围内推动负责任的人工智能使用，以及增强人工智能安全保障的合作努力。通过政策、资金支持、参与、伙伴关系及其他机制来应对《手册》中列出的挑战，美国希望鼓励和吸引更多利益相关方承诺采用安全、可靠且可信赖的人工智能技术的设计、部署和使用。

图片来源 ： Andr é Josu é Anchecta Oseguera

INTRODUCTION

为了最大限度地发挥人工智能的优势并减轻其风险，支持设计、部署和使用人工智能系统的生态系统及人员至关重要。在适当的安全保障和有利条件下，人工智能可以推动各行业的更高效率，并加速包括健康、教育、农业和粮食安全、能源、治理和公共服务交付在内的各个领域的开发成果。

推广AI的好处需要多利益相关者和跨学科的合作，利用当地社区、私营部门、工人及其组织、学术界、民间社会以及其他方面的专业知识来识别机会、导航障碍并管理风险，以培育负责任的AI生态系统。

Playbooks 是美国的关键交付成果 关于安全、可靠和可信的人工智能开发和使用的行政命令 ，发布于2023年10月30日。这份Playbook旨在 characterization 超越美国边境的AI机会和挑战，特别是在低收入和中等收入国家（LMICs），并提供建议以确保参与AI生态系统各方能够合作、建立伙伴关系，并将优先事项与可持续发展目标对齐，包括保护人权的自由行使。 [2]

世界各地在安全、可靠和值得信赖的人工智能方面存在显著差距 ， 几乎 全世界有 60 亿人 生活在缺乏稳健负责任人工智能治理措施的国家。在隐私、数据保护、问责制、透明度、偏见、歧视、网络安全和语言多样性等领域，世界还有很长的路要走，以确保人工智能系统以负责任的方式设计和部署。

负责任的 AI

一种将人工智能的设计、部署和使用与民主、可靠性、安全性、安全性、可信度、包容性、透明性、隐私权、网络安全、公平性、人权和可问责性等核心价值观相结合的方法。 [3]

2 人工智能是指基于机器的系统，在给定一组由人类定义的目标时，可以进行预测、建议或决策，从而影响现实或虚拟环境。人工智能系统利用机器和人类输入来感知现实和虚拟环境；通过自动化分析将这些感知抽象为模型；并利用模型推理来制定信息或行动的选项。

本地 AI 生态系统是什么样的 ？

在任何国家，人工智能生态系统都是更广泛的数字生态系统的重要组成部分。这包括数字数据的收集、存储、管理和共享所涉及的过程和政策；参与研究、开发和私营部门活动的当地劳动力；积极参与的社会团体；以及治理和政策框架。该生态系统涉及多样化的利益相关方，包括政府机构、企业和初创公司、大学和培训机构，以及社会团体组织。

理解人工智能对发展的影响需要承认各国、各地区、各社区和不同情境下的差异性。即使在同一地区的国家或具有相似收入水平的国家之间，也存在显著的人工智能技术获取差异以及人工智能应用方式的不同，最终以多样化的方式影响个人和社区。

全球发展的有意义的方法必须优先考虑由当地领导并相关的解决方案，而不仅仅是国家层面，还应包括各个社区内部。这涉及理解并适应每个背景下独特的政治、社会、文化、经济和环境条件，认识到没有一种“一刀切”的方法。特别是在推动负责任的人工智能部署方面，这一点尤为重要，因为它确保人工智能技术能够针对不同人群的具体需求和挑战进行定制。此外，这种方法还允许实施针对每个背景下特定风险的防护措施。

有多种现有的工具可供利益相关者使用，以更好地了解AI在当地情境中的作用，包括指数等指标。 负责任的 AI 全球指数 ，the 斯坦福大学的 AI 指数 ，the 教科文组织准备情况评估方法 ，the 政府 AI 就绪指数 ，以及微软AI参与指数。这些指标在不同层面追踪AI的发展，并有助于了解哪些AI生态系统方面的内容可以优先考虑或加强。

进近

在起草《全球发展 playbook》中的AI章节之前，美国国际开发署（USAID）、美国国务院和美国商务部（以及其他跨机构合作伙伴）的领域专家发布了相关材料。 信息请求(RFI) 旨在从公众和人工智能发展生态系统中的相关合作伙伴收集广泛的观点。超过60个组织和个人提供了回应。本次RFI中确定的主题受到关于全球发展和人道主义援助中人工智能的相关文献、报告、倡议、案例研究以及数据的桌面研究结果的影响。研究团队随后与来自世界各地不同发展阶段国家政府、民间社会、发展组织、慈善机构和私营部门的代表进行了13场虚拟咨询会议。《手册》是对通过研究、信息请求以及咨询会议收集的信息进行提炼和分析的结果。 [4] .

虽然《全球发展 playbook 中的人工智能》旨在解决从咨询中收集到的许多基础问题，但我们认识到它无法涵盖所有可能的人工智能关注领域，并且会选择最具相关性的领域，同时尽可能提供全面的方法。我们鼓励读者通过 aiplaybook@usaid.gov 向美国政府就人工智能方面的一般反馈、建议或未来应纳入的方法重点与我们联系。

《手册》探讨了在全球发展背景下，确保人工智能（AI）的安全、安全和可信设计、部署和使用所面临的生态系统级机遇与挑战。基于《人工智能风险管理框架》中识别的风险，存在一些独特的问题需要考虑：构建安全、安全和可信的人工智能系统以应对全球发展挑战所需具备哪些必要的技能？哪些合作伙伴关系和倡议可以弥补低收入和中等收入国家（LMICs）在安全、安全和可信的人工智能开发过程中存在的基础设施缺口？有限互联网连接社区中的数据多样性和平等代表性意味着什么？如何设计和部署人工智能系统，使其能够体现人类体验的丰富多样性，并且包括当地用户和受影响社区的参与与合作？在评估人工智能模型时存在哪些差距，这些差距可能会使它们对当地社区最有用？如何防止和减轻有害的人工智能应用？可以采取哪些激励措施来确保人工智能的适当开放性、透明性和可解释性？最后，如何使人工智能的设计、部署和使用与环境可持续性和应对气候变化的目标相一致？

这些考虑不仅影响构建AI系统的组织，还影响塑造这些系统设计、部署、使用及其最终影响的生态系统中的所有参与者。因此，《playbook》从广泛的视角出发，涵盖了各种行业和议题领域。[5] 旨在将整体格局转变为利用未来AI Actors面临的机会并减轻相关风险。

美国努力在全球发展中促进负责任的人工智能

鉴于人工智能固有的交叉性质及其许多交叉效应，美国政府创造了 AI. gov 网站，全面审查为支持安全、可靠和值得信赖的人工智能而采取的行动。

特别是在全球发展的背景下，美国国际开发署发起了 AI 行动计划 在2022年，美国政府制定了战略愿景，旨在推动负责任的人工智能发展，以应对全球关键挑战。行动计划概述了美国国际开发署（USAID）如何将人工智能整合到不同领域的开发项目中，如何支持数字生态系统以促进负责任的人工智能，以及如何与其他机构合作制定人工智能在发展领域的全球议程。该行动计划重点关注在合作伙伴国家建立当地能力、确保尊重人权以及与多元利益相关方开展协作。在这些合作伙伴关系中，美国国际开发署联合起来 与其他国际发展援助机构合作，支持全球负责任的人工智能生态系统，包括英国的Foreign, Commonwealth and Development Office (FCDO)、加拿大的International Development Research Centre (IDRC)、德国的Corporation for International Cooperation (GIZ)以及比尔及梅琳达·盖茨基金会。此外，USAID与IDRC共同启动了 数字时代的捐助者人权原则 在联合国互联网治理论坛上，该论坛旨在将人权和民主价值置于与数字参与和投资相关的政府行动中心位置。这些原则特别强调在发展和人道主义援助项目中确保安全和安全性的必要性，并指出捐助者有责任进行人权尽职调查评估。

通过研究和咨询过程，最为相关的问题领域包括：医疗领域的AI解决方案；教育领域的AI解决方案；农业和粮食安全领域的AI解决方案；能源领域，包括能源消耗和生产；气候领域，包括气候变化的缓解与适应；交通领域，包括道路、港口、机场、铁路等；以及公共服务业的提供。此外，我们还听到了对金融sector和工业制造领域解决方案的需求，但程度较低。

2023 年 9 月 ， 布林肯国务卿在一次关于“ 人工智能加速可持续发展目标的进展：应对社会最大的挑战 在美国联合国高级周会议上与国会合作的边缘 ，美国承诺在未来五年内提供1500万美元的外国援助，以促进全球范围内AI的负责任使用和治理，其中包括利用AI工具帮助各国实现可持续发展目标（SDGs）的相关项目。此外，美国私营部门承诺投入2100万美元，利用AI技术增强SDGs，并为发展中国家的四百万人提供培训，其中至少包括200万人在非洲地区。除了美国之外，一些共同主办国也宣布了未来的重要资金和资源投入，总计超过90亿美元，用于支持AI初创企业和成长型企业以及全球风险分析。该活动突显了在健康、教育、农业和粮食安全、能源等领域，发展中国家私营部门的倡议。

图片来源 ： 美国国务院

The 关于安全、可靠和可信的人工智能开发和使用的行政命令 发布的于2023年10月概述了政府在人工智能领域的整体策略。这包括通过多边论坛推广全球人工智能实践、加强与盟友和国际组织的合作以应对共同的人工智能挑战，并在全球范围内支持能力建设项目。《人工智能促进全球发展手册》是根据行政命令制定的一项成果，旨在推动负责任的人工智能使用以促进全球福祉，同时减轻潜在的风险和挑战。 全球 AI 研究议程 另一个在AI行政命令中识别的交付物是Playbook的配套文件。研究议程指导AI相关研究的目标和实施，在美国边境以外的背景下，并包括旨在确保AI的安全、负责任、有益且可持续开发和采用的原则、指南、优先事项和最佳实践。

2024 年 3 月 21 日 ， 联合国大会一致通过了以美国为首的决议。 把握安全、可靠且可信赖的人工智能系统机遇以促进可持续发展 “这一决议强调了全球致力于利用人工智能潜力实现可持续发展目标的承诺。总体而言，该决议呼吁通过国际合作确保人工智能系统的负责任开发，并解决与隐私、安全和不公平相关的风险。”

最后，作为支持行政命令第11条努力的一部分，美国国务院与NIST和USAID协调开发了一个“ 人工智能与人权风险管理简介 “本NIST人工智能风险管理框架（AI RMF）概况借鉴了多利益相关方咨询，以识别可能由人工智能的设计、部署和使用引发的潜在无意和故意的人权影响案例。随后，该概况为全球政府、私营部门和民间社会提供了自愿建议，指导如何将人权考量纳入AI RMF中的特定行动，以管理此类风险。

全球发展中人工智能的机遇、挑战和建议

增强能力 ， 在所有部门和级别推广与 AI 相关的技能 ， 并保护劳动力

在许多 LMIC 中 ， 有一个关键的 短缺 拥有AI专长，阻碍了AI为可持续发展带来的潜在好处。在低收入和中等收入国家（包括公共部门）投资负责任的AI技能对于构建AI workforce并确保安全、可靠和可信赖的AI的好处能够公平分配至关重要。

机遇与挑战 ：

许多国家在发展强大的人工智能劳动力以及在面对技术变革时构建韧性劳动力市场方面面临挑战。然而，通过投资技能提升并促进尊重劳动者权益，各国有机会培养多元的人才库，并创建韧性劳动力市场，从而能够根据各自独特的背景开发和适应定制化的人工智能工具，推动可持续发展。

