



人权理事会

第四十七届会议

2021年6月21日至7月9日

议程项目3

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

虚假信息与意见和表达自由

促进和保护意见和表达自由权特别报告员伊雷内·汗的报告

概要

本报告中，促进和保护意见和表达自由权特别报告员探讨了虚假信息给人权、民主机构和发展进程带来的威胁。特别报告员肯定道，在数字时代，虚假信息带来了多种复杂因素与挑战，同时认为，国家和公司的应对存在问题，措施不足且有损人权。她呼吁以国际人权框架为坚实基础的多层面多利益攸关方应对，并敦促公司审查自身商业模式，敦促国家重新调整本国应对虚假信息的措施，同时提升自由、独立和多元媒体的作用，在媒体和数字素养方面进行投入，增强个人权能并重建公众信任。



一. 引言

1. 两千多年前，屋大维为了摧毁对手马克·安东尼，发起了一场恶意散布虚假信息运动，最终成为罗马首位皇帝奥古斯都·凯撒。自那样久远的年代起，就有捏造和操纵信息，以此赢得战争、推进政治野心、复仇、伤害弱势者和谋取经济利益的行为。
2. 虚假信息并非新生事物。新的是数字技术，它为各种行为体提供了途径，使它们得以出于政治、意识形态或商业动机，以前所未有的规模、速度和范围创造、传播和放大虚假或被操纵的信息。线上虚假信息与现实世界中政治、社会和经济现状的不满情绪，如相互推波助澜，就可能对民主和人权产生严重影响，最近的选举、对 2019 冠状病毒病(COVID-19)大流行的应对以及对少数群体的攻击已体现出这一点。网上虚假信息在政治上造成两极化，阻碍人们真正行使人权，破坏他们对政府和体制的信任。
3. 找到应对虚假信息的适当方法并非易事，尤其是因为这一概念没有定义，容易被滥用，并且在缺少数据和研究的情况下，这一问题的规模和性质尚有争议。国家的应对往往存在问题并且措施强硬，有损人权。公司在传播虚假信息方面具有重要影响，但它们解决这一问题的努力严重不足。
4. 问题的核心是人权挑战，而信息失序加剧了这一挑战。越来越多的证据表明，如果人权受到限制，公共信息制度不健全，媒体质量不高、多样性和独立性不足，虚假信息就容易盛行。反之，意见和表达自由受到保护时，民间社会、记者和其他各方则能够质疑谎言并提出不同观点。这就使国际人权成为适合应对虚假信息的有力框架。
5. 本报告中，促进和保护意见和表达自由权特别报告员肯定了虚假信息的复杂性，并从概念和背景方面概述了这一问题所具有的困难之处。她阐述了相关国际法律标准，分析了国家和公司应对这一现象的措施，然后提出了多元素应对措施的建议，这些措施的基础是国际人权法、透明度和问责以及国家、公司和民间社会的多方利益攸关方参与。前任特别报告员在数字平台的人权责任方面做了有开创性的工作，在此基础上，现任特别报告员呼吁，应当审视数字平台的商业模式并重新调整国家应对虚假信息的措施。
6. 撰写报告时，特别报告员参考了 119 个民间社会组织和学术实体、三个国际组织、三个成员国和三家公司提交的材料，并参考了与民间社会组织的多次线上磋商以及与成员国、社交媒体公司和专家的会议。
7. 本报告无意于其内容或建议的完整无缺。例如，报告没有涵盖国家或国家赞助的行为体向他国人民主导虚假信息运动的问题，因为这一复杂问题所需磋商和思考超出了本报告时限内可能达到的水平。
8. 本报告意在开启与有关利益攸关方，包括与成员国、公司和民间社会的对话，并为正在各种论坛进行的讨论作出贡献，以期进一步完善并落实结论和建议。

二. 概念与背景方面的困难

A. 虚假信息的概念

9. 虚假信息没有普遍接受的定义。无法达成共识导致难以采取全球应对措施，同时缺乏协商一致也突显了这一概念的复杂性、内在政治属性和争议性。

10. 部分问题在于，无法明确区分事实与虚假，或者伤害意图的有无。利用虚假信息的各行为体可能持有截然相反的目的。真实信息可能被标签为“假新闻”从而变为非法。对于意见、信仰、不确定的知识和其他形式的表达，例如戏仿和讽刺而言，不容易进行真假的二元分析。此外，线上传播的意图造成伤害的内容(虚假信息)可能被无意造成伤害的不知情的第三方看到并分享(错误信息)，不知情的媒介促进了传播，同时增加了恶意宣传方的可信度。不论有意无意，伤害都会产生。某些形式的虚假信息可能构成煽动仇恨、歧视和暴力，这是国际法所禁止的。

11. 按照欧洲联盟委员会的描述，虚假信息是经核实属于不实或误导的信息，这些信息的发布和传播是为了经济利益或故意欺骗公众，并且可能造成公众伤害。¹ 另一方面，按照宽带促进可持续发展委员会的描述，虚假信息是有潜在影响的不实或误导的内容，不论信息的发布和传播背后是何种意图或行为。² 处理虚假信息的国家法律法规涵盖了多种组合，包括不实或误导的信息、有无伤害意图以及所致或意图造成之伤害的性质。用于描述虚假信息的术语通常宽泛、定义模糊且不合国际法律标准。

12. 学界对信息失序进行了分类，根据其中的描述，“虚假信息”是出于伤害之意图有意发表的不实信息，“错误信息”是无意传播的不实信息，“恶意信息”则是出于伤害之意图发表的真实信息。³ 信息失序框架展现了一个相互关联的全景图，从而支持以多维度、多样化和情境化的方法看待虚假信息问题。

13. 有学者将虚假信息现象界定为“病毒式欺骗”，它有三个媒介：操纵行为、欺骗行为与有害内容。⁴ 重点在于线上行为，而非内容真实与否。Facebook 等一些大型社交媒体平台以这些媒介作为参考，制定各自应对协作虚假行为的政策。

14. 归根结底，对于何为虚假信息的认识并不清晰，也不一致，例如错误信息一词通常被互换使用，这降低了回应措施的效力，⁵ 还导致采取危及意见和表达自由权的方法。务必在国际人权法框架内澄清虚假信息和错误信息的概念。

15. 本报告中，虚假信息意为有意传播的造成严重社会危害的不实信息；错误信息意为在不知情的情况下传播的不实信息。两个词语不可互换使用。

¹ 欧盟委员会，《虚假信息行为守则》(2018年)。

² 宽带促进可持续发展委员会，《平衡行为：在尊重表达自由的同时打击数字虚假信息》(国际电信联盟和联合国教育、科学及文化组织，2020年9月)，第8、第18页和第25页ff。

³ Claire Wardle and Hossein Derakhshan, *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking* (Council of Europe, 2017), p. 5.

