



## 人权理事会

### 第四十六届会议

2021 年 2 月 22 日至 3 月 19 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

## 少数群体问题

### 少数群体问题特别报告员费尔南·德瓦雷纳的报告\*\*\*

#### 概要

在本报告中，少数群体问题特别报告员费尔南·德瓦雷纳概述了他自上次报告(A/HRC/43/47)以来开展的活动，并就社交媒体上广泛针对少数群体的仇恨言论作了专题报告。他描述了相关现象，包括国家当局普遍否认或不承认少数群体受到所禁止的仇恨言论之害，或未能有效保护他们免受这些言论之害。他最后强调，国家、民间社会和社交媒体平台有责任承认仇恨言论主要是少数群体问题，并有责任作为紧急事项，采取进一步措施，充分有效地履行相关人权义务。

\* 本报告逾期提交，以纳入最新信息。

\*\* 本报告附件不译，原文照发。



## 一. 导言

1. 少数群体问题特别报告员的任务由人权委员会 2005 年 4 月 21 日第 2005/79 号决议确立，人权理事会后来在一系列决议中延长了任务期限。
2. 特别报告员费尔南·德瓦雷纳于 2017 年 6 月 26 日获人权理事会任命，并于 2017 年 8 月 1 日就职。2020 年，理事会第 43/8 号决议将任务期限延长三年。
3. 特别报告员很荣幸被委任这项任务，并感谢人权理事会对他的信任。他还要感谢联合国人权事务高级专员办事处(人权高专办)对履行该任务的支持。

## 二. 特别报告员的活动

4. 特别报告员谨提请人权理事会注意该任务的网页，其中提供该任务相关活动的一般信息，包括往来信函、新闻声明、公开露面、国别访问和专题报告。<sup>1</sup> 向大会第七十五届会议提交的报告(A/75/211)概述了任务负责人在 2020 年 1 月 1 日至 7 月 1 日期间的活动。
5. 特别报告员继续把重点放在提高联合国各机构和会员国，以及更广泛地说，提高普通公众及其他区域和国际组织对少数群体问题的关注度和认识上——特别是与他的主题优先事项有关的问题(无国籍状态；少数群体的教育、语言和人权；社交媒体上针对少数群体的仇恨言论；以及预防族裔冲突)。他还把重点放在改善该任务下各项活动，如少数群体问题论坛可及性的新方法上。
6. 这项工作已包括两项主要举措：
  - (a) 与汤姆·兰托斯研究所及许多区域少数群体和人权组织合作，巩固每年继续就与少数群体问题论坛相同的主题举办的区域论坛；
  - (b) 为了本任务的目的，澄清少数群体概念的工作定义，以及联合国文书承认的四类少数群体(民族或族裔、宗教和语言少数群体)的重要性和范围。
7. 特别报告员已将纪念《在民族或族裔、宗教和语言上属于少数群体的人的权利宣言》通过三十周年作为一个优先事项，预计 2022 年将在纽约举行的大会第七十七届会议期间举办一次高级别活动，并开展相关活动。

### A. 国别访问

8. 特别报告员期待继续与喀麦隆、印度、约旦、肯尼亚、尼泊尔、俄罗斯联邦、南非、南苏丹、阿拉伯叙利亚共和国、东帝汶、美利坚合众国和瓦努阿图进行对话，他已向这些国家提出访问请求。他感谢巴拉圭政府表示愿意在 2021 年接受特别报告员的访问。
9. 在访问中，特别报告员的重点是处理无国籍人、达利特人和罗姆人等特别弱势的少数群体、遭受双重甚至三重边缘化的少数群体妇女，以及聋人和重听者(作为手语使用者，他们属于语言少数群体)所面临的歧视、排斥和其他侵犯人权行为的重要性。

<sup>1</sup> [www.ohchr.org/EN/Issues/Minorities/SRMinorities/Pages/SRminorityissuesIndex.aspx](http://www.ohchr.org/EN/Issues/Minorities/SRMinorities/Pages/SRminorityissuesIndex.aspx).

10. 应吉尔吉斯斯坦政府邀请，特别报告员于 2019 年 12 月 6 日至 17 日对吉尔吉斯斯坦进行了正式访问(见 A/HRC/46/57/Add.1)。

## B. 往来信函

11. 特别报告员根据从不同来源收到的关于侵犯民族、族裔、宗教和语言少数群体人权的资料，向有关会员国发出了指控信和紧急行动信。这些信函及其答复可公开查阅。

12. 自 2020 年 1 月以来，共向各国政府发送了 78 封信函，与上一年的信函总量相比有大幅增加。所有信函都是与其他特别程序任务负责人联合发出的，其中 16 封是紧急呼吁，53 封是指控信，9 封是表达立法和政策关切的其他信件。

13. 发送给亚洲和太平洋国家的信函最多(43 封)，其次是欧洲和中亚(16 封)、中东和北非(9 封)、美洲(5 封)和撒哈拉以南非洲(4 封)。其中一封信函是致一家私人公司的。

## C. 会议和提高认识活动

14. 特别报告员自 2017 年 6 月获人权理事会任命以来，一再强调提高对少数群体人权的认识 and 关注度，这是其任务的一个重要方面。因此，特别报告员经常参与和推动在世界各地以及与各种政府和非政府组织举办的国际、区域和国家研讨会和会议。他还经常就少数群体的人权问题接受媒体采访。

15. 特别报告员在 2020 年 7 月至 2020 年 12 月期间开展的主要活动概要载于本报告附件一。特别报告员提交大会第七十四届会议的报告(A/74/160)概述了此前开展的活动。

## 三. 2020 年少数群体问题论坛的最新情况

16. 特别报告员在 2018 年向人权理事会提交的第一份报告中提出，需要采取更具区域性的办法举办少数群体问题论坛，以方便世界不同地区的少数群体参与，并更多地体现区域关切和背景(A/HRC/37/66, 第 64 段)。2019 年，为实施这一办法采取了第一步措施，举办了三个区域论坛。2020 年，由于冠状病毒病(COVID-19)大流行造成的不确定性，所设想的四个区域论坛中只有两个得以举行，其主题优先事项是处理通过社交媒体发布对少数群体的仇恨言论和煽动对他们的仇恨问题。尽管如此，2020 年仍有 400 多人参加了在欧洲和亚洲及太平洋举行的区域论坛。区域论坛的举行得益于汤姆·兰托斯研究所的协调以及许多非政府组织、国家代表及区域和国际组织，如联合国教育、科学及文化组织、欧洲安全与合作组织(欧安组织)、欧洲联盟、欧洲委员会和其他组织的贡献和协助。

17. 人权理事会在 2007 年第 6/15 号决议中设立了少数群体问题论坛，并在 2012 年第 19/23 号决议中重申了其重要作用。该论坛的任务是提供一个平台，在民族或族裔、宗教和语言少数群体问题上促进对话与合作，并为少数群体问题特别报告员的工作提供专题意见和专业知识。特别报告员负责指导论坛的工作，筹备年度会议，并向人权理事会报告论坛的专题建议。论坛每年在日内瓦举行为期两个工作日的会议，进行专题讨论。近年来，论坛参加者通常超过 600 人。

18. 论坛第十三届会议于 2020 年 11 月 19 日和 20 日在线举行，主题为“仇恨言论、社交媒体和少数群体”。参加者不到 400 人，与前几年相比大幅减少，原因在于大流行造成的举办论坛的不确定性，以及人权高专办很长一段时间没有顾问，无法组织论坛。缺乏全职顾问意味着，与特别报告员一道工作但以前从未参加论坛的人权高专办工作人员被指派在举行为期两天的会议之前，确保论坛取得成功。正是由于在 11 月论坛之前可用的工作人员有限，缺乏一名具有论坛经验的顾问，以及论坛采取了未曾尝试过的在线形式，导致参加人数大大减少。

19. 关于论坛及其建议的完整报告将向人权理事会单独提交(A/HRC/46/58)。

#### 四. 专题报告：仇恨言论、社交媒体和少数群体

20. 特别报告员遗憾地表示，也许具有讽刺意味的是，关键时刻的一次计算机故障导致数据大量丢失，专题报告编写工作延迟。

##### A. 引言

21. 2020 年有两种大流行病迅速蔓延：一种是身体上的疾病，另一种是心灵上的疾病。COVID-19 大流行的肆虐有时掩盖了社交媒体上的心灵疾病，这种疾病往往会导致暴行、屠杀，甚至是种族灭绝呼吁。虽然 COVID-19 疫情正在得到处理，并可能会在一年内得到控制，但心灵疾病并没有减弱的迹象。相反，社交媒体上的仇恨言论正在传播和加强。少数群体是仇恨及煽动暴力和歧视的主要受害者。在关于社交媒体上的仇恨言论或关于仇恨犯罪的现有分类数据中，少数群体约占目标对象的 70%或以上。