多样化人群的多样化技能

开发广泛领域的本地相关AI专长可以导致针对当地和区域挑战量身打造的创新解决方案的创造。 医疗保健 ， 农业 ， 教育 和其他领域。

The development and use of safe, secure, and trustworthy AI technologies benefits from diverse pools of talent, representing all segments of the population. 如上所示： UNESCO ，来自低收入和中等收入国家（LMICs）的声音和视角对于识别和纠正偏见、确保AI解决方案适应当地环境至关重要。然而，总体而言，这些声音在国际讨论中的参与度仍然有限。 缺乏多样性 在不同阶段的AI设计、部署和使用所覆盖的技术生态系统中。例如，在AI与数据科学领域中，涉及该技术研究和开发的专业人士。 少于 22% 在人工智能与数据科学领域，专业人士中女性的比例较低。这种缺乏多样性不仅基于性别——种族或 Ethnic 少数民族、残疾人、农村背景的人群、老年人以及其他群体也面临着障碍。

相关学校课程也在这一背景下显得尤为重要。虽然负责任的人工智能技能应包括数据科学、机器学习、软件开发、网络安全以及特定行业和技术领域的相关技术技能，但负责任的人工智能教育还需要对人工智能的理解。 社会影响 ，包括探讨偏见、公平性和透明性等问题如何影响全球社区、地区和行业中的数字鸿沟。从性别研究、历史、网络安全与隐私、行为经济学、心理学、政治科学、人类学、社会学和语言学等领域的专家视角出发，这些观点对于指导系统的设计以符合预期用户的需求、管理风险以及解析和应对人工智能的社会影响至关重要。这些都是负责任地开发人工智能的教育基础。

随着人工智能的不断演进，投资于再培训、重新技能化、提升技能以及职业发展项目以帮助工人过渡到新角色变得至关重要。政府、企业和教育机构可以合作制定全面的 再培训计划 使工人具备由人工智能驱动的经济所要求的技能。这些计划应该双管齐下，既关注技术技能的提升，也关注通用技能的培养；具体而言，应努力教育个人在数字素养、AI的应用影响、AI工具的应用与开发方法（特别是在传统行业）、预见、识别、情境化和应对AI系统影响的方法、利益相关者和社区参与、解决问题以及适应性等方面的知识。在线课程、职业培训以及与私营部门和学术界的合作伙伴关系可以提供便捷且灵活的学习机会。例如，Equitech Futures 的项目就是一个很好的例子。 公民技术学院 一个为期10周的项目，针对政府、非营利组织或社会企业的新兴领导者，旨在让他们“实践实现规模化影响所需的数据和科技技能”。通过将雇主的需求与愿意学习并提升自身技能的人士连接起来，此类努力有潜力支持良好的就业机会和 resilent劳动力市场。

人工智能的权利和尊重劳动的方法

AI 还为不断变化的工作和劳动力市场提出了新的问题。 AI 伙伴关系 ，例如，一个由人工智能生态系统各利益相关方组成的非盈利联盟进行了广泛的“数据丰富”经济研究——该经济由收集和标注用于AI系统的数据的人构成，并且已经开发了 指导方针 为了确保数据增强服务的负责任采购。这些准则旨在确保数据增强工作者获得公平报酬、公正的劳动实践以及更好的工作条件。国际劳工组织和联合国秘书长数字技术特使办公室也发布了一份报告，强调了构建人工智能技能和鼓励与工人进行社会对话的必要性，以确保技术进步尊重工人的权利并提高工作质量。

保护工人也意味着保护他们的劳动基本权利，包括结社自由和集体谈判权，这可以使工人及其代表能够就工作场所中人工智能的发展、使用和监控进行谈判并提供有意义的输入，尤其是在工作性质发生变化的情况下。

AI对劳动力市场潜在影响也强化了支持自动化进程中受影响工人的稳健系统的重要性。确保工人能够受益于由AI创造的新机会并受到其潜在危害的保护，将需要持续关注工人赋权，正如建议所示。 AI 和工人的原则 由美国劳工部提出。这可能意味着优先加强失业救济、医疗保健 accessibility、培训机会以及社会服务，以在过渡和失业期间支持工人并为其提供安全网。基于社区的支持系统、心理健康服务和职业咨询也能够帮助工人应对职业生涯转变的挑战并维持身心健康。 全球 AI 研究议程 进一步探讨这些问题，并为人工智能对劳动市场、人工智能价值链以及受人工智能变化影响的工人的影响提供研究和政策方向。

负责任的计算挑战 - 培训下一代技术人员

Mozilla基金会的责任计算挑战（Responsible Computing Challenge, RCC）强调了在低收入和中等收入国家（LMICs）投资责任人工智能技能的重要性。通过针对大学和教育机构，RCC旨在将责任人工智能实践纳入技术课程中，确保未来的技术人员具备构建道德人工智能系统的能力，并理解其工作的更广泛的社会、政治和环境影响。

RCC强调跨学科教育，促进与行业的连接和合作伙伴关系，并根据当地社区需求定制课程。这种方法旨在推动整个行业向负责任的人工智能转变。

例如，在肯尼亚基斯伊大学，本科人工智能课程重新设计以整合伦理原则，并通过行业实践者共同授课来促进从教育到就业的过渡。RCC还通过研讨会和活动触及传统计算机科学项目之外的学生，影响了成千上万名学生。从2018年到2023年，RCC已资助了美国、印度和肯尼亚的50所高等院校，并计划扩展至更多国家。RCC还整理了相关经验教训，编制了负责任计算挑战手册和课程库，旨在推动超越资助机构的变革。

通过提供资源和支持，RCC使新兴技术专家能够开发出针对其社区的AI解决方案，确保这些技术具有社会和文化相关性，最终促进可持续发展。

Karya - “创建高质量数据集 ， 同时为印度农村地区创造经济机会 ”

Karya 是一家专注于在人工智能经济中创造公平和可持续就业机会的印度社会企业。该组织旨在通过提供有尊严且公正报酬的工作来赋能社区，这些工作对于人工智能的发展至关重要，包括对图像进行分类、音频转录和数据标注等数字任务。

Karya 提供培训项目，旨在帮助工人掌握数字素养和专业技术技能，涵盖基本计算机技能、数据管理以及数据标注中使用的特定工具。这些项目的目的是提升工人的技术能力，并为他们未来在全球数字经济中的就业机会做好准备。

Karya 通过再投资为工人提供公平的补偿 利润转化为工资 通过此举，促进数字经济中被排除在外的人们实现财务稳定和经济赋权。

中央凯亚的方法核心在于直接与当地社区开展合作进行培训。通过与社区领袖和组织的合作，凯亚旨在从基层推动经济发展，确保其项目和就业机会能够支持经济的发展。这种以社区为中心的方法意在培养信任与合作，从而为工人创造一个支持性的环境，并进一步确保其项目具有文化相关性并满足其所服务社区的具体需求。

卡亚拉的举措特别旨在惠及女性，她们往往因根深蒂固的性别规范而面临获得有偿就业的障碍。通过提供灵活的家庭工作机会，在数据标注和无需大量基础设施发展的数字任务（在这种情况下使用现成的智能手机），卡亚拉支持女性在传统上被低估其工作的社区中获得公平薪酬。此外，卡亚拉强调数字素养和技能培训，以赋能女性，使她们具备在全球人工智能经济中取得成功的所需技术专长。

根据 Karya 截至目前，他们已在印度农村的27个州达到了40,000人，并为4000万任务分配了超过100万美元的工资。

图片来源 ： Riaz Jahanpour 为美国国

非洲内容主持人联盟 - 保护 AI 生态系统中工人权利的草根努力

数据标注员、调解员及其他完成人工智能设计、部署和使用所必需任务的工作者通常由科技公司外包。

The 非洲内容主持人联盟 并且其他基层组织长期致力于倡导安全的就业实践和环境。他们在改善劳动条件的集体谈判中处于前沿位置，并得到了Foxglove的支持，而Foxglove又得到了福特基金会、Luminate和开放社会基金会等机构的支持。此外，非洲内容审核员工会试图加强政府对有害劳动实践的回应，并支持针对他们认为创造了不安全工作环境的雇主提起的法律纠纷。

联合组织倡导公平支付合同工的劳动报酬，强调对这些工人提供更大支持有望整体上促进人工智能价值链的积极变革，并推动与这些科技公司互动的社区的经济社会发展。

赋能和支持基层和地方主导的努力有能力在整个AI价值链中推动变革。通过诸如加强政府保护、资金支持和技术援助（如法律支持的访问）等举措，基层努力可以有助于缓解与数据标注经济相关的许多劳动挑战，并促进社区层面的经济发展。

关于增强能力、促进人工智能相关技能和保护劳动力的建议

政府

- 将负责任且相关情境的AI教育整合到各级（从基础教育到大学教育，再到面向公众或职业导向的再培训和教育努力）的国家、区域和地方课程中。鼓励教育机构提供必要的资源，如教师培训和与相关政府部门的合作，以教授安全、安全和可信赖的AI，并特别关注缩小数字鸿沟。

美国的支持

美国支持一系列人工智能能力提升项目，包括AI Connect、负责任计算挑战、与未来增长倡议下的谷歌合作项目，以及公平人工智能挑战。

美国致力于扩大这些现有项目，开发专注于AI政策能力培养的新项目，并与私营企业及通过多边机构合作，以增加针对当地Context的全球AI能力建设努力。

- 这些措施可能会激励私营部门与教育机构和职业发展机构合作，开展人工智能教育和技能提升项目以及职业发展。这可以包括为积极贡献于构建安全、可靠和可信的人工智能能力的企业提供税收优惠、拨款和公共认可。

- 优先重视培训、失业救济、医疗保健 доступ以及社会服务，以支持受人工智能导致的劳动中断影响或易受影响行业的工人。

- 严格执行并监控遵守国内劳动法律和国际劳工标准，以防止工人被剥削并帮助确保人工智能价值链内外的体面工作。

- 特别努力确保代表性不足的群体——尤其是女性、女孩和性别多样的人——能够参与人工智能项目和教育，并旨在缩小STEM教育中的差距。

私营部门

- 开发人工智能培训和课程、认证项目以及资质认证计划，并与发展专家、包括教育和劳动力发展组织在内的民间社会组织以及当地社区进行有意义的合作，重点关注低收入和中等收入国家（LMICs）的需求以及这些国家的人工智能生态系统加强所需的技能。

- 在组织文化中优先确保AI系统的安全、可靠和可信赖的发展与实施，以便在全球范围内部署AI时能够满足当地需求并融入当地视角。

- 为人工智能价值链中的数据丰富工作者提供体面的工作，并支持和尊重工人集体谈判的权利。

- 支持和尊重工人集体谈判的权利，使他们能够在工作中就人工智能系统的设计、部署和使用以及工作场所的数据收集和使用进行协商。

图片来源 ： Jack Gordon for USAID

- 支持包容性招聘、晋升、职业发展和技能提升努力，以支持所有人群，尤其是女性、女孩和性别多样人群，以及STEM相关领域中常被排除在外的其他少数群体。

发展捐助者和慈善组织

- 向中低收入国家（LMICs）的大学和教育机构分配资金，以开发和实施负责任的人工智能课程，重点关注技术对社会的影响、偏见与公平性、透明度、网络安全以及与人工智能关键领域相关的专业知识（例如健康、教育、农业和粮食安全、能源、气候变化、交通以及公共服务交付）。这些课程应由了解当地背景的人设计，以确保其适用性。

