联合国 A/79/170

大 会

Distr.: General
17 July 2024
Chinese
Original: English

第七十九届会议

临时议程* 项目 71(b)

促进和保护人权：人权问题，包括改善人权和基本自由切实享受的各种途径

人权与国际团结

秘书长的说明

秘书长谨向大会转递人权与国际团结独立专家塞西莉亚·贝利艾特根据人权理事会第 53/5 号决议提交的报告。

* A/79/150。

24-13137 (C) 310724 140824

人权与国际团结独立专家塞西莉亚·贝利艾特的报告

人工智能与国际团结——以人为本的人工智能国际团结设计

摘要

本报告是人权与国际团结独立专家塞西莉亚·贝利艾特为大会编写的第一份报告。本报告依照人权理事会第 53/5 号决议提交。独立专家讨论了当前全球不平等加剧的挑战，并解释了人工智能国际团结设计战略如何能够阐明国家、公司和民间社会加强平等获得技术、非歧视地包容弱势群体和个人的义务。独立专家在报告中建议，人工智能治理应支持国家和企业尽职调查程序机制，让直接和间接利益攸关方参与人工智能生命周期中的数据处理和决策。

一. 导言

1. 国际团结是国际法的一项原则，也是一项普遍价值。它吸引了那些关心解决大流行病、气候变化、不平等和清洁能源等全球挑战的政策制定者的注意力。[1] 世界各地正在出现一种新的意识，意在从新自由主义的不平等转向由基于价值观的革命所驱动的管理、团结和集体行动。[2] 人工智能必须用于团结人类而不是分裂人类。根据人权原则，人工智能应促进跨国界的集体的国际团结行动。[3] 2022 年，联合国教育、科学及文化组织(教科文组织)发布了《人工智能伦理问题建议书》。该建议书认识到需要加强团结，以促进公平获取人工智能技术，并应对人工智能技术给文化和伦理体系的多样性和互联性带来的挑战。此外，团结可以支持减轻人工智能的潜在滥用，帮助实现人工智能可以带来的全部潜力，并帮助确保国家人工智能战略以伦理原则为指导。

2. 有一项研究指出，人工智能正越来越多地融入社会，包括创建“智能城市”，以促进对公民生活各个方面的跟踪和监控，增加生物特征登记的使用，并作为国际发展和应对流离失所和其他危机的人道主义行动的工具。[4] 政府机构和外包机构(包括公司和民间社会组织)使用人工智能来处理和进行利用个人数据的案件和索赔的决策，预计将在全球范围内显著增加。人工智能监控特别影响穷人，因为国家机构利用人工智能作为“过度监管”边缘化社区的工具，并在跟踪非正常移民的背景下，来查明滥用社会福利的行为。[5] 人工智能存在跨越各部门的歧视，包括基于种族、民族、宗教、性别、地点、国籍和社会经济地位的歧视，而认识到这一点弱点，十分重要。[6]

3. 此外，《2024 年人工智能指数报告》的结论是：“人工智能开发人员缺乏透明度，特别是在披露训练数据和方法方面。这种缺乏开放性的现象阻碍了进一步了解人工智能系统的稳健性和安全性的努力”。[7] 此外，信息和通信技术(信通技术)公司更多地与来自全球北方的民间社会组织接触，这可能会削弱与全球南方民间社会组织的团结。[8] 由于全球南方团体参与解决人工智能对侵犯人权行为问责网络的机会有限，解决这一不平衡的努力变得错综复杂。

---

[1] 见全球民族，《2023 年全球团结报告》(2023 年 9 月)。

[2] Kurt April，“人工智能引发的团结经济：需要管理指导”，《有效执行》，第 26 卷，第 3 号(2023 年)。

[3] 互联网治理论坛提供的资料。另见 https://intgovforum.org/en/content/pnai-report。

[4] Linnet Taylor，“什么是数据正义？全球连接数字权利和自由的案例”，《大数据与社会》，第 4 卷，第 2 号(2017 年 7 月至 12 月)。

[5] 同上。

[6] 同上。

[7] 斯坦福大学，以人为本人工智能研究所，《2024 年人工智能指数报告》(斯坦福大学，加利福尼亚州，2024 年)

[8] 见 https://www.business-humanrights.org/en/from-us/briefings/dismantling-the-facade-a-global-south-perspective-on-the-state-of-engagement-with-tech-companies/dismantling-the-facade-a-global-south-perspective-on-the-state-of-engagement-with-tech-companie/。

4. 在编写本报告时，独立专家征求了会员国、民间社会组织、公司和学术界的意见。2024 年 7 月举行了民间社会利益攸关方的磋商。审查了学术文献、联合国条约和宪章机构的报告，并分析了国际、区域和国家法律标准。

二. 负责任的人工智能团结标准：透明、公平、非歧视和包容

5. 鉴于人工智能越来越多地被授予数据处理和决策权，因此从根本上需要透明度，以确保公平、非歧视和包容。《2024 年人工智能指数报告》将公平定义为：“创建公平的算法，避免偏见或歧视，并考虑所有利益攸关方的不同需求和情况，从而与更广泛的社会公平标准保持一致。”[9] 该报告进一步指出：“结果显示，虽然大多数公司已经全面实施了至少一项公平措施，但仍然缺乏全面的整合。在作为全球负责任人工智能调查的一部分收集数据的五项措施中，采用公平措施的全球平均值为 1.97。世界基准联盟指出，截至 2023 年，全球 200 家最具影响力的科技公司中，只有四分之一达到了在数字包容方面采用人工智能伦理原则的最低披露标准。[10]

6. 《罗马人工智能伦理呼吁》强调包容是一项关键原则，而在 2024 年广岛人工智能伦理促进和平活动上，阿布扎比和平论坛主席谢赫·阿卜杜拉·本·巴耶(Shaykh Abdallah Bin Bayyah)强调，“由于人工智能的利益、危害和好处混在一起，合作、团结和共同努力对于处理人工智能的发展是必要的，以确保其系统和产品不仅在技术上先进，而且在道德上也是合理的”。[11] 2024 年 6 月 26 日的《全球数字契约》草案提出了一个专门的包容性目标：“我们的合作将会缩小国家内部和国家之间的数字鸿沟，并推进一个促进和实现多样性的数字环境”。[12]

7. 根据《工商企业与人权指导原则》，科技公司应进行人权影响评估，并将风险识别作为其质量控制流程的一部分，并与外部利益攸关方接触，作为人权评估的一部分。[13] 可以建议，人工智能国际团结的方法将侧重于需要具体建立程序性方法，以解决歧视和加强包容。建议将代际团结纳入人工智能法规。[14]

---

[9] 《2024 年人工智能指数报告》

[10] 见 https://www.worldbenchmarkingalliance.org/impact/investor-statement-for-ethical-ai-2024/。

[11] 见 https://www.romecall.org/和 https://www.romecall.org/ai-ethics-for-peace-hiroshima-july-9th-2024/。

[12] 见 6 月 26 日草案，可查阅 https://www.un.org/techenvoy/global-digital-compact。

[13] Kate Jones，“人工智能治理与人权：重置关系”，研究论文，《国际法方案》(伦敦，皇家国际事务研究所，2023 年 1 月 10 日)。

[14] Sébastien Fassiaux，“通过欧盟对人工智能的监管保护消费者自主权：长期方法”，《欧洲风险监管学刊》，第 14 卷，特刊第 4 号(2023 年 12 月)。另见 Jon Truby 等人，“监管高风险人工智能应用的沙箱方法”，《欧洲风险监管学刊》，第 13 卷，第 2 期(2022 年 6 月)，承认团结是欧洲联盟人工智能监管的一项原则。

8. 社会团结可以用来分享人工智能部署的益处和成本，促进人工智能发展轨迹的多样性，并促进透明度和合规性，以纠正人工智能信息不对称。[15] 独立专家强调，在评估人工智能的设计和实施时，必须将团结作为一个视角。