22. 仇恨言论的威胁首先影响到少数群体。无论是否出于疏忽，该领域的许多行为者都没有系统地在名义上承认谁是种族主义、偏见、寻找替罪羊、甚至是社交媒体上煽动暴力行为的主要目标。由于没有具体提到少数群体，仇恨言论的严重程度和残忍性遭到忽视，甚至被泛泛而谈的迷雾所掩盖。从某种意义上说，如果不指出主要受害者，每个人都会成为仇恨的帮凶。结果是滋生不容忍和排斥提供了肥沃的土壤，而这二者是仇恨少数民族的根源。雪上加霜的是，仇恨对一些人来说已经变得极其有利可图，对另一些人来说也是有用的。

23. 本专题报告审查了打击社交媒体上的仇恨言论方面的一些不足，包括少数群体在多大程度上是社交媒体上仇恨言论的主要目标，以及社交媒体平台的所有者在多大程度上仍然忽视涉及国际法中人权义务的各种形式的仇恨言论，或没有采取足够或有效的措施来遏制这些言论的传播。更为复杂的是，社交媒体上的仇恨言论实际上涉及这些平台——即全球最大和盈利最多的公司——实际所有者中的非国家私人第三方。

24. 仇恨言论、社交媒体和少数群体的主题将非洲、亚洲、美洲和欧洲许多最弱势的群体——例如非洲人后裔、亚洲群体、达利特人、罗兴亚人和罗姆人等少数民族，以及巴哈伊教徒、穆斯林、犹太人和基督教徒等宗教或信仰少数群体——与一些最强大的企业实体的利益对立起来，而国家则为难地夹在二者之间。

25. 在使用社交媒体等在线服务和在这些媒体中的形象方面，少数群体尤其处于弱势或不利地位。除了成为社交媒体上仇恨言论的主要目标外，他们还可能受到

内容审查系统中的限制和删除措施不成比例的影响，这是由于用户和第三方对通知系统本身的偏见，以及内容审查工具中使用的算法所固有的偏见，因为每个算法的背后都是人。

26. 这种情况与约 20 年前所预期的不同，当时人们认为——这或许有些天真——互联网及其产物社交媒体只会为人类提供令人惊叹的机会，以快速、低成本的方式进行沟通和交流，并支持几乎即时的信息和教育跨境传输。善恶二分法与世界本身的故事一样古老，但它被扩大了，而且由于社交媒体平台一些所有者的商业模式，遭到蓄意的引导，使之货币化和更加有利可图。

27. 社交媒体的阴暗面现在不容忽视，因为它们已成为危险论调、阴谋论、错误信息、虚假信息、种族主义和将少数群体作为替罪羊的温床及宣传工具。例子包括一些最原始的反犹太宣传形式的死灰复燃，甚至回溯到一个多世纪前的锡安长老协议。2021 年 1 月美国国会大厦发生的事件表明，社交媒体可被用来直接破坏民主稳定的社会，并作为暴力宣传工具。

28. 本报告审议了其中一些方面，以期更好地了解 and 强调国家、公司和其他行为者在处理社交媒体上最普遍的仇恨言论，即针对民族或族裔、宗教和语言少数群体的仇恨言论方面的人权义务，并提供相关指导。

## B. 国际人权法：以表达自由为起点

29. 甚至在对人权的现代表述从第二次世界大战及其恐怖的灰烬中产生之前，意见和表达自由就已经被视为自由、稳定和民主社会的支柱之一。它们提供了一个安全的空间，让个人可以做自己，自由表达其不同和多样性。从某种意义上说，意见和表达自由是民主的氧气。如人权事务委员会所指出的，它们是人的全面发展不可或缺的条件，是每个自由民主社会的基石。<sup>2</sup>

30. 意见和表达自由在国际和区域文书中得到广泛承认，是所有人的基本权利。<sup>3</sup> 促进和保护意见和表达自由权特别报告员在其 2019 年提交大会的关于网上仇恨言论的专题报告(A/74/486)中明确指出，意见和表达自由应被视为一个默认起点，对其设置限制时应非常谨慎。

31. 本报告没有重复促进和保护意见和表达自由权特别报告员在其报告中已经有力而雄辩地对意见和表达自由的核心重要性和作用所作的详细而明确的阐述。相反，关于社交媒体的本报告侧重于这位同事和其他同事以及联合国实体在很大程度上忽略或未涉及的内容，具体而言，即社交媒体上最普遍的仇恨言论形式，如针对少数群体的仇恨言论，以及是否可以、可在什么情况下和何种程度上对表达自由加以限制。

32. 应当明确指出的是，甚至在 2013 年的一份报告中，联合国人权事务高级专员就对许多国家的媒体、政界人士和社会其他行为者对移民和少数群体所作的往往非常负面的描述表示震惊，她呼吁采取措施遏制日益增长的仇外态度，并强调

<sup>2</sup> 关于见解自由和言论自由的第 34(2011)号一般性意见，第 2 段。

<sup>3</sup> 《消除一切形式种族歧视国际公约》，第五条；《儿童权利公约》，第 13 条；《残疾人权利公约》第二十一条；《保护所有移徙工人及其家庭成员权利国际公约》，第 13 条；《美洲人权公约》，第 13 条；《非洲人权和民族权宪章》，第 9 条；《非洲儿童权利与福利宪章》，第 7 条；《阿拉伯人权宪章》，第 32 条；以及《保护人权与基本自由公约》，第 10 条。

需要迅速谴责仇恨言论，起诉那些涉嫌煽动种族暴力和实施了种族主义和暴力行为的人(A/HRC/22/17/Add.4, 附件, 第 3 段)。她的报告载有一系列专家研讨会的建议和结论，专家们在这些研讨会之后通过了《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》(同上, 附件, 附录)。

33. 这一过程中出现的仍需大力强调的一点是：社交媒体上绝大多数仇恨言论案件的受害者和目标是少数群体。因此，从逻辑上讲，打击仇恨言论方面的澄清、优先事项和努力应以少数群体为重点。但很大程度上情况并非如此，在大多数联合国倡议和其他倡议中，很少提到少数群体——那些最常受到仇恨言论影响的群体。认为不用提及仇恨言论主要涉及少数群体这一点就有可能解决世界大约四分之三仇恨言论案件的根源，并从而制定针对这一现象的相关战略，这种想法至少是令人费解的。

34. 无论如何，处理社交媒体上仇恨言论的出发点必须是充分尊重表达自由。因此，无论仇恨言论的性质或内容如何，国际法只对表达自由规定了有限的可允许限制。故如本报告所述，必须明确规定最直接相关的各方特别是国家和社交媒体平台所有者的人权义务和责任，因为它们与社交媒体上最常见的仇恨言论，即针对少数群体的仇恨言论有关。

## C. 考虑具体背景：仇恨的大流行和采取行动打击社交媒体上所禁止形式的仇恨言论的义务

### 1. 蔓延的思想毒药

35. 除绝对必要外，表达自由不应受到限制。遗憾的是，到目前为止，各国和社交媒体平台的所有者基本上无法或不愿承认和应对世界上流传的针对少数群体的最恶劣和最危险的仇恨形式，这些仇恨不仅毒害人的心灵，还威胁到整个国家的和平与稳定。

36. 秘书长确实认识到，这种大流行引发了新一轮的仇恨、仇外、找替罪羊和危言耸听等行为，网上和街头的排外情绪高涨，反犹太阴谋论蔓延，并发生了与 COVID-19 有关的反穆斯林袭击事件。<sup>4</sup> 更准确地说，这种仇恨主要针对少数群体。一些人将这种大流行武器化，另一些人则将其货币化，过去十年愈发如此。少数群体和其他弱势群体占煽动仇恨受害者的大多数。

37. 除秘书长提到的群体外，欧洲许多地区的罗姆人也一直是社交媒体平台传播仇恨和暴力毒药的首选目标。南亚的达利特人也是受害者，此外还有移民和难民，无论如何，他们也属于少数群体。反犹太主义和仇视伊斯兰现象在网上迅速发展和扩大。<sup>5</sup>

38. 在社交媒体上，世界经历的事情包括 2017 年呼吁对尼日利亚北部的伊格博少数民族，2018 年呼吁对缅甸的罗兴亚少数民族实行种族灭绝。还有些关于在

<sup>4</sup> 见“United Nations guidance note on addressing and countering COVID-19 related hate speech”，11 May 2020。

<sup>5</sup> Savvas Zannettou and others, “A quantitative approach to understanding online antisemitism”, 24 November 2019. 可查阅 [www.arxiv.org/pdf/1809.01644.pdf](http://www.arxiv.org/pdf/1809.01644.pdf)。

某些国家对巴哈伊人、穆斯林和基督教少数群体，在另一些国家对达利特人实施暴力、谋杀、强奸和其他暴行的呼吁，以及在欧洲煽动对罗姆少数群体的暴力，在美洲许多地区煽动对非洲人后裔的暴力等行为。移民也在少数群体之列，他们仍然被视为犯罪和疾病的来源，或被称为威胁或败坏风气者。“在少数几家堪称历史上最大的宣传机器的互联网公司推动下”，<sup>6</sup> 针对少数群体的仇恨言论、仇恨犯罪和暴行激增。