- 投资专注于负责任人工智能的研究项目，并鼓励低收入和中等收入国家（LMICs）的机构与世界其他地区成熟的AI研究中心进行合作。这可以通过提供资助、联合研究项目、奖学金和国际会议来促进。

- 在人工智能背景下，支持倡导工人权益的工会和其他民间社会组织，提高对剥削性劳动实践的认识，并推广工作中的基本原则和权利。

构建可信和可持续的数字基础设施

安全、可靠和值得信赖的 AI 技术依赖于更广泛的可信的数字基础设施 bridging 数字鸿沟和解决这些数字基础设施挑战，包括人工智能对环境的影响，对于促进当地创新和经济发展至关重要，并可以使得人工智能技术得以包容性和可持续地开发和部署。

机遇与挑战 ：

许多低收入和中等收入国家（LMICs）面临有限的互联网连接和不稳定的电力供应。解决这些基础设施缺口，尤其是在人工智能（AI）具有显著能源成本的情况下，不仅支持技术进步和经济增长，还能够开发和利用针对可持续发展目标的针对性AI解决方案。

电力和连通性挑战

许多低收入国家的互联网普及率仍然有限。例如 ， 虽然 全球互联网使用情况 达到了63%的比例，而在非洲和南亚地区， respective比率分别为36%和43%。同时，需要解决若干连接性和数字鸿沟问题： 性别数字鸿沟 ，城乡之间 ， 以及 与残疾有关的分歧 ， 在许多其他人中 . 扩展连接性可以使个人、组织和企业能够访问并构建AI技术，并允许生成更多本地相关的数据以提高这些AI模型的质量。

频繁停电和不稳定 电力供应 在许多低收入和中等收入国家（LMICs），断电也是人们尝试构建和使用AI系统的一个障碍（此外还有其更广泛的负面影响）。断电中断了AI系统的稳定运行，妨碍了数据处理和模型训练，并 complicates 努力建立可靠且可持续的AI基础设施。

AI 的能源成本

AI与气候可持续性在多个方面相互作用。训练和部署AI技术，尤其是广泛应用于生成式AI的基础模型，需要大量的计算资源，这反过来又消耗自然资源。这些计算需求带来了电网的巨大电力需求，如果不能用清洁能源来满足，将会导致温室气体排放量增加。增加清洁能源的部署对于确保AI系统在长期内具有可持续性至关重要，并且至少在足够强大的模型能够以较低能耗进行训练之前，可以使得各规模组织能够构建和部署自定义AI系统。 .

在整个AI系统的生命周期中，需要能源来供电数据中心，并且需要水资源来进行冷却系统的工作。随着AI的发展加速，这种能源和水资源的消耗也同步增加。在某些情况下，数据中心的碳足迹可能比运行它们所消耗的电力更大。国际能源署估计，“到2026年，来自数据中心、人工智能（AI）和加密货币领域的电力消耗可能会翻倍。”由于全球迫切需要从化石燃料向可再生能源过渡，以将全球变暖限制在1.5摄氏度以内并避免最严重的气候变化影响，因此满足这一增加的能源需求可持续性至关重要。

因此，研究人员正在积极研究减少AI系统能耗的方法，以在计算量较少的情况下达到相同模型性能。例如，模型剪裁、量化和知识蒸馏等技术可能能够在不显著牺牲性能的情况下大幅减少许多AI应用所需的计算资源。

尽管应持续追求效率提升措施，部署AI系统的公司还应探索扩大清洁能源供应和使用以 powering 这些系统的选择。通过清洁能源为AI系统供电可避免因燃烧未加限制的化石燃料而产生的温室气体排放。支持清洁能源的扩展，并在相关情况下推动清洁能源基础设施的建设，可以避免将能源系统中化石能源的排放转移到其他地方。此外，利用AI系统庞大的能源需求可以促进清洁电力生成和存储技术的创新、商业化和成本降低。同样，根据可变可再生能源输出调整运营可以协助电网运营商整合更多的风能和太阳能。

尽管已经努力测量各种AI系统的能耗，但这些方法和指标尚未标准化，而且从设计和部署AI的企业获取能耗数据并不总是可行的。使这些数据公开，并过渡到使用清洁能源来供电计算基础设施，可以显著减少与AI相关的碳排放。类似的举措也需用于衡量并解决用于服务器冷却的数据服务器的水资源消耗问题。

另一种减少AI能耗的关键机制是为特定任务选择最合适的AI模型。虽然多用途生成模型具有高度的适应性并受到广泛关注，但它们往往比同等性能的任务专用模型计算密集度高几个数量级。这一现象不仅在模型开发和训练阶段观察到，在推理阶段也同样存在——即在生产运营中模型做出决策时。通过精心选择和设计模型，并不断用更高效的替代模型替换不高效的模型，从业者可以在减少碳排放的同时，使其系统更容易在计算资源有限的环境中部署。

微软的 Phi - 3 模型 - 在 AI 中进行创新 ， 同时降低能源需求

微软开发的Phi-3语言模型代表了更加节能的AI技术的进步。该模型旨在保持高性能的同时显著降低能耗，通过系统性地消除模型中的不必要的参数，从而减少所需的计算功率。这种方法不仅有助于减少与系统相关的碳足迹，还可能使模型在计算资源受限的环境中部署更为便捷，或在需要快速响应时间的情况下更具优势。

通过专注于减少AI模型的碳排放，Phi-3项目不仅旨在解决当前的环境问题，还为未来的AI研究和开发树立了先例。通过将能源效率融入AI技术的设计中，开发者可以为全球可持续发展努力做出贡献，同时也能增强AI应用在各种环境下的实用性。这种做法突显了在平衡技术创新与环保责任方面持续创新的重要性。

关于构建可信和可持续数字基础设施的建议

政府

- 创建促进政府、私营部门和民间社会之间合作的政策和制度框架，以提高可信赖的数字基础设施水平，并为连接性和能源投资创造有利的投资环境。
- 提供激励措施以降低投资风险，例如税收优惠、赠款、提前市场承诺、担保合同以及对公司在可信赖且可持续基础设施方面投资进行的公共认可。
- 与私营部门合作，为未覆盖地区提供负担得起的互联网接入，并支持当地能力提升以利用这些连接资源。

私营部门

- 定期发布有关人工智能系统从训练到部署整个过程中的能源使用和环境成本的数据，并投资于研发以减少这些系统的能源需求。努力改进并标准化测量和披露方法。
- 与政府、慈善组织和当地组织合作，投资扩大可信赖的宽带网络和能源基础设施。

美国的支持

美国支持一系列旨在直接投资或降低私营部门投资风险的基础设施项目，包括Power Africa、Digital Invest、亚洲开放无线接入网络（Open RAN）学院、与印度尼西亚亚马逊Web服务的合作，以及数字连接与网络安全合作伙伴关系。

国家标准与技术研究院（NIST）正召集专家和利益相关者社区，包括通过人工智能安全研究院联盟内的新工作组，研究人工智能对环境的影响测量方法，包括能源和水资源的使用。该项目将调研影响测量方法的现状，识别存在的差距，并可能建立一个专家小组以制定严格的共享测量标准。

图片来源 ： Riaz Jahanpour 为美国国际开发署

发展捐助者和慈善组织

- 支持推动AI能源消耗透明化的倡导倡议，鼓励公司披露其能源使用数据，从而促进行业更好地理解和承担责任。

- 继续在现有投资的基础上发展并分配资金以扩展可持续的宽带网络
  改善服务不足地区的互联网接入 ， 特别是在最后一英里和低收入国家。

- 建立监测和报告数字基础设施需求及障碍的机制，特别是那些有助于持续数字鸿沟的问题。利用这些信息支持民间社会的倡导努力，并将其与政府和行业分享，以供其参考。

扩大对数据存储和计算资源的访问

在许多地区，云访问和本地计算在获取上往往存在困难或负担不起。相关的计算资源，包括访问受信任的云提供商以及在适当情况下使用本地计算，可以促使低收入和中等收入国家（LMICs）开发并使用符合当地挑战的AI系统。访问相关计算资源的成本可能阻碍当地AI开发者进行实验、创新，并参与全球AI生态系统。缺乏数据中心和基础设施的访问可能会导致更高的延迟和运营成本。

为了开发符合当地需求的AI系统，来自低收入和中等收入国家（LMICs）的创新者需要获得用于训练和推理的计算资源。使计算资源更加普及和经济实惠可以降低进入AI创新领域的障碍，从而带来更多实用且相关的人工智能应用。

促进公私合作伙伴关系、区域合作以及创新融资模式（如共享基础设施项目或补贴云服务访问）的举措可以帮助克服这些障碍。此外，通过有针对性的培训计划和能力建设倡议来培养地方专业知识，可以使社区有能力管理和优化这些资源的使用，从而确保人工智能技术不仅可获得，还能有效利用以解决当地需求。

公平过渡计划 - 扩大计算渠道

公平过渡倡议（由Uwazi基金会、NVIDIA和HP联合发起）使访问高性能计算（HPC）以支持AI模型的开发成为可能。

公平过渡倡议的行动计划包括多项战略举措，旨在弥合这一差距。其中一个显著的例子是与巴巴多斯政府合作开展的国家计算基础设施项目，该项目旨在将该岛国转变为全球技术中心。通过建立最先进的数据中心，巴巴多斯可以为当地企业家、学生和研究人员提供必要的计算资源，以开发和部署人工智能技术。这些数据中心将能够处理和存储大量数据集，并作为创新试验场。此外，该倡议还致力于赋予肯尼亚青年使用计算资源的机会。通过此次合作，本地数据科学家获得了配备GPU的工作站，使他们能够将AI模型的微调和部署速度提高23倍。这一合作伙伴关系增强了当地的能力，并确保了AI解决方案符合低收入和中等收入国家的具体需求和背景。

这类专注于根植于当地需求的公私合营的努力，有能力改变技术与人工智能 доступ性的格局，使人工智能的益处可以直接应用于满足当地社区的需求和利益。在肯尼亚，年轻人能够微调、适应并部署一个地球科学和气候相关的人工智能模型，以尝试解决当地的挑战。未来在计算方面的努力应采取类似的方法：与当地利益相关者和社区组织进行彻底的需求评估，然后发展公私合营以实质性地解决这些背景下存在的计算差距。

关于扩大对数据存储和计算资源的访问的建议

政府

- 创建促进政府、私营部门和民间社会之间合作的政策和机构框架，以扩大对计算资源的访问，包括可信赖的云服务提供商。

私营部门

- 考虑针对本地需求、条件和使用模式的产品包装和分层定价模型。

发展捐助者和慈善组织

- 提供云计算访问的资助和资源，并降低投资高性能计算设施的风险，以服务低收入和中等收入国家（LMICs）。开发能力建设计划，使这些设施能够最好地支持当地研究人员、开发者和企业进行创新，构建安全、可靠和可信赖的AI系统以应对当地挑战。

美国的支持

美国国家科学基金会正在通过National AI Research Resource试点项目来扩大参与前沿AI研究的人群，在该项目中，政府机构与政府支持的和非政府组织合作伙伴合作，将研究人员和教育者连接到计算、数据和培训资源。此外，美国国际开发署（USAID）将与国际伙伴合作，扩大全球创新者和企业家获取计算资源的途径。

创建具有代表性的本地有用数据集和保存文化遗产

机遇与挑战：低收入和中等收入国家（LMICs）本地相关、多样化、多语言和多元文化的数据集相对稀缺。但当这些数据集是在与当地社区有意义的合作基础上创建时，它们可以促进更适合当地情境的AI模型的发展和应用，并推动公平发展的进程。

无代表性数据的后果...

