

超越南北差距的AI治理之路： 促进民主和分配诚信的行动计划

报告 | 2022年3月





巴黎和平论坛为法国于2018年发起的一项倡议，旨在在巴黎创建一个致力于开发协调性、建立规则和培养能力以解决全球治理问题的多方平台。全年中，巴黎和平论坛与来自世界各地的行动者携手合作，加强全球公域的治理，包括气候、公共卫生、外太空和数字问题

巴黎和平论坛的年度项目汇集了各国国家元首、政府首脑和国际组织，以及民间社会和私营部门的领导人，就改善全球治理的具体解决方案展开讨论。



软件企业Initiate：社会数字权使全球性的多方利益相关者进程得以建立，以解决算法治理所带来的问题。这项工作吸引了来自欧洲、北美、特别是南半球的政府和决策者、公民社会组织、科学家和工程师。

本组织正在努力为国际数字权利框架制定可能的路径，让公民社会参与确定自动化技术及其应用的轨迹，并帮助制定加强世界各地集体权利的监管方法。

软件企业Initiate是一个位于马萨诸塞州剑桥市的非营利组织，由巴黎和平论坛孵化，并得到了Luminate的慷慨资助。



超越南北差距的AI治理之路： 促进民主和分配诚信的行动计划¹

报告 | 2022年3月

¹ 本文认为，南北方的划分并不严密，应将地缘政治和地缘经济实力置于后殖民发展的历史中。

关于作者

本文内容来自于专家工作组21名成员的讨论和磋商，该工作组探讨了南、北半球在算法治理方法上的趋同和分歧。该小组在2022年下半年定期召开会议。其成员包括

编写委员会

Anita Gurumurthy	IT for Change (印度)
Chioma Agwuegbo	TechHer (尼日利亚)
Nandini Chami	IT for Change (印度)
Virgilio Almeida	哈佛大学伯克曼·克莱因互联网与社会中心

工作小组

Abdul-Hakeem Ajjola	国家网络安全政策和战略审查委员会 (尼日利亚)
Abhishek Gupta	蒙特利尔人工智能伦理研究所、微软、绿色软件基金会
Anita Gurumurthy	IT for Change (印度)
Baobao Zhang	雪城大学马克斯韦尔公民与公共事务学院
Chioma Agwuegbo	TechHer (尼日利亚)
Derya Özkul	牛津大学难民研究中心
Emre Eren Korkmaz	牛津大学国际发展部
Francisco Brito Cruz	因特网实验室 (巴西)
Jean F. Queralt	IO Foundation (马来西亚)
Juan Carlos Lara	Derechos Digitales (智利)
Jun-E Tan	独立研究员 (马来西亚)
Malavika Jayaram	亚洲数据中心 (香港)
Moussa Kondo	Accountability Lab (马里)
Nagla Rizk	开罗美国大学
Nandini Chami	IT for Change (印度)
Nighat Dad	Digital Rights Foundation (巴基斯坦)
Peggy Hicks	联合国人权事务高级专员办事处
Roxana Radu	牛津大学布拉瓦尼克政府学院
Shashi Jayakumar	国家安全卓越中心(新加坡)
Urvashi Aneja	Digital Futures Lab (印度)
Virgilio Almeida	哈佛大学伯克曼·克莱因互联网与社会中心

研究团队

Sadhana Sanjay	IT for Change (印度)
Tanay Mahindru	IT for Change (印度)

秘书处

Adrien Abécassis	巴黎和平论坛
Merih Angin	Koç University (土耳其)
Elina Noor	Initiate: 社会数字权
Jack Loveridge	Initiate: 社会数字权

目录

执行摘要	6
建议	8
I. 处于十字路口的人工智能治理：分散与协调	9
II. 公民公共空间的侵蚀：为什么人工智能治理需要范式转变	12
III. 人工智能经济中的正义：公平价值分配和对发展的影响	16
IV. 人工智能宪政：推进数字经济治理方面的合作	19
结语	21
参考书目	24

执行摘要

人工智能(AI)改变世界的速度，超过了世界缓解日益加剧的地缘政治分歧和社会经济差距的速度。由于技术变革的速度超过了监管政策的速度，目前还没有一个共同的平台来协调不同国家间不同的治理方法。决策者必须优先考虑南半球国家公民和公民社会的关切和利益——广而言之，即拉丁美洲、加勒比地区、非洲、中东、南亚和中亚以及亚太地区的后殖民国家，以扭转全世界算法平台和人工智能系统治理日益分化的局面。我们必须特别关注各国政府和跨国公司部署此类系统以监控、管理和操纵全球南方的公民公共空间的各种方式。

南半球是人工生成数据的主要来源，也是复杂计算机网络和人工智能系统所依赖的原材料。因此，南半球社会有权获得公平的经济利益和保护，免受强大平台和工具的影响，这些平台和工具主要由北半球的公司和大国控制，特别是美国、中国及欧盟。这种公平必须基于我们所定义的“AI宪政”，即将人工智能和大数据作为现代经济中的基本资源，就像电力和水一样，这是21世纪经济和社会发展的重要组成部分。



我们将正在进行的人工智能伦理讨论转向以权利为基础的范式，并针对各国国情制定具体的政策原则，这为国际治理框架提供了最好的前景，使南半球的利益与大国的利益处于平等地位。根据倡议组织“社会中的数字权利”(Digital Rights in Society)和巴黎和平论坛在2021年下半年定期召集的21人专家工作组的讨论，本文提出，人工智能的宪政主义和基于权利的方法应指导制定高级别国际协议和公约，为全球人工智能的开发和部署制定决策标准。

无论是在国际还是在国家层面，都必须制定特殊保护措施，以保障南半球公民的公共空间，并在部署人工智能服务之前，概述评估其潜在影响的最佳做法。更概括地说，必须建立国际规范和协议，以确保公平分配人工智能平台和算法系统带来的利益，避免南半球陷入依赖性循环中。我们必须在南半球和新兴人工智能大国之间寻求并建立起“公平价值分配”机制。



