联合国 E/C.16/2024/7

经济及社会理事会

Distr.: General
29 January 2024
Chinese
Original: English

公共行政专家委员会

第二十三届会议

2024 年 4 月 15 日至 19 日，纽约

临时议程* 项目 9

数字政府

人工智能治理加强《2030 年议程》，不让任何人掉队

秘书处的说明

秘书处谨向公共行政专家委员会转递委员会成员谢里法·谢里夫和卡洛斯·桑蒂索编写的文件。

* E/C.16/2024/1。

24-01544 (C) 080224 200224

人工智能治理加强《2030 年议程》，不让任何人掉队

摘要

人工智能正在全世界得到越来越多的应用，带来了巨大的潜在惠益，增强了人类的能力，提高了人们的福祉，促进了社会的进步。然而，随着技术继续以前所未有的速度发展，许多挑战、风险和道德问题仍然存在，必须紧急解决。政府作为人工智能的监管者和使用者可以发挥特别重要的作用，考虑到政府对人们生活的巨大影响，这种作用尤其突出。

作者们详细阐述了人工智能的演变，指出人工智能以从前无法想象的方式改变了各行各业和人们的日常生活。他们还概述了人工智能在加速落实《2030 年可持续发展议程》方面的潜力，同时简要介绍了人工智能对某些可持续发展目标的惠益。

作者们随后详细阐述了从长期看使用人工智能带来的诸多风险和挑战，特别是发展中国家的风险和挑战，从而引发了对道德、公平、透明度以及遵守现有和新兴法规的关切。

作者们呼吁开展人工智能治理，其目的是确保平等获得人工智能的各种惠益，保护数字权利，防止伤害。他们概述了现有的人工智能治理实践，同时强调了构建强大的人工智能治理框架所需的主要内容。

最后，作者们强调有必要继续正在进行的全球讨论，并对人工智能的积极和消极影响建立共识。必须确定原则、设立政策优先事项并确保政策的一致性，以使人工智能能够加强《2030 年议程》，不让任何人掉队。

一. 导言

1. 在技术发达的当今世界，人工智能得到政府、组织和民众越来越多的使用，带来巨大的潜在惠益。人工智能系统越来越能干并日益融入我们的日常生活，可以不断增强人类能力、提高人民福祉、促进社会进步。人工智能可以促进可持续经济增长、提高创新和生产力、助力应对全球挑战。例如，人工智能驱动的气候建模可以助力解决紧迫的气候变化问题，而在教育领域，人工智能可以赋能个性化的学习体验，并使偏远或服务不足地区的人们更容易获得教育机会。

2. 政府作为人工智能的监管者和使用者发挥特别重要的作用，考虑到人工智能对人们生活的影响及其确保人民福祉的责任，这种作用尤其突出。[1] 政府已开始投资于人工智能技术，以支持智慧决策、提高运营效率，从而解决诸如交通流量、获得教育和医疗服务机会、基础设施监测、网络攻击等各种公共部门问题。目前，人工智能被用于公共部门的不同领域，如执法、司法行政、预防欺诈、税收和反腐等。[2] 人工智能可以成为政府创新的神奇助推器，改变公共行政部门的思维和运作方式，并有可能提高内部运营效率和决策效力(包括更好地确定公共开支和社会转移支付的针对对象)、改善公共服务的提供和响应能力(包括转向更加灵活、个性化、积极主动和以人为本的公共服务)、促进政府廉政和预防欺诈以及公共政策评估。

3. 然而，随着技术的不断演变发展，许多挑战和道德问题仍然存在，必须紧急解决。隐私和安全问题以及人工智能可能被滥用等需要得到认真考虑和监管。一项具体的挑战涉及人工智能正在改变公共治理的方式以及公共部门本身对人工智能的负责任使用情况，而公共部门在保护人们的数字权利方面负有特殊责任。在公共部门部署人工智能，特别是在社会福利、发现欺诈、执法和司法判决以及个性化服务等敏感政策领域部署人工智能，需要有强有力的防护栏。使用人工智能开展治理必将成为数字时代的决定性挑战。

4. 展望未来，有必要继续进行全球对话，并就人工智能对政府机器的积极和消极影响达成共识。必须确定原则、设立政策优先事项、确保政策一致性，以加强落实《2030 年可持续发展议程》。人工智能治理方面的集体行动对于尽量减少其负面后果至关重要，重点是不让任何人、任何国家掉队，并确保平等获得技术。在促进以人为本、以权利为基础的人工智能部署以及支持建立亟需的人工智能全球治理和监管方面，联合国可以发挥举足轻重的关键作用。

