联合国 A/HRC/48/31

大 会

Distr.: General
13 September 2021
Chinese
Original: English

人权理事会

第四十八届会议

2021 年 9 月 13 日至 10 月 1 日

议程项目 2 和 3

联合国人权事务高级专员的年度报告和
联合国人权事务高级专员办事处的报告
以及秘书长的报告

促进和保护所有人权——公民权利、
政治权利、经济、社会及文化权利，
包括发展权

数字时代的隐私权*

联合国人权事务高级专员的报告

概要

本报告是人权理事会第 42/15 号决议授权编制的。高级专员在报告中分析了各国和工商企业广泛使用人工智能，包括特征分析、自动化决策和机器学习技术，如何影响隐私权及相关权利的享有。在概述了国际法律框架之后，高级专员强调指出了人工智能为干涉隐私提供方便的各个方面，并举例说明了隐私权和相关权利在四个重要领域所受的影响。高级专员随后探讨了应对挑战的方法，就设计和实施保障措施向各国和工商企业提出了一套建议，以防止和尽可能减少有害后果并促进充分享受人工智能可带来的惠益。

* 本报告逾期提交，以纳入最新信息。

一. 导言

1. 人权理事会在其第 42/15 号决议中请联合国人权事务高级专员组织一次为期一天的专家研讨会，讨论人工智能(包括特征分析、自动化决策和机器学习技术)在没有适当保障的情况下可能会如何影响隐私权的享有，就此问题编写一份专题报告，提交理事会第四十五届会议。[1]

2. 近年来，没有其他技术的发展比人工智能(特别是机器学习技术)更引发公众的想象力。[2] 的确，这些技术可成为一股巨大的向善力量，帮助社会克服当前时代的一些重大挑战。然而，如果在运用这些技术时没有充分考虑其对人权的影响，也可能产生负面，甚至是灾难性的影响。

3. 本报告虽然不侧重 2019 冠状病毒病(COVID-19)大流行，但当前的全球健康危机提供了一个强有力、高能见度的实例，说明人工智能在全球生活不同领域的发展速度、规模和影响。使用多种类型数据(地理位置、信用卡、交通系统、健康和人口统计)以及有关个人网络信息的接触者追踪系统已被用于追踪疾病的传播。人工智能系统已被用于将有关个人标识为潜在受感染者或传染源，要求他们进行隔离或检疫。用于预测成绩分布的人工智能系统导致出现了歧视公立学校和贫穷社区学生的结果。这些发展表明了人工智能系统对人们日常生活的广泛影响。隐私权在所有这些情形中都受到影响，人工智能使用个人信息，常常做出对人们生活有切实影响的决定。然而，与隐私问题息息相关的是对享受其他权利的各种影响，如健康、教育、行动自由、和平集会自由、结社自由和表达自由等权利。

4. 2019 年，联合国秘书长在“最高愿望：人权行动呼吁”中确认，数字时代开辟了人类福祉、知识和探索的新疆域。他强调，数字技术为倡导、捍卫和行使人权提供了新的手段。然而，新技术常常被用于侵犯权利，特别是那些已经很脆弱或落在后面的人的权利，例如通过监视、压制、审查和网上骚扰，包括针对人权捍卫者的此类做法。福利系统的数字化虽然具备提高效率的潜力，但却带来了把最需要帮助的人们排除在外的风险。秘书长强调，绝不能利用新技术的进步来损害人权、加深不平等或加剧现有歧视。他强调指出，人工智能的治理需要确保公平、问责、可解释性和透明度。在安全领域，秘书长再次呼吁在全球禁止致命自主武器系统。

5. 本报告以高级专员关于数字时代隐私权问题的前两次报告为基础。[3] 报告中还纳入了根据理事会第 42/15 号决议于 2020 年 5 月 27 日和 28 日举办的虚拟专家研讨会的结论意见以及各方对高级专员要求对本报告提供投入的呼吁作出的回应。[4]

[1] 该报告的编写被推迟。见 A/HRC/45/26 和 A/HRC/47/61。

[2] “人工智能”一语尚无一个普遍认同的定义。在本报告中，它由一系列程序和技术构成，使计算机能够辅助或取代人工完成具体任务，例如作出决定和解决问题(A/73/348，第 3 段)，其中包括但不限于机器学习和深度学习。

[3] A/HRC/27/37 和 A/HRC/39/29。

[4] 关于征求意见呼吁以及所收集意见，见 www.ohchr.org/EN/Issues/DigitalAge/Pages/cfi-digital-age.aspx。

二. 法律框架

6. 《世界人权宣言》第十二条、《公民及政治权利国际公约》第十七条以及若干其他国际和区域人权文书均确认隐私权为一项基本人权。[5] 隐私权在国家和个人之间的权力平衡中起着关键作用，它是民主社会的一项基础性权利。[6] 在当今日益以数据为中心的世界里，隐私权对线上和线下享有和行使其他人权的重要性与日俱增。[7]

7. 隐私权是人的尊严的一种表现，牵涉对人的自主性和个人身份认同的保护。[8] 在使用人工智能的背景下，隐私的一些方面包括信息隐私特别重要，因为它们涵盖关于个人及其生活的现有或可推导得出的信息，以及在这些信息基础上作出的决定，[9] 此外也牵涉就个人身份认同作出决定的自由。

8. 对隐私权的任何干涉都不得任意或非法。[10] “非法”一词是指各国只能在法律基础上依照相关法律干涉隐私权。法律本身必须符合《公民及政治权利国际公约》的规定、目标和目的，并且必须具体说明在什么情况下才容许这种干涉。[11] 使用干涉这个概念的用意是确保法律所规定的干涉都符合《公约》的规定、目标和目的，而且无论如何都要在特定情形中合情合理。[12] 因此，对隐私权的任何干涉都必须服务于合法目的，是实现该合法目的所必需，而且是相称的。[13] 任何限制也都必须是现有可用的侵扰性最小的选择，而且不得损害隐私权的本质。[14]

9. 隐私权适用于每个人。假如基于种族、肤色、性别、语言、宗教、政治或其他见解、民族血统或社会出身、财产、门第或其他地位而对其进行差别保护，那就不符合《公民及政治权利国际公约》第二条第 1 款和第三条规定的不歧视原则。基于这些原因的歧视也违反了《公约》第二十六条所载法律面前平等的权利。