A. 作为人工智能团结目标的包容和算法不歧视

9. 监管人工智能以纠正偏见和歧视并确保安全的必要性至关重要。人工智能可以通过促进偏远地区教育而具备解放力，可以通过翻译功能提供语言便利，可以用来打击陈规定型观念和仇恨言论，因此仍然需要支持研究人工智能的社会影响。在决策过程(包括司法和行政领域)中设计人工智能的各机构，应该从规划到应用的所有阶段寻求弱势群体和以民主为导向的民间行为体的意见，以防止侵犯人权和减轻伤害。此外，必须建立一个独立的监督机制来解决数据保护问题，以便在规划和部署人工智能的不同阶段规范和发布有关收集和处理个人数据的指导方针。独立专家支持有条件人口差异模型，可作为自动歧视案件的一套标准统计证据。[16]

10. 一份提交给本报告的文件提出了一个包容性的人工智能愿景，通过创建平台来扩大传统上代表性不足的人口在全球对话中的观点，例如人工智能翻译、活动家为联络而使用的社交媒体平台，以及赋予民间社会权力以追究政府的责任，例如在腐败案件中。[17]

B. 关于包容和不歧视的国家标准

11. 在国家一级，有许多标准(大多数是愿望性的、不具约束力的或草案形式)宣布恪守平等、包容和不歧视原则，但往往缺乏单独的机构能力和具体的程序机制来确保遵守并为违反行为提供补救措施。

12. 澳大利亚有自愿、不具约束力的人工智能伦理原则，承认人工智能系统应有利于个人、社会和环境，并尊重人权、多样性和个人自主权。[18] 同样，中国有生成式人工智能服务管理临时措施，要求防止基于民族、宗教、国籍、地域、性别、年龄、职业和健康状况的歧视，以及采取措施增加培训数据的多样性和非歧视内容。[19] 印度有一项国家人工智能战略，解决平等和非歧视问题。[20] 该战略规定，人工智能系统必须平等对待处于与决定相关的相同情况下的个人，

---

[15] Juan C. Mateos-García，《人工智能的复杂经济学》(2018 年 12 月 2 日)(未经发表的工作论文)。

[16] Sandra Wachter，Brent Mittelstadt 和 Chris Russell，“为什么公平不能自动化：弥合欧盟非歧视法和人工智能之间的差距”，《计算机法律和安全评论》，第 41 卷(2021 年)。

[17] 加拿大多伦多城市大学林肯-亚历山大法学院助理教授 Jake Okechukwu Effoduh 提供的资料。

[18] 见 https://www.industry.gov.au/publications/australias-artificial-intelligence-ethics-framework/australias-ai-ethics-principles。

[19] https://www.chinalawtranslate.com/en/generative-ai-interim/.

[20] 见 https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf；https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf；https://www.niti.gov.in/sites/default/files/2021-08/Part2-Responsible-AI-12082021.pdf。

人工智能系统不应根据其身份拒绝合格人员接受教育、就业或进入公共场所机会，并且不应加深基于宗教、种族、种姓、性别、血统、出生地或居住地的有害的历史和社会分裂。此外，该战略还提出了一项目标，即应努力防止被不公平地排除在服务或福利之外。巴西提出了一项关于人工智能法规的提案，其中载有一项非歧视标准，包括质疑决定和要求人为干预的权利，以及纠正直接、间接、非法或滥用歧视性偏见的权利。[21] 西班牙建立了一个独立的平等待遇和不歧视机构，负责监督和促进使用符合伦理、值得信赖和尊重基本权利的人工智能。[22]

13. 美利坚合众国有一份“人工智能权利法案蓝图：让自动化系统服务于美国人民”，支持一项行政命令，要求进行独立评估和报告，以确保不歧视。[23] 肯尼亚有一份人工智能业务守则草案，确切要求建立一个执行不歧视标准的机制，包括通过记录公平性评估、记录为解决偏见而采取的步骤以及记录反对歧视性结果的政策。[24]

C. 带有偏见的人工智能应用的风险

14. 人们认识到，“以负担得起的价格和有知识的方式使用互联网已成为充分实现所有人权和基本自由、民主、发展和社会正义的基本需要”。[25] 这导致了一系列私人数字权利倡议，从增加可访问性的角度寻求解决互联网基础设施、应用和使用问题。人工智能偏见通过使用人工智能算法将个人或群体与他人进行比较，使其受到刻板印象或偏见的影响。人工智能偏见出现在人工智能系统的设计、数据的收集和解释以及直接和间接的利益攸关方互动中。当代形式种族主义、种族歧视、仇外心理和相关不容忍行为特别报告员确认了“数据问题、算法设计问题、故意歧视性地使用人工智能和问责问题”的持续挑战(见A/HRC/56/68)。教科文组织 2019 年关于人工智能的报告将团结视为创建“知识社会”的一个要素，并解释了解决有意或无意的歧视性编程、机器学习算法训练数据的偏见或其他因素导致的人工智能歧视的紧迫性：[26]

> 许多类型的歧视可能是间接的；例如，如果某算法通过某人手机的使用模式来确定其信贷价值，认定以下情况的女性具有较高的信用风险：㈠ 手

[21] 见 2023 年第 2338 号法律草案。可查阅 https://www25.senado.leg.br/web/atividade/materias/-/materia/157233。

[22] 西班牙提供的资料。

[23] 见 https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf；另见 https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/。

[24] 见 https://www.dataguidance.com/sites/default/files/kebs-tc_094_n66_public_review_kenya_standard_dks_3007_ai_code_of_practice.pdf。

[25] George A. Walker，“技术法，权利和道德：一个选择，一个未来”，《国际法学者》，第 56 卷，第 1 期(2023 年)。

[26] 联合国教育、科学及文化组织(教科文组织)，《引领人工智能与先进信息传播技术构建知识型社会：权利、开放、可及、多方的视角》(巴黎，2019 年)，第 63 页。

机使用率较低，或㈡ 没有手机，则该算法具有歧视性。算法所适用的条件可能看起来是平等和公平的，但它对特定群体不利。算法可能导致并加剧以上多种形式的歧视。现有的社会和政治偏见正以多种方式在机器学习算法中变得系统化。此外，人工智能带来的新型潜在歧视也值得研究，例如基于统计相关性的排他性，这种统计相关性不一定与社会显著特征相符，但却与个人身份密切相关。

15. 《2024年人工智能指数报告》描述了语言的标记化(将语言分解为组件进行分析)如何对非西方语言产生负面影响，因为人工智能的效率将低于西方语言。[27] 还有一些民间社会组织直接处理算法偏见、排斥和对种族、族裔或宗教少数群体的歧视，以及基于性别或其他身份的歧视，并致力于改善无障碍环境。[28] 大赦国际和全国有色人种协进会对预测性警务工具、[29] 使用自动化系统确定获得医疗保健和社会服务的机会、监视难民和移民流动以及用于面部识别和欺诈检测的人工智能工具对种族化社区的偏见影响表示担忧。[30] 接受这些人工智能程序的人可能不知道也不了解人工智能的使用，他们往往对国家机构的歧视性待遇缺乏任何补救措施。民间社会团体强调，需要将利益攸关方纳入人工智能法规的设计。有人建议，尽管受到批评，但欧洲联盟仍在歧视性监控系统中使用人工智能，如风险评估系统和预测分析，以便利推回，[31] 而在移民方面，正在为不同的目而开发人工智能：(a) 通过开发面部识别技术进行身份验证；(b) 风险评估；(c) 拘留评估；(d) 监控。在这种情况下，强调了人为控制被尽量减少的风险。[32]

16. 鉴于利用人工智能进行数据处理和决策的国家机构和公司在解释其开发、培训、使用或应用方面并不透明，声称侵犯人权的个人或团体往往无法提供将人工智能与侵权行为联系起来的证据。侵权行为首先可能表现为通过人工智能连接的不同国家系统之间的快速反馈循环，导致国家机构的有害做法或决策。犯罪分子还利用人工智能在原籍国和居住国欺诈性地挪用移民给家人的汇款，或家人汇给移民的钱。