---

[1] 例如，见经济合作与发展组织(经合组织)和拉丁美洲开发银行，《在拉丁美洲和加勒比公共部门战略性和负责任地使用人工智能》(巴黎，经合组织，2022 年)；Jamie Berryhill and others，“Hello, world: artificial intelligence and its use in the public sector”, OECD Working Papers on Public Governance, No. 36 (OECD, 2019)。

[2] 例如，见世界经济论坛，“数字时代的黑客腐败：技术如何在危机时期塑造诚信的未来”，2020 年 5 月；Carlos Santiso，“Trust with integrity: harnessing the integrity dividends of digital government for reducing corruption in developing countries”, DESA Working Paper, No. 176 (New York, United Nations, Department of Economic and Social Affairs, 2022)。

5. 本文件在公共行政专家委员会以往关于数字政府的工作基础上编写。委员会在第二十二届会议上指出，展望未来，迫切需要做出努力，通过采用新的政策和监管制度及标准，除其他外消除数字鸿沟、改善数据治理并减轻使用人工智能和社交媒体等新技术所带来的风险。

6. 会员国在 2023 年 9 月在大会主持下召开的可持续发展高级别政治论坛的政治宣言(第 78/1 号决议)中同意，它们将力求更好地实现人工智能的惠益并应对其挑战。

7. 2023 年，秘书长宣布创立一个新的咨询机构，以评估人工智能的风险、机遇和国际治理，支持国际社会治理人工智能的各项努力。人工智能高级别咨询机构在其第一份中期报告中确认，全球人工智能治理对于抓住人工智能的重大机遇、同时降低其在今天和未来给国家、群体和个人带来的风险至关重要。秘书长在题为“我们的共同议程”的报告(A/75/982)中还建议在 2024 年未来峰会上商定一项“全球数字契约”，除其他外，该契约可以促进对人工智能的监管，以确保其符合全球共同价值观。[3]

二. 人工智能的定义和演变

8. 目前，人工智能没有普遍认可的定义。就本文件而言，人工智能被定义为机器通过执行各种认知任务(如感知、处理口头语言、推理、学习、决策和展示相应的操纵物体能力)来模仿人类智能行为的能力。[4] 人工智能本质上是使用算法来模仿人脑的操作和程序，目的是使计算机像人类一样思考和行动。这些算法的主要功能是模式识别、预测和控制，人工智能因此成为各国政府、区域和国际组织议程上的一个关键问题。人工智能有能力处理和分析巨量数据，可以用来辅助日常任务、解决长期以来人类无法理解的复杂问题。

传统人工智能与生成式人工智能

9. 传统的人工智能接收输入并产生输出，数据得到分析并被用于决策和预测。传统的人工智能仍然非常受欢迎，被用于为聊天机器人和预测分析等大量人工智能系统提供动力。传统的人工智能依赖基于规则的方法，通过对明确的指令和预定义的规则进行编程，使系统能够执行具体任务并生成输出。这些规则是由人类专家根据他们对问题领域的理解设计的。

10. 另一方面，生成式人工智能采用数据驱动的办法，使用机器学习技术从大型数据集中学习模式和结构。生成式人工智能模型不依赖明确的规则，而是从数据中学习，并通过捕获数据内的底层模式和关系来生成新内容。生成式人工智能为用户提供了更具创造性和创新性的机会，从而减少了花在构思过程上的时间。ChatGPT 是生成式人工智能工具的一个例子。

[3] 如需详细信息，请查阅 https://www.un.org/techenvoy/global-digital-compact。

[4] 另见亚洲及太平洋经济社会委员会，《亚洲及太平洋人工智能》，政策简报，2017 年 11 月。

演变

11. 人工智能的演变是一个非凡的历程，无数的突破和创新加速并推动了这一领域的发展。从 20 世纪 50 年代的卑微开端到今天看到的复杂深度学习模型，人工智能以从前不可想象的方式改变了各行各业和人们的日常生活。

12. 20 世纪 50 年代，阿兰·图灵发明了图灵测试，以确定机器能否模仿人类智能，从而引入了人工智能。20 世纪 60 年代，约翰·麦卡锡开发了第一个人工智能编程语言 LISP。早期的人工智能系统侧重于符号推理和基于规则的系统，这最终导致了 20 世纪 70 年代和 80 年代专家系统即模仿人类专家决策能力的计算机系统的开发。

13. 20 世纪 90 年代，人工智能的重点转向机器学习和数据驱动办法，这是数字数据可用性增加和计算机发展的结果。神经网络的兴起使人工智能系统能够从数据中学习，从而带来了更好的表现和适应性。在 21 世纪，人工智能研究进入新的领域，包括自然语言处理、计算机视觉和机器人学，这为今天的人工智能革命铺平了道路。