---

[5] 见《儿童权利公约》第 16 条、《保护所有移徙工人及其家庭成员权利国际公约》第 14 条、《残疾人权利公约》第 22 条、《非洲儿童权利与福利宪章》第 10 条、《美洲人权公约》第 11 条和《保护人权与基本自由公约》(《欧洲人权公约》)第 8 条。

[6] A/HRC/39/29，第 11 段。

[7] 儿童权利委员会，第 25(2021)号一般性意见，第 67 和 68 段；A/HRC/39/29，第 11 段。

[8] 儿童权利委员会，第 25(2021)号一般性意见，第 67 段；欧洲人权法院，*Goodwin* 诉联合王国，申请编号 28957/95，2002 年 7 月 11 日判决，第 90 段。

[9] A/HRC/39/29，第 5 段。

[10] 关于“任意”和“非法”两个用语的详细分析，见 A/HRC/27/37，第 21 至 27 段。

[11] 人权事务委员会，第 16(1988)号一般性意见，第 3 和 8 段。

[12] 同上，第 4 段。

[13] *Toonen* 诉澳大利亚(CCPR/C/50/D/488/1992)，第 8.3 段；*Van Hulst* 诉荷兰(CCPR/C/82/D/903/1999)，第 7.3 和 7.6 段；*Madhewoo* 诉毛里求斯(CCPR/C/131/D/3163/2018)，第 7.5 段；CCPR/C/USA/CO/4，第 22 段。另见儿童权利委员会第 25(2021)号一般性意见，第 69 段。

[14] 人权事务委员会，第 31(2004)号一般性意见，第 6 段；A/HRC/27/37，第 22 段，以及 A/HRC/39/29，第 10 段。

10. 《公民及政治权利国际公约》第二条第 1 款要求各国尊重和保障其本国境内和管辖范围内所有人的《公约》所赋权利，不受歧视。换言之，各国不仅不得侵犯《公约》承认的权利，而且还有义务采取积极步骤保护这些权利的享受。[15] 这意味着有义务采取适当的立法和其他措施，保护个人的隐私不受干涉，无论这种干涉是来自国家当局，还是来自自然人或法人。[16] 这一义务也反映在《工商企业与人权指导原则》第一支柱中，其中述及各国有义务提供保护，避免出现与公司有关的负面人权影响。

11. 工商企业有责任尊重所有国际公认的人权。这意味着应当避免侵害他人的人权，并在自身卷入时消除负面人权影响；《工商企业与人权指导原则》第二支柱为所有企业提供了关于如何履行这一责任的权威蓝图。[17] 尊重的责任贯穿企业的所有活动和商业关系。

三. 人工智能对隐私权和其他人权的影响

A. 人工智能系统的相关特征

12. 人工智能系统的运行可以不同方式便利和强化对隐私的侵犯以及对各种权利的其他干涉。其中包括全新的应用程序以及人工智能系统的功能，这些程序和功能可以扩大、强化或鼓励对隐私权进行干预，最明显的是通过增加个人数据的收集和使用。

13. 人工智能系统通常依赖大数据集，常常包括个人数据。这鼓励了广泛的数据收集、存储和处理。许多工商企业优化服务，以收集尽可能多的数据。[18] 例如，像社交媒体公司这样的在线企业依赖收集大量的互联网用户数据并出售这些数据盈利。[19] 所谓的物联网是企业和国家都在利用的快速增长的数据来源。数据收集在私密场合以及私人和公共空间进行。[20] 数据经纪人获取、合并及分析个人数据，并与无数接收者分享。这些数据交易在很大程度上不受公众监督，只受到现有法律框架的轻微限制。[21] 所产生的数据集是巨大的，所收集的信息也是空前大量的。

14. 除了将人们的私生活暴露给公司和国家政府之外，这些数据集还在许多其他方面使个人很容易受到影响。数据泄露事件一再导致数百万人的敏感信息外

[15] 人权事务委员会，第 31(2004)号一般性意见，第 6 段。

[16] A/HRC/39/29，第 23 段。另见人权事务委员会，第 16(1998)号一般性意见，第 1 和 9 段，以及第 31(2004)号一般性意见，第 8 段；儿童权利委员会，第 25(2021)号一般性意见，第 36 至 39 段。

[17] 在其第 17/4 号决议中，人权理事会一致通过了《工商企业与人权指导原则》。

[18] Wolflie Christl, *Corporate surveillance in everyday life* (维也纳，Cracked Lab-Institute for Critical Digital Culture，2017 年)。

[19] 《数字化权利排名》提供的材料。

[20] 德里国立法学院传播治理中心、Derechos Digitales(数字权利)、数字权利观察、全球合作伙伴数字、国际非营利法中心和乌贝兰迪亚联邦大学提交的材料。

[21] Aaron Rieke 和其他人，《开放社会中的数据经纪人》(伦敦，开放社会基金会，2016 年)。

泄。[22] 大型数据集使得能够开展无数形式的数据分析以及与第三方进行共享，而这往往会进一步侵犯隐私，并导致其他负面的人权影响。例如，允许政府机构直接获取工商企业所持此类数据集的安排增加了任意或非法干涉有关个人隐私权的可能性。[23] 一个特别令人担忧的问题是，不同来源数据融合之后有可能促成去匿名化。[24] 与此同时，数据集的设计可能会对个人身份认同产生影响。例如，一个数据集如果以二元方式记录性别，就会错误标识那些没有注明男性或女性者。个人数据的长期存储也产生特别的风险，因为数据的未来利用形式在最初收集数据时并未设想到。[25] 随着时间的推移，数据可能会变得不准确、不相关或延续历来的错误识别，进而导致未来数据处理产生偏颇或错误结果。[26]

15. 应当注意的是，人工智能系统并不完全依赖个人数据的处理。不过，即使不涉及个人数据，包括隐私权在内的各种人权依然可能受到此类数据使用的不利影响，如下所示。[27]

