

28 April 2023
Russian
Original: English

Генеральная Ассамблея
Семьдесят седьмая сессия
Пункт 126 а) повестки дня
Укрепление системы Организации Объединенных Наций:
укрепление системы Организации Объединенных Наций

«Наша общая повестка дня»

Концептуальная записка 8. Информационная добросовестность на цифровых платформах

Резюме

Проблемы, с которыми мы сталкиваемся, могут быть решены только путем укрепления международного сотрудничества. Саммит будущего, который состоится в 2024 году, — это возможность договориться о многосторонних решениях ради лучшего завтра на основе укрепления глобального управления на благо нынешнего и грядущих поколений (резолюция 76/307 Генеральной Ассамблеи). Мне как Генеральному секретарю было предложено внести вклад в подготовку к Саммиту путем представления ориентированных на конкретные действия рекомендаций на основе предложений, содержащихся в моем докладе под названием «Наша общая повестка дня» (A/75/982), который, в свою очередь, стал ответом на декларацию о праздновании семьдесят пятой годовщины Организации Объединенных Наций (резолюция 75/1 Генеральной Ассамблеи). Настоящая концептуальная записка является одним из материалов, представляемых в ответ на это предложение. В ней нашли свое развитие идеи, впервые предложенные в «Нашей общей повестке дня», с учетом последующих указаний государств-членов и результатов продолжавшихся более года межправительственных и многосторонних консультаций, а также опирающиеся на цели и принципы Устава Организации Объединенных Наций, Всеобщей декларации прав человека и других международных документов.

Настоящая концептуальная записка посвящена тому, как угрозы для информационной добросовестности сказываются на прогрессе в решении глобальных, национальных и местных проблем. В «Нашей общей повестке дня» я призвал к эмпирически выверенному консенсусу в отношении фактов, научных выкладок и знаний. Поэтому в настоящей записке излагаются потенциальные принципы для кодекса поведения, который поможет сориентировать государства-члены, цифровые платформы и другие заинтересованные стороны в их



усилиях, призванных сделать цифровое пространство инклюзивнее и безопаснее для всех, энергично защищая при этом право на свободу мнений и их свободное выражение и право на доступ к информации. В контексте подготовки к Саммиту будущего разрабатывается кодекс поведения для обеспечения информационной добросовестности на цифровых платформах. Надеюсь, он станет золотым стандартом в качестве ориентира для действий по упрочению информационной добросовестности.

Цифровые платформы — это солиднейшие инструменты, вызвавшие повсеместную трансформацию социальных, культурных и политических взаимодействий. Они соединяют по всему миру граждан земного шара, которых волнуют значимые проблемы. В условиях, когда мы стремимся обеспечить мир, достоинство и равенство на здоровой планете, эти платформы помогают Организации Объединенных Наций напрямую информировать и заинтересовывать людей. Они давали людям надежду в периоды кризиса и борьбы, усиливали голоса, которые были ранее неслышны, и вдыхали жизнь в глобальные движения.

Но при этом те же самые платформы обнажили более темную сторону цифровой экосистемы. Они создали возможность для быстрого расползания лжи и ненависти, что оборачивается реальным вредом в глобальном масштабе. Оптимизм по поводу способности социальных сетей соединять и увлекать людей оказался приглушен тем, что с периферии цифрового пространства в мейнстрим хлынули неверная информация, дезинформация и язык ненависти. Возникшую опасность невозможно переоценить. Проникающие через социальные сети дезинформация и ненавистническая риторика могут приводить к насилию и гибели^a. Способность распространять крупномасштабную дезинформацию, чтобы торпедировать научно установленные факты, представляет для человечества экзистенциальный риск (A/75/982, п. 26) и подвергает опасности демократические институты и основные права человека. Эти риски еще больше усилились из-за стремительного развития таких технологий, как генеративный искусственный интеллект. Организация Объединенных Наций повсеместно отслеживает, как неверная информация, дезинформация и язык ненависти могут угрожать продвижению к целям в области устойчивого развития. Стало ясно, что жить как ни в чем не бывало не получится.

^a A/HRC/42/50; A/77/287; A/HRC/51/53; United Nations, “Statement by Alice Wairimu Nderitu, Special Adviser on the Prevention of Genocide, condemning the recent escalation of fighting in Ethiopia”, press release, 19 October 2022; Office of the United Nations High Commissioner for Human Rights (OHCHR), “Myanmar: Social media companies must stand up to junta’s online terror campaign say UN experts”, press release, 13 March 2023; OHCHR, “Freedom of speech is not freedom to spread racial hatred on social media: UN experts”, statement, 6 January 2023; Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, “#JournalistsToo: women journalists speak out”, 24 November 2021; and OHCHR, “Sri Lanka: Experts dismayed by regressive steps, call for renewed UN scrutiny and efforts to ensure accountability”, press release, 5 February 2021.

I. Что такое информационная добросовестность?

1. Под информационной добросовестностью понимаются точность, последовательность и достоверность информации. Ей угрожает распространение дезинформации, неверной информации и языка ненависти. Общепринятых дефиниций этих терминов не имеется, но структурами Организации Объединенных Наций составлены рабочие определения.
2. Специальный докладчик по вопросу о поощрении и защите права на свободу мнений и их свободное выражение называет дезинформацией «ложную информацию, намеренно распространяемую с целью причинения серьезного социального вреда»¹. Организация Объединенных Наций по вопросам образования, науки и культуры (ЮНЕСКО) описывает дезинформацию как ложный или вводящий в заблуждение контент, который может причинить конкретный вред, независимо от мотивации, осознанности или поведения².
3. Для целей настоящей концептуальной записки различие между неверной информацией и дезинформацией заключается в наличии намерения.
4. Дезинформация — это информация, которая не только неправдива, но и распространяется с намерением обмануть и с целью причинить вред. Дезинформация может распространяться государственными или негосударственными субъектами во многих контекстах, включая контекст вооруженного конфликта, и затрагивать все области развития — и мир с безопасностью, и права человека, и охрану здоровья людей, и оказание гуманитарной помощи, и противодействие изменению климата.
5. Под неверной информацией понимается непреднамеренное распространение неточных сведений, которыми одни люди честно делятся с другими, не осознавая при этом, что они передают неправду. Неверная информация может происходить из дезинформации — по мере того как преднамеренная ложь и вводящий в заблуждение нарратив превращаются со временем в оружие, вливаются в публичный дискурс и невольно тиражируются³. На практике бывает трудно провести различие между неверной информацией и дезинформацией.
6. Язык ненависти — согласно рабочему определению, содержащемуся в Стратегии и плане действий Организации Объединенных Наций по борьбе с языком ненависти, — представляет собой «любого рода коммуникацию в устной, письменной или поведенческой форме, которая носит враждебный характер или использует уничижительные или дискриминационные формулировки в отношении какого-либо лица или группы лиц по признаку того, кто они есть, или, иными словами, по признаку их религии, этнического происхождения, национальности, расы, цвета кожи, социального происхождения, пола и других факторов идентичности»⁴.
7. Неверная информация, дезинформация и язык ненависти — это смежные, но разные явления, характеризующиеся отдельными совпадениями и

¹ A/HRC/47/25, п. 15.

² Kalina Bontcheva and Julie Posetti, eds., *Balancing Act: Countering Digital Disinformation While Respecting Freedom of Expression – Broadband Commission Research Report on “Freedom of Expression and Addressing Disinformation on the Internet”* (Geneva, International Telecommunication Union (ITU); Paris, United Nations Educational, Scientific and Cultural Organization (UNESCO), 2020).

³ См. Организация Объединенных Наций, «Борьба с дезинформацией» (URL: www.un.org/ru/countering-disinformation), и A/77/287.

⁴ URL: https://www.un.org/en/genocideprevention/documents/advising-and-mobilizing/Action_plan_on_hate_speech_RU.pdf.

различиями в том, как их можно идентифицировать, смягчить и преодолевать. Вся эта тройка загрязняет информационную экосистему и угрожает человеческому прогрессу⁵.

8. Угрозы для информационной добросовестности не новы. Ложь и ненависть издавна распространяются ради политической или финансовой выгоды. Однако в цифровую эру эти операции могут совершаться с немыслимым ранее размахом. Сейчас мощные коммуникационные инструменты могут мгновенно распространять контент по всему земному шару, создавая при этом настолько разрастающуюся проблему, что сами онлайн-платформы не в состоянии подчас полностью осознать ее масштабы. Отсутствие у этих терминов определений, согласованных на государственном уровне, не должно приводить к инертности. Мы должны сделать всё возможное, чтобы обуздать причиняемый ими вред.

II. Информационная добросовестность и цифровые платформы

9. Цифровые платформы должны быть обязательными действующими лицами движения за информационную добросовестность. Источниками неверной информации и дезинформации могут являться и отдельные традиционные средства массовой информации, однако скорость, объем и вирусность их распространения по цифровым каналам требуют экстренного и отлаженного реагирования. Для целей настоящей записки термином «цифровая платформа» описывается цифровой сервис, облегчающий взаимодействие между двумя или более пользователями, и охватывается широкий спектр услуг — от социальных сетей и поисковых систем до приложений для обмена сообщениями. Как правило, эти платформы собирают данные о своих пользователях и их взаимодействиях⁶.

10. Неверная информация и дезинформация создаются широким кругом субъектов, которыми движут различные мотивы и которым удается в целом сохранять анонимность. Государственными и негосударственными субъектами проводились скоординированные дезинформационные кампании, в которых для распространения вредоносных нарративов, влекущих серьезные последствия, использовались изъяны в цифровых системах.

11. Многие государства развернули инициативы по регулированию цифровых платформ, и за последние четыре года было принято или рассмотрено как минимум 70 соответствующих законов⁷. Суть законодательных подходов заключается обычно в появлении узкого набора правовых средств, предусматривающих определение вредоносного контента и его удаление. Сосредоточившись на удалении вредоносного контента, некоторые государства ввели ущербное и чересчур широкое законодательство, которое стало фактически заглушать то, что международное право признает подпадающим под свободу слова. Имели место и другие меры (например, повальное отключение Интернета и запрещение

⁵ В настоящее время Организация Объединенных Наций проводит исследование для изучения взаимосвязей и соотношений между неверной информацией, дезинформацией и языком ненависти, а также того, в чем эти смежные, но разные явления сходятся и расходятся как на концептуальном, так и на оперативном уровне.

⁶ Европейская комиссия определяет онлайн-платформы в следующем документе: “Shaping Europe’s digital future: online platforms”, 7 June 2022. URL: <https://digital-strategy.ec.europa.eu/en/policies/online-platforms>.

⁷ См. Office of the United Nations High Commissioner for Human Rights (OHCHR), “Moderating online content: fighting harm or silencing dissent?”, 23 July 2021.

платформ), которые, возможно, лишены законных оснований и посягают на права человека.

12. Многие государства и политические фигуры использовали якобы заботу об информационной добросовестности как предлог для ограничения доступа к информации, для дискредитации и ограничения репортерства и для притеснения журналистов и оппонентов⁸. Государственные субъекты также понуждали платформы к выполнению их распоряжений, выдаваемых за борьбу с неверной информацией и дезинформацией⁹. Экспертами по свободе выражения мнений подчеркивалось, что государственные субъекты несут особую ответственность в этом контексте и что им «не следует предоставлять, спонсировать, поощрять или поддерживать ложную информацию» (A/77/287, п. 45).

13. Для уменьшения рисков, сопряженных с регулированием выражения мнений, необходим тщательно продуманный подход, отвечающий требованиям законности, необходимости и соразмерности в соответствии с правом прав человека, — даже при наличии обоснованной цели, связанной с государственными интересами (там же, п. 42).

14. Дезинформация — это еще и большой бизнес. К ключевым источникам ложного и вводящего в заблуждение контента относятся как «теневые», так и мейнстримные пиар-фирмы, подряжаемые государствами, политическими фигурами и частным сектором¹⁰. Одним из тактических приемов становилась публикация контента на подделанных, клонированных версиях новостных сайтов с тем, чтобы статьи выглядели происходящими как бы из легитимных источников¹¹. Этот темный бизнес крайне сложно отследить и исследовать, так что истинный масштаб проблемы неизвестен. Распространением ложных утверждений занимаются и отдельно взятые личности, которые приторговывают продуктами или услугами ради наживы, зачастую нацеливаясь на уязвимые группы в кризисное или беспокойное время.

15. Один из доминирующих подходов в текущих бизнес-моделях большинства цифровых платформ завязан на «экономике внимания». Алгоритмы проектируются так, чтобы выпячивать контент, удерживающий внимание пользователей, максимально усиливая тем самым увлеченность аудитории и рекламную выручку. Наибольшую же увлеченность зачастую вызывает недостоверный и ненавистнический контент, призванный поляризовать пользователей и вызывать сильные эмоции, и в результате получается так, что алгоритмы вознаграждают и амплифицируют неверную информацию, дезинформацию и язык ненависти¹².

16. Сталкиваясь со снижением рекламной выручки, цифровые платформы ищут альтернативные каналы получения прибыли, не связанные с экономикой внимания. Например, появляются платные планы верификации (позволяющие купить для аккаунта отличительный знак, ранее использовавшийся для обозначения аутентичности), которые заставили серьезно опасаться за информационную добросовестность ввиду потенциала для злоупотреблений со стороны дезинформаторов¹³.

⁸ См. Организация Объединенных Наций, «Борьба с дезинформацией», и A/77/287.

⁹ A/HRC/47/25.

¹⁰ Stephanie Kirchgaessner and others, “Revealed: the hacking and disinformation team meddling in elections”, *The Guardian*, 14 February 2023.

¹¹ Alexandre Alaphilippe and others, “Doppelganger – media clones serving Russian propaganda”, EU DisinfoLab, 27 September 2022.

¹² United Nations Economist Network, “New economics for sustainable development: attention economy”.

¹³ Twitter, “About Twitter Blue”, и Meta, “Testing Meta Verified to help creators establish their presence”, 17 March 2023.

Рисунок I
Социальными сетями пользуются миллиарды людей



Источник: Kepios, Digital 2023: Global Overview Report (2023).

III. Какова соответствующая международно-правовая база?

17. Содействие информационной добросовестности должно всецело опираться на подходящие международные нормы и стандарты, включая право прав человека и принципы суверенитета и невмешательства во внутренние дела. В августе 2022 года я направил Генеральной Ассамблее доклад под названием «Борьба с дезинформацией в целях поощрения и защиты прав человека и основных свобод»¹⁴. В этом докладе я привел те положения международного права прав человека, которые применимы к проблеме дезинформации, включая Всеобщую декларацию прав человека и Международный пакт о гражданских и политических правах. В соответствии с этими международно-правовыми документами каждый человек имеет право на свободное выражение своего мнения¹⁵.

18. Статья 19 Всеобщей декларации прав человека и пункт 2 статьи 19 Пакта защищают право на свободное выражение мнений, включая свободу искать, получать и распространять всякого рода информацию и идеи независимо от государственных границ и любыми способами. Право человека на свободу выражения мнений распространяется не только на положительно воспринимаемую информацию (A/77/287, п. 13). Связанная со свободой выражения мнений, свобода информации сама по себе является правом. Генеральная Ассамблея постановила, что «свобода информации является основным правом человека и представляет собой критерий всех видов свободы, защите которых Объединенные Нации себя посвятили» (там же, п. 14). Свобода выражения мнений и доступ к информации могут подвергаться определенным ограничениям, которые отвечают конкретным критериям, изложенным в пункте 3 статьи 19 Пакта¹⁶. Государства не могут вводить добавочные основания или ограничивать свободу выражения мнений сильнее, чем разрешено международным правом.

19. Принятый в 2012 году Рабатский план действий по запрещению пропаганды национальной, расовой или религиозной ненависти, представляющей собой подстрекательство к дискриминации, вражде или насилию, служит для государств практическим установочно-правовым ориентиром в отношении того, как лучше всего осуществлять пункт 2 статьи 20 Пакта и статью 4 Международной конвенции о ликвидации всех форм расовой дискриминации, которые запрещают определенные формы языка ненависти. Рабатский план действий уже использовался государствами-членами в разных контекстах¹⁷.

20. Язык ненависти становился предтечей совершения зверских преступлений, включая геноцид. Конвенция 1948 года о предупреждении преступления геноцида и наказании за него запрещает «прямое и публичное подстрекательство к совершению геноцида».

¹⁴ A/77/287.

¹⁵ По состоянию на февраль 2023 года участниками Международного пакта о гражданских и политических правах являлось 173 государства-члена.

¹⁶ Ограничения на свободу выражения мнений должны соответствовать следующим прочно устоявшимся условиям: законность, т. е. ограничения должны быть установлены законом так, чтобы можно было с достаточной точностью отличить правомерное выражение мнения от неправомерного; необходимость и соразмерность, т. е. очевидна минимальность бремени, создаваемого ограничением для осуществления права, и налицо факт или вероятность того, что это ограничение защищает легитимные интересы государства, о которых идет речь; легитимность, т. е. для того, чтобы быть правомерными, ограничения должны защищать только те интересы, которые перечислены в пункте 3 статьи 19 Международного пакта о гражданских и политических правах.

¹⁷ К таким контекстам относятся каналы аудиовизуальной коммуникации в Кот-д'Ивуаре, Марокко и Тунисе и отслеживание случаев подстрекательства к насилию, осуществляемое Многопрофильной комплексной миссией Организации Объединенных Наций по стабилизации в Центральноафриканской Республике.

21. Генеральная Ассамблея в своей резолюции 76/227, принятой в 2021 году, подчеркнула, что все формы дезинформации могут негативно сказываться на осуществлении прав человека и основных свобод, а также на достижении целей в области устойчивого развития. В свою очередь, Совет по правам человека в своей резолюции 49/21, принятой в 2022 году, подтвердил, что дезинформация может негативно воздействовать на осуществление и реализацию всех прав человека.

IV. Каков вред от распространения неверной информации, дезинформации и языка ненависти в онлайн-среде?

22. Распространение неверной информации, дезинформации и языка ненависти в онлайн-среде вызывает серьезную озабоченность у мировой общественности. Изучение данных одного опроса респондентов из 142 стран показало, что 58,5 процента регулярных пользователей Интернета и социальных сетей во всем мире обеспокоены тем, что в онлайн-среде они сталкиваются с неверной информацией, причем молодежь и люди с низким доходом ощущают себя значительно уязвимее¹⁸. Сегодняшняя молодежь — это «цифровые аборигены»; по сравнению с остальным населением у молодежи с большей вероятностью есть выход в онлайн, что делает ее самым «цифropодключенным» поколением в истории¹⁹. В мире каждые полсекунды происходит первый выход какого-нибудь ребенка в Интернет, где он рискует подвергнуться онлайн-овому вреду и языку ненависти, что в некоторых случаях сказывается на психическом здоровье детей²⁰.

23. Воздействие появляющихся в онлайн-среде неверной информации, дезинформации и языка ненависти можно наблюдать повсюду в мире, в том числе в таких сферах, как здравоохранение, противодействие изменению климата, демократия и выборы, гендерное равенство, безопасность и гуманитарное реагирование. В ходе опроса, проведенного в 2021 году, 75 процентов страновых отделений Программы развития Организации Объединенных Наций назвали серьезной проблемой информационное загрязнение. Оно серьезно сказывается на доверии, безопасности, демократии и устойчивом развитии, что подтверждается обзором более 800 академических, журналистских, корпоративных документов и документов гражданского общества, проведенным недавно по заказу ЮНЕСКО²¹.

24. Неверная информация и дезинформация могут быть опасными и потенциально смертоносными, особенно в периоды кризисов, чрезвычайных ситуаций или конфликтов. Во время пандемии коронавирусного заболевания (COVID-19) в Интернете начал циркулировать целый поток неверной информации и дезинформации о вирусе, здравоохранительных мерах и вакцинах²². Некоторые субъекты сыграли на сумятице в своих целях: так, агитаторы-антиваксеры направляли пользователей на сайты, продающие якобы целебные или профилактические средства²³. Многие жертвы COVID-19 отказывались вакцинироваться или

¹⁸ Aleksi Knuutila, Lisa-Maria Neudert and Philip N. Howard, “Who is afraid of fake news? Modeling risk perceptions of misinformation in 142 countries”, *Harvard Kennedy School (HKS) Misinformation Review*, vol. 3, No. 3 (April 2022).

¹⁹ ITU, *Measuring the Information Society* (Geneva, 2013).

²⁰ United Nations Children’s Fund (UNICEF), “Protecting children online”, 23 June 2022. URL: www.unicef.org/protection/violence-against-children-online.

²¹ UNESCO, “Digital governance and the challenges for trust and safety”, working papers. URL: www.unesco.org/en/internet-conference/working-papers.

²² See Julie Posetti and Kalina Bontcheva, “Disinfodemic: deciphering COVID-19 disinformation”, policy brief 1 (Paris, UNESCO, 2020), and “Disinfodemic: dissecting responses to COVID-19 disinformation”, policy brief 2 (Paris, UNESCO, 2020).

²³ Center for Countering Digital Hate, *Pandemic Profiteers: The Business of Anti-Vaxx* (2021).

принимать элементарные предосторожности для охраны здоровья после того, как получали неверную информацию и дезинформацию в Интернете²⁴.

25. Дезинформация может также оказываться смертоносной в социальных и политических контекстах, уже характеризующихся нестабильностью. В своем докладе 2022 года Специальный докладчик по вопросу о поощрении и защите права на свободу мнений и их свободное выражение проанализировала последствия ситуаций, когда информация используется в качестве оружия, с тем чтобы сеять смуту, разжигать ненависть, подстрекать к насилию и затягивать конфликты²⁵. В другом докладе, вышедшем в 2022 году, приводилось заключение, согласно которому «частью дезинформации могут быть неприятие инакомыслия и язык ненависти, направленный против меньшинств, женщин и любых так называемых “других”, что ставит под угрозу не только тех, против кого непосредственно направлена дезинформация, но и принципы инклюзии и социальной сплоченности. Во время чрезвычайных ситуаций, кризисов, ключевых политических моментов или вооруженных конфликтов дезинформация может порождать больше напряженности и разногласий»²⁶.

26. Одни из наихудших последствий онлайн-вреда проявляются в контекстах, которые цифровые платформы обходят вниманием, даже если у них имеется высокий уровень проникновения на рынок. Страны, где бушует конфликт или создалась какая-то иная нестабильная обстановка, зачастую являются менее прибыльными рынками, и на них не выделяется достаточно ресурсов для модерации контента или оказания помощи пользователям. Если традиционные средства массовой информации остаются для большинства людей в зонах конфликтов важным источником новостей, то на цифровых платформах происходило распространение ненависти, также разжигавшее и подпитывавшее насилие²⁷. Некоторые цифровые платформы подвергались критике за их роль в конфликтах, включая войну, происходящую на Украине²⁸.

27. В свою очередь, неверная информация и дезинформация по поводу чрезвычайной климатической ситуации задерживают действия, срочно необходимые для сохранения нашей планеты пригодной для жизни в будущем. Неверную информацию и дезинформацию о климате можно трактовать как ложный или вводящий в заблуждение контент, который разъедает научно согласованную базу, свидетельствующую о существовании, причинах и последствиях антропогенного изменения климата. Ведутся скоординированные кампании с целью отрицать научный консенсус, выработанный Межправительственной группой экспертов по изменению климата, преуменьшать этот консенсус, отвлекать от него и срывать экстренные действия по достижению целей Парижского соглашения 2015 года. Небольшое, но громкоголосое меньшинство отрицателей климатологии²⁹ продолжает

²⁴ Michael A. Gisoni and others, “A deadly infodemic: social media and the power of COVID-19 misinformation”, *Journal of Medical Internet Research*, vol. 24, No. 2 (February 2022).

²⁵ A/77/288.

²⁶ A/77/287, п. 6.

²⁷ В 2018 году независимая международная миссия по установлению фактов, назначенная Советом по правам человека, сочла, что «больше всего случаев ненавистнической риторики в Мьянме» приходится на Facebook (A/HRC/42/50, п. 72).

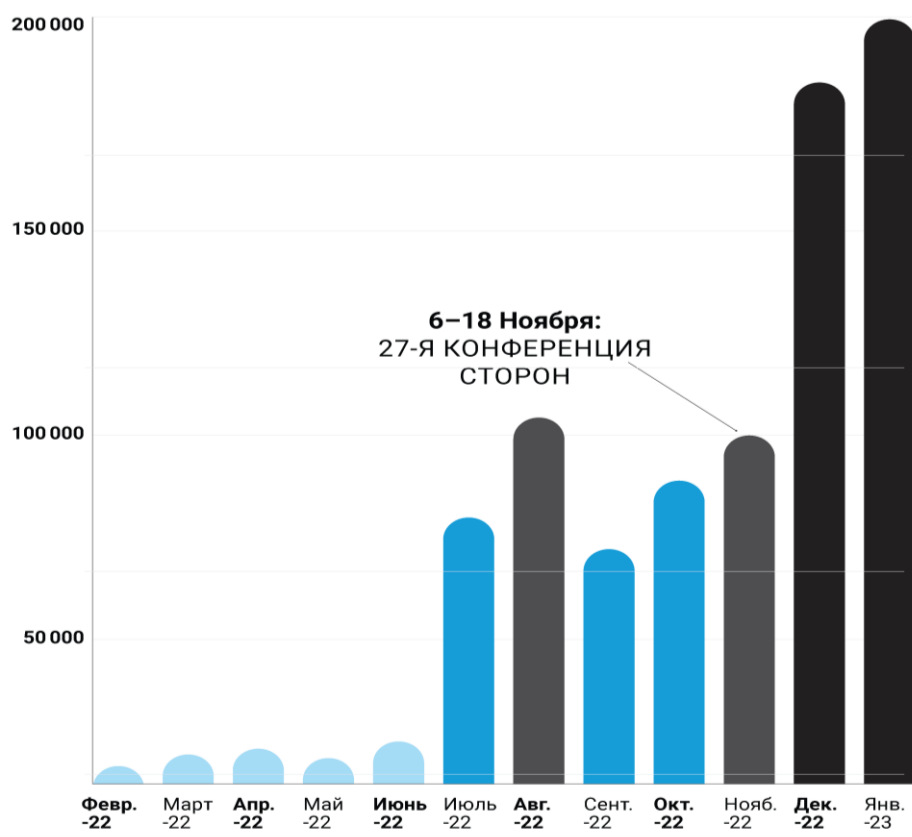
²⁸ См. United Nations News, “Hate speech: a growing, international threat”, 28 January 2023, и “Digital technology, social media fuelling hate speech like never before, warns UN expert”, 20 October 2022.

²⁹ См. John Cook, “Deconstructing climate science denial”, in *Research Handbook in Communicating Climate Change*, David C. Holmes and Lucy M. Richardson, eds. (Cheltenham, United Kingdom, Edward Elgar, 2020). Кук упоминает об опубликованном (Abraham et al., 2014) сводном обзоре того, как в научной литературе решительно опровергались статьи с отрицательскими утверждениями, например о спутниковых измерениях, указывающих на охлаждение, или о расчетах, указывающих на низкую чувствительность климата. Еще в одной публикации (Benestad et al., 2016) сообщается, что авторы попытались

отвергать консенсусную позицию и пользоваться чересчур заметным присутствием на некоторых цифровых платформах. Так, в 2022 году организациями гражданского общества была проведена имитация произвольных поисков, в ходе которой обнаружилось, что алгоритм Facebook рекомендует контент, отрицающий изменение климата, в ущерб климатологическим выкладкам³⁰. В Twitter количество использований хештега #climatescam («климатическая афера») выросло с менее чем 2700 в месяц в первой половине 2022 года до 80 000 в июле и 199 000 в январе 2023 года. Эта фраза также была занесена платформой в число топ-результатов поиска по запросу «климат»³¹. В феврале 2022 года Межправительственная группа экспертов по изменению климата впервые озвучила проблему климатической дезинформации, заявив, что «сознательный подрыв науки» способствует «превратным представлениям о научном консенсусе, неопределенности, игнорированию риска и экстренности, а также расколу»³².

Рисунок II

Помесячное количество использований хештега #climatescam в Twitter



Источник: Департамент глобальных коммуникаций (с использованием данных от Talkwalker).

воспроизвести результаты, приводимые в статьях отрицателей, и выявили ряд изъянов, таких как неподходящие статистические методы, ложные дихотомии и выводы, основанные на неверном понимании физики.

³⁰ Global Witness, "The climate divide: how Facebook's algorithm amplifies climate disinformation", 28 March 2022.

³¹ Анализ Департамента глобальных коммуникаций с использованием данных Talkwalker.

³² Jeffrey A. Hicke and others, "North America", in Intergovernmental Panel on Climate Change, *Climate Change 2022: Impacts, Adaptation and Vulnerability*, Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (Cambridge, United Kingdom, Cambridge University Press, 2022).

28. У некоторых компаний, работающих в сфере ископаемого топлива, вошло в обычай применять стратегию «гринвошинга», ложно убеждая общественность в том, что компания или предприятие делают для защиты окружающей среды больше и наносят ей вреда меньше, чем на самом деле. Компании действуют не в одиночку. Усилиям, призванным запутать публику и отвлечь внимание от ответственности индустрии ископаемого топлива, потакают и способствуют рекламодатели, пиарщики, размещающие рекламу технологические компании, новостные агентства и цифровые платформы³³. Занимающиеся рекламой и пиаром фирмы, которые создают гринвошинговый контент, и третьи стороны, которые его распространяют, коллективно зарабатывают миллиарды на этих усилиях по ограждению индустрии ископаемого топлива от придирчивого внимания и подотчетности. Пиар-фирмы провели сотни кампаний для угольных, нефтяных и газовых корпораций³⁴.

29. Неверная информация и дезинформация оказывают глубокое воздействие на демократию, ослабляя доверие к демократическим институтам и независимым средствам массовой информации и ослабляя участие в политических и общественных делах. Когда на протяжении предвыборного цикла на избирателей обрушивается лживая и вводящая в заблуждение информация, она может лишить их возможности сделать осознанный выбор. Распространение неверной информации и дезинформации способно подрывать общественное доверие к избирательным институтам и самому избирательному процессу (например, к регистрации избирателей, подсчету голосов и оглашению результатов) и может в потенциале приводить к апатии избирателей или неприятию достоверных результатов выборов. Государства и политические лидеры показали себя мощными источниками дезинформации, умышленно и стратегически распространяя неправду, чтобы сохранить или закрепить за собою власть либо чтобы подрывать демократические процессы в других странах³⁵.

30. Маргинализованные и уязвимые группы тоже часто становятся объектами неверной информации, дезинформации и ненавистнической риторики, что приводит к их дальнейшему социальному, экономическому и политическому отчуждению. Кандидатки, избирательницы, работницы избирательных органов, журналистки и представительницы гражданского общества делаются объектами гендеризированной онлайн-дезинформации³⁶. Такие атаки подрывают участие в

³³ Mei Li, Gregory Trencher and Jusen Asuka, “The clean energy claims of BP, Chevron, ExxonMobil and Shell: a mismatch between discourse, actions and investments”, *PLOS ONE*, issue 17, No. 2 (February 2022).

³⁴ Robert J. Brulle and Carter Werthman, “The role of public relations firms in climate change politics”, *Climatic Change*, vol. 169, No. 1–2 (November 2021). По данным «Глобального индекса дезинформации» (некоммерческая организация, следящая за ситуацией), в 2021 году рекламодатели от технологической отрасли перечислили 36,7 млн долл. США 98 веб-сайтам, на которых размещена климатическая дезинформация на английском языке. Центр противодействия цифровой ненависти (инициативная группа) обнародовал в ноябре 2022 года доклад, где показано, что на одном только Google почти половина из 23,7 млн долл. США, потраченных нефтегазовыми компаниями за последние два года на поисковую рекламу, была привязана к поисковым запросам, касающимся экологической устойчивости. Исследование, проведенное InfluenceMap, позволило выяснить, что в 2020 году на платформах Facebook в Соединенных Штатах Америки было размещено 25 147 вводящих в заблуждение рекламных объявлений от 25 организаций нефтегазового сектора, потративших на это в общей сложности 9 597 376 долл. США. До сих пор реакция на эту проблему была несоизмерима с масштабами последней.

³⁵ См. резолюцию 76/227 Генеральной Ассамблеи, резолюцию 49/21 Совета по правам человека и European Union External Action, “Tackling disinformation, foreign information manipulation and interference”, 27 October 2021.

³⁶ Lucina Di Meco, “Monetizing misogyny: gendered disinformation and the undermining of women’s rights and democracy globally”, #ShePersisted, February 2023.

политической жизни и ослабляют демократические институты и права человека, включая свободу выражения мнений этими группами и их доступ к информации³⁷. Эта проблема должна становиться для международного сообщества всё более настойчивым приоритетом, не в последнюю очередь потому, что в 2024 году на избирательные участки по всему миру придет более 2 миллиардов избирателей.

31. Кроме того, между цифровыми платформами и традиционными средствами массовой информации и внутри них происходит «перекрестное опыление» неверной информации и дезинформации, что еще больше затрудняет их отслеживание и устранение — если только они не обнаруживаются прямо у источника. Дезинформация может становиться сознательной тактикой идеологизированных медиа, кооптированных кругами, которые преследуют политические и корпоративные интересы³⁸. В то же время подъем цифровых платформ стал толчком к резкому упадку независимых, заслуживающих доверия средств массовой информации. Новостные аудитории и рекламные поступления в массовом порядке мигрировали на интернет-платформы, и пандемия COVID-19 эту тенденцию усугубила. В некоторых регионах или странах можно наблюдать «вымирание средств массовой информации» или появление «новостных пустынь», когда сообщества теряют достоверные местные источники новостей³⁹, что способствует загрязнению информационной экосистемы. «Обновоствление», при котором спонсируемый контент делают выглядящим так, будто это публикация новостного сюжета, зачастую неадекватно обозначается при размещении материала на цифровых платформах, что придает ему налет легитимности. После того как материал подхватывается другими медиа, цитируется политиками или широко расходуется по разным платформам, происхождение информации всё более затуманивается и у потребителей новостей не остается возможности отличить этот контент от объективного факта.

32. Дезинформация также непосредственно воздействует на работу Организации Объединенных Наций. Со стороны координаторов-резидентов, посланников, посредников и миротворцев звучала озабоченность по поводу влияния дезинформации на оперативную безопасность, эффективность и результативность Организации. В 2022 году был проведен опрос, в ходе которого 70 процентов миротворцев Организации Объединенных Наций заявили, что неверная информация и дезинформация тяжело, критически или умеренно сказываются на их работе, а 75 процентов заявили, что они сказываются на их безопасности и защищенности. Неверная информация и дезинформация могут также становиться орудием, бьющим по гуманитарным организациям, и осложнять операции по спасению жизни людей в конфликтных зонах.

V. Информационная добросовестность и цели в области устойчивого развития

33. Как показано на рисунке III, угрозы для информационной добросовестности могут негативно отражаться на достижении целей в области устойчивого развития.

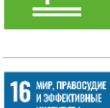
³⁷ См. Andrew Puddephatt, “Social media and elections”, Cuadernos de Discusión de Comunicación e Información, No. 14 (Montevideo, UNESCO, 2019), и Julie Posetti and others, “The chilling: global trends in online violence against women journalists”, research discussion paper (UNESCO, 2021).

³⁸ EU DisinfoLab, “The role of ‘media’ in producing and spreading disinformation campaigns”, 13 October 2021.

³⁹ См. United Nations News, “Social media poses ‘existential threat’ to traditional, trustworthy news: UNESCO”, 10 March 2022, и Anya Schiffrin and others, “Finding the funds for journalism to thrive: policy options to support media viability”, World Trends in Freedom of Expression and Media Development (Paris, UNESCO, 2022).

