

Сентябрь 2024 года

# УПРАВЛЕНИЕ ИСКУССТВЕННЫМ ИНТЕЛЛЕКТОМ В ИНТЕРЕСАХ ЧЕЛОВЕЧЕСТВА



United  
Nations



AI  
Advisory  
Body



**Управление искусственным интеллектом в интересах человечества:  
Заключительный доклад**

Авторское право © Организация Объединенных Наций, 2024 год  
Все права защищены.

Запрещается полное или частичное воспроизведение или передача настоящего издания в какой бы то ни было форме или какими бы то ни было средствами, электронными или механическими, включая фотокопирование, запись или использование в любых существующих или изобретенных в будущем системах хранения и поиска информации, без письменного разрешения издателя.

Заявки на получение разрешения на частичное воспроизведение или фотокопирование просьба направлять в Центр выдачи разрешений в области авторского права по адресу: URL: [copyright.com](http://copyright.com)

Со всеми запросами относительно прав и лицензий, включая производные права, просьба обращаться по адресу: United Nations Publications, 405 East 42nd Street, S-11FW001, New York, NY 10017, United States of America. Электронная почта: [permissions@un.org](mailto:permissions@un.org); веб-сайт: URL: [shop.un.org](http://shop.un.org).

Употребляемые обозначения и изложение материала в настоящем издании не означают выражения со стороны Секретариата Организации Объединенных Наций какого бы то ни было мнения относительно правового статуса той или иной страны, территории, города или района, или их властей или относительно делимитации их границ.

eISBN: 9789211068870

# Управление искусственным интеллектом в интересах человечества

Заключительный доклад



**United  
Nations**



AI  
Advisory  
Body

## О Консультативном органе высокого уровня по искусственному интеллекту

Консультативный орган высокого уровня по искусственному интеллекту с участием многих заинтересованных сторон, предложение о создании которого впервые прозвучало в 2020 году в Дорожной карте по цифровому сотрудничеству Генерального секретаря Организации Объединенных Наций (A/74/821), был сформирован в октябре 2023 года в целях проведения анализа и выработки рекомендаций относительно международного управления искусственным интеллектом.

Члены Консультативного органа участвовали в его работе в личном качестве, а не как представители своих соответствующих организаций. Настоящий доклад представляет собой консенсус большинства; не ожидается, что кто-либо из членов согласится с каждым пунктом, содержащимся в этом документе. Члены Консультативного органа подтверждают свое широкое, но не универсальное согласие с приводимыми в докладе выводами и рекомендациями. Формулировки, включенные в настоящий доклад, не подразумевают институционального одобрения со стороны соответствующих организаций.

# Оглавление

|  |           |
|--|-----------|
| <b>О Консультативном органе высокого уровня по искусственному интеллекту</b> | <b>4</b>  |
| <b>Резюме</b>  | <b>7</b>  |
| 1. Необходимость глобального управления                                      | 7         |
| 2. Глобальные пробелы в управлении ИИ  | 8         |
| 3. Укрепление глобального сотрудничества                                     | 10        |
| А. Общее понимание   | 10        |
| В. Общие позиции   | 12        |
| С. Общие выгоды  | 15        |
| D. Согласованность усилий  | 21        |
| E. Размышления на тему организационных моделей                               | 24        |
| 4. Призыв к действиям  | 24        |
| <b>1. Введение</b>   | <b>26</b> |
| А. Возможности и предпосылки   | 28        |
| В. Основные предпосылки использования ИИ в интересах человечества            | 32        |
| С. Управление как одна из ключевых предпосылок                               | 32        |
| D. Риски и вызовы  | 33        |
| E. Риски, связанные с ИИ   | 33        |
| F. Вызовы, требующие реагирования  | 40        |
| <b>2. Необходимость глобального управления</b>                               | <b>44</b> |
| А. Руководящие принципы и функции в контексте международного управления ИИ   | 45        |
| В. Формирующийся ландшафт международного управления ИИ                       | 48        |

|  |            |
|--|------------|
| <b>3. Пробелы в глобальном управлении ИИ</b>   | <b>51</b>  |
| А. Пробелы в представительстве   | 51         |
| В. Пробелы в координации   | 53         |
| С. Пробелы в реализации  | 55         |
| <b>4. Укрепление глобального сотрудничества</b>  | <b>58</b>  |
| А. Общее понимание   | 59         |
| Международная научная группа по вопросам ИИ  | 59         |
| В. Общие позиции   | 64         |
| Политический диалог по вопросам управления ИИ  | 65         |
| Обмен стандартами в области ИИ   | 69         |
| С. Общие выгоды  | 72         |
| Сеть для наращивания потенциала  | 79         |
| Глобальный фонд для ИИ   | 81         |
| Глобальная система данных ИИ   | 84         |
| D. Согласованность усилий  | 88         |
| Подразделение по вопросам ИИ в рамках Секретариата Организации<br>Объединенных Наций               | 88         |
| E. Размышления на тему организационных моделей   | 92         |
| Международное агентство по ИИ?   | 92         |
| <b>5. Заключение: призыв к действию</b>  | <b>98</b>  |
| <b>Приложения</b>  | <b>100</b> |
| Приложение А: члены Консультативного органа высокого уровня по<br>искусственному интеллекту        | 100        |
| Приложение В: круг ведения Консультативного органа высокого уровня по<br>искусственному интеллекту | 101        |
| Приложение С: список консультативных мероприятий в 2024 году                                       | 102        |
| Приложение D: список «глубоких погружений»   | 103        |
| Приложение E: ответы, полученные в рамках Глобальной оценки рисков,<br>связанных с ИИ              | 104        |
| Приложение F: ответы на вопросы, касающиеся возможностей   | 114        |
| Приложение G: список сокращений  | 120        |