数据质量和可用性对于AI的发展和应用至关重要，然而许多低收入和中等收入国家（LMICs）缺乏本地相关数据集，或者数据尚未数字化。

以下是你提供的英文句子的中文翻译，保持了原有的格式和其他符号，并且风格专业严谨：

To take one example, off-the-shelf healthcare AI tools may not accurately diagnose conditions prevalent in LMICs without sufficient local, diverse, and relevant data—a 现象 这与许多部门和环境有关。

机遇与挑战 ：

本地相关、多样、多语言和多元文化的数据集在低收入和中等收入国家（LMICs）相对稀缺。但是，当这些数据集是在与当地社区有意义的合作基础上创建时，它们可以促进更适合当地环境的AI模型的发展和使用，并推动公平的发展。

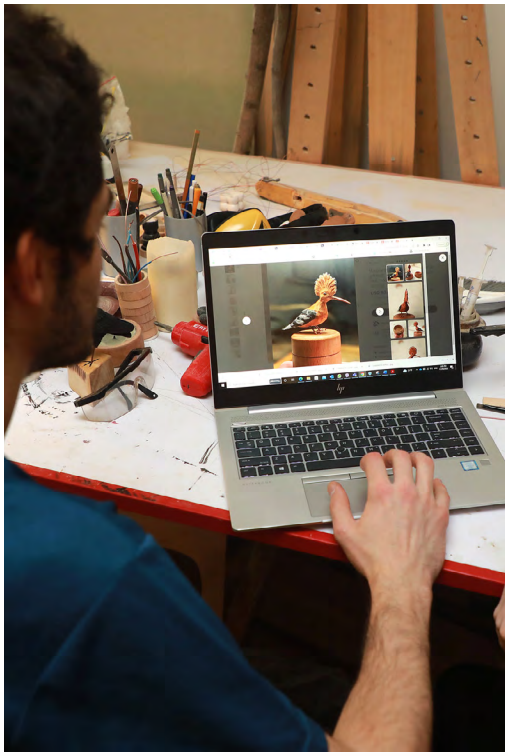
这些数据差距的潜在差异凸显了探索AI模型的代表性与偏见的重要性，以提升AI解决方案的有效性。在2023年，美国国际开发署（USAID）资助了一个项目，该项目由一组民间社会组织与墨西哥的一个州政府合作，对基于AI的技术进行了审查。 早期预警系统 . This 系统 旨在通过识别处于风险中的学生并为他们提供支持来提高学校留校率和毕业率。然而，系统中关键的性别偏见问题使其无法识别出超过4000名需要帮助才能继续上学的女孩。此次评估使系统得以更新，能够识别和减轻源数据集中的性别偏见，从而使其更有效地实现预期目标。重要的是，该项目早期发现了这一性别偏见，避免了进一步将AI工具制度化。通过本地数据集进行早期和定期的偏见检测对于解决这些迅速扩大的信息缺口至关重要，而该项目为此类工作提供了范例。
tool 为了支持这些努力。类似的偏见在其他背景下也有所体现；例如，评估简历的语言模型可能会对提到自己有残疾的人产生偏见。不仅需要检查数据集以避免特定群体的代表性不足，还需要考虑数据可能带来的其他影响方式。 反映不等式 在社会层面（例如，与性别、种族、 Ethnicity、种姓或能力相关）使得AI工具从中学习并强化这些不平等现象。

... 以及投资代表性数据的好处

投资更具代表性的数据集对于多个原因至关重要。首先，这有助于确保人工智能技术更好地应对低收入和中等收入国家（LMICs）面临的特定挑战，从而实现更有效且文化相关性的解决方案。通过捕捉这些地区内语言、方言和文化背景的多样性，人工智能模型可以在更广泛的场景下进行更准确和公平的训练。此外，代表性数据集还为全球人工智能生态系统提供了更完整和深入的理解，有助于避免在广泛使用的AI工具中延续偏见和不准确性。这不仅提高了全球AI应用的包容性和公正性，还使各国能够更全面地参与AI技术的设计、部署和使用，促进本国境内的创新和经济增长。

数据的代表性(或缺乏) 也在当今的基础模型中发挥作用 已评估 基准测试通常用于评估生成式AI模型在做出真实陈述、避免问题输出以及抵御破解等类似攻击方面的表现。不同语言的测试结果比较可以揭示这些模型的能力和局限性，但创建本地相关的基准则需要 当地相关 数据集。许多组织现在也开始为不同语言开发评估基准，以在各种地方背景下评估模型。

有一系列组织正在加紧加强和创建这些数据集。例如 ， Masakhane —a 旨在集中数据收集和研究努力于资源不足的非洲语言的协作性非洲研究人员社区—拥有多种项目以生成和使用高质量的文本和语音数据集。另一个例子是 Sustainbench ，一套包含十一项数据集的工具，用于推动和衡量七个可持续发展目标（SDGs）的进展，并促进更多学术和研究方面的投入。



图片来源 ： Beso Gulashvili 为美国国际开发署

在实现可持续发展目标（SDGs）的斗争中，包括通过提供“多种SDGs任务上的机器学习模型评估标准基准”，来评估机器学习模型在各类SDGs任务上的表现。除了民间社会主导的这些努力之外，政府和私营部门也扮演着至关重要的角色。 政府 并且通过公私合营在创建开放数据系统方面进行合作，这些系统可供创新者和大学用于构建本地相关的AI模型。

支持土著社区保存和促进其文化

在此背景下，AI在支持和推动当地社区发展方面的作用尤为关键，尤其是对于世界各地的原住民社区。

例如，AI可以支持原住民语言的保护与复兴， enabling 原住民语言之间的翻译服务，并通过支持原住民社区控制自己的数据以及根据原住民社区成员的决定利用AI为不同的目的服务来赋能原住民社区。

公共-private合作伙伴关系以及使用奖金等创新融资举措也开启了利用人工智能技术进行文化保护的新途径。例如，OpenAI、冰岛政府以及当地的冰岛科技公司 partnered 为了改善OpenAI的GPT-4模型在冰岛语中的性能， Incorporating 文化上敏感的本地社区和个体反馈。类似地，Anthropic与新加坡的信息通信媒体发展局（IMDA）和AI Verify基金会合作，通过与当地人群互动来开展能力提升工作。 测试(或 “红色团队 ”) AI 系统 为了提升语言能力并针对与其社区相关的话题。此次红队评估活动使用了英语、泰米尔语、普通话和马来语（新加坡的官方语言），重点关注与这些人口群体相关的话题。

拉库纳基金 — “将机器学习的优势带给全球的数据科学家、研究人员和社会创业者”

成立于 2020 年，Lacuna 基金 填补了限制针对低收入和中等收入国家（LMICs）需求定制的人工智能解决方案发展的数据空白。通过为数据集的创建和维护提供财务支持和技术援助，这一由资助者和数据科学家组成的联盟为当地相关的人工智能设计和使用奠定了基础。目前，该基金由梅里迪安研究所管理，并得到了洛克菲勒基金会、加拿大国际发展研究委员会（IDRC）、Google.org、德国联邦经济合作与发展部（GIZ）、Wellcome、戈登与贝蒂·摩尔基金会、帕特里克·J·麦格芬基金会和罗伯特·伍德·约翰逊基金会的资源支持。

拉库纳基金会在健康、农业、欠资源语言和气候等领域支持了数据集的创建。例如，在农业领域，拉库纳基金會支持了小农户农田的数据集标签，这些数据集包括“地理参考作物图像以及关于投入使用、作物管理、物候学、作物损害和产量的标签，这些数据来自肯尼亚8个县。”通过聚焦本地相关数据，人工智能工具将更好地适应该地区农民所面临的特定挑战，从而促进更有效的农业实践。

秉承Lacuna基金的使命，每个数据集都是在当地开发和拥有，并向国际社区开放。Lacuna基金还强调负责任的数据管理实践和社区参与的重要性。由Lacuna基金资助的项目必须遵守严格的规定，确保数据的收集和使用尊重个人的权利、自主性和隐私。此外，该基金促进当地研究人员和社区参与数据采集工作，以确保数据集在文化和上下文方面是合适的。

lacuna基金会在“可访问性、公平性、伦理观、参与性方法、质量以及变革性影响”方面的原则指导着其项目选择，涵盖了超过30个数据集。截至目前，这些数据集已被下载超过1,600,000次。


图片来源 ： Emily Mahoney

通用语音 - 收集包容性语音数据

语音识别是一个迅速增长的市场，而大型语言模型（LLMs）和生成式AI的发展进一步加速了这一增长。尽管世界上有超过7,000种语言，但只有少数几种被语音助手所识别。专有的数据集创建、购买、维护和适应成本高昂，通常专注于资源丰富的语言如英语。此外，现有的语音训练数据集往往嵌入与文化、种族、性别和社会经济地位相关的偏见，倾向于教育程度较高、居住在城市地区的用户。

Mozilla 普通语音 (MCV) 旨在通过提供开源的高质量语音语料库来解决这些问题，该语料库适用于全球多种语言。MCV是最大的、最多样化的公共参与多语言语音语料库，专注于包容性、社区驱动的努力以及权利意识的发展。目前，MCV拥有超过 30, 000 小时记录 ， 120 + 语言 ，以及约250万次下载量。MCV还采取了性别响应的方法，以确保数据集的性别代表性，措施包括制定性别行动计划并扩展平台上的性别选项。

在2024年，MCV将其核心的剧本平台扩展到包括特定领域的语料库，并通过提供语言变体和口音的端到端支持来增加语言多样性。这一扩展使社区成员能够贡献被排除在外的变体和口音，从而确保边缘化声音得到更充分的代表性。

MCV还推出了自发演讲新平台的beta版本，不仅涵盖了朗读演讲，还捕捉了更加自然、有机的演讲模式，包括代码切换和根植于口说为主的语言文化。MCV计划在未来两年内实现10万小时的众包多样化语音，并继续探索在大模型和AI时代支持边缘化语言社区的方式。

Mozilla社区基金（MCV）与GIZ、比尔及梅琳达·盖茨基金会以及FCDO合作，投资于构建三种东非语言的数据集和应用场景——卢旺达语基尼亚卢瓦语、乌干达语鲁加纳语以及以刚果民主共和国、肯尼亚和坦桑尼亚为重点的斯瓦希里语。这项工作基于Mozilla的资助和拨款方法，旨在培养领导力和社区。该工作由包括Digital Umuganda和Makerere AI实验室在内的研究员和合作伙伴主导。这些数据集是在COVID-19疫情期间建立的，并依赖于建立社区信任来确保成功。通过这一投资，已经开发出解决方案，以当地语言提供有关COVID-19的信息，以及关于土地权利、金融服务和农业服务的信息。

Te Hiku Media - 毛利人数据保存

Te Hiku Media 是一家位于新西兰的毛利族所有和运营的非营利媒体组织。该组织专注于保护和推广毛利语言（te reo Māori）和文化。2018年，他们推出了 竞争
在新西兰，毛利语说者记录了超过300小时标注的音频。基于这些数据，他们构建了用于自动语音识别和毛利语（te reo Māori）发音的AI工具。 如所述 泰胡媒体（Te Hiku Media）首席技术官凯奥尼·马赫洛纳表示：“必须意识到数据访问量巨大……全球对毛利语（te reo Māori）的兴趣日益增加，毛利人需要按照自己的条件和要求来引领这一趋势。”