建议

为了努力构建一个浑然一体的人工智能政策格局，将南半球国家的利益和关切纳入其中，工作组建议采取几项关键行动*，包括：

- 1** | 基于人权是完整的且不可分割的愿景，开展民主、多层次的对话，以实现人工智能治理的具体监管原则。
- 2** | 建立一个全球数据库，用于追踪和监控人工智能立法对人权和发展的影响，并促进相关政策制定。
- 3** | 创建南半球国家在基于人工智能的服务领域对跨国公司的问责制。
- 4** | 建立激励机制，以留住南半球的年轻人工智能科学家和工程师。

* 有关建议的更多信息见本报告第22页。

I. 处于十字路口的人工智能治理： 分散与协调

当今新兴的人工智能 (AI) 治理格局高度分散。²目前仅有160多套人工智能伦理和治理原则，但没有一个共同的平台将这些不同的倡议结合起来 (Report of the Secretary-General, 2020; Radu, 2021)。私营部门和政府在这些人工智能治理举措中的投入相对均衡，而公民社会组织的代表性则不强 (Ulnicane et al., 2021)。此外，围绕人工智能的规范制定存在着巨大的地域性差异。³

值得注意的是，这些指导方针大多来自富裕的经济合作与发展组织 (OECD) 国家，而南半球的声音仍然缺乏代表性 (Haas et al., 2020)。对现有框架的回顾表明，平等和非歧视、透明度、问责制、安全、社会福利、隐私、人类尊严和自治构成了关于人工智能治理的全球对话中规范性问题的共同核心 (Fukuda-Parr et al., 2021)。

尽管最初有关人工智能治理的对话大多在秘密的情况下进行，而技术专家专注于“机器学习模型、输入和输出”的解决方案 (Aizenberg et al., 2020)，但最近的关键事件⁴为伦理转向铺平了道路。在这一转向中，不仅有技术专家，还有公共政策参与者、公民社会活动家和大型科技公司也积极参与其中。不幸的是，由于缺乏可强制执行的标准和问责措施，人权话语中体现的道德价值往往被作为这些准则中的修辞手段来使用 (Fukuda-Parr et al., 2021)——从而导致一种开放式的、随大流的道德实践。

幸运的是，最近在人工智能治理方面的概念探索反映了必要的技术社会跨学科性，尽管是在少数工业化背景下，例如连接智能自动化和未来工作；公共领域算法、民

² 参见 <https://oecd.ai/repo>。

³ 参见 <https://www.technologyreview.com/2020/09/14/1008323/ai-ethics-representation-artificial-intelligence-opinion>。

⁴ 如剑桥分析丑闻 (2016) 揭开了算法化公共领域对民主的风险；Project Maven和Project Dragonfly的曝光 (2018-19) 提醒广大公众警惕新军工综合体，并对福利系统中的算法歧视和联合国特别报告员Philip Alston对数字福利国家调查 (2019) 越来越感到不安。

主生活、公民权利和数字福利国家 (Gurumurthy et al., 2019)。然而，在人工智能治理生态系统中，如果缺少明确和可执行的义务和承诺的相应制度安排，这种伦理转向的政策影响可能很有限。因此，当务之急是建立一个基于权利的人工智能治理范式⁵，为消费者和公民——尤其是世界各地被卷入人工智能系统的弱势个人和群体——提供可行的补救措施。

在当前的超级资本主义人工智能范式所带来的结构转型中，社会经济不平等和劳资差距的加剧，对全世界大多数人的社会经济权利构成了双重关切。(Acemoglu et al 2020; Bughin et al 2019)。新证据还表明，殖民主义的历史和地理环境构建了人工智能时代的国际政治经济秩序 (Mohamed et al 2020)，对南半球国家人民和发展的权利产生了不可磨灭的影响。在当今的人工智能经济中，大多数发展中国家仅仅是新数据原材料的来源，同时也依靠着北半球的人工智能基础设施和服务 (Feijóo et al., 2020)。关键的是，这些国家也是用于创建和驱动人工智能系统的物质原材料的来源。



⁵ 必须认识到，在体制和监管能力薄弱的国家，基于权利的体制存在局限性。基于权利的观点也可能无法充分处理结构性和集体性伤害。



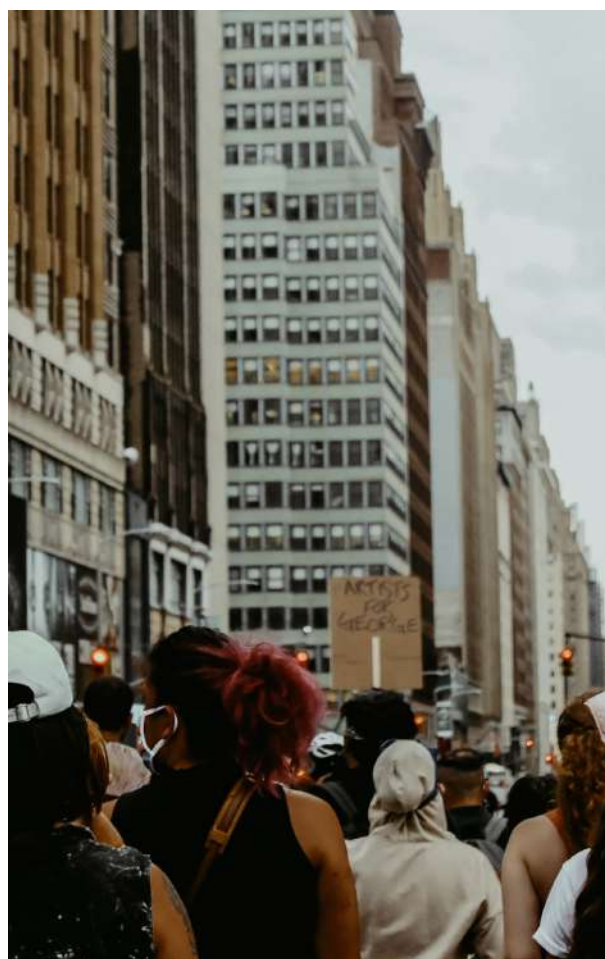
在南北问题的背景下，对算法系统的批评多种多样，包括：算法决策系统压倒性的“白化” (Cave et al, 2020)；在推动人工智能商业模式的跨国数据价值链中，全球劳工等级制度的强化；向立法框架不健全的国家输出可疑的、侵犯权利的人工智能产品测试，都是“算法殖民”的表现 (Mohamed et al., 2020)，体现了在人工智能驱动的国际新秩序中对南半球的剥削和侵占。因此，以权利为中心的人工智能治理系统，必须特别关注在国际发展政治经济中产生的社会经济权利，因为它们出现在国际发展的政治经济中，这些权利跨越所有世代的人权。