⁴ Camille François, “Actors, behaviors, content: a disinformation ABC” (Transatlantic Working Group, September 2019)。

⁵ 联合国教育、科学及文化组织提交的材料。

B. 参与方和媒介

16. 虚假信息通过社交媒体和消息平台迅速广泛传播，甚至在世界上一些最偏远和最脆弱的地区也是如此，⁶ 尽管总体而言它在流通的所有信息中仍然只是一小部分。⁷ 数字技术使人们有可能以新方式分享可能扭曲现实的文本、图像和视频，包括“深度伪造”和“浅度伪造”。⁸ 一些推广耸人听闻之内容以吸引用户的平台为此设计的算法和商业模式放大了不实信息。线上环境鼓励放大时，虚假信息容易盛行，同时多元和多样化信息源的获得也有所减损。

17. 全球虚假信息系统是一项利润丰厚的业务，受到商业动机推动，并且日益专业化。⁹ 科技公司据称还准许错误信息的传播者将内容货币化，例如准许传播与 COVID-19 相关的阴谋的垃圾新闻网站在其平台上发布广告。¹⁰ 本质上，虚假信息是通过故意散布谎言获利，是数字时代的一种现代牟利方式。¹¹

18. 利用新技术生产虚假信息或引起争议之内容的行为体多种多样，包括国家、政党、政客、其他有权势的个人或企业，他们有网络水军或公关公司的支持，其动机也多种多样，包括政治、意识形态或商业动机。¹² 这些煽动者生产的虚假信息通常在有意或无意的情况下通过传统媒体、¹³ 名人或普通用户及他们的端到端和好友到好友网络传播，是线上和线下世界交流的复杂组合。

19. 意识形态驱动的非国家行为体，包括极端主义或恐怖主义分子，在招募成员和使之激进化的宣传中也经常散布虚假新闻和故事。¹⁴ 安全维度和国家对此的过度反应加重了人权关切。¹⁵

20. 尽管如此，近年来虚假信息的增长不能完全归因于技术或恶意行为体。理解这一问题还应结合其他背景因素，包括：传统媒体部门因面临数字转型和来自线上平台的竞争而处于困境，在世界某些地区还受到国家压力的威胁；有力的新闻制度缺失；大众的数字和媒体素养不高；感到沮丧不满的人越来越多，数十年的

⁶ 燕子基金会提交的材料。

⁷ 路透新闻学机构(牛津大学)和社交媒体与政治中心(纽约大学)提交的材料。

⁸ 对传播虚假信息的各种方法的说明，见，例如，Kate Jones, “[Online disinformation and political discourse: applying a human rights framework](#)”, Chatham House Research Paper (November 2019), pp. 11–12。

⁹ Samantha Bradshaw, Hannah Bailey and Philip N. Howard, “[Industrialized disinformation: 2020 global inventory of organized social media manipulation](#)”. (Computational Propaganda Project, University of Oxford, 2021), p. 21。

¹⁰ 例如，见 United Kingdom of Great Britain and Northern Ireland, Parliament, “[Misinformation in the COVID-19 Infodemic](#)” (20 July 2020)。

¹¹ 埃迪斯科文大学提交的材料。

¹² 例如，见 Dhanaraj Thakur 和 Devan L. Hankerson, “[Facts and their discontents: a research agenda for online disinformation, race and gender](#)” (Center for Democracy and Technology, February 2021)。另见美国笔会提交的资料。

¹³ Media Matters for Democracy, “[Disorder in the newsroom: the media’s perceptions and response to the infodemic](#)” (December 2020)。

¹⁴ 例如，见联合国区域间犯罪和司法研究所, “[消除虚假信息的病毒: COVID-19 期间恶意使用社交媒体的风险以及应对的技术选择](#)” (2020 年 11 月), 第 15-17 页。

¹⁵ 斯坦福大学提交的材料。

经济匮乏、市场失灵、政治权利被剥夺和社会不平等又令沮丧和不满加剧，从而使一些人更容易受到操纵。¹⁶

21. 虚假信息并非社会危机和公众对机构信任崩溃的原因，而是其后果。如果不对这些潜在因素给予更多关注，应对虚假信息的战略不太可能成功。

C. 目标与受害者

22. 虽然实证研究显示，接触虚假信息只是一小部分人，¹⁷ 但机构、社区和个人受到的影响是真实、广泛的并且在情理之中。特别报告员编写本报告时收到的 100 多份材料中有许多实例。¹⁸ 这些实例表明，目标所指有许多是出于政治动机，针对的是处境脆弱的机构和个人，并影响到广泛的人权，包括经济、社会、文化、公民和政治权利。

23. 有明确证据表明，有力的新闻制度和独立的新闻报道是对治虚假信息的良药。因此，社交媒体网络上针对记者的诽谤运动的加剧倍加令人不安。一些政治领导人给媒体贴上“人民公敌”的标签，¹⁹ 或者支持整个“假新闻产业”淹没媒体的报道。²⁰ 这种攻击既侵蚀了公众对新闻业的信任，也让记者对自己的工作更加担忧。²¹ 一份报告显示，2020 年至少 34 名记者因“假新闻”的罪名而入狱，而 2012 年只有 1 名。²² 虚假信息不仅对记者的安全构成威胁，也对记者所在的媒体生态系统构成威胁，²³ 迫使传统媒体将宝贵的资源从报道转用于驱散和揭穿谎言。

24. 多个国家曾极为明显地使用虚假信息损害自由和公正选举的权利。²⁴ 例如美利坚合众国最近的三次重要选举中，曾使用有种族针对性的虚假信息运动压制有色人种社区的选票。²⁵ 2020 年总统大选期间，时任总统唐纳德·特朗普及其代理人多次试图削弱人们对邮政投票系统的信心，并在社交媒体上发表关于选举舞弊的毫无根据的说法。²⁶ 出于政治动机之虚假信息的有害影响波及了许多其他国家的民主机构，压制了自由言论，降低了公众对公共领域这一民主议事空间的信任程度，放大了反民主的论述，助长了两极分化，促进了威权和民粹主义议程。²⁷

¹⁶ 第十九条组织提交的材料：International Centre against Censorship and the Internet Governance Project。

¹⁷ 社交媒体与政治中心和路透新闻学机构提交的材料。

¹⁸ 这些材料将公布在 www.ohchr.org/EN/Issues/FreedomOpinion/Pages/Report-on-disinformation.aspx。

¹⁹ 保护记者委员会，“The Trump Administration and the media: attacks on press credibility endanger US democracy and global press freedom” (April 2020)。

²⁰ 欧洲议会，“Disinformation and propaganda: impact on the functioning of the rule of law in the EU and its Member States” (2019)。

²¹ Media Matters for Democracy, “Disorder in the newsroom”。

²² 保护记者委员会提交的材料。

²³ 同上。

²⁴ 民主和技术中心与美国笔会提交的材料。

²⁵ Young Mie Kim, “Voter suppression has gone digital”, Brennan Center for Justice, 20 November 2018。

²⁶ 国际非营利法中心提交的材料。

²⁷ 例如，见进步通信协会和 Global Partners Digital 提交的材料。

25. 过去一年中，一些政府试图隐瞒或伪造信息，这种背景下，非国家行为体传播的虚假信息和错误信息给健康权和应对 COVID-19 大流行带来了巨大挑战。²⁸ 如同全球卫生，在气候变化方面，有组织的线上虚假信息运动导致科学信息的可信度减弱，也令环保活动人士受到攻击。

26. 意识形态和基于认同的虚假信息煽动了对少数群体、移民和其他边缘化社区的歧视和仇恨，²⁹ 造成了族裔或宗教冲突，³⁰ 有时积累成为线下暴力，例如埃塞俄比亚³¹ 和缅甸³² 的情况。民间社会组织呼吁开展更多研究，以全面了解关于弱势和少数群体的虚假信息的严重程度。³³

27. 涉及性别问题的线上虚假信息运动越来越多地被用于威胁阻止女性介入公共领域，并将“根深蒂固的陈旧性别偏见与社交媒体的匿名性和影响力结合在一起，试图破坏女性的声誉，将她们赶出公共生活”。³⁴ 就女权主义问题发声的女性记者、政治家和性别平等倡导者尤其成为针对目标。³⁵ 围绕性健康和生殖健康问题也存在严重的虚假信息。³⁶