# Резюме

- i** Искусственный интеллект (ИИ) преобразует наш мир. Этот комплекс технологий обладает огромным потенциалом для благих целей: от открытия новых областей научных исследований и оптимизации энергетических сетей до улучшения здоровья населения и сельского хозяйства, а также содействия более широкому прогрессу в достижении целей в области устойчивого развития (ЦУР).
- ii** Вместе с тем, если оставить ИИ без управления, его возможности могут не реализоваться или же распределяться неравномерно. Увеличение цифрового разрыва может привести к тому, что преимуществами ИИ смогут воспользоваться лишь несколько государств, компаний и отдельных лиц. Упущенные шансы – неспособность воспользоваться преимуществами ИИ и поделиться ими из-за отсутствия доверия или неблагоприятных условий, таких как изъядны потенциала и неэффективное управление, – могут ограничить сферу возможностей.
- iii** ИИ несет в себе и другие риски. К предвзятости ИИ и его использованию для целей слежки добавляются новые проблемы, такие как конфабуляции (или «галлюцинации») больших языковых моделей, создание и распространение дезинформации с помощью ИИ, риски для мира и безопасности, а также энергопотребление систем ИИ в период климатического кризиса.
- iv** Быстрые, непрозрачные и автономные системы ИИ бросают вызов традиционным системам регулирования, а все более мощные системы могут радикально изменить мир трудовой деятельности. Автономное оружие и использование ИИ в целях обеспечения

общественной безопасности ставят серьезные вопросы в том, что касается права, безопасности и гуманитарной сферы.

- v** Сегодня на глобальном уровне существует дефицит управления в отношении ИИ. Несмотря на активное обсуждение этических стандартов и принципов, система норм и институтов еще только зарождается и полна пробелов. Зачастую бросается в глаза отсутствие подотчетности, в том числе при развертывании закрытых систем ИИ, которые воздействуют на других. Соблюдение норм часто основывается на добровольности; практика не соответствует риторике.

- vi** Как отмечалось в нашем промежуточном докладе<sup>1</sup>, управление ИИ имеет ключевое значение не только для решения проблем и нейтрализации рисков, но и для обеспечения того, чтобы мы использовали потенциал ИИ таким образом, чтобы никто не остался в стороне.

## 1. Необходимость глобального управления

- vii** В частности, настоятельная необходимость глобального управления является неоспоримой. Сырье для ИИ – от важнейших полезных ископаемых до обучающих данных – добывается в глобальном масштабе. ИИ общего назначения, развернутый на трансграничном уровне, порождает множество способов применения во всем мире. Ускоренное развитие ИИ концентрирует власть и богатство в глобальном масштабе, что влечет за собой геополитические и геоэкономические последствия.

<sup>1</sup> См. <https://un.org/ai-advisory-body>.

**viii** Более того, в настоящее время никто не понимает всех внутренних механизмов ИИ настолько, чтобы полностью контролировать результаты его применения или прогнозировать его развитие. Лица, принимающие решения, также не могут нести ответственность за разработку, развертывание или использование систем, которые они не понимают. Между тем негативные побочные эффекты и последствия таких решений для смежных областей также могут быть глобальными.

**ix** Разработка, внедрение и использование такой технологии не могут быть предоставлены только прихотям рынков. Национальные правительства и региональные организации будут играть решающую роль, но сама природа данной технологии – трансграничной по структуре и применению – требует глобального подхода. Управление также может стать ключевым фактором, способствующим внедрению инноваций в сфере ИИ для достижения ЦУР на глобальном уровне.

**x** Таким образом, ИИ создает проблемы и возможности, которые требуют целостного, глобального подхода, охватывающего все политические, экономические, социальные, этические, правозащитные, технические, экологические и другие области. Такой подход может превратить множество развивающихся инициатив в комплексную, взаимосвязанную систему, которая будет основана на международном праве и ЦУР и сможет приспособиться к разным условиям и обстоятельствам, меняющимся с течением времени.

**xi** В нашем промежуточном докладе мы изложили принципы<sup>2</sup>, которыми следует руководствоваться при формировании новых международных институтов управления ИИ.

Эти принципы исходят из того, что управление ИИ не осуществляется в вакууме, и что к ИИ применяются нормы международного права, особенно в том, что касается прав человека.

## 2. Глобальные пробелы в управлении ИИ

**xii** Нет недостатка в документах и диалогах, посвященных управлению ИИ. Правительствами, компаниями и консорциумами, а также региональными и международными организациями были разработаны сотни руководств, рамочных документов и принципов.

**xiii** Однако ни один из них не может быть по-настоящему глобальным по сфере применения и всеобъемлющим по охвату. Это порождает проблемы представительства, координации и реализации.

**xiv** Что касается представительства, то целые регионы мира остались за рамками международных дискуссий по вопросам управления ИИ. На рисунке (а) показаны семь известных инициатив в области ИИ, реализуемых вне системы Организации Объединенных Наций<sup>3</sup>. Семь стран участвуют во всех вышеуказанных проектах в сфере управления ИИ, тогда как 118 стран не участвуют ни в одном (в основном речь идет о странах глобального Юга).

**xv** Справедливость требует, чтобы большее число людей играло значимую роль в принятии решений о том, как управлять технологиями, которые влияют на нас. Концентрация принятия решений в технологическом секторе ИИ не может быть оправдана; мы также должны признать, что исторически многие сообщества были полностью исключены из обсуждения вопросов управления ИИ, которые их затрагивают.