39. 社交媒体还将少数群体描绘成一种威胁或危险，从而助长了针对他们的暴力行为。例如，2018 年，一名枪手在美国匹兹堡市的一个犹太教堂杀害了 11 位礼拜者，他利用社交媒体网络 Gab 传播一种阴谋论，即犹太人将移民带到美国，是为了消灭美国的白人。犹太人是“大取代”论的目标对象，匹兹堡枪击事件就是一个例证，这种论调似乎近来源于法国的极右翼，按照该论调，某些非白人或非基督教少数群体最终将通过移民和高出生率，从人口学的角度征服并威胁“欧洲”文化。

40. 在印度阿萨姆邦，正在进行一项更新国家公民登记册的活动。Avaaz 组织最近对 Facebook 上的 800 个帖子进行了研究，发现针对孟加拉移民的仇恨言论占主导地位，这些移民被公开称为“罪犯”、“强奸犯”、“恐怖分子”、“猪”和其他非人化称谓。这些帖子已经被分享了近 10 万次，暴力仇恨言论的浏览量达到约 540 万次。“平等实验室”(Equality Labs)在全印度范围内对脸书上的仇恨内容开展了另一项类似研究，并将该平台上的仇恨内容分类如下：37%的帖子系仇视伊斯兰(包括反罗兴亚人的材料、提到“爱情圣战”的帖子、对早期暴力侵害穆斯林事件的颂扬，以及仇视伊斯兰的亵渎性语言)，16%是假新闻，13%针对性别或性行为，13%针对种姓少数群体(这些帖子中有 40%针对基于种姓的名额保留，其余的则包括种姓诽谤、反安贝德卡的信息、谴责种姓间个人关系的帖子等)，9%针对其他宗教少数群体。

41. 就仇恨言论的形式和内容而言，上述例子是全世界的普遍现象，而不是例外。欧安组织 2018 年涵盖欧洲国家和其他欧安组织国家的仇恨犯罪数据显示，76%以上的仇恨犯罪涉及犹太人、穆斯林、罗姆人及其他族裔和宗教少数群体。

42. 根据向特别报告员提供的资料，令人难过的是，在社交媒体上，少数群体越来越多地成为讽刺和仇恨的受害者，其残忍和敌意令人发指。

43. 仇恨言论会引发仇恨犯罪，就如错误信息和虚假信息一样。正如提交特别报告员的一份材料所指出的，大屠杀不是始于毒气室，而是始于针对一个少数群体的仇恨言论。虚假信息最终可能会对少数群体有害，甚至致命，并导致所谓的社交媒体私刑暴力或“WhatsApp 私刑”。2014 年，法国发生了一起众所周知的案件，造成对一些罗姆少数民族的人身攻击，此前社交媒体上有假信息称，罗姆人绑架了一辆白色面包车上的一名儿童。2018 年在斯里兰卡发生的一起更致命的案件中，主要在脸书上流传的关于穆斯林少数群体密谋对占人口多数的僧伽罗人实行绝育的传言造成了死亡，几个城镇的暴民焚烧了穆斯林少数群体的清真寺、商店和房屋。这类针对少数群体的例子不胜枚举。

<sup>6</sup> Anti-Defamation League, “Sacha Baron Cohen’s keynote address at ADL’s 2019 Never Is Now summit on anti-semitism and hate”, 21 November 2019.

44. 特别报告员提到一份来文中表示的关切，即非人化的语言往往将少数群体贬称为动物或昆虫，使针对这些群体的暴力行为正常化，并使得对他们的迫害和排除变得可以接受，当这些侵犯行为带有歧视或偏见的意图时，就会成为妖魔化和非人化的途径，从而可能导致种族灭绝。个人可能会发现自己被社交媒体拉到非人化的语言和仇恨环境中，并最终被观点相似的人包围。因此，他们可能陷入社交媒体中的证实偏向，这种“孵化”环境变得特别有利于表达——实际上是加强和巩固了——针对某些被作为替罪羊的少数群体的种族主义、不容忍甚至暴力观点。

## 2. 法律和监管环境

45. 一份关于全球网上仇恨监管的重要汇编显示，在对仇恨言论的法律禁止、概念界定、限制类型以及提供补救措施方面，有各种各样的方法。<sup>7</sup> 虽然不可能进行归纳，但该简报确实表明，没有处理仇恨言论的单一方法，而且往往没有一种方法专门针对社交媒体上仇恨言论的特殊性及其带来的挑战。

46. 向特别报告员提交的一些材料突出表明，限制社交媒体上仇恨言论的措施，特别是那些旨在保护少数群体的措施，往往得不到执行。据指出，在一些国家，没有关于社交媒体上仇恨言论案件的数据，在有些情况下，关于仇恨犯罪的现有法律从未使用过，或因过于繁琐或模糊而无法成功援引进行起诉。一些材料表明，少数群体在向有关当局提起仇恨言论案件时犹豫不决，因为他们自己的经验表明，公共当局不会干预，违法者不会承担任何后果，或者对社交媒体使用审查或投诉机制不大可能纠正这种情况。公共当局和社交媒体平台所有者缺乏有效的法律和其他应对措施，其后果可能是悲剧性的，甚至是致命的，并导致大规模暴行和侵犯人权行为，脸书在缅甸未能解决煽动反对罗兴亚少数民族的问题就证明了这一点。在某次族裔暴力运动之前和期间，社交媒体被用来妖魔化罗兴亚少数民族，缅甸问题国际独立实况调查团证实，脸书一直是寻求传播仇恨者的一个有用工具(A/HRC/39/64, 第 74 段)。结果发生了一场可预见和有计划的人权灾难，作为种族清洗和可能的种族灭绝企图的一部分，实施了轮奸，有数千人被杀，学校、市场、住宅和清真寺被烧，所有这些导致了一场可怕的人道主义危机，波及数十万罗兴亚少数民族的男子、妇女和儿童。

47. 从国际法的角度来看，公司，如社交媒体平台，并不负有与政府相同的义务。但政府至少有禁止煽动种族灭绝和构成煽动歧视、敌对或暴力的鼓吹行为的直接义务。鉴于社交媒体如今在传播仇恨言论方面有很大影响，严重侵犯了数百万人的权利，甚至威胁到他们的生命和安全，各国政府有义务采取措施，确保禁止煽动种族灭绝或构成煽动歧视、敌意或暴力的鼓吹仇恨行为。

48. 一个反复出现的问题是，在让社交媒体平台成为煽动种族灭绝或构成煽动歧视、敌意或暴力的鼓吹行为的工具或沃土时，它们是否应像主流传统媒体那样承担后果和受到惩罚。上述关于全球网络仇恨监管的汇编表明，社交媒体平台往往不会面对这样的后果和惩罚。

<sup>7</sup> Hogan Lovells and PeaceTech Lab, *The Global Regulation of Online Hate: A Survey of Applicable Laws* (December 2020). 可查阅 [www.hoganlovells.com/~media/hogan-lovells/pdf/2020-pdfs/2020\\_12\\_15\\_the-global-regulation-of-online-hate\\_a-survey-of-applicable-laws\\_special-report\\_december-2020.pdf?la=en](http://www.hoganlovells.com/~media/hogan-lovells/pdf/2020-pdfs/2020_12_15_the-global-regulation-of-online-hate_a-survey-of-applicable-laws_special-report_december-2020.pdf?la=en)。

49. 特别报告员获悉了一些平台所有者最近为应对可能构成煽动仇恨或伤害的信息而采取的措施，例如亚马逊最近暂停了 Parler 平台的网络托管服务，因为 Parler 未能监管并删除暴力内容。

50. 特别报告员还注意到并欢迎设立了一个由独立成员组成的监督委员会，该委员会将在尊重表达自由和人权义务的基础上，就脸书和 Instagram 应当允许或删除的内容作出最终和具有约束力的决定，以消除任何煽动种族灭绝或构成煽动歧视、敌意或暴力的鼓吹民族、种族或宗教仇恨言论。

51. 然而，提交材料中反复提到的一点是，许多针对社交媒体上仇恨言论的立法和其他措施过于模糊或薄弱，或者不够充分。仇恨言论这一祸害及其引发的伤害和暴力不仅在继续，还在加剧，有可能破坏社会稳定，最近对美国国会大厦和德国国会大厦的袭击就是明证。