28. 人权维护者和民间社会组织，特别是代表边缘化和受歧视群体的组织，受到线上虚假信息运动的猛烈攻击以及辱骂和贬损。³⁷

29. 虚假信息的负面影响不可否认，必须解决。国际人权法提供了一种有效的应对办法和制定对策的框架。

三. 适用的国际法律框架

30. 人权理事会确认，应对虚假信息和错误信息传播的措施必须以国际人权法为基础，包括合法性、正当性、必要性和相称性原则。³⁸ 虚假信息对广泛的人权皆有影响，但本报告侧重于意见和表达自由，因为这项权利对于打击虚假信息的努力而言格外重要。

31. 《世界人权宣言》第十九条和《公民及政治权利国际公约》第十九条保障保持意见不受干预之权利以及通过任何媒介，不分国界，寻求、接受及传播各种消息及思想之权利。意见自由是绝对的，表达自由在特定情况下则可能受到限制。国家有义务不干涉这一权利，也有义务确保包括企业在内的其他各方不干涉这一权利。

²⁸ A/HRC/44/49, 第 45-47 段。

²⁹ A/HRC/46/57。

³⁰ Ahmadiyya Muslim Lawyers Association 提交的材料。

³¹ 见 www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=26483&LangID=E。

³² 见 www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=26808&LangID=E。

³³ 见立即普及组织与民主和技术中心提交的材料。

³⁴ Nina Jankowicz, “How disinformation became a new threat to women”, *Coda Story*, 11 December 2017。

³⁵ Media Matters for Democracy 提交的材料。另见 Maria Ressa 案，AL [PHL 12/2018](#) 中有所提及。

³⁶ MSI Reproductive Choices 提交的材料。

³⁷ 数字权利基金会提交的材料。

³⁸ 第 44/12 号决议。

32. 上一任任务负责人曾敦促数字技术公司在其商业惯例中适用国际人权标准。³⁹ 在虚假信息和错误信息问题的背景下，这些敦促变得更加紧迫。

A. 意见自由

33. 意见自由权有两个维度：向内的维度与隐私权和思想自由权密切相关，向外的维度则与表达自由相关。后者经常被讨论，前者则最近才开始受到关注，因为人们越来越了解并认识到，社交平台、国家行为体和其他各方在线上使用操控技术影响他们的方式可能侵犯他们的意见自由。⁴⁰

34. 《世界人权宣言》第十九条和《公民及政治权利国际公约》第十九条保护持有主张而不受干涉的权利。这项权利是绝对的，没有例外，也不得加以限制。尽管这项权利具有绝对性和广泛性，但现实中，人类的思想 and 意见不断受到他人影响，“受各种影响的自由本身就是我们的自主性的一个维度”。⁴¹ 因此，判定网上虚假信息是否以及如何侵犯意见自由时，关键在于权利持有人的知情及同意。

35. 形成并通过推理发展自己的意见的权利是意见自由的基本要素。既定原则是，意见自由包括不发表意见的权利，也包括个人自由选择随时或出于任何理由改变见解的权利。⁴² 也就是说，禁止非自愿披露，同时肯定精神自主。禁止任何胁迫他人持有或者不持有任何意见的行为。

36. 因他人持有某种意见而进行惩处、骚扰、恐吓和污名化，包括以强迫、非自愿或未经同意的方式操控形成意见的思考过程，都是对意见权利的侵犯。⁴³ 当前的认识是，胁迫或操纵行为包括灌输、“洗脑”、“以精神活性药物或其他操控手段”影响“意识或潜意识”。⁴⁴ 体现在数字领域就是国家和非国家行为体使用各种技术在人们不知情或不同意的情况下掌握和影响他们的想法和意见，例如借助强大的平台建议或精准投放进行内容集展。此类技术在传播虚假信息方面具有重要作用，在非自愿或未经同意的情况下操纵思考过程，有悖言论自由权。⁴⁵

³⁹ A/HRC/38/35, 第 9-12 段; A/74/486, 第 40-55 段。

⁴⁰ Evelyn Aswad, “Losing the freedom to be human”, *Columbia Human Rights Law Review*, vol. 52 (2020); Susie Alegre, “Rethinking freedom of thought for the 21st Century”, *European Human Rights Law Review* (2017); and Kate Jones, “Protecting political discourse from online manipulation: the international human rights law framework”, *European Human Rights Law Review* (2021)。

⁴¹ Kate Jones, “[Online disinformation and political discourse: applying a human rights framework](#)”, p. 33。

⁴² 人权事务委员会第 34 号一般性意见(2011 年), 第 9 段。

⁴³ “以未经同意的手段故意施加影响的行为，如果达到使他人丧失精神自主或操控他人思维的程度，则侵犯了这项权利。”见 Evelyn Aswad, “Losing the freedom to be human”, p. 329。

⁴⁴ Manfred Novak, *UN Covenant on Civil and Political Rights: CCPR Commentary*, 2nd ed. (Kehl am Rhein, N.P. Engel, 2005)。

⁴⁵ Evelyn Aswad, “Losing the freedom to be human”。另见 Susie Alegre 对《保护人权与基本自由公约》(《欧洲人权公约》)第 9 条的类似评论, “Rethinking freedom of thought for the 21st Century”。

B. 表达自由

37. 表达自由权广泛而具有包容性，它涵盖了通过任何媒体，不论线下或线上，不分国界地寻求、接受及传播各种消息及思想之自由。⁴⁶

38. 在虚假信息方面应当注意两点。首先，表达自由权适用于各种信息和思想，包括可能造成震撼、冒犯或纷扰的信息和思想，⁴⁷ 无论内容真假。⁴⁸ 根据国际人权法，人们有权按自己所愿发表根据不足的意见和言语或者尽情戏仿或讽刺。⁴⁹ 其次，信息自由流动是表达自由的一个要素，国家因此负有积极义务，主动将关乎公共利益的信息公之于众并促进信息来源多元化，包括媒体自由。信息自由流动可以作为打击虚假信息的有力工具。

39. 对表达自由的限制必须遵守《公民及政治权利国际公约》第十九条第三款，该条规定，所有限制均须经法律规定，并且为尊重他人权利和名誉以及保障国家安全、公共秩序、公共卫生或风化所必要。这项权利对享有所有其他人权而言具有根本的重要性，因此对它的限制必须是例外并且须加以严格解读。

40. 罪刑法定原则要求法律的范围、意义和效力足够明确、准确和公开。赋予过度酌处权的模糊的法律可能导致任意决策，从而有悖《公约》第十九条第三款。凡是对虚假信息的限制，必须与保护第十九条第三款所述合法目标具有紧密的具体关联。根据国际人权法，禁止不实信息本身并非合理目标。

41. 评估限制是否必要的关键考量，是言论和伤害之间因果关系的直接程度以及伤害的严重性和紧迫性。根据必要性原则的要求，限制应当对于达到合法目的而言适当且相称，同时使用限制最少的保护手段。刑事制裁构成对表达自由的严重干预，并且除了对于最恶劣的案件，都是不相称的应对。

42. 表达自由对于民主和享有所有其他人权与自由具有根本的重要性，因此国际人权法对关乎公共利益的事务的表达，包括批评政府和政治领导人以及政客和其他公众人物的言论，以及媒体自由，给予了格外有力的保护。⁵⁰ 这并不意味着政治言论中的虚假信息从不会受到限制，而是意味着，任何此类限制都须符合高标准的合法性、正当性、必要性和相称性。例如，选举法可以合理地禁止传播与选举诚信有关的虚假信息，但这种限制必须加以严格解读、有时间限制和针对性，以避免局限政治辩论。

43. 虚假信息经常被用来煽动仇恨和暴力。《公民及政治权利国际公约》第二十条第二款规定，任何鼓吹民族、种族或宗教仇恨之主张，构成煽动歧视、敌视或强暴者，应以法律禁止之。该条款既未要求刑罪化，也丝毫未提及不实信息。