在全球数据资本主义下，国家与市场之间的契约是一个重要的政治舞台，在这个舞台上，对公正世界秩序的争论已经出现。本文主张恢复人工智能范式，并将其转向民主和分配诚信，追踪共同关注的问题，并确定南北半球进步公民社会必须关注的隐忧。

II. 公民公共空间的侵蚀： 为什么人工智能治理需要范式转变

人工智能正在改变集体选择的结构，通过这种结构，当代民主产生了社会政策，重塑了国家行使政治权力的方式 (Risse, 2021)。这种转变有可能将更大的权力集中在更少的人手中。自动化的公共领域充斥着虚假信息、仇恨言论、计算式宣传和信息战。大量证据表明，作为社交媒体商业模式核心的用户参与度最大化算法正在放大极端内容和仇恨言论 (Dasgupta, 2021)。社交媒体平台上的仇恨、仇外情绪和煽动暴力的行为正在上升。正如联合国 (UN) 少数群体问题特别报告员在2021年初观察到的，四分之三或更多的在线仇恨言论的受害者是少数群体成员 (Office of the United Nations High Commissioner for Human Rights 2021)。在冠状病毒病 (COVID-19) 全球大流行期间，在线歧视和性别歧视仇恨也如滚雪球般，达到了前所未有的程度 (Dehingia et al 2021)。

在确保迅速删除有害内容方面，依靠人类和人工智能相结合的平台自治表现不佳 (Lyons, 2021)。在这方面，南半球的司法管辖区面临着额外的风险。脸书 (Facebook) 吹哨人弗朗西斯·豪根于



2021年9月通过《华尔街日报》发布的内部文件表明，该公司没有建立有效的服务条款和条件，没有修订现有的商业模式，也没有投资开发AI系统来过滤发展中国家的当地语言仇恨言论和错误信息，即使内部团队已经将其列为高风险内容 (Elliot et al., 2021)。然而，脸书绝非全球北方公司中因看似对北半球公民和公民社会有害的算法和做法而面临审查的唯一公司。

政府和政党的社交媒体操纵和数字监控策略，也是破坏数字媒体论坛公共话语的罪魁祸首(Neudert et al., 2019)。牛津互联网研究所2019年的一项研究显示，在45个民主国家，政客和政党部署了网络宣传，传播被操纵的媒体，以积累虚假的追随者争取选民支持 (Bradshaw et al., 2019)。此外，以色列网络武器公司NSO Group的Pegasus间谍软件至少从2011年开始在全球部署，对政治家、记者和活动家进行监视，其动机各异，并产生了广泛的不利影响 (Marczak et al. 2018)。这些案例揭示了数字系统的广泛脆弱性，并应揭示出算法(尤其是人工智能平台)可能如何被不受约束的政府 and 不法行为者滥用。

此外，人工智能监控技术的滥用不仅仅局限于不自由的国家。卡内基 (Carnegie) 的2019全球人工智能监测指数对全球176个国家进行了调查，结果显示，包括51%的先进民主国家在内的75个国家正进行人工智能监测。研究显示，56个国家部署了智能城市/安全城市平台，64个国家推出了面部识别系统，52个国家采用了智能警务实践 (Feldstein, 2019)。执法机构在没有保障措施的情况下部署面部识别技术，不仅

在南半球——印度 (IFF, 2020)、乌拉圭 (Datysoc, 2020)、巴西 (Network Rights Coalition, 2019 & 2020)和南非 (Lekabe, 2021)——而且同样在北半球——美国(New America, 2021)、英国 (Privacy International, 2021)、欧盟成为争议的主要焦点⁶。尽管欧盟是人工智能治理的“值得信赖的基于人权的方法”的倡导者，但欧盟对执法机构基于人工智能的监控仍然存在很大的顾虑 (Vincent, 2021)。



⁶ 参见 <https://panoptic.in/central/FRT-000025>; <https://reclaimyourface.eu>.