#### D. 为防止伤害而允许对表达自由进行的限制

52. 如表达自由问题特别报告员在其 2019 年报告(A/74/486, 第 26 段)中所强调和国际人权机构判例广泛承认的那样，有些内容对许多人来说可能是冒犯性的、有争议的、不受欢迎或令人不安的，但表达自由保护表达此类信息和想法的权利。少数群体尤其有可能成为极其广泛的可能具有冒犯性、不宽容的言论甚至是种族主义言论的目标：政府是否应当禁止任何此类形式的言论，或政府有权在多大程度上严厉或令人厌恶地限制表达自由，都是有待讨论的问题。然而，作为一个基本原则，对任何形式表达的限制仍然必须是一个例外，只有在国际人权法承认的三种情况下才能允许这样做，具体如下所述。

53. 根据以往的报告——如表达自由特别报告员 2019 年的报告、前少数群体问题特别报告员 2015 年的报告(A/HRC/28/64)以及其他不胜枚举的报告——确定可以限制表达自由的例外情况是有可能的。在国际人权法中，如《公民及政治权利国际公约》第十九条和第二十条(广义地说，在其他人权条约中所重申的)和《防止及惩治灭绝种族罪公约》所载，可在三种严格界定的情况下限制表达自由：

(a) 各国必须将构成煽动种族灭绝的言论定为犯罪；<sup>8</sup>

(b) 各国义务以法律禁止——尽管不一定将其定罪——任何构成煽动歧视、敌意或暴力的鼓吹民族、种族或宗教仇恨言论；<sup>9</sup>

(c) 各国可进一步限制表达自由权，但须以经法律规定并为尊重他人的权利和名誉或为保护国家安全、公共秩序、或公共卫生或道德所必要者为限。<sup>10</sup>

54. 值得注意的是，尽管在很大程度上同样未被承认，但国际法允许限制表达自由的前两种情况与少数群体更易受到多数人口和国家政府的仇恨和暴力有关。例如，现代历史上大多数种族灭绝的目标都是少数群体，如犹太人、罗姆人、波什尼亚克人和图西人。构成煽动的鼓吹“民族、种族或宗教”仇恨这种提法再次指出了世界各地最有可能受到歧视、敌视和暴力的少数群体的主要特征。

<sup>8</sup> 根据《防止及惩治灭绝种族罪公约》第三条(c)项，“直接公然煽动灭绝种族”应予惩治。

<sup>9</sup> 《公约》第二十条第二款。

<sup>10</sup> 同上，第十九条第三款。

55. 虽然允许限制表达自由的三种情况看似不相关，但实际上三者之间有一个未被重视的连接线：即伤害，包括实际和可能的伤害。这三种情况都涉及严重程度不同的伤害性言论，其中显然最有害的威胁是煽动种族灭绝，即毁灭一个民族、族裔、种族或宗教群体——通常是少数群体——的生命。这种最恶劣形式的伤害意味着要承担最大的义务：将这种形式的表达定为犯罪。

56. 第二严重的伤害形式被认为是任何基于民族血统、种族或宗教的任何鼓吹或表达仇恨的行为。这一限制并非针对所有不容忍、偏见或令人不适的成见。它的范围非常窄。各国义务以法律禁止达到构成煽动歧视、敌意或暴力这一门槛的表达形式；也就是说，鼓动另一人以歧视、有敌意或暴力的形式对他人实施伤害的行为。只有最严重的煽动歧视、敌意或暴力的形式才应定罪。如《拉巴特行动计划》(A/HRC/22/17/Add.4, 附件, 附录, 第34段)所述, 在不太严重的情况下, 应考虑民事制裁和补救措施, 包括金钱和非金钱损害赔偿, 以及更正权和答辩权。还应考虑行政制裁和补救措施, 包括由各专业机构和监管机构确定并实施的制裁和补救措施。

57. 第一类情况涉及将某些表达形式定为犯罪, 第二类涉及国家予以禁止, 第三类则涉及允许但不强制各国以有限的理由监管和限制言论的情况。在这种情况下, 因有可能造成伤害, 可根据人权法允许各国对行使表达自由权施加有限的限制。

58. 允许这种限制的条件严格而苛刻, 如表达自由特别报告员在其 2019 年报告(A/74/486, 第6段)中所指出的: “各国只有在法律规定和为尊重他人的权利或名誉, 或保护国家安全、公共秩序、公共卫生或道德所必需的情况下, 才可……限制表达。这些都是狭义的例外情况……限制言论的当局应负责证明限制是合理的, 而不是由发言者来证明他们有权发表这样的言论。”他接着指出, 任何限制都必须满足三个条件: 即合法性, 正当性, 以及必要性和相称性。

59. 因此, 社交媒体上的仇恨言论可能涉及根据国际人权法可允许限制的三种情况之一。在国际法中的表达自由这一基本支柱下, 不得限制任何其他可能应受谴责、令人不适或令人反感的表达。

60. 在考虑各种形式的仇恨言论和社交媒体时, 一些与承认和保护少数群体人权有关的具体问题往往没有充分涉及:

(a) 各国在多大程度上遵守其法律义务, 将在社交媒体和其他媒体上煽动种族灭绝定为犯罪;

(b) 各国在多大程度上遵守其法律义务, 禁止社交媒体和其他媒体上构成煽动歧视、敌意或暴力的鼓吹民族、种族或宗教仇恨言论;

(c) 制定确保各国遵守这些国际法律义务的立法。除了合法性、正当性及必要性和相称性的一般条件外, 防止一切形式仇恨言论的立法还必须谨慎而严格, 仅在所允许的范围内限制表达自由。许多现有的仇恨言论法律含糊不清, 可以用来、而且确实被用来限制少数群体和人权维护者的表达自由, 包括阻止他们表达合理的不满。含糊不清会使立法变得毫无用处, 甚至不适当地威胁到表达自由, 促使社交媒体平台删除内容, 而不是冒被罚款的风险(这种现象被称为“过度封锁”), 从而对表达自由造成不允许的限制;

(d) 各国在多大程度上起诉违反对社交媒体和其他媒体上的言论自由予以可允许的限制这一规定的行为；

(e) 各国为履行将社交媒体上不允许的有害言论定为犯罪或加以禁止的法律义务而制定的立法，其中规定社交媒体平台所有者的义务，包括可能的经济责任。这些平台是传播最有害形式的仇恨言论，包括煽动种族灭绝和构成煽动歧视、敌意或暴力的鼓吹民族、种族或宗教仇恨言论的有利生态系统；

(f) 制定要求社交媒体平台所有者通过相关政策和规则的立法，这些政策和规则应承认并考虑到国家有义务禁止所述非法形式的有害言论，并可根据国际法就其他形式的言论进行立法。虽然社交媒体平台是私人实体，但这种立法至关重要，因为社交媒体已成为 21 世纪仇恨的主要载体，特别是针对少数群体的仇恨；

(g) 制定要求社交媒体公司参考其产品的人权影响来管理其平台上仇恨言论的立法。因此，这些公司必须制定人权审查程序，具体提到并侧重于有关国家最易受仇恨言论影响的群体，特别是少数群体、妇女和男女同性恋、双性恋、跨性别者和间性者群体；

(h) 制定必须允许公众监督的透明度要求。这种要求在两个主要层面上运作：(一) 国家机构和社交媒体平台关于强制收集煽动种族灭绝和构成煽动歧视、敌意或暴力的鼓吹行为数据的程序和机制。要求各国确保这些数据按照与弱势少数群体有关的仇恨(无论是民族、族裔、种族还是宗教仇恨)或基于性别和性取向等因素影响其他弱势群体的仇恨分类；(二) 为确保有关仇恨言论的现行立法得到切实执行而采取的措施。

61. 只有在法律规定的情况下，即上述每一个具体问题都得到解决的情况下，国家才能对行使言论自由权施加第三类限制，而且仅以处理足以对尊重他人的权利或名誉、保护国家安全、公共秩序或公共卫生或道德构成的危害所必要者为限。不允许对言论自由进行除这些类型的危害之外的限制。

62. 虽然经常有人说，何为不允许这一点太不确定或太依赖于具体情况，但这种说法并不完全正确。《拉巴特行动计划》包括一个由六部分组成的门槛标准，用于评估根据国际法可被判定为非法仇恨言论的表达中的仇恨程度。同样，《关于表达自由和平等的卡姆登原则》的原则 12 对主要相关术语，包括“仇恨”、“鼓吹”和“煽动”，提供了简单明确的法律定义。<sup>11</sup>

63. 对表达自由的任何限制都必须在国际人权法，特别是《防止及惩治灭绝种族罪公约》、《消除一切形式种族歧视国际公约》和《公民及政治权利国际公约》严格界定的范围内。言论自由意味着必须能够表达观点，进行公开辩论以及对意见和机构进行批评，即便是具有冒犯性，或者令人不安。关键在于，必须将构成直接和公开煽动种族灭绝的言论定为犯罪，必须禁止构成煽动歧视、敌意或暴力的鼓吹仇恨言论。

64. 如上所述，根据国际法，一些其他形式的言论——其中一些可被理解为不太严重的仇恨言论——可作为例外予以限制，但须以经法律规定并为尊重他人的权利或名誉或为保护国家安全、公共秩序或公共卫生或道德所必要者为限。