⁴⁶ 人权事务委员会，第 34 号一般性意见(2011 年)第 12 段，以及人权理事会第 20/8 号决议。

⁴⁷ 人权事务委员会，第 34 号一般性意见(2011 年)，第 11 段。另见欧洲人权法院，*Handyside* 诉联合王国，第 5493/72 号申诉，1976 年 12 月 7 日的判决，第 49 段。

⁴⁸ 人权事务委员会，第 34 号一般性意见(2011 年)，第 47 和第 49 段。另见欧洲人权法院，*Salov* 诉乌克兰，第 65518/01 号申诉，2005 年 9 月 6 日的判决，第 113 段：“《[欧洲人权]公约》第 10 条(关于表达自由)不禁止讨论或传播收到的信息，即使这些信息极有可能并不真实。”

⁴⁹ 只有符合《公民及政治权利国际公约》第十九条第三款所列限制和标准的虚假陈述才能予以限制。

⁵⁰ 人权事务委员会，第 34 号一般性意见(2011 年)，第 38 段。

《关于禁止构成煽动歧视、敌意或暴力的鼓吹民族、种族或宗教仇恨言论的拉巴特行动计划》是解读第二十条第二款的权威路线图，它提出了六要素，用以判定何种严重程度的煽动有必要刑罪化：背景；发言者的身份；意图；言论的内容和形式；言论的影响范围；风险的可能性。特定情况下，它可以作为处理虚假信息的有用框架。⁵¹

44. 《消除一切形式种族歧视国际公约》第四条禁止与种族或族裔血统相关的仇恨言论。人权事务委员会和消除种族歧视委员会都曾澄清道，必须严格遵循《公民及政治权利国际公约》第十九条说明禁止是合理的，并且刑罪化只可用于最严重的情况。

45. 《关于表达自由与“假消息”、虚假信息和宣传问题的联合声明》借鉴国际人权法，列出了关键原则，用以指导国家、公司和其他各方。⁵²

四. 国家应对虚假信息的措施：主要关切

46. 国家应对虚假信息的措施可以有各种形式，从采取干扰互联网到审查，从惩处或限制传播到监管社交媒体平台。这一系列措施的一端，是国家表面上试图制止虚假信息，实则予以赞助；另一端(积极的)是，国家推行各种措施，鼓励信息自由流动，加强媒体多样性，支持媒体信息和数字素养，以此为手段打击虚假信息。

A. 国家赞助的虚假信息

47. 国家赞助的虚假信息可能直接来自国家机构或来自代理人，它们具有政治和战略目的，指向本国境内或境外的受众。⁵³ 在数字时代，新技术显著扩大了此类操作的规模、速度和传播范围。它们一旦与一国的权力、手段和影响力结合，可能对人权产生破坏的影响。如果国家一贯在宣传自己的虚假叙述的同时压制其他来源，就剥夺了个人根据《公民及政治权利国际公约》第十九条第二款享有的寻求和接受信息的权利。⁵⁴

48. 国家主导虚假信息的一个著名实例是缅甸的“政府军真实新闻信息团队”，该团队在线上发布了与罗兴亚危机有关的修改过并且具有误导性的照片。2018年8月，Facebook以散布仇恨言论为由封禁了该“团队”的账号。⁵⁵ 另一实例是菲律宾国家特工在线上使用“红色标签”，将活动家、记者和政治对手错误地标为左翼、共产主义、恐怖主义分子或颠覆分子，从而增加了这些人遭受逮捕、袭击或杀害的风险。⁵⁶

⁵¹ A/HRC/22/17/Add.4, 附件, 附录。

⁵² 见 www.osce.org/files/f/documents/6/8/302796.pdf。促进和保护意见和表达自由权特别报告员、欧洲安全与合作组织媒体自由代表、美洲国家组织表达自由特别报告员以及非洲人权和民族权委员会表达自由和信息获取问题特别报告员于2017年3月3日通过了这项联合声明。

⁵³ 本报告不探讨国家试图在境外散布虚假信息的举动。

⁵⁴ Mark Milanovic and Michael N. Schmitt, “Cyber attacks and cyber (mis)information operations during a pandemic”, *Journal of National Security Law and Policy*, vol. 11 (2020)。

⁵⁵ 见 <https://about.fb.com/news/2018/08/removing-myanmar-officials/>。

⁵⁶ AL PHL 1/2021。

49. 近年来，在一些国家，由国家主导的虚假信息运动试图影响选举和其他政治进程，控制公开辩论的叙述，或遏制对政府的抗议和批评。在 COVID-19 大流行的背景下发生了多种事件，事关国家行为体散布未经核实的关于造成 2019 冠状病毒病的病毒的说法，否认疾病传播或提供关于感染率、死亡人数和卫生保健建议的虚假信息。这种虚假信息有损控制大流行的努力，危及健康权和生命权，也危机人们对公共信息和国家机构的信任。⁵⁷

B. 断网

50. 过去两年中，白俄罗斯、⁵⁸ 刚果民主共和国、⁵⁹ 埃塞俄比亚⁶⁰ 和缅甸⁶¹ 等多个国家在选举前夕或选举期间都曾断网，表面上是为了防止可能引发暴力的虚假信息在线上传播。一些政府还曾在示威期间或是为了压制异见而下令断网，例如在巴林和委内瑞拉玻利瓦尔共和国。⁶² 塔吉克斯坦也曾有模糊的法律赋予政府广泛的酌处权，使之得以关闭或以其他方式中断互联网连接和电信接入。⁶³ 与之形成对比的是，印度尼西亚一家法院判定，不得在巴布亚和西巴布亚政治动荡期间断网以防“假新闻”传播，理由是该决定过度且非必要。⁶⁴

51. 人权理事会曾强烈谴责通过断网有意和任意地阻止或干扰线上获取信息。⁶⁵ 断网一并封禁了互联网的其他多重用途，所以这种行为本身就是不相称的应对。这样做违反了国际人权法规定的必要性和相称性要求，剥夺了人们可在线上获得的全部信息和服务，阻碍选民获取选举信息，阻碍人权维护者记录和分享人权关切，阻碍记者和媒体报道关乎公共利益的问题。断网剥夺了人们的信息来源，这样做非但不能遏制虚假信息，反而会阻碍实况调查，并且可能助长谣言。许多情况下，关闭的目的似乎是压制少数人的声音，剥夺他们获取重要信息的机会。⁶⁶

C. 刑法

52. 各国早已谨慎立法，以解决不实信息所致危害，例如关于诽谤、消费者保护和金融欺诈的立法。以刑法惩处散布定义宽泛的关乎公共利益问题的虚假信息则更成问题。其中一些法律可追溯至殖民时代，乌干达、⁶⁷ 赞比亚⁶⁸ 和津巴布韦⁶⁹

⁵⁷ A/HRC/44/49, 第 45 段。

⁵⁸ 见 www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=26164&LangID=E。

⁵⁹ 见 www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=24057&LangID=E。

⁶⁰ 见 www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=26483&LangID=E。

⁶¹ 见 www.ohchr.org/en/NewsEvents/Pages/DisplayNews.aspx?NewsID=26431&LangID=E。

⁶² A/HRC/35/22, 第 11 段。

⁶³ A/HRC/35/22/Add.2, 第 29 段。

⁶⁴ Moch. Fiqih Prawira Adjie, “Jokowi ‘violates the law’ for banning internet in Papua, court declares”, *Jakarta Post*, 3 June 2020.