美国和欧盟犯了中国在国际政策论述中经常被批评的毛病——出口的人工智能监控技术可能会威胁其他国家公民和政治自由 (Greco, 2021)。2020年隐私国际 (PI) 的一项研究发现，欧盟一直将援助资金用于在整个非洲大陆建立大规模、高风险的生物识别身份系统，以管理移民流动，但没有进行任何数据保护和人权影响评估 (Privacy International, 2020)。社交媒体上的外国影响力活动是另一个威胁，社交媒体公司至少在七个国家发现了从事此类行为的水军：中国、印度、伊朗、巴基斯坦、俄罗斯、沙特阿拉伯和委内瑞拉 (Bradshaw et al., 2019)。“喷子”农场和机器人的部署使得这种宣传战更难追踪和解决 (Barsotti, 2018)。

联合国极端贫困与人权问题特别报告员2019年的报告指出，在北半球和南半球出现的另一个新问题是福利国家算法化 (Secretary-General, 2019)。在没有考虑公民权利的情况下，对公民进行排名和排序，以确定获得福利的资格：维多利亚时代的贫民窟在数字时代的升级，自动将贫困的公民分为那些“值得”和“不值得”国家慷慨解囊 (Eubanks, 2018)。这一趋势给消费

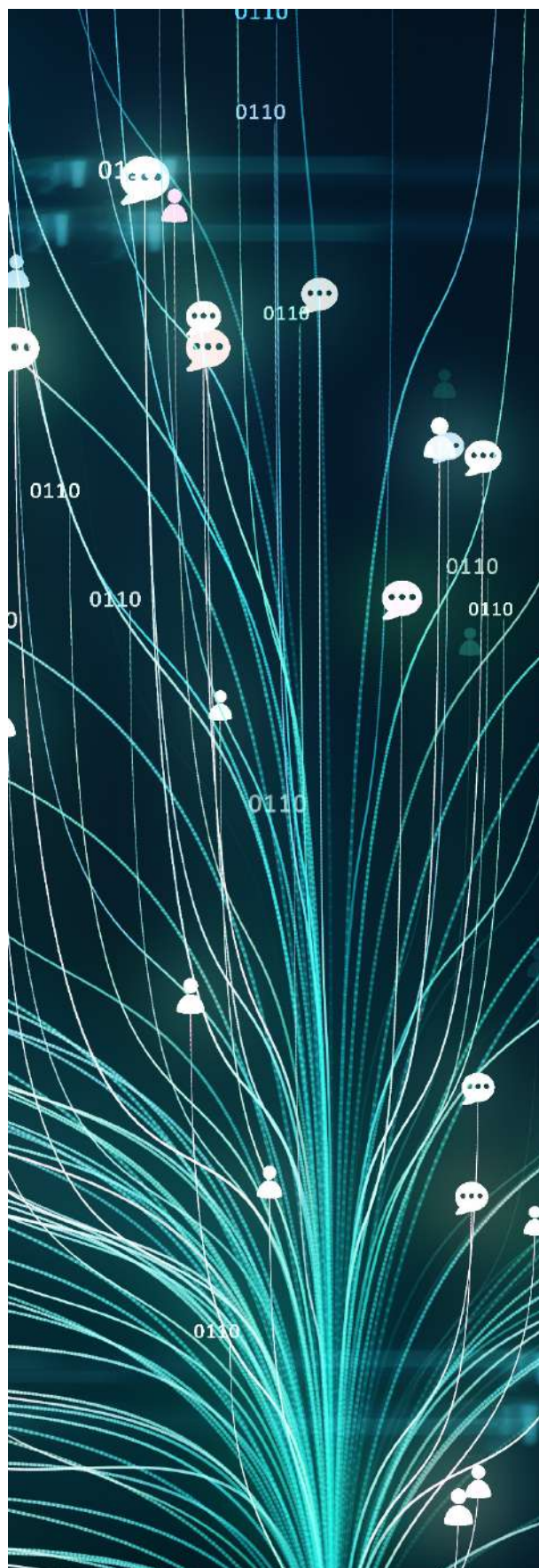


者增加了一层长期的脆弱性，这是由于消费者需要在很长时间内创建和维护一个或多个在线身份，以访问默认数字服务 (Kira et al. forthcoming)。此外，由于政府的人工智能系统经常从北半球引进，在部署时没有考虑地理因素，南半球的公民也处于不利地位 (Secretary-General, 2019)。⁷

⁷ 也有少数例外，如印度在国内使用Aadhaar并出口到其他国家。

由于缺乏关于社交媒体治理的全球协议，在很大程度上使拥有这些平台的公司能够不受惩罚地运作，特别是在南半球。关于在线内容应该如何被节制的《基督城倡议》(Christchurch Call, 2019)也许是在这个问题上最接近全球共识的声明。然而，《基督城倡议》仍然不是一个多边协议，缺乏任何对数字公司具有法律约束力的义务 (Pandey, 2020)。一个被操纵和武器化的网络空间会侵蚀实质性的民主，掩盖国家和市场的勾结，公然无视人权和法治。关于网络安全的国际公约陷入僵局 (Clarke, 2021)，也意味着政治主权和国家安全利益在国际秩序中受到威胁，外国暗中的人工智能信息战正在成为常态 (Ördén et al., 2021)。在国家福利系统中采用人工智能而不对其必要性、相称性和合法性进行适当的测试，可能预示着公民权的危机，而在国际人权基准中没有任何追索或补救措施。

现状表明，目前的体制框架不足以通过对人工智能的意义、使用和限制进行适当的政治调解来保护和培育社会的民主内容。因此，人工智能治理的当务之急是在当前的情况下恢复社会秩序的民主诚信。



III. 人工智能经济中的正义： 公平价值分配和对发展的影响

人工智能对数字时代的影响就像电力对工业革命的影响一样：一种改变范式的通用技术，它的传播带来了生产力的指数级增长。这种指数级的增长来自于生产过程中固定资本和人的能力的增强、劳动力的替代以及产品和服务的创新 (Bughin et al., 2018; Zuboff 2018)。据估计，到2030年，人工智能将为全球经济产出增加13万亿至15.7万亿美元 (Rao & Verweij, 2017)。正如联合国贸易和发展会议 (UNCTAD) 2021年数字经济报告所指出的，如果不控制为这些模型提供的数据，围绕人工智能的商业模式就无法存在 (UNCTAD, 2021)。“智能溢价” (Gurumurthy et al., 2019) 的产生是以对社会数据的不断获取为前提的。⁸这就解释了为什么来自美国和中国的先行数字平台控