16. 人工智能工具被广泛用于寻求深入了解人类行为模式。有了正确的数据集，就有可能得出结论，判断某个特定社区有多少人可能会去某个礼拜场所，他们会更喜欢什么电视节目，甚至他们起床和睡觉的大致时间。人工智能工具可以对个人做出深远的推论，包括关于其精神和身体状况的推论，而且可借以识别群体，比如识别有特殊政治或个人倾向的人。[28] 人工智能也被用于评估未来行为或事件的可能性。人工智能所作的推论和预测尽管具有概率性质，但却可能成为会影响人们权利的决定的依据，有时是完全自动化作出决定。

17. 许多推论和预测都深刻影响到隐私权的享受，包括人们的自主性及其确定自己身份细节的权利。它们还提出了许多与其他权利有关的问题，如思想自由权、意见自由权、表达自由权和公正审判权及相关权利。

18. 基于人工智能作出的决定并非没有错误。事实上，人工智能解决方案的可扩缩性可以急剧增加看似很小的出错率的负面影响。[29] 人工智能系统的错误输出有多种来源。首先，人工智能算法的输出带有概率元素，这意味着它们的输出附带着不确定性。[30] 此外，所用数据的相关性和准确性常常令人怀疑。此外，不切实际的期望可能会导致部署没有能力实现预期目标的人工智能工具。例如，一项对

[22] 例如，见 www.ftc.gov/news-events/press-releases/2019/07/equifax-pay-575-million-part-settlement-ftc-cfpb-states-related。

[23] 全球网络倡议提供的材料。

[24] 德里国立法学院传播治理中心、数字权利和隐私国际提供的材料。

[25] OVD-Info 提供的材料。

[26] 消除种族歧视委员会，第 36(2020)号一般性建议第 33 段。

[27] 欧洲委员会，“关于解决算法系统对人权的影响的准则”，(部长理事会就算法系统对人权的影响向成员国提出的 CM/Rec(2020)1 号建议附录)，A 节，第 6 段。

[28] 数字权利和隐私国际提供的材料。

[29] 德国提交的材料。

[30] 欧洲联盟基本权利署，“#BigData：由数据支持的决策中的歧视”(维也纳，2018 年)，第 4 页。

数以百计诊断和预测2019冠状病毒病风险的医疗人工智能工具所作的分析显示，其中无一适合临床使用，而这些工具在开发时曾被寄予厚望。[31]

19. 依赖错误数据的人工智能系统所产生的输出可能会以多种方式助长侵犯人权行为，例如错误地标记一个人，断定其可能是恐怖分子或犯有福利欺诈行为。导致作出基于人工智能系统的歧视性决策的偏向性数据集尤其令人担忧。[32]

20. 许多人工智能系统的决策过程是不透明的。人工智能系统开发和运作背后的数据环境、算法和模型很复杂，而且政府和私营行为体故意隐秘，这些因素妨碍公众以切实方式了解人工智能系统对人权和社会的影响。机器学习系统增加了一个重要的不透明因素；他们可能有能力识别相关模式并开发出难以甚至不可能解释的解决方案。[33] 这常常被称为“黑匣子”问题。[34] 这种不透明状况使得难以对人工智能系统进行有意义的审查，而且可能成为人工智能系统造成伤害后进行有效问责的障碍。[35] 然而，应当指出的是，这些系统无需完全不可理解。[36]

B. 对关键部门人工智能系统的担忧

21. 本节探讨人工智能工具应用引发关注的四个关键领域，以说明在具体实践中是如何经历这些关切的。

执法、国家安全、刑事司法和边境管理中的人工智能

22. 各国正日益把人工智能系统融入执法、国家安全、刑事司法和边境管理系统。[37] 其中许多应用可能确实令人担忧，但本节将重点讨论几个特定的例子，它们代表着一些新出现的不同人权问题。

23. 人工智能系统经常被用作预测工具。他们使用算法来分析大量数据，包括历史数据，以评估风险和预测未来趋势。视目的而定，培训数据和所分析的数据可以包括诸如犯罪记录、逮捕记录、犯罪统计、警察在特定居民区出警记录、社交媒体帖子、通信数据和旅行记录等。[38] 这些技术可能被用来创建人们的档

[31] 见 www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/。

[32] 消除种族歧视委员会，第36(2020)号一般性建议，第31至36段；数字合作高级别小组，“数字相互依存的时代”(2019年6月)，第17和18段。

[33] 德国提交的材料。

[34] 见 www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/。

[35] 见 www.cambridge.org/core/journals/international-and-comparative-law-quarterly/article/international-human-rights-law-as-a-framework-for-algorithmic-accountability/1D6D0A456B36BA7512A6AFF17F16E9B6；www.cambridge.org/core/journals/international-review-of-the-red-cross/article/abs/ai-for-humanitarian-action-human-rights-and-ethics/C91D044210CADF7A0E023862CF4EE758。

[36] 例如，见 Inioluwa Deborah Raji 等人，“弥合人工智能责任差距：为内部算法审计界定一个端到端的框架”，2020年1月3日。

[37] 关于人工智能和其他数字技术在边界管理中所涉人权问题的深入分析，见 A/75/590。

[38] 隐私国际提交的材料。另见 A/HRC/44/57，第35段。

案，识别犯罪或恐怖活动增加的可能地点，甚至将个人标记为可能的嫌疑人和未来的惯犯。[39]

24. 这些活动对隐私和更广泛人权的影响是巨大的。首先，使用的数据集包含关于大量个人的信息，因此牵涉到他们的隐私权。其次，它们可能引发国家干预，如搜查、讯问、逮捕和起诉，尽管由于所作预测的概率性质，人工智能评估本身不应被视为合理怀疑的依据。受影响的权利包括隐私权、公平审判权、不受任意逮捕和拘留的权利以及生命权。第三，人工智能决策的固有不透明性引起了特别紧迫的问题，亦即人工智能为强制性措施提供信息时的国家问责问题，在通常普遍缺乏透明度的领域，如反恐部队活动中，更是如此。[40] 第四，预测性工具带有延续甚至加剧歧视的固有风险，反映所用数据集内含的历史种族和族裔偏见，例如对某些少数群体的治安巡查重点不成比例。[41]

25. 生物识别技术领域的发展已导致执法和国家安全机构越来越多地使用这种技术。生物特征识别依赖个人某些特征(例如面部、指纹、虹膜、语音或步态)的数字代码与数据库中其他此类代码的比较。[42] 通过比较，可以推断出该人确实是要鉴别或识别的人的概率高低。这些过程越来越多地以实时和远程方式进行。特别是，全球各地当局越来越多地启用远程实时面部识别。[43]