Рисунок III

Информационная добросовестность и цели в области устойчивого развития

	Неверная информация и дезинформация продолжают сказываться на усилиях по ликвидации нищеты и на глобальной экономике. Экономические невзгоды также могут способствовать тиражированию поляризующих и ненавистнических измышлений, в том числе о маргинализованных группах. Кризис стоимости жизни стал особенно благодатной почвой для распространения дезинформации, согласно которой переход на возобновляемую энергетику якобы виноват, например, в стремительном росте цен на энергию или в утрате рабочих мест.		Неверная информация, дезинформация и язык ненависти, а также чересчур неразборчивое реагирование на эти явления могут пагубно сказываться на инновациях, в том числе ограничивая потенциал маргинализованных групп и делая цифровые пространства менее равноправными и инклюзивными.
	Угрозы для информационной добросовестности могут обострять глобальный голод, в том числе путем усугубления конфликтов, климатических изменений, бедствий, бедности и неравенства. Дезинформация может отвлекать внимание от вызовов для глобальной продовольственной безопасности, порождаемых конфликтами, и затушевывать эти вызовы.		Распространение неверной информации, дезинформации и языка ненависти в Интернете поляризует общества, прицельно бьет по и без того маргинализованным и уязвимым слоям населения и способно приводить к их дальнейшему социальному, экономическому и политическому отчуждению.
	Во время пандемии коронавирусного заболевания (COVID-19) происходила «инфодемия» неверной информации и дезинформации по ее поводу, подрывавшая здравоохранительные меры и вакцинационные кампании. Из-за подверженности вредному контенту сохраняется угроза для здоровья и благополучия детей.		Усилия, направленные на то, чтобы сделать города и общества более устойчивыми, могут подрываться дезинформацией, которая отрицает воздействие человеческой активности на окружающую среду или отвлекает внимание от него.
	Распространение неверной информации, дезинформации и языка ненависти может отрицательно сказываться на доступе к качественному образованию, в частности для маргинализованных групп, включая молодых женщин и девочек. Одну из ключевых ролей в сдерживании социальных последствий онлайн-вреда будут играть доступ к информации и проведение акций по повышению цифровой медийной грамотности с целью усилить сопротивляемость.		Активисты, стоящие за инициативами по развитию экономики замкнутого цикла и внедрению методов нулевой отходности, становились объектами ненавистнической риторики и дезинформации в Интернете.
	Ненавистническая риторика и дезинформация на гендерной почве призваны систематически закабалить женщин, принуждая их к молчанию и вытесняя их из публичной сферы. Они могут иметь разрушительные последствия — от подавления женских голосов и нагнетания самоцензуры до причинения профессионального и репутационного ущерба и подстрекательства к физическому насилию.		Климатическая дезинформация и поощряемая ею инертность подрывают усилия по принятию экстренных мер к преодолению климатического кризиса. В частности, они мешают принципиальному переходу от загрязняющих ископаемых видов топлива на возобновляемые энергоисточники и срочному инвестированию в климатическую выносливость.
	Два миллиарда человек живут без безопасно организованного снабжения питьевой водой. Неверная информация и дезинформация относительно безопасности питьевой воды и санитарии могут иметь вредные последствия для здоровья^a.		Неверная информация и дезинформация могут негативно сказываться на усилиях по сохранению и устойчивому использованию океанов, морей и морских ресурсов.
	Неверная информация и дезинформация о климате, которые во многом насаждаются индустрией ископаемого топлива, саботируют экстренный переход на более чистые формы производства энергии, сужая закрывающуюся перспективу добиться устойчивого будущего для всех.		Экологические активисты, занимающиеся защитой экосистем суши, становились объектом ненавистнических и дезинформационных онлайн-кампаний, оборачивавшихся последствиями в реальной жизни. Неверная информация и дезинформация о климате подрывают усилия по противодействию климатическим изменениям^c.
	Исследования говорят о пагубном воздействии неверной информации, дезинформации и языка ненависти на экономику^b.		Дезинформация и язык ненависти использовались, чтобы оказывать влияние на выборы и публичные нарративы и чтобы запугивать людей. Они использовались для компрометации оппонентов, торпедирования миротворческих усилий, подстрекательства к насилию, затягивания конфликтов и подрыва доверия к верховенству права. В результате серьезно страдали усилия по выстраиванию мирных и инклюзивных обществ, а также роль Организации Объединенных Наций в деле поддержки мира и безопасности^d.
			Неверная информация, дезинформация и язык ненависти могут мешать налаживанию значимых партнерств по достижению целей в области устойчивого развития, а отвлечение ресурсов на устранение этой проблемы может ослаблять усилия, призванные не обойти никого вниманием.

^a World Health Organization and United Nations Children's Fund, *Progress on Household Drinking Water, Sanitation and Hygiene 2000–2020: Five Years into the SDGs* (Geneva, 2021).

^b См. Roberto Cavazos and CHEQ, "The economic cost of bad actors on the Internet: fake news, 2019", и London Economics, "The cost of lies: assessing the human and financial impact of COVID-19 related online misinformation on the UK", December 2020.

^c Global Witness, *Last Line of Defence: The Industries Causing the Climate Crisis and Attacks against Land and Environmental Defenders* (2021).

^d A/77/288.

VI. Как нам усилить информационную добросовестность?

34. Неверная информация, дезинформация и язык ненависти существуют не в вакууме. Эти явления распространяются, когда люди чувствуют себя отчужденными и неслышанными, когда они сталкиваются с проявлениями экономического неравенства и когда они чувствуют политическое разочарование. При реагировании на указанные явления следует учитывать эти вызовы, порождаемые реальной жизнью. Основополагающее значение для построения мира, в котором может быть восстановлено доверие, имеют усилия по достижению целей в области устойчивого развития.

35. Продумывая меры такого реагирования, важно не упускать из виду ту колоссальную ценность, которую цифровые платформы приносят в мир. Эти платформы революционизировали происходящую в реальном времени массовую коммуникацию, позволив распространять спасительную для людей информацию во время стихийных бедствий и пандемий. Они помогли мобилизовать поддержку в интересах тех целей, к которым стремится Организация Объединенных Наций, часто оказываясь позитивными силами, выступающими за инклюзию и участие в общественной жизни. Они соединили географически разрозненные человеческие сообщества, которые иначе оставались бы неохваченными, включая людей, страдающих редкими расстройствами здоровья, и огромную массу активистов, работающих над тем, чтобы сделать мир лучше.

Регуляторное реагирование

36. Вопрос о том, могут и должны ли цифровые платформы нести юридическую ответственность за размещенный на них контент, стал предметом продолжительных дебатов. В некоторых контекстах для противодействия угрозам, которым подвергается информационная добросовестность, эффективно используются существующие законы по борьбе с диффамацией, кибербуллингом и домогательствами, без введения новых ограничений на свободу выражения мнения ([A/77/287](#), п. 44).

37. Кроме того, в последнее время имели место некоторые законодательные начинания, призванные решать этот вопрос на региональном и национальном уровнях. К их числу относятся рамочные установления, которые были утверждены в 2022 году Европейским союзом и включают Закон о цифровых услугах, инициативу «Прозрачность и таргетирование политической рекламы» и Кодекс практики по борьбе с дезинформацией. Закон о цифровых услугах устанавливает новые правила для пользователей, цифровых платформ и работающих в онлайн-среде бизнесов в Европейском союзе. Предусматриваемые меры нацеливаются на борьбу с незаконными онлайн-товарами, услугами и контентом и снабжают пользователей механизмом как для мечения незаконного контента, так и для оспаривания модераторских решений, которые приняты не в их пользу. Эти меры требуют, чтобы цифровые платформы повышали прозрачность, особенно в части использования и характера алгоритмов рекомендаций, и чтобы более крупные платформы предоставляли исследователям доступ к данным.

38. Кодекс практики по борьбе с дезинформацией устанавливает принципы и обязательства онлайн-платформ и рекламного сектора по противодействию распространению онлайн-дезинформации в Европейском союзе, которые подписанты согласились выполнять⁴⁰. К ним относятся добровольные обязательства помогать с демонетизацией дезинформации — как путем предотвращения

⁴⁰ European Commission, “Shaping Europe’s digital future: the 2022 Code of Practice on Disinformation”, 4 July 2022.

распространения рекламы, содержащей дезинформацию, так и путем недопущения размещения рекламы рядом с контентом, содержащим дезинформацию. Подписанты также согласились четче маркировать политическую рекламу, сообщая при этом детали о спонсоре, рекламных затратах и продолжительности демонстрации, и создать базы данных о политической рекламе, дающие возможность поиска. Кроме того, они обязались делиться информацией о случаях, когда на их платформах обнаруживается использование злонамеренных манипуляторских приемов (как то: подделка аккаунтов, амплификация с привлечением ботов, выдача себя за другое лицо и вредоносные дипфейки), и регулярно осуществлять и обновлять политику борьбы с ними. Другие обязательства посвящены наделению пользователей способностью распознавать, понимать и помечать дезинформацию, а также укреплению сотрудничества со специалистами по проверке фактов и расширению доступа к данным для исследователей. Реальным тестом для этих новых механизмов станет их осуществление.

39. Одна из ключевых целей Кодекса практики по борьбе с дезинформацией заключается в повышении прозрачности платформ. В феврале 2023 года подписанты Кодекса практики опубликовали свои первые, базисные отчеты о ходе выполнения обязательств. Эти отчеты позволили получить представление о том, насколько удалось предотвратить попадание рекламных доходов к дезинформаторам, и об обнаружении других манипуляторских приемов, включая широко-масштабные скоординированные усилия по манипулированию в нескольких европейских странах общественным мнением о войне на Украине⁴¹.

Реагирование цифровых платформ

40. Цифровые платформы весьма разнообразны по размеру, функциям и структуре и применяют широкий спектр мер по борьбе с причиняемым вредом. Несколько более крупных платформ публично обязались соблюдать Руководящие принципы предпринимательской деятельности в аспекте прав человека⁴², но пробелы в политике, прозрачности и реализации сохраняются. Некоторые платформы не обеспечивают соблюдение их собственных стандартов и в той или иной степени допускают и амплифицируют ложь и ненависть⁴³. Алгоритмы, созданные для продвижения той модели получения прибыли, которой придерживаются цифровые платформы, настроены на умышленное нагнетание заинтересованности и монополизацию внимания, что имеет тенденцию подталкивать пользователей к поляризующему или провокационному контенту.

41. У большинства цифровых платформ есть та или иная система механизмов саморегулирования, модерации или надзора, однако сохраняется такая проблема, как прозрачность политики и практики в отношении удаления контента⁴⁴. Инвестирование в эти механизмы осуществляется крайне неоднородно, разнясь в зависимости от региона и языка, и сосредоточено в основном на глобальном Севере, где наблюдается и более строгое обеспечение платформами соблюдения их собственных правил. Недавно проведенное обследование показало, что перевод модераторских инструментов и надзорных механизмов на местные языки

⁴¹ См. замечания Веры Йоуровой, заместителя Председателя Европейской комиссии по ценностям и прозрачности (European Commission, “Code of Practice on Disinformation: new Transparency Centre provides insights and data on online disinformation for the first time”, daily news, 9 February 2023. URL: https://ec.europa.eu/commission/presscorner/detail/en/mex_23_723).

⁴² URL: <https://unglobalcompact.org/library/2>.

⁴³ Center for Countering Digital Hate and Human Rights Campaign, “Digital hate: social media’s role in amplifying dangerous lies about LGBTQ+ people”, 10 August 2022.

⁴⁴ См. Andrew Puddephatt, “Letting the sun shine in: transparency and accountability in the digital age”, World Trends in Freedom of Expression and Media Development (Paris, UNESCO, 2021).

осуществлен на разных платформах не полностью⁴⁵. В то же время модерация часто отдается на аутсорсинг, а ресурсов для ее обеспечения на языках помимо английского плачевно мало⁴⁶. Из свидетельств модераторов вырисовываются тревожные вопросы, касающиеся плохого обращения, соблюдения трудовых стандартов и вторичной травматизации⁴⁷. Модераторы сообщают, что постоянно сталкиваются с жестоким и будоражающим контентом и что им даются считанные секунды, чтобы определить, нарушает ли политику компании опубликованный пост, о котором им просигнализировали. Немаловажную роль могут играть автоматизированные системы модерации контента, но в них может быть заложена предвзятость из-за данных и структур, используемых для их обучения. Кроме того, у них высок процент ошибок на английском языке, становящийся еще более высоким на других языках. У ряда цифровых платформ функционируют команды, занимающиеся вопросами доверия и безопасности, прав человека и информационной добросовестности, однако входящие в них эксперты нередко не участвуют в ранних стадиях разработки продукта, а при введении режима экономии их должности зачастую первыми попадают под сокращение.

Доступ к данным

42. Еще одним настойчивым приоритетом глобального масштаба является обеспечение доступа к данным для исследователей. В существующих исследованиях и ресурсах по-прежнему наблюдается сильный перекос в сторону Соединенных Штатов Америки и Европы. Если не считать нескольких примечательных исключений (включая доклады о деятельности Организации Объединенных Наций по поддержанию мира в Африке и о работе независимой международной миссии по установлению фактов в Мьянме⁴⁸, а также некоторые журналистские расследования и репортажи⁴⁹), то исследований о воздействии на остальной мир было опубликовано немного. Отчасти это объясняется тем, что у исследователей недостаточно доступа к платформам и их данным. Кроме того, инструменты, необходимые для эффективного изучения того ограниченного круга данных, который предоставляется платформами, спроектированы обычно с расчетом на маркетинг и являются в основном запредельно дорогими. Возможность правильно оценивать проявления вреда могла бы появиться у исследователей в случае, если бы платформы перенастроились с «доступа по запросу» на «раскрытие по умолчанию», сопроводив это введением необходимых гарантий конфиденциальности.

Расширение прав и возможностей пользователей

43. Группами гражданского общества и академическими кругами проведены обширные исследования о том, как лучше всего бороться с неверной информацией, дезинформацией и языком ненависти, защищая при этом свободу выражения мнений. В ряде таких исследований подчеркнута необходимость решений, которые работают по принципу «снизу вверх», а именно наделяют пользователей Интернета способностью ограничить воздействие онлайн-вреда на их сообщества и ликвидировать централизацию власти в руках платформ.

⁴⁵ Whose Knowledge?, Oxford Internet Institute and The Centre for Internet and Society, *State of the Internet's Languages Report* (2022).

⁴⁶ [A/HRC/38/35](#).

⁴⁷ Billy Perrigo, "Inside Facebook's African sweatshop", *Time*, 17 February 2022.

⁴⁸ [A/HRC/42/50](#).

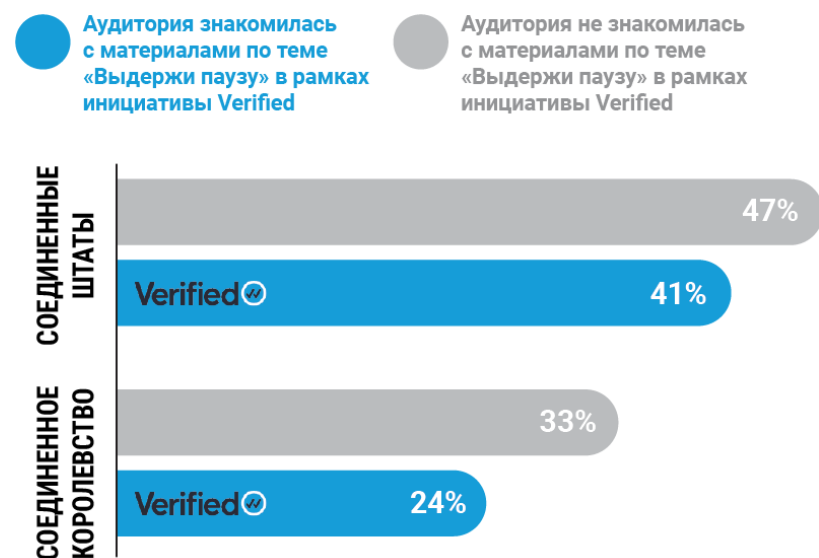
⁴⁹ К числу примечательных публикаций можно отнести следующие: Maria Ressa, *How to Stand up to a Dictator* (New York, HarperCollins, 2022), и Max Fischer, *The Chaos Machine* (New York, Little, Brown and Company, 2022).

44. Пользователей платформ, в том числе из состава маргинализированных групп, следует побуждать к присутствию в политическом пространстве, приобщать к этому пространству и вовлекать в него. При этом молодежь обладает богатыми и глубокими навыками. Будучи «цифровыми аборигенами», молодежь (особенно девушки) и дети уже часто становятся мишенью для неверной информации, дезинформации и языка ненависти и будут непосредственно затрагиваться новыми и формирующимися платформами. Более юные пользователи могут по собственному опыту говорить о дифференцированном воздействии различных предложений и их потенциальных недостатках. Кроме того, они вносили активный вклад в разъяснительную работу и проверку фактов в Интернете⁵⁰.

45. Совершенствование навыков критического мышления может делать пользователей менее восприимчивыми к цифровому манипулированию. В частности, цифровая грамотность учит пользователей полнее оценивать информацию, с которой они сталкиваются в Интернете, и ответственно относиться к ее передаче другим. Разными структурами Организации Объединенных Наций накоплен ценный опыт в этой области. Реализуемая Организацией Объединенных Наций инициатива Verified⁵¹ успешно внедрила ряд приемов, включая направление таргетированных сообщений пользователям, упреждающее разоблачение ложных утверждений еще до того, как пользователи с ними сталкиваются, и проведение акций по повышению цифровой грамотности.

Рисунок IV

Эффективность кампаний Организации Объединенных Наций по противодействию неверной информации и дезинформации: вероятность обмена фейковыми новостями (2021 год)^a



^a На основе исследования, проведенного в марте 2021 года Массачусетским технологическим институтом.

⁵⁰ См. UNICEF, “Young reporters fact-checking COVID-19 information”.

⁵¹ См. <https://shareverified.com/>.

Меры дестимулирования

46. В текущих бизнес-моделях большинства цифровых платформ стремление увлечь аудиторию ставится выше прав человека, конфиденциальности и безопасности. Это включает монетизацию персональных данных ради наживы, происходящую несмотря на растущие доказательства социального вреда, который причиняется такой бизнес-моделью.

47. Некоторые группы гражданского общества и исследователи изучали способы, позволяющие демонетизировать, а следовательно, дестимулировать генерирование и распространение в Интернете неверной информации, дезинформации и языка ненависти, отмечая при этом, что если свобода выражения мнений является одним из основных прав человека, то нажива на ней таковым не является⁵². Озвучиваемые предложения направлены на то, чтобы бороться с прибыльностью дезинформации, добиваться полной прозрачности по части монетизации контента и независимых оценок рисков, а также лишать тех, кто занимается онлайн-рекламой, стимулов к потворствованию дезинформации.

48. Бренды, размещающие рекламу рядом с контентом, содержащим неверную информацию, дезинформацию и ненавистническую риторику, рискуют подорвать эффективность своих рекламных кампаний, а в конечном счете — и свою репутацию. Рекламодатели могут выработать политику, которая четко предписывала бы избегать нечаянного финансирования и легитимации неверной информации, дезинформации и языка ненависти, а также помогала бы делать такой контент убыточным. Способы реализации этой политики могут включать ведение актуализируемых «черных» и «белых» списков и применение инструментов верификации рекламы. Рекламодатели могут также оказывать давление на цифровые платформы, побуждая их к активизации действий по защите информационной добросовестности, и воздерживаться от размещения рекламы на медиа-площадках, которые разжигают ненависть и распространяют дезинформацию⁵³.

Независимые средства массовой информации

49. Введенные в десятках стран новые меры продолжают подрывать свободу прессы. Согласно глобальному докладу за 2022 год из флагманской серии ЮНЕСКО «Глобальные тенденции в области свободы выражения мнений и развития средств массовой информации», 85 процентов населения мира столкнулось за предшествующие пять лет с упадком свободы прессы в своей стране⁵⁴. В условиях, когда 2,7 миллиарда человек до сих пор не в сети⁵⁵, еще одним приоритетом становится укрепление независимых средств массовой информации, повышение распространенности инициатив по проверке фактов и поддержка

⁵² Некоммерческая группа «Глобальный индекс дезинформации» отслеживает рекламу, размещаемую вместе с дезинформацией. Жертвой этой практики оказалась и Организация Объединенных Наций: «Глобальный индекс дезинформации» обнаружил случаи, когда рекламные объявления Детского фонда Организации Объединенных Наций были размещены рядом с антивакцинаторскими статьями, а рекламные объявления Управления Верховного комиссара Организации Объединенных Наций по делам беженцев — рядом с контентом, направленным против беженцев.

⁵³ Conscious Advertising Network, “The Manifestos”. URL: www.consciousadnetwork.com/the-manifestos/.

⁵⁴ UNESCO, *Journalism is a Public Good: World Trends in Freedom of Expression and Media Development – Global Report 2021/2022* (Paris, 2022).

⁵⁵ ITU, “Facts and figures 2021: 2.9 billion people still offline”, 29 November 2021. В глобальном цифровом договоре, который государствам-членам предстоит обсудить на Саммите будущего в 2024 году, будут изложены общие принципы открытого, свободного и защищенного цифрового будущего для всех (см. www.un.org/techenvoy/global-digital-compact).

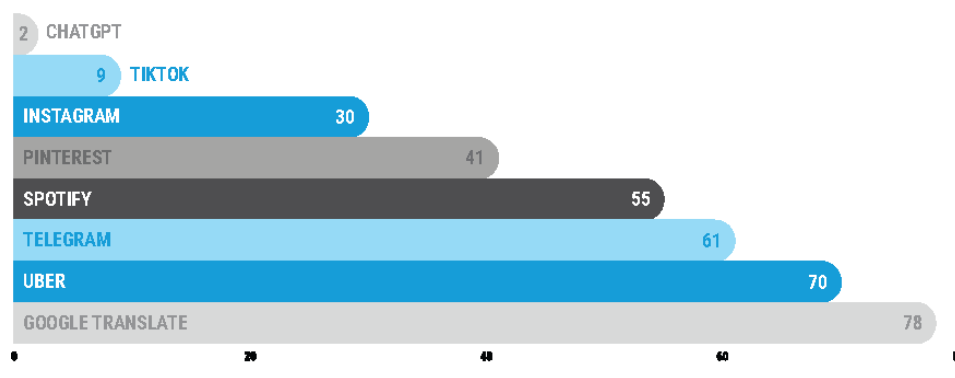
достоверного и точного репортерства в общественных интересах. Настоящие публичные дебаты опираются на факты, которые ясно изложены и этично и независимо донесены. Этичные репортеры, имеющие качественное обучение и качественные условия труда, обладают навыками для того, чтобы выправить баланс перед лицом неверной информации и дезинформации. Они способны предлагать жизненно важную услугу: предоставление точной, объективной и достоверной информации по значимым вопросам.

Рассчитанность на будущее

50. Пока мы ищем решения, позволяющие защитить информационную добросовестность при нынешнем ландшафте, мы должны проследить за тем, чтобы рекомендации были рассчитаны на будущее, распространяясь на формирующиеся и еще не появившиеся технологии. В ноябре 2022 года компанией OpenAI была запущена платформа ChatGPT-3, которая к январю 2023 года набрала 100 миллионов пользователей, что сделало ее самым быстрорастущим в истории потребительским приложением⁵⁶, и многие другие компании спешат разработать конкурирующие с ней инструменты. В искусственном интеллекте заложен почти невообразимый потенциал для преодоления глобальных вызовов, однако есть серьезные и настойчивые опасения по поводу того, что последние достижения в области искусственного интеллекта, включая генераторы изображений и дипфейк-видео, обладают не менее мощным потенциалом как угроза для информационной добросовестности. Как показывают недавние отчеты и исследования, инструменты генеративного искусственного интеллекта формировали неверную информацию, дезинформацию и ненавистническую риторику, которые убедительно выдавались пользователям за факт⁵⁷.

Рисунок V

Количество месяцев, понадобившихся для прибавления 100 миллионов активных пользователей в месяц у ChatGPT — в сопоставлении с другими популярными приложениями



Источник: Similarweb (с использованием данных от Sensor Tower).

51. Мой Посланник по вопросам технологий возглавляет усилия по оценке последствий появления генеративного искусственного интеллекта и других формирующихся платформ. При этом мы должны учиться на ошибках прошлого.

⁵⁶ Krystal Hu, "ChatGPT sets record for fastest-growing user base - analyst note", Reuters, 2 February 2023.

⁵⁷ См. Center for Countering Digital Hate, "Misinformation on Bard, Google's new AI chat", 5 April 2023; и Tiffany Hsu and Stuart A. Thompson, "Disinformation researchers raise alarms about A.I. chatbots", *The New York Times*, 13 February 2023.

Цифровые платформы были вброшены в мир без достаточного осознания или оценки потенциального ущерба для общества и отдельно взятых людей. Сейчас у нас есть возможность обеспечить, чтобы с формирующимися технологиями эта история не повторилась. Эпоху философии «двигайся быстро и ломай вещи», которую провозгласили в Силиконовой долине, нужно заканчивать. Совершенно необходимо, чтобы во все новые технологии и продукты были с самого начала встроены такие параметры, как конфиденциальность пользователей, защищенность, прозрачность и безопасность по дизайну.

Реагирование Организации Объединенных Наций

52. Предпринимаются также (в том числе по линии операций Организации Объединенных Наций в пользу мира и ее страновых отделений) шаги, предусматривающие мониторинг угрозы, которую неверная информация и дезинформация создают для выполнения мандатов Организации Объединенных Наций, анализ этой угрозы и реагирование на нее. В Стратегии и плане действий Организации Объединенных Наций по борьбе с языком ненависти приводятся стратегические указания в отношении того, как Организации надлежит противодействовать языку ненависти на национальном и глобальном уровнях. В феврале 2023 года ЮНЕСКО устроила конференцию «Интернет доверия», чтобы обсудить проект глобального руководства по регулированию цифровых платформ, который предстоит окончательно оформить до конца этого года⁵⁸.

53. Совокупность этих инициатив и подходов помогает наметить курс дальнейшей проработки основополагающих принципов для кодекса поведения Организации Объединенных Наций.

VII. К кодексу поведения Организации Объединенных Наций

54. Выдвигаемый мною кодекс поведения Организации Объединенных Наций для обеспечения информационной добросовестности на цифровых платформах будет строиться на следующих принципах:

- приверженность информационной добросовестности;
- уважение прав человека;
- поддержка независимых средств массовой информации;
- усиление прозрачности;
- расширение прав и возможностей пользователей;
- совершенствование исследований и доступа к данным;
- наращивание мер реагирования;
- более активное дестимулирование;
- повышение доверия и безопасности.

55. Эти принципы почерпнуты из стержневых идей, обсуждаемых в настоящей концептуальной записке, и они согласуются и взаимосвязаны с моей концептуальной запиской о глобальном цифровом договоре. Государствам-членам будет предложено внедрить кодекс поведения на национальном уровне. Будут продолжены консультации с заинтересованными сторонами, призванные

⁵⁸ Проект руководства имеется по адресу: www.unesco.org/en/internet-conference.

дополнительно уточнить содержание кодекса поведения, а также определить конкретные методологии претворения его принципов в жизнь.