制着庞大的数据附件，也在引领人工智能的投资和研究。⁹

如果全球人工智能的采用沿着同样的轨迹继续下去，它可能会扩大业绩差距，不仅是在公司层面和工人个人层面，而且是在国家层面。走在前沿的人工智能公司可能会受益匪浅，到2030年，他们的回报可能会翻倍，而那些后来者将被远远抛在后面 (Bughin et al., 2018)。同样，在工人层面，对于少数拥有数字和认知技能以及在难以实现自动化的任务方面具有专长的知识工人来说，工作和工资的需求可能会增长，但对于大多数从事重复性和低数字技能工作的人来说，需求将萎缩 (Acemoglu et al., 2020)。美国和中国在全球人工智能经济中占主导地位：这两个国家占过去五年人工

⁸ 大规模的数字化，在20世纪90年代随着互联网的发展而扩大，在2000年随着数据中心的发展而升级，已经提供了大量的数据资源。一个知识提取制度——建立在大数据之上——逐渐采用高效的算法，通过捕捉这些开放的数据来源来提取“情报”，主要是为了预测消费者行为和销售广告。知识经济已经演变成一种新颖的资本主义形式，其中对基于数据情报的单边控制是利润的来源。

⁹ 正如UNCTAD (2021) 所观察到的：关于活跃在人工智能领域的初创企业的并购 (M&A)，在2016年1月1日至2021年1月22日期间，有308宗并购交易，价值284亿美元。按同期被收购的人工智能初创企业数量计算，世界前五名是来自美国的大型科技公司，其次是中国的百度（排名第六）和腾讯（排名第八）。苹果排名第一，其次是谷歌和微软。

智能初创企业所有资金的94%以上，拥有70%的世界顶级人工智能研究人员 (UNCTAD, 2021)，占世界70家最大数字平台公司市值的90%，这些公司控制着互联网上相当一部分跨境数据流 (UNCTAD, 2019)。美国和中国的参与者在制定标准的行业机构中也更有代表性，为整个世界创造了对基本技术协议的长期依赖。

随着从物联网中获取的大数据变得至关重要，欧盟、韩国和日本凭借其强大的制造业基础、相关的计算能力和人力资源能力，有很大机会迎头赶上 (UNCTAD, 2021)。赢家很可能在未来几十年的地缘经济和地缘政治中占据主导地位 (Feijóo, et al., 2020)。

获取有效的国内人工智能能力取决于三个因素：大数据、计算能力，以及杰出的人工智能研究人员和工程师。不幸的是，发展中国家在融入用户产生数据流的互联网经济方面处于不利地位，而且转向智能制造的工业能力有限，因此很有可能永远被归入人工智能经济的低价值部分。按照目前的配置，人工智能竞赛有可能将撒哈拉以南非洲和大多数发展中国家抛在后面 (UNCTAD, 2021)，财富将空前地集中在中国和美国的少数公司手中。发展中国家在

历史上享有的“廉价劳动力”的竞争优势可能因此变得完全没有意义 (Lee, 2018)。





在一个细微的层面上，数据所有权集中在大型技术跨国公司手中，助长了它们在南半球经营国家的地方不平等现象。这种数据所有权的不对称性代表着较小的本土初创企业的进入壁垒，并在当地立法基础设施薄弱、竞争法和数据保护法（如果有的话）仍处于起步阶段的情况下助长了市场集中 (Rizk, 2019)。这加剧了不平等，并导致南半球国家中较不富足的人被进一步排除在外。

缺乏全球公认的数据经济资源治理制度将加剧经济集中和加深人工智能范式中的不平等。¹⁰全球经济中的跨境数据流动规则是由少数强国决定的，这些国家的企业将来自世界各地的数据作为商业机密封存起来，(James, 2021) 对这些持有的数据主张事实上的永久所有权 (Fia, 2021)。在这种智能经济中，缺乏数据处理和人工智能能力的南半球国家和社区面临着一个危险的、难以维持的矛盾现象。他们不仅必须放弃对锁在跨国资本人工智能系统中的自身数据的任何要求，而且他们也无法合法地从数据产生的利益中获得公平的份额。这导致了全球数字经济中的严重经济不公平。因此，算法殖民就此被自然化了 (Gurumurthy & Chami, 2021)。

¹⁰ 也就是说，在我们开始建立将数据作为一种经济资源的制度之前，可能还需要建立其他基于权利的制度。

IV. 人工智能宪政： 推进数字经济治理方面的合作

今天，标志着人工智能范式的治理缺陷，已经引发了民主和分配诚信的危机。如果没有一个以人和地球为中心的整体愿景，危机就无法弥补，这并非没有先例。发展人工智能治理路线图需要我们避免技术悲观主义和技术原教旨主义。对人工智能的全面禁令和下意识反应或许不是解决方案 (Schwartz & Sheard, 2021; Paz, 2021)，对假设的技术奇点的过度关注 (Lacker, 2021) 会分散对人工智能治理辩论中权力的关键叙述。同时，一切如常的乐观主义将无法持续。

我们所看到的是，在以人工智能为媒介的社会性中，公民公共空间被不断侵蚀，以及人工智能经济的不公正秩序，这不利于构建一个和平与公正的未来。正如联合国言论自由问题特别报告员 (2018) 所强调的那样，普遍的人权框架可以为演变出有效的人工智能治理方法提供一个强有力的起点。然而，这种必要的转变不能仅仅是对



人工智能时代全球宪政的名义上的重新表述。全球人工智能宪政¹¹必须承载国际社会的愿望，优先考虑构建一个无操纵、拥有自主权、没有剥削和不公的未来社会。它必须以重新平衡的多边主义为基础，建立一个新的情报范式，为国家对自由的义务设定一个高标准。

¹¹ 我们所说的“全球人工智能宪政”是指一套尚未制定的普遍原则，可以作为各国人工智能监管政策和立法的参考。在这种情况下，联合国贸发会议数字经济报告 (2021) 强调，需要一个全球数据流治理新框架，以确定“在什么条件下，谁可以获得数据，用于什么用途” (UNCTAD, 2021)。