26. 远程实时生物识别在国际人权法下引起了严重关切，高级专员以前曾强调过这一点。[44] 其中一些关切反映了与预测性工具相关的问题，包括错误识别个人的可能性以及对某些群体成员的不成比例影响。[45] 此外，面部识别技术可以用于根据族裔、种族、民族血统、性别和其他特征，进行个人资料分析。[46]

27. 远程生物识别与对隐私权的深度干扰存在关联。一个人的生物识别信息是其个性的重要特征之一，因为它揭示了该人区别于其他人的独特特征。[47] 此外，远程生物识别极大提高了国家当局在公共场所有系统识别和追踪个人的能力，削弱了人们在不被监视情况下生活的能力，并且对行使表达自由权利、和平集会权利

[39] 见 www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf。

[40] A/74/335 和 A/HRC/43/46，第 37 至 38 段。

[41] 科技蜂巢顾问有限公司提交的材料。另见：消除种族歧视委员会，第 36(2020)号一般性建议，第 33 段；联合国人权事务高级专员关于促进和保护非洲人和非洲人后裔人权与基本自由以使其不受执法人员过度使用武力和其他侵犯人权行为侵害的会议室文件(A/HRC/47/CRP.1)，可查阅 www.ohchr.org/Documents/Issues/Racism/A_HRC_47_CRP_1.pdf，第 15 和 19 段。

[42] A/HRC/31/64，第 14 段。

[43] 数字权利和国际非营利法中心提交的材料。

[44] A/HRC/44/24。

[45] 隐私国际提交的材料。

[46] A/HRC/44/57，第 39 至 40 段。

[47] A/HRC/44/24，第 33 段。另见欧洲人权法院，*Reklos* 和 *Davourlis* 诉希腊(第 1234/05 号申诉)，2009 年 4 月 15 日的判决，第 40 段。

和结社权利以及行动自由产生了直接负面影响。[48] 因此，在此背景下，高级专员欢迎最近为限制或禁止使用实时生物识别技术而开展的努力。[49]

28. 人工智能工具的开发据称也是为了可以根据人的面部表情和其他“预测性生物特征”来推断其情绪和精神状态，从而确定其是否构成安全威胁。[50] 人脸情绪识别系统的运行前提是能够自动、系统地从人的面部表情中推断出其情绪状态，这是缺乏坚实科学基础的。[51] 研究人员发现，情绪与面部表情之间只有微弱的联系，[52] 并强调面部表情在不同文化和背景下存在差异，[53] 这使得情绪识别容易受到偏见和误解的影响。鉴于这些关切，公共当局使用情绪识别系统，例如专门挑出特定人员由警察拦截或逮捕，或评估审讯期间所作陈述的真实性，有可能会损害人权，如隐私权、自由权和公平审判权。

人工智能系统和公共服务

29. 人工智能系统越来越多地被用来帮助提供公共服务，常常号称是为了开发更有效的系统，以便及时而准确地提供服务。这也越来越多地出现在人道主义场合中，在此种场合中，人道主义货物和服务的交付可能与人工智能系统联系在一起。尽管这些是合法的，甚至值得称道的用途，但如果没有适当的保障措施，在提供公共和人道主义服务时部署人工智能工具就有可能对人权产生不利影响。

30. 人工智能被用于各种公共服务，从进行各种福利决策到确定儿童保育机构需进行家访的家庭，不一而足。[54] 这些决定是利用大型数据集做出的，其中不仅包括国家持有的数据，而且还可包括从私人实体(如社交媒体公司或数据经纪人)获取的信息，此类信息常常是在保护性法律框架之外收集的。[55] 此外，由于人工智能系统的计算知识和能力往往掌握在私营公司手中，这些安排常常意味着私营公司可以获得包含众多人口信息的数据集。这既引起了对隐私的担忧，也引发了对于数据中嵌入的历史偏见会如何影响公共当局决策的担忧。

---

[48] 见欧洲数据保护委员会和欧洲数据保护监管局，第 5/2021 号联合意见，第 30 段；国际非营利法中心和隐私国际提交的材料。另见 A/HRC/44/24，第 34 段，以及 A/HRC/41/35。

[49] 欧洲联盟提交的材料。另见 https://nymag.com/intelligencer/2020/01/why-we-should-ban-facial-recognition-technology.html；欧盟委员会，关于制定欧洲议会和理事会关于人工智能的统一规则(人工智能法案)和修订联盟某些立法法案的条例提案，2021 年 4 月 21 日 COM(2021)206 最终稿，第 5(1)(D)条。

[50] 见 www.ohchr.org/Documents/Issues/Terrorism/biometricsreport.pdf。

[51] 见 www.nature.com/articles/d41586-020-00507-5。

[52] 见 https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780190613501.001.0001/acprof-9780190613501-chapter-7。

[53] 见 https://journals.sagepub.com/doi/10.1177/1529100619832930；以及 https://pubmed.ncbi.nlm.nih.gov/22509011/。

[54] 见 A/74/493。

[55] 隐私国际提交的材料。

31. 关于将人工智能用于公共服务的一个重大关切是，它可能具有歧视性，特别是对于边缘化群体而言。[56] 极端贫困与人权问题特别报告员警告说，可能会出现一种“数字福利敌托邦”，即不受限制的数据匹配被用于曝光、调查和惩罚福利受益者，而且受益者被强加条件，使个人的自主权和选择受到损害。[57] 这些担忧最近在荷兰得到了体现，那里有一项被广泛报道的法院裁决，其中禁止使用一种数字福利欺诈检测系统，因为该系统被认定侵犯了隐私权。该案所涉系统为中央和地方当局提供了广泛的操纵力，可以分享和分析以前单独存放的数据，包括就业、住房、教育、福利和医疗保险等数据以及其他形式可识别数据。此外，该工具针对的是低收入和少数族裔社区，导致基于社会经济背景的事实上歧视。[58]

人工智能在工作场合的使用

32. 各种类型和规模工商企业的广泛雇主都表明，使用数据驱动技术(包括人工智能系统)监控和管理员工的需求日益增长。所谓的人员分析声称能提供更有效、更客观的员工信息。这可包括招聘、晋升计划或解雇方面的自动化决策。

33. 虽然此类技术的大部分重点是在于监控与工作相关的行为和表现，但人工智能系统的一系列应用也延伸到与工作无关的行为和数据。[59] COVID-19 大流行在两个方面加速了这一趋势。首先，一些为工人提供预防性健康计划的公司越来越多地收集与健康相关的数据。其次，随着越来越多的流程在人们在家工作时以数字方式执行，人工智能系统对工作场所的监控被引入人们家中。这两种趋势都增加了将工作场所监测数据与非工作相关数据输入合并的风险。这些基于人工智能的监控做法在整个数据使用周期构成巨大的隐私风险。除此之外，数据还可以用于除最初传达给员工之外的其他目的，这可能导致所谓的功能蠕变。[60] 与此同时，许多用于人员管理的人工智能系统的定量社会科学基础并不牢固，容易产生偏差。例如，如果一家公司使用一种在历史数据集基础上生成的人工智能招聘算法，而那些数据集偏向男性、白人和中年男子，那么由此产生的算法将不利于女性、有色人种以及同样有资格填补空缺的年轻人或老年人。[61] 与此同时，保护工人的问责结构和透明度往往缺乏，工人们日益面临着很少或根本无法得到有关人工智能监测做法的解释。[62] 虽然在某些情况中，公司真心

[56] 数字权利观察提交的材料。有关自动化在福利系统中的不同影响的深入分析，请参阅 Virginia Eubanks，*Automating Inequality* (纽约，圣马丁出版社，2018 年)。

[57] 见 A/74/493。另见特别报告员的信(IRL 1/2020)，其中他指出了对一种数字服务卡的类似关切以及与此相关的答复。本报告数处提到了人权理事会各特别程序任务负责人发出的信函。所有此类信函和相关答复均可查阅 https://spcommreports.ohchr.org/Tmsearch/TMDocuments。

[58] 见 www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25522。

[59] 见 https://journals.sagepub.com/doi/10.1177/20539517211013051。

[60] Christl, *Corporate surveillance in everyday life*.

[61] 波兰提交的材料。另见 www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G。

[62] 见 https://journals.sagepub.com/doi/full/10.1177/2053951720938093。

希望防止工作场所不当行为，但为维护这种利益而采用的措施常常不能证明量化工作场合社交互动模式和相关绩效目标的广泛侵入性做法是合理的。在工作场所，鉴于雇主和雇员之间的权力关系，人们也可以想象工人被迫放弃隐私权以换取工作的潜在情形。[63]

用于在线管理信息的人工智能

34. 社交媒体平台使用人工智能系统来为内容管理决定提供支持。[64] 公司使用这些系统对内容进行排名，决定放大哪些内容，缩减哪些内容，包括根据不同用户的个人资料对这些决定进行个性化调整。在对内容实行限制时也使用自动化工具，包括在司法管辖区内和司法管辖区之间应对不同的法律要求。[65] 对中介媒体设定与所感知网上伤害相关的过滤义务，可能导致扩大对人工智能的广泛依赖，而不考虑这些系统在地方和全球层面对隐私权和表达自由权的严重影响。

35. 管理、放大和调节系统所依赖的海量数据集是通过对平台用户及其个人网络进行广泛在线监控和分析而创建及不断扩展的。[66] 这种收集信息并从中推断的持续不断过程，再加上极端的市场集中度，导致出现这样一种状况：少数几家公司在全球持有并控制着数十亿个人和整个网络化公共领域的信息资料。

36. 正如两位先后担任过促进和保护意见和表达自由权特别报告员的任务负责人指出的那样，由拥有巨大市场力量的公司进行的人工智能辅助内容管理引起了人们对个人形成和加深见解的能力所受影响的担忧。[67] 此外，平台推荐系统往往专注于最大限度地提高用户参与度，同时依赖对人们偏好、人口组成和行为模式的深入分析，而情况表明，这常常助长耸人听闻的内容，潜在地加剧两极分化的趋势。[68] 此外，有目标地提供信息有可能不受欢迎，甚至会导致危险的侵犯隐私做法。例如，推荐系统导致暴力受害者发现，社交媒体平台向其推荐施暴者作为潜在朋友，反之亦然，使受害者面临危险。此外，搜索结果所得数据中反映的多数群体或主导群体的偏见已被证明会影响少数群体或弱势群体分享的或有关少数群体或弱势群体的信息。例如，研究表明，谷歌搜索结果中存在令人不安的重大性别[69] 和种族偏见。[70]

---

[63] 见 www.californialawreview.org/print/3-limitless-worker-surveillance/。

[64] 见 www.theverge.com/2020/11/13/21562596/facebook-ai-moderation；https://journals.sagepub.com/doi/full/10.1177/2053951720943234。

[65] 见 OTH 71/2018 和 OTH 73/2020。关于自动内容过滤的深入分析，另见 https://journals.sagepub.com/doi/full/10.1177/2053951720920686。

[66] A/73/348，第 17 段。

[67] 同上，第 25 段，以及 A/HRC/47/25，第 36 段。

[68] 见 www.brookings.edu/techstream/how-youtube-helps-form-homogeneous-online-communities/。

[69] 奥地利和德国提交的材料。

[70] Safiya Umoja Noble, *Algorithms of Oppression* (New York, New York University Press, 2018).

四. 应对挑战

37. 越来越多的专家、利益攸关方和国际社会已认识到，有必要对新技术总体，特别是人工智能，采取基于人权的方法。[71] 基于人权的方法可提供一个工具箱，帮助社会找到预防和减轻伤害的办法，同时最大限度地实现技术进步的惠益。

A. 基本原则

38. 基于人权的人工智能应用方法要求实行一些核心原则，包括平等和不歧视、参与和问责制，这些原则也是可持续发展目标以及《工商企业与人权指导原则》的核心。此外，合法性、正当性、必要性和相称性的要求必须连贯一致地适用于人工智能技术。[72] 此外，人工智能的运用应能促进落实经济、社会和文化权利，确保实现这些权利的关键要素：可提供性、可负担性、可获得性和良好质量。[73] 那些因人工智能的使用导致人权遭受侵犯和践踏的人应当能够获得有效的司法和非司法补救。[74]

39. 如上所述，对隐私权的限制必须经法律规定，是实现合法目标所必需，而且与该目标相称。在实践中，这意味着各国必须审慎地确定某项措施是否能够达到既定目标，该目标有多重要以及那项措施将产生什么影响。各国还应确定侵入程度较低的方法是否可以取得同样的结果和同样的实效；倘若可以，那就需要采取这样的措施。高级专员已经概述了情报机构和执法部门监控活动中的此类必要限制和保障措施。[75] 应该指出的是，必要性和相称性检验也可能导致得出不能采取某些措施的结论。例如，强加于电信和其他公司的全面、不加区分地保留通信数据的要求将无法通过相称性测试。[76] 同样，如果没有提供其他选择，对福利受惠人强加生物特征识别要求也是不相称的。此外，至关重要的是，对有关措施的评估不能是孤立的，而是适当考虑到不同但相互作用的措施的积累影响。例如，一个国家在决定部署新的基于人工智能的监视工具之前，必须评估现有能力及其对享受隐私权和其他权利的影响。

B. 法律和法规

40. 隐私权和相互关联权利的有效保护取决于各国确立的法律、法规和体制框架。[77]

---

[71] 大会第 75/176 号决议，第 6 段；人权理事会第 47/16 号决议第 8(d)段和第 47/23 号决议序言部分第 16 段；A/73/348 第 47 至 60 段、A/75/590 第 57 段以及 A/HRC/43/29；奥地利、加拿大隐私专员、数字权利观察、全球网络倡议和隐私国际提交的材料。

[72] A/HRC/43/29，第 41 段。

[73] 见 A/HRC/43/29 中关于新技术对实现经济、社会和文化权利的作用的详细分析。

[74] 《公民及政治权利国际公约》第二条第 3 款以及《工商企业与人权指导原则》原则 15(c)和支柱三。

[75] A/HRC/39/29，第 34 至 41 段。

[76] 同上，第 18 段；欧洲联盟法院，*Digital Rights Ireland and Others*, C-293/12 和 C-594/12，第 69 段。另见欧洲联盟法院，*Maximilan Schrems* 诉数据保护专员案，C-362/14，第 94 段，裁定“允许公共当局广泛获取电子通信内容的立法必须被视为有损私生活受尊重权利的实质”。

[77] A/HRC/39/29，第 26 段。

41. 随着数据驱动的人工智能系统的出现，数据隐私法下进行有效法律保护的重要性与日俱增。这些保护措施应符合高级专员上次关于隐私权的报告中确定的最低标准。[78]

42. 数据隐私框架应该考虑到与使用人工智能相关联的新威胁。[79] 例如，法律可能对可合法推断和/或进一步使用和共享的数据类型施加限制。立法者也应考虑加强个人的权利，包括给予他们获得明确解释的权利以及反对影响其权益的完全自动化决定的权利。[80] 随着人工智能技术的不断发展，有必要继续在数据隐私框架内制定更多保障措施。

43. 应对全球数据环境(包括其巨大信息不对称)日益复杂和不透明的一个关键要素是独立的数据隐私监督机构。这些机构需要拥有有效的执法权力并获得足够的资源。应当使民间社会组织具备能力，以支持执行数据隐私法，包括通过建立强有力的投诉机制。

44. 除了数据隐私立法，还需要审查并在可能情况下通过更广泛的法律，以尊重权利的方式应对人工智能的挑战。[81]

45. 考虑到人工智能应用、系统和使用的多样性，条例法规应当足够具体，以解决特定行业的问题，并针对所涉及的风险定制应对措施。[82] 人权风险越高，使用人工智能技术的法律要求就应该越严格。因此，对个人利害关系特别大的部门，如执法、国家安全、[83] 刑事司法、社会保障、就业、医疗、教育和金融部门，应优先考虑。对立法和条例法规采取与风险相称的方法将要求禁止某些人工智能技术、应用或用例，因为它们会造成无法在国际人权法下合理解释的潜在或实际影响，包括那些不能通过必要性和相称性测试的影响。此外，不应允许使用与禁止歧视有内在冲突的人工智能。例如，应按照这些原则要求政府[84] 或人工智能系统不得基于已被禁止的歧视理由，[85] 对个人进行社会评分，将个人归类于特定集群。

[78] 同上，第 28 至 33 段。

[79] 例如，2018 年通过的欧洲委员会《修正〈关于在自动处理个人数据方面保护个人的公约〉的议定书》就是对所出现的新数据处理做法的回应。

[80] 见包含此类权利的欧洲联盟《一般数据保护条例》和授权监管机构实施此类规则的《加利福尼亚隐私权法案》。

[81] 见欧洲委员会，部长理事会关于算法系统的人权影响的 CM/Rec(2020)1 号建议。

[82] 拟议中的欧洲联盟人工智能法案采用了此一基于风险的方法。自由在线联盟、全球网络倡议和全球伙伴数字提交的材料中表达了对基于风险的条例法规的支持。

[83] 高级专员在 A/HRC/27/37 和 A/HRC/39/29 中澄清了在刑事调查背景下为保护国家安全而采取的监视措施的各项要求，这些要求应当指导这方面的立法。

[84] 欧洲联盟提交的材料。Catelijne Muller, “The impact of artificial intelligence on human rights, democracy and the rule of law”, report to the Council of Europe, Ad hoc Committee on Artificial Intelligence (CAHAI(2020)06-fin)，2020 年 6 月 24 日，第 75 段；联合国教育、科学及文化组织(教科文组织)，“Draft text of the recommendation on the ethics of artificial intelligence”(SHS/IGM-AIETHICS/2021/JUN/3 Rev.2)，2021 年 6 月 25 日，第 26 段。

[85] 见欧洲数据保护委员会和欧洲数据保护监管局，第 5/2021 号联合意见，第 33 段。

对于在某些情况下使用会给人权带来风险的系统，各国需要对其使用和销售进行管制，以防止和减轻在本国境内和境外对人权的不利影响。[86] 在有可能发生不利的人权影响时，应规定强制要求人工监督和决策。[87] 鉴于评估和解决风险可能需要时间，各国还应暂停使用远程实时面部识别等潜在高风险技术，直到确保这些技术的使用不会侵犯人权为止。

46. 各国还应当对监测技术的跨境贸易实行强有力的出口管制制度，以防止此类技术在存在被用于侵犯人权的风险情形下出售，包括以人权捍卫者或记者为目标的风险情形下。[88]

47. 人工智能系统引发的各种风险表明，有必要对人工智能系统的开发、部署和使用进行充分、独立、公正的监督。监督工作可以由行政、司法、准司法和(或)议会监督机关结合开展。[89] 例如，除了数据隐私管理当局之外，消费者保护机构、部门监管单位、反歧视机构和国家人权机构应当构成监督系统的组成部分。此外，专门监督人工智能使用的跨部门监管单位可以帮助设定基本标准，并确保政策和执法的一致性。

C. 人权尽责调查

48. 国家和工商企业应确保在获取、开发、部署和运行人工智能系统时，以及在共享或使用所持涉及个人的大数据之前，进行全面的人权尽职调查。[90] 除了为此类进程提供资源和引导之外，各国也可要求或以其他方式鼓励各公司进行全面的人权尽职调查。

49. 人权尽职调查程序的目的是确定、评估、预防和减轻一个实体可能造成的或可能促成或有直接关联的对人权不利影响。[91] 如果尽职调查过程显示，由于缺乏切实的减轻伤害途径，人工智能的使用与人权不符，就不应当进一步寻求使用这种人工智能。评估人权影响是人权尽职调查程序的一项基本内容。[92] 应当在人工智能系统的整个使用周期进行尽职调查。[93] 应特别注意妇女和女童、男女同性恋、

[86] 数字权利提交的材料。

[87] 见 www.cambridge.org/core/journals/international-and-comparative-law-quarterly/article/international-human-rights-law-as-a-framework-for-algorithmic-accountability/1D6D0A456B36BA7512A6AFF17F16E9B6。

[88] A/HRC/41/35，第 49 段；A/HRC/44/24，第 40 段。在这些报告中，促进和保护意见和表达自由权特别报告员和高级专员还呼吁暂停发放监视技术出口许可证。

[89] 见 https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64。

[90] 联合国人权事务高级专员办事处(人权高专办)的 B-Tech 项目正在制定关于在技术行业落实《工商企业与人权指导原则》的指导准则，包括应对人工智能技术使用对人权所产生影响的措施。见 www.ohchr.org/EN/Issues/Business/Pages/B-TechProject.aspx。

[91] 关于人工智能背景下的人权尽职调查概述，见 https://international-review.icrc.org/articles/ai-humanitarian-action-human-rights-ethics-913，第 174 至 178 页。

[92] 关于人权影响评估方法的简明摘要，见 https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64。

[93] A/HRC/43/29，第 62(g)段，以及 A/HRC/44/24，第 38、53(j)㈠和 54(c)段。

双性恋、跨性别者和性别奇异者、残疾人、属于少数群体者、老年人、贫困者和其他弱势者受到的不成比例影响。

50. 应当与有可能受影响的权利持有人和民间社会进行有意义的协商，同时应让具备跨学科技能的专家参与影响评估，包括参与制定和评估缓解措施。国家和工商企业应持续监测其所使用人工智能系统的影响，以核实这些系统是否对人权产生不利影响。人权影响评估的结果、为应对人权风险而采取的行动以及公开进行的协商本身都应公开。[94]

D. 国家-工商企业关系

51. 在国家和技术公司之间关系密切的情况下，需要予以专门的关注。[95] 国家是一个重要的经济行为者，它可以塑造人工智能的发展和使用方式，而不仅仅是国家在法律和政策措施中扮演角色。在国家与私营部门人工智能开发者和服务提供者合作的情形中，各国应采取额外步骤，确保人工智能不被用于与人权不符的目的。这些步骤应广泛适用于国有公司管理、研发资金、各国向人工智能技术公司提供的资金和其他支持、私有化努力和公共采购活动。

52. 在国家作为经济行为者开展运作的情形下，国家仍然是国际人权法规定的主要义务承担者，必须积极履行其义务。与此同时，工商企业在与国家政府合作时，仍然有责任尊重人权，在面临与人权法相抵触的政府要求时，应设法维护人权。[96] 例如，在面临与人权标准不符的获取个人数据要求时，它们应该利用自身影响力，阻止或减轻可能造成的伤害。[97]

53. 国家政府可以通过连贯一致地要求工商企业负责任行事，强化对人权的保护。例如，出口信贷机构向人工智能技术公司提供支持时，应确保这些公司在尊重权利方面有良好的行为记录，并且可以通过强有力的尽职调查程序证明这一点。

54. 当国家政府依赖人工智能企业提供公共产品或服务时，必须确保它们能够监督人工智能系统的开发和部署。为此可以要求和评估有关人工智能应用程序的准确性和风险的信息。如果风险无法有效缓解，政府不应使用人工智能来提供公共产品或服务。

E. 透明度

55. 人工智能系统的开发者、营销者、运营者和用户应该大幅增强其在人工智能使用透明度方面的努力。作为第一步，国家政府、工商企业和人工智能其他用户应该提供关于其所使用系统的类型和用途以及系统开发者和运营者的身份的信

[94] A/73/348，第 68 段。

[95] 见 www.ohchr.org/Documents/Issues/Business/B-Tech/b-tech-foundational-paper-state-duty-to-protect.pdf。

[96] 《工商业与人权指导原则》原则 23(b)。

[97] A/HRC/32/38，第 58 段。另见 www.ohchr.org/Documents/Issues/Terrorism/biometricsreport.pdf，第 39 和 40 段。

息。[98] 在即将或者已自动或在自动化工具帮助下做出决策时，应当有系统地通知受影响的个人。[99] 当个人提供的本人数据将成为人工智能系统所用数据集的一部分时，也应该通知个人。[100] 对于与人权关系重大的应用，政府可以采用载有人工智能工具及其使用方面关键信息的登记册。[101] 应确保有效落实数据隐私框架中包含的透明度义务以及数据访问、删除和更正权利。应特别注意使个人能够更好地理解和控制所汇编的关于他们的个人资料。[102]

