



# Открытые данные для искусственного интеллекта (ИИ)

Перевод исследования подготовлен РЦИС



PDF

# ЮНЕСКО

## Открытые данные для искусственного интеллекта (ИИ)

### Что дальше?

Настоящая публикация доступна в открытом доступе по лицензии attribution-Share Alike 3.0 IGO (CC-BY-SA 3.0 IGO) по ссылке <http://creativecommons.org/licenses/by-sa/3.0/igo>. Используя содержание данной публикации, пользователи соглашаются с условиями работы с Репозиторием Открытого Доступа ЮНЕСКО

(<https://en.unesco.org/open-access/terms-use-ccbysa-en>).

Используемые обозначения и представление материала в данной публикации не подразумевают выражения определенной позиции со стороны ЮНЕСКО относительно правового статуса какой-либо страны, территории, города или района, их органов власти или относительно делимитации их границ или зон ответственности.

Идеи и мнения, выраженные в данной публикации, принадлежат авторам; они не обязательно принадлежат ЮНЕСКО и не обязывают к чему-либо Организацию.

Изображения из фотобанка Shutterstock в настоящей публикации не подпадают под действие лицензии CC-BY-SA и не могут использоваться или воспроизводиться без предварительного разрешения правообладателей.

Автором данной публикации является Зунке Зиск (Soenke Ziesche).

Графический дизайн, дизайн обложек и верстка: ООО «Кроматик» Cromatik Ltd.

ePub: ООО «Кроматик» (Cromatik Ltd.)

Отпечатано ЮНЕСКО в Париже (Франция).

# КРАТКОЕ СОДЕРЖАНИЕ

## ОТКРЫТЫЕ ДАННЫЕ – НЕОБХОДИМОСТЬ ДЛЯ ИИ

В целях борьбы с эпидемией, во всем мире осуществлялся обмен своевременными, актуальными и высококачественными открытыми данными. Это совместное и эффективное начинание могло бы послужить планом для лучшего и более устойчивого будущего для всех.

Наличие необработанных данных – это первый важный шаг при их обработке и преобразовании в полезную информацию. Но требуемые данные должны обладать определенными характеристиками: точность, своевременность и надежность. **Еще одним важным аспектом данных является то, что их можно легко найти, их доступность, интероперабельность и пригодность для повторного использования (FAIR) кем угодно и для чего угодно.**

Цель руководящих принципов – ознакомить государства-члены с ценностью открытых данных и указать, как данные курируются и открываются. Эти руководящие принципы определяют конкретные шаги, которые государства-члены могут предпринять для открытия своих данных в три этапа: подготовка, открытие, последующие действия.

Эти руководящие принципы являются продолжением **Рекомендации ЮНЕСКО по этическим принципам работы с искусственным интеллектом**, которая, среди прочего, включает призыв к созданию и работе с открытыми данными для ИИ. Эти руководящие принципы также будут играть решающую роль в поддержке Рекомендаций ЮНЕСКО по открытой науке, облегчая обмен данными, повышая их воспроизводимость и прозрачность, продвигая совместимость данных и стандарты их качества, поддерживая сохранение данных и долгосрочный доступ к ним.

**У 120 стран отсутствует политика в отношении открытых данных.**

**«Из-за того, что войны начинаются в умах мужчин и женщин, именно в умах мужчин и женщин должна выстраиваться система обеспечения мира и безопасности»**

ЮНЕСКО



<b>Предисловие</b> .....	5
<b>Благодарность</b> .....	6
<b>Основные положения</b> .....	7
<b>1. Данные в процессе борьбы с глобальной пандемией – пример из практики</b> .....	9
1.1 Введение .....	9
1.2 Трудности в отношении применения искусственного интеллекта .....	10
1.3 Трудности при работе с данными .....	10
1.4 Пути преодоления .....	11
1.5 Открытые данные – обзор извлеченных уроков .....	11
<b>2. Видение вопроса и перспективы</b> .....	13
2.1 Введение .....	13
2.2 Открытые данные .....	15
2.2.1 Предпосылки .....	15
2.2.2 Определение, принципы, план «пять звезд» и лицензии .....	16
2.2.3 Категории открытых данных .....	18
2.2.4 Плюсы и минусы открытых данных .....	18
2.2.5 Существующие общие инициативы и ресурсы по открытым данным .....	19
2.2.5.1 Инициативы по открытым данным, связанные с ЦУР .....	20
2.2.5.2 Государственные инициативы по открытым данным .....	20
2.2.5.3 Инициативы по открытым данным в науке .....	21
2.2.5.4 Частные инициативы по открытым данным .....	21
2.2.5.5 Инициативы по открытым данным COVID-19 .....	22
2.2.6 Индексы и барометры .....	23
2.3 Ии и открытые данные .....	23
2.3.1 Современный контекст .....	23
2.3.2 Синергия между ИИ и открытыми данными для ответа на глобальные вызовы .....	24
2.3.3 Потенциально возросшие риски для ИИ с учетом открытых данных .....	26
Другие темы .....	27
<b>3. Руководящие принципы</b> .....	28
3.1. Предисловие .....	28
3.2. Подготовка .....	28
3.3 Раскрытие данных .....	29
3.4 Сопровождение повторного использования и устойчивости .....	32
<b>Эпилог</b> .....	33
<b>4. Список литературы</b> .....	33

# Предисловие

Процессы, производственные циклы и связанные с ними исследования генерируют большие объемы данных в различных форматах и последовательностях. Эти факторы приводят к сложным проблемам при обработке данных и их использовании для генеративных систем ИИ. Поскольку мы все больше осознаем роль искусственного интеллекта, доступность и доступ к данным важны как никогда.

ЮНЕСКО проводит исследования в области ИИ, чтобы сформулировать идеи и практические решения для содействия цифровой трансформации и создания инклюзивных обществ знаний. Мы создали настоящие рекомендации, чтобы способствовать общему пониманию открытости данных, их преимуществ и затруднений при работе с ними. Наша цель состоит в том, чтобы проинформировать читателей, особенно государства-члены ЮНЕСКО, о важности открытых данных для ИИ, а также поддержать государства-члены при принятии соответствующих мер посредством ряда шагов.

В руководящих принципах описывается тип и объем экосистемы данных, а также заново формулируются концепции данных FAIR (легко находимые, доступные, совместимые и пригодные для повторного использования данные) для ИИ.

Помимо этого, в руководящих принципах перечислены двенадцать конкретных шагов, которые могут предпринять государства-члены Организации, чтобы сделать свои данные открытыми и доступными в формате FAIR. Настоящие инструкции включают три основных раздела: подготовка, открытие данных и последующие меры по повторному использованию и обеспечению устойчивости, — каждый из которых, в свою очередь, включает четыре шага.

Я надеюсь, что настоящая публикация послужит призывом к действию, побуждая государства-члены не только поддерживать открытость качественных данных, но и внедрять культуру открытости данных для поддержки приложений ИИ.

## **Тауфик Джеласси**

Помощник Генерального директора по коммуникации  
и информации ЮНЕСКО

## Благодарность

ЮНЕСКО благодарит ведущего автора г-на Зунке Зиска за то, что он взялся за разработку руководящих принципов.

ЮНЕСКО также искренне благодарит основную группу, которая руководила подготовкой настоящего документа, а именно:

Проф. Джона Шоу-Тейлора,

Проф. Марию Фасли,

Г-н Давора Орлича,

Г-н Алеша Вершича

г-н Бхану Неупане

Среди рецензентов публикации были:

Г-жа Мариэлза Оливейра, директор по партнерству и оперативной программе мониторинга, коммуникации и информации, ЮНЕСКО

Г-н Жако дю Туа, начальник Секции всеобщего доступа к информации, Сектор коммуникации и информации, ЮНЕСКО

Г-жа Дороти Гордон, председателя Программы ЮНЕСКО Информация для всех

Г-жа Эльза Эстевес, кафедра ЮНЕСКО по обществам знаний и цифровому управлению Южного национального университета и профессор Национального университета Ла-Платы, Аргентина

Г-жа Ирина Кучма, руководитель программы открытого доступа Stichting EIFL.net

ЮНЕСКО также выражает благодарность коллегам из Сектора коммуникации и информации, Сектора науки, Сектора культуры, Сектора образования, Сектора социальных и гуманитарных наук, Отдела гендерного равенства и Департамента Африки, которые входят в Целевую группу ЮНЕСКО по ИИ за их поддержку в оформлении документа.

## Основные положения

Несмотря на то, что пандемия COVID-19 стала глобальным кризисом, почти остановившим мир, она также объединила национальные и международные усилия, от правительств и политиков до академического и исследовательского сообществ и широкой общественности для решения ее проблем. Для борьбы с пандемией с самого начала необходимы не только своевременные, актуальные, качественные, но и открытые данные, чтобы их можно было бы использовать для научных исследований и для обоснования необходимых вмешательств и разработки превентивной политики.

Были переданы большие объемы открытых данных, связанных с COVID-19, что оказало значительное влияние на ситуацию, включая поддержку приложений искусственного интеллекта (ИИ) и демонстрацию потенциала использования данных для решения других глобальных проблем. Действительно, это совместное и эффективное начинание может послужить планом для сбора надежных данных, а также предложения, изложенного в «Нашей общей повестке» Генерального секретаря Организации Объединенных Наций, в котором содержится призыв к модернизации ООН, включая новые принципы работы с данными и стратегические предвидение.

Наличие необработанных данных – это первый важный шаг при их обработке и преобразовании в полезную информацию. Однако указанные данные должны обладать определенными характеристиками: точность, своевременность и надежность. Еще одним важным критерием данных является их беспрепятственная находимость, доступность, интероперабельность и возможность вторичного использования (FAIR) кем угодно и для чего угодно. Согласно Рекомендации ЮНЕСКО об открытой науке (2021b), эти данные включают «открытые исследовательские данные и, среди прочего, цифровые и аналоговые данные, как необработанные, так и обработанные, сопутствующие метаданные, а также числовые оценки, текстовые записи, изображения, звуки, протоколы, коды анализа и рабочие процессы, которые могут открыто и повторно использоваться, сохраняться и распространяться кем угодно при условии соответствующего подтверждения» (ЮНЕСКО, 2021b, стр. 9). Помимо того, что открытые данные могут быть свободно использованы, модифицированы и распространены кем угодно и для чего угодно<sup>1</sup>, они могут способствовать укреплению взаимного доверия, обеспечению повторного использования и внедрению инноваций, а также организации системы подотчетности.

Значительный объем данных об окружающей среде, промышленности, сельском хозяйстве во всем мире в настоящее время собирается с помощью автоматических процессов, включая датчики. Такие данные могут быть легко доступны, но также потенциально слишком велики, чтобы пользователи их эффективно обрабатывали или анализировали, тем не менее они могут служить входными данными для систем ИИ. Методы искусственного интеллекта и науки о данных продемонстрировали свою способность анализировать большие объемы данных, что в настоящее время иллюстрируется генеративными системами ИИ, они же помогают обнаруживать ранее неизвестные или скрытые закономерности для предоставления полезной информации в режиме реального времени. Однако многие современные системы ИИ работают на проприетарных наборах данных; при этом если они соответствуют критериям открытости, то принесут дополнительную пользу системам ИИ и снизят потенциальные риски, такие как: отсутствие справедливости, подотчетности или прозрачности.

Цель настоящих руководящих принципов – ознакомить государства-члены с ценностью открытых данных и указать, как данные курируются и открываются. Государствам-членам рекомендуется не только поддерживать открытость высококачественных данных, но и использовать технологии ИИ и способствовать наращиванию потенциала и обучению, включая инклюзивные открытые данные, а также грамотность в области ИИ.

1. <https://opendefnition.org/>

Отчет был подготовлен на основе обширного обзора литературы и консультаций с заинтересованными сторонами с последующим процессом экспертной оценки. В нем изложены конкретные шаги, которые смогут помочь государствам-членам сделать доступные им данные максимально открытыми, в три этапа: подготовка; открытие данных; и мониторинг их повторного использования и устойчивости; каждый этап, в свою очередь, состоит из четырех шагов. Этап подготовки помогает государствам-членам делать данные открытыми и включает следующие шаги: разработка политики открытых данных, сбор и высококачественных данных, развитие потенциала открытых данных и подготовка данных к использованию ИИ.

Этап по открытию данных подразумевает следующие шаги: выбор категорий данных, их открытость, закрепленная юридическим статусом, техническое открытие данных, а также создание культуры, основанной на открытых данных. Последующий этап повторного использования и обеспечения устойчивости данных состоит из следующих шагов: поддержка со стороны граждан, выход на международный уровень, привлечение в процесс искусственного интеллекта и поддержание высокого качества данных.

Указанные выше руководящие принципы являются продолжением Рекомендации ЮНЕСКО по этическим принципам работы с искусственным интеллектом, которая, среди прочего, включает призыв к открытым данным (ЮНЕСКО, 2021d). Если государства-члены ЮНЕСКО будут следовать настоящим руководящим принципам, открывать свои данные и поддерживать их устойчивость, а также создавать потенциал и культуру открытых данных и соответствующие инициативы, то государства-члены смогут использовать те же принципы для решения других глобальных, региональных и национальных проблем. Хотя руководящие принципы описывают конкретные шаги, которые помогут государствам-членам в их усилиях по созданию, открытию, предоставлению и использованию данных, — данные процессы и работа с ИИ потребуют разрешения трудностей, возникающих в ряде областей, в том числе вопросы конфиденциальности, этики и наращивания потенциала и компетенции. ЮНЕСКО будет оказывать дальнейшую поддержку государствам-членам в реализации настоящих руководящих принципов, а также предоставлять отчеты специалистов, охватывающие конкретные сферы.



# 1. Данные в процессе борьбы с глобальной пандемией – пример из практики

## 1.1 Введение

---

Пандемию COVID-19 можно охарактеризовать как «наиболее серьезный глобальный кризис в области здравоохранения со времен пандемии гриппа 1918 года» (Cascella et al., 2022). К ноябрю 2021 года в мире насчитывалось более 251 миллиона подтвержденных случаев. Пять миллионов человек скончались от COVID-19<sup>2</sup>, значительное число людей страдает от стойких симптомов заболевания (Raveendran et al., 2021). Почти все население мира пострадало от частично затянувшихся мер по борьбе с болезнью, таких как локдауны, которые имели также психологические последствия (Odrizola-González et al., 2020; Panchal et al., 2020). Более того, пандемия оказала значительное влияние на национальные системы образования (ЮНЕСКО, 2021а), а также на экономику<sup>3</sup>.