14. 如今，全球范围内人工智能方面的政府支出正在增加，在加拿大、中国、大不列颠及北爱尔兰联合王国和美利坚合众国这点尤其突出。2020 年，美国政府为人工智能项目提供了超过 10 亿美元的资金。2021 年 3 月，加拿大政府承诺斥资 5 亿多美元推动人工智能举措。继 ChatGPT 发布之后，市场上又涌现出数百项人工智能产品，令人难以置信。

15. 预计人工智能将在以下领域产生主要影响：医疗保健；汽车；财务服务；零售和消费者；技术、通讯和娱乐；制造业；能源；运输和物流。[5] 例如，在汽车行业，人工智能可用于改善车辆性能、司机安全和乘客体验，而人工智能机器人正被用于装配线。在医疗保健领域，人工智能可以助力减少人为错误、协助医疗专业人员、提供全天候患者服务。

16. 人工智能演变的速度是前所未有的。部分原因是因为技术可以自我增强，从而提高自身能力。在所谓的“数据革命”背景下，人工智能还可以利用不断扩大的新数据来源。[6] 随着物联网和非结构化大数据等新技术的使用日益增多，产生的数据也越来越多，而随着互联网的速度日益加快，人工智能的发展速度只会越来越快。此外，人工智能是一种具有无限应用范围的通用技术。虽然大多数人更熟悉 ChatGPT 或文本到图像生成器，但人工智能也可用于能源系统或供水等关键基础设施，因此，对其进行负责任的管理并设置防护栏是不可或缺的。最近在联合王国举行的 2023 年人工智能安全峰会上强调的一个重要风险是，生成式人工智能可能会失控，超出人类的预期或意图开展的监督，做出技术开发者未曾预见或不打算做出的决定或采取的行动，从而造成潜在的破坏性后果。

[5] 同上。

[6] 数据革命促进可持续发展问题独立专家咨询小组，“世界需要数据：发动数据革命促进可持续发展”，2014 年 11 月。

17. 随着人工智能所能实现的边界不断被突破，新的挑战和伦理困境将不可避免地出现。然而，通过促进在政府、国际组织、企业和研究人员等不同行为体之间建立合作环境，人工智能的演变可以集体愿景为指导，优先考虑世界各地的社会改良和个人福祉。随着落实《2030 年议程》的最后期限迅速临近和向前推进，应继续投资于研究和开发，确保以负责任和合乎道德的方式利用人工智能的潜力，继而应对全球挑战、为所有人创造更美好的世界，不让任何人、任何国家掉队。

三. 人工智能在加速实现可持续发展目标方面的潜力

18. 人工智能可以增强创造力和解决问题的能力，并有望为设计可持续解决方案开辟新途径，从而加速落实《2030 年议程》，促进不同行业和部门的创新。本节简要介绍人工智能在支持加快实现某些可持续发展目标方面的潜在惠益，同时铭记所有这些目标是相互关联的。[7]

目标 1(消除贫困)

19. 《2030 年议程》旨在确保在世界各地消除一切形式贫困，不让任何人掉队。人工智能可以通过以下方式支持这一承诺：

- **支持确定弱势群体、跟踪贫困水平。**这样可以更好地确定扶贫和基于公平的政策和计划的针对对象，确保援助和资源分配到最需要的地方。此外，还可以预测未来的趋势和需求，以更好地规划未来的干预措施。
- **改善获得基本服务的机会。**人工智能可以改善穷人和弱势群体获得教育和医疗保健等基本服务的机会。例如，人工智能可以评估巨量医疗保健数据和人口信息，以确定需要医院、流动医疗小组或远程医疗服务的区域。
- **促进农业发展。**例如，人工智能可以预测作物产量和价格，让农民获得最大利润，同时还可以监测作物和土壤的健康状况。此外，人工智能机器人可以更快的速度收获更多的作物，增加农民收入。
- **促进金融普惠。**人工智能可以通过为弱势群体提供负担得起、容易获得的银行服务来促进金融普惠；可以更准确地评估信用，使个人和小企业更容易获得贷款和金融服务；可以在数据分析的基础上确定最值得资助的受助人，从而助力高效分配小额信贷。

目标 4(优质教育)

20. 人工智能有潜力应对当今教育领域一些最紧迫的挑战，实现教学和学习转型，为所有人创建一个更加个性化、更加有效、更加容易获得的教育系统，并加快实现目标 4。人工智能对教育部门的潜在积极影响包括：

[7] 另见 Ricardo Vinuesa and others, “The role of artificial intelligence in achieving the Sustainable Development Goals”, Nature Communications, vol. 11, 2020。