泰胡克媒体开发了多种针对毛利人的AI技术与工具（例如，自动语音识别、自动标签标注、实时语言反馈），并通过其数字语言平台提供了许多工具的API接口。 爸爸 Reo ．他们还建立了新的数据保护协议，并与毛利人领导的组织合作，开发支持社区的AI工具。同时，他们与其他土著社区合作，利用Te Hiku语言数据库训练其他土著群体的语言模型——例如，在库克群岛，语言模型覆盖了大约
70% 精度 只有 10 个小时的数据。

基于这一由社区主导的解决方案的强大基础，公私合作伙伴关系也发挥了作用：Te Hiku Media 与相关方进行了合作。
NVIDIA 为了获得更可负担的计算能力，使他们能够构建和训练这些AI系统。这表明了采取负责任的AI方法和当地社区与技术公司之间有意义的合作如何有助于缩小数字鸿沟。

第一语言 AI 现实 - 振兴土著语言

第一语言 AI 现实(FLAIR) ，来自 魁北克人工智能研究所(Mila) 应对语言灭绝的紧迫威胁，特别是迅速消失的原住民语言。

FLAIR 's 使命是通过提供工具和自主权，赋能原住民社区，使他们能够维持与这些语言同步发展起来的语言和文化。这包括开发针对原住民语言的语音识别系统，可用于语言学习、音频转录和语音控制技术。FLAIR的目标是即使对于数据有限且剩余使用者稀少的语言，也能迅速创建定制模型——这适用于世界上许多语言的特点。FLAIR工作的核心在于，在设计人工智能系统时考虑到北美原住民语言的独特细微之处，与欧洲语言相比，并创建由社区主导且基于社区同意的数据收集方法，以实现包容性。

FLAIR 强调开源解决方案和社区参与的重要性。所有通过 FLAIR 开发的工具和模型均公开共享，确保其他社区能够从中受益并利用这些进步。这种开放的方法旨在赋能全球原住民社区，促进数千种资源不足的语言的保存与复兴。通过专注于灵活的模型开发方法而非仅为单一语言量身定制的模型，FLAIR 有助于快速扩展，从而确保该技术能够在各种情境中部署，进而减轻语言灭绝的风险。

关于创建具有代表性的本地有用数据集和保存文化遗产的建议

政府

- 通过以下方式鼓励和支持地方数据收集工作提供资金、资源和政策支持。合作伙伴与教育机构、当地领导人和民间社会组织聚集和策划全面具有有意义的知情同意的数据集。

- 制定并应用确保数据安全处理的政策。这些政策应促进数据隐私、保护个人和社区，并推动允许可信方之间无缝数据共享的数据治理方法。

美国的支持

美国开放政府数据门户提供了数十万份数据集的访问途径，并支持使用AI保护藏族文化的一项计划，同时正在乌克兰创建一个基于AI的系统，用于记录文化遗产和国家基础设施的损坏情况。

美国还计划采取更多措施来加强与合作伙伴的数据质量和可用性，包括投资创建更具包容性的AI系统基准，并确保现成的AI工具尽可能对全球各地的社区有用。

- 倡导保护原住民和历史上被边缘化群体的语言和文化，并确保这些群体能够自主管理其数据的收集、所有权和应用。

私营部门

- 持续在部署前后评估AI系统，以识别和减轻文化、区域和语言偏见以及歧视性、有害的输出。
- 与政府、非营利组织和当地技术公司合作，开发支持文化保护的AI工具。与当地利益相关者合作，以文化敏感且有利于保护当地语言的方式改进AI模型。
- 支持并提供技术支援以促进社区主导的项目，确保人工智能工具的发展和使用与社区需求相一致。这可以包括提供资源、培训和开发人工智能技术的平台接入。

发展捐助者和慈善组织

- 分配资金用于专注于收集和整理高质量、当地相关、领域特定的数据集、资源不足的语言数据集以及基础模型基准（所有工作均需获得有效的知情同意并采取适当的隐私保护措施）的项目，在低收入和中等收入国家（LMICs）。确保资金投入包括维持数据的相关性、更新和可用性的资金。同时，支持当地社区进行数据收集，以确保数据收集的文化相关性和知情同意的有效性。
- 投资于发展和实施数据治理框架，以促进跨境数据的自由流通，并确保信任、安全和保障。同时，努力确保数据生产者能从这些数据流中获益。提供资源用于培训和能力建设，以提升数据隐私与安全实践。倡导负责任的数据行为，并确保地方声音及民间社会在数据治理讨论中的参与。

- 支持土著群体、边缘化人群以及文化遗产机构召开会议并讨论其与人工智能技术相关的目标和愿望，随后资助他们的提案。这可能包括开发和整理数据集及人工智能技术，并支持关于如何使用和分发这些数据集的选择。

- 使土著社区和边缘化群体有机会与政策制定者或AI开发者合作，以积极的方式参与AI政策制定和/or AI开发，从而突出强调其赋权和自主性。

制定战略以在实践中实现 AI 的承诺

持续改进和评估

人工智能有可能改变多个部门， 推动实现可持续发展目标的进展 然而 ，扩展 AI 用例需要确凿的证据来衡量有效性。 系统评估 如何人工智能系统有助于应对发展挑战，可以增强对应投资或终止的干预措施的信心。虽然对全球发展项目进行严格的评估需求并非特定于人工智能，但从这些系统中仍可以学到很多。 发展史评价 并且涉及实施科学。评估项目影响往往意味着将评价的“终点线”进一步向下游推移。一个错误率低的AI模型可能不够充分；测量实际生活中的结果和后果，或者最接近这些测量的代理指标，可能会更具意义。

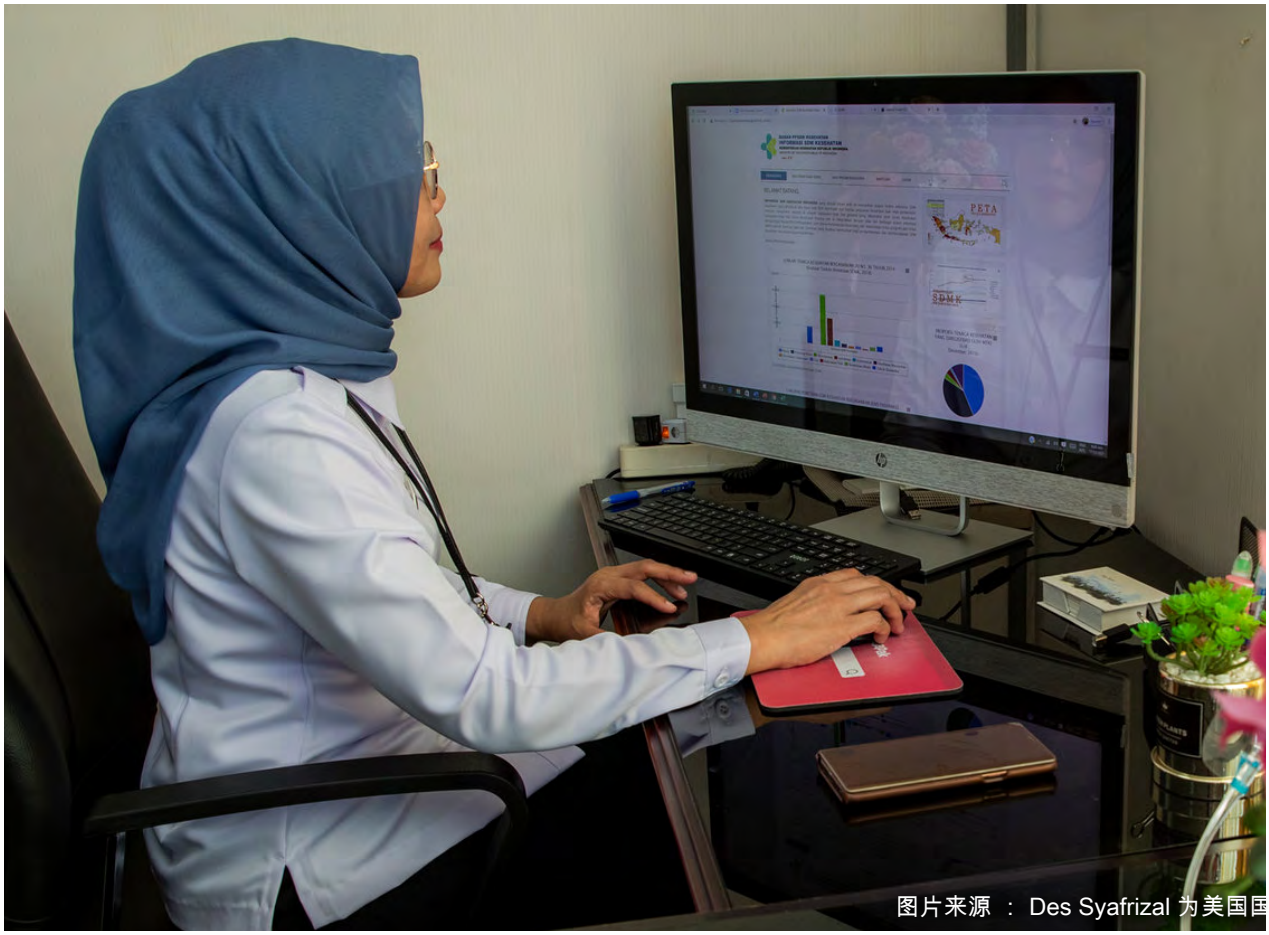
例如，由人工智能驱动的诊断工具可能在实验室环境中提高诊断准确性，但其在实际应用中的表现应受到监控，以确保它们能够正确使用并整合到医疗工作流程中，真正改善患者结果、节省医疗专业人员的时间或降低医疗成本。当这类人工智能技术被部署时，应跟踪诸如减少诊断错误、改善健康生物标志物、增加获取医疗信息和服务的机会等指标，以验证特定的人工智能系统是否提供了边际价值。

机遇与挑战 ：

许多AI项目未能达到规模并产生实质性影响，原因在于要么未能 robustly 评估AI系统在当地环境下的有效性，要么缺乏持续的投资。基于坚实证据的AI干预措施并扩大成功方法的规模对于解决全球发展挑战和最大化AI的社会影响至关重要。

在教育领域，人工智能有潜力个性化学习体验并提高学生参与度。但应通过衡量其对学生成绩和高质量教育资源获取的实际影响来评估AI驱动干预措施的有效性，以便这些措施能够在整个教育领域推广。辍学率、毕业率以及在欠发达地区教育成果的改进都是可以综合使用的指标，用以评估引入和使用AI系统的影响。

同样的原理适用于其他AI应用领域，如优化农业产量或预测自然灾害并优化救援行动。在这些每个领域中，开发人员和部署者必须问自己：这个AI系统相对于相关替代方案提供了多少边际价值？是否存在可以增强其效果或减轻负面影响的配套政策或工具？通过关注透明且基于证据的结果，我们可以帮助确保AI干预措施在个人或项目层面是有益的，并值得进一步投资。