⁶⁵ 第 44/12 号决议。

⁶⁶ 见立即普及组织与法律和民主中心提交的材料。

⁶⁷ 乌干达，最高法院，Charles Onyango Obbo 和另一人诉总检察长，2002 年第 2 号宪法上诉，2004 年 2 月 10 日的判决。

⁶⁸ 赞比亚，高等法院，Chipenzi 等诉 the People, HPR/03/2014, 2014 年 12 月 4 日。

⁶⁹ 津巴布韦，最高法院，Chavunduka 诉内政部长，案件编号 2000 JOL 6540 (ZS), 2000 年 5 月 22 日。

的国内法院曾认定，在现代民主社会，这些法律违宪并且不合理。在冈比亚，煽动、假新闻和刑事诽谤是刑事犯罪，西非地区法院于 2016 年认定，这些罪名违反了表达方面的国际法，并下令将之撤销。⁷⁰

53. 过去十年中忽然出现大量法律，禁止互联网和社交媒体平台上各种形式的“假新闻”，仅过去一年，至少 17 个国家通过了立法，以处理与大流行病有关的不实信息。⁷¹

54. 很多关于“假新闻”的法律并不符合《公民及政治权利国际公约》第十九条第三款关于合法性、必要性及合理性的三重检验标准。这些法律通常并未足够准确地界定何为虚假信息或法律试图防范的损害，也未要求证明施行的行为和所致损害之间存在较强的具体关联。使用诸如“错误”、“虚假”或“有偏见”等词而不详加说明，并基于循环逻辑作出断言(例如，“一项陈述是错的或误导性的，因为它是错的或误导性的，无论是全部还是部分，也无论它本身是错的还是在其所在上下文中是错的”)。⁷² 某些情况下，危害的定义过于宽泛。⁷³ 在一些国家的立法，特别是马来西亚⁷⁴ 和新加坡⁷⁵ 的立法，在无司法监督的情况下赋予行政机关不受约束的酌处权，从而带来了滥用权力和任意决策的可能。规定的处罚通常过于严厉且不相称，可能造成寒蝉效应，影响表达自由。⁷⁶

55. 这类法律模糊而宽泛，让各国政府得以将之用于惩处记者、政治对手和人权维护者。例如，土耳其《刑法》和反恐立法中模糊而宽泛的条款将广大类别的言论定为犯罪，包括“诋毁土耳其民族”或“侮辱总统”的表达，这些条款曾被用于惩处土耳其的许多政治活动家和记者。⁷⁷ 在埃及，人权维护者和记者曾在发表有关该国人权状况的报道后因散布“假消息”而受到起诉。⁷⁸ 孟加拉国曾依《数字安全法案》拘留漫画作者、博主和记者，引起了关于拘留期间遭受酷刑和死亡的指控。⁷⁹ COVID-19 大流行初期，亚洲许多国家利用“假新闻”法律压制对政府的批评的做法显著增加，联合国人权事务高级专员曾对此表示担忧。⁸⁰

⁷⁰ 西非国家经济共同体社区法院、非洲记者联合会和其他四人诉冈比亚，第 ECW/CCJ/JUD/04/18 号判决。

⁷¹ 见 <https://ipi.media/rush-to-pass-fake-news-laws-during-covid-19-intensifying-global-media-freedom-challenges/>。见法律和民主中心提交的材料。

⁷² 见新加坡《防止在线虚假和操纵法》，第 2(2)(b)条。

⁷³ 例如，见卡塔尔《刑法》第 136 条和 [QAT 1/2020](#)。

⁷⁴ [OL MYS 5/2021](#)。

⁷⁵ [OL SGP 3/2019](#)。

⁷⁶ [OL BFA 2/2020](#)。

⁷⁷ 例如，见 [OL TUR 13/2020](#)、[AL TUR 18/2020](#) 和 [AL TUR 20/2020](#)。

⁷⁸ 例如，见 [AL EGY 19/2020](#)、[AL EGY 15/2020](#)、[AL EGY 10/2020](#)、[UA EGY 6/2020](#) 和 [UA EGY 1/2020](#)。

⁷⁹ [OL BGD 4/2018](#) 和 [AL BGD 7/2020](#)。

⁸⁰ 见 www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=25920&LangID=E。

D. 社交媒体监管

56. 线上虚假信息或错误信息关切的核心问题是，此类信息可能会以极快的速度病毒式传播。国家应对这一挑战的方式各不相同，一些国家出台了法规，责成主管机构直接管控线上内容，另一些国家关注的并非内容，而是公司处理虚假信息或错误信息的方法。一些法律或提案引起了隐私关切，因为即时通讯服务可能属于其管辖范围，并且它们要求用户提供身份信息，追踪数据或使用过滤器。⁸¹

57. 近年来，多个国家出台的法律赋予了主管机构过度酌处权，主管机构可以要求社交媒体平台删除它们认为非法的内容，包括它们认为是虚假信息或“假新闻”的内容。如不删除则将被处以高额罚款和/或被封禁。例如，肯尼亚、⁸² 巴基斯坦⁸³ 和俄罗斯联邦⁸⁴ 就发生了这种情况。实际上，这些法律令合法线上表达受到压制，同时正当程序有限或缺失或者无法院事先命令，有悖《公民及政治权利国际公约》第十九条第三款的要求。⁸⁵ 在拉丁美洲，关于虚假信息的法律在没有司法命令的情况下强迫平台决定是否删除内容，这有悖《美洲人权公约》第 13 条。⁸⁶

58. 各国将传统上属于法院的“言论警察”职能委托给线上平台的趋势仍在继续。这类法律的风险在于，媒介可能因担心受制裁，为保险起见“过度删除”内容。德国《网络执法法》于 2018 年通过并于最近修订，该法允许用户标记他们认为根据《刑法》某些条款属于非法的内容，并要求平台在短时间内删除“违规内容”，否则将面临高额罚款。⁸⁷ 尽管德国的法律没有禁止或惩处虚假信息，但其他国家曾援引德国的法规，希望出台限制过度的媒介法律或社交媒体法规，准许在没有司法甚至准司法命令的情况下删除“假新闻”。

E. 新出现的趋势

59. 没有明文禁止虚假信息或错误信息的国家一般依靠社交媒体平台的服务条款处理虚假信息。然而，欧洲联盟似乎正在发展的新趋势是，数字服务法草案一旦通过并执行，将要求平台和其他媒介采取透明度和正当程序措施，除其他外，这些措施可有助于解决虚假信息问题。此类监管建议侧重于透明度和正当程序义务，而非基于观点或内容的监管，可以为保护人权和加强平台的公共问责做出积极贡献。然而，为使监管措施正常运作，必须保证并严格尊重监督机构或监管者的独立性。⁸⁸

⁸¹ 例如，见 [BRA 6/2020](#)；另见数字权利组织提交的资料。

⁸² [OL KEN 10/2017](#)。

⁸³ [OL PAK 3/2020](#)。

⁸⁴ [OL RUS 4/2019](#)。

⁸⁵ 见 <https://transparency.facebook.com/content-restrictions>。

⁸⁶ [OL BRA 6/2020](#)。

⁸⁷ [OL DEU 1/2017](#)。

⁸⁸ 例如，见 Eleonora Maria Mazzoli and Damian Tambini, “[Prioritisation uncovered: the discoverability of public interest content online](#)”, Council of Europe study DGI(2020)19, pp. 40–43。