人类面临着不负责任的权力滥用人工智能的紧急情况——将双重用途的人工智能技术武器化，对缔约国主权的攻击，以及企业权力不受限制的膨胀，特别是在南半球地区。各国必须立即履行其职责，保护其公民的权利，并维护其对所有人民人权的域外义务。然而，除此之外，所涉及的不仅仅是一个国家的承诺。这是一种现在就采取行动意愿，奠定正确的基础，就人工智能时代维护权利的框架达成共识——并进行适当的调整，以填补国际人权法的空白，并发展实施补救措施的能力。

致力于全球人工智能宪政并不等同于呼吁制定一个普遍的公式来解决人工智能系统设计中的偏见、歧视和不平等。对人权在不同背景下所代表的道德价值的严肃解释仍然需要被采纳。例如，隐私问题以独特的方式出现在南半球，这源于跨越个人信息的自主权、共享数字人工制品的现实和社区身份的复杂和定位概念 (Ahmed et al., 2017)。



结论

由于人工智能在经济繁荣和发展途径方面具有决定性的作用，以及明显的、日益扩大的人工智能鸿沟，特别是在Covid-19大流行之后，迫切需要重新定义人工智能治理的辩论。关于南北半球之间的“人工智能鸿沟”已经有很多讨论。因此，对适当应用人工智能促进社会转型和实现可持续发展目标 (SDGs) 的呼吁，取决于在人工智能范式中解决全球不平等问题的前景。

我们认为，人工智能的全球治理框架不能回避数据政治经济和人工智能在构建世界经济秩序中的新作用。联合国的“人工智能造福人类” (AI For Good) 应公平地分配人工智能范式的收益，以鼓励嵌入当地的发展创新，而不仅仅是将知识和创新从北半球单向地分享到南半球。在国际发展的背景下，人工智能驱动的技术和相关数据的所有权和分配，绝不能让南方国家和目前在该领域占主导地位的大国(中国、美国，及某种程度上的欧盟)之间形成依赖循环。人工智能分配和利益的条件必须有利于南半球。通过突出人工智能治理辩论中政治领域的共同点和分歧，我们的目标是让人们关注公民社会的进步行为者，即活动

家、学者、技术专家，以“人工智能的公平价值”的概念为基础的新议程要素。

本文的目的是明确规范方向和核心原则，以解决人工智能治理范式中的政策碎片化问题，并为以南半球为中心的未来发展奠定基础。这种国际制度的具体内容只能通过对话和辩论来制定——正如《突尼斯信息社会议程》所概述的那样——需要加强各国政府之间的合作，并与所有利益相关者协商 (WSIS Executive Secretariat 2006)。尽管如此，工作组建议采取一些关键行动，将南半球国家的利益和关切纳入其中，具体讨论如下。

工作组建议构建一个浑然一体的人工智能政策格局，纳入南半球的观点。

1. 为国际人工智能治理进行民主的、多层次的对话

人工智能治理的监管原则需要基于完整、不可分割，以及促进民主和分配诚信的人权愿景。这需要一个多层次的对话过程，在国际最高层达成具体的共识，优先考虑处于政治经济秩序边缘的人群的需求和利益，尤其是南半球人民的需求和利益。尽管此类对话由政府主导，但仍需要各方的协商，并让全球到地方各级的各种利益攸关方参与进来，包括私营部门、技术社区、传统发展支持者和数字权利团体。

2. 建立追踪和监测人工智能立法的全球数据库

建立一个关于拟议和实施的人工智能立法的全球数据库，作为跟踪世界各地的最佳做法和监管盲点的首个预警系统，有助于对人工智能政策的人权和发展影响进行持续评估。经合组织的人工智能政策观察站为此提供了一个重要和值得赞赏的开端。¹² 然而，为了各种不同的情况下为政策进程提供信息，并建立公民社会的人工智能审计能力，有必要更加专注于南半球。

3. 针对人工智能服务企业问责的国家措施

南半球国家应强制要求跨国公司提供基于人工智能的服务的事前社会影响评估报告。根据国内法律，跨国公司可能需要向有关当局披露源代码，以保护人权和防止市场滥用。

4. 建立激励机制，以留住南半球的人工智能人才

采取激励措施防止早期职业人工智能科学家和工程师流向北半球，对于发展中国家跨入人工智能范式至关重要。确保全球科技公司投资于南半球的国内研发设施的结构性的措施，对于提升当地的人工智能能力和促进人工智能发展生态系统的多样性极为重要。

不可或缺的是，明确和一致拒绝反公民滥用人工智能技术、算法系统和软件的行为。利用人工智能在世界各地侵犯公民权利和民主公共话语的前景是一个真正的威胁。特别是在南半球国家的成果方面，这些涉及人工智能的关键网络安全问题必须在最高级别得到解决，目的是通过现有的政府间组织和国际机构（如联合国），公约、议定书和其他有约束力的程序，将保护措施编纂成法典。

在没有治理的情况下，人工智可能有加剧社会经济不平等、侵蚀公民权利、破坏世界各地和平与经济风险。如果以合作的方式开发、监测和应用，并符合广泛认可的人权标准，人工智能就有可能刺激可持续的经济增长，减少不平等现象，促进整个世界——南半球和北半球——的持久和平。



参考书目

Acemoglu, D., & Restrepo, P. (2020). The Wrong Kind of AI? Artificial Intelligence and the Future of Labour Demand. *Cambridge Journal of Regions, Economy and Society*, 13 (1), 25–35.

Ahmed, S. I., Haque, M., Guha, S., Rifat, M. R., & Dell, N. (2017). Privacy, Security, and Surveillance in the Global South: A Study of Biometric Mobile SIM Registration in Bangladesh. *the 2017 CHI Conference*.