图片来源 ： Des Syafrizal 为美国国

AI 应用程序的可持续扩展

严格评估本身不足以使人工智能在实践中兑现其承诺。为了扩大那些证明有效的干预措施的应用范围，跨公司和组织的持续投资与合作至关重要。通过促进和支持这些投资，我们可以帮助确保证明有效的AI解决方案不仅能够维持其影响力，还能扩大其应用范围。加大对各国和地区本地初创企业的投资将鼓励针对当地问题的可持续解决方案——在这方面不存在短缺。 有前途的研究实验室和初创公司 在 LMICs 中。

解决资本障碍至关重要。风险投资和影响力投资基金可以重新设计以更好地满足当地企业家的需求，包括提供较小的投资规模、更长的投资周期以及更大的灵活性，以适应这些市场所面临的独特挑战。协作模式，如创新 hub 和加速器，这些模式提供导师指导、 networking 机会以及对全球市场的曝光，也可以在规模化成功的 AI 解决方案中发挥关键作用。

最后，促进本地创新者与全球科技公司的合作伙伴关系可以加速人工智能解决方案的设计、部署和使用。这些合作关系可以提供访问全球专业知识、先进工具和平台的机会，这些工具和平台在没有这种合作的情况下可能难以获得。此外，这样的合作还可以帮助弥合本地需求与全球技术趋势之间的差距，确保人工智能解决方案不仅具有可扩展性，而且在本地背景下也相关且具有影响力。

TRACE - TB 项目 - 在抗击结核病中扩展 AI 解决方案

The TRACE - TB 项目 由USAID资助并由Wadhwani AI实施的该项目是精心评估AI系统可以推动大规模变革的一个典型案例。结核病（TB）仍然是全球性的重要健康挑战，特别是在像印度这样的国家。该项目的目标是将AI解决方案集成到TB护理生命周期中，以提高预防、检测和治疗效果，最终旨在消除印度的TB。

The TRACE-TB 项目包括一个用于分析咳嗽声音以预测结核病（TB）可能性的筛查工具、一个计算机视觉程序用于解释耐药性结核病实验室结果、以及一个评估患者是否完成 TB 治疗风险的预测工具。这些工具专门针对结核病护理中特定的挑战进行设计，如早期检测和治疗方案的依从性。例如，咳嗽声音筛查工具利用经过大量音频记录数据训练的深度学习神经网络，能够在医疗机构和家庭环境中实现快速且非侵入性的 TB 筛查。

该项目的评估过程强调了在扩大规模之前衡量AI解决方案有效性的重要性。 TRACE - TB 已经通过其咳嗽检测工具筛查了超过100,000人，识别出超过15,000例疑似结核病例，从而将患者的诊断率提高了12%。此外，该工具还帮助识别出超过26,000名需要更 intensive 照顾的患者，导致负面结果（如治疗中断、永久性肺损伤和死亡）减少了16%。目前，美国国际开发署（USAID）和Wadhwani AI正与印度政府合作，将该项目在全国范围内推广。

总体而言，TRACE-TB项目展示了如何负责任地评估和扩展AI系统。通过关注有效性并整合现有的医疗保健基础设施，该项目展示了利用AI应对全球复杂健康挑战的途径。

制定战略以在实践中实现 AI 承诺的建议

政府

- 促进并 facilitation 最佳实践和经验教训的共享平台的创建，以开发和评估人工智能干预措施。
- 与独立评估机构、学术机构或审计机构合作，使用与实际结果挂钩的指标评估AI干预措施的影响。与外部专家共享数据和见解，以提高影响评估的可信度、透明度和有效性。扩大有前景的干预措施，并将资源从无效项目中分配出去。
- 投资政府员工的技能提升和持续学习，以确保员工具备知识和技能，能够就AI技术作出知情决策。

美国的支持

美国支持
评估 AI 现实的项目范围 -
持续的世界影响和支持
对干预措施的投资表明
承诺 ： 评估 AI 中的性别偏见 -
基于财富估计 ， 包容性金融 ，
基于 AI 的读写能力的测试和扩展
方案 ， 改善孕产妇和新生儿
结果 ， 并使用 AI 减少艾滋病毒
治疗中断等。

今年，USAID 将设立负责任人工智能基金，该基金将支持并严格评估强化全球人工智能生态系统的方法。这还将补充现有努力，对人工智能进行评估。 医疗保健设置 ，in 教育计划 和其他部门。

私营部门

- 创建或贡献专门的AI影响基金，旨在支持具有 proven socioeconomic benefits 的AI技术初创企业和组织，特别是来自低收入和中等收入国家（LMICs）。

- 优先投资和合作伙伴关系，重点支持已证明可推动可持续发展目标的AI解决方案。

发展捐助者和慈善组织

- 与持久且持续的项目资金配套提供评估资金和技术支持，以衡量人工智能项目的影响力。部分基于项目部署、扩展和评估计划的质量来评估项目。支持建立标准化指标和方法论，以评估人工智能对发展干预措施的贡献。确保使用的指标反映当地社区的需求和优先事项。

制定良好的治理框架 ， 以开发和使用安全和尊重权利的人工智能

人工智能有可能被滥用，从而危害人们和社会，削弱人权并破坏民主规范。人工智能系统的发展加速了这些风险，包括国家和非国家行为者利用人工智能进一步开展虚假信息campaigns、恶意网络活动、深度合成内容（deepfakes）、不公的结果、侵犯知识产权以及非法或任意的监视行为。 预防和缓解 这些滥用情况——尽管促进了负责任的人工智能——应成为促进全球发展和提供人道主义援助工作的首要优先事项。确保人工智能是负责任的包括保护隐私权这一核心需求。采取这种做法需要在人工智能设计、部署和使用的所有层面嵌入人权、隐私和安全原则，如合法性、公平性、数据最小化和目的限制、完整性和保密性。

治理是构建安全且尊重人权的AI的关键组成部分，政府和政策制定者肩负着重要的责任，以有效开发并实施平衡而有效的框架来利用AI带来的好处并管理其风险。这些措施可以培养公民的信任；这种信任可以促进采用，而采用又可以推动创新。在治理的这一背景下，行业特异性以及与现有监管机制的整合同样至关重要。由于AI是一种通用技术，其带来的利益和风险取决于该技术开发和使用的具体背景。因此，许多政府都有能力利用现有的监管机构针对特定行业进行AI治理，从而补充现有行业专业知识和能力。在AI系统的背景下，尤其是在合成内容（包括虚假信息和误导性信息以及隐私问题）等新且跨领域的挑战面前，政策制定者仍在收集信息并咨询相关方，以便更好地定义和应对这些问题。这些领域也是自愿指导可以为私营部门和其他参与者提供安全保障的区域，在技术专家和政策制定者逐步了解AI技术的过程中，过早地实施监管是不适当的。

降低虚假和误导性信息传播的风险

AI可以用于创建虚假或误导性信息，而算法内容推荐往往通过利用受知识产权保护的内容来放大不实信息的传播，这在国际范围内尤其成为一个重大问题，增加了对民主和人权侵犯与滥用的可能性。在存在预存的政治和社会紧张局势、脆弱的制度结构或有限的媒体和数字素养的背景下，由AI驱动的不实信息尤为令人担忧。生成模型可能进一步降低大规模推广有说服力的不实信息活动的门槛，使资金和资源有限的行动者能够在线上拥有相当大的影响力。最终，这可能会利用AI来压倒和歪曲民主参与，例如通过过度定向或选民压制运动，这对民主价值观的维护产生了深远的影响。正如大型语言模型所揭示的那样，这一趋势正在全球范围内迅速发展。

机遇与挑战 ：

人工智能有可能以损害社会的方式被滥用，包括通过信息操纵活动、政治操控、深度假象、非法或任意的监控以及侵犯人权。然而，通过支持对抗这些侵害行为的利益相关方、推动负责任的人工智能实践，并促进国际合作，我们可以通过利用人工智能产生积极的社会影响，并保护民主价值观和国际人权来实现这一目标。

模型可以生成令人信服的文本，图像生成器可以生成逼真但虚构的音频、图像或视频——有时称为“深度假象”。在摩尔多瓦， 摩尔多瓦总统 Maia Sandu 的 deepfake 视频 推广亲俄候选人一事在Facebook上迅速传播，被超过一百万用户看到。在另一次事件中，在斯洛伐克总统选举前的几天里， AI 生成的音频 据报道，其中一名候选人描述了如何操纵选举的方法。这些工具在资源较少的背景下尤其令人担忧，例如地方选举或技术平台用户数量较少的国家，因为这些地方的内容审核较少且独立的事实核查速度较慢。

一种最普遍且最具危害性的生成AI工具的应用是创建女性和女童的假象及淫秽图像——这是一种技术辅助下的基于性别的人身暴力形式。 女性占 99% 针对深伪色情中目标个体。此外，深伪技术也被用于政治和公共生活中针对女性和女孩的目标，常常利用性别歧视或厌女症的刻板印象来制造 性别虚假信息 旨在遏制言论自由的行使。随着深度伪造内容创建变得越来越容易，恶意行为者操纵公众舆论、剥削和伤害个人，以及破坏民主进程的可能性正在同步增长。在日益全球化的时代背景下，人工智能驱动的内容影响远远超越了任何单一国家的边界，威胁着..... 阻碍进步 在关键问题上取得进展，如民主韧性、公共卫生、经济发展、性别平等、隐私保护、知识产权保护以及社会凝聚力。

在此潜在由合成内容引发的危害背景下，值得注意的是知识产权问题。与人工智能相关的知识产权问题复杂且迅速演变，涉及法律、政策、伦理和技术等多方面的考量。因此，联合国世界知识产权组织已草拟了相关文件。 原则 为了应对一些问题并强调在设计、部署和使用AI技术时尊重知识产权的重要性。

减轻人工智能驱动的虚假信息和深度合成的风险意味着进一步投资于技术保障措施，如来源验证、数据追踪和水印等，以及一系列非技术措施。能够为用户提供标签的技术方法必须与帮助用户理解并适当应对这些标签的努力相结合。更广泛地说，在机构层面（如AI取证、数据隐私保护工具、调查 journalism、独立媒体和事实核查组织）增强本地能力，尤其是具有区域重点的组织，至关重要。与技术保障措施和标签的发展与部署进步相配合，媒体和数字素养培训应教导公民如何识别由人工智能操纵的媒体的特征，并理解技术保障所能和不能做到的事情，以及现有社会偏见和分歧如何被利用。

解决隐私和监控问题

随着人工智能虚假信息的增加 ， 人工智能监控系统 正在全球部署 ，经常是非法且任意地进行 ，且政府和私营部门存在重大投资和滥用问题。这些系统利用AI分析来自各种来源的数据，包括监控摄像头和社会媒体数据。当前的局势尤为令人担忧的是，这一趋势呈现出日益加剧的趋势。 模糊线 在公共健康领域实现广泛社会效益的监控（例如疾病追踪）与非法或任意使用人工智能进行监控之间 目标 政治活动家、工人和工会成员、记者、少数群体或其他人。此外，还存在一个重要的问题，即意外危害，如此类人工智能系统对已经处于脆弱状态或边缘化的社区产生的不成比例的影响，这往往源于训练数据中的偏见。

图片来源 ： Amir Mohebali

根据一些观察者的意见，在许多国家，政府和私营部门采用人工智能技术进行不道德、非法和任意的监视。 成长 ，而且这些技术的使用往往超前于国内相关法律框架的发展和应用，以保护工会、民间社会和个人的权利。这种不平衡可能使得非法或任意监视变得更加容易。人工智能在监视中的使用可能会不成比例地影响边缘化社区，原因可能是明确的目标选择或由于数据集中的隐性偏见而引起的无意间的影响，从而加剧现有的社会不平等，并引发恐惧和不信任。因此，人工智能监视技术仅应在能够负责任地使用且具备最严格保障措施的情况下部署。