---

[27] 《2024 年人工智能指数报告》。

[28] 见 https://digitalrightsfoundation.pk/wp-content/uploads/2021/03/Policy-Upload-2.pdf；https://www.hrw.org/report/2023/12/21/metas-broken-promises/systemic-censorship-palestine-content-instagram-and；https://www.apc.org/en/member/7amleh-arab-center-social-media-advancement；https://smex.org/；https://www.derechosdigitales.org/；见 https://www.accessnow.org/keepiton。

[29] 关于预测性警务导致对黑人社区的不成比例的监视和警务，见：https://naacp.org/resources/artificial-intelligence-predictive-policing-issue-brief。

[30] 见 https://www.amnesty.org/en/latest/campaigns/2024/01/the-urgent-but-difficult-task-of-regulating-artificial-intelligence/。

[31] 格拉斯哥卡利多尼亚大学 Indira Boutier 提供的资料。

[32] 同上，另见 Petra Molnar 和 Lex Gill，《门口的机器人：加拿大移民和难民系统自动决策的人权分析》，(国际人权课程(多伦多大学法学院)和公民实验室（多伦多大学蒙克全球事务和公共政策学院)，2018 年)，第 31-34 页。

17. 大会已认识到有必要打击算法歧视，[33] 查明脆弱性，改善无障碍环境，并为侵犯人权行为提供补救措施。[34] 联合国人权条约机构和特别程序注意到，在负责处理移民、犯罪、保健和老年人护理问题的国家机构内都使用人工智能。消除种族歧视委员会发表了结论性意见，其中对在庇护背景下使用人工智能的歧视性影响表示关切。[35] 消除对妇女歧视委员会发表结论性意见，呼吁制定适当的保障措施，防止执法部门在预防和调查犯罪中使用的生物识别、监控和算法特征分析系统产生性别成见，并采取措施消除与人工智能和算法服务有关的算法偏见。[36] 老年人享有所有人权问题独立专家对访问期间提供的机会表示赞赏，以了解人工智能如何在老龄化、护理和医疗服务方面发挥作用，但她也建议使用数据收集进行审查，以确保此类数据的使用能够维护与不同背景的老年人相关的不歧视义务，从而暗示人了工智能团结的观点。[37]

18. 民间社会团体强调，人工智能对男女同性恋、双性恋、跨性别者、性别奇异者和间性者等积极分子参与团结行动的能力产生了负面影响。

三. 解决机构人工智能进程的区域倡议

19. 经济合作与发展组织(经合组织)于 2024 年更新了人工智能原则，强调了必须促进包容代表性不足的人群，减少经济、社会、性别和其他不平等现象，必须促进多样性、公平、社会正义、透明度、人类代理和监督。[38] 国家机构越来越有兴趣使用人工智能进行案件处理和决策，但问题是，许多这些系统是由公司创建的，因此存在缺乏透明度的风险。国家机构可能没有意识到人工智能问题的风险，例如“幻觉”，即人工智能创建并非基于内容的内容，以及必须有人为控制以监控信息使用并识别算法偏见。经合组织的可信人工智能工具和指标目录提供了各种技术、教育和程序方法的例子，以追求人工智能团结。[39]

20. 《欧洲联盟人工智能法》第 5(1)(b)、(c)、(g)条通过确定禁止的人工智能做法来处理国际团结问题。欧盟委员会提出了一项旨在由各国法院实施的责任指

[33] 大会第 78/265 号决议，第 6(h)段。

[34] 大会第 78/213 号决议。

[35] CERD/C/DEU/CO/23-26，第 45-46 段。

[36] CEDAW/C/ITA/CO/8，第 26 段。另见 CEDAW/C/DEU/CO/9 号文件中类似的结论性意见，第 27-28 段。委员会在 CEDAW/C/ESP/CO/9 号文件第 21、23、31 和 33 段所载结论性意见中对人工智能和性别暴力表示关切。

[37] A/HRC/45/14/Add.1，第 93 段。

[38] 见 https://oecd.ai/en/ai-principles；另见 https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449。

[39] 参见 https://oecd.ai/en/。

令，因此它规定了证据标准。[40] 欧盟委员会于 2018 年通过了关于在司法系统及其环境中使用人工智能的《欧洲伦理宪章》，其中呼吁对数据处理进行外部审计，以筛查歧视。[41] 欧盟委员会认识到，在移民、庇护和边境控制管理以及司法和民主程序中使用人工智能具有很高的风险，因此在 2024 年通过了《人工智能法》。该法呼吁在委员会内设立欧洲人工智能办公室，以监督通用人工智能模型提供商有效实施和遵守该法的情况。欧洲人工智能委员会将由成员国和各公司的代表组成。人工智能办公室可以邀请通用人工智能模型提供商和相关国家主管部门参与制定行为守则，而民间社会、工业界、学术界、下游提供商和独立专家可以支持这一过程。[42] 这一标准在民间社会直接参与影响评估的能力方面显得薄弱。

21. 欧洲法院在一份战略文件中指出了未来使用人工智能的歧视风险：“采用人工智能技术的主要风险之一是在人工智能模型的训练过程中引入非自愿偏见的可能性，造成无意的歧视”。[43] 欧洲议会人权小组委员会对人工智能进行了一项分析，强调了结构性歧视的风险：“人工智能系统可以使偏见永久化并扩大偏见，导致包括就业、执法和信用评分在内的各个部门的歧视。有大量证据证明，人工智能可以通过反映其训练数据或设计中存在的偏见来巩固社会经济差异。[44]

22. 欧洲委员会部长理事会于 2024 年 5 月 17 日通过了《人工智能与人权、民主和法治框架公约》，将于 2024 年 9 月开放供签署。《框架公约》关于平等和不歧视的第 10 条提出了采用生命周期办法系统地查明和纠正偏见的想法；应根据适用法律在整个使用期间审查人工智能系统的歧视。[45]

1. 各缔约方应采取或维持措施，以确保人工智能系统生命周期内的活动尊重平等，包括性别平等，并根据适用的国际和国内法律禁止歧视；

2. 各缔约方承诺采取或维持旨在克服不平等的措施，在涉及人工智能系统生命周期内的活动中实现公平、公正的结果，符合其适用的国内和国际人权义务。

---

[40] 见 https://artificialintelligenceact.eu/the-act/。另见欧洲议会和理事会关于使非合同民事责任规则适用于人工智能的指令提案(人工智能责任指令)。可在 https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022PC0496。

[41] 见 https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment。

[42] 见 https://artificialintelligenceact.eu/high-level-summary/。

[43] 见 https://curia.europa.eu/jcms/upload/docs/application/pdf/2023-11/cjeu_ai_strategy.pdf。

[44] 见 https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf；另见 Ahmet Bilal Aytekin，“欧盟反歧视指令背景下的算法偏见”，2023 年 6 月 7 日至 9 日在瑞士温特图尔举行的欧洲算法公平问题研讨会上提交的论文；可查阅 https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-bias-in-algorithms_en.pdf。

[45] 见 https://rm.coe.int/1680afae3c。

23. 还有其他促进数字包容的区域举措，如非洲人权和民族权委员会通过的关于非洲互联网关闭和选举的第 580(LXXVIII)2024 号决议。[46] 2023 年 11 月，美洲人权法院与法官举行对话，讨论了教科文组织“关于人工智能和司法法治的全球工具包”。工具包介绍了排斥问题：

> 问题是，[.....]数据可能会被注入偏见[.....]。例如，临床试验往往排除妇女和有色人种，导致数据代表性不足。如果使用这些数据训练的算法被用于分析皮肤图像或优先考虑患者护理，这就可能产生严重的后果。因此，确保人工智能算法使用代表性数据进行训练，以避免此类偏见并确保所有人的公平结果至关重要。[47]