<sup>11</sup> 第十九条，《关于表达自由和平等的卡姆登原则》(伦敦，2009 年)。

65. 以各国必须禁止并在一些情况下将其定为犯罪的仇恨言论形式(煽动种族灭绝和构成煽动暴力、敌意或歧视的鼓吹仇恨言论)为重点,应当明确指出的是,各国必须确保其法律义务得到社交媒体平台的尊重,因为后者是为所禁止的仇恨言论提供沃土的主要生态系统。

66. 另一方面,波兰最近提出立法,禁止社交媒体平台所有者删除内容或不许并未违反波兰法律的用户使用。虽然该措施已经提出,并可视为是为了保护表达自由,但问题依然在于,波兰法律是否符合该国的国际法律义务,即将煽动种族灭绝定为犯罪,并禁止构成煽动歧视、敌意或暴力的鼓吹仇恨言论。换言之,这类立法如以不符合国际法的方式限制社交媒体所有者、保护言论自由,就可能违反人权法。

## E. 社交媒体所有者的角色和责任

67. 社交媒体平台作为私营公司,目前在一些国家享有很大程度的财务豁免和其他豁免权,因为它们是仇恨和暴力的载体,甚至是对少数群体和其他人进行种族灭绝的呼吁的载体。然而,尽管财产和社交媒体等平台的私人所有者有权决定其服务和接纳对象,但如果其行动造成了根据国际人权法必须禁止的伤害形式,也不能凌驾于法律之上。

68. 世界上的少数社交媒体平台,如脸书、谷歌、YouTube 和推特,几乎可以瞬间接触到数十亿人。它们是最富有的公司。其商业模式依赖于参与,以及故意扩大内容以保持用户参与的算法,但它们也是“回音室”,往往是范围狭窄的信息来源,并集中了偏见和成见。

69. 在线平台的商业模式依赖于通过阅读或观看时间的最大化来实现利润最大化。它们通过让广告商精确定位受众来赚钱,因此设计算法将用户引向他们将花费最多时间的社区和主题,是符合其利益的。研究证实,这产生了一个意想不到的后果,那就是让人“掉进兔子洞”,即将他们引向极端或强迫性的内容,如宣扬阴谋论或其他分裂、误导或虚假且令人生厌的视频。例如,2019年6月,为消除新纳粹和白人至上主义视频,YouTube 改变了一种算法,将被视为用于传播错误信息的“踩线内容”视频的浏览量减半。

70. 这些算法助长、集中并传播仇恨和不宽容:几乎三分之二的人加入极端团体(其中大多数人会再次传播针对少数群体的仇恨言论),是因为社交媒体上的算法所推送的建议。<sup>12</sup> 其结果是推动了社交媒体上仇恨、激进化、非人化、找替罪羊、煽动种族灭绝和构成煽动针对少数群体的暴力、敌意或歧视的鼓吹仇恨行为的爆发,导致仇恨犯罪和暴行惊人地增加。<sup>13</sup> 报道广泛证实,网上的仇恨言论与全球针对少数群体的暴力事件增加有关联,包括大规模枪击、私刑和种族清

<sup>12</sup> Guillaume Guichard, “Facebook a minimisé des initiatives internes visant à affaiblir les contenus extrémistes”, *Le Figaro*, 27 May 2020.

<sup>13</sup> 根据欧洲委员会人权事务专员 Dunja Mijatović 在其 2019 年年度活动报告(2020 年 4 月 21 日)中所述,“原住民主义、仇视伊斯兰和反吉普赛主义已达到令人震惊的程度……针对罗姆人的仇恨言论和犯罪也(仍然)普遍存在。”

洗。<sup>14</sup> 仇恨带来了营利，少数群体却受到伤害：社交媒体平台利润丰厚，而少数群体通过这些平台遭到的仇恨和煽动暴力行为越来越多。

71. 在认识到社交媒体上针对少数群体的仇恨言论规模之后，情况就开始有了变化——尽管有些延迟。如提交给特别报告员的材料中所指出的，即使社交媒体平台制定了审查政策，这些政策也可能过于缓慢、无效或宽泛。<sup>15</sup> 其实施往往不能保护最弱势的人免受伤害。此外，少数群体发布的内容尤其容易受到仇恨言论的攻击、被妖魔化或成为替罪羊，与多数人发布的包含歧视性言论或种族主义言论的内容相比，更常被删除。许多关于仇恨言论的研究指出，针对少数群体的事件被政府和执法部门视为无关紧要或恶作剧，并被随意驳回。<sup>16</sup> 具有讽刺意味的是，在一些国家，达到煽动种族灭绝或构成煽动歧视、敌意或暴力的鼓吹仇恨这一门槛的事件的制造者往往不会受到起诉或惩罚，然而，如《拉巴特行动计划》(第 11 段)所指出的：

与此同时，由于滥用模糊的国内立法、判例和政策，少数群体成员事实上受到迫害，对其他人产生了寒蝉效应。这种(1) 不起诉“实际的”煽动案件与(2) 借国内煽动法之名迫害少数群体之间的二分法似乎很普遍。世界各国的反煽动法可以说是各不相同，有时过于狭隘或模糊。……虽然有几个国家采取了……政策，但这些政策大多过于笼统，没有系统地采取后续行动，缺乏重点，并且没有进行适当的影响评估。

72. 有关变化的一个最近的例子表明，人们认识到针对少数群体的仇恨言论的规模，并认识到需要采取一种办法，反映少数群体特别易受仇恨言论影响的情况以及他们遭受的更大伤害。这个例子就是，脸书表示其对仇恨言论的广泛定义需要考虑到那些被特别针对和伤害的人。脸书在 2020 年改变算法，以优先处理针对非洲人后裔、穆斯林和犹太人等少数群体的仇恨言论。

73. 其他许多例子说明了正在开发的算法和使用人工智能的危险，在缺乏规则和适当考虑到少数群体的脆弱性及其成为社交媒体上攻击目标这一点的人权影响评估的情况下，这往往会助长和加剧少数群体遭受的仇恨和伤害。一个由人工智能驱动的名为 Lee Luda 的韩国聊天机器人拥有近 100 万用户，2021 年 1 月，该机器人在推出几周后就因为发表针对少数民族的仇恨言论而被撤下。<sup>17</sup> 这种情况不是第一次发生了。另一款人工智能机器人，即微软的 Tay 在 2016 年引发了争议，并在发布仅 16 小时后就被关闭，因为它受到操纵，对少数群体使用了仇视伊斯兰和白人至上的诋毁性语言。互动平台中使用的人工智能有着同样的缺陷：它们的算法使机器人可以从过去的聊天记录或与当前用户的互动中学习。鉴于社

<sup>14</sup> Zachary Laub, “Hate speech on social media: global comparisons”, Council on Foreign Relations, 11 April 2019.

<sup>15</sup> 欧盟委员会在对《欧洲联盟打击网上非法仇恨言论行为守则》执行情况的第四轮监测中强调，该守则“补充了打击种族主义和仇外的(国家)立法，要求切实起诉线上或线下非法仇恨言论的作者。”见 European Commission, “How the Code of Conduct helped countering illegal hate speech online”, February 2019. 可查阅 [https://ec.europa.eu/info/sites/info/files/hatespeech\\_infographic3\\_web.pdf](https://ec.europa.eu/info/sites/info/files/hatespeech_infographic3_web.pdf).

<sup>16</sup> Mari J. Matsuda, “Public response to racist speech: considering the victim’s story”, *Michigan Law Review*, vol. 87, No. 8 (August 1989).

<sup>17</sup> Justin McCurry, “South Korean AI chatbot pulled from Facebook after hate speech towards minorities”, *Guardian*, 14 January 2021.

交媒体上针对少数群体的仇恨言论大规模流行，以及种族主义和不容忍用户对人工智能机器人的“教导”或操纵，这些机器人很容易被引向各种形式的不容忍和歧视，它们自己也开始发布针对少数群体的仇恨言论。

74. 在保护少数群体的人权方面，还有其他一些令人严重关切的人工智能问题，这些问题不在本报告范围之内。其中一个问题是警察和安全部队在人工智能和数字技术中对面部识别的使用，及其如何干预禁止歧视的规定以及行动、言论和结社自由，特别是少数群体和土著人民等特定群体的权利。<sup>18</sup> 另一个问题是警察部队通过面部识别技术锁定特定的少数群体，这种技术可以进行种族定性，并专门识别少数群体成员。<sup>19</sup> 这些问题和其他问题可在今后的专题报告中讨论。

75. 总的来说，感觉对于仇恨、错误信息和虚假信息的爆发，以及它们日益造成的伤害、激进化、两极分化和危险，必须以某种方式加以解决，从而提出更多的建议、提案，加大监管压力和监管措施。<sup>20</sup>

76. 德国等国已经通过了关于网上仇恨言论的法律，要求社交媒体平台在严格的时限内删除仇恨言论，否则将面临巨额罚款。<sup>21</sup> 但立法的某些方面引起了对言论自由的严重关切，包括根据模糊和不明确的标准如“侮辱”或“诽谤”来禁止传播信息。<sup>22</sup>

77. 2016 年，欧盟出台了一项关于仇恨言论的自愿行为守则。<sup>23</sup> 在编写本报告时，欧盟还在起草一项数字服务法案，<sup>24</sup> 该法案可能会规定法律义务和责任，认识到尽管社交媒体和其他在线平台已经对非法和有害内容进行了监管，但透明度、问责度或(许多情况下的)有效性仍然很低。到目前为止，脸书、微软、推特、YouTube、Instagram、Snapchat、DailyMotion、jeuxVideo.com 和 TikTok 都加入了该行为准则。2020 年对该守则执行情况进行的第五轮(也是最新一轮)监测显示，这些公司目前平均在 24 小时内对 90% 的标记内容进行评估，并将 71% 被

<sup>18</sup> 研究证实，许多面部识别系统中都存在对少数群体的偏见。一项研究表明，根据特定的算法和搜索类型，美国的亚裔和非裔少数群体被错误识别的可能性是白人男性的 100 倍(Drew Harwell, “Federal study confirms racial bias of many facial-recognition systems, casts doubt on their expanding use”, *Washington Post*, 20 December 2019)。人工智能中的这种偏见，以及由此产生的假肯定和假否定信息，对少数群体造成明显的严重后果。

<sup>19</sup> 中国的几家科技公司申请了各种设备和系统的专利，其中包括维吾尔族和其他少数民族的识别。见 IPVM, “Patenting Uyghur tracking – Huawei, Megvii, more”, 12 January 2021。

<sup>20</sup> 例如见消除种族歧视委员会，关于打击种族主义仇恨言论的第 35 号一般性建议(2013 年)；欧洲委员会，部长委员会向会员国提出的关于“仇恨言论”的 R(97)20 号建议，1997 年 10 月 30 日；欧洲反对种族主义和不容忍委员会，关于打击仇恨言论的第 15 号一般性政策建议，2015 年 12 月 8 日；以及欧安组织少数民族事务高级专员，“关于数字时代少数民族与媒体的塔林准则及解释性说明”(海牙，2019 年)。

<sup>21</sup> 见《网络执行法》，2017 年 9 月 1 日。可查阅 [www.gesetze-im-internet.de/netzdg/BJNR335210017.html](http://www.gesetze-im-internet.de/netzdg/BJNR335210017.html) (仅有德语)。

<sup>22</sup> 2017 年 6 月 1 日促进和保护意见和表达自由权特别报告员致德国常驻联合国日内瓦办事处和日内瓦其他国际组织代表的信。可查阅 [www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf](http://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf)。

<sup>23</sup> 可查阅 [https://ec.europa.eu/newsroom/just/document.cfm?doc\\_id=42985](https://ec.europa.eu/newsroom/just/document.cfm?doc_id=42985)。该守则非常简短，且十分笼统，没有界定仇恨言论，也没有提到少数群体。

<sup>24</sup> 欧盟委员会，欧洲议会和理事会关于数字服务单一市场并修正第 2000/31/EC 号指令的条例提案(《数字服务法》)，2020 年 12 月 15 日。可查阅 <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0825&from=en>。

视为非法仇恨言论的内容予以删除。<sup>25</sup> 但同样，仍然令人关切的是，目前的程序没有如实记录针对少数群体的许多形式的仇恨言论。

78. 虽然可以说，社交媒体上约三分之二的仇恨言论是针对少数群体的，但大多数社交媒体平台的社区标准或内容审查准则很少直接关注少数群体，甚至没有具体提到他们。例如，TikTok 的社区准则提到诸如“描绘或助长偷猎或非法交易野生动物的内容”等问题，并将仇恨言论或行为界定为基于种族、族裔和宗教等属性，“攻击、威胁个人或群体、煽动对他们的暴力或以其他方式将其非人化的内容”。<sup>26</sup> 但该准则中根本没有提及“少数群体”一词。虽然提到“野生动物”是合理的，但不提及少数群体令人遗憾且与直觉相反，尽管事实上少数群体是社交媒体上仇恨言论的最大目标和受害者。

79. 以上并不是说没有积极的事态发展。大多数社交网络只在最近几年才临时制定了内容审查规则。脸书在 2009 年首次制定了全面的内部审查准则，但直到 2018 年才发布公开准则。越来越多的迹象表明，大多数社交媒体平台的所有者正在努力提高透明度，加强与民间社会组织在内容审查方面的合作，其中可能包括一些领域的人权影响评估。这些努力值得赞扬。<sup>27</sup> 然而，最明显的不足仍然是，在对作为仇恨言论主要目标的少数群体予以更多关注的问题上相对沉默，保护他们的政策措施很少，或根本没有。一个值得注意的例外是来自脸书的上述信号，即它正在改变算法，以优先处理针对少数群体的仇恨言论。然而，目前还不知道这一变化是全球性的，还是范围比较有限。

## F. 少数群体与民间社会组织

80. 民间社会的倡议、存在和参与在通过社交媒体传播信息的现代世界中至关重要，特别是因为仇恨言论主要针对的是少数群体。事实上，联合国也澄清，必须把重点放在那些由于根深蒂固的或历史性的污名化、歧视、长期冲突以及被排除在社会政治、经济和社会生活之外而处于弱势的群体。<sup>28</sup> 这些群体大多是少数群体，因此社交媒体平台必须更多地关注少数群体，因为他们面临的风险最大。

81. 由于这个原因，人们越来越认识到，有必要引入或加强少数群体在社交媒体中的结构化代表。因此，平台所有者需要与少数群体接触，并确保他们参与其咨询、协商和其他机构，特别是与人权影响评估和仇恨言论问题有关的机构。直接和透明地纳入少数群体更有可能使社交媒体平台的所有者更好地理解 and 考虑社交媒体中仇恨、种族主义和不容忍的根本原因，并更有效地管理和运营其平台和产品，避免利用仇恨获利。

<sup>25</sup> 见 [https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/eu-code-conduct-counteracting-illegal-hate-speech-online\\_en](https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/eu-code-conduct-counteracting-illegal-hate-speech-online_en)。

<sup>26</sup> 见 [www.tiktok.com/community-guidelines?lang=en#38](https://www.tiktok.com/community-guidelines?lang=en#38)。

<sup>27</sup> 例如见人权高专办，“Public comment by UN Special Rapporteur on minority issues relating to cases on hate speech and minorities”，23 December 2020。

<sup>28</sup> 《联合国消除仇恨言论战略和行动计划：联合国驻地机构详细执行指南》(2020 年)，第 11 页。

## 五. 建议

82. 特别报告员请各国、区域组织和非政府组织回顾表达自由特别报告员在其 2019 年报告(A/74/486)中提出的建议, 不要对社交媒体平台施加任何不符合国际人权法所规定的标准的限制。

83. 少数群体问题特别报告员提醒各国, 对社交媒体表达形式的限制只有在经法律规定并为尊重他人的权利或名誉或为保护国家安全、公共秩序或公共卫生或道德所必要的情况下才是允许的。这些限制还必须是明确和加以严格界定的, 应针对紧迫的社会需要, 是对表达自由有效和约束最小的干预措施, 不能过于宽泛或模糊, 并且具有相称性, 即对受保护利益的好处大于对言论自由的损害。

84. 特别报告员还请各国和其他有关各方参考关于在不允许的仇恨言论领域, 哪些表达形式可根据国际人权法加以限制的现有指导, 特别是《拉巴特行动计划》中的六部分门槛标准和《关于言论自由和平等的卡姆登原则》的原则 12。

85. 特别报告员提醒各国, 为了履行其国际义务, 必须通过立法禁止某些形式的仇恨言论, 包括禁止任何构成煽动歧视、敌意或暴力的鼓吹民族、族裔、种族或宗教仇恨言论, 并将在社交媒体上煽动种族灭绝定为犯罪。必须根据《拉巴特行动计划》中的六部分门槛标准以及《联合国消除仇恨言论战略和行动计划》中的详细指导, 对这些被禁止的仇恨言论形式作狭义的解释。为履行这些国际义务, 请各国研究并详细说明有效执行这些被禁止形式的仇恨言论的情况, 并考虑社交媒体平台在不予遵守时可能承担的责任。

86. 各国应审查其法律框架, 确保纳入一系列受保护的特征, 或确定国际人权法中承认的少数群体(民族或族裔、宗教和语言少数群体), 特别是因为世界各地大多数不允许的仇恨言论事件都涉及少数群体。

87. 特别报告员请国家机构和社交媒体平台建立程序和机制, 强制收集关于仇恨言论的数据, 至少收集关于煽动种族灭绝和构成煽动歧视、敌意或暴力的鼓吹行为的数据。这些数据应按照与弱势少数群体有关的仇恨(无论是民族、族裔、种族还是宗教仇恨)或基于性别和性取向等因素影响其他弱势群体的仇恨分类。

88. 特别报告员建议制定国家立法, 要求社交媒体平台通过政策和规则, 识别基于民族血统、族裔、宗教和语言以及性别和性取向等其他因素的仇恨言论, 同时尊重个人的隐私权。

89. 特别报告员鼓励社交媒体公司参照其产品(包括算法和聊天机器人等人工智能程序)对人权的影响来管理其平台上的仇恨言论。因此, 这些公司必须制定人权审查程序, 具体提到并侧重于有关国家最易受仇恨言论影响的群体, 特别是少数群体、妇女和男女同性恋、双性恋、跨性别者和间性者群体。

90. 特别报告员请各国采取果断、迅速和有效的行动, 处理和打击在线交流中针对少数群体的仇恨言论, 包括迅速和有效地调查和起诉责任人, 追究他们的责任, 并确保受害者能够有效诉诸司法和获得补救。

91. 特别报告员请各国、联合国及其实体, 特别是人权高专办, 启动一项进程, 为社交媒体平台制定打击仇恨言论的全球自愿行为守则。还请它们支持他在 2021-2022 年作为紧急事项起草打击社交媒体上针对少数群体的仇恨言论准则的努力。

92. 特别报告员强调，社交媒体的内容审查系统和社区标准以及任何监督或上诉实体都应明确承诺保护弱势和边缘化少数群体及其他群体。应特别将少数群体定为社交媒体平台的优先群体。

93. 特别报告员请社交媒体平台的所有者在其各自的运营中适用《工商企业与人权指导原则》。应将人权标准系统地纳入社交媒体平台的内容政策和决策机制，以便按照《指导原则》的规定，其所有者“遵守所有适用法和尊重国际公认的人权，无论其在何处经营”，并“在其经营地，将(其可能负有责任的)造成或加剧严重侵犯人权的行为视为是否守法的问题”。

94. 特别报告员促请各国建立符合国际标准的独立和权威的专门机构，开展打击仇恨言论的工作，并建立民间社会组织在线举报仇恨言论的便利机制。

95. 特别报告员建议各国向执法机构和司法机构提供关于少数群体权利，特别是关于网上针对少数群体的仇恨言论的充分和专门的培训。

96. 特别报告员鼓励各国和社交媒体平台全面处理对犹太人和穆斯林的歪曲和系统性偏见问题，因为有证据表明，反犹太主义和仇视伊斯兰现象是少数群体权利面临的紧迫挑战。

97. 特别报告员促请各国采取关于少数群体权利的人权教育举措，包括将其纳入学校课程；促进多样性和多元化；通过传播能够激发对话的积极、替代性和包容性的信息，打击歧视、陈规定型观念、仇外、种族主义和仇恨言论。

98. 为了改进处理仇恨内容的机制和内容审查政策，并确保纳入社交媒体上仇恨言论主要目标群体的关切，特别报告员促请在咨询委员会和其他相关委员会中要有作为最大目标群体和最弱势的少数群体的代表。

99. 特别报告员邀请各国、国家人权机构和民间社会代表利用联合国人权机制，包括人权理事会特别程序、普遍定期审议和人权条约机构，以及区域人权机制和其他适当论坛，打击网上针对少数群体的仇恨言论。

100. 特别报告员鼓励所有利益攸关方研究注重保护并尊重不同社区的创新、教育和预防战略，打击仇恨言论。

## Annex I

1. On 9 July 2020, the Special Rapporteur took part in a webinar entitled ‘Debating Challenges on Minority Protection’, organised by the Tom Lantos Institute, in cooperation with the Human Rights Consortium at the University of London and other organisations. The webinar focussed on the interrelationship between conflicts, minority rights and the promotion of inclusiveness and stability.
2. On 21 July 2020, he participated in a webinar on “Keeping the Faith in Times of Hate: The Practical Utility of Human Rights” co-hosted by Religions for Peace, UN Human Rights and the Convention on the Elimination of all Forms of Discrimination Against Women, raising the implications of COVID-19 for minorities who simultaneously face an increase in hate speech in social media. On 21 July 2020, the Special Rapporteur also addressed a webinar organised by the OSCE Transnational Threats Department’s Strategic Police Matters Unit in collaboration with the OSCE High Commissioner on National Minorities entitled “Towards resilient and cohesive societies: community policing to prevent and combat human trafficking” where he spoke on the particular vulnerability to trafficking of minority women and girls.
3. On 31 July 2020, the UN Special Rapporteur gave the keynote addresses at the plenary session of the Unrepresented Peoples Organisation (UNPO) General Assembly on the impact of COVID-19 on minorities and indigenous peoples worldwide.
4. On 11 August 2020, the Special Rapporteur in the webinar celebrating International Youth Day organised by the National Youth Council of Moldova, in cooperation with the Youth Platform of Interethnic Solidarity of Moldova and former UN Minority Fellows Programme.
5. On 31 August 2020, he gave an online lecture to students in the Human Rights programme at Vesalius College in Brussels on his mandate on minority issues and the protection of minority rights in international law.
6. On 10 September 2020, the Special Rapporteur spoke on ‘Inequalities, disenfranchisement and frailty: fault lines in global human rights protection and challenges faced by minorities’ during a webinar session of the online international Conference on Global Justice, Rule of Law and Human Rights on Pandemic Perspectives, organised by the Research Group Culture, Law and Society of the State University of Maranhão, Brazil.
7. On 14 September 2020, the Special Rapporteur participated in a webinar on “Situation of Human Rights Violation of Dalits in Nepal and Role of the International Community”.
8. On 21 and 22 September 2020, the Special Rapporteur participated in his mandate’s two-day online regional forum for Europe on hate speech, social media and minorities. More than 200 experts, state, human rights and minority representatives participated in the event during four panels looking more closely to the vastly disproportionate, and growing, targeting and scapegoating of minorities in social media for hate and incitement to violence and discrimination.
9. On 23 September 2020, Dr de Varennes gave the plenary presentation for the World Federation of the Deaf webinar and annual meeting for the International Day of Sign Languages. The Special Rapporteur addressed the language rights of users of sign languages as members of linguistic minorities.
10. On 25 September 2020, the Special Rapporteur gave the keynote speech for the 2020 Symposium on French and the Law organised by Massey College at the University of Toronto, on equity and access to education and the legal system.
11. On 8 October 2020, the Special Rapporteur participated online with discussions with and a presentation to students in the Global Minority Rights Summer School 2020 organised by the Tom Lantos Institute in Budapest.
12. On 19 and 20 October 2020, the Special Rapporteur participated in his mandate’s two-day online regional forum for Asia-Pacific on hate speech, social media and minorities. Some 200 experts, state, human rights and minority representatives participated in the event during

four panels looking more closely to the vastly disproportionate, and growing, targeting and scapegoating of minorities in social media for hate and incitement to violence and discrimination.

13. On 23 October 2020, the Special Rapporteur participated in the annual meeting of the European Language Equality Network Steering Committee Plus and discussed on the pandemic and minority languages, as well as the conclusions and recommendations of his 2019 mission report to Spain.

14. On 26 October 2020, the Special Rapporteur made introductory remarks and participated in the online conference “Minority Protection and Ethnic Groups’ Rights in Central and Central Eastern Europe” organised by the Federal Union of European Nationalities and the Cultural Foundation of German Expellees for Science and Research (Kulturstiftung der deutschen Vertriebenen für Wissenschaft und Forschung).

15. On 12 November 2020, the Special Rapporteur participated in the online consultation on anti-Muslim hatred online and Islamophobia organised by his colleague, the Special Rapporteur on freedom of religion or beliefs.

16. On 18 November 2020, the Special Rapporteur participated in Mexico’s National Transparency Week in an international conference on “Public Health and Transparency, the Importance of Public Information to Face Sanitary Crises” organised by the Mexican National Institute for Transparency, Access to Information and Personal Data Protection.

17. On 19 and 20 November 2020, the Special Rapporteur guided the 13th Forum on Minority Issues, held virtually this year exceptionally because of the pandemic on the theme of Hate Speech, Social Media and Minorities. Some 300 participants discussed during four panels the most significant areas of concern in relation to hate speech, social media and minorities, including on the causes, scale and impact of hate speech targeting minorities in social media, the international legal and institutional framework, the regulation of online hate speech, the role and responsibility of intergovernmental organisations, States, internet companies and social media platforms; and positive initiatives to address online hate speech.

18. On 8 December 2020, the Special Rapporteur made introductory remarks at the presentation of the results of the Justice and Social Equality Programme of the OHCHR Regional Office for Central Asia in collaboration with the Government of the Republic of Kyrgyzstan as well as the support of the European Union and the United Nations Peace Fund. He also on the same day participated online in an expert panel on discrimination in the justice system during the regional dialogue for the Latin America and Caribbean region on Promoting and Protecting the Rights of People of African Descent and their Communities in the Context of COVID-19 and Implementing the International Decade for People of African Descent, the Montevideo Consensus on Population and Development and the 2030 Agenda.