随着隐私侵犯和监控担忧的加剧，出现了严重限制言论自由的重要问题，即具有高度针对性和侵略性的内容审查。此类由人工智能驱动的工具增强了专制国家及其非国家行为者监控、控制和抑制信息的能力，这减少了真正自由开放交流、异议和分歧的机会，从而可能限制获取信息和做出明智决策的可能性。

AI 与国家安全

人工智能带来的挑战不仅限于上述内容——必须采取有意识的努力来确保在设计、部署和使用人工智能以及其未预见的后果方面进行全面考虑，以解决问题并减轻负面影响。在国家安全背景下，人工智能系统的实施，特别是在自主武器和伤亡评估等关键领域，可能会引发潜在的伦理或法律问题。为了保护国际人道法和国际人权，类似美国国务院发布的《负责任使用人工智能与自主性的政治宣言》等统一努力已获得超过50个国家的支持，变得越来越必要。这些努力旨在保护平民和民用物体免受武装冲突的影响，减少未预见的偏见，采用和保险关键的安全功能，并确保相关国防人员的透明度和监督。人权评估也可以帮助识别和解决风险。需要注意的是，这里概述的挑战并不全面，在不断变化的人工智能设计、部署和使用环境中，必须持续进行全面的努力，以促进良好的治理和尊重人权的使用。

图片来源 ： Angela Rucker

非洲代码 - 在 AI 时代保护信息生态系统

非洲代码 (CfA) 是一个市民科技倡议项目，利用数据和科技促进透明度、打击虚假信息并加强非洲的数字民主。

CFA 发起了各种举措 ， 以打击虚假信息并促进准确的信息。例如 ， 其 PesaCheck 项目 是非洲最大的本土事实核查项目。PesaCheck验证多种语言下的新闻和社交媒体内容的准确性，帮助遏制虚假信息的传播。通过系统性的过程，包括识别声明、查找数据、验证数据以及发布调查结果，PesaCheck能够高效地识别并驳斥基于多编辑审核的误导性信息。

作为另一个例子 ， CfA 的 WanaData 网络 是非洲大陆的一项倡议，旨在支持女性数据科学家和记者撰写数据驱动的故事。该网络通过利用数据来揭露和报道腐败、治理和社会正义等方面的问题，从而增强透明度和责任感。通过为本地记者提供分析和报道数据的工具与技能，CfA 促进了一个更加知情和积极参与的公民社会。这种相互协作的技能提升和对本地专家有意义的支持方式，证明了在应对诸如打击虚假信息等社区需求方面具有实用性。

最后 ， CfA 还致力于通过以下项目提高数字素养和公民参与 非洲调查报告中心网络(ANCIR) ANCIR为调查记者提供培训、资源和协作平台，帮助他们掌握有效使用数字工具的技能。这一举措有助于记者产出高质量的调查作品，从而监督权力并告知公众。

尤其是通过许多CfA的项目可以明显看出其在合作方面的有意义的方法——将社区领导者、领域专家、新兴领导者、民间社会组织以及私营部门聚集在一起，倡导非洲的共同目标和愿景，并促进数字民主。

关于开发安全且尊重权利的人工智能的治理框架的发展与使用，建议

政府

- 建立健全涵盖政府各部门的治理和监督框架，专门针对AI技术的滥用问题，包括虚假信息、深度合成、未经同意的亲密图像、网络攻击、言论审查、公共安全、个人伤害以及国家及非国家行为者的非法和任意监视滥用。这些法律和框架应保护隐私权、知识产权，保障民主机构的安全，并维护人权。此外，这些法律和框架还应纳入性别视角，以反映对女性和女孩，以及LGBTQI+ 政治和公众人物进行基于深度合成图像性虐待的独特且不成比例的攻击。

- 发展、实施并推进国际标准和兼容性框架，以促进人工智能设计、部署和使用相关的国际合作、尊重权利的采购、伙伴关系和协调。

美国的支持

美国正在支持一系列旨在促进尊重人权的人工智能的政策措施、资助项目和工具。这包括《人工智能权利法案蓝图》、《人工智能和人权风险管理概况》、国家电信和信息管理局发布的《人工智能问责报告》、一份专注于减少合成内容风险的报告以及一个关于构建负责任、包容性和尊重人权的人工智能方法能力的项目。

在今后几年里，USAID的“推进数字民主”（ADD）倡议将支持数十个国家建立AI生态系统，防范AI的滥用和潜在危害。ADD将支持当地各方加强尊重人权的AI治理法律框架，并促进对初创企业、科技孵化器及其他推动负责任、可信赖且尊重人权的AI系统的投资。

- 支持资源受限环境下外国投资者和开发者，为其提供技术援助以导航当地与AI相关的法规，例如通过指定的AI官员提供实地联系和支持，促进负责任的AI创新。

- 确保政府采购、合同签订和人工智能技术使用能够促进增强隐私权、福祉和人权的努力，并内置技术性和非技术性保障措施。

私营部门

- 支持政府官员、政策制定者、工人及其组织与其他相关利益方之间的信息交流，就AI在低收入和中等收入国家（LMICs）的设计与部署所引发的关切进行发声，并将此指导因素纳入产品开发过程中。

- 投资于增强隐私和透明度的技术及其他能最大限度减少潜在滥用风险的技术解决方案。

- 设计AI工具以使尊重权利的使用更加容易，而不尊重权利的使用更加困难。在发布AI工具之前进行人权尽职调查，并内置相应的防护措施，如NIST AI RMF生成式人工智能配置文件中提到的针对已知危害和滥用（例如AI生成的非自愿亲密影像及儿童色情材料）的安全保障。

- 充分验证客户以防止此类技术被恶意使用，包括非民主政府和恶意非国家行为体的使用，尤其是如果此类技术具有侵犯公民自由、人权和个人隐私的能力。负责任地采集源数据，包括使用知情同意的方式。

发展捐助者和慈善组织

- 支持研发工具以检测和缓解AI滥用，并采取措施防止滥用。投资促进透明度、问责制以及负责任使用AI技术的倡议。
- 投资当地新闻媒体、事实核查组织及监督机制，特别是在选举等脆弱背景下，以识别和应对深度假信息及其他形式的虚假信息。
- 支持内容来源和认证工具的发展与使用，并开展公共意识宣传活动以推广媒体素养和虚假信息检测技能，同时清晰传达这些技术的局限性。
- 支持民间社会组织对抗虚假信息对女性及LGBTQI+政治和公共人物的性别化影响，特别是在选举背景下。
- 倡导保护隐私和促进在人工智能技术使用中尊重人权的政策。开发和支持现有的监控人工智能治理框架实施的机制，并支持民间社会的倡导努力以提高这些框架的遵守程度。

通过开放性、透明度和可解释性培养对人工智能的信任

透明度 (或缺乏透明度) 与以下方面相关 多层次 在用于训练AI模型的数据中、在AI模型内部、在AI的实施方式中、在组织层面以及最终在政策制定层面。确保透明性对于保障AI的安全、安全和可信至关重要。

机遇与挑战 ：

缺乏透明度的AI可能会对信任和问责制造成障碍。通过促进AI生态系统及技术本身的开放性，我们可以培养更大的信任、包容性的创新以及公众在AI政策和治理中的参与。

可解释的 AI 系统

AI模型，尤其是深度学习神经网络，通常被描述为“黑箱”。这种不透明性——在模型层面，常被称为“可解释性”——可能会加剧关于潜在偏见、可靠性、公平性、知识产权保护和尊重的问题；问责制问题；或AI系统的意外行为。黑箱行为可能会影响这些方面。 难以避免 对于某些类型的算法或性能水平 ， 以及 其他机制 可以常常提供人们期望从解释中获得的好处。但在某些领域和应用中，AI模型的可解释性可以帮助接收方理解AI的输出结果，并了解和判断何时应该信任这些结果。为此，寻找解释AI模型行为的技术变得尤为重要。 活跃的研究领域 应该进一步投资。

模型设计和部署中的透明度

即使对于黑盒模型而言，也可以采取措施来提升AI系统开发与部署过程中的透明度。透明度有助于人们理解模型在特定情境下的表现，其内容可能包括AI系统的数据来源、模型训练流程、验证过程，以及模型适用或应避免的情境等。 模型卡 and 数据集的数据表 是两个有用的工具)。

另一种开放性也可能有帮助 ： AI 模型可以发布 “打开 ” 或 “ 关闭 ” 以不同的方式和不同的程度，不同的方式带来不同的上下文优势和劣势。开放模型(通常被称为 “开源 AI ” 模型，尽管这些术语可能意味着不同的东西) 可以提供显著的好处，特别是对于中小型企业。一个好处是开放模型通常允许研究人员进行审计的更大能力。另一个是开发人员可以更好。 适应和定制
将模型应用于他们的用例 ， 并以一小部分的成本来做到这一点。

图片来源 ： 非洲力量

由于这些原因，开源AI模型也可能有助于赋能更广泛的AI开发者群体。通过提供更大的自定义和适应性，开源模型能够帮助社区根据其特定需求和背景定制模型。这是探索这一概念的一个理由。 人工智能系统作为数字公共产品 (DPG) 。

同时，由于双用途基础模型具有广泛可用的权重，它们可能会引发一系列伦理和政策挑战，原因在于这些模型存在被滥用的风险，且在开源发布后几乎不可能“收回”模型。美国国家电信和信息管理局近期发布的一份报告审查了这些政策关切，并建议在采取措施确保政府能够应对潜在风险增高的情况下，积极监控风险。

组织内部和 AI 治理中的透明度

组织层面的透明度涉及对AI系统的设计、部署和使用进行清晰的沟通。组织应当建立相关政策和实践以确保AI的责任使用并保证可追溯性。这包括提供关于AI系统目标、限制以及潜在影响的信息，以及为满足相关标准或政策要求所采取的步骤。工具如 NIST AI 风险管理框架 和 AI 权利法案的蓝图 对此尤其有帮助。

上一级 ， 人工智能的政策制定也应该是 开放和透明 为了确保监管和标准的制定过程具有包容性。透明的政策制定涉及与多元利益相关方开展开放咨询，包括行业代表、学术界、民间社会、工人和工会以及普通公众。这种参与式方法有助于识别人工智能的社会影响，并在保护公共利益的同时促进创新。例如，当NIST在制定 AI RMF ，而且当白宫制定《人工智能权利法案蓝图》时，他们举行了一系列听证会和公开公众研讨会，发布了信息请求，并提供了公开草案供评论。总之，这种方法产生了一份包含了广泛视角的文件。

数据营养项目 - 提高数据透明度以实现更好的 AI 系统

The 数据营养项目 这是一个旨在提高AI领域中使用的数据集质量与透明度的项目。该项目受到食品产品营养标签的启发，创建了“数据集营养标签”，提供了关于数据集组成、来源以及潜在偏差的详细信息。这些标签有助于从业者为特定的AI应用做出知情决策，从而促进更加负责任的AI开发。

数据营养标签包括多种指标和度量标准，从多个维度评估数据集的质量，如伦理考量、数据组成以及潜在危害。这种标准化的数据集评估框架可以促进透明度，使用户能够更好地理解特定于某些基于这些底层数据集训练的AI系统的优缺点，并且可能还能够根据已知的数据缺陷确定它们适用和不适用的具体情境。像数据营养项目这样的倡议为如何在AI领域增加透明度提供了模型，强调了审查和改进支撑AI系统的数据的重要性。