60. 数字服务法草案还将要求大型在线平台对“平台服务的运作和使用产生的重大系统性风险”进行年度审查，包括“故意操纵平台服务”，以免对保护公共健康、未成年人、公民言论、选举进程和公共安全造成或可能造成负面影响。⁸⁹ 随后须采取相应的适当缓解措施，并进行独立审计。超大型在线平台也须按要求遵守行为准则，包括关于虚假信息的准则，并接受独立的监管监督。这些尽职措施在欧洲联盟成员国能否保护人权并解决虚假信息的问题，最终将取决于措施起草成法律后的清晰和精确程度以及监管机构的效力和独立性。

61. 另一新兴趋势是，一些法律草案或提案寻求限制非法或其他有害内容放大，例如巴西⁹⁰、法国⁹¹ 和美国⁹² 的趋势。为使此类措施有助于减少虚假信息的影响，法律应当以清楚准确的语言拟就，从而不侵犯言论自由，还应当只寻求制止不当影响。

62. 2021 年 1 月 6 日美国国会大厦事件后，美国前总统特朗普在 Twitter、Facebook、YouTube、Reddit 和其他平台上的账号被永久停用，由此催生了一些关于禁止公司删除合法内容的监管提议。波兰一项新的拟议法律可能要求社交媒体公司恢复经大体由政府控制的机构认定为合法的内容。⁹³ 在巴西，允许基于观点删除内容的服务条款可能遭致暂停服务或其他制裁。这种“强制载播”义务或相关义务的特征是“公正义务”，如果不符合国际人权法，则可能带来损害，国际人权法在坚定维护表达自由的同时禁止仇恨言论。⁹⁴

五. 公司的应对：主要关切

63. 公司与国家的人权义务有所不同。然而，依照《工商企业与人权指导原则》，公司在其活动和业务中须尊重人权。公司至少应定期对其产品、业务和政策进行人权影响评估，并实施尽职调查程序，以便发现、预防或缓解任何实际或潜在的对人权的不良影响。他们还应该为用户设置补救程序。《关于表达自由与“假消息”、虚假信息和宣传问题的联合声明》以及上一任任务负责人的报告也提供了重要指导，帮助公司了解适用于其政策和内容审查做法的人权标准。

64. 为应对虚假信息和错误信息带来的挑战，美国最大的几家社交媒体平台⁹⁵ 采用了一系列政策和工具。这些平台一般禁止它们视为“假新闻”的内容和有损其平台真实性和诚信的各种欺骗做法。一些平台针对与 COVID-19 相关的错误

⁸⁹ 见 <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en>。

⁹⁰ OL BRA 6/2020。

⁹¹ OL FRA 5/2018。

⁹² 立即普及组织提交的材料。

⁹³ 例如，见 Richard Wingfield, “Poland: draft law on the protection of freedom of speech on online social networking sites”, 8 February 2021。

⁹⁴ 例如，见罗马一家法院在 Facebook 诉 CasaPound 案中的裁决，案件编号 80961/19，2020 年 4 月 29 日。

⁹⁵ 由于缺少易于获取的信息，本章节主要指 Facebook、YouTube 和 Twitter，一定程度上也指 TikTok。

信息⁹⁶ 以及公民诚信⁹⁷ 出台了具体政策。政策的执行有多种形式，包括标记、警告、删除内容和关闭账号。这样做使得存在问题的信息可见度降低或者影响范围缩小。⁹⁸ 各家公司还提及推广权威内容的努力。⁹⁹ Facebook 已出台第三方事实核查计划。¹⁰⁰ 最近，Twitter 宣布采用新的以社区为基础的办法打击虚假信息。¹⁰¹ 社交媒体平台以外，WhatsApp 等即时通讯服务限制了用户向无限数量的收件人转发消息的可能，特别是在选举等敏感时期。¹⁰²

65. 这些措施总体上是积极的，但不足以应对虚假信息带来的挑战。如果不认真审查虚假信息和错误信息的大部分驱动因素背后的商业模式，被动的内容审查工作根本不足以带来真正变化。¹⁰³ 此外，内容审查工作不断展现出长期以来的问题，包括公司服务条款的应用不一致，补救机制不完善，以及透明度和数据获得的缺失有碍于对已采取措施的效力进行客观评估。此外，这些平台虽然是全球企业，但它们在各个地理区域适用的政策似乎并不一贯，在各个辖区维护人权的程度似乎也不相同。

A. 广告驱动的商业模式

66. 在将用户导向损害持有见解权和表达自由的“极端主义”内容和阴谋论方面，最大的社交媒体公司的算法、定向广告和数据收集做法负有主要责任。¹⁰⁴ 一个真实的关切是，系统地收集用户线上活动数据和定向广告可能侵犯用户根据《公民及政治权利国际公约》第十九条第一款享有的意见自由权。公司自动策划线上内容的做法缺乏透明度，是对个人形成观点不受操纵的权利和隐私权的干涉，令人难以接受。公司设计内容高度个性化的产品，鼓励上瘾参与，从而进一步推行一种十分有损人们在信息吸收结构方面的主体能动性和选择的制度。¹⁰⁵ 最后，有证据表明，对人们通过线上搜索和其他活动表示的私人想法进行记录，商业行为体或政府可能会选择性地予以用来加害于他们。¹⁰⁶ 另一关切是，有政治暴力史的国家无限制或不加保障地收集族裔或政治派别的数据，这可能带来危险。

⁹⁶ 例如 Twitter 和 YouTube。

⁹⁷ 见 <https://help.twitter.com/en/rules-and-policies/election-integrity-policy>。

⁹⁸ 例如，见 Tessa Lyons, “Hard questions: what’s Facebook’s strategy for stopping false news?”, 23 May 2018。另见 Twitter 的执法方案，参阅 <https://help.twitter.com/en/rules-and-policies/enforcement-options>。

⁹⁹ 例如，见 www.youtube.com/howyoutubeworks/product-features/news-information/。

¹⁰⁰ 见 www.facebook.com/journalismproject/programs/third-party-fact-checking。

¹⁰¹ Keith Coleman, “Introducing Birdwatch, a community-based approach to misinformation”, 25 January 2021.

¹⁰² 例如，见 Les Femmes, la force du changement 提交的材料。

¹⁰³ 社交媒体和政治中心提交的材料。

¹⁰⁴ 例如，见沃达丰提交的材料。

¹⁰⁵ A/HRC/44/49, 第 60 段。

¹⁰⁶ Evelyn Aswad, “Losing the freedom to be human”, p. 363。

67. 尽管有这些关切，尚不清楚社交媒体平台是否作为人权尽职工作的一部分对自身的商业模式做了审查。也没有充足的公开信息，让用户、研究人员和活动人士了解算法如何推广特定类型的内容。

68. 最大的公司已采取措施的主要领域似乎与政治广告有关。与选举相关的虚假信息问题严重，社交媒体公司才总算创建了广告档案或图书馆，以便对其平台上的政治广告进行一些审查。然而，可用信息往往十分有限，并且对政治广告主的检查并非必须。¹⁰⁷ 此外，用户默认收到政治广告，而非通过选择机制。总体上，并不清楚定向广告采用何种标准或宗旨，以及这些标准或目标是否符合人权标准。

69. 这一问题需要更大的透明度，包括关于定向、实际覆盖范围和广告支出金额的信息。¹⁰⁸ 就政治广告而言，提高透明度应当与国家的选举法改革的同时进行，以确保线上政治广告和数字宣传运动无损选举和民主进程之健全。

B. 规则的适用

70. 社交媒体公司依靠一系列内容政策处理线上虚假信息或错误信息。适用的规则可能难以找到，因为它们通常散布于公司网站的各处，包括社区标准、政策、领导声明、新闻资讯、产品信息页面和商业帮助中心。¹⁰⁹ 一个普遍关切是，定义通常过于宽泛：它们并非总是清楚地阐明何种伤害以及伤害之可能性将导致内容被删除、标记或受到其他制裁。实施协作虚假行为政策也可能对表达自由，特别是对合法竞选活动产生不利影响。¹¹⁰ 公司的定义不清晰，影响了公司规则和政策执行的一贯性。