Aizenberg, E., & van den Hoven, J. (2020, July). Designing for human rights in AI. *Big Data & Society*, <https://doi.org/10.1177/2053951720949566>.

Allmann, K. & Radu, R. (forthcoming). Missteps on the path to digitizing welfare: ‘digital footprints’ as access currency for e-government services. *Global Policy*.

Barsotti, S. (2018). *Weaponizing Social Media: Heinz Experts On Troll Farms And Fake News*. Retrieved from Heinz College of Information Systems and Public Policy: <https://www.heinz.cmu.edu/media/2018/October/troll-farms-and-fake-news-social-media-weaponization>

Bissio, R. (2018). *Spotlight on Sustainable Development 2018*. Civil Society Report, IT for Change.

Bradshaw, S., & Howard, P. N. (2019). *2019 Global Inventory of Organised Social Media Manipulation*. The Computational Propaganda Project at the Oxford Internet Institute.

Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018). *Notes From the AI Frontier: Modeling The Impact Of AI On The World Economy*. Discussion Paper, McKinsey Global Institute.

Bughin, J., Seong, J., Manyika, J., Hämläinen, L., Windhagen, E., & Hazan, E. (2019). *Notes From the AI Frontier: Tackling Europe’s Gap in Digital and AI*. Discussion Paper, McKinsey Global Institute.

Cave, S., & Kanta, D. (2020). The Whiteness of AI. *Philosophy & Technology*, 33, 685-703.

Christchurch Call. (2019). *Christchurch Call*. Retrieved from <https://www.christchurchcall.com/call.html>

Clarke, L. (2021, April 22). *UN countries agreed to a more peaceful cyberspace – but it won’t stop state-sponsored attacks*. Retrieved from TechMonitor: <https://techmonitor.ai/policy/geopolitics/un-countries-cybersecurity-deal-state-sponsored-attacks>

Dasgupta, B. (2021, October 23). *Twitter says its algorithm amplifies right wing political content*. Retrieved from Hindustan Times: <https://www.hindustantimes.com/india-news/twitter-says-its-algorithm-amplifies-right-wing-political-content-101634926182240.html>

Datysoc. (2020, November 17). *Organizaciones de la sociedad civil y académicas expresan su preocupación por reconocimiento facial en el Proyecto de Ley de Presupuesto de Uruguay*. Retrieved from Datysoc: <https://datysoc.org/2020/11/17/organizaciones-de-la-sociedad-civil-y-academicas-expresan-su-preocupacion-por-reconocimiento-facial-en-el-proyecto-de-ley-de-presupuesto-de-uruguay/>

Dehingia, N., Lundgren, R., Dey, A., & Raj, A. (2021). *Trends in online misogyny before and during the COVID-19 pandemic: Analysis of Twitter data from five South-Asian countries*. Center on Gender Equity and Health, University of California San Diego.

Elliot, V., Christopher, N., Deck, A., & Schwartz, L. (2021, October 26). *The Facebook Papers reveal staggering failures in the Global South*. Retrieved from Rest Of World: <https://restofworld.org/2021/facebook-papers-reveal-staggering-failures-in-global-south/>

- Eubanks, V. (2018). *The Digital Poorhouse*. Retrieved from Harper's Magazine: <https://harpers.org/archive/2018/01/the-digital-poorhouse/>
- Feijóo, C., Kwon, Y., Bauer, J., Bohlin, E., Howell, B., Jain, R., . . . Xia, J. (2020). Harnessing artificial intelligence (AI) to increase wellbeing for all: The case for a new technology diplomacy. *Telecommunications Policy*, 44 (6).
- Feldstein, S. (2019). *The Global Expansion of AI Surveillance*. Working Paper, Carnegie Endowment for International Peace.
- Fia, T. (2021). An Alternative to Data Ownership: Managing Access to Non-Personal Data through the Commons. *Global Jurist*, 21 (1), 181-210. <https://doi.org/10.1515/gj-2020-0034>.
- Fukuda-Parr, S., & Gibbons, E. (2021). Emerging Consensus on 'Ethical AI': Human Rights Critique of Stakeholder Guidelines. *Global Policy*, 12, 32-44. <https://doi.org/10.1111/1758-5899.12965>.
- Greco, M. (2021, October 7). *What Biden Needs to Say About Surveillance Tech and Foreign Policy*. Retrieved from Just Security: <https://www.justsecurity.org/78467/what-biden-needs-to-say-about-surveillance-tech-and-foreign-policy/>
- Gurumurthy, A., & Chami, N. (2019). *The Wicked Problem of AI Governance*. Friedrich-Ebert Stiftung India Office. Artificial Intelligence in India.
- Gurumurthy, A., & Chami, N. (2021, May 3). Towards a Global Digital Constitutionalism: A Radical New Agenda for UN75. *Development*, 64, 29-38.
- Gurumurthy, A., & Chami, N. (2021). *Governing the Abstract Object of Data: Towards a Distributive Integrity Framework*. IT for Change.
- Gurumurthy, A., Bharthur, D., Chami, N., Vipra, J., & Anwar, I. A. (2019). *Platform Planet: Development in the Intelligence Economy*. IT for Change.
- Haas, L., & Gießler, S. (2020, April 28). Retrieved from Algorithm Watch: <https://algorithmwatch.org/en/ai-ethics-guidelines-inventory-upgrade-2020/>
- IFF. (2020). *Introduction to Facial Recognition Projects in India*. Retrieved from Internet Freedom Foundation: <https://internetfreedom.in/facial-recognition-in-india-part-i/>
- James, W. (2021, October 22). *G7 countries reach breakthrough on digital trade and data*. Retrieved from Reuters: <https://www.reuters.com/world/g7-countries-agree-shared-position-digital-trade-communique-2021-10-22/>
- Lacker, K. (2021, October 25). *Thoughts On AI Risk*. Retrieved from Kevin Lacker's blog: <https://lacker.io/ai/2021/10/25/thoughts-on-ai-risk.html>
- Lee, K.-F. (2018). *AI superpowers: China, Silicon Valley, and the new world order*. Houghton Mifflin Harcourt.
- Lekabe, T. (2021, April 23). *AfriForum wants assurance on Sanral facial recognition tech*. Retrieved from TheCitizen: <https://www.citizen.co.za/news/south-africa/government/2478929/afriforum-wants-assurance-on-sanral-facial-recognition-tech/>
- Lyons, K. (2021, October 17). *Facebook disputes report that its AI can't detect hate speech or violence consistently*. Retrieved from The Verge: <https://www.theverge.com/2021/10/17/22731214/facebook-disputes-report-artificial-intelligence-hate-speech-violence>