24. 该工具包还解决了人工智能代表性问题：

> 许多全球南方国家的数字鸿沟导致了“数据不可见性”，而这可能会影响到历史上被边缘化的群体，如妇女、种姓、部落族群、宗教和语言少数群体以及移民劳工。基于现成数据开发的人工智能算法的有用性和有效性可能会受到数据不可见性所带来的偏见的限制。这强调了算法透明度和问责制的要求。

工具包强调了代理歧视问题，例如银行在评估可能与少数民族或少数种族有关的贷款申请时使用邮政编码、教育水平或收入。这种方法本身就有可能使交叉歧视政策和做法长期存在下去。它表明，人工智能可能有助于确定包括司法机构在内的公共和私人行为体对边缘化群体和个人的偏见模式。

25. 美洲人权法院与法官的对话还讨论了教科文组织“人工智能伦理影响评估工具”。工具呼吁考虑“人工智能项目团队的多样性，特别是在性别、年龄、种族、肤色、血统、语言、宗教、民族血统、族裔血统、社会出身、经济或社会状况、残疾和性取向方面，但不限于此，包括这如何反映预期用户群体的复杂性和多样性，以及这如何可能引入偏见”。[48]

A. 团结促进数字素养以打击虚假信息

26. 越来越多的人呼吁对“极端分子引发的选举误导运动”进行监管。[49] 人工智能生成的内容质量的提高使人们难以识别虚假信息和虚假内容，并且有越来越多的国家关闭互联网的报告，这对信息的获取产生了不成比例的影响。有必要制定一项国际团结战略，纳入立法、技术和教育合作，以提高数字素养水平。[50] 响应式团结被用作处理与恶意使用人工智能有关的紧急情况的一种手

[46] 见 https://achpr.au.int/en/adopted-resolutions/580-internet-shutdowns-elections-africa-achprres580-lxxvii。

[47] 见 https://nataliazuazo.com/2023/11/20/ai-and-the-rule-of-law-at-the-inter-american-court-of-human-rights/。另见 https://unesdoc.unesco.org/ark:/48223/pf0000387331。

[48] 见 https://unesdoc.unesco.org/ark:/48223/pf0000386276。

[49] Jake Okechukwu Effoduh 提供的资料。

[50] 同上。

段，例如“在一个有灭绝种族历史的国家的选举日，成千上万的带有种族暴力的深度虚假视频流传”。[51]

27. 欧洲联盟在 2022 年通过了《数字服务法》。该法于 2024 年生效，要求大型在线平台和搜索引擎使用行为准则，包括在适用中对虚假信息(包括虚假问题和政治广告)实施严格必要和有针对性的对策。[52] 该法支持《2022 年反虚假信息强化行为守则》(适用于自我监管实体)，其中包括关于禁止传播虚假信息、保证政治广告的透明度、加强与事实核查人员的合作以及便利研究人员获取数据的指导。[53] 然而有人批评说，“虽然这些发展可能具有统一效果，但目前欧洲大陆对虚假信息的处理方法是支离破碎的”，[54] 此外，人们担心行为守则与欧洲人权法院的裁决之间可能存在冲突。[55]

28. 欧洲联盟《人工智能法》也包含了遏制深度造假的影响和冲击的规定。根据第 50 条第(4)款，部署者，即使用人工智能系统的人，通常必须披露某内容由人工智能生成。然而，法律本身包含了这一披露义务的例外情况，意味着必须对其有效性进行审查。

29. 非洲联盟发展署呼吁建立法律保护和监管框架，以打击算法歧视。[56] 非洲联盟执行委员会支持根据执行委员会通过的人工智能概念框架制定非洲大陆战略。[57] 在该区域，有一种趋势是对“技术内容”进行监管，包括审查或阻止内容的获取，以防止假新闻和(或)仇恨言论并对“法律内容”进行监管，包括讨论、起草和通过法案，以监管假新闻和仇恨言论。这两项措施都有适用范围过广或针对政治反对派从而削弱选举的独立性的风险。

30. 沙特阿拉伯向独立专家通报了沙特数据和人工智能管理局的创建情况。这是阿拉伯语大语言模型(称为 ALLaM)的一个重大发展。该举措旨在利用人工智能技术和数字应用促进文化多样性，造福全人类。其中一个项目 SauTech，专注于将人工智能技术本地化和保护当地文化，能够识别各种阿拉伯方言的语音并

---

[51] Miguel Luengo-Oroz，“团结应成为人工智能的核心伦理原则”，《自然机器智能》，第 1 卷，第 494 页(2019 年 11 月)。另见 Patrik Hummel 和 Matthias Braun，“公正数据？数据驱动医学中的团结与正义”，《生命科学、社会与政策》，第 16 卷，第 8 号(2020 年)。

[52] 欧洲议会和欧盟理事会 2022 年 10 月 19 日关于单一数字服务市场的第(EU)2022/2065 号条例，并修正了第 2000/31/EC 号指令(《数字服务法》)。

[53] 见 https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation。

[54] 华盛顿哥伦比亚特区，乔治·华盛顿大学法学院国际法和比较法方案人权研究员 Dorothy Andersen 提供的资料。

[55] Paolo Cavaliere，“假新闻中的真相：虚假信息法如何在数字平台上重构真相和准确性概念”，《欧洲人权法公约评论》，第 3 卷，第 4 号(2022 年 11 月)。另见 Ethan Shattock，“斯特拉斯堡的假新闻：选举假信息和表达自由”，《欧洲法律与技术学刊》，第 13 卷，第 1 号(2022 年)。另见 https://edmo.eu/wpcontent/uploads/2022/01/Case-law-for-policy-making-Report-2022.pdf。

[56] 见 https://www.nepad.org/publication/ai-and-future-of-work-africa-white-paper。

[57] 见 https://au.int/en/newsevents/20240419/multistakeholder-consultative-sessions-development-continental-strategy。

将其转换为文本，以便找到适合当地背景的解决办法。[58] 沙特数据和人工智能管理局通过培训方案提高人们对人工智能的认识。同样，哥伦比亚提供的资料介绍了不同级别的人工智能培训。[59]

B. 人工智能中的性别团结

31. 人们也越来越关注国际团结方法在打击与人工智能有关的性别歧视方面的潜力。[60] 《2024 年人工智能指数报告》发现，欧洲国家报告的信息技术相关专业毕业生中，男子多于女性，在缩小性别差距方面进展缓慢。[61] 消除对妇女歧视委员会对一些缔约国的报告发表了结论性意见，其中对人工智能领域的女童和妇女代表性不足以及缺乏具体措施来防止将性别陈规定型观念和算法性别歧视和暴力纳入人工智能编码表示关切，并呼吁建立有效的保障措施。[62]

32. 沙特阿拉伯在其提供的资料中报告说，妇女占沙特数据和人工智能管理局技术人员的 50%以上。[63] 该国的人工智能研究与伦理问题国际中心与谷歌云合作，推出了 Elevate 倡议，旨在培训 25 000 名数据和人工智能领域的妇女。秘鲁民主和人权研究所报告说，秘鲁总统支持了一项用人工智能制定的商业项目培训方案。自 2024 年 5 月以来登记的公民有机会申请学习奖学金，其中 50%提供给处境脆弱的妇女。[64]

33. 企业与人权资源中心描述了目前人工智能促进针对妇女的反团结挑战：[65]

> 妇女和非常规性别者继续感受到设计、开发和部署不当技术所造成的不成比例的影响。这一事实证明，最新的技术工具，包括生成式人工智能，都带有缺陷和偏见，造成了危害，包括放大了性别歧视和性别陈规定型观念，促进了对妇女的社会控制，剥夺了妇女患者的权力，歧视女性求职者，助长了针对妇女人权维护者的攻击，使妇女在获得福利方面处于不利地位。