71. 过度依赖无法捕捉细微差别或理解上下文的自动过滤器加剧了这一问题。技术有局限，对何为虚假信息的判定带有政治性质，加之内容审核决策缺乏透明度，增加了受到允许的内容被删除的可能。由此凸显了人类参与内容删除决策的需要，特别是在现实世界中有伤害或暴力风险时。¹¹¹

C. 补救

72. 公司一直没有针对基于虚假信息或错误信息的过失提供足够的补救。大型社交媒体公司使用不完善的过滤器删除内容，这种做法本身有重大风险，针对错误决定的上诉机制对于抵消这种风险至关重要。然而，公司采取的标记和降级等执法行动似乎不可上诉。似乎也无法通过上诉质疑基于协作伤害或虚假行为政策作出的决定。此外，尚不明确是否有多种语言的上诉机制可用。

¹⁰⁷ 隐私国际提交的材料。

¹⁰⁸ Kate Jones, “Online disinformation and political discourse”, p. 54. 另见 Nathalie Maréchal, Rebecca MacKinnon and Jessica Dheere, “Getting to the source of infodemics: it’s the business model” (New America, May 2020), pp.55-56.

¹⁰⁹ 例如，见 Facebook 监督委员会在案件 2020-006-FB-FBR 中的决定。

¹¹⁰ 见 www.accessnow.org/rights-groups-to-facebook-on-tunisia-disappeared-accounts-were-still-waiting-for-answers/。

¹¹¹ Global Partners Digital 提交的材料。

73. 除内部投诉机制外，设立第三方监督机构也可作为加强补救措施的重要手段。Facebook 监督委员会是外部投诉机制，是一个新尝试。虽然目前评估其效力为时尚早，但应当在适当时候通过透明的、多方利益攸关方的参与式进程对其加以评估，因为它可以为该部门提供宝贵经验。还有必要考虑全行业、多方利益攸关方的措施，如建立社交媒体理事会，特别是对于规模较小的参与者而言。这样的多方利益攸关方机构可以提供政策建议，包括关于虚假信息和错误信息的建议，并审议参与公司所作决定产生的上诉。

D. 地域差异

74. 就公司的内容审核方法而言，一大问题是，公司在世界不同地区的政策实施显然有差别。¹¹² 美国设有专门的选举或投票信息中心或广告图书馆，但多数国家的投入似乎并未达到相同水平。¹¹³ 即便在美国，西班牙语资源所获资助也不及英语资源。一些公司仅在它们的业务所在地针对国别逐步制定临时的选举错误信息政策，¹¹⁴ 但这些公司似乎是少数。此外，标记内容的做法似乎只有在美国完全可用，在拉丁美洲国家则少得多。¹¹⁵

75. 事实核对服务在世界许多地区较为有限。¹¹⁶ 从吹哨人报告的信息来看，公司在较不富裕市场中的故意不作为令人关切。一经证实，这将说明内容审查的质量有很大差异，有损发展中国家的公民空间。¹¹⁷

76. 最大的几家公司总部设在美国，受美国的政治和公众舆论影响，似乎受到美国和欧洲的优先事项的推动。他们没有投入充足资源以了解世界其它地区，特别是发展中国家，催生线上虚假信息的当地因素。¹¹⁸ 在虚假信息较为盛行的国家，有必要透彻了解当地的政治、社会和经济背景，掌握当地语言，并与民间社会密切合作。

E. 政治压力

77. 对于公众人物，公司似乎没有一贯适用其服务条款或社区标准。它们似乎也没有制定明确的政策，用以防止政治内容被审查定为“假新闻”，或用以处理公众人物制作的鼓吹煽动暴力的内容。

¹¹² 例如，见 Intervozes 提交的材料。

¹¹³ 隐私国际提交的材料。

¹¹⁴ 见 Facebook 就其在缅甸的战略提交的材料。

¹¹⁵ 数字权利组织提交的资料。

¹¹⁶ 见 Mahsa Alimardani and Mona Elswah, “Trust, religion and politics: coronavirus misinformation in Iran”, in *2020 Misinfodemic Report: COVID-19 in Emerging Economies* (Meedan, 2020)。

¹¹⁷ 见 www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo。

¹¹⁸ 法律和民主中心提交的资料。

78. 2020 年美国总统选举期间，社区标准适用不一致招致了对大型平台的批评。Facebook 在 2021 年 2 月缅甸政变后封禁了与军方有关联的账号，但并未承诺在其他情况下也这样做。当局对平台施压，要求平台关闭报道今年早些时候印度农民抗议活动的记者和人权维护者的账号，平台似乎照做了。¹¹⁹ 2020 年 4 月，Facebook 的服务器在越南被下线，导致该平台减速并且 7 周无法运行，据称之后 Facebook 同意遵照越南政府的要求，大幅增加对“反国家”内容的审查。¹²⁰

79. 由于没有明确政策，公司很容易受到压力，被要求打击合法政治言论，并为国家煽动的虚假信息提供便利。公司的社区标准应以国际人权标准为基础，包括与表达自由权利相关的标准，根据这些标准，政客、政党和其他公众人物及实体的言论将受到高度保护。国际人权法只准许出于非常有限的原因，包括煽动暴力和仇恨，删除公众人物发布的内容。公司如能以国际标准作为服务条款的坚实基础，将更有能力抵御要求其删除合法言论的压力，因为它们遵从的文书和原则正是国家所正式承诺的。

F. 透明度、问责与数据获得

80. 在几乎所有与虚假信息和错误信息有关的问题中，缺乏透明度和数据获取一直是公司的主要不足之处。它们阻碍了独立审查，并且影响问责和信任。对用户而言，不透明削弱了他们的权能，否认了它们的主体能动性。

81. 最大的社交媒体公司大多每年发布两次透明度报告，但并不公开更准确、更有意义的信息，说明为解决虚假信息或错误信息而采取的行动。¹²¹ 例如，Facebook 的透明度报告只提供关于删除虚假账号的资料，而不提供关于内容的资料。同样，这些公司不提供资料，说明标记的内容数量或与虚假信息政策有关的对删除内容或暂停账号之决定的上诉。也不提供关于用户参与虚假信息或错误信息的数据，包括转发数量、浏览数量、覆盖范围以及投诉或删除请求的数量。为发现和删除内容而部署的人工智能系统的可靠性和准确性方面，也需要更多资料。¹²² 公司的内容审核系统和流程整体缺乏透明度，因此无法评估公司所采取措施的效力及其对人权的影响。研究人员、学者或民间社会无法充分获取数据，因而无法做出更加独立客观的评估，令这种评估更加困难。

82. 公司与政府间的协议也需要提高透明度，特别是如果协议涉及更加突出政府的信息传递或删除内容或其他言论限制。

¹¹⁹ IND 2/2021。

¹²⁰ 见 www.reuters.com/article/us-vietnam-facebook-exclusive-idUSKCN2232JX。

¹²¹ 例如，见 Working Group on Infodemics, Policy Framework (Forum on Information and Democracy, November 2020), pp. 17 ff., 另见进步通信协会提交的材料。

¹²² 例如，见进步通信协会提交的材料。

六. 结论和建议

83. 在一份关于虚假信息的专门报告中，要忽略数字技术对民主、可持续发展和人权的重要意义，或者忽略意见和言论自由权在这方面的关键地位，这很容易，但很危险。因此，通过损害人权打击虚假信息是短视的做法，并且适得其反。意见和言论自由权不是其中的问题，而是打击虚假信息的目标和手段。COVID-19 大流行暴露出维护权利的紧迫性以及应对虚假信息和错误信息的挑战。