В то время как ВОЗ объявила чрезвычайную ситуацию в области общественного здравоохранения, имеющую международное значение, в связи с COVID-19 30 января 2020 г. и пандемией 11 марта 2020 г., ученые опубликовали открытое письмо и призвали к всеобщему обмену данными о SARS-CoV-2 уже 29 января 2020 года<sup>4</sup>. 16 марта 2020 г. специалисты по работе с данными призвали к созданию инфраструктуры и экосистемы данных, подходящих для борьбы с пандемиями<sup>5</sup>. 30 марта 2020 г. генеральный директор ЮНЕСКО г-жа Одри Азуле создала национальные правительства и выступила за научное сотрудничество и интеграцию программы Открытая наука в исследовательские программы государств<sup>6</sup>. В апреле 2020 года была запущена инициатива о выдаче открытых лицензий (Open COVID Pledge), призывающее организации по всему миру сделать свои патенты и авторские права бесплатными и доступными, которую подписали многие, в том числе, крупные корпорации<sup>7</sup>.

Пандемия COVID-19 также послужила предпосылкой для разработки международного нормативного инструмента Open Science (открытая наука) в форме Рекомендации ЮНЕСКО (ЮНЕСКО, 2021b). Рекомендация настоятельно призывает национальные правительства приложить усилия для расширения возможностей открытой науки и исследовательских данных, как необработанных и обработанных, а также сопутствующих метаданных, числовых оценок, текстовых записей, изображений и звуков, протоколов, кодов анализа и рабочих процессов.

Для борьбы с пандемией COVID-19, с самого начала были необходимы своевременные, актуальные и качественные данные. Настоящее тематическое исследование показывает, что в дополнение к упомянутым особенностям решающее значение имеет открытость данных. Во-вторых, действительно были переданы большие объемы открытых данных, связанных с COVID-19, и это оказало ключевое влияние. Можно провести различие в отношении того, применялись ли доступные данные людьми или приложениями ИИ. Ниже приведены некоторые приложения ИИ, трудности, а также выводы (уроки), характерные для общего движения по внедрению открытых данных.

---

2. <https://covid19.who.int/>

3. <https://www.statista.com/topics/6139/covid-19-impact-on-the-global-economy/#dossierKeyfigures>

4. <https://www.covid19dataportal.org/support-data-sharing-covid19>

5. <https://thegovlab.org/static/files/publications/ACallForActionCOVID19.pdf>

6. <https://en.unesco.org/news/unesco-mobilizes-122-countries-promote-open-science-and-reinforced-cooperation-face-covid-19>

7. <https://opencovidpledge.org>

## 1.2 Трудности в отношении применения искусственного интеллекта

---

Технологии машинного обучения ИИ<sup>8</sup> были разработаны для поддержки борьбы с пандемией в различных аспектах. Модель искусственного интеллекта на основе данных была разработана для города Валенсия в Испании группой исследователей, выигравшей конкурс XPRIZE Ответ на вызов пандемии (Pandemic Response Challenge) на 500K (Lozano et al., 2021). Вычислительная эпидемиологическая модель основана на открытых данных, направлена на прогнозирование уровня заражения COVID-19 и предписывает планы немедикаментозных вмешательств. Дополнительные системы искусственного интеллекта были применены к имеющимся данным о COVID-19 для поддержки таких областей, как диагностика, сдерживание и мониторинг, разработка и применение лекарств и вакцин, лечение а использовать ИИ для расширения прогнозирования и раннего предупреждения будущих эпидемий и пандемий<sup>10</sup>.

Системы искусственного интеллекта сыграли важную роль в некоторых областях, таких как разработка вакцин<sup>11</sup>. Однако, несмотря на большое количество диагностических и прогностических инструментов, большинство из них по разным причинам оказались неудовлетворительными<sup>12</sup>. Большинство этих проблем было связано с данными низкого качества, нестандартизированными, неправильно маркированными и часто полученными из неизвестных источников<sup>13</sup>. Более того, данные часто не были инклюзивными, например, из-за того, что меньшинства не были в них адекватно представлены, что приводило к предвзятости и перекосам. Принимая во внимание все вышесказанное, появились модели, обученные на этих данных, которые не давали точных результатов. Другая проблема заключалась в том, что часто ни данные, ни модели обучения не использовались совместно, поскольку у академических исследователей обычно мало стимулов для карьерного роста.

## 1.3 Трудности при работе с данными

---

Возникли пробелы в данных, особенно в развивающихся странах и среди некоторых групп риска и уязвимых групп населения (Milan & Tere, 2020)<sup>14</sup>, а также проблемы, связанные с отсутствием дезагрегации данных, особенно по полу<sup>15</sup>. Указанные проблемы, среди прочих, были связаны с дефицитом финансирования – с ним столкнулись более 60 процентов стран с низким уровнем дохода и уровнем дохода ниже среднего при финансировании своих данных о COVID-19 и соответствующей статистики<sup>16</sup>.

В целом, данные часто не отвечают так называемым принципам FAIR (т.е. не могут быть найдены, недоступны, не интероперабельны и не пригодны для повторного использования), в то время как аббревиатура также интерпретируется как «объединенные, готовые к работе с ИИ» (Federated, AI-Ready – FAIR), чтобы проиллюстрировать то, что данные могут быть использованы системами ИИ<sup>17</sup>. В основном нет четкого различия, предназначены ли данные при работе с ИИ для борьбы с пандемией или нет. Другими словами: взаимозаменяемы ли

---

8. Машинное обучение связано с возможностями компьютера адаптироваться к новым обстоятельствам, а также обнаруживать и экстраполировать закономерности (Russell and Norvig, 2015).

9. См. обзоры: Harrus & Wyndham (2021), Hussain et al. (2020) и Кемасуван и Кольт (2021).

10. <https://pandemichub.who.int/>

11. См., например: Ong et al. (2020).

12. См. обзоры: Heaven (2021), von Borzyskowski et al. (2021), Wynants et al. (2020) и Робертс и др. (2021).

13. См. обзоры: OECD (2020) и <https://opendatawatch.com/whats-being-said-resource/data-in-the-time-of-covid-19/>

14. Также: <https://www.npr.org/sections/goatsandsoda/2021/03/15/977455005/covid-19-data-is-missing-a-lot-of-people-and-raising-a-lot-of-questions?t=1622558115396>

15. <https://data2x.org/tracking-the-gender-impact-of-covid-19/>

16. <https://paris21.org/news-center/news/press2020-under-covid-19-worrying-stagnation-funding-despite-growing-data-demand>

17. <https://www.go-fair.org/implementation-networks/overview/vodan/>

просто открытые данные и открытые данные для искусственного интеллекта? Если да, то какие это влечет последствия?

Еще одной проблемой является баланс между открытыми данными о COVID-19 и правом на неприкосновенность частной жизни<sup>18</sup>, поскольку были случаи утечки данных пациентов с COVID-19<sup>19</sup>.

Пандемия COVID-19 также сопровождалась дезинформацией; явление, которое ЮНЕСКО (2020 г.) окрестило «дезинфодемией». Одним из аспектов этого явилось то, что скептики COVID-19 создали визуализацию данных, якобы показывающую, что ответные меры правительств на пандемию были непропорциональными и что факты и серьезность, связанная с пандемией, была либо занижена, либо завышена. Эта дезинформация часто распространялась в социальных сетях (Lee et al., 2021).

## 1.4 Пути преодоления

---

С одной стороны, пандемия COVID-19 объединила мир для решения этой глобальной проблемы. Произошли прорывы благодаря своевременному открытому обмену данными, сотрудничеству и прогнозированию, что позволило использовать инструменты диагностики, а также вакцины, разработанные в рекордно короткие сроки. С другой стороны, пандемия усугубила поляризацию не только в отношении открытого обмена данными<sup>20</sup>, но и в отношении справедливого распределения вакцин, а также недоверия к вакцинации – вопросы, выходящие за рамки настоящего доклада. В целом, положительные аспекты, изложенные выше, с точки зрения сотрудничества и эффективности, должны развиваться и служить образцом для сбора надежных данных.

Однако следует отметить, что большинство инициатив по открытым данным были ситуативными и плохо скоординированными, поскольку мир не был готов к пандемии. Таким образом, как следует из извлеченного урока, необходимо разработать нормативно-правовую базу и модели управления данными, опираясь на достаточную инфраструктуру, человеческие ресурсы и институциональные возможности для решения проблем, связанных с открытыми данными, изложенных выше, чтобы лучше подготовиться к пандемиям и другим глобальным вызовам (OECD, 2020 г.). Кроме того, необходимо дополнительно уточнить взаимосвязь между открытыми данными и ИИ, включая то, какие функции открытых данных необходимы, чтобы они были «готовы к работе с ИИ».

Как будет более подробно описано ниже, это тематическое исследование полезно, поскольку многие возможности и проблемы, кратко проиллюстрированные здесь, применимы к открытым данным в целом.

## 1.5 Открытые данные – обзор извлеченных уроков

---

- Должна существовать политика управления данными, сотрудничества и обмена данными для исследований, а также для государственных учреждений, хранящих или обрабатывающих данные, связанные со здоровьем населения, при обеспечении конфиденциальности данных путем анонимизации. Это должно включать, в частности, сотрудничество между исследователями ИИ и клиницистами.

---

18. <https://en.unesco.org/covid19/communicationinformationresponse/opensolutions>

19. <https://www.hrw.org/news/2020/12/15/personal-data-thousands-covid-19-patients-leaked-moscow>

20. См., например: Waltman et al. (2021) для сторонников обмена данными и тех, кто подчеркивает важность патентования.

- Государственным служащим, работающим с данными, которые имеют или могут стать актуальными в связи с пандемиями, может потребоваться обучение для осознания важности таких данных, а также необходимости делиться ими.
- Необходимо собрать как можно больше качественных данных. Характеристики «качественных» данных – точные, не устаревшие и исчерпывающие; таким образом, данные должны поступать из различных заслуживающих доверия источников, которые, однако, должны быть этичными, т. е. не должны включать виды данных с предвзятостью и вредоносным содержанием и собираться только с согласия граждан не нарушающими конфиденциальность способами. Данные о людях должны быть разбиты там, где это уместно, в идеале, по доходам, полу, возрасту, расе, этнической принадлежности, миграционному статусу, инвалидности и географическому местоположению (Межведомственная и экспертная группа по показателям достижения целей в области устойчивого развития, 2017 г.). Кроме того, пандемии обычно представляют собой быстро развивающиеся процессы; таким образом, постоянное обновление данных имеет важное значение.
- Такие функции данных особенно необходимы для улучшения неадекватных инструментов диагностики и прогнозирования ИИ в будущем. Кроме того, могут быть учтены данные, которые кажутся незначительными, поскольку сила систем искусственного интеллекта с машинным обучением заключается в обнаружении соответствующих закономерностей и корреляций в данных, которые люди не в состоянии распознать.
- Требуется усилие для преобразования соответствующих данных в машиночитаемый формат, что включает в себя обработку собранных данных, т. е. очистку и маркировку.
- Следует открыть широкий спектр данных, связанных с пандемиями, в соответствии с принципами FAIR, которые являются легко находимыми, доступными, совместимыми и многоразовыми, при этом аббревиатура также интерпретируется как «объединенные, готовые к работе с ИИ».
- Целевая аудитория открытых данных, связанных с пандемией, включает в себя научные и академические круги, лиц, принимающих решения в правительствах, частный сектор для разработки соответствующих продуктов, а также широкую общественность, которая должна быть проинформирована об имеющихся данных. Общественность играет важную роль в движении за открытые данные и была вовлечена в пандемию COVID-19 через инициативы под общим названием «наука о гражданах»<sup>21</sup>.
- Инициативы по открытым данным, связанные с пандемией, должны быть институционализированы, а не формироваться ad hoc (от случая к случаю), и, таким образом, должны быть введены в действие для обеспечения готовности к дальнейшим пандемиям. Такие инициативы должны носить инклюзивный характер и объединять различных производителей и пользователей данных.
- Кроме того, полезное использование данных, связанных с пандемией, для методов машинного обучения ИИ должно регулироваться с целью предотвращения неправомерного использования для разработки искусственных пандемий, т. е. биологического оружия, с помощью систем ИИ.

21. См, например: <https://citizenscience.org/covid-19/>

## 2. Видение вопроса и перспективы

### 2.1 Введение

---

Мир сталкивается с несколькими взаимосвязанными кризисами. Помимо продолжающейся пандемии COVID-19 и разрушительных бедствий, ухудшения состояния окружающей среды и утраты биоразнообразия, существуют различные другие природные и антропогенные опасности, такие как техногенные, в том числе ядерные, аварии, а также угрозы космического базирования, которые создают усугубляющие, каскадные или даже системные риски бедствий.

В то же время в мире преобладают неравенство, бедность и голод, которые усугубляются вышеуказанными кризисами и опасностями, в частности, чрезвычайной климатической ситуацией.

Для обеспечения устойчивого развития важно постоянно наблюдать за миром, а также измерять и проверять прогресс. Чтобы постоянно наблюдать за миром, а также измерять и проверять прогресс, необходимы данные о мире.

Это было проиллюстрировано приведенным выше примером: прогресс в борьбе с пандемией COVID-19 был значительно облегчен благодаря наличию данных. Данные способствовали интерпретации сценариев, разработке и последующей корректировке политики, а также эффективным нефармацевтических, медицинских и вакцинных вмешательств. Более того, данные позволили тщательно изучить политику правительств, что привело к повышению прозрачности и подотчетности.

Если бы соответствующие данные были общедоступны, можно было бы ожидать аналогичного эффекта, что, в конечном итоге, поможет справиться с глобальными кризисами. Общества находятся в ключевом моменте времени, когда мир объединился для преодоления одного кризиса, пандемии COVID-19, и должен справиться с различными аспектами поляризации, чтобы воспользоваться еще одной исторической возможностью путем передачи и распространения передового опыта, применявшегося во время пандемии, особенно в сфере обмена открытыми данными, при поиске ответов и на другие вызовы.

Доступность необработанных данных — это только первый шаг. Второй шаг включает в себя очистку данных и отделение необъективных, несогласованных и вредоносных данных, которые нельзя использовать по этическим соображениям. На третьем этапе данные преобразуются в полезную информацию. Кроме того, как указано в приведенном выше тематическом исследовании, данные также могут подгружаться в системы ИИ, способного выявлять неочевидные закономерности.

Благодаря значительным разработкам в области датчиков и так называемого Интернета вещей (Международный союз телекоммуникаций и систем CISCO, 2016 г.), замеряется все больше показателей о мире, процессах в природе, а также деятельности человека, что создает большой объем данных, который слишком велик для обработки или анализа без помощи автоматических систем, в том числе систем обработки данных и ИИ.

В то же время многие характеристики мира не поддаются измерению, т.к. либо датчики для этой конкретной функции не разработаны, либо они существуют, но не применяются повсеместно, либо есть другие проблемы, например, конфиденциальность данных. В этом заключается трудность — никто не должен выпадать из поля зрения. Кроме того, большая часть собранных данных не всем доступна, по целому ряду причин<sup>22</sup>.