34. 妇女越来越多地受到人工智能仇恨言论、网络霸凌、非自愿的性内容、报复性色情和跟踪的影响，而所有这些都对她们实现个人和职业抱负的能力产生

[58] 沙特阿拉伯提供的资料。

[59] 哥伦比亚提供的资料。

[60] Keratso Georgiadou，“人工智能时代的团结、性别平等和对话”，载于 *Kritische Pädagogik und Bildungsforschung: Anschlüsse an Paulo Freire*，Wassilios Baros、Rita Braches-Chyrek、Solvejg Jobst 和 Joachim Schroeder 编辑(德国 Wiesbaden，Springer Fachmedien，2024 年)，第 477-489 页。

[61] 《2024 年人工智能指数报告》

[62] 见 CEDAW/C/ITA/CO/8，CEDAW/C/DEU/CO/9，第 27 和 28 段；CEDAW/C/ESP/CO/9，第 21、23、31 和 33 段；CEDAW/C/TJK/CO/7，第 47、48 和 55 段。另见以下文件所载结论性意见：CEDAW/C/TUR/CO/8，第 46 段；CEDAW/C/CRI/CO/8，第 27、28 和 37 段。

[63] 沙特阿拉伯提供的资料。

[64] 秘鲁民主和人权研究所提供的资料。

[65] 见 https://media.business-humanrights.org/media/documents/BHRRC_Submission_Gender_and_Role_of_Business_2023_j2HkLgP.pdf。

了负面影响。民间社会组织介绍了一项旨在结束性别暴力的创新举措，其中包括 2023 年在坎帕拉国际大学举行的数字安全培训。[66] 他们还指出，使用社交媒体至关重要，有助于国际社会团结一致，要求起诉杀害妇女行为。他们介绍了肯尼亚进行的立法改革，宣布杀害妇女和暴力侵害妇女行为为国家紧急状态，并设立一个专门负责处理这些罪行和打破有罪不罚循环的委员会。[67] 他们呼吁加强科技公司、非政府组织和政府之间的国际合作，开发有效和合乎伦理的人工智能内容审核工具，适用于不同的语言和文化背景，以识别和标记针对妇女的在线暴力行为。

C. 人工智能设计中的团结

35.《全球数字契约》第二次修订版宣布，将本着全球团结的精神执行该契约，并特别强调包容，重点是数字技术公司和开发人员需要与各种背景和能力的用户接触，将他们的观点和需求纳入数字技术的生命周期，追求数字利益的公平分配，并支持数字空间中的数字无障碍性和语言及文化多样性。[68] 此外，国际社会呼吁在制定和执行国家和地方数字连通性战略时考虑到处境脆弱者以及服务不足、农村和偏远地区人们的需要，并强调需要为妇女和女童、儿童和青年以及老年人、残疾人和处境脆弱者开展有针对性的能力建设，并确保他们切实参与方案的设计和执行。此外，《契约》还寻求制定和开展国家数字包容性调查，系统地按年龄、残疾和性别分列数据，以查明学习差距，并通报具体情况下的优先事项。

36.《全球数字契约》支持采取国家人工智能立法以及有效的监督和补救机制。《契约》还呼吁科技公司和人工智能开发人员认识到尊重人权的义务，并实施尽职调查和影响评估。《契约》还呼吁科技公司和人工智能开发人员(与政府和其他利益攸关方协商)共同开发行业问责制框架，以提高其系统和流程的透明度，明确责任，并致力于标准和可审计的公共报告。《契约》呼吁社交媒体平台建立安全、可靠和可用的报告机制，供用户及其倡导者报告违反政策的行为(包括适合儿童和残疾人的特别报告机制)。它主张人工智能技术公司和社交媒体平台加强其系统的透明度和问责制，包括其服务条款、内容审核和算法以及以当地语言处理用户个人数据，以使用户能够做出知情选择并提供或撤回知情同意。

37. 教科文组织强调在响应式人工智能利益攸关方机制的设计中纳入下列因素，将之作为透明度的一个要素：[69]

- 政府是否鼓励其他利益攸关方参与国家人工智能治理？
- 人工智能专业人士、消费者和其他利益攸关方社区是否有活跃的协会？

[66] “我们的声音，我们的未来”联盟提供的资料。

[67] 同上。

[68] 可查阅 https://www.un.org/techenvoy/global-digital-compact。

[69] 教科文组织，《指导人工智能和先进的信通技术的知识社会》。

- 政府是否积极让其他利益攸关方参与制定全球人工智能治理政策？

《契约》呼吁实施指标，通过衡量不同利益攸关方群体(妇女和性别多元化者、青年和边缘化群体)参与人工智能政策制定以及参与国际和区域人工智能论坛、进程和机制的程度，确保包容性。

38. 有人认为：“社会技术系统中的人工智能技术可能有助于促进或塑造社会互动，因此可能破坏或促进团结。”[70] 注意到利益攸关方的相关性可能会随着时间的推移而改变，可能会出现不可预见的团结问题，因此将他们纳入人工智能设计的决策中至关重要。人权是衡量与人工智能有关的团结的尺度。[71] 人们可能会认为，人工智能的使用可能会导致由于面部识别而侵犯隐私权的情况，或违反招聘中对妇女或少数民族的不歧视，或拒绝或自我实现生活目标(或生活项目)，例如拒绝住房或教育贷款的申请。人权影响评估应被视为人工智能团结设计的核心要素。此外，在设计过程中应确定并考虑到用户的利益和需要。参与式设计将邀请用户讨论设计选项并参与决策过程，其中应包括能够使群体内团结一致的集体决策。然而，仅仅关注用户并没有认识到间接的利益攸关方，例如那些无意中受到影响的人。人工智能团结的观点需要超越对用户的关注，以关注社会技术系统中所有受影响的利益攸关方。这不仅需要评估总体的危害和益处，而且还需要评估这些危害和益处如何在社会和利益攸关方群体中分布。根本问题是如何让科技公司接受团结的义务，以及在分享益处的同时分担风险。人工智能的团结方法将进行人权风险评估，确定个人自我实现的权利，并规范伤害责任的分配，以及数据处理和决策的益处的分配。[72] 因此，支持集体决策的政治团结以及基于团结的人工智能治理框架也是应对人工智能影响所必需的。

39. 人工智能的国际团结方法可能需要在不同领域(如卫生或能源)制定法规。人工智能设计和开发中的对话应该促进用户、人工智能技术开发人员和其他利益攸关方之间的合作，以找到人工智能解决办法。有必要创建参与途径，以支持与人工智能用例开发人员、人工智能解决办法开发人员或任何其他适用的人工智能/机器学习开发人员的对话，以实现持续更新，以应对人工智能解决办法使用的不断变化的风险和益处。建立基于风险的、量身定制的沟通和参与计划，使客户能够轻松理解有关人工智能解决办法如何开发、其性能和维护以及如何与最新的最佳实践和监管要求保持一致的说明。

40. 承认和尊重应被视为数据团结的要素，使“在数据密集背景下的具体边缘化经历和不公正经历具有概念地位”。[73]团结作为人工智能的原则，应该意味着可持续的公平影响，以分享繁荣和负担，并防止不平等。[74] 有一个学术建议是

[70] Catharina Rudschies，“在人工智能背景下探索团结的概念：数字社会设计方法的伦理”，《数字社会》，第 2 卷，第 1 号(2023 年)。

[71] 同上，第 12 页。

[72] 同上，第 13 页。

[73] Hummel 和 Braun，“公正数据？数据驱动医学中的团结和正义”。

[74] Luengo-Oroz，“团结应成为人工智能的核心伦理原则”。

在防止不平等的激励措施之上建立一个总体的人工智能团结框架，例如每次使用人工智能系统时支付版税，使用人工智能模型进行诊断的医生获得奖励，或者每次机器人撰写公开文章时为人工智能自动文本生成器制作文本的人可获得报酬。[75] 其他人则呼吁采取团结的方法，通过倡导个人控制正在处理的数据的保障措施来衡量对社会的外部影响。[76] 另一个观点是建立一个数据团结框架，使数据进程对公共利益可见。[77] 他们认为，这可以促使企业和公共数据利益攸关方分担数据访问、生产和共享的风险和益处。数据团结是为了提高数据集进程的可见度，以查明和纠正基于边缘化的歧视。[78] 他们呼吁采取集体行动，利用团结作为创建公共数据集的数据治理原则，以开始建立信任和问责制。有一种观点认为，国际社会正面临一个转折点，数据团结应促进基础设施的建设，并应根据包容性等民主价值观加以塑造。[79] 此外，这可能需要建立一个独立的全球治理实体，其成员由行业、国家、民间社会、国际组织和学术界的代表组成，以促进基于人权的人工智能规则。[80] 人们可以考虑 Anthropic 公司的“克劳德宪法”，其中除了人权之外还包括鼓励考虑非西方观点的原则。[81]

41. 一些科技公司提供了投入，描述人工智能国际团结的设计的实施：

(a) 培训课程以无障碍、文化敏感和包容的方式设计；

(b) 算法在不同的数据集上进行训练，以代表它们所服务的人；

(c) 定期审计，在不同的人口群体中进行测试，并实施公平意识机器学习方法；

(d)制定符合伦理的人工智能准则，优先考虑算法开发和部署中的公平性、透明度和问责制；

(e) 持续监测和调整人工智能系统以防止歧视；

(f) 与倡导团体、非政府组织和弱势群体的社区代表接触，了解他们的关切，收集关于技术影响的反馈意见；

(g) 对团队进行关于算法公平性和人工智能伦理影响的持续培训；

(h) 制定禁止传播虚假信息、仇恨言论和有害内容的内容审查政策；

[75] 同上。

[76] Hummel 和 Braun，“公正数据？数据驱动医学中的团结和正义”。

[77] Mercedes Bunz 和 Photini Vrikki，“从大数据到民主数据：为什么人工智能的崛起需要数据团结”，载于《民主前线：算法与社会》，Michael Filimowicz 编辑(伦敦，Taylor & Francis 出版社，2022 年)。

[78] 同上。

[79] 同上。

[80]Ana Beduschi，“人权与人工智能治理”，研究简报，(日内瓦国际人道法与人权学院，2020 年)。

[81] 见 https://www.anthropic.com/news/claudes-constitution。

(i) 通过自动化工具和经培训的人工主持人来识别和删除虚假信息；

(j) 与信誉良好的事实核查组织合作；

(k) 在算法中优先考虑透明度，减轻虚假信息的无意放大；

(l) 进行算法设计，促进可靠的来源，减少误导性内容的可见度；

(m) 通过信息活动和平台内通知，教育用户识别和报告虚假信息；

(n) 与政府和非政府机构以及学术机构合作，分享有效打击虚假信息的见解和最佳做法。

四. 面向积极做法的国家指导

42. 西班牙报告说，该国有一个国家人工智能战略和一个监督该战略的机构。[82] 西班牙表示，它将建立一个伦理和监管框架，加强对个人和集体权利的保护，以保障包容和社会福利。西班牙计划制定数字权利宪章，并通过人工智能咨询委员会与数字转型咨询委员会合作，推出人工智能伦理的国家治理模式。2022 年 7 月 12 日第 15/2022 号法律第 23 条规定，必须采取措施减少偏见，同时在公共机构决策中使用人工智能时促进更大的透明度和问责制。这些措施包括分析设计和培训数据，并评估它们是否具有歧视性影响。[83] 此外，该法还寻求在人工智能领域建立对话、提高认识以及国家和国际参与的论坛，以促进政府、科学界、私营部门和民间社会之间的沟通。

43. 沙特阿拉伯报告说，沙特数据和人工智能管理局有一个组织、开发和处理人工智能数据并提供政府服务的系统，通过让利益攸关方参与人工智能的开发和实施，寻求可靠地支持数字化转型和数据保存。[84] 沙特阿拉伯提倡七项人工智能伦理原则，包括：公平；隐私和安全；人道；社会和环境效益；可靠性和安全性；透明度和可解释性；问责制和责任制。[85] 该机构宣布了一项激励计划，以帮助公司自愿遵守人工智能伦理。该过程首先确定和评估所有潜在风险及其影响的严重性。

44. 马来西亚提供的资料介绍了其拟议的监督机制，即人工智能协调和执行单位。该单位将作为处理与人工智能有关的所有事项的政府机构。[86] 马来西亚计划建立一个前瞻委员会，负责进行前景扫描、前瞻和政策倡导。

---

[82] 见 https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/National-Strategy-on-AI.pdf。

[83] 西班牙提供的资料。

[84] 沙特阿拉伯提供的资料。

[85] 同样，哥伦比亚在其意见中指出，尊重透明、隐私、人为控制和不歧视等原则，确保技术具有包容性和公平性。

[86] 马来西亚提供的资料。

45. 德国利用人工智能作为其“匹配”方案的一部分，通过与难民协商，确定他们对住房、就业/专业经验、爱好、娱乐、保健、家庭状况、宗教社区和其他事项的偏好，改善难民融入城市。[87] 这一做法促进了对难民自我实现权利的承认，符合国际团结的做法。人们注意到，“匹配”方案是在设计方案中纳入难民观点的唯一方案。[88]

46. 多米尼加共和国表示，该国制定了一项国家人工智能战略，优先考虑在司法、卫生、教育、环境和安全等关键部门的公共行政中使用人工智能，并将其作为预测分析模型，为公民设计服务。该战略的目标是制定包括预防行动、程序保障和问责机制的守则，以确保负责任地实施人工智能。此外，还将建立监督机制，以核查伦理遵守情况。YoSoyFuturoRD 人力资源和创新中心将优先考虑脆弱部门。它提议创建一个强大且协作的区域人工智能生态系统，为该区域的技术进步、经济和社会发展以及合作做出重大贡献。它表示，它将建立对人工智能系统造成损害的监督、赔偿和追索机制，以保障其公民的权利。

47. 阿根廷 Rawson 市政府利用人工智能促进社会援助方案和激励措施，如“Red de Economía Social y Solidaria”(社会和团结经济网)，在粮食主权框架内建立可持续的粮食循环。它还为罗森市(Municerca)设立了邻里服务中心，负责处理改善公共空间和街道的请求、案件的发起和咨询以及市政索赔的接收。

A. 卫生保健团结和人工智能

48. 卫生保健团结应成为关于使用人工智能的法规的基础，以确保利用和公平，但由于许多国家的卫生保健私有化，实现这一目标变得复杂。[89] 由于种族、阶级、年龄和其他因素而存在的卫生差异可能会在人工智能医疗系统中复制，而这些系统并不能解决结构性不公正问题。用于提供卫生保健的人工智能需要在设计时考虑到社区医疗保健，更具体地说，通过促进低资源环境中的健康保险，将其部署在低资源阶层。一些人提出了人工智能卫生团结的愿景，即通过控制数据流，通过创建可控性基础设施，以及通过专注于治理中的输出导向，使个人能够共享或撤回数据，以防止和减轻不公正。[90] 巴西有一个 Bolsa Família 方案，利用数字化工具向低收入家庭发放援助。该方案通过将现金转移与上学和体检等具体条件挂钩，减少了贫困，促进了人力资本的发展。马来西亚使用人工智能来协助医疗诊断，开发个性化治疗，通过聊天框提供信息，并协助预测分析，以确定积极的措施。[91]

[87] 见 https://matchin-projekt.de/en/。

[88] 见 https://www.rsc.ox.ac.uk/files/files-1/automating-immigration-and-asylum_afar_9-1-23.pdf。

[89] Nicolas Terry，“卫生保健监管中的人工智能与机器”，《耶鲁法律与技术学刊》，第 133 号特刊(2019 年)。

[90] Hummel 和 Braun，“公正数据？数据驱动医学中的团结和正义”。

[91] 马来西亚提供的资料。

49. 沙特阿拉伯介绍了对人工智能卫生保健团结的投资。它有一个卫生部门人工智能卓越中心。沙特专家和工程师开发了 Eyenai，以在地区层面上彻底改变诊断医学。糖尿病视网膜病变是由 1 型和 2 型糖尿病引起的疾病；它是沙特阿拉伯的主要失明原因之一。早期诊断对于减轻后期并发症的可能性至关重要，Eyenai 通过提供准确的检测和诊断来促进这一点。由沙特数据和人工智能管理局(Tawakkalna)开发的应用程序于 2022 年获得联合国公共服务奖，以创新手段应对冠状病毒病(COVID-19)大流行。

B. 工人团结和人工智能

50. 使用人工智能导致的劳动力分散被视为阻碍了工人组织团结的能力：“平台公司可以轻松雇用有时甚至来自世界各地的新工人，而这会产生竞争和孤立，并严重阻碍集体认同的形成。简而言之，基于平台的工作割裂了工人的集体身份，破坏了集体行动，特别是基于团结和信任的行动。”[92]

51. 然而有人建议，团结中介人可以在危机时期帮助工人。[93] 一个核心问题是人工智能是否可以被重新设想为促进工人团结的工具。[94] 乌拉圭社会保险银行为雇主创建了集中的自动化程序，通过在线服务、手机应用程序和多渠道援助促进自我管理。据报道，聊天机器人回答了 97%的查询，满意率为 100%。其成果包括社会保险缴款逃避率大幅下降 24.4%，有 57%的雇主在线注册，42%的雇主在线付款。[95] 阿根廷职业风险监督办公室成功实施了一个名为 Julieta 的人工智能聊天机器人。[96] 一方面，出现了明显更加开放和多样化的新的横向团结形式。还有一种“分布式话语”的趋势，“在这种话语中，活动家和工会官员之间的官僚障碍被消除，透明度得到提高，日常成员有更大的权力挑战和重新制定寡头工会结构”。[97] 此外，还出现了“加速多元化”，其定义为“基于利益的群体政治正在分裂，有利于更多基于问题和流动的群体政治，以及妇女和其他以前被传统工会结构边缘化的工人进行民主讨论的新的安全空间”。[98] 人们不妨将将提高妇女农业工人获得人工智能的能力以预测收获和天气模式作为人工智能团结

[92] Tammy Katsabian，“技术规则——技术如何被用来扰乱基本劳动法保护”，《刘易斯和克拉克法律评论》，第 25 卷，第 3 号(2021 年)。

[93] Saiph Savage 和 Mohammad H. Jarrahi，“COVID-19 期间向人群工作过渡的团结和人工智能”，为“工作新未来”虚拟研讨会撰写的论文，2020 年 8 月。另见 Kurt April，“人工智能引发的团结经济：需要管理指导”。

[94] Frances Flanagan 和 Michael Walker，“工会如何利用人工智能来建立权力？在美国和澳大利亚使用人工智能聊天机器人组织劳工”，《新技术，工作和就业》，第 36 卷，第 2 号(2021 年)。

[95] 见 https://www.issa.int/sites/default/files/documents/2024-06/2-AI%20in%20SecSoc%202024.pdf。

[96] 同上。

[97] 见 Andy Hodder 和 David Houghton，“工会使用社交媒体：Twitter 上的大学和学院工会研究”，《新技术、工作和就业》，第 30 卷，第 3 号(2015 年 11 月)。

[98] 见 Anne-Marie Greene 和 Gill Kirton，“远程参与工会的可能性：动员妇女活动家”，《劳资关系杂志》，第 34 卷，第 4 号(2003 年 10 月)。

的一个范例。[99] 另一个例子是由技术领域的工人领导的反种族隔离运动。[100] 另一方面，加州护士协会熟练地参与要求人工智能保障措施，以保护患者免受治疗不足的影响。[101] 尽管如此，据观察，“雇主在网络空间进行了反动员，而社交媒体言论的法律地位并不确定，对在线员工言论产生了额外的寒蝉效应。”[102]

52. 因此，工会活动类型的背景以及劳工运动的政治和组织差异反映群体身份、内部凝聚力、文化、战略、治理和社区，将影响人工智能在权力争夺问题上与国际团结有关的使用。[103]

五. 尽职调查申诉程序

53. 对透明度的日益关注促使人们要求对企业在访问和使用个人数据方面的过度行为进行审查。例如，Meta 公司宣布它计划实施一个默认设置，用户的内容用于训练人工智能模型。挪威消费者委员会认为选择退出过程非常复杂，并对该公司提出法律投诉，称其违反了欧洲联盟《通用数据保护条例》。[104] 《工商企业与人权指导原则》阐明了原则 15 规定的公司的尽责义务，其中包括要求采取预防性办法以及补救程序；原则 25 规定了诉诸司法的框架；原则 29 要求工商企业建立补救机制。

54. 《联合国全球信息完整性原则》呼吁建立监督机制，并委托定期进行外部和独立的人权审计，审计范围除其他外包括：服务条款、社区标准、广告政策、内容审查、投诉程序、研究人员的数据访问、对脆弱性和边缘化、性别平等和儿童权利的影响评估。审计应当公开，使所有用户都可以查阅和理解。[105]

55. 教科文组织在其《人工智能伦理问题建议书》中指出，各国应建立包容所有利益攸关方的监督机制，并鼓励所有利益攸关方发展人权。2024 年 6 月 13 日欧洲议会和欧洲理事会关于企业可持续性尽职调查的(EU)2024/1760 号指令第 14 条规定，各国有义务确保公司为个人、民间社会组织、人权维护者、工会和其他人提供公平、公开、可利用、可预测和透明的投诉机制，以解决公司、其子公司或活动链中的其他实体侵犯人权和环境影响。[106] 这些措施应坚持保密标准，防止对提出申诉的个人或实体进行报复。该指令建立了一个框架，补充了国际

[99] 见 CEDAW/C/OMN/CO/4，关于人工智能和天气预报，见 https://deepmind.google/discover/blog/graphcast-ai-model-for-faster- and-more-accurate-global-weather-forecasting/。

[100] 见 https://www.notechforapartheid.com/。

[101] 见 https://www.nationalnursesunited.org/press/cna-demand-patient-safeguards-against-artificial-intelligence。

[102] Flanagan 和 Walker：“工会如何利用人工智能来建立权力？在美国和澳大利亚使用人工智能聊天机器人开展劳工组织”。另见 Louise Thornthwaite，“令人发凉的时代：社交媒体政策、劳工法和就业关系”，《亚太人力资源学刊》，第 54 卷，第 3 号(2015 年 8 月)。

[103] 同上。

[104] 见 https://www.forbrukerradet.no/side/legal-complaint-against-metas-use-of-personal-content-for-ai-training/。

[105] 见 https://www.un.org/sites/un2.un.org/files/un-global-principles-for-information-integrity-en.pdf。

[106] 见 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401760。

团结权宣言修订草案(第 6 条第 3 款)所要求的沟通机制。此外，修订后的宣言草案(第 8 条第 3 款)为普遍定期审议中可能设计和分享人工智能国际团结政策和做法制定了灵活的框架。根据宣言修订草案第 8 条第 3 款，各国有义务在各自能力范围内采取步骤，促进保护实际和虚拟通信空间，包括利用互联网和基础设施，以便使个人和人民能够分享团结的思想。这些规范为支持创造性的人工智能国际团结监督机制提供了基础。这些机制可以促进民间社会的参与性包容。

六. 结论

56. 独立专家同意必须建立一个全球多利益攸关方治理模式，以防止和纠正人工智能系统内的歧视和其他侵犯人权行为。[107] 在治理方面，人工智能高级别咨询机构建议采取一种可互操作的方法，涵盖公共和私营行为体，并延伸到包括国际、区域和国家层面在内的各个司法管辖区，例如欧洲联盟的《人工智能法》。[108] 尽管如此，不断发展的系统的一个关键标志是使用自愿代码和协议，例如由科技公司、人工智能开发人员和安全公司签署的旨在打击选举干预的反深度造假协议。[109] 目前，国际社会缺乏公平分享人工智能团结的利益和风险的做法，而全球南方的民间社会组织与来自全球北方的技术公司和人工智能开发人员之间似乎缺乏基本信任。此外，大多数人工智能伦理原则是全球北方起草的，可能缺乏对全球南方特定背景问题的关切，即使后者提供了人工智能数据基础设施发展所必需的矿产和能源。[110] 科技公司和人工智能开发人员之间的权力集中，加剧了与人工智能相关的反团结措施得到加强的风险，从而加剧了国家之间和国家内部以及社会不同部门之间的数字鸿沟。互联网治理论坛强调了全球南方的代表被排除在全球协商之外的问题，如缺乏资金、连通性不稳定以及优先使用英语作为对话语言。[111] 人工智能的国际团结设计，可以形成未来人工智能发展的关键价值，并实施监督机制，以维护程序正义和所有人的包容性参与。为了支持政策变革以实现可持续的人工智能国际团结，独立专家提出了应在未来五年内紧急采取的建议。

七. 建议

57. **独立专家建议各国、公司和民间社会行为体：**

---

[107] 见 https://www.unwomen.org/en/news-stories/explainer/2024/05/artificial-intelligence-and-gender-equality?gad_source=1&gclid=CjwKCAjw4f6zBhBVEiwATEHFVvzWNyAYvJV56epBlSUMBQTVZ4hm_tCsn_VJGGvIzoRMMXfutYebdhoCzqEQAvD_BwE。

[108] 见 https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf。另见 https://oecd.ai/en/accountability。

[109] 见 https://www.techradar.com/pro/top-tech-companies-ai-developers-and-security-firms-sign-anti-deepfake-agreement-to-combat-election-interference。

[110] 互联网治理论坛提供的资料。

[111] 同上。

(a) 确保政府机构、技术公司和民间社会团体，在实施人工智能进行数据处理和决策时，根据教科文组织《人工智能伦理问题建议书》，包容所有个人和团体的积极参与，不论其种族、肤色、血统、性别、年龄、语言、宗教、政治观点、民族或族裔出身、社会或经济背景、残疾或任何其他因素；

(b) 采取国家法规，设计促进人工智能国际团结，采用非歧视方法，确保民间社会和其他相关利益攸关方参与数据处理和决策的政府机构和技术公司的人工智能规划、选择、设计和实施；

(c) 政府和公司应在收集、处理或部署数据之前进行持续的人权影响评估，并由民间社会直接提供意见，并提供充分的透明度，包括人工智能的培训；

(d) 确保将人工智能用于数据处理和决策的政府机构和技术公司建立独立、外部和系统的透明度审计和人权影响评估(以查明培训、算法和决策中的偏见)，并将国际团结作为人工智能整个生命周期的价值和目标；[112]

(e) 确保在国家和区域立法中纳入独立、透明、无障碍、有效的申诉和上诉机制，以便通过政府机构、技术公司或参与数据处理或决策的民间社会团体使用的人工智能，对歧视和(或)排斥行为追究责任；

(f) 认识到举证责任应放在相关政府机构和(或)技术公司身上，以证明人工智能技术如何被利用于涉及因人工智能数据处理或决策而导致的歧视或其他侵犯人权行为的案件中；

(g) 在开发涉及敏感数据(例如健康数据)或大量数据(例如国家数据库)时，在机构数据处理和决策影响评估和权利保护机制中采用明确的人工智能管制；

(h) 确保国家机构和技术公司或民间社会团体以易于理解的方式告知所有处于弱势地位的个人和群体：人工智能将用于数据处理或决策，并会事先征得他们的同意，而不会有直接或间接胁迫；

(i) 确保能够诉诸独立的司法机制，以解决与人工智能有关的歧视或其他侵犯人权行为；

(j) 政府和技术公司应以可解释的方式向弱势群体提供法律信息，说明他们在遭受与使用人工智能有关的侵权行为时的权利和补救机制；

(k) 投资于公众数据素养教育，以解决人工智能深度造假、虚假信息和仇恨言论，从而创造更具复原力的社会；

(l) 承认所有人追求自我实现的权利作为通过设计方法实现人工智能国际团结的关键标准，其进程除其他外涉及获得教育、住房、就业、保健的机会；

(m) 在收集和处理数据时尊重数据主体的自主权；

[112] 见 https://digitalrightscheck.bmz-digital.global/；另见 https://digital-strategy.ec.europa.eu/en/library/ethicsguidelines-trustworthy-ai 和 https://ieeexplore.ieee.org/document/8058187。

(n) 政府和科技公司应告知数据主体其数据保护权利，并尽一切必要努力确保每个人都有数据保护权利，包括删除权、访问权和选择退出权；

(o) 政府和科技公司处理超出原始目的的数据(数据的二次使用)时，不论该数据是如何获得的，都应寻求个人的自由的事先知情同意；安全部门根据与必要性、相称性、合法性和不歧视相关的法律标准打击极端主义或反恐活动时，不应使用人工智能工具针对国际团结活动人士；

(p) 建立法律框架和系统程序，以确定公共利益的范围，并对数据处理的紧迫性和必要性进行分类；个人应保留选择退出的权利以保护其隐私；

(q) 国家和公司应根据符合性别平等的人工智能国际团结，禁止创造合成或操控的、未经同意的深度私密造假。

58. 独立专家建议民间社会：

(a) 继续在创建符合伦理的由人工智能驱动的内容审核工具时倡导包容，以防止、识别和消除对妇女和遭受跨部门歧视的人的在线暴力和歧视；

(b) 继续向联合国、各国政府和科技公司或人工智能开发人员提供算法歧视的例子，以便提供改进实践的建议；

(c) 工会应促进工人在人工智能共同设计、培训和风险评估中的参与和协商，以及获得数字素养。

59. 独立专家建议各公司：

(a) 建立具体机制，扩大传统上被边缘化的群体的声音，以促进安全、包容的在线环境，承认相互尊重多样性的原则，包括以不同语言提供协商，并设立赠款，以促进全球南方利益攸关方的参与；

(b) 科技公司和人工智能开发人员应该采用旨在通过设计尊重地球界限的人工智能国际团结；进行独立、透明和持续的环境风险评估；推广节能算法；通过使用可再生能源支持可持续数字开发；为创建数据提供人工智能团结经济融资，以支持公众可访问的公共区域(如海底和海洋)的环境和气候变化知识；管理电子废物；奉行包容当地社区的循环经济原则；

(c) 人工智能的资助者和人工智能开发人员应致力于人工智能的国际团结，进行预防性和系统性的人权评估，以查明在数据的整个生命周期中使用算法偏见违反非歧视和平等的风险，并要求将直接和间接的民间社会利益攸关方包括在内(包括在结构上沉默的社区)，参与独立的监督机制，以查明和应对反性别平等和反民主的图谋；

(d) 科技公司应建立预防性和反应式团结机制，以解决与使用人工智能有关的侵犯人权行为，例如导致社会暴力或骚扰、监视、歧视或对结构性沉默社区进行过度审查的虚假信息和错误信息活动；[113]

(e) 技术公司应提供快速反应团队。这些团队应拥有充足的资源，并拥有及时回应申诉和提供解决办法的决策权。

60. 独立专家建议联合国：

(a) 创建一个数字团结平台，使(来自全球北方和全球南方)民间社会组织通过该平台交流国际团结的想法，此外，还可以根据独立专家关于民间社会和国际团结的报告中概述的建议，与技术公司代表、人工智能开发人员和国家官员讨论人工智能包容性和非歧视政策和实践、挑战及创新；[114]

(b) 设立一个基金，支持为所有语言的数据建立大型语言模型，以保护文化多样性。

[113] 该条建议与《联合国全球信息完整性原则》相关：“提高危机应对能力。与在高风险地区开展业务的利益攸关方合作，建立预警和升级流程，在危机和冲突情况下加快及时反应速度。建立机制，以便能够突出、及时地获得服务于公共利益的可靠、准确的信息。”

[114] A/HRC/56/57。