56. 应进一步提高透明度，包括为此持续努力克服上述“黑箱”问题。开发和有系统采用各种方法，使人工智能系统更容易解释——常常称作算法透明度。这对于确保充分保护权利至关重要。[103] 当人工智能被用来决定司法程序中或与实现经济、社会和文化权利所必需的社会服务有关的关键问题时，这一点极为重要。研究人员已经开发出一系列方法，可推进这一目标，因此增加这一领域的投资至关重要。[104] 国家政府还应采取措施，确保知识产权保护不会妨碍对影响人权的人工智能系统进行有意义的审查。[105] 采购规则应当更新，以反映对透明度的需要，包括人工智能系统的可审计性。[106] 尤其是，政府应避免使用可能对人权产生重大不利影响，但又无法进行切实审计的人工智能系统。[107]

五. 结论和建议

A. 结论

57. 本报告强调指出，人工智能技术对行使隐私权和其他人权的影响不可否认，而且稳步增长，其结果有好有坏。它指出了令人担忧的发展，包括一个庞大的生态系统，其中大部分是不透明的个人数据收集和交换，这些生态系统塑造了被广泛使用的部分人工智能系统。这些系统影响政府处理治安和司法事务的方法，决定公共服务的可获性，决定谁有机会被招聘工作，而且也影响人们可在网上看到

[98] A/HRC/43/29，第 52 段；A/73/348，第 49 段。

[99] 欧洲委员会，《关于消除算法系统对人权所产生影响的准则》，(部长委员会向成员国提出的关于算法系统所产生人权影响的 CM/Rec(2020)1 号建议附录)，B 节，第 4.2 段。

[100] A/73/348，第 49 段。

[101] A/HRC/43/29，第 52 段。欧洲联盟关于人工智能法的提案包含了建立高风险人工智能系统登记册的条款。

[102] 见 https://link.springer.com/article/10.1007/s12394-008-0003-1，第 67 段。

[103] 关于算法透明度各要素的概述，见 www.cambridge.org/core/journals/international-and-comparative-law-quarterly/article/international-human-rights-law-as-a-framework-for-algorithmic-accountability/1D6D0A456B36BA7512A6AFF17F16E9B6，第 320 至 323 页。

[104] 见 https://arxiv.org/abs/2001.00973 and https://arxiv.org/pdf/1711.01134.pdf。

[105] 欧洲委员会，《关于消除算法系统对人权所产生影响的准则》，(部长委员会向成员国提出的关于算法系统所产生人权影响的 CM/Rec(2020)1 号建议附录)，B 节，第 4.1 段。

[106] 见德国、数字权利、自由在线联盟和全球合作伙伴数字提交的材料。

[107] A/73/348，第 55 段，以及 A/HRC/43/29，第 54 段。

和分享哪些信息。此外，与基于人工智能的决策相关的歧视风险是实际存在的。本报告概述了解决人工智能所涉根本问题的一系列途径，强调只有全面的基于人权的方法才能确保可持续解决问题，造福所有人。

58. 然而，鉴于人工智能背景下出现的新问题多种多样，本报告仅简单概述不断演变的人工智能格局。需要进一步分析的领域包括卫生、教育、住房和金融服务。生物识别技术正日益成为国家政府、国际组织和技术公司的首选解决方案，这是一个迫切需要进行更多人权指导的领域。此外，从人权角度来看，今后工作的一个重点应该是设法填补全球数据环境中的巨大问责差距。最后，应当紧急确定并实施克服人工智能所致歧视的解决方案。

B. 建议

59. 高级专员建议国家政府：

(a) 充分认识到需要在人工智能开发、使用和治理过程中把保护和加强所有人权作为一项核心目标，并确保线上和线下平等尊重和落实所有人权；

(b) 确保人工智能的使用符合所有人权，并确保通过使用人工智能对隐私权和其他人权的任何干涉均有法律依据，用于合法目的，符合必要性和相称性原则，不损害有关权利的实质；

(c) 明令禁止无法在符合国际人权法情况下运行的人工智能应用程序，并暂停销售和使用给享受人权带来高风险的人工智能系统，除非订立适当保障措施来保护人权；

(d) 暂停在公共场合使用远程生物识别技术，至少在主管当局能够证明符合隐私权和数据保护标准和不存在重大准确性问题和歧视性影响之前，以及在A/HRC/44/24 第 53(j)(1-5)段所述所有建议都得到落实之前暂停使用；

(e) 通过独立、公正的主管机构，采纳并有效执行公共和私营部门数据隐私立法，以此作为在人工智能背景下保护隐私权的一个基本前提；

(f) 实施立法和监管框架，充分预防和减轻与公共和私营部门使用人工智能有关的多方面不利人权影响；

(g) 确保与使用人工智能系统有关联的侵犯和践踏人权行为受害者能够获得有效补救；

(h) 要求对所有可能严重影响人权，特别是在公共部门严重影响人权的人工智能所致决定作出充分解释；

(i) 加强努力，抵御与国家政府和工商企业使用人工智能系统有关联的歧视，包括对人工智能系统得出的结论以及其部署后产生的影响进行、要求并支持有系统评估和监测；

(j) 确保在提供和使用人工智能技术方面的公私伙伴关系是透明的，并接受独立人权监督，且不会导致政府推卸对人权的责任。

60. 高级专员建议国家政府和工商企业：

(a) 在其设计、开发、部署、销售、获取或运行的人工智能系统整个使用周期有系统地进行人权尽职调查。其人权尽职调查的一个关键要素应该是定期、全面的人权影响评估；

(b) 极大提高其使用人工智能的透明度，包括充分告知公众和受影响个人，并为自动化系统的独立、外部审计提供支持。与人工智能使用相关的潜在或实际人权影响越有可能、越严重，就越需要更多的透明度；

(c) 确保所有相关利益攸关方参与关于人工智能开发、部署和使用的决策，特别是受影响个人和群体；

(d) 提高基于人工智能的决策的可解释性，包括为实现这一目标提供资金和进行研究。

61. 高级专员建议工商企业：

(a) 尽一切努力履行其尊重所有人权的责任，包括充分实施《工商企业与人权指导原则》；

(b) 强化努力，抵御与其开发、销售或运行人工智能系统有关联的歧视，包括对人工智能系统得出的结论以及其部署后产生的影响进行有系统评估和监测；

(c) 采取果断步骤，确保负责开发人工智能的劳动队伍具备多元性；

(d) 通过合法程序，包括通过有效的业务层面申诉机制，对企业自身造成或促成的不利人权影响提供补救或进行合作。