---

22. См. также «Общесистемную дорожную карту по обновлению данных и статистики ООН», разработанную Комитетом Главных статистиков системы ООН при участии Комитета высшего уровня по программам ООН: <https://unstats.un.org/unsd/unsystem/documents/Roadmap-Innovating%20UN%20Data%20and%20Statistic.pdf>

Представьте себе сценарий, при котором данные по каждому 231 показателю ЦУР (цель устойчивого развития), которые отслеживают достижение 169 задач по 17 ЦУР, доступны в открытой форме и в режиме реального времени, чтобы каждый мог проверить их в любое время через онлайн-платформу<sup>23</sup>. Это был бы отличный пример доступности соответствующих данных.

Очевидно, что для решения глобальных проблем требуемые данные и их источники должны обладать определенными характеристиками, такими как точность, своевременность, надежность, полнота и инклюзивность. Тем не менее, этих характеристик самих по себе может быть недостаточно для решения задач. Еще одним важным аспектом данных является то, что они должны быть легко находимыми (F), доступными (A), интероперабельными (I) и могли бы быть повторно использованы (R) кем угодно для любых целей. Такие данные называются данными, соответствующими критериям FAIR, в то время как «открытые» данные отвечают характеристике «свободно используются, модифицируются и распространяются кем угодно для любых целей»<sup>24</sup>.

Открытые данные находятся в центре внимания настоящего отчета, поскольку они рассматриваются как необходимое условие для обоснованных планов, решений и конкретных действий. Таким образом, в отчете утверждается, что государства-члены должны обмениваться данными и идеями, обеспечивая прозрачность и подотчетность, а также возможности для всех использовать данные.

Основная цель настоящих руководящих принципов – способствовать всеобщему доступу к информации и знаниям с помощью открытых решений для инклюзивной цифровой трансформации и развития искусственного интеллекта. В то же время, существуют обоснованные опасения по поводу безопасности ИИ, которыми нельзя пренебрегать, но которые из-за своей сложности выходят за рамки настоящего руководства (например, Бостром, 2014; Ямпольский, 2018). Кроме того, отчет направлен на повышение осведомленности о технологиях и инструментах, которые позволяют осуществлять информированное планирование и принятие решений в отношении различных видов деятельности, а также осуществлять мониторинг и независимые проверки развития процесса.

Здесь утверждается, что главное преимущество открытых данных среди многих заключается в том, что они наглядны и укрепляют доверие и солидарность. Например, решение проблемы изменения климата требует совместных усилий многих заинтересованных сторон, в том числе граждан из разных стран мира. Предварительным условием является то, что все заинтересованные стороны не только имеют доступ, но и доверяют информации, связанной с изменением климата, которая скомпрометирована угрозой фальшивых новостей. Тем не менее, если правительства не продемонстрируют полную прозрачность в отношении информации, связанной с климатом в своих странах<sup>25</sup> то не будут достигнуты ни доверие, ни заинтересованность, ни вовлеченность, ни ответственность сторон. Следовательно, потенциал данных не будет использован в полной мере всеми заинтересованными сторонами; что является императивом, учитывая огромные проблемы и последствия изменения климата (например, Ratcliffe & Tuzeneu, 2019; Bodor et al., 2020).

Цель настоящих руководящих принципов заключается в том, чтобы помочь государствам-членам ЮНЕСКО не только в измерении этих аспектов мира, но и в предоставлении открытого доступа к данным, которые могут способствовать выработке общедоступных решений (не забывая при этом о правах человека на неприкосновенность частной жизни), что-

23. Хотя с этими же целями были разработаны платформы (<https://unstats-undesa.opendata.arcgis.com/>, <https://sdg.tracking-progress.org/>, <https://unstats.un.org/sdgs/dataportal>), они ограничены значительными пробелами в данных и временными задержками, поскольку страны не имеют возможности собирать косвенные данные в режиме реального времени (например, Sachs et al., 2021).

24. <https://opendefinition.org/>

25. Дополнительным вызовом является так называемый «побочный эффект», когда действия одной страны оказывают положительное или отрицательное влияние на способность других стран достичь ЦУР, особенно в решении вопросов, связанных с климатом. См. <https://www.unsdsn.org/spillover-effects> и <https://dashboards.sdgindex.org/map/spillovers>.

бы их можно было обрабатывать и анализировать в дальнейшем на благо общества любым лицом, в том числе, с использованием систем искусственного интеллекта.

Методология, использованная для настоящего руководства, состояла из тщательного обзора литературы, а также консультаций с заинтересованными сторонами, с последующей экспертной оценкой.

Оставшаяся часть руководства структурирована таким образом: в следующем разделе более подробно представлены открытые данные. Затем следует раздел об искусственном интеллекте и о том, почему открытые данные особенно важны для ИИ. Затем излагаются смежные темы, которые не могут быть рассмотрены в данном отчете, но, тем не менее, относятся к данной тематике. Во второй части отчета представлены действия и шаги, которые необходимы в отношении открытых данных.

## 2.2 Открытые данные

### 2.2.1 Предпосылки

---

Технологические разработки и достижения последних десятилетий, впоследствии названные «Четвертой промышленной революцией», привели к беспрецедентному увеличению объема и сложности генерируемых данных, а также их сбора в большие объемы/наборы данных (также называемые «большими данными»). Подсчитано, что в 2020 году цифровая вселенная содержала 44 зеттабайта<sup>26</sup> данных. Тем не менее, на фоне постоянно растущего потока данных возникает ряд проблем, связанных с наличием надежных и своевременных данных для разработки, таких как:

- Данные не были собраны.
- Данные были собраны, но не разглашаются.
- Данные были собраны, но разрознены без ведома других сторон.
- Данные были собраны, но они устарели, ненадежны или неточны.
- Данные были собраны, но в несовместимых форматах, включая, например, печатный формат.
- Данные были собраны, но не помечены соответствующей лицензией.

Огромное количество данных открывает новые возможности, поскольку такие данные, например, могут быть переданы в системы ИИ для анализа и дальнейшего использования. Но даже данные в больших объемах могут не обеспечить адекватного отображения картины мира, как хотелось бы, поскольку многие соответствующие данные не собираются, а многие собранные данные недоступны.

Стало очевидным, что данные должны обладать определенными свойствами, которые делают их полезными для решения глобальных проблем, изложенных во введении, таких как: их наличие, быстрый поиск, открытый доступ к ним, функциональная совместимость, модифицируемость и возможность повторного использования, — которые будут описаны ниже. Поэтому к заинтересованным сторонам, особенно правительствам, а также научному сообществу и частному сектору, был обращен призыв предоставлять открытые данные. В 2012 году изобретатель всемирной паутины Бернерс-Ли в своей статье «Необработанные данные, немедленно!» выразил просьбу и обозначил возможности, которые появляются из использования дополнительных данных: «инновации, прозрачность, подотчетность, более

---

26. Один зеттабайт = 1021. См. <https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/>

эффективная управляемость и экономический рост» (Бернерс-Ли, 2012). Кроме того, бывший Генеральный секретарь ООН Пан Ги Мун признал актуальность данных и заказал отчет под названием «Мир, который имеет значение», в котором также подчеркивалась ценность открытых данных и рекомендовалось партнерство между международными организациями, правительствами, частными компаниями и организациями гражданского общества для обмена данными и совместного мониторинга (Независимая консультативная группа экспертов по информационной революции в целях устойчивого развития, 2014 г.).

В настоящее время возникает третья волна использования открытых данных, характеризующаяся сосредоточением внимания на эффективном повторном использовании данных<sup>27</sup>, как это подчеркивается в Рекомендациях ЮНЕСКО по открытой науке (ЮНЕСКО, 2021b). Предлагаемый в настоящее время Закон ЕС об управлении данными также поддерживает расширенное повторное использование данных государственного сектора<sup>28</sup>. Другими передовыми практиками являются, например, Альянс больших данных для миграции, который организовал виртуальные семинары по проблемам для Западной Африки, в которых участвовали заинтересованные стороны в правительстве, международных организациях, исследованиях, гражданском обществе и государственном секторе и определяли совместные подходы к инновациям данных (Альянс больших данных для миграции, 2021 г.), или Директива ЕС INSPIRE, которая привела к унифицированным моделям данных и спецификациям для 34 тем пространственных данных<sup>29</sup>.

## 2.2.2 Определение, принципы, план «пять звезд» и лицензии

В то время как Рекомендации по открытой науке не содержат определения открытым данным, но уточняют их содержание и условия использования (ЮНЕСКО, 2021b), определение открытых данных, данное Фондом открытых знаний, является наиболее точным и гласит: «Открытые данные и контент могут свободно использоваться, изменяться и распространяться кем угодно для любых целей»<sup>30</sup>. Поскольку основное внимание в настоящем руководстве уделяется открытым правительственным данным (OGD), которые представляют собой «философию – и все чаще набор политик, – которая способствует прозрачности, подотчетности и созданию стоимости путем обеспечения доступности государственных данных для всех»<sup>31</sup>, имеет смысл утверждать, что правительства не только собирают данные и разрабатывают нормы для обмена данными, но также выступают в качестве поставщиков данных.

Важным дополнением к этому определению являются шесть принципов Хартии открытых данных, которые представляют собой «согласованный на глобальном уровне набор желаемых правил публикации данных»: 1) открытость по умолчанию, 2) своевременность и полнота, 3) доступность и удобство использования, 4) сопоставимость и совместимость, 5) улучшение управления и вовлечение граждан. 6) инклюзивное развитие и инновации<sup>32</sup>.

- Еще один важный набор принципов, называемый FAIR, уже упоминался. Название расшифровывается, как находимые (Findable), доступные (Accessible), функционально совместимые (Interoperable) и повторно используемые (Reusable) данные. Руководство по внедрению принципов FAIR изложено в так называемых Принципах FAIRфикации<sup>33</sup>:

27. Первая волна сосредоточилась на регулировании и законодательстве, в то время как вторая волна выступала за открытые данные по умолчанию, но без конкретной цели. См. <https://opendatapolicylab.org/images/odpl/third-wave-of-opendata.pdf>

28. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020PC0767&from=EN>

29. <https://inspire.ec.europa.eu/inspire-directive/2>

30. <https://opendefinition.org/>

31. <https://www.oecd.org/gov/digital-government/open-government-data.html>

32. <https://opendatacharter.net/principles/>

33. <https://www.go-fair.org/fair-principles/>



- Для обеспечения возможности как ручного, так и машинного поиска, важно присвоить данным глобальные уникальные и постоянные идентификаторы, разработать для них подробные метаданные и обеспечить их наличие в доступном для поиска ресурсе.
- Для доступности крайне важно, чтобы данные можно было получить по их идентификатору через стандартизированный протокол связи, который является открытым, бесплатным и универсально применимым и может включать процедуру аутентификации и авторизации.
- Для функциональной совместимости важно, чтобы данные были представлены на формальном, доступном, совместно используемом и широко применимом языке и были интегрированы с другими данными посредством квалифицированных ссылок.
- Для повторного использования важно, чтобы данные были хорошо описаны метаданными и соответствовали стандартам сообщества, а также чтобы была предоставлена простая и доступная лицензия на использование данных, а также информация о происхождении данных, методе их сбора и их обслуживании.

Набор принципов FAIR был дополнен принципами CARE для управления данными коренных народов. Аббревиатура CARE расшифровывается, как коллективная выгода (Collective benefit), полномочия по контролю (Authority to control), ответственность (Responsibility) и этика (Ethics)<sup>34</sup>. Использование принципов FAIR и CARE является краеугольным камнем всех действий по управлению данными, направленных на установление стандартов обмена данными и их этического использования.

Что касается технического открытия данных, Бернерс-Ли разработал «пятизвездочный» план открытия данных<sup>35</sup>:

- Одна звездочка: сделать данные доступными онлайн в любом формате;
- Две звездочки: предоставлять данные в структурированном формате (например, MS Excel вместо отсканированной таблицы);
- Три звездочки: предоставлять данные в непатентованном открытом формате (например, CSV вместо MS Excel);
- Четыре звездочки: использовать URI для обозначения объектов<sup>36</sup>, чтобы другие пользователи могли указывать на данные;
- Пять звездочек: связать данные с другими данными, чтобы обеспечить контекст.

Существует различие между технически открытыми и юридически открытыми данными. С юридической стороны необходимо указать, как данные могут быть (повторно) использованы и требуется ли указание авторства.

Существует ряд лицензий на открытые данные, ярким примером которых являются лицензии Creative Commons<sup>37</sup>. Относительно важного вопроса о том, можно ли и в какой степени использовать открытые данные для обучения систем ИИ, ведутся постоянные споры<sup>38</sup>.

34. <https://www.gida-global.org/care>

35. <https://5stardata.info/en/> и <https://www.w3.org/DesignIssues/LinkedData.html>

36. URI означает «универсальный идентификатор ресурса», который универсально идентифицирует физический или логический ресурс физического мира. Примерами ресурсов являются объекты реального мира, включая людей, местоположения и концепции, или информационные ресурсы, включая веб-страницы, документы и книги.

37. <https://creativecommons.org/about/>

38. <https://creativecommons.org/2021/03/04/should-cc-licensed-content-be-used-to-train-ai-itdepends/>

## 2.2.3 Категории открытых данных

---

В дополнение к вышеприведенному «пятизвездочному» плану открытые данные также можно классифицировать в соответствии с другими критериями, которые проиллюстрированы ниже в контексте открытых данных, связанных с COVID-19:

- Типы данных: преобладают числовые данные (с геотегами), но для исследований COVID-19 также важны и другие типы данных, поэтому желательно делиться ими, например изображения (КТ и рентген) и звуки (дыхание и кашель) (Shuja et al., 2021)<sup>39</sup>.
- Источники данных. Основными источниками открытых данных, связанных с COVID-19, являются открытые правительственные данные, а также данные научных кругов и исследований. Хотя это вряд ли имело место в случае пандемии COVID-19, в будущем желательно, чтобы и частный сектор открывал соответствующие данные.
- Использование данных: Предполагаемое использование данных зависит от контекста и спецификаций конкретного проекта, которые могут быть разными. Для пандемии COVID-19 требовались данные, которые поддерживают следующие два варианта использования: 1) чтобы понять, как распространяется коронавирус, таким образом, чтобы информировать о мерах системы здравоохранения по его сдерживанию, и 2) для борьбы с вирусом, чтобы обеспечить более эффективную разработку вакцины.
- Продукты данных: различные форматы могут определять способ совместного использования данных, например банки данных, информационные панели данных, которые обеспечивают взаимодействие или визуализацию данных.
- Целевые аудитории: это относится к субъектам, которые, как ожидается, будут использовать данные, которые были проанализированы ОЭСР и GovLab (2021 г.) а именно: обычные люди (71%), гражданское общество (68%), государственные органы (41%), исследовательские/академические круги (38%) и частный сектор (22%).