- **个性化动态学习。**人工智能可以为每个学生量身定制教育内容，从而提高学习成果和学生参与度，还可以创建互动教育工具。

- **包容。**凭借先进的翻译和理解能力，人工智能技术可以助力弥合知识和语言差距，使更广泛的受众能够获得优质教育和资源，而不受语言或地方/区域限制。生成式人工智能驱动的平台可以为学习者提供全天候的协助，使优质教育更加普及。此外，人工智能驱动的工具可以为残疾学生提供支持(如语音到文本的转换)。这一点非常重要，因为“不让任何人掉队”必须意味着每个人都能接触到当前的技术革命并从中受益。

目标 8(体面工作和经济增长)

21. 人工智能正在成为经济活动不可或缺的工具，其对经济发展的巨大贡献正在迅速显现，人工智能与经济活动各部门之间存在着明显的交叉点。例如，人工智能可以改善生产、提高效率、增进生产线的安全性，同时降低成本，从而使不同部门能够以具有竞争力的价格提供良好服务。

22. 人工智能算法是数据驱动的，能够随着时间的推移学习数据趋势，因此非常适合预测增长率、利率、汇率和通货膨胀率等经济指标，而这些指标对货币政策管理和经济稳定至关重要。对这些指标的准确预测可以为政策制定者提供支持，使他们能够更积极主动地预测下一次金融危机等即将到来的挑战。在股票和债券等交易型资产方面，人工智能技术还可以预测价格走势，使决策者能够在最佳时间进行交易。

23. 在全球层面，生成式人工智能可以弥合语言和知识差距，促进更大的国际合作。在国家层面，人工智能可以通过以下方式促进体面工作和经济增长：

- **提高生产力，协助完成复杂任务。**生成式人工智能可以协助人类执行复杂的任务，提高产出和效率。与人类相比，生成式人工智能能够更有效地管理、分析和处理信息，从而有可能提高整体生产力。据估计，到 2030 年，生成式人工智能对生产力的影响可能会为全球经济增加合计 15.7 万亿美元的价值。[8] 人工智能有能力节省数百万工时，这也是在政府程序中使用人工智能的一个重要理由。因此，员工可以腾出手来，专注于更重要的智力任务。近期估算得出的结论是，政府工作人员的任务自动化每年可节省 33 亿至 411 亿美元。[9]

- **降低成本、提高效率。**人工智能和自动化可以加快处理速度、降低成本、加快服务的提供。人工智能驱动的机器人还可以全天候工作，确保产出的持续供应。大约 33%的制造商已经通过人工智能和其他技术降低了劳

[8] PricewaterhouseCoopers, “Sizing the prize: what’s the real value of AI for your business and how can you capitalize?”, 2017; 另见 Michael Chui and others, The Economic Potential of Generative AI: The Next Productivity Frontier (McKinsey & Company, 2023).

[9] Deloitte, “AI-augmented government: using cognitive technologies to redesign public sector work”, 2017.

动力成本。人工智能和自动化还有望将计划外停机时间和产品缺陷减少50%、实现高达 20%的制造业增效。[10]

- **创造新的就业机会。**随着生成式人工智能的发展，可能会出现新的行业和职业，就像当初信息技术革命导致产生大量技术工作一样。世界经济论坛《2020 年未来就业报告》估计，到 2025 年，人工智能和技术的发展将创造约 9 700 万个新的就业机会。

- **改善获得专门知识和服务的机会。**在缺乏专家的领域，生成式人工智能可以提供必要的专业知识和服务。

24. 人工智能相关活动将成为许多国家进一步经济发展的推动力，并将导致生产结构和方法以及消费数量和质量的根本转变。随着技术形势的快速推进，人工智能将重塑经济、劳动力市场和各行各业，为不同部门带来革命性的变化。鉴于其对全球劳动力队伍和经济差距的影响，需要制定深思熟虑的政策。

目标 9(产业、创新和基础设施)

25. 人工智能有可能通过以下方式加速实现目标 9：

- **加速创新。**生成式人工智能可以通过分析巨量数据、预测结果和生成创新解决方案来加速各行各业的研发。

- **管理基础设施。**生成式人工智能可以通过预测潜在的基础设施故障、实时优化交通系统以及更有效地管理大型公用事业来加强基础设施管理。

- **强化制造业。**先进的人工智能系统可以推动自动化、优化供应链、预测机械维护问题，从而提高制造业的整体效率。

26. 政府应努力引进和推广新技术、为国际贸易提供便利、使资源得到高效利用，并加大对科研和创新的投资。

目标 13(气候行动)

27. 人工智能可以通过支持减缓和适应措施来助力应对气候变化，其中包括：

- **改进气候变化模式的建模和预测。**人工智能可以根据秘书长的“全民预警”倡议助力社区和主管部门起草更有效的适应和减缓战略，让他们为即将发生的极端天气事件(如热浪、干旱和洪水)更充分地做好准备。

- **改善城市规划和交通管理。**人工智能可以减少温室气体排放，使城市更加可持续、更加宜居。人工智能还可以跟踪污染水平，使地方政府能够在出现危险水平时向公众发出警报。

[10] Saxon, “Impact of AI in manufacturing: improved quality and efficiency”, December 2022.