- Marczak, B., Scott-Railton, J., McKune, S., Abdul Razzak, B., & Deibert, R. (2018). Hide and seek: tracking NSO Group's Pegasus spyware to operations in 45 countries. *Citizen Lab Research Report*. No. 113, University of Toronto.
- Mhlambi, S. (2020). *From Rationality to Relationality: Ubuntu as an Ethical and Human Rights Framework for Artificial Intelligence Governance*. Carr Center for Human Rights Policy.
- Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy & Technology*, 33, 659-684. <https://doi.org/10.1007/s13347-020-00405-8>.
- Muchiri, C. W. (2020). *Impact of Artificial Intelligence on IP Policy: Call for Comments- Response from a Global South Perspective*. Strathmore Law School.
- Network Rights Coalition. (2019, November 27). *IGF 2019: Open Letter*. Retrieved from Coalizão Direitos na Rede: <https://direitosnarede.org.br/2019/11/27/igf-2019-open-letter/>
- Network Rights Coalition. (2020, November 17). *Open Letter from Brazilian Civil Society on the occasion of the 15th edition of the United Nations Internet Governance Forum*. Retrieved from Coalizão Direitos na Rede: <https://direitosnarede.org.br/2020/11/17/open-letter-from-brazilian-civil-society-on-the-occasion-of-the-15th-edition-of-the-united-nations-internet-governance-forum/>
- Neudert, L.-M., & Howard, P. (2019, November 18). *Online politics needs to be cleaned up – but not just by Facebook and Twitter, say Oxford academics*. Retrieved from Oxford Internet Institute: <https://www.oii.ox.ac.uk/news-events/news/online-politics-needs-to-be-cleaned-up-but-not-just-by-facebook-and-twitter-say-oxford-academics/>
- New America. (2021, June 3). *Civil Rights Concerns Regarding Law Enforcement Use of Face Recognition Technology*. Retrieved from New America: <https://www.newamerica.org/oti/briefs/civil-rights-concerns-regarding-law-enforcement-use-of-face-recognition-technology/>
- Office of the United Nations High Commissioner for Human Rights. (2021, March 23). *Report: Online hate increasing against minorities, says expert*.
- Ördén, H., & Pamment, J. (2021). *What Is So Foreign About Foreign Influence Operations?* Carnegie Endowment for International Peace.
- Pandey, P. (2020, August 3). One year since the Christchurch Call to Action: A Review. *ORF Issue Brief*(389).
- Paz, A. W. (2021, July 7). *We need to regulate mind-reading tech before it exists*. Retrieved from Rest Of World: <https://restofworld.org/2021/chile-neuro-rights/>
- Privacy International. (2020, November 10). *Here's how a well-connected security company is quietly building mass biometric databases in West Africa with EU aid funds*. (International Privacy) Retrieved from Privacy International: <https://privacyinternational.org/news-analysis/4290/heres-how-well-connected-security-company-quietly-building-mass-biometric>
- Privacy International. (2021, August). *Civil Society Groups: Live Facial Recognition Technology should not be used in public spaces*. Retrieved from Privacy International: <https://www.privacyinternational.org/sites/default/files/2021-08/LFRT%20Open%20Letter%20Final.pdf>
- Radu, R. (2021). Steering the governance of artificial intelligence: national strategies in perspective. *Policy and Society*, 40 (2), 178-193.
- Rao, A., & Verweij, G. (2017). *Sizing the prize What's the real value of AI for your business and how can you capitalise?* PWC.

Report of the Secretary-General. (2020). *Roadmap for Digital Cooperation*. United Nations.

Risse, M. (2021). *Artificial Intelligence and the Past, Present, and Future of Democracy*. Discussion Paper, Carr Center for Human Rights Policy.

Rizk, N. (2019). Artificial Intelligence and Inequality in the Middle East: The Political Economy of Inclusion. *The Oxford Handbook of Ethics of AI*.

Schwartz, A., & Sheard, N. (2021, January 20). *Why EFF Doesn't Support Bans On Private Use of Face Recognition*. Retrieved from Electronic Frontier Foundation: <https://www.eff.org/deeplinks/2021/01/why-eff-doesnt-support-bans-private-use-face-recognition>

Secretary-General. (2019). *Report of the Special rapporteur on extreme poverty and human rights*. Advance Unedited Version.

Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W.-G. (2021). Framing governance for a contested emerging technology: insights from AI policy. *Policy and Society*, 40 (2), 158-177.

UNCTAD. (2021). *Digital Economy Report 2021*.

UNCTAD. (2019). *Digital Economy Report 2019*.

UNCTAD. (2021). *Technology and Innovation Report*.

Vincent, J. (2021, April 21). *EU outlines wide-ranging AI regulation, but leaves the door open for police surveillance*. Retrieved from The Verge: <https://www.theverge.com/2021/4/21/22393785/eu-ai-regulation-proposal-social-credit-ban-biometric-surveillance-exceptions>

WSIS Executive Secretariat. (2006). "Report of the Tunis phase of the World Summit on the Information Society, Tunis, Kram Palexpo, 16-18 November 2005."

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs.