

# 多伦多宣言：保护机器学习系统中的平等权和不歧视权

## 序言

1. 随着机器学习系统日趋先进，其应用日益增多，我们必须检视这项技术对人权产生的影响。我们承认机器学习及相关系统在促进人权方面的潜力，但也越来越关注这些系统有可能助长针对某些个人或群体有意无意的歧视行为。我们亟需处理这些技术影响人类及其权利的问题。在机器学习系统的世界里，谁会承担侵害人权的责任？
2. 在围绕伦理与人工智能的讨论持续进行的同时，本《宣言》旨在提请大家注意那些相关和完善确立的国际人权法框架及国际人权标准。这些普世、具有约束力且能促成行动的法律及标准提供了具体的方式，保护个人免受歧视；促进包容性、多样性及公正性；并维护平等。人权是“普遍、不可分割、相互依存、相互联系。”<sup>1</sup>
3. 本《宣言》以现有探讨这项技术所带来之害处的讨论、原则及文章为基础，而众多专家在这一领域所做的重要工作，则有助于提升对机器学习系统可能带来之歧视风险的认识，并为有关的讨论提供信息。<sup>2</sup> 我们希望重申人权法及人权标准在任何情况下保护个人与群体免受歧视所起的作用，以补充现有的工作。本《宣言》引用的人权法及人权标准为制定机器学习方面的伦理框架提供了坚实的基础，其中包含有关问责的条文及补救方式。

---

<sup>1</sup> 联合国人权事务委员会，《维也纳宣言和行动纲领》，1993年，<http://www.ohchr.org/EN/ProfessionalInterest/Pages/Vienna.aspx>。

<sup>2</sup> 例如，参见机器学习公平、透明、问责制（FAT/ML），《负责任算法原则及算法社会影响宣言》（Principles for Accountable Algorithms and a Social Impact Statement for Algorithms）；电气电子工程师学会自动化及智能系统伦理问题全球倡议（IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems），《人工智能设计的伦理准则》（Ethically Aligned Design）；《蒙特利尔人工智能负责任发展宣言》（The Montreal Declaration for a Responsible Development of Artificial Intelligence）；以及由生命未来研究所（Future of Life Institute）制定的《阿西洛马人工智能原则》（Asilomar AI Principles）。

4. 比方说，警务工作、福利体系、医疗卫生服务、网络讨论平台等运用了机器学习技术的系统，能够以前所未有的幅度大大和迅速地巩固或改变权力结构，而且对人权造成重大损害，特别是平等权。此外，有越来越多的实质证据显示，在缺乏必要保障措施的情况下引进和应用机器学习系统，可能会导致歧视或压制行为。这种系统一般晦涩难懂，而且包含令人费解的程序。
5. 国家和私营部门行为者应在机器学习和相关技术帮助人们行使与享有人权的情况下，推动其开发及使用。例如，在医疗卫生领域中，机器学习系统可以推进诊断和治疗方面的发展，同时能使医疗卫生服务的提供更广泛和普及。更广泛而言，关于机器学习及人工智能系统方面，各国应促进人们享受科技进步成果这一积极权利，以此作为对经济、社会及文化权利的肯定。<sup>3</sup>
6. 在本《宣言》中，我们着眼于平等及不歧视的权利。另外还有多项人权可能因机器学习系统的应用及不当使用而受到负面影响，其中包括隐私权及数据保护权、言论及结社自由权、参与文化生活的权利、法律面前人人平等及获得有效补救的权利。做决定及处理数据的系统亦有可能损害经济、社会及文化权利；例如，它们可能会影响到医疗卫生与教育等关键服务的提供，并限制就业等方面的机会。
7. 尽管本《宣言》侧重于机器学习技术，但当中的许多规范与原则也同样适用于人工智能的广泛概念之下的科学技术。

---

<sup>3</sup>《经济、社会、文化权利国际公约》，第 15 条，  
<https://www.ohchr.org/EN/ProfessionalInterest/Pages/CESCR.aspx>。

## 目录

序言 .....	1
采用国际人权法框架 .....	4
平等及不歧视的权利 .....	4
防止歧视 .....	5
保护所有人及群体的权利：促进多样性及包容性 .....	5
国家的责任：人权义务 .....	6
国家使用机器学习系统 .....	6
促进平等 .....	9
向私营部门行为者追究责任 .....	9
私营部门行为者的责任：人权尽责调查 .....	10
获得有效补救的权利 .....	12
结论 .....	13

## 采用国际人权法框架

8. 国家有义务促进、保护和尊重人权；包括公司在内的私营部门行为者有责任在任何时候尊重人权。我们提出这项《宣言》，以申明这些义务和责任。
9. 目前，在超国家、国家及地区层面，在科技公司、学术机构、公民社会及以外的地方都有许多讨论，集中于人工智能的伦理问题，以及如何令这一领域的科学技术以人为本。这些问题必须从人权角度分析，以评估这项技术现时和将来在人权方面可能造成或助长的损害，并采取切实的措施应对任何可能导致损害的风险。
10. 人权法是一个基于法治的普世价值观体系，它提供已确立的手段，以确保包括平等及不歧视的权利在内的人权得到维护。由于它是一套有普遍约束力且可促成行动的标准，因此尤其适用于无边无界的科技。人权法设定了相关标准，并提供了相关机制，在公共及私营部门行为者没有履行自身的义务和责任保护及尊重人权时，追究其责任。它亦要求人人必须能够在自身权利被剥夺或侵犯时获得有效补救与救济。
11. 政府以及构思、开发与部署机器学习系统的私营部门行为者必须紧急检视和处理这些系统带来的风险。重要的是，要查明和处理可能导致的损害，并设立相关机制，向造成损害的人追究责任。政府的措施应具有约束力，而且能充分保护和促进人权。学者以及法律与公民社会领域的专家应能充分参与这些讨论，并能就这些技术的使用提出批评和建议。

## 平等及不歧视的权利

12. 本《宣言》着眼于平等和不歧视的权利，而这也是各项人权所依据的重要原则。
13. 按照国际法的定义，歧视是指“任何基于种族、肤色、性别、语言、宗教、政治或其他见解、国籍或社会出身、财产、出生或其他身份的任何区域、排斥、限制或优惠，其目的或效果为否认或妨碍任何人在平等的基础上认识、享有或行使一切

权利和自由。”<sup>4</sup> 以上的定义并非详尽无遗，联合国人权事务高级专员早已确认有**必要防止**针对其他阶层的歧视行为。<sup>5</sup>

## 防止歧视

14. 政府有义务，而私营部门行为者亦有责任，积极预防歧视的发生，以遵守现有的人权法及人权标准。当预防措施不充分或未尽人意，并出现歧视行为时，该制度应当受到审查，且应立即处理所造成的损害。
15. 在部署新技术的过程中，国家及私营部门行为者皆有可能因为有新挑战（包括关于平等、不同个人及群体的代表性、对不同个人及群体的影响）出现，而需要探寻新的方法来保护人权。
16. 在涉及这些技术所特有的情况中，现有结构性歧视的模式可能会再现或恶化，例如，设定了自行实现的成功标志和加剧不平等模式的机器学习系统目标，或是使用不具代表性或偏倚的数据集所产生的问题。
17. 所有公共和私营行为者必须在设计、开发及运用机器学习技术的过程中，防止和减少歧视出现的风险。他们亦须确保，在部署这一系统前和使用系统的整个生命周期中，设有机制让人们能够有机会获得有效补救。

## 保护所有人及群体的权利：促进多样性及包容性

18. 本《宣言》指出，包容、多元及公正是保障和维护平等权及不歧视权的关键要素。在开发和部署机器学习系统的过程中，必须考虑所有这些要素，以防止歧视出现，尤其是针对边缘化群体的。
19. 尽管数据搜集或有助于减少歧视，但对某些群体而言，收集有关歧视的数据尤其会带来困难，因此必须对这些群体提供额外的保护，包括保护敏感数据。

---

<sup>4</sup> 联合国人权事务委员会，《第 18 号一般性意见》，联合国文件编号：HRI/GEN/1/Rev.9 Vol.I (1989 年)，第 7 段。

<sup>5</sup> 联合国人权事务高级专员办事处，《处理针对男女同性恋、双性恋、跨性别者及双性人的歧视：商业行为标准》（Tackling Discrimination against Lesbian, Gay, Bi, Trans, & Intersex People Standards of Conduct for Business），<https://www.unfe.org/standards/>

20. 在设计过程中存有隐性及无意识的偏见，也会造成歧视。在此过程中，机器学习系统的构思、开发及终端应用主要由社会上某一阶层监督。目前，这项技术主要由某些国家及地区的企业开发、运用及审核；这项技术背后的人带着自身的偏见，同时，他们从不同种族、文化、性别及社会经济背景的群体所收到的意见可能非常有限。
21. 所谓包容、多元及公正，是指在设计和运用机器学习系统的过程中，让包括终端用户在内的不同社群积极参与并与之进行有意义的协商，以便确保这些系统以尊重人权的方式开发和使用，尤其是涉及易受歧视的边缘化群体的权利。

## 国家的责任：人权义务

22. 国家负有促进、保护、尊重和实现人权的主要责任。根据国际法，国家在公共体系中或通过公私合作方式设计或实施机器学习系统时，不得参与或支持歧视或其他侵犯权利的行为或做法。
23. 国家必须遵守相关的国内及国际法律法规（例如数据保护法及隐私法），当中写入与落实保护个人或群体免受歧视及其他权利侵害的人权义务。
24. 国家负有积极义务，包括利用具有约束力的法律，保护个人或群体免受私营部门行为者的歧视，同时促进平等及其他权利。
25. 本部分所列举的国家义务亦适用于公共与私营部门行为者合作使用机器学习的情况。

## 国家使用机器学习系统

26. 国家必须确保，在考虑到机器学习技术所带来的风险和应对有关风险的情况下，更新现有防止歧视及其他侵害权利现象出现的措施。

27. 机器学习系统日益被公共机构应用或实施于各个领域，对于行使和享有人权、法治、正当程序、言论自由、刑事司法、医疗卫生服务、获取社会福利及住房非常重要。尽管这项技术可以给这些领域带来益处，但亦极有可能导致歧视或其他权利侵害后果。因此，至关重要的是，国家在这些问题出现时提供切实的机会，对所造成的损害进行有效补救及救济。
28. 人权事务委员会申明，《公民权利和政治权利国际公约》第 26 条“禁止公共当局管理和保护的领域中法律上或事实上的歧视。”<sup>6</sup> 其他涉及特定形式歧视的国际公约也进一步阐述了这一点，在这些公约中，国家承诺不参与歧视行为，并确保公共机关及机构“遵守此项义务行事”。<sup>7</sup>
29. 国家必须彻底避免利用或要求私营部门使用具有歧视性、导致歧视后果或侵害人权的工具。
30. 国家必须采取下列措施，缓解和减少公共部门系统中机器学习带来的歧视损害：

## 一. 识别风险

31. 任何部署机器学习技术的国家必须在开发或采购前，并尽可能在使用前及持续在这些技术使用的整个生命周期中，在其部署的范围内彻底调查是否有歧视及其他权利侵害的风险，其中可包括：
- a) 政府在采购前和开发期间，在各个重要阶段，以及在部署和使用机器学习系统的整个过程中，定期进行影响评估，以识别可能造成歧视或其他权利侵害后果的因素，例如：是在算法模型的设计、监管程序抑或数据处理环节；<sup>8</sup>

<sup>6</sup> 联合国人权事务委员会，第 18 号一般性意见（1989 年），第 12 段。

<sup>7</sup> 例如，《消除一切形式种族歧视国际公约》，第 2 条 a 款，以及《消除对妇女一切形式歧视公约》，第 2 条 d 款。

<sup>8</sup> 人工智能研究所（AI Now Institute）为公共机构进行算法影响评估提供了一个实用框架，<https://ainowinstitute.org/aiareport2018.pdf>。欧盟的《通用数据保护条例》（General Data Protection Regulation）第 35 条规定需要开展数据保护影响评估（Data Protection Impact Assessment）；此外，《通用数据保护条例》第 25 条规定，从产品或服务的构思阶段起并在其作用周期始终，设计和预设（by design and by default）采用数据保护原则。

- b) 采取适当措施，减缓在影响评估中识别到的风险，例如：减少无意的歧视或是数据或系统中的代表不足问题；采用动态测评的方法，以及在发布前进行测试；确保可能受影响的团体及领域内的专家能够在设计、检测及审查阶段积极参与并享有决定权；在适当时候将系统交付独立专家检视；
- c) 对系统进行实时、定期的检测与审计；检视成功标志，以防偏见及自动完成的反馈循环出现；并确保就现实环境中发生的人权侵害对系统进行全面、独立审核；
- d) 披露有关系统已知的局限，例如：录得的成功率、已知的失效情况，以及适当的使用限制。

## 二. 确保透明度和问责性

32. 在公共部门使用机器学习系统时，国家必须确保和要求问责及最大可能的透明度。其中须包括清楚说明这些技术如何运作，以便独立机构能够有效检视相关个人及团体所受的影响，确定责任归属，并向肇事者追究责任。国家应当：
- a) 公开披露哪些公共领域使用了机器学习系统；提供信息，以清晰易懂的方式解释自动化及机器学习的决策过程是如何达成；并记下在识别、记录及减少歧视或其他权利侵害影响方面所采取的措施；
  - b) 使用可审计的系统，以便能够进行独立的分析和监管；
  - c) 避免使用无法符合有效问责及透明性标准的“黑箱系统”，并在高风险的情况下完全停止使用这些系统。<sup>9</sup>

## 三. 执行监管

33. 国家必须采取措施，以确保政府官员了解和注意到机器学习系统中的歧视及其他权利侵害风险。国家应：
- a) 积极推行不同的雇用政策和进行协商，以确保纳入多样化观点，使参与设计、实施和检视机器学习的人来自不同背景与身份；
  - b) 确保公共机构对参与采购、开发、使用及检视机器学习工具的官员开展有关人权及数据分析的培训；

---

<sup>9</sup> 纽约大学人工智能研究所，《人工智能研究所 2017 年报告》，  
[https://ainowinstitute.org/AI\\_Now\\_2017\\_Report.pdf](https://ainowinstitute.org/AI_Now_2017_Report.pdf)。



- c) 设立独立监督机制，包括在必要时由司法机构进行审查；
  - d) 确保支持机器学习的决定符合国际认可的正当程序标准。
34. 鉴于机器学习系统的研发工作主要由私营领域推动，各国实际上往往依靠私人承包商在公共系统中设计并应用这些技术。在此情况下，国家不得背弃其在不同方面的义务，包括防止歧视出现，以及在提供这些服务的过程中确保就人权侵害行为实行问责与救济。
35. 任何从私营领域采购机器学习技术的国家机关均应对系统的使用进行相关监督和管理，并要求第三方开展人权尽责调查，以识别、防止和减少歧视及其他人权侵害，并公开说明其在这一方面所做的工作。

## 促进平等

36. 国家有责任采取积极措施消除歧视。<sup>10</sup>
37. 在机器学习及广大科技发展方面，国家的其中一个当务之急是推广那些在科学、技术、工程及数学领域（通常被称为 STEM 领域）提升多样性、包容性及公正性的计划。这些措施本身并非最终结果，但可能有助于减缓歧视性的后果。国家亦应投入资金，研究有哪些方式可以减少机器学习系统中的人权侵害。

## 向私营部门行为者追究责任

38. 国际法明确规定国家有责任保护人权，包括确保私营部门行为者尊重不歧视的权利。

---

<sup>10</sup> 联合国经济、社会和文化权利委员会申明，除了避免采取歧视性行动外，“国家还应采取具体、谨慎及有针对性的措施，确保消除落实《公约》权利方面的歧视。”——联合国经济、社会和文化权利委员会，《第 20 号一般性意见》，联合国文件编号：E/C.12/GC/20 (2009 年)，第 36 段。

39. 根据联合国经济、社会和文化权利委员会，“缔约国必须采取包括立法在内的措施，确保私人领域的个人及实体不因禁止的理由而歧视。”<sup>11</sup>
40. 国家应制定符合人权法的法规，以便在可能造成歧视或其他权利侵害后果的情况中对私营领域使用机器学习实施监督，并认识到技术标准可以与法规相辅相成。此外，在国家及区域层面，关于不歧视、数据保护、隐私及其他领域的法律可以进一步阐释，并加强适用于机器学习的国际人权义务。
41. 国家必须保障所有因这些技术的使用而权利受到侵犯或侵害的人有机会获得有效补救。

## 私营部门行为者的责任：人权尽责调查

42. 私营部门行为者有责任尊重人权；这一责任独立于国家义务之外。<sup>12</sup> 作为履行这责任的一部分，私营部门行为者需持续采取积极主动的措施，以确保其不会造成或助长人权侵犯，此程序被称为“人权尽责调查”。<sup>13</sup>
43. 开发和部署机器学习系统的私营部门行为者应当遵循人权尽责调查的框架，以免在使用其系统的过程中造成或加深歧视，并更广泛地尊重人权。
44. 人权尽责调查程序有三大核心步骤：
- i. 识别可能出现的歧视后果；
  - ii. 采取有效措施，以防止和减少歧视，并追踪应对措施；
  - iii. 就识别、防止和减少机器学习系统中的歧视行为所采取的措施保持公开透明。

<sup>11</sup> 联合国经济、社会和文化权利委员会，《第 20 号一般性意见》，联合国文件编号：E/C.12/GC/20 (2009 年)，第 11 段。

<sup>12</sup> 参见《联合国工商业与人权指导原则》及附加支持文件。

<sup>13</sup> 参见欧洲委员会部长理事会就互联网中间商的角色与责任对成员国的第 CM/Rec(2018)2 号建议 ( Council of Europe’s Recommendation CM/Rec(2018)2 of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries ) ， [https://search.coe.int/cm/Pages/result\\_details.aspx?ObjectID=0900001680790e14](https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=0900001680790e14)。

## 一. 识别可能出现的歧视后果

45. 在开发和部署任何新的机器学习技术时，非国家及私营部门行为者应评估该系统带来的歧视风险。运用于不同用途时，所导致的歧视风险及歧视所造成的损害亦不尽相同，而应对歧视所需的措施取决于具体情况。各方必须谨慎行事，不仅要确认直接歧视，亦需识别间接形式的差别待遇，因它表面看来或许不偏不倚，却会导致歧视。
46. 在勘查风险时，私营部门行为者应当考虑通常与机器学习系统有关的风险，例如：建立在不完整或缺乏代表性的数据之上的培训系统，或体现了长期或系统性歧视的数据集。私营行为者需以包容的方式咨询利益攸关方，包括受影响群体；致力于人权、平等及反歧视的组织；以及独立人权专家和机器学习专家。

## 二. 采取有效措施，以防止和减少歧视，并追踪应对措施

47. 在识别人权风险后，第二步则是防止这些风险的出现。机器学习系统的开发者需要：
  - a) 在设计模型与决定使用哪些数据时，以及针对系统造成的影响方面，纠正歧视问题。
  - b) 在机器学习开发团队中贯彻多元化、公平公正及包容理念，以便有计划地辨识偏见，并防止出现无意的歧视。
  - c) 将极有可能导致侵犯人权行为的系统交予独立第三方审计。
48. 私营部门行为者经过评估认为出现歧视或其他权利侵犯问题的风险极高，或无法缓解这些风险时，不应在此情况下部署机器学习系统。
49. 这一步骤的另一重要一环是，私营部门行为者追踪自身所采取的应对措施（针对应用过程中及随着时间推移而出现的问题），其中包括评估其应对措施的有效性。这要求其在设计、测试及部署阶段定期及持续进行质保检查，并开展实时审计工作，以监测系统在各种情况下及实际产生的歧视性影响，并视情况而定，纠正错误与弥补损害。鉴于反馈循环有可能加剧和加深歧视性后果，这一环尤为重要。

### 三. 就识别、防止和减少机器学习系统中的歧视行为所采取的措施保持公开透明

50. 透明度是人权尽职调查的关键要素，涉及“沟通，向受影响的个人或团体，以及向其他相关利益攸关方，提供透明化及问责措施。”<sup>14</sup>
51. 开发和实施机器学习系统的私营部门行为者应公开其识别风险的程序、已确认的风险，以及在防止及缓解已确认之人权风险方面所采取的切实措施，其中可包括：
- a) 披露企业已经确认的风险及歧视的具体实例，例如：与某个机器学习系统的设计方式相关之风险，或在特定情况下使用机器学习系统涉及的风险；
  - b) 在可能出现歧视的情况下，公布机器学习及其功能的技术规范细节，包括培训所使用的数据样本及数据源之细节；
  - c) 建立机制，以确保在使用机器学习系统的过程中出现歧视时，包括受影响个人在内的相关方得知所造成的损害，以及他们能如何对决定或后果提出异议。

## 获得有效补救的权利

52. 获得司法公正的权利是国际人权法的关键要素。<sup>15</sup> 根据国际法，人权遭到国家或非国家行为体侵犯的受害人须及时获得有效的补救，而对侵犯行径负责的人须被追究责任。
53. 设计和实施机器学习系统的企业及私营部门行为者应采取行动，以确保个人及团体能够有机会获得有意义且有效的补救与救济。例如，这可以包括：基于对个人或社会产生的负面影响建立明确、独立、明显的补救程序，以及在机构中指定负责对此类问题迅速提供补救之岗位，各方应当能够对任何补救的决定进行有效的申诉，而这些补救的决定必须接受司法审查。

<sup>14</sup> 联合国《联合国工商业与人权指导原则》，原则 21。

<sup>15</sup> 例如，参见：《世界人权宣言》，第 8 条；《公民权利和政治权利国际公约》第 2 条第 3 款；《经济、社会、文化权利国际公约》，第 2 条；经济、社会和文化权利委员会，《第 3 号一般性意见：缔约国义务的性质（《公约》第 2 条第 1 款）》，联合国文件编号：E/1991/23 (1990 年)；《消除一切形式种族歧视国际公约》，第 6 条；《消除对妇女一切形式歧视公约》，第 2 条；以及联合国经济、社会和文化权利委员会，《第 9 号一般性意见：〈公约〉在国内的适用》，E/C.12/1998/24 (1998 年)，<http://www.refworld.org/docid/47a7079d6.html>。

54. 若使用的机器学习系统危及个人权利，可能会对确保人们获得补救的权利构成挑战。由于某些系统不透明，这意味着个人可能不知道那些影响他们权利的决定是如何被做出的，以及这一程序是否具有歧视性。在某些情况下，涉及的公共机构或私营部门行为者本身可能无法解释有关的决策程序。
55. 当提出各种决定的建议、作出或执行决定的机器学习系统在司法系统中使用时，挑战尤为严峻，因为司法系统正是负责保障各项权利，包括获得有效补救的权利。
56. 在识别、记录和应对歧视问题上所提出的措施，以及就这些措施保持公开透明并实施问责，将有助国家确保个人有机会得到有效补救。此外，国家应当：
- a) 确保若公共部门部署机器学习系统时，其使用须遵循正当程序的标准；
  - b) 鉴于机器学习系统对公正审判及诉讼当事人的权利构成的风险，在司法部门使用这一系统时要谨慎行事；<sup>16</sup>
  - c) 明确列出开发及实施机器学习方面的责任，并阐明哪些机构或个人对这些系统的使用过程中所做的决定负有法律责任。
  - d) 为受害人（在公共或私营机构使用机器学习系统期间出现歧视行为而蒙受损害）提供有效补救，包括采取弥补措施，在适当情况下，这可包括赔偿、对责任人实施制裁，并确保不再出现同样问题。利用现有法律法规有可能可行，但也有可能需要制定新的法律法规。

## 结论

57. 本《宣言》的签署者呼吁公共及私营部门行为者履行其在人权法及人权标准之下所承担的义务和责任，尽可能避免在使用机器学习系统的过程中造成歧视。在出现歧视行为时，必须有措施实现让人们获得有效补救的权利。
58. 我们呼吁各国及私营部门行为者通力协作，在保护个人及团体不受歧视方面扮演积极的角色。在开发和部署机器学习系统时，他们必须按照其在国际人权法与国

---

<sup>16</sup> 例如，参见：Julia Angwin、Jeff Larson 及 Surya Mattu，ProPublica，《机器的偏见》（Machine Bias），2016年，<https://www.propublica.org/article/machine-bias-risk-assessments-incriminal-sentencing>。

际人权标准下所承担的义务及责任，**采取切实措施**以加强问责及促进人权，**当中包括但不限于平等及不歧视**的权利。

59. **科技进步不得损害我们所享有的人权。我们正处于十字路口，掌权者必须立即采取行动以保护人权，并帮助保障当下我们所有人及我们的子孙后代所享有的权利。**

### 起草委员会成员

国际特赦组织：Anna Bacciarelli 及 Joe Westby

即时访问（Access Now）：Estelle Massé、Drew Mitnick 及 Fanny Hidvegi

尼日利亚范式倡议（Paradigm Initiative Nigeria）：Boye Adegoke

隐私国际（Privacy International）：Frederike Kalthener

数字亚洲中心（Digital Asia Hub）：Malavika Jayaram

研究员 Yasodara Córdova

康奈尔大学（Cornell University）：Solon Barocas

人权数据分析小组（The Human Rights Data Analysis Group）：William Isaac

本《宣言》由国际特赦组织及即时访问于 2018 年 5 月 16 日发布，并于 2018 年在加拿大多伦多召开的世界数字人权大会 RightsCon 上公布。