- **支持碳中和。**人工智能在支持各国走上碳中和道路方面发挥着关键作用。例如，人工智能可以优化制造过程，从而减少制造过程对环境的影响、减少交通流量、提高可再生能源来源的效率。

四. 人工智能的风险和挑战

28. 尽管人工智能的广泛应用可能会促进短期经济繁荣并带来诸多潜在惠益，但从长远看，过度依赖人工智能可能会带来诸多风险和挑战。

过度依赖技术

29. 过度依赖人工智能可能会减少人类的互动和联系、削弱批判性思维和其他基本的软技能，从而导致创造力、社交技能和同理心的丧失。此外，技术难题和故障可能会扰乱教育、学习和生产力。

工作岗位流失和不断变化的要求

30. 人工智能最令人关切的问题是非自愿失业，更笼统地说，是人工智能时代工作的未来。人工智能有可能通过模仿人类的认知过程、以更快的速度和更低的运营成本完成目前由员工完成的一些日常活动，从而消除工作岗位。

31. 体力劳动和涉及重复性任务或可以系统化的工作最有可能被自动化，其中不仅包括蓝领工作，而且包括某些白领职业(如会计、编辑、零售和快递服务人员、保安人员甚至医生)。人工智能的普及还可能导致中等工资就业机会的流失，在依赖人类创造力和内容生成的行业这点尤其突出。根据世界经济论坛《2020 年未来就业报告》，由于人工智能和相关技术的发展，到 2025 年可能会流失约 8 500 万个工作岗位。

32. 随着某些形式人类劳动的价值因生成式人工智能的各种能力而降低，可能会出现一个动荡的调整阶段。传统角色可能会遭遇需求减少、报酬降低问题，而其在许多国家劳动力队伍内部的差距会扩大、国际不平等有可能会加剧。此外，对新技能的需求可能会增加，同时会转向更灵活的工作安排。各国政府可能需要更积极主动地进行干预，例如制定新的劳工政策、支持工人向新行业过渡、甚至探索普遍基本收入等新理念，以应对劳动力成本大幅下降的情况。它们还必须建设公共部门员工队伍的内部能力，以免将所有基于人工智能的技术开发都外包给私营部门合作伙伴，从而得以更好地理解、开发和管理这些技术。

缺乏技能

33. 虽然一些职业会流失，但预计人工智能也将对就业市场产生有利的整体影响，为有能力的工作人员创造充分的机会。但人工智能的这种积极影响只有在各国对劳动力进行技能再培训并提高其所需技能和能力的前提下才可行。据 IBM 估计，

在未来三年内，大约 40%的工作人员(全球 34 亿劳动力中的 14 亿)将需要重新掌握技能。[11] 数字文盲问题也需要解决。

传统产业流失

34. 人工智能的使用可能会导致传统行业的流失。在经济依赖传统产业的发展中国家，人工智能驱动的快速自动化可能导致经济不稳定。

缺乏优质数据

35. 人工智能程序的好坏取决于其运行所依赖的底层算法和数据。如果不使用含有代表性准确数据的可靠登记册对算法进行微调和道德审查，就会对结果产生负面影响、增加偏倚风险，从而可能导致新形式的排斥和歧视。这个问题在“数据贫乏”的发展中国家尤其突出，因为这些国家缺乏关于其公民的优质数据。因此，经济合作与发展组织(经合组织)建议更加重视数据的道德问题，特别是在公共部门重视这个问题。[12] 人工智能战略应进一步融入数据治理和基础设施战略以及高效的数字化行政登记册，或与之挂钩。然而，这仍然是一个普遍的挑战。

经济差距和公平问题

36. 数字技术的进步令人震惊，数字工具的使用日益增加，互联网基础设施得到改善，产生了明显的社会效应。然而，不同国家、不同行业、不同部门、不同社会制度对人工智能的使用和发展情况大相径庭。数字革命的惠益需要在各经济体和个人之间平均分配，增加获得数字可能性的机会、弥合数字鸿沟。

37. 如果生成式人工智能技术主要由少数国家或公司开发和拥有，则可能会导致全球经济严重失衡、国家之间的经济差距进一步扩大。发展中国家缺乏获得新技术的机会，有可能加剧国家之间的不平等。配备人工智能技术的制造业巨头将加速增长，而无法获得此类先进技术的发展中国家将被甩在后面。同样，如果只有某些区域或群体才能获得先进的人工智能教育工具，则可能扩大国家内部和国家之间的教育差距。如果不加以控制，这些先进技术造成的破坏可能会产生巨大的社会后果。因此，重要的是要确保所有国家和人民都能从人工智能发展中受益。

38. 各国政府应努力最大限度地减少经济两极分化，以防止人工智能的惠益分配不公，同时防止经济鸿沟扩大、导致更大比例的财富被拥有和管理生成式人工智能系统者所控制。

道德和伦理问题

39. 虽然新技术可以带来经济效益，但也可能因助长偏见和歧视而触发对某些群体的不利影响。这种两极分化的情况要求采取精明的政策干预措施。例如，机器学习偏见、特别是在种族定性方面的机器学习偏见可能会错误地识别用户的基本信息，从而可能导致不公平地剥夺其获得医疗保健和贷款的机会，或在识别犯罪

[11] IBM Institute for Business Value, “Augmented work for an automated, AI-driven world: boost performance with human-machine partnerships”, 2023.

[12] 经合组织公共部门数据道德良好实践原则。

嫌疑人时误导执法部门。然而，在人工智能系统中灌输道德和伦理价值观，特别是对具有重大后果的决策过程进行这样做，仍然是一个相当大的挑战。

40. 尽管如此，一些国家正在制定或已经制定了人工智能治理程序和规则，以最大限度地减少算法中的偏见或歧视，如通过提高透明度和建立开放的公共算法登记册来做到这一点。例如，哥伦比亚是最早为其人工智能战略采用道德框架的经合组织国家之一。

错误信息和操纵

41. “深度伪造”[13] 等人工智能生成的内容日益助长虚假信息的传播和对舆论的操纵。发现和打击这种错误信息的努力至关重要，因为错误信息可能会损害公共机构在人民心中的合法性、加深政治两极分化、助长民粹主义运动。

隐私问题和安全风险

42. 由于生成式人工智能系统需要巨量数据，因此人们对谁控制数据以及可能的数据垄断或滥用问题表示关切。需要防止人工智能的潜在滥用(如网络攻击)，并解决对人工智能监控的关切，即出于安全、执法和营销目的使用人工智能技术监控和分析人类行为。需要严格的数据保护法规和安全的数据处理实践；还需要制定防止人工智能安全威胁的全球规范和条例，考虑到对流氓国家或非国家行为体使用人工智能驱动的自动化武器的日益关切，尤其需要制定此等规范和条例。

缺乏透明度、产生意外后果、存在潜在生存风险

43. 人工智能系统缺乏透明度，深度学习模型尤其缺乏透明度，这是一个难以解释的复杂问题，需要迫切解决，因为，如果人们无法理解人工智能系统如何得出结论或解决方案，就可能对采用这项技术产生不信任和阻力。此外，当一个算法是一个“黑匣子”时，很难对其进行有效监督。

44. 由于其复杂性和缺乏人类监督，人工智能系统也可能表现出意想不到的行为或做出具有不可预见后果的决定，对个人或整个社会产生负面影响。此外，超越人类智能的人工通用智能(自学、能够自主执行各种任务)可能得到开发开始引起人们的关切, 即，如果这些先进的人工智能系统不符合人类的价值观或优先事项，则可能产生意想不到的潜在灾难性后果。[14]

五. 人工智能治理确保可持续发展，不让任何人掉队

45. 人工智能在世界范围内的迅速采用引发了对道德、公平、透明度和遵守其他法规的各种关切。如第四节所示，如果没有适当的治理，人工智能系统可能会带来巨大风险，特别是给发展中国家带来巨大风险。

[13] 对某人的面部或身体进行数字修改、使其看起来像是另一个人的视频通常被恶意使用或用于传播虚假信息。

[14] 例如见 Bernard Marr, “The 15 biggest risks of artificial intelligence”, Forbes, 2023。

46. 政府需要评估和盘点人工智能技术，承认其潜在的惠益和固有的风险。如果得不到解决，短期挑战可能发展成长期系统性问题。因此，迫切需要全面重新思考和重新设计政策、社会保障制度、劳动力市场和税收框架，同时确保透明度、问责制和人类监督，并尊重共同的规范和价值观，如《联合国宪章》、《世界人权宣言》和国际法所载的规范和价值观。[15]

47. 值得进一步考虑在公共部门负责任和合乎道德地使用人工智能这一问题。许多数字技术先进的经合组织国家先行在政府流程和服务中使用了人工智能。作为人工智能的监管者和重要用户，它们更加关注人工智能给公共部门带来的惠益及其具体的挑战和风险。

现有的人工智能治理实践

48. 各国在制定和执行国家人工智能政策和战略方面处于不同阶段。[16] 包括加拿大和芬兰在内的一些国家早在 2017 年就开始制定本国的人工智能计划，法国、德国、日本和英国则在 2018 年紧随其后。包括巴西、埃及、匈牙利、波兰和西班牙在内的其他国家最近也推出了国家人工智能战略。如今，60 多个国家已经制定专门的人工智能战略，[17] 其他若干国家也在推进人工智能政策的制定和咨询程序。

49. 近年来，越来越多的国家政府还出台了旨在公共部门开展人工智能治理的具体政策和标准。公共部门的人工智能部署高度分散，其高效治理需要政府中心的有力指导和监督，以确保统一规则和一致标准。例如，《加拿大自动决策系统指令》介绍了加拿大政府如何利用人工智能指导多个部门决策的情况：其正在使用评分系统来评估需要何种人类干预、同行审议、监测和应急规划才能实现旨在为其公民服务的人工智能工具。

50. 智利和法国等许多国家正在通过开放登记公共算法以及对政府实体使用的算法执行信息获取规则，强制要求公共部门实体使用的人工智能算法具有透明度。智利还在努力制定拉丁美洲和加勒比第一个关于公共部门算法透明度的法规。包括阿姆斯特丹、巴塞罗那和赫尔辛基在内的若干城市也建立了开放登记制度。此外，欧洲联盟委员会还设立了欧洲算法透明度中心。

51. 包括联合王国和美国在内的其他若干国家正在使用政府采购规则，将核心(道德)原则嵌入公共部门实体的人工智能采购解决方案。通过坚持对承包商采用某些标准，可以让承包商树立榜样，影响更广泛的市场行为。

52. 一些国家和区域还致力于通过法规和政策保护数字权利。例如，西班牙通过了《数字权利宪章》，欧洲联盟签署了《数字权利和原则宣言》。

---

[15] 另见人工智能高级别咨询机构 2023 年中期报告。

[16] 另见经合组织关于人工智能的国家政策和战略的实时存储库，可查阅 https://oecd.ai/en/dashboards/overview。

[17] Carlos Santiso, “Public governance in the age of artificial intelligence”, Governance Matters (Chandler Institute of Governance, 2023).

53. 关于私营部门使用人工智能问题，亚洲及太平洋的一些国家发布了解决道德问题的若干政策和法规(如新加坡的人工智能治理和道德举措)。包括联合王国和美国在内的其他国家则正在借鉴七国集团监管人工智能的办法、推动私营部门使用自愿行为守则。

54. 一些国家还设想建立类似于数据保护机构的人工智能机构，这些机构可以牵头进行详细的影响评估、测试潜在的解决方案并在推出之前对其潜在的积极和消极影响进行研究。例如，西班牙在 2023 年创建了欧洲第一个人工智能监督机构。

55. 此外，还有一些区域举措旨在实现人工智能方面的监管趋同，包括在非洲、亚洲和欧洲实现监管趋同。例如，2023 年 12 月，欧洲联盟就其《人工智能法案》达成一致，该法案将于 2026 年生效。该法案规定，根据不同人工智能系统对用户构成的风险对其进行分类并赋予其不同的监管级别。除其他外，该法案禁止在欧洲联盟部署构成“不可接受风险”的人工智能系统，并对被归类为“高风险”或“有限风险”的人工智能系统(如“深度伪造”)规定不同程度的义务。2023 年 10 月，在联合国教育、科学及文化组织和拉丁美洲开发银行的支持下，20 个国家通过了《促进拉丁美洲和加勒比发展符合道德的人工智能的圣地亚哥宣言》。

下一步行动

56. 人们对人工智能治理的兴趣与日俱增，关注焦点是我们的日常生活有多少应该由算法来塑造以及确定由谁来控制对算法的监测。统一的全球人工智能治理办法应防止监管碎片化，并允许建设性地使用人工智能技术，同时确保平等获取、保护人权(“数字权利”)并防止伤害。[18] 这一点尤其重要，因为技术空间是跨境运作的，因此国际协调合作必不可少。这也符合将在 2024 年未来峰会上通过的拟议全球数字契约，该契约旨在概述为所有人创造开放、自由和安全的数字未来的共同原则。全球人工智能伙伴关系和经合组织人工智能政策观察站等倡议可以支持各国和不同利益攸关方之间进行必要的信息交流、对话和协作。

57. 以下是人工智能治理应旨在实现的目标内容：

- 为人工智能技术的应用建立体制框架和法律框架。
- 遵守数据治理规则和隐私法规；概述查阅及管理个人资料的准则。
- 解决与人工智能相关的道德、伦理和安全问题。
- 防止错误信息和操纵。
- 促进安全、信任和透明度。
- 确保人工智能不侵犯公民自由和法治。
- 预见和预防使用人工智能的意外后果。

[18] 另见经合组织人工智能原则。

- 利用人工智能扩大平等机会、促进生产力和可持续经济增长并使人们能够获得新的就业岗位、产业、教育和创新。
- 促进建立在循证办法、分析研究和多利益攸关方参与基础上的国际协作和伙伴关系。
- 确保人工智能的研究和开发是为了帮助人类以合乎道德和负责任的方式采用和使用这些系统。

58. 建立强大的人工智能治理框架还需要：

- **人类问责。**人类构建算法，并在算法基础上根据人工智能系统提供的信息做出决策。因此，人类问责是合乎道德的人工智能办法的关键所在。
- **监管合规。**人工智能法规有助于保护用户数据、确保负责任地使用人工智能。公共和私营组织必须遵守旨在保护信息的数据隐私要求、准确性标准和存储限制。可考虑按照人工智能高级别咨询机构的建议，设立一个人工智能中央监管机构，以确保负责任地使用人工智能，包括在公共部门本身负责任地使用人工智能。[19]
- **风险管理和预测。**人工智能治理应包括有效的风险管理策略，如选择适当的培训数据集、实施网络安全措施、解决人工智能模型中的潜在偏见或错误。此外，还需要加大努力，预测人工智能在未来可能产生的破坏性影响，辅之以灵活的前瞻性监管办法。[20]各国政府还应制定适当的体制机制，预测人工智能带来的风险和机遇。例如，联合王国设立了监管视野理事会，这是一个独立的专家委员会，负责确定技术创新的影响。
- **健全的监督和监测机制。**应建立有效的监督和监测机制，确保安全和负责任地使用人工智能。应该对每种算法进行上游和下游影响评估，包括事前社会和伦理影响评估。[21]
- **决策和可解释性。**人工智能系统的决策能力应该公正客观。为了培养责任感和信任，合理化或有能力理解人工智能输出的背后原因至关重要。
- **利益攸关方的参与。**有效治理人工智能需要利益攸关方参与决策和监督，以确保人工智能技术得到负责任的开发和使用。

59. 人工智能治理的未来将取决于所有会员国和其他利益攸关方之间的合作。其成功与否则取决于能否制定全面的人工智能政策和法规，这些政策和法规在促进创新、弥合法律框架在人工智能问责、公平、透明和诚信方面所存在差距的同时保护公众。这很可能也会影响未来对生物技术和神经技术等其他新兴技术的监管。

[19] 另见人工智能高级别咨询机构 2023 年中期报告。

[20] 例如见经合组织，“理事会关于开展灵活监管治理以驾驭创新的建议”，2021 年 10 月。

[21] 另见联合国教育、科学及文化组织，2021 年 11 月 23 日通过的《人工智能伦理问题建议书》，2022 年。

六. 结论和建议

60. 传统人工智能和生成式人工智能是人工智能领域的两种不同方式。虽然生成式人工智能的优势在于创造力、处理不确定性和新颖的应用程序，但传统人工智能在效率、可解释性和特定任务解决方面表现出色。这两种方式各有优势和局限性，它们在人工智能领域的未来潜力巨大，可以带来突破性的进步和变革性的应用。

61. 尽管人工智能有多项惠益，但从长远来看，也可能带来诸多风险和挑战。在一个技术时代即将到来之际，为了确保人工智能促进落实《2030 年议程》，迫切需要开展全球合作、具备战略远见、坚定不移地致力于确保公平分配人工智能的惠益，同时解决其潜在的负面外部效应。只有制定积极主动、具有包容性的决策路线图，才能实现这一变革性技术的潜力。

62. 人工智能对公共部门本身具有特殊意义，影响着政策设计和政府服务提供的方式以及决策的透明度。展望未来，需要进一步考虑在公共部门负责任和合乎道德地使用人工智能。

63. 需要对人工智能进行治理，其中应包括一个法律框架，确保人工智能技术的研究和开发目标是帮助人类以合乎道德和负责任的方式采纳和使用这些系统。人工智能治理应旨在弥合技术进步中问责、透明度、道德和诚信方面的差距。

64. 各国政府、联合国系统和其他利益攸关方正在进行的努力(包括 2024 年未来峰会可能通过的全球数字契约)应继续进行，同时应促进全球对话，以建立监管人工智能所需的证据基础，确保人工智能符合全球共同价值观，促进落实《2030 年议程》，不让任何人掉队。巴西将于 2024 年担任二十国集团轮值主席国，这可能是又一次重要机会，有助于推动旨在实现公正数字化过渡、负责任使用人工智能的全球议程，从而促进更大的社会包容、减少国家之间和国家内部的不平等。2024 年是一场更加公平的数字革命的关键之年，其将更多地纳入弱势群体和发展中国家。