## 2.2.4 Плюсы и минусы открытых данных

---

Есть аргументы за и против открытых данных. Например, одним из аргументов в пользу<sup>40</sup> открытых данных является то, что в любом случае не может быть авторского права на фактологические данные и что каждый должен иметь право доступа к ним. Кроме того, отличительной чертой демократии должно быть то, что деятельность государственных органов является прозрачной благодаря открытым данным, что также способствует укреплению доверия к государству. Кроме того, возможность повторно использовать, перекомпоновывать и комбинировать их, а также потенциально извлекать из них новые (научные) идеи позволяет вовлекать граждан, а также создавать новые инновационные услуги и продукты, тем самым способствуя созданию социальной, экономической и экологической ценности<sup>41</sup>. Не только граждане могут создавать новые знания из открытых данных, но и системы ИИ, которые, как утверждается здесь, могут способствовать преодолению мировых кризисов.

Противники предоставления данных в открытый доступ утверждают, например, что данные могут нарушать конфиденциальность заинтересованных лиц, которые вправе самостоятельно решать, какие активно или пассивно собранные данные о них раскрывать. В некоторых случаях это также может нарушать авторские права и/или права на интеллектуальную собственность, если данные собираются и используются без согласия, указания авторства или выплаты вознаграждения. Кроме того, во многих случаях сбор, очистка и распространение

---

39. Другие инициативы предлагали открытый доступ к научным статьям, что не является предметом данного тематического исследования. См. например: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-onnovel-coronavirus-2019-ncov> и <https://www.semanticscholar.org/cord19>

40. Например, см. <https://okfn.org/opendata/why-open-data/>, <https://opendatahandbook.org/guide/en/whyopen-data/>

41. Например, см. <http://opendatatoolkit.worldbank.org/en/essentials.html#uses>

данных является трудоемким и дорогостоящим процессом, который необходимо оплачивать, и от этих услуг будут отказываться, если данные открыто распространяются. Кроме того, наборы данных могут содержать систематическую ошибку данных, возникающую на различных этапах, в том числе во время составления отчетов человеком, отбора данных, маркировки и классификации данных, вплоть до интерпретации результатов моделирования. Существует также опасение, что данные могут быть использованы не по назначению со злым умыслом, особенно с учетом того, что многие новые технологии, включая ИИ, могут иметь двойное назначение. В связи с этим Бостром (2011, стр. 3) определил «опасность данных» следующим образом: «Конкретные данные, такие как генетическая последовательность смертельного патогена или проект создания термоядерного оружия, если они будут распространяться, создают риски».

## 2.2.5 Существующие общие инициативы и ресурсы по открытым данным

---

За последние годы было создано несколько инициатив и ресурсов в области открытых данных, которые будут кратко представлены в хронологическом порядке и по темам:

В 2012 г. Бернерс-Ли и Найджел Шадболт основали Институт открытых данных для создания открытой и надежной экосистемы данных<sup>42</sup>.

В 2013 г. Департамент ООН по экономическим и социальным вопросам опубликовал «Руководство по открытым правительственным данным для вовлечения граждан»<sup>43</sup>.

Также в 2013 году лидеры G8 одобрили Хартию открытых данных G8, призывающую к тому, чтобы открытость данных стимулировала прозрачность, инновации и подотчетность. В 2015 году на полях Генеральной Ассамблеи ООН была запущена усовершенствованная и более инклюзивная версия Международной хартии открытых данных (Хартия открытых данных, 2015).

Кроме того, в 2013 г. был создан Альянс по Исследовательским Данным, который представляет собой общественную инициативу Европейской комиссии, Национального научного фонда США, Национального института стандартов и технологий США, а также Департамента инноваций правительства Австралии<sup>44</sup>.

В 2014 году была создана глобальная сеть «Открытые данные для развития» (OD4D). Ее задачей стало продвижения экосистем открытых данных в развивающихся странах для поддержки социальных изменений<sup>45</sup>.

В 2015 году Рабочая группа по борьбе с коррупцией (ACWG) «Большой двадцатки» опубликовала антикоррупционные принципы открытых данных, которые показывают роль открытых данных как инструмента обеспечения прозрачности, подотчетности и доступа к информации<sup>46</sup>.

В 2016 году в журнале Scientific Data была опубликована статья, описывающая важность того, чтобы данные можно было легко находить, чтобы они были доступными, функционально совместимыми и могли повторно использоваться (Wilkinson et al., 2016). Лидеры G20 одобрили эти так называемые принципы FAIR, подробно описанные выше, на своем саммите в том же году.

---

42. <https://theodi.org/>

43. <https://publicadministration.un.org/en/ogd>

44. <https://www.rd-alliance.org/about-rda>

45. <https://www.od4d.net/>

46. <http://www.g20.utoronto.ca/2015/G20-Anti-Corruption-Open-Data-Principles.pdf>

В 2018 г. Статистическая комиссия Департамента ООН по экономическим и социальным вопросам (ДЭСВ) учредила «Рабочую группу по открытым данным», занимающуюся выработкой принципов, рекомендаций и поддержкой внедрения открытых данных в государствах-членах<sup>47</sup>.

В 2020 году Генеральный секретарь ООН представил «Дорожную карту цифрового сотрудничества» для решения проблем, связанных с новыми технологиями, на пути к более безопасному и справедливому цифровому миру (Генеральная Ассамблея ООН, 2020 г.).

Также в 2020 г. Координационный совет руководителей системы ООН (2020 г.) утвердил «Общесистемную дорожную карту по обновлению данных и статистики ООН», целью которой является «способствовать улучшению мира посредством своевременных, достоверных данных и статистики для всех».

В 2021 г. 193 государства-члена ЮНЕСКО приняли «Рекомендацию по этике искусственного интеллекта», которая стала важной вехой не только потому, что это был первый глобальный нормативный документ по этике искусственного интеллекта, но и призвала государства-члены продвигать открытые данные (ЮНЕСКО, 2021d).

### 2.2.5.1 Инициативы по открытым данным, связанные с ЦУР

---

Глобальная система показателей для целей устойчивого развития (ЦУР) включает 17 целей, 169 связанных задач и 231 показатель, которые в значительной степени зависят от их реализации и мониторинга данных. Проблема заключается в том, что многие из необходимых данных в настоящее время не собираются странами на регулярной основе, что влияет на достоверность 95 показателей по состоянию на февраль 2022 года<sup>48</sup>. Кроме того, Sachs et al. (2021) указали на значительные пробелы в данных, особенно по ЦУР 4 (Качественное образование), 5 (Гендерное равенство), 12 (Ответственное потребление и производство), 13 (Борьба с изменением климата) и ЦУР (Жизнь под водой), и призвали к улучшению статистического потенциала в соответствующих странах. Также утверждается (например, Петров и др., 2016), что достижению целей устойчивого развития способствуют не только такие показатели, как доступность, точность, своевременность и разукрупнение данных<sup>49</sup>, но также и открытость данных. В связи с этим Статистическая комиссия ДЭСВ ООН создала «Открытый центр данных по ЦУР»<sup>50</sup>. Кроме того, Сеть решений ООН в области устойчивого развития поддерживает платформу для мониторинга ЦУР<sup>51</sup>.

### 2.2.5.2 Государственные инициативы по открытым данным

---

Правительства публикуют большие объемы данных, которыми в прошлом практически не делились, не говоря уже о том, что ранее было невозможно получить их в функционально совместимом формате и использовать для других целей. Уже существует множество инициатив национальных правительств, межправительственных организаций, субнациональных регионов и муниципалитетов в области открытых данных<sup>52</sup>. Партнерство «Открытое пра-

---

47. <https://unstats.un.org/open-data/>

48. <https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification/>

49. Хотя это и не является основным предметом исследования данного документа, разукрупнение данных имеет большое значение для достижения целей устойчивого развития. Межведомственная и экспертная группа по показателям достижения целей в области устойчивого развития (2017 г.) призывает к тому, чтобы показатели ЦУР «разбивались, где это уместно, по доходу, полу, возрасту, расе, этнической принадлежности, миграционному статусу, инвалидности и географическому положению или другим характеристикам».

50. <https://unstats-undesa.opendata.arcgis.com/> и также <https://sdg.tracking-progress.org/>

51. <https://www.unsdnsn.org/sdg-index-and-monitoring>

52. См. по ссылке (динамичный) обзор: [https://en.wikipedia.org/wiki/List\\_of\\_open\\_government\\_data\\_sites](https://en.wikipedia.org/wiki/List_of_open_government_data_sites)

вительство» – это инициатива, которая поддерживает страны, в том числе муниципалитеты, в обеспечении прозрачности, включая предоставление открытых данных<sup>53</sup>. Существуют также региональные инициативы, такие как openAFRICA – «крупнейшая в Африке платформа открытых данных, управляемая добровольцами»<sup>54</sup>, Азиатское партнерство по открытым данным<sup>55</sup>, Латиноамериканская инициатива по открытым данным<sup>56</sup> и официальный портал европейских данных<sup>57</sup>. В 2021 году завершился второй этап Инициативы G20 по пробелам в данных (МВФ и Совет Финансовой Стабильности, 2021).

### 2.2.5.3 Инициативы по открытым данным в науке

---

Научная деятельность производит большие объемы данных. Для развития науки, а также в интересах прогресса эти данные должны быть доступны для совместного использования, особенно если они получены за счет государственного финансирования. При этом не стоит забывать о проблеме возможной опасности данных, упомянутой выше. Понимание важности открытых данных в области науки и создание инициатив в этой области не новы. Например, Комитет по данным Международного совета по науке был создан еще в 1966 году<sup>58</sup>. Несмотря на то, что было подписано Соглашение об обмене научными данными и работают различные платформы<sup>59</sup>, получить доступ к ним трудно из-за большого объема данных, а также из-за неадекватного управления платформами.

Эта проблема решается в Руководящих принципах FAIR, инициативе по улучшению управления данными и рационального использования, с упором на автоматический поиска данных (Wilkinson et al., 2016). Еще одной вехой стала Рекомендация ЮНЕСКО об открытой науке, принятая Генеральной конференцией ЮНЕСКО в ноябре 2021 г. (ЮНЕСКО, 2021b).

### 2.2.5.4 Частные инициативы по открытым данным

---

Корпорации также собирают и владеют большими объемами данных, как правило, с единственной целью увеличить свои доходы. Однако значительная часть этих данных имеет отношение к устойчивому развитию. Раньше компании отказывались делиться своими данными, однако теперь появилась возможность делиться данными и не остаться в убытке: несколько корпораций присоединились к инициативе UN Global Pulse под названием «Благотворительность в области данных». Они взяли на себя обязательство делиться своими данными на благо общества<sup>60</sup>. Многие другие корпорации сотрудничают с Учебным и исследовательским институтом ООН в рамках инициативы Программы оперативных спутниковых приложений ООН, чтобы обмениваться спутниковыми снимками в гуманитарных целях<sup>61</sup>. Хотя эти данные теперь можно использовать для реализации и мониторинга устойчивого развития, у корпораций также есть множество стимулов для предоставления доступа к своим данным, таких как взаимность обмена данными; проведение ими исследований, набора персонала и доступ к новым идеям; улучшение репутации и связей с общественностью; увеличение дохода; соответствие нормативным требованиям; а также ответственность и корпоративная благотворительность (Klein & Verhulst, 2017). В предлагаемом в настоящее время Законе ЕС

53. <https://www.opengovpartnership.org/>

54. <https://africaopendata.org/>

55. <https://www.linkedin.com/company/aodp/>

56. <https://idatosabiertos.org/en/>

57. <https://data.europa.eu/en>

58. <https://codata.org/>

59. Например, см.: <https://opendatascience.com/>

60. <https://www.unglobalpulse.org/2011/09/data-philanthropy-public-private-sector-data-sharing-for-global-resilience/>

61. <https://www.unitar.org/sustainable-development-goals/united-nations-satellite-centre-UNOSAT>

об управлении данными вводится аналогичный механизм, называемый «альтруизмом в области данных», который поощряет и предоставляет возможности отдельным лицам и компаниям делиться данными для общественного блага<sup>62</sup>.

## 2.2.5.5 Инициативы по открытым данным COVID-19

---

На призывы открыть данные, связанные с COVID-19, ответили различные платформы по обмену данными, от открытых государственных данных до частных проектов<sup>63</sup>. Например:

- Информационная панель COVID-19 Центра системных наук и инженерии Университета Джона Хопкинса показывает количество случаев заболевания, смертей и вакцинированных по странам, и является ресурсом для общественности, а также для политиков<sup>64</sup>.
- Worldometer предоставляет глобальную статистику COVID-19 в режиме реального времени путем ручного анализа, проверки и агрегирования данных из более чем 5.000 источников в режиме реального времени<sup>65</sup>.
- Европейская платформа данных COVID-19 облегчает обмен данными и их анализ для ускорения исследований по борьбе с коронавирусом<sup>66</sup>.
- Репозиторий открытых данных GitHub COVID-19 пытается стать крупнейшей эпидемиологической базой данных по COVID-19<sup>67</sup>.
- На платформе GISAID размещено более 450.000 геномов вируса<sup>68</sup>, но ее критикуют за то, что пользователи не могут повторно публиковать геномы без разрешения (Van Noorden, 2021).
- Nextstrain – это проект с открытым исходным кодом для отслеживания эволюции патогена в режиме реального времени<sup>69</sup>.
- Global.health – это хранилище данных и платформа визуализации, которая обеспечивает открытый доступ к анонимным эпидемиологическим данным в режиме реального времени<sup>70</sup>.
- Vivli – это платформа, которая обменивается анонимными данными участников завершённых клинических испытаний<sup>71</sup>.
- Coronavirus Watch – это платформа, которая предоставляет информацию о новых тенденциях, связанных с COVID-19, во всем мире в виде карт, визуализаций статуса заболевания и информационных панелей<sup>72</sup>.
- Универсальный шлюз ресурсов Covid-19 (CURE) поддерживается ЮНЕСКО и представляет собой портал для ресурсов с открытым доступом, публикующих последние статьи, информационные панели данных, а также образовательные и обучающие материалы по COVID-19<sup>73</sup>.

62. <https://www.consilium.europa.eu/en/press/press-releases/2021/10/01/eu-looks-to-make-datasharing-easier-council-agrees-position-on-data-governance-act>

63. См. обзор: Alamo et al. (2020).

64. <https://coronavirus.jhu.edu/map.html>

65. <https://www.worldometers.info/coronavirus/>

66. <https://www.covid19dataportal.org/>

67. <https://github.com/GoogleCloudPlatform/covid-19-open-data>

68. <https://www.gisaid.org/>

69. <https://nextstrain.org/>

70. <https://global.health/>

71. <https://search.vivli.org/>

72. <https://www.unesco.org/en/articles/corona-virus-media-watch-launched-unescos-internationalresearch-centre-artificial-intelligence>

## 2.2.6 Индексы и барометры

---

Кроме того, существует несколько индексов и барометров, которые оценивают и ранжируют страны в соответствии с их усилиями по открытию данных, например следующие:

- Барометр открытых данных<sup>74</sup>,
- Глобальный барометр данных<sup>75</sup>,
- Инвентаризация открытых данных / наблюдение<sup>76</sup>,
- Глобальный индекс открытых данных<sup>77</sup>,
- Открытые правительственные данные ОЭСР / Индекс OURdata<sup>78</sup>,
- Зрелость открытых данных в Европе<sup>79</sup>.

## 2.3 ИИ и открытые данные

### 2.3.1 Современный контекст

---

Хотя открытые данные могут использоваться во многих областях, они критически важны для искусственного интеллекта, о чем речь пойдет далее в этой части.

Существуют различные определения ИИ из-за того, что понятие «интеллект» не является однозначным. Легг и Хаттер (2007, стр. 12) дают следующее общее определение: «Интеллект измеряет способность субъекта достигать целей в самых разных условиях». Если под «субъектом» в этом определении понимать человека или животное, то это обычный интеллект, а если «субъект» — это машина, то мы имеем искусственный интеллект. Область ИИ можно разделить на взаимодействующие подразделы, из которых Рассел и Норвиг (2015) перечисляют следующие: обработка естественного языка, накопление знаний, автоматизированное формирование рассуждений, машинное обучение, компьютерное зрение и робототехника.

Недавний бум ИИ в значительной степени основан на значительных достижениях в области машинного (и особенно глубокого) обучения, которое требует больших объемов данных, в то время как некоторые другие подразделы не требуют такого большого количества данных. Генеративный ИИ — это общий термин, обозначающий очень успешные в настоящее время алгоритмы машинного обучения, которые также привлекли большое внимание средств массовой информации и которые генерируют искусственный цифровой контент, такой как текст, изображения, аудио- и видеоконтент, на основе больших объемов обучающих данных. Вместо простой обработки существующих данных генеративный ИИ позволяет создавать программы, которые могут генерировать новый контент, например, изображения, видео, музыку или текст. Глубокое обучение — это тип машинного обучения, в котором используются нейронные сети для понимания закономерностей и создания новых данных. Приложения, созданные на основе генеративного ИИ, могут создавать реалистичные изображения или видео, генерировать ответы на обычном разговорном языке, придумывать дизайн новых товаров и сочинять музыку. Существуют опасения по поводу возможного неправомерного использования генеративного ИИ, например, для создания ложных новостей или дипфейков. Следовательно, важно тщательно рассмотреть этические последствия применения гене-

73. <https://www.goap.info/cure>

74. <https://opendatabarometer.org>

75. <https://globaldatabarometer.org/>

76. <https://opendatawatch.com/>

77. <http://index.okfn.org/>

78. <https://www.oecd.org/gov/digital-government/open-government-data.html>

ративного ИИ и разработать правила его ответственного использования. Ведь качество этого контента становится настолько высоким, что люди не могут различить, был ли он создан машиной или человеком, что вызывает опасения и споры.

Таким образом, прогресс в машинном обучении, с одной стороны, основан на развитии инновационных технологий, а с другой – на постоянно растущем избытке доступных данных, в то время как их недостаток в предыдущие десятилетия мешал развитию ИИ.

Однако, когда мы обсуждаем ИИ и данные, необходимо отметить проблемы в его развитии, существующие в настоящее время:

- Поток данных настолько велик, что многие данные пока невозможно проанализировать.
- Существуют инструменты ИИ, для работы которых не хватает данных, поскольку соответствующие данные не были собраны, переданы, или же существуют в отрыве от остальных данных, являются устаревшими, ненадежными, неточными, имеют нерабочий формат, или не размечены.
- Более того, в некоторых сферах имеются ограниченные, но критически важные данные, однако пока не существует инструментов ИИ, которые могли бы адекватно проанализировать небольшие объемы данных.

Было уже сказано, что чем быстрее обрабатываются данные, тем быстрее работают организации, использующие эти данные. Однако, если скорость обработки данных снижается из-за проблем с их доступностью, из-за снижения способности обрабатывать большие объемы данных и задержек, вызванных устаревшими версиями программ, организации могут не достичь желаемых результатов и утратить конкурентное преимущество. Кроме того, объем потенциально доступных необработанных данных не имеет большого значения, пока эти данные не преобразованы из различных разношерстных форматов в формат, который можно обработать и использовать для принятия наилучших решений.

ИИ способен анализировать большие (или, если нет другого варианта, небольшие) объемы данных, выявлять ранее неизвестные или скрытые закономерности и предоставлять полезную информацию в режиме реального времени. Эта информация должна использоваться, в частности, для реализации целей устойчивого развития на период до 2030 года, а также для поиска решений глобальных проблем. Об этом говорится в «Дорожной карте по цифровому сотрудничеству» Генерального секретаря ООН. В ней содержится призыв укреплять цифровой потенциал для расширения цифрового сотрудничества и интеграции в это сотрудничество новых стран (цель 4), а также поддерживать глобальное сотрудничество в области ИИ (цель 6) (Генеральная Ассамблея ООН, 2020 г.).

### **2.3.2 Синергия между ИИ и открытыми данными для ответа на глобальные вызовы**

---

Работа многих систем ИИ основана на использовании проприетарных данных. Однако было признано, что данные, соответствующие критериям открытых, могли бы в дальнейшем принести дополнительную пользу системам ИИ и снизить потенциальные риски для этих систем, такие как отсутствие беспристрастности, подотчетности и прозрачности (Davies, 2019). Это понимание отражено в рекомендации ЮНЕСКО по открытой науке (ЮНЕСКО, 2021b), рекомендации ЮНЕСКО по этике искусственного интеллекта (ЮНЕСКО, 2021d), а также в теме Текущей Волны открытых данных, поскольку они одинаково поощряют эффективное повторное использование данных.

Это также отражено в рекомендациях, содержащихся в отчете Администрации президента США, а именно Национального совета по науке и технологиям, Комитета по технологиям,





под названием «Подготовка к будущему искусственного интеллекта», в котором говорится: «Федеральные агентства должны уделять приоритетное внимание открытым данным для обучения и стандартам открытых данных в ИИ. Правительство должно сделать упор на выпуск наборов данных, которые позволят использовать ИИ для решения социальных проблем. Потенциальные шаги могут включать разработку инициативы «Открытые данные для ИИ» с целью высвобождения значительного количества набора правительственных данных для ускорения исследований в области ИИ и активизации использования стандартов открытых данных и передового опыта в правительстве, научном сообществе и частном секторе». (Администрация президента США – Национальный совет по науке и технологиям, Комитет по технологиям, 2016 г., стр. 14).

Однако также было признано, что режим совместного использования существующих данных, основанный на принципах доступности, совместимости и многократного использования, не является достаточным, и что необходимо также решить другие проблемы, перечисленные ранее, а именно: отсутствие, устаревание, ненадежность, неточность данных или отсутствие у них маркировки. Беспокойство вызывает, например, отсутствие данных о маргинальных сообществах или неформальной экономике, что, согласно Access Partnership (2018), является политической проблемой, поскольку системы ИИ даже бы не столкнулись с этими аспектами общественной жизни.

Эта ситуация может быть проиллюстрирована следующим примером: системы перевода между английским и французским языками в настоящее время достаточно развиты, поскольку для обучения этих систем доступны большие массивы обоих языков, чего нельзя сказать о множестве других языков. Другими словами, разработка успешного машинного обучения ИИ зависит от доступности данных, которые следует рассматривать как возможность для всех государств-членов: чтобы разрабатывать приложения ИИ для справедливого процесса принятия решений, во благо общества и т. д., данные не должны быть предвзятыми или не должны дискриминировать определенные страны, культуры или группы. Поэтому необходимо всесторонне учесть и обеспечить доступ к данным о забытых странах, культурах или группах<sup>80</sup>.

Существуют инициативы для решения этих проблем, такие как FAIR Forward<sup>81</sup> и Open for Good Alliance<sup>82</sup>. Эти инициативы направлены на улучшение локализованных данных надлежащего качества для обучения ИИ с целью внедрения локализованных инноваций в сфере ИИ. Например, FAIR Forward собрала 1200 часов голосовых записей на руандийском языке – киньяруанда. FAIR Forward также проверила индийские языки и отметила проблемы, препятствующие поиску для них универсальных наборов речевых данных в хорошем качестве и в стандартизованном формате (GIZ, 2020). Хотя такие отдельные инициативы и приветствуются, но открытые обучающие информационные порталы, посвященные ИИ, немногочисленны, а общая информационная стратегия, которая бы затрагивала оба вопроса: как сделать данные доступными для всех, и как небольшие доступные наборы данных могут привести к реальным идеям, – в них отсутствует.

Технологии на базе ИИ все больше проникают во все сферы жизни общества. Таким образом, взаимосвязь ИИ и гендерных проблем становится все более важной темой в рамках борьбы за гендерное равенство. Имеются опасения, что прикладываемые усилия в сфере гендерных вопросов и открытых данных преимущественно разобщены (GODAN, 2018), поскольку, во-первых, те группы, которые обсуждают проблемы женщин, и те, которые обсуждают открытость данных, редко взаимодействуют. Таким образом, реализация потребности в открытии данных может служить важным шагом на пути к гендерному равенству и расширению

79. <https://data.europa.eu/en/impact-studies/open-data-maturity>

80. Следует отметить, что, хотя для уменьшения предвзятости и дискриминационности систем ИИ необходимы более разнообразные данные, повышение разнообразия данных само по себе не сможет в полной мере разрешить эту сложную проблему.

81. <https://www.giz.de/expertise/html/61982.html>

прав и возможностей женщин. В период с 2022 по 2023 год ЮНЕСКО сделало ставку на то, чтобы возглавить всеобъемлющий и комплексный подход к развитию цифровых компетенций, включая сферу ИИ, а также чтобы устранить цифровой и интеллектуальный разрыв. Концепция ЮНЕСКО об универсальности интернета (ROAM-X (ПОДУ-К) ратует за открытость данных для преодоления гендерного цифрового разрыва.

Итак, у ИИ есть много возможностей для измерения и достижения устойчивого развития при условии, что соответствующие данные будут собираться, а также открыто предоставляться. Как отмечалось, предвзятость представляет собой риск для приложений на базе ИИ, который необходимо снизить, предоставляя максимально всеобъемлющие наборы данных. Как следствие возникает необходимость призвать государства-члены не только поддерживать открытость высококачественных данных, но и использовать технологии ИИ и содействовать наращиванию потенциала, обучению и образованию в этой сфере, включая всестороннее повышение грамотности в сфере открытых данных с учетом ИИ.

### **2.3.3 Потенциально возросшие риски для ИИ с учетом открытых данных**

---

ИИ – это технология двойного назначения, поэтому она связана не только с возможностями, но и с рисками. Важно рассматривать открытые данные не только как фактор риска для ИИ. Наборы данных, которые загружаются в системы ИИ, могут представлять собой угрозу безопасности, поскольку они могут привести к нежелательным результатам или реакциям ИИ. Хотя и упомянутые выше проблемы, связанные с предвзятостью и дискриминацией, являются нежелательными, но в значительной степени вызваны непреднамеренными действиями, однако в отношении систем ИИ могут совершаться и злоумышленные деяния, которые становятся результатом неправомерного использования открытых данных. Такие атаки получили название «яд в колодце» (poison in the well) или «атаки уклонения» (evasion attacks) и включают в себя различные методы (Lohn, 2021). Одним из таких способов может выступать, например, простая замена меток в наборах данных, тогда как другой, более изощренный способ нацелен на изображения с помощью вредоносных образцов (adversarial examples), которые тщательно продуманы и трудно распознаваемы для человека, но для ИИ приводят к сбою систем распознавания (Biggio & Roli, 2018). Появление общедоступных наборов данных непреднамеренно повышает риск таких атак, потому как происходит рост объема данных, потенциально доступных для атаки, но в то же время большой объем данных снижает шансы на выявление манипуляций с данными.

## Другие темы

Открытые данные и ИИ тесно связаны с рядом других тем, а также с проблемами, которые не могут быть подробно рассмотрены в этом отчете, но, тем не менее, актуальны и приводятся ниже для общего обзора:

- **Конфиденциальность данных:** большинство государственных данных касается граждан. Таким образом, баланс между раскрытием этих данных и частной жизнью граждан является предметом постоянных дискуссий (например, Scassa, 2019).
- **Этика данных:** эта тема связана с конфиденциальностью и «оценивает методы работы с данными, которые могут негативно повлиять на людей и общество – при сборе, обмене и использовании данных»<sup>83</sup>.
- **Правовые вопросы:** использование данных обычно регулируется законодательством. Например, Статья 8 Хартии Европейского союза об основных правах (2012 г.) гласит: «Каждый человек имеет право на защиту данных личного характера, относящихся к нему или ней»<sup>84</sup>.
- **Управление данными:** эта тема, например, рассматривается рабочей группой<sup>85</sup> в рамках Глобального партнерства по искусственному интеллекту (GPAI), которая представляет собой «многостороннюю инициативу, направленную на преодоление разрыва между теорией и практикой в области ИИ»<sup>86</sup>.
- **Информационная грамотность:** информационная грамотность является неперенным условием для использования (открытых) данных. Однако имеются определенные признаки того, что уровни информационной грамотности в отношении данных в определенных группах низки. Поэтому представляется целесообразным укрепление потенциала в данном направлении (например, Montes & Slater, 2019).
- **Инфраструктура данных:** еще одной предпосылкой для предоставления, а также использования открытых данных является доступность аппаратного компонента инфраструктуры данных, который неравномерно распределен по всему миру (Dodds & Wells 2019).
- **Гендерный аспект:** как указано выше, гендерное равенство связано с аспектами жизни общества, а значит, и с (открытыми) данными. Несбалансированность данных о женщинах сказывается, в частности, на системах ИИ, которые отражают эту ситуацию в своих выводах<sup>87</sup>.
- **Данные о коренных народах:** информация о коренных народах требует дополнительного осмысления, включая данные по их суверенитету, и может быть в общем виде выражена как ЗАБОТА (CARE): общее благо (Collective benefit), право на контроль (Authority to control), ответственность (Responsibility), этика (Ethics)<sup>88</sup>.
- **Малые данные:** как упоминалось ранее, зачастую в отношении маргинализированных групп населения доступен лишь ограниченный набор данных. Тем не менее, важно, чтобы ИИ мог работать с ними, например, как в подходе «быстрое обучение» (few-shot learning) (Wang et al., 2020).
- **Семантическая сеть:** семантическая сеть – это способ сделать интернет-данные машиночитаемыми. Чтобы увязать открытые данные с этим подходом, была придумана и официально оформлена концепция Сопряженных открытых данных (Linked Open Data) (например, Bauer & Kaltenböck, 2011).

82. <https://www.openforgood.info>

83. <https://theodi.org/service/consultancy/data-ethics/> См. например, также: <https://dataethics.eu/>

84. См. также Европейский парламент или Европейский совет (2016).

85. <https://gpai.ai/projects/data-governance/>

86. <https://gpai.ai>

87. <https://aplusalliance.org>

## 3. Руководящие принципы

### 3.1. Предисловие

---

Цель руководящих принципов состоит в том, чтобы государства-члены разработали план действий, а также политику устойчивого развития открытых данных. Приведенная выше концепция дает общую картину, а руководящие принципы позволят экспертам получить более четкое представление и «погрузиться» в конкретные темы.

Ниже приведены практические и одновременно с этим высокоуровневые шаги по раскрытию данных, основанные на существующих руководящих принципах. Можно выделить три стадии: подготовки, раскрытия данных и наблюдения за повторным использованием и устойчивостью, каждую из которых можно разбить на четыре этапа.

Важно отметить, что некоторые из этапов могут выполняться одновременно, т. е. не обязательно последовательно.

### 3.2. Подготовка

#### Разработка политики управления и обмена данными

---

Политика управления и обмена данными является важным предварительным условием перед раскрытием данных, поскольку такая политика определяет заинтересованность государства в обмене данными. Институт открытых данных предлагает следующие элементы политики открытых данных: определение термина «открытые данные», общую декларацию принципов, перечень типов данных и ссылки на соответствующее законодательство, правила или иные нормативные акты<sup>89</sup>. Государству рекомендуется придерживаться принципа «настолько открыто, насколько возможно, настолько закрыто, насколько необходимо» (Landi et al., 2020). Если данные не могут быть раскрыты с учетом проблем юридического характера, проблем, связанных с неприкосновенностью частной жизни или по иным причинам, например, персональные или охраняемые данные, то это должно быть четко оговорено. Политика управления и обмена данными должна также соответствовать принципам СПРАВЕДЛИВОСТИ (FAIR: Findability (находимости), Accessibility (доступности), Interoperability (совместимости), Reusability (повторного использования)) и ЗАБОТЫ (CARE: Collective benefit (общее благо), Authority to control (право на контроль), Responsibility (ответственность), Ethics (этика) для управления локальными данными. Кроме того, государство должно поощрять развитие политики управления и обмена данными, проводимой исследователями и частным сектором, на территории своих странах, придерживаясь этих принципов.

#### Сбор и объединение высококачественных данных

---

Сбор и хранение имеющихся данных должно производиться в одном хранилище, например, от различных государственных ведомств, где они могли храниться разрозненно. Более того, пробелы в данных должны быть заполнены за счет сбора новых данных. Данные должны быть точными и актуальными. Кроме того, данные должны быть полными и не должны, например, игнорировать меньшинства или неформальные сектора экономики. Данные о людях должны быть организованы, где это уместно, в том числе по доходу, полу, возрасту, расе, этнической принадлежности, миграционному статусу, инвалидности и географическому положению (Межучрежденческая рабочая группа по глобальному обзору достижения целей в области устойчивого развития, 2017 г.).

88. <https://www.gida-global.org/care>

## Развитие потенциала открытых данных

---

Развитие потенциала должно проходить по двум направлениям: государственные служащие и потенциальные пользователи данных. Ключевое слово – «информационная грамотность», которое входит в компетенцию, связанную с медийной и информационной грамотностью, предложенной ЮНЕСКО<sup>90</sup>, и для которой существуют курсы и образовательные программы. Для государственных служащих развитие потенциала включает в себя понимание выгод от открытых данных и преодоление распространенных страхов и непонимания. В это связи был придуман термин «управление данными» (data stewardship), который означает «систематическое, устойчивое и ответственное управление данными на благо общества» (GovLab, 2020). Для потенциальных пользователей развитие потенциала включает в себя демонстрацию возможностей открытых данных, например, возможность их повторного использования, а также возможность принятия осознанного (информированного) выбора.

## Подготовка данных для использования в ИИ

---

Если данные используются не только людьми, но и предполагается их использование в системах ИИ, то они должны соответствовать еще нескольким критериям. Первым шагом на пути к этому является подготовка данных в машиночитаемом формате. Системы ИИ могут воспринимать некоторые форматы лучше, чем остальные<sup>91</sup>. Если данные, кроме того, будут использоваться в качестве обучающих данных для обучения с учителем (supervised learning)<sup>92</sup> (в отличие от обучения без учителя (unsupervised learning)<sup>93</sup> или обучения с подкреплением (reinforcement learning)<sup>94</sup>), то они должны быть очищены и промаркированы, что зачастую требует много времени, а, следовательно, и средств. Успех системы ИИ зависит от качества обучающих данных, в том числе от их согласованности и актуальности. Необходимый объем обучающих данных трудно предугадать заранее, но его необходимо контролировать с помощью проверок производительности. Данные должны охватывать все сценарии, для которых была выстроена система ИИ.

## 3.3 Раскрытие данных

### Выберите наборы данных, которые необходимо открыть

---

Первым шагом в открытии данных является принятие решения о том, какие наборы данных следует открыть. Критерии в пользу открытия следующие: существовали ли ранее запросы на открытие этих данных, или эти данные были открыты другими государствами, и принесло ли это использование пользу. Критериями против открытия являются вопросы, связанные с конфиденциальностью, правами на интеллектуальную собственность или национальной безопасностью. Другими словами, открытие данных не должно нарушать национальные законы, такие как законы о защите персональных данных.

89. <https://theodi.org/article/how-to-write-a-good-open-data-policy/>

90. <https://iite.unesco.org/mil/>

91. См. обзор: <http://opendatahandbook.org/guide/en/appendices/file-formats/>

92. Обучение с учителем основано на помеченных данных, например, на фотографиях людей, которые были отмечены человеком. Затем строится модель, которую можно применять к схожим данным, например, для автоматической идентификации тех же людей на новых фотографиях (ЮНЕСКО, 2021с).

93. Обучение без учителя основано на данных, которые не были классифицированы или помечены. Цель состоит в том, чтобы выявить в них скрытые закономерности, которые используются для классификации новых данных. Примером может служить автоматическая идентификация букв и цифр в рукописном тексте путем поиска шаблонов в большом количестве образцов (ЮНЕСКО, 2021с).

94. Обучение с подкреплением включает в себя постоянное улучшение модели посредством обратной связи. ИИ конструирует из исходных данных модель, которая оценивается как правильная или неправильная и соответственно вознаграждается или наказывается. Затем это подкрепление используется для обновления модели, которая многократно повторяется с течением времени (ЮНЕСКО, 2021с).

## Открывайте наборы данных легально

---

Перед раскрытием наборов данных государство должно точно указать, на каких условиях, если таковые имеются, могут использоваться данные. Было разработано множество лицензий на открытые данные, и государство может выбирать, какая из них лучше всего соответствует его целям. Широко известны лицензии категорий Creative Commons и Open, «которые предоставляют каждому человеку или организации в мире возможность бесплатного, простого и стандартизованного способа получения разрешений на использование авторских прав на творческие и академические произведения; гарантируют указание авторства; и позволяют другим копировать, распространять и использовать эти произведения»<sup>95</sup>. Для всех данных, на которые распространяется действие лицензии Creative Commons, помимо отказа от всех прав могут быть установлены следующие ограничения:

- Атрибуция: При использовании данных необходимо указывать источник.
- Некоммерческий характер использования: данные не должны использоваться в коммерческих целях.
- Без производных: если данные изменены, они не могут быть переданы для совместного использования.
- Сохранение Условий (Share Alike): измененные данные могут использоваться совместно, но на условиях той же лицензии.

Другой вариант заключается в том, чтобы государства разрабатывали свои собственные лицензии, как сделало, например, правительство Великобритании<sup>96</sup>.

## Открывайте наборы данных технично

---

Наиболее распространенным способом открытия данных является их публикация в электронном формате для загрузки на веб-сайте, будь то государственный или сторонний веб-сайт. Кроме того, данные могут быть также опубликованы в архиве или репозитории данных, который должен, согласно Рекомендации ЮНЕСКО об открытой науке, «поддерживаться и обслуживаться академическим учреждением, научным обществом, государственным учреждением или другой хорошо зарекомендовавшей себя некоммерческой организацией, работающей на общее благо, что обеспечивает открытый доступ, неограниченное распространение, функциональную совместимость и долгосрочное цифровое хранение и архивирование» (ЮНЕСКО, 2021b, стр. 9). Данные должны быть представлены в формате, позволяющем осуществлять их поиск, обеспечивающем доступность, совместимость и повторное использование, что соответствует принципу СПРАВЕДЛИВОСТИ (FAIR), описанному выше. Для реализации принципов «находимости» и доступности существуют различные системы управления данными с открытым исходным кодом, которые позволяют сделать портал открытых данных в виде универмага «одного окна»<sup>97</sup>. Наборы данных должны сопровождаться метаданными, которые дополнительно уточняют данные и для которых существуют

95. <https://creativecommons.org/about/>

96. <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>

97. См. напр.: <https://ckan.org/>, <https://magda.io/> или <https://www.re3data.org/>

98. <http://dublincore.org/>

99. <https://www.w3.org/TR/vocab-dcat-2/>

100. <https://datacite.org/>

101. <https://schema.org/>

102. <https://5stardata.info/en/> и <https://www.w3.org/DesignIssues/LinkedData.html>

103. См. обзор: <http://opendatahandbook.org/guide/en/appendices/file-formats/>

104. См. напр.: <https://en.unesco.org/themes/access-information-laws>

четыре универсальных стандарта: Dublin Core (The Dublin Core Metadata Initiative)<sup>98</sup>, Data Catalog Vocabulary<sup>99</sup>, DataCite<sup>100</sup> и Schema.org<sup>101</sup>. Что касается конкретных форматов данных, то пятизвездочный план открытых данных был представлен ранее в этом отчете. Чем больше звездочек, тем более открытыми являются данные<sup>102</sup>. Форматы данных должны поддерживать совместимость, например, с интерфейсами прикладного программирования (API) и возможность повторного использования, например, другими пользователями, а также машинами<sup>103</sup>. Кроме того, необходимо помнить о читабельности ИИ, обсуждавшейся на шаге 4. Кроме того, меры безопасности имеют решающее значение для предотвращения манипулирования данными со стороны неавторизованных пользователей.

## Создавайте культуру открытых данных

---

Опыт показал, что помимо раскрытия данных с правовой и технической точки зрения, для создания культуры, основанной на открытых данных, необходимо выполнить по крайней мере еще две вещи: зачастую сотрудники, особенно в государственных ведомствах, не привыкли делиться данными, а скорее были обучены сохранять «бункерный менталитет». Более того, данные должны стать, по возможности, исключительной основой для принятия решений; другими словами, решения должны основываться на данных. Обе темы также поднимаются в «Шаге 3: Развитие потенциала открытых данных», однако помимо потенциала от всего вовлеченного персонала требуется совершить культурные изменения. Необходимо поощрять проактивное раскрытие данных, которое может гарантировать, что данные будут доступны еще до выставления запроса на них. Этого можно добиться путем включения подобных положений в законы о порядке предоставления доступа к информации<sup>104</sup>.

105. <http://opendatatoolkit.worldbank.org/en/demand.html#stage-3>

106. <https://www.licenses.ai/about>

## 3.4 Сопровождение повторного использования и устойчивости

### Поддержка участия граждан

---

После раскрытия данных они должны быть обнаружены потенциальными пользователями. Для этого необходимо разработать коммуникативную стратегию, которая может включать анонсирование раскрытия данных в сообществах открытых данных и соответствующих каналах социальных сетей, например, с использованием хэштега #opengov. Еще одним важным видом деятельности являются ранние консультации и взаимодействие с потенциальными пользователями. В дополнение к информированию потенциальных пользователей об открытых данных их также следует стимулировать к использованию и повторному использованию данных, а также к дальнейшему участию. Механизмы, которые доказали свою эффективность – это хакатоны или другие соревнования по использованию открытых данных для решения конкретных задач и проблем<sup>105</sup>.

### Поддержка международного участия

---

Международные партнерства могли бы еще больше увеличить преимущества открытых данных, например, благодаря сотрудничеству юг-юг и север-юг. Особенно важны партнерские отношения, которые позволяют поддерживать и наращивать потенциал для эффективного повторного использования данных как с использованием ИИ (см. ниже), так и без него. Альянс Open for Good, Глобальное партнерство по искусственному интеллекту (GPAI), а также межправительственные и региональные инициативы, которые были представлены выше, заслуживают высокой оценки.

### Поддержка полезного взаимодействия с ИИ

---

Открытые данные предоставляют множество возможностей для систем ИИ. Чтобы полностью раскрыть потенциал данных необходимо убедить разработчиков использовать открытые данные и разрабатывать системы ИИ под эти цели. В то же время необходимо предотвращать злоупотребление открытыми данными со стороны ненадежных и вредоносных приложений с использованием ИИ. Вот пример – Надежные Лицензии ИИ (RAIL)<sup>106</sup>. Этот шаг связан с предыдущим. Также рекомендуется вести публичный учет того, какие данные использовались системами ИИ и каким образом.

### Высокое качество поддержки данных

---

Многие данные быстро устаревают. Поэтому наборы данных должны регулярно обновляться.

**Шаг «Высокое качество поддержки данных» превращает это руководство в цикл, поскольку он ссылается на шаг «Сбор и объединение высококачественных данных».**



## Эпилог

Настоящие руководящие принципы служат призывом к действию в соответствии с Рекомендацией ЮНЕСКО об этике искусственного интеллекта (ЮНЕСКО, 2021d). Если государства-члены ЮНЕСКО будут следовать этим руководящим принципам и раскрывать данные на постоянной основе, а также создавать соответствующие возможности и культурную среду для открытых данных, то концепция со всеми ее преимуществами, изложенная в предыдущей части, может стать реальностью.

Открытые данные являются необходимым условием по обеспечению контроля, а также по достижению устойчивого развития. Из-за масштабности задач государства должны не только принять процесс раскрытия данных, но и создать благоприятные условия для продуктивного использования ИИ, который извлекает новые знания из открытых данных, а также для принятия решений на основе фактических данных.

В двух словах: сделайте ваши данные доступными для поиска, совместимыми и многообразными, а также готовыми к взаимодействию с ИИ, словом, ЧЕСТНЫМИ, чтобы их мог обрабатывать и анализировать кто угодно на благо общества.

## 4. Список литературы

Access Partnership (2018). Artificial Intelligence for Africa: An opportunity for growth, development and democratisation. South Africa: University of Pretoria. (Партнерство доступа (2018). Искусственный интеллект для Африки: возможности роста, развития и демократизации. Южная Африка: Университет Претории.)

[https://www.up.ac.za/media/shared/7/ZP\\_Files/ai-for-africa.zp165664.pdf](https://www.up.ac.za/media/shared/7/ZP_Files/ai-for-africa.zp165664.pdf)

Alamo, T., Reina, D. G., Mammarella, M., & Abella, A. (2020). Covid-19: Open-data resources for monitoring, modeling, and forecasting the epidemic. *Electronics*, 9(5), 827. (Аламо, Т., Рейна, Д.Г., Маммарелла, М., и Абелла, А. (2020). Covid-19: ресурсы с открытыми данными для мониторинга, моделирования и прогнозирования эпидемии. *Электроникс*, 9(5), 827.)

<https://www.mdpi.com/2079-9292/9/5/827/html>

Bauer, F., & Kaltenböck, M. (2011). *Linked open data: The Essentials*. Edition mono/monochrom, Vienna, 710. (Бауэр, Ф., и Кальтенбек, М. (2011). *Связанные открытые данные: основы*. Издание моно/монохром, Вена, 710 г.)

<https://www.reeep.org/LOD-the-Essentials.pdf>

Berners-Lee, T. (2012). Raw data, now! *Wired*. (Бернерс-Ли, Т. (2012). Исходные данные, немедленно! *Ваярд*.)

<https://www.wired.co.uk/article/raw-data>

Big Data for Migration Alliance (2021). *Designing Data Collaboratives to Better Understand Human Mobility and Migration in West Africa*. (Большие данные для миграционного альянса (2021 г.). Проектирование совместного использования данных для лучшего понимания мобильности и миграции населения Западной Африки.)

<https://odpl.thegovlab.com/uploads/ccm-directus/originals/aa45abf2-743d-4d58-8471-e25bff956ba5.pdf>

Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition*, 84, 317-331. (Биджио, Б., и Роли, Ф. (2018). Дикие паттерны: десять лет спустя становления состязательного машинного обучения. *Распознавание паттернов*, 84, 317-331.)

<https://arxiv.org/pdf/1712.03141.pdf>

Bodor, Á., Varjú, V., & Grünhut, Z. (2020). The effect of trust on the various dimensions of climate change attitudes. *Sustainability*, 12(23), 10200. (Бодор, А., Варью, В., и Грюнхут, З. (2020). Эффект доверия применительно к различным аспектам отношений к изменению климата. *Устойчивое развитие*, 12(23), 10200.)

<https://www.mdpi.com/2071-1050/12/23/10200>

von Borzyskowski, I., Mazumder, A., Mateen, B., & Wooldridge, M. (2021). Data science and AI in the age of COVID-19. The Alan Turing Institute. (фон Боржисковски И., Мазумдер А., Матин Б. и Вулдридж М. (2021). Наука о данных и искусственный интеллект в эпоху COVID-19. Институт Алана Тьюринга.)

<https://www.turing.ac.uk/research/publications/data-science-and-ai-agecovid-19-report>

Bostrom, N. (2011). Information hazards: A typology of potential harms from knowledge. Review of Contemporary Philosophy, (10), 44-79. (Бостром, Н. (2011). Информационные угрозы: Типология потенциальных угроз от знаний. Обзор современной философии, (10), 44-79.)

<https://www.nickbostrom.com/information-hazards.pdf>

Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford: Oxford University Press.

(Бостром, Н. (2014). Сверхразум: пути, опасности, стратегии. Оксфорд: Издательство Оксфордского университета.)

Cascella, M., Rajnik, M., Aleem, A., Dulebohn, S. C., & Di Napoli, R. (2022). Features, evaluation, and treatment of coronavirus (COVID-19). Statpearls [internet]. Каскелла, М., Райник, М., Алим, А., Дульбон, С.С., и Ди Наполи, Р. (2022). Особенности, оценка и лечение коронавируса (COVID-19). Statpearls [интернет].

<https://www.ncbi.nlm.nih.gov/books/NBK554776/>

Davies, T. (2019). Issues in Open Data – Algorithms and AI. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), The State of Open Data: Histories and Horizons. Cape Town and Ottawa: African Minds and International Development Research Centre. (Дэвис, Т. (2019). Проблемы открытых данных – алгоритмы и искусственный интеллект. В рамках Т. Дэвис, С. Уокер, М. Рубинштейн и Ф. Перини (ред.), Состояние открытых данных: истории и горизонты. Кейптаун и Оттава: Центр исследований африканских рынков и международного развития.)

<https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/57585/The%20State%20of%20Open%20Data.pdf>

Dodds, L. & Wells, P. (2019). Issues in Open Data – Data Infrastructure. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), The State of Open Data: Histories and Horizons. Cape Town and Ottawa: African Minds and International Development Research Centre. (Доддс, Л. и Уэллс, П. (2019). Проблемы открытых данных – инфраструктура данных. В рамках Т. Дэвис, С. Уокер, М. Рубинштейн и Ф. Перини (ред.), Состояние открытых данных: истории и горизонты. Кейптаун и Оттава: Центр исследований африканских рынков и международного развития.)

<https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/57585/The%20State%20of%20Open%20Data.pdf>

European Parliament and Council (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. (Европейский парламент и Совет (2016 г.). Регламент (ЕС) 2016/679 Европейского парламента и Совета от 27 апреля 2016 г. о защите физических лиц в отношении обработки персональных данных и о свободном перемещении таких данных.)

<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02016R0679-20160504&from=EN>

European Union (2012). Charter of Fundamental Rights of the European Union. Official Journal of the European Union, C 326/391. (Европейский союз (2012 г.). Хартия Европейского Союза об основных правах. Официальный журнал Европейского Союза, C 326/391.)

<https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012P/TXT&from=EN>

Executive Office of the President of the United States of America – National Science and Technology Council Committee on Technology (2016). Preparing for a Future of Artificial Intelligence. (Администрация президента США – Комитет по технологиям Национального совета по науке и технологиям (2016 г.). Подготовка к будущему искусственного интеллекта.)

[https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/microsites/ostp/NSTC/preparing\\_for\\_the\\_future\\_of\\_ai.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf)

Giovannini, G., Haick, H., & Garoli, D. (2021). Detecting COVID-19 from breath: A game changer for a big challenge. ACS sensors, 6(4), 1408-1417. (Джованнини Г., Хейк Х. и Гароли Д. (2021). Обнаружение COVID-19 по выдыхаемому воздуху: переломный момент в большой игре. ACS sensors, 6(4), 1408-1417.)

<https://pubs.acs.org/doi/10.1021/acssensors.1c00312>



GIZ (2020). A Study on Open Voice Data in Indian Languages. (ГИЗ (2020). Исследование открытых голосовых данных на примере индийских языков.)

[https://toolkit-digitalisierung.de/app/uploads/2021/02/Study-on-Open-Voice-Data-in-Indian-Languages\\_GIZ-BizAugmentor.pdf](https://toolkit-digitalisierung.de/app/uploads/2021/02/Study-on-Open-Voice-Data-in-Indian-Languages_GIZ-BizAugmentor.pdf)

GovLab (2020). Wanted: Data Stewards. (ГовЛаб (2020). Разыскиваются: распорядители данных.)

<https://thegovlab.org/static/files/publications/wanted-data-stewards.pdf>

Harrus, I., & Wyndham, J. (2021). Artificial intelligence and COVID-19: applications and impact assessment. AAAS AI Report. (Харрус, И., и Виндхэм, Дж. (2021). Искусственный интеллект и COVID-19: приложения и оценка воздействия. Отчет AAAS по ИИ.)

[https://www.aaas.org/sites/default/files/2021-05/AIandCOVID19\\_2021\\_FINAL.pdf](https://www.aaas.org/sites/default/files/2021-05/AIandCOVID19_2021_FINAL.pdf)

Heaven, W.D. (2021). Hundreds of AI tools have been built to catch covid. None of them helped. MIT Technology Review. (Хэвен В.Д. (2021). Сотни инструментов ИИ были созданы для борьбы с коронавирусом. Ни один из них не помог. Обзор технологий Массачусетского технологического института)

<https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/>

Hussain, A. A., Bouachir, O., Al-Turjman, F., & Aloqaily, M. (2020). AI techniques for COVID-19. IEEE access, 8, 128776-128795. (Хуссейн А.А., Буашир О., Аль-Турджман Ф. и Алокайли М. (2020). Методы искусственного интеллекта для COVID-19. IEEE access, 8, 128776-128795.)

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9136710>

Independent Expert Advisory Group on a Data Revolution for Sustainable Development (2014). A world that counts: Mobilizing a Data Revolution for Sustainable Development. Независимая консультативная группа экспертов по информационной революции для устойчивого развития (2014 г.). Мир, который имеет значение: мобилизуя революцию данных для устойчивого развития.

<http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>

Inter-Agency and Expert Group on Sustainable Development Goal Indicators (2017). Revised list of global Sustainable Development Goal indicators. Annex III. E/CN.3/2017/2. (Межведомственная и экспертная группа по достижению целевых показателей в сфере устойчивого развития (2017 г.). Пересмотренный список глобальных целевых показателей в области устойчивого развития. Приложение III. E/CN.3/2017/2.)

<https://unstats.un.org/sdgs/indicators/official%20revised%20list%20of%20global%20sdg%20indicators.pdf>

International Monetary Fund and Financial Stability Board (2021). G20 Data Gaps Initiative (DGI-2). The Sixth Progress Report – Countdown to December 2021. (Международный валютный фонд и Совет по финансовой стабильности (2021 г.). Инициатива G20 по пробелам в данных (DGI-2). Шестой отчет о ходе работы – обратный отчет до декабря 2021 года.)

<https://www.fsb.org/wp-content/uploads/P081021-1.pdf>

International Telecommunication Union and CISCO Systems (2016). Harnessing the Internet of Things for Global Development. (Международный союз электросвязи и CISCO Systems (2016 г.). Покорение интернета вещей для глобального развития.)

<http://www.itu.int/en/action/broadband/Documents/Harnessing-IoT-Global-Development.pdf>

Khemasuwan, D., & Colt, H. G. (2021). Applications and challenges of AI-based algorithms in the COVID-19 pandemic. BMJ Innovations, 7(2). (Кхемасуван, Д., и Колт, Х.Г. (2021). Приложения и вызовы алгоритмов на основе ИИ в условиях пандемии COVID-19. BMJ Innovations, 7 (2).)

<https://innovations.bmj.com/content/bmjinnov/7/2/387.full.pdf>

Klein, T., & Verhulst, S. (2017). Access to new data sources for statistics: Business models and incentives for the corporate sector. (Кляйн, Т., и Ферхюльст, С. (2017). Доступ к новым источникам данных для статистики: бизнес-модели и стимулы для корпоративного сектора.)

<https://www.thegovlab.org/static/files/publications/paris-21.pdf>

Landi, A., Thompson, M., Giannuzzi, V., Bonifazi, F., Labastida, I., da Silva Santos, L. O. B., & Roos, M. (2020). The “A” of FAIR—as open as possible, as closed, as necessary. *Data Intelligence*, 2(1-2), 47-55. (Ланди, А., Томпсон, М., Джаннуцци, В., Бонифази, Ф., Лабастиди, И., да Силва Сантос, Л. О. Б., и Рус, М. (2020). «А» (доступность) из FAIR – максимально открытый, настолько закрытый, насколько это необходимо. *Data Intelligence*, 2(1-2), 47-55.)

<https://direct.mit.edu/dint/article/2/1-2/47/9998/The-A-of-FAIR-As-Open-as-Possible-as-Closed-as>

Lee, C., Yang, T., Inchoco, G. D., Jones, G. M., & Satyanarayan, A. (2021). Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1-18). (Ли, К., Ян, Т., Инчоко, Г.Д., Джонс, Г.М., и Сатьянараян, А. (2021). Вирусные визуализации: как скептики коронавируса используют ортодоксальные методы работы с данными для продвижения неортодоксальной науки в Интернете. На материалах конференции CHI 2021 г., посвященной человеческому фактору в вычислительных системах (стр. 1-18).)

<https://arxiv.org/pdf/2101.07993.pdf>

Legg, S., & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence. *Minds and Machines*, 17(4). (Легг, С., и Хаттер, М. (2007). Универсальный разум: определение искусственного интеллекта. *Разум и машины*, 17 (4)).

<https://arxiv.org/pdf/0712.3329.pdf>

Lohn, A.J. (2021). Poison in the Well – Securing the Shared Resources of Machine Learning. CSET Policy Brief. (Лон, А.Дж. (2021). Яд в колодце – защита общих ресурсов машинного обучения. Краткий обзор политики Джорджтаунского центра безопасности и новых технологий (CSET)).

<https://cset.georgetown.edu/publication/poison-in-the-well/>

Lozano, M. A., Piñol, E., Rebollo, M., Polotskaya, K., Garcia-March, M. A., Conejero, J. A., ... & Oliver, N. (2021). Open Data Science to Fight COVID-19: Winning the 500k XPRIZE Pandemic Response Challenge. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 384-399). Springer, Cham. (Лозано, М. А., Пиньоль, Э., Реболло, М., Полоцкая, К., Гарсия-Марч, М. А., Конехеро, Дж. А., ... и Оливер, Н. (2021). Наука открытых данных для борьбы с COVID-19: выигрыш приза XPRIZE в размере 500 тыс. долларов в соревновании по реагированию на пандемию. По материалам совместной европейской конференции по машинному обучению и открытию знаний в базах данных (стр. 384-399). Спрингер, Чам.)

[https://2021.ecmlpkdd.org/wp-content/uploads/2021/07/sub\\_638.pdf](https://2021.ecmlpkdd.org/wp-content/uploads/2021/07/sub_638.pdf)

Milan, S., & Treré, E. (2020). The Rise of the Data Poor: The COVID-19 Pandemic Seen From the Margins. *Social Media+ Society*, 6(3), 2056305120948233. (Милан, С., и Трепре, Э. (2020). Восстание скудных данных: пандемия COVID-19 – взгляд со стороны. *Соцсети + общество*, 6(3), 2056305120948233.)