56. Кодекс поведения может основываться на следующих рекомендациях:

Приверженность информационной добросовестности

а) всем заинтересованным сторонам следует воздерживаться от использования, поддержки или амплификации дезинформации и языка ненависти с любыми намерениями, в том числе для достижения политических, военных или других стратегических целей, подстрекательства к насилию, подрыва демократических процессов или нападок на гражданское население, уязвимые группы, сообщества или отдельных лиц;

Уважение прав человека

б) государствам-членам следует:

i) обеспечивать, чтобы реагирование на неверную информацию, дезинформацию и язык ненависти происходило в соответствии с международным правом, включая международное право прав человека, и не перерождалось в блокирование какого-либо легитимного выражения взглядов или мнений, в том числе путем повального отключения Интернета либо запрещения платформ или средств массовой информации;

ii) ввести регуляторные меры для защиты основных прав пользователей цифровых платформ, включая правоприменительные механизмы, обеспечив при этом полную прозрачность требований, предъявляемых к технологическим компаниям;

с) всем заинтересованным сторонам следует соблюдать Руководящие принципы предпринимательской деятельности в аспекте прав человека;

Поддержка независимых средств массовой информации

д) государства-члены должны гарантировать свободный, жизнеспособный, независимый и плюралистичный медиаландшафт, отмечающийся прочной защитой журналистов и независимых средств массовой информации, и поддерживать создание, финансирование и обучение независимых организаций по проверке фактов на местных языках;

е) средства массовой информации должны обеспечивать достоверное и этичное независимое репортерство, подкрепляемое качественным обучением и адекватными условиями труда в соответствии с международными нормами и стандартами в области труда и прав человека;

Усиление прозрачности

ф) цифровые платформы должны:

i) обеспечивать значимую прозрачность в отношении алгоритмов, данных, модерации контента и рекламы;

ii) публиковать и обнародовать доступное изложение политики в отношении неверной информации, дезинформации и языка ненависти, а также сообщать о частоте случаев скоординированной дезинформации на их сервисах и об эффективности политики противодействия таким операциям;

g) новостные средства массовой информации должны обеспечивать значимую прозрачность источников финансирования и рекламной политики, а также четко отграничивать редакционный контент от платной рекламы, в том числе при публикации на цифровых платформах;

Расширение прав и возможностей пользователей

h) государства-члены должны обеспечивать публичный доступ к точной, прозрачной и поступающей из достоверных источников правительственной информации, особенно информации, которая служит общественным интересам, включая все аспекты целей в области устойчивого развития;

i) цифровые платформы должны предусматривать прозрачный порядок обеспечения прав и защиты пользователей, предоставляя людям больше выбора в отношении того, какой контент они видят и как используются их данные. Они должны давать пользователям такие способы доказывания и аутентификации их личности, которые не сопряжены с монетизацией или уступками по части конфиденциальности, и устанавливать прозрачные процедуры для подачи пользователями жалоб и сигнализирования ими о нарушениях, опирающиеся на независимые, доступные и преданные широкой гласности механизмы рассмотрения жалоб;

j) всем заинтересованным сторонам следует инвестировать в активные кампании за цифровую грамотность, вырабатывающие у пользователей всех возрастов способность лучше разбираться в том, как работают цифровые платформы, как могут использоваться их персональные данные, а также как выявлять неверную информацию, дезинформацию и язык ненависти и реагировать на них. Особое внимание следует уделять обеспечению того, чтобы молодежь, подростки и дети были полностью осведомлены о своих правах на онлайн-площадках;

Совершенствование исследований и доступа к данным

k) государствам-членам следует поддерживать и инвестиционно подпитывать независимые исследования, призванные выяснять распространенность и воздействие неверной информации, дезинформации и ненавистнической риторики в разных странах и на разных языках, особенно в контекстах, где сервисного обслуживания недостаточно, и на языках помимо английского, позволяя при этом гражданскому обществу и академическим кругам действовать в условиях свободы и безопасности;

l) цифровые платформы должны:

i) давать исследователям и ученым доступ к данным, уважая при этом конфиденциальность пользователей. Исследователи должны иметь возможность собирать примеры и качественные данные в отношении групп и отдельно взятых лиц, ставших объектами неверной информации, дезинформации и ненавистнической риторики, чтобы лучше уяснить масштаб и характер причиняемого вреда, уважая при этом защищенность данных и права человека;

ii) обеспечивать всестороннее участие гражданского общества в усилиях по преодолению неверной информации, дезинформации и языка ненависти;

Наращивание мер реагирования

- m) все заинтересованные стороны должны:
 - i) выделять ресурсы для выяснения и оглашения истоков, каналов и последствий распространения неверной информации, дезинформации и языка ненависти, соблюдая при этом нормы и стандарты в области прав человека, и дополнительно инвестировать в потенциал для проверки фактов в разных странах и контекстах;
 - ii) формировать широкие коалиции за обеспечение информационной добросовестности, позволяющие соединять разные подходы и экспертные наработки для того, чтобы способствовать преодолению разрыва между местными организациями и действующими в глобальном масштабе технологическими компаниями;
 - iii) способствовать обучению и наращиванию потенциала, чтобы развивать понимание того, как неверная информация, дезинформация и язык ненависти проявляют себя, и укреплять стратегии их профилактики и смягчения их последствий;

Более активное дестимулирование

- n) цифровым платформам следует уйти от бизнес-моделей, в которых стремление увлечь аудиторию ставится выше прав человека, конфиденциальности и безопасности;
- o) рекламодатели и цифровые платформы должны следить за тем, чтобы рекламные объявления не размещались рядом с неверной информацией, дезинформацией или ненавистническими высказываниями в онлайн-среде и чтобы не поощрялась реклама, содержащая дезинформацию;
- p) новостные средства массовой информации должны обеспечивать, чтобы весь платный рекламный и рекламодательский контент был четко обозначен как таковой и не содержал неверной информации, дезинформации и языка ненависти;

Повышение доверия и безопасности

- q) цифровые платформы должны:
 - i) следить за тем, чтобы дизайн всех продуктов был рассчитан на безопасность и конфиденциальность, в том числе путем адекватного ресурсного обеспечения внутренних экспертных кадров, занимающихся вопросами доверия и безопасности, и чтобы это сопровождалось последовательным применением действующей политики в разных странах и на разных языках;
 - ii) инвестировать в системы модерации контента человеком и искусственным интеллектом на всех языках, используемых в странах функционирования, и обеспечивать прозрачность механизмов сигнализирования о контенте и ускорение процедур отклика на поступающие сигналы, особенно в обстановке конфликта;
- r) всем заинтересованным сторонам следует принять экстренные и незамедлительные меры для того, чтобы обеспечивать защищенное, безопасное, ответственное, этичное и согласующееся с правами человека использование искусственного интеллекта, и заниматься вопросом о том, что последние достижения в этой области означают для распространения неверной информации, дезинформации и языка ненависти.

VIII. Следующие шаги

57. Секретариат Организации Объединенных Наций проведет с различными заинтересованными сторонами широкие консультации о разработке кодекса поведения Организации Объединенных Наций, включая механизмы отслеживания и осуществления. Это могло бы включать создание независимой обсерватории, состоящей из признанных экспертов и занимающейся оценкой мер, принимаемых субъектами, которые обязались придерживаться кодекса поведения, и других отчетных механизмов.

58. Чтобы поддерживать кодекс и способствовать его содержательности, Секретариат Организации Объединенных Наций может проводить углубленные исследования, позволяющие усилить понимание информационной добросовестности в глобальных масштабах, особенно в недостаточно исследованных районах мира.

59. Генеральный секретарь создаст в Секретариате Организации Объединенных Наций специализированное подразделение для наращивания мер реагирования на такие проявления неверной информации, дезинформации и языка ненависти в онлайн-среде, которые сказываются на выполнении мандатов Организации Объединенных Наций и на ее субстантивных приоритетах. Опираясь на экспертный мониторинг и анализ, такое подразделение занималось бы разработкой отлаженных коммуникационных стратегий для прогнозирования угроз и/или быстрого реагирования на них до того, как они перерастут в онлайн- и офлайн-вред, и поддерживало бы развитие потенциала у сотрудников Организации Объединенных Наций и у государств-членов. Оно стало бы поддерживать усилия, прилагаемые государствами-членами, цифровыми платформами и другими заинтересованными сторонами для соблюдения и внедрения кодекса после его окончательного оформления.

IX. Заключение

60. Усиление информационной добросовестности на цифровых платформах является для международного сообщества настойчивым приоритетом. Принимаемые в столь разных сферах, как здравоохранение, гендерное равенство, мир, справедливость, образование и противодействие изменению климата, меры по сдерживанию воздействия неверной информации, дезинформации и языка ненависти будут способствовать усилиям, призванным достичь устойчивого будущего и ни о ком не забыть. Даже если на национальном уровне действия налажены, полноценно заниматься этими проблемами можно будет только при более активном глобальном сотрудничестве. Стержневые идеи, изложенные в настоящей концептуальной записке, демонстрируют, что продвижение к большей информационной добросовестности должно основываться на правах человека и быть многосторонним и многоплановым. Эти идеи скомпонованы в ряд принципов, который предлагается рассмотреть на предмет отражения в кодексе поведения Организации Объединенных Наций для обеспечения информационной добросовестности на цифровых платформах, а сам этот кодекс задумывается как схема для упрочения информационной добросовестности, сопровождаемого энергичным отстаиванием прав человека. Я рассчитываю на сотрудничество с государствами-членами и другими заинтересованными сторонами в том, чтобы превратить эти принципы в осязаемые обязательства.

Приложение

Консультации с государствами-членами и другими соответствующими заинтересованными сторонами

1. Изложенные в настоящей концептуальной записке мысли опираются на предложения, приводимые в докладе под названием «Наша общая повестка дня» ([A/75/982](#)), составлению которого способствовали обширные консультации с государствами-членами, системой Организации Объединенных Наций, идейными лидерами, молодежью и деятелями гражданского общества со всего мира. Концептуальная записка является, в частности, ответом на развернутые и содержательные соображения относительно «Нашей общей повестки дня», озвученные государствами-членами и другими заинтересованными сторонами в ходе 25 дискуссий на Генеральной Ассамблее.
 2. Перед опубликованием настоящей концептуальной записки были проведены консультации с государствами-членами, в том числе в рамках неофициального брифинга для Комитета по информации, на который были приглашены и все страны, не являющиеся членами Комитета. Состоялись также обсуждения с партнерами из гражданского общества, представителями академических кругов, экспертами и частным сектором, включая технологические компании.
 3. В преддверии Саммита будущего будут проведены широкие консультации по выработке кодекса поведения.
-