84. 虚假信息是一个复杂而多方面的现象，其影响很严重。它破坏了人们对民主体制的信任。在公共信息制度薄弱，独立调查新闻受到限制的环境中，虚假信息容易盛行。它削弱了个人的权能，剥夺了人们在搜索、接收和分享信息以及形成观点方面的自主权。在平台的世界，个人被视作用户，而非具有主体能动性的权利持有人。

85. 虚假信息存在问题，但国家和公司的应对也存在问题。制定法律和政策时依据的往往并非对线上危害的最优认识，并且数据、研究或公众磋商都不充分。国家采用不相称的措施，例如断网或以模糊宽泛的法律将线上言论定为刑事犯罪，予以封禁、审查和压制并缩小公民空间。这些措施不仅不符合国际人权法，而且更加放大误解，催生恐惧，加深公众对体制的不信任。

86. 公司的应对一直是被动反应，并且不充分也不透明。大型平台专注于改善内容审核，忽视了其商业模式、透明度缺失和用户正当程序权利不足引起的人权关切。

87. 国家、公司和媒体面临的根本挑战是恢复公众对健全的信息秩序的信任。处理虚假信息需要多层面、多利益攸关方的应对措施，这些措施应当以全面的人权和国家、公司、国际组织、民间社会及媒体的积极参与为坚实基础。需要多方利益攸关方对话和伙伴关系，这一点至关重要。

88. 国家是尊重、保护和实现人权之义务的主要承担者。依照尊重人权的义务，国家不应发表、赞助、鼓励或传播它们明知或理应知道为虚假的声明，也不应准许以断网作为打击虚假信息的手段。国家不应限制线上或线下表达自由，除非按照对《公民及政治权利国际公约》第十九条第三款和第二十条第二款之要求的严格精确解读予以限制。

89. 只有在极特殊和最恶劣的煽动暴力、仇恨或歧视的情况下才可使用刑法。刑事诽谤法是殖民历史的遗留，在现代民主社会中没有立足之地。应将之废除。

90. 国家有责任确保公司尊重人权。不应强迫公司删除或屏蔽符合国际人权法的内容，也不应要求公司根据国内法律判定内容的合法性，而应当由法院判定。公司对于国家的此类要求应透明处理，不应进行非公开交易。

91. 国家对社交媒体的监管应侧重于提升透明度，用户的正当程序权和企业的人权尽职，以及确保以法律明确界定、保障和限制监管机构的独立性和职权。

92. 数据保护是重新导向广告驱动的数字经济商业模式之关键，这种模式助长了信息无序和相关侵犯人权行为。国家应出台强有力的数据保护法，更新选举法和其他相关法律，以限制对个人及其线上活动的大面积跟踪和锁定目标。

93. 多样化和可靠的信息显然是对治虚假信息和错误信息的良药。国家应履行其确保信息权的义务，首先应提高自身透明度，在线上 and 线下主动公开官方数据，其次应重申其对媒体自由、多样性和独立性的承诺。恢复公众对公共领域作为安全的民主议事场所的信心之关键是在线上和线下确保记者的安全，消除对威胁、恐吓、骚扰、攻击和杀害记者(包括女记者)、博主、漫画作者和人权维护者的行为有罪不罚的现象。

94. 大会最近指出，媒体信息和数字素养增强了人们的权能，增强了他们对虚假信息和错误信息的应对能力。¹²³ 它应成为国家学校课程的内容，并让年轻人和老年人都参与其中。除了数字素养，还必须更加关注数字包容，以便使目前完全依靠社交媒体平台和通讯应用进行连通(采用零费率)的发展中国家的人民能够真正获得免费、开放、可互操作、可靠且安全的互联网接入。

95. 根据国际人权法，公司有义务尊重人权。数字平台虽是私人行为体，却在公共空间对人权有深远影响。因此，他们不仅应对用户负责，也应对整个社会负责。人们日益关切最大的公司对市场的主导以及它们目前商业模式的有害影响。公司应积极回应这些关切，不止应改善内容审核，还应审查自身商业模式，承认用户作为权利持有人的主体能动性和自主权，并且应增加透明度、监管和选择，确保正当程序，从而增强用户权能。

96. 根据《工商企业与人权指导原则》，社交媒体公司应审查其商业模式，并确保其业务运营、数据收集和数据处理做法符合国际人权标准，包括《公民及政治权利国际公约》第十九条，以及数据保护原则和相关的国家消费者保护标准。它们还应对其产品进行人权影响评估，特别是评估算法和排名系统在放大虚假信息或错误信息方面的作用。此类评估应定期进行，全国选举或 COVID-19 大流行等重大危机前后也应当进行。

97. 公司应审查自己的广告模式，以确保其不对观点和想法的多样性产生负面影响，并确保其中的定向广告有清晰标准。应在线上广告库中提供关于广告主的有意义的信息，并让用户能够手动选择打开接受广告投放。

98. 公司应与所有相关利益攸关方协商，而后遵循国际人权法，采用清晰精准的内容以及关于虚假信息和错误信息的广告投放政策。还应采取符合国际人权标准的关于公众人物的清晰政策，并在各个地理区域一贯适用。应确保所有政策便于用户查阅和理解，并且因地制宜，一贯执行。

99. 公司应提供关于其算法或推荐系统参数的清晰而有意义的信息，并确保这些系统默认让用户接收多样化的观点，同时让用户能够选择影响自己线上体验的变量。

100. 公司应发布全面、详细且适合特定背景的透明度报告，包括在发生 COVID-19 大流行等特殊情况下发布单独报告，报告中应细分针对与虚假信息或错误信息相关的内容采取的行动以及对这些行动的上诉，应包括转发数量、观点、影响范围、投诉和删除请求这些信息。

¹²³ 第 75/267 号决议。

101. 用户必须有适当追诉权。公司应当设置内部上诉机制，将之用于更广泛的内容审核决定和内容类型，例如协作虚假行为。还应探索建立外部监督机制，例如社交媒体理事会。

102. 在 Me Too 运动的时代，国家和公司应优先应对线上性别虚假信息，并特别关注其在现实世界中的影响。公司应针对平台体验的各个面向，与受这种恶行影响的人协商，出台从性别视角出发制定的适当政策、补救措施和机制。国家还应将顾及性别平等的视角充分纳入其政策和方案，以应对虚假信息和错误信息，包括在媒体、信息和数字素养方案中。

103. 公司作为全球舞台上的行为体，应投入更多资源并借鉴当地民间社会和虚假信息之目标群体的观点，加深对助长虚假信息和错误信息的当地背景的了解，解决公司在知识、语言、政策及服务方面与发展中国家、少数群体和其他弱势群体的不对等。

104. 在处理虚假信息和错误信息方面的一个主要挑战是：无法获得数据，特别是与发展中国家有关的数据而产生的知识差距。应更加重视为研究、决策、监测和评估提供数据。“差分隐私”的概念¹²⁴或许是一个方向，可以在为研究目的获取海量数据的同时尊重用户的人权与安全关切。

105. 最后，同样重要的是，在确保应对虚假信息和错误信息的所有努力以国际人权法为坚实基础，包括尊重意见和言论自由方面，联合国人权系统，特别是人权理事会，具有重要作用。人权理事会应考虑与国家、公司、民间社会组织及相关国际和区域行为体定期进行多方利益攸关方协商，并考虑针对在数字空间保障和促进人权这一议题制定举措。

¹²⁴ Working Group on Infodemics, *Policy Framework*, p. 125.