[https://journals.sagepub.com/doi/full/10.1177/2056305120948233?fbclid=IwAR0jUNmHgCSSZmOXaByn3X4EyK oWRDCxM6qXnCA1FuN-yrgF\\_s8rnoCgVqM&](https://journals.sagepub.com/doi/full/10.1177/2056305120948233?fbclid=IwAR0jUNmHgCSSZmOXaByn3X4EyK oWRDCxM6qXnCA1FuN-yrgF_s8rnoCgVqM&)

Montes, M. & Slater, D. (2019). Issues in Open Data – Data Literacy. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), *The State of Open Data: Histories and Horizons*. Cape Town and Ottawa: African Minds and International Development Research Centre. (Монтес, М. и Слейтер, Д. (2019). Проблемы открытых данных – информационная грамотность. В Т. Дэвис, С. Уокер, М. Рубинштейн и Ф. Перини (редакторы), *Состояние открытых данных: истории и горизонты*. Кейптаун и Оттава: Центр исследований африканских рынков и международного развития).

<https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/57585/The%20State%20of%20Open%20Data.pdf>

Odrizola-González, P., Planchuelo-Gómez, Á., Iruña, M. J., & de Luis-García, R. (2020). Psychological effects of the COVID-19 outbreak and lockdown among students and workers of a Spanish university. *Psychiatry research*, 290, 113108. (Одриосола-Гонсалес, П., Планчуэло-Гомес, А., Ируна, М. Дж., и де Луис-Гарсия, Р. (2020). Психологические последствия вспышки COVID-19 и самоизоляции среди студентов и сотрудников Испанского университета. *Психиатрические исследования*, 290, 113108).

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7236679/>

OECD (2020). Why open science is critical to combatting COVID-19. *OECD Policy Responses to Coronavirus (COVID-19)*. (Организация экономического сотрудничества и развития (2020 г.). Почему открытая наука имеет решающее значение для борьбы с COVID-19. Политические ответные меры Организации экономического сотрудничества и развития на коронавирус (COVID-19))

[https://read.oecd-ilibrary.org/view/?ref=129\\_129916-31pgjnl6cb&title=Whyopen-science-is-critical-to-combatting-COVID-19](https://read.oecd-ilibrary.org/view/?ref=129_129916-31pgjnl6cb&title=Whyopen-science-is-critical-to-combatting-COVID-19)



OECD & GovLab (2021). Open Data in action. Initiatives during the initial stage of the COVID-19 pandemic. (Организация экономического сотрудничества и развития и GovLab (2021). Открытые данные в действии. Инициативы на начальном этапе пандемии COVID-19).

<https://www.oecd.org/gov/digital-government/open-data-in-action-initiativesduring-the-initial-stage-of-the-covid-19-pandemic.pdf>

Ong, E., Wong, M. U., Huffman, A., & He, Y. (2020). COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *Frontiers in immunology*, 11, 1581. (Онг, Э., Вонг, М.У., Хаффман, А., и Хе, Ю. (2020). Разработка вакцины против коронавируса COVID-19 с использованием обратной вакцинологии и машинного обучения. *Границы в иммунологии*, 11, 1581).

<https://www.frontiersin.org/articles/10.3389/fimmu.2020.01581/full>

Open Data Charter (2015). International Open Data Charter. (Хартия открытых данных (2015 г.). Международная хартия открытых данных).

[https://opendatacharter.net/wp-content/uploads/2015/10/opendatachartercharter\\_F.pdf](https://opendatacharter.net/wp-content/uploads/2015/10/opendatachartercharter_F.pdf)

Pahar, M., Klopfer, M., Warren, R., & Niesler, T. (2021). COVID-19 cough classification using machine learning and global smartphone recordings. *Computers in Biology and Medicine*, 135, 104572. (Пахар, М., Клоппер, М., Уоррен, Р., и Нислер, Т. (2021). Классификация кашля COVID-19 с использованием машинного обучения и глобальных записей смартфонов. *Компьютеры в биологии и медицине*, 135, 104572).

<https://arxiv.org/pdf/2012.01926.pdf>

Panchal, N., Kamal, R., Orgera, K., Cox, C., Garfield, R., Hamel, L., & Chidambaram, P. (2020). The implications of COVID-19 for mental health and substance use. *Kaiser family foundation*, 21. (Панчал, Н., Камаль, Р., Оргера, К., Кокс, К., Гарфилд, Р., Хамел, Л., и Чидамбарам, П. (2020). Последствия COVID-19 для психического здоровья и употребления психоактивных веществ. *Фонд семьи Кайзер*, 21).

<https://www.kff.org/coronavirus-covid-19/issue-brief/the-implications-of-covid-19-for-mental-health-and-substance-use/>

Petrov, O., Gurin, J., & Manley, L. (2016). Open data for sustainable development. (Петров О., Гурин Дж. и Мэнли Л. (2016). Открытые данные для устойчивого развития).

<https://openknowledge.worldbank.org/bitstream/handle/10986/24017/Open0data0for0sustainable0development.pdf?sequence=4&isAllowed=y>

Ratcliffe, J. & Tuzeneu, M.-C. (2019). of Climate Data by Building Trust. Blog post. (Рэтклифф, Дж. и Тузенуе, М.-К. (2019). Максимизация возможностей климатических данных за счет укрепления доверия. *Сообщения блога*).

<https://www.climatelinks.org/blog/maximizing-power-climate-data-building-trust>

Raveendran, A. V., Jayadevan, R., & Sashidharan, S. (2021). Long COVID: an overview. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. (Равиндран, А.В., Джаядеван, Р., и Шашидхаран, С. (2021). Долгий COVID: обзор. *Диабет и метаболический синдром: клинические исследования и обзоры*).

<https://www.sciencedirect.com/science/article/pii/S1871402121001193>

Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., ... & Schönlieb, C. B. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nature Machine Intelligence*, 3(3), 199-217. Робертс М., Дриггс Д., Торп М., Гилби Дж., Юнг М., Урспрунг С., ... и Шенлиб С. Б. (2021). Типичные ошибки и рекомендации по использованию машинного обучения для выявления и прогнозирования COVID-19 с использованием рентгенограмм грудной клетки и результатов компьютерной томографии. *Nature Machine Intelligence*, 3(3), 199-217).

<https://www.nature.com/articles/s42256-021-00307-0>

Russell, S. J. & Norvig, P. (2015). *Artificial Intelligence - A Modern Approach*, Third Edition. Upper Saddle River: Pearson. Sachs, J. D., Kroll, C., Lafortune, G., Fuller, G. & Woelm, F. (2021). *Sustainable Development Report 2021*. Cambridge: Cambridge University Press. (Рассел, С. Дж. и Норвиг, П. (2015). *Искусственный интеллект – современный подход*, третье издание. Река Аппер-Сэдл: Пирсон. Сакс, Дж. Д., Кролл, К., Лафортун, Г., Фуллер, Г. и Велм, Ф. (2021). *Отчет об устойчивом развитии за 2021 год*. Кембридж: Издательство Кембриджского университета).

<https://s3.amazonaws.com/sustainabledevelopment.report/2021/2021-sustainable-development-report.pdf>



Scassa, T. (2019). Issues in Open Data – Privacy. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), *The State of Open Data: Histories and Horizons*. Cape Town and Ottawa: African Minds and International Development Research Centre. (Скаска, Т. (2019). Проблемы открытых данных – конфиденциальность. В составе Т. Дэвис, С. Уокер, М. Рубинштейн и Ф. Перини (ред.), *Состояние открытых данных: истории и горизонты*. Кейптаун и Оттава: Центр исследований африканских рынков и международного развития).

<https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/57585/The%20State%20of%20Open%20Data.pdf>

Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., ... & Shen, D. (2020). Review of artificial intelligence techniques in imaging data acquisition, segmentation, and diagnosis for COVID-19. *IEEE reviews in biomedical engineering*, 14, 4-15. (Ши, Ф., Ван, Дж., Ши, Дж., Ву, З., Ван, К., Тан, З., ... и Шен, Д. (2020). Обзор IEEE применения методик искусственного интеллекта при сборе данных визуализации, сегментации и диагностики COVID-19. *Обзоры IEEE по биомедицинской инженерии*, 14, 4-15).

<https://arxiv.org/ftp/arxiv/papers/2004/2004.02731.pdf>

Shuja, J., Alanazi, E., Alasmay, W., & Alashaikh, A. (2021). COVID-19 open source datasets: a comprehensive survey. *Applied Intelligence*, 51(3), 1296-1325. (Шуджа, Дж., Аланази, Э., Аласмари, В., и Алашайх, А. (2021). Наборы данных из открытых источников по COVID-19: всестороннее исследование. *Эпплайд интеллидженс*, 51 (3), 1296–1325).

<https://link.springer.com/article/10.1007/s10489-020-01862-6>

UN General Assembly (2020). Road map for digital cooperation: implementation of the recommendations of the High-level Panel on Digital Cooperation. Report of the Secretary-General. A/74/821. (Генеральная Ассамблея ООН (2020 г.). Дорожная карта цифрового сотрудничества: выполнение рекомендаций Группы высокого уровня по цифровому сотрудничеству. Доклад Генерального секретаря. A/74/821).

<https://documents-dds-ny.un.org/doc/UNDOC/GEN/N20/102/51/PDF/N2010251.pdf?OpenElement>

UNESCO (2020). Disinfodemic. Deciphering COVID-19 disinformation. Policy brief 1. (ЮНЕСКО (2020 г.). Дезинформация. Расшифровка дезинформации о COVID-19. Аналитическая записка 1.)

[https://en.unesco.org/sites/default/files/disinfodemic\\_deciphering\\_covid19\\_disinformation.pdf](https://en.unesco.org/sites/default/files/disinfodemic_deciphering_covid19_disinformation.pdf)

UNESCO (2021a). One year into COVID: prioritizing education recovery to avoid a generational catastrophe. ED/ADG/2021/01. (ЮНЕСКО (2021а). Год после COVID: приоритизация восстановления образования, чтобы избежать катастрофы поколений. ED/ADG/2021/01).

<https://unesdoc.unesco.org/ark:/48223/pf0000376984>

UNESCO (2021b). Recommendation on Open Science. SC-PCB-SPP/2021/OS/UROS. (ЮНЕСКО (2021b). Рекомендация по открытой науке. SC-PCB-SPP/2021/OS/UROS.)

<https://unesdoc.unesco.org/ark:/48223/pf0000379949>

UNESCO (2021c). AI and Education Guidance for Policy-makers. Paris, UNESCO. (ЮНЕСКО (2021c). ИИ и руководство по образованию для политиков. Париж, ЮНЕСКО).

<https://unesdoc.unesco.org/ark:/48223/pf0000376709>

UNESCO (2021d). Recommendation on the Ethics of Artificial Intelligence. Paris, UNESCO. (ЮНЕСКО (2021d). Рекомендация по этике искусственного интеллекта. Париж, ЮНЕСКО).

<https://unesdoc.unesco.org/ark:/48223/pf0000381137>

UN System Chief Executives Board for Coordination (2020). Summary of deliberations. Addendum. System-wide Road Map for Innovating United Nations Data and Statistics. CEB/2020/1/Add.1. (Координационный совет руководителей системы ООН (2020 г.). Резюме обсуждений. Дополнение. Общесистемная дорожная карта по обновлению данных и статистики ООНаций. CEB/2020/1/Add.1)

[https://unsceb.org/sites/default/files/2021-01/CEB\\_2020\\_1\\_Add1.pdf](https://unsceb.org/sites/default/files/2021-01/CEB_2020_1_Add1.pdf)

Van Noorden, R. (2021). Scientists call for fully open sharing of coronavirus genome data. *Nature*, 590(7845), 195-197. (Ван Норден, Р. (2021). Ученые призывают к полностью открытому обмену данными о геноме коронавируса. *Нейчер*, 590 (7845), 195-197).

<https://www.nature.com/articles/d41586-021-00305-7>



Waltman, L., Pinfield, S., Rzaeva, N., Henriques, S. O., Fang, Z., Brumberg, J., ... & Swaminathan, S. (2021). Scholarly communication in times of crisis: The response of the scholarly communication system to the COVID-19 pandemic. (Вальтам Л., Пинфилд С., Рзаева Н., Энрикес С.О., Фанг З., Брумберг Дж., ... и Сваминатан С. (2021). Научная коммуникация во время кризиса: ответ системы научной коммуникации на пандемию COVID-19).

[https://ori.figshare.com/articles/report/\\_/17125394](https://ori.figshare.com/articles/report/_/17125394)

Wang, Y., Yao, Q., Kwok, J. T., & Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. ACM Computing Surveys (CSUR), 53(3), 1-34. (Ван Ю., Яо К., Квок Дж. Т. и Ни Л. М. (2020). Обобщение на основе нескольких примеров: обзор быстрого обучения. Вычислительные исследования ACM (CSUR), 53(3), 1-34).

<https://arxiv.org/pdf/1904.05046.pdf>

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18. (Вилкнсон, М.Д. и соавт. (2016). Руководящие принципы FAIR по управлению научными данными. Науч. данные 3:160018 doi: 10.1038/sdata.2016.18.)

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4792175/>

Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E., ... & van Smeden, M. (2020). Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *bmj*, 369. (Винантс, Л., Ван Калстер, Б., Коллинз, Г.С., Райли, Р.Д., Хайнце, Г., Шуйт, Э., ... и ван Смеден, М. (2020). Модели прогнозирования для диагностики и прогноза covid-19: системный обзор и критическая оценка. *bmj*, 369).

<https://www.bmj.com/content/369/bmj.m1328>

Yampolskiy, R. V. (Ed.). (2018). Artificial intelligence safety and security. CRC Press. Yan, L., Zhang, H. T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., ... & Yuan, Y. (2020). An interpretable mortality prediction model for COVID-19 patients. *Nature machine intelligence*, 2(5), 283-288. (Ямпольский, Р. В. (ред.). (2018). Безопасность и безопасность искусственного интеллекта. CRC Press. Ян, Л., Чжан, Х.Т., Гонсалвес, Дж., Сяо, Ю., Ван, М., Го, Ю., ... и Юань, Ю. (2020). Интерпретируемая модель прогнозирования смертности для пациентов с COVID-19. *Природа машинного интеллекта*, 2(5), 283-288).

<https://www.nature.com/articles/s42256-020-0180-7.pdf>

Ziesche, S. (2017). Innovative big data approaches for capturing and analyzing data to monitor and achieve the SDGs. Report of the United Nations Economic and Social Commission for Asia and the Pacific: Subregional Office for East and North-East Asia (ESCAP-ENEА). (Зиеше, С. (2017). Инновационные подходы к большим данным для сбора и анализа данных для мониторинга и достижения ЦУР. Доклад Экономической и социальной комиссии ООН для Азии и Тихого океана: Субрегиональное отделение для Восточной и Северо-Восточной Азии (ESCAP-ENEА)).

<https://www.unescap.org/publications/innovative-big-data-approachescapturing-and-analyzing-data-monitor-and-achieve-sdgs>