19. On 15 December 2020, the Special Rapporteur made the closing remarks and was presented the recommendations for the advocacy training workshop held online and in Baghdad on the preservation and protection of minority languages in Iraq organised by the JSSOR Youth Organization with the support of the United Nations Assistance Mission in Iraq and the OHCHR.

20. On 23 December 2020, the Special Rapporteur participated in the webinar entitled “Covid-19: I am minority do I matter?”, organised by the Swedish Institute Network of Future Global Leaders Malmö.

21. The Special Rapporteur also intervened in various capacities, including in an intervention to the Facebook Oversight Board in relation to its first 6 cases, three of which concerned hate speech and had some relation to minority issues. He drew special attention to the need to incorporate a minority rights approach and retain focus on the overarching aim of protecting vulnerable and marginalised groups which are in most cases minorities.

22. The Special Rapporteur also conducted numerous media interviews concerning his mandate and minority rights issues globally.

## Annex II

### Hate Speech, Social Media and Minorities: sample questionnaire and list of contributors

#### A. Sample questionnaire

##### Call for Submissions by 30 November 2020

In accordance with his mandate pursuant to Human Rights Council resolution 34/6, the Special Rapporteur on minority issues, Dr. Fernand de Varennes, will present a thematic report at the 46th Session of the UN Human Rights Council, which will provide a detailed analysis and highlight the issue of “Hate speech, social media and minorities”. Previously, in his first report to the UN Human Rights Council in March 2018, the Special Rapporteur identified the issue of ‘hate speech’, xenophobic rhetoric and incitement to hatred against minorities as one of the thematic priorities of his mandate. In this regard, his report to the Human Rights Council will build and further expand upon the work carried out by his predecessors on this particular area (see report [A/HRC/28/64](#)), as well as the contributions by other human rights mechanisms and the work done by civil society in this field.

Inspired by the [Rabat Plan of Action](#), the [Beirut Declaration](#) and its [18 commitments](#) on “Faith for Rights” and the 2019 launch by the Secretary-General of the [UN Strategy and Plan of Action on Hate Speech](#), the present report will refer to the global discussions on ‘hate speech’ and incitement to discrimination, hostility or violence through social media and will highlight the significance of the promotion and protection of the human rights of minorities, whilst respecting the right of freedom of expression.

##### Context

Across the world, discriminatory practices and hateful messages serve to stigmatize and vilify those perceived to be the “other”, the “foreigner” or the “one who does not belong”. Most of those targeted by these messages are minorities – overwhelmingly so. Disaggregated data in some countries suggest that perhaps three-quarters or more of hate speech, for example, target ethnic, religious or linguistic minorities. ‘An ethnic, religious or linguistic minority is any group of persons which constitutes less than half of the population in the entire territory of a State whose members share common characteristics of culture, religion or language, or a combination of any of these. A person can freely belong to an ethnic, religious or linguistic minority without any requirement of citizenship, residence, official recognition or any other status’ according to the concept adopted in the UN Special Rapporteur’s [2019 annual report](#) to the UN General Assembly.

Expressions of hate and discrimination are increasingly present in certain political agendas and discourses and can contribute to a climate of fear amongst individuals and communities. They can at times also create a climate of rejection, exclusion and even scapegoating, threatening societal values and undermining the respect of human dignity and the protection of human rights. A state’s international obligations are thus central in any effort to address the direct challenges to and even rejection of the fundamental values at the core of the United Nations’ human rights’ architecture.

In the digital age, internet companies and social media platforms present opportunities for the online expression, association, participation and empowerment for members of ethnic, religious and linguistic minorities around the world, as well as for advocacy for the protection and promotion of their human rights. However, it has become increasingly obvious that there have also been numerous and flagrant examples of social media platforms being exploited to spread hatred, unfounded and stigmatizing rumours, fostering a climate of insecurity and animosity, and in the most extreme cases, leading to violence against members of minorities. Thus, while platforms offer members of minorities expanding opportunities for generating, framing and sharing information and reporting on issues that affect these individuals and groups, the same platforms can also host online ‘hate speech’, including incitement to discrimination, hostility or violence against those perceived as “others”, mainly members of

minorities. In extreme cases, they can even be used to propagate calls for genocide against minorities. Thus, unregulated online expressions of hate can increase the chances of human rights violations taking place offline against some of the most marginalised segments of society.

At the same time, some internet companies are responding to pressure to remove online content that is deemed harmful. This can occur through the imposition of intermediary liability, the application of filters, as well as automated tools. Some companies have also set their own content standards in this regard. Frequently, however, these measures have the purpose or effect of unlawfully, illegitimately and unnecessarily restricting the exercise of human rights – especially freedom of opinion and expression – online and operate in the absence of any meaningful oversight mechanisms. Furthermore, such measures may have a disproportionate impact upon or even target individuals from marginalised groups, particularly persons belonging to minorities (as well as political opponents, critics, and human rights defenders), while limiting the possibilities for information-sharing, awareness-raising and advocacy for civil society organisations, human rights defenders and representatives of persons belonging to minorities. Moreover, companies' and social media platforms' online content moderation policies can lack transparency and any precise and meaningful basis in international human rights standards, raising the possibilities that the decisions made by these actors undermine the human rights of individuals, including those belonging to minority groups.

Thus, the fundamental, two-pronged concern first raised by the Rabat Plan of Action in October 2012 ([A/HRC/22/17/Add.4](#)) – that members of minorities are effectively persecuted through the abuse of vague domestic legislation, jurisprudence and policies on 'hate speech', whereas actual incidents which meet the threshold of incitement to discrimination, hostility or violence under international human rights law are not addressed – has become an even more pressing issue at the beginning of the new decade, one that requires effective and urgent responses from States, social media platform owners, and other stakeholders which are based on international human rights law and standards.

Freedom of expression and the essential communication tools and services provided by or dependent on an open and accessible internet must be protected, as minorities and others who are marginalised or vulnerable must be protected from hate speech, incitement to discrimination, hostility or violence, and even calls to genocide.

### **Call for submissions**

In accordance with the established practice of thematic mandate-holders, the Special Rapporteur welcomes inputs by States, UN agencies, regional and international organizations, national human rights institutions, civil society, scholars and research institutions, private companies including those from the ICT sector, and others who may wish to submit for this purpose. Such submissions may include, for instance, recommendations, evidence and case studies. The following questions are intended to guide submissions:

1. Please provide annual disaggregated data since 2017 if possible on hate speech in social media, and in particular hate speech targeting minorities (national or ethnic, religious and linguistic minorities). Please additionally indicate whether there are future plans to include specifically disaggregated data on hate speech targeting minorities, considering that in most countries, the victims of hate speech on social media are usually members of minorities.
2. Please identify the mechanisms and processes in place to remove, penalise or address hate speech in social media targeting minorities. Please also specify and include any studies or reports assessing their implementation and effectiveness.
3. Please provide (legal and non-legal) examples of good practices of appropriate responses developed by States, internet companies, civil society and other relevant stakeholders to address online 'hate speech', including incitement to discrimination, hostility or violence, against persons belonging to minorities. Please include assessments, if any, on the effectiveness of these examples.

4. Please identify legal, institutional and policy challenges to address online ‘hate speech’, particularly incitement to discrimination, hostility or violence, on social media against minorities, as well as how specifically is it being addressed so as not to infringe freedom of expression.

5. Please identify good practices to strengthen the participation of persons belonging to minorities and their representatives in the development of laws and policies dealing with hate speech and social media, including their representation in advisory or other similar bodies, as well as their participation in national and other forums on online ‘hate speech’ and the promotion of diversity.

6. Please provide examples of trainings or other initiatives to strengthen partnerships among various stakeholders and to build their capacity to address ‘hate speech’ against minorities specifically, especially incitement to discrimination, hostility or violence on social media platforms.

## **B. List of contributors**

### **States:**

Azerbaijan, Croatia, Colombia, Cyprus, Greece, Guatemala, Hungary, India, Italy, Kenya, Kyrgyzstan, Latvia, Lebanon, Mexico, Poland, Portugal, Qatar, Singapore, Slovenia, Switzerland and the European Union

### **International and Regional Organizations:**

UNESCO

### **National Human Rights Institutions:**

Australian Human Rights Commission

Comisión de Derechos Humanos de la Ciudad de México

Commission nationale des droits de l’homme Togo

Czech Public Defender of Rights

Defensoría del Pueblo de la Nación de Argentina

Ombudsperson of Azerbaijan

Office of the Croatian Ombudsman

Slovak National Centre for Human Rights

### **Civil society organizations and Academia:**

Anhad

Baha’i International Community

Citizens against Hate

National Christian Evangelical Alliance of Sri Lanka (NCEASL)

Rights for Peace

RyangOk Ku and other lawyers in Japan

University of Minnesota

---