项目 AEDES - 为公共卫生创建开放 ， 适应性强的 AI 系统

项目 AEDES 专注于通过利用数字和气候数据来改善菲律宾的登革热公共卫生应对措施，并被认定为一项数字公共产品。 数字公共产品联盟 （ DPGA) 项目 AEDES 致力于创建解决关键健康挑战的开源解决方案。

项目AEDES作为一个先进的早期预警和探索服务，旨在通过气候和数字数据预测登革热病例，并利用卫星数据识别潜在的热点区域。鉴于该疫情的影响，这一举措尤为重要。 高利率 菲律宾的登革热。

一个区别于Project AEDES的特点是其对开放数据和开放模型的承诺，这使其能够为公众所使用并具有灵活性。该项目利用 实时数据 从各种来源进行建模和预测登革热 outbreak，包括气候读数和谷歌搜索趋势。此外，它还利用卫星图像来定位可能存在静止水体的区域，这些区域可能是携带登革热病毒的蚊子的繁殖地。

公共部门的算法问责制

在2021年，Ada Lovelace研究所、AI Now研究所和开放政府伙伴关系（OGP）启动了一个合作项目，旨在增加公共部门的算法问责制。鉴于算法在福利资格认定、失业欺诈检测和城市规划等领域越来越多地用于辅助和自动化决策，该项目重点关注这些系统中透明度和问责制的必要性。

该项目强调了人工智能政策制定和AI工具实施中透明度的重要性。它指出，许多当前的算法系统缺乏透明度和问责制，这可能导致无意中加剧歧视等问题。

为了应对这些问题，该项目审查和分析了各种旨在增强透明度和问责制的政策机制。这些机制包括影响评估、审计和监管检查、外部监督机构、听证权和上诉权，以及采购条件。这些机制旨在提供算法系统运行方式的透明度，识别并减轻潜在的偏见或缺陷，并强化其使用时的问责制。同时，该项目强调了公共透明的重要性，通过向公众提供关于算法系统的相关信息。这种透明度使个人和团体能够了解这些系统，并要求对特定背景下使用AI系统的原因进行说明。此外，本报告还强调了公众参与政策制定的重要性，以确保受影响社区的需求得到满足，并确保政策既有效又公平。

通过汇总各个司法辖区首次实施算法问责制政策的经验教训，该项目为政策制定者和公共部门工作人员提供了实用指导。其目标是支持问责机制的一致和有效实施，促进公众参与，并鼓励跨部门和不同治理层级的协调。

关于通过开放性、透明度和可解释性培养人工智能信任的建议

政府

- 实施促进人工智能设计、部署和使用透明性的治理框架，包括定期审计和适当披露数据来源、算法、部署流程以及部署后监督流程的相关信息。确保与透明性、准确性和公平性相关的这些政策和框架通过广泛的利益相关者开放咨询来制定。

美国的支持

通过与全球最大的科技公司进行互动，包括获得自愿承诺以管理AI带来的风险，美国正致力于在AI生态系统中实现开放、透明和可解释性。此外，USAID支持获奖者增加AI应用的透明度，并与哥伦比亚政府合作，构建一个AI系统以提高透明度并改善社区组织对基础设施项目访问的便利性。

私营部门

- 在AI设计、部署、使用及监督过程中实施透明化实践，包括明确记录数据来源和算法。投资于可解释AI技术及其他帮助用户解读、信任并采取行动的AI系统输出的方法。以用户易于理解的方式对最终用户保持透明和清晰，包括通过参与式的研究与开发以及用户体验设计与最终用户合作。

发展捐助者和慈善组织

- 支持专注于开发实用透明工具和可解释AI系统的研究倡议，包括针对不同受众（例如技术性和非技术性）开发的工具，并评估其可解释性或透明性特征的下游影响。

- 投资促进适当、尊重权利的开放模型开发的项目，并共享最佳实践和透明与合作的机会；支持倡导透明人工智能治理的民间社会组织。

为气候行动部署 AI

尽管与人工智能相关的能源成本和温室气体排放是重大关切，但也确实存在这样的事实，即人工智能技术，尤其是更为传统的预测机器学习系统（即非生成式AI），保持巨大的潜力为了加速适应气候变化和减少碳排放的努力，包括在低收入和中等收入国家（LMICs）。例如，存在一个活跃的研究领域，专注于使用 AI 提高能源电网效率和效率数据中心本身。对于能源网格，AI 可以帮助需求预测和电网监测减少停电并 potentially 增强能源分配效率。AI 被用于预测森林砍伐，使各国能够在问题发生之前解决潜在的土地清理问题。

机遇与挑战：

广泛的人工智能-enable的应用目前正在支持气候行动的进步，并有望进一步加速这一进程。在气候领域的有前景的人工智能应用领域包括能源效率、清洁能源部署以及其它气候缓解目标，还包括气候适应性和韧性。

此外 ， AI 正在积极 正在使用 在管理气候变化的影响 ， 支持气候适应和复原力。例子包括 正在进行的使用 在目标自然灾害和应急响应中应用AI，包括预测洪水和野火，并制定应对计划。AI还被用于加速应急响应的部署。 清洁能源基础设施 .

天气 - 气候基础模型 ， 如 NASA 和 IBM 的 Prithvi 可以应用于更好地追踪气象相关灾害并提高预测准确性。人工智能还被用于保护自然资源。例如，GIZ的FAIR Forward倡议正在与当地社区合作，开发和部署AI应用。 保护印度尼西亚的林地 。与此同时 ， IDRC 正在支持人工智能和气候变化创新 研究网络 ，旨在开发和规模化AI解决方案以推动气候行动，并为人工智能与气候change交叉领域的学生提供资金支持。

2021 年 ， 全球人工智能伙伴关系(GPAI) 与气候变化人工智能和人工智能与气候中心一起发布了一份题为 “ 气候变化与人工智能 ： 政府行动建议。 这项研究突出了人工智能在气候行动中可以做出贡献的关键领域，并提供了政府如何更好地合作以减少人工智能对气候的负面影响的建议。同时，它还指出人工智能在预测太阳能发电量、优化供暖和冷却系统以及从卫星图像中检测森林砍伐等方面特别有用。此外，人工智能在科学研究中也可能大有作为，例如通过建模气候系统、发现新的能源存储材料、模拟日益复杂的气候情景以及绘制地图等。 气候脆弱人群 .

AI + 亚洲气候期货 - AI 与气候交汇处的行动路线图

The AI + Climate Futures in Asia项目由Digital Futures Lab实施并得到了Rockefeller基金会的支持，开展研究以全面了解AI与气候行动之间的交集。该项目提供了有关我们如何利用AI支持气候缓解、适应和韧性方面的见解和建议。

这在许多高度易受气候变化影响的亚洲地区尤为重要：洪水、频发的风暴、热浪、干旱、空气污染以及其他影响——所有这些都对人类生计构成风险。

AI与亚洲气候未来聚焦于9个亚洲国家的几个关键领域：农业和食品系统、能源转型以及灾害响应与准备。总体而言，该研究建议了一系列行动，这些行动与本手册中的推荐高度一致：开放数据共享、严格评估试点项目、跨学科合作、本地化和参与式数据收集，以及投资绿色计算等。

为气候行动部署 AI 的建议

美国的支持

美国一直走在多边合作利用人工智能技术支持气候行动和清洁能源部署的前沿，特别是在发展中国家通过联合国气候变化#AI4ClimateAction倡议和美国发起的净零世界倡议。此外，通过一个名为SERVIR的项目，NASA和USAID利用地理空间技术和人工智能增强可持续性和气候韧性。通过COOLERCHIPS计划，美国正在投资研发以减少数据中心的总制冷能耗。

通过清洁能源需求倡议（CEDI），美国国务院将科技公司与关键市场的政府连接起来，以支持清洁能源政策的采纳，确保公司能够获取并购买到其运营所需的清洁能源。

在2024年7月，美国能源部（DOE）宣布了AI for Science, Security, and Technology (FASST)计划，该计划利用DOE及其17个国家实验室，通过AI测试平台支持高效能的AI硬件，开发安全可靠的模型并注重能源效率，以及加速AI技术以帮助科学家发现新的能源材料。此外，美国能源部还利用其国家实验室在全球范围内支持清洁能源的生成，包括通过国家太阳能辐射数据库促进全球清洁能源的开发和应用。

政府

- 实施促进清洁能源发电投资的政策和监管框架存储、电网基础设施和能效措施 ， 并使私营部门采购清洁能源 ， 以减少与 AI 系统相关的碳排放。
- 促进本地企业、研究机构与国际合作伙伴之间的合作，开发旨在应对气候变化的AI解决方案。
- 将直接的科研资金投入用于减轻人工智能对气候的影响，并应用人工智能于与气候相关的问题。
- 鼓励投资可再生能源以减少与人工智能技术相关的碳排放。

私营部门

- 努力确保AI的发展具有可持续性并尽可能减少温室气体排放，通过投资和采用清洁能源解决方案。
- 选择致力于具有气候效益应用的AI部署项目和合作伙伴。避免选择可能进一步加剧气候变化的项目。

发展捐助者和慈善组织

- 研发能源高效的人工智能技术以减少训练和运行人工智能模型所需的计算资源，并确保其可持续性。
- 支持将AI解决方案应用于气候缓解和适应努力的倡议，并特别与低收入和中低收入国家（LMIC）的利益相关者开展合作。

Conclusion

人工智能为促进可持续发展提供了重要的机遇。本手册作为拜登总统关于安全、可靠和可信赖的人工智能开发与使用的行政命令的关键成果之一而制定，旨在强调人工智能作为一种具有深远影响和潜在风险的独特全球机遇。手册指出，安全、可靠和可信赖的人工智能的设计、部署和使用不仅是可能的，而且是必要的。在整个案例研究和示例中的一贯主题是如何通过积极且多样化的利益相关者参与来实现这一负责任的设计、部署和使用，确保涉及或可能受人工智能系统影响的人群的有意义代表性。

这份 playbook 也标志着美国在全球范围内负责任地部署和使用人工智能方面的又一承诺，根植于人权理念。

美国当前的努力基于这样一个信念：当负责任地开发和部署时，人工智能可以成为实现可持续发展目标、应对世界上最紧迫挑战的强大动力。展望未来，美国将继续通过资金支持、倡导活动和召集努力来支持低收入和中等收入国家（LMICs），共同 navigating 数字时代的复杂性，并致力于一个技术发展利益广泛共享的未来。

我们承认在这方面仍有许多工作需要完成，而这份手册仅是推进可持续发展所需采取的众多措施之一。认识到国际合作和多利益相关方合作伙伴关系在取得进展方面的重要性，我们邀请其他人贡献他们的专业知识、资源和视角，以丰富并扩展这一框架。负责任的人工智能真正的进步标准不在于我们机器的复杂性，而在于技术提升的生活质量。我们共同可以致力于确保人工智能的潜力服务于这一目标。

附件

Resources

- 美国国际开发署数字生态系统框架
- 美国国际开发署人工智能行动计划
- 人工智能和人权的风险管理简介
- 人工智能风险管理框架 ： 生成人工智能简介
- AI 权利法案的蓝图
- 反映过去 ， 塑造未来 ： 让人工智能为国际发展服务
- 国务院的人工智能
- AI. gov
- 全球 AI 研究议程

AI
在全球发展中的
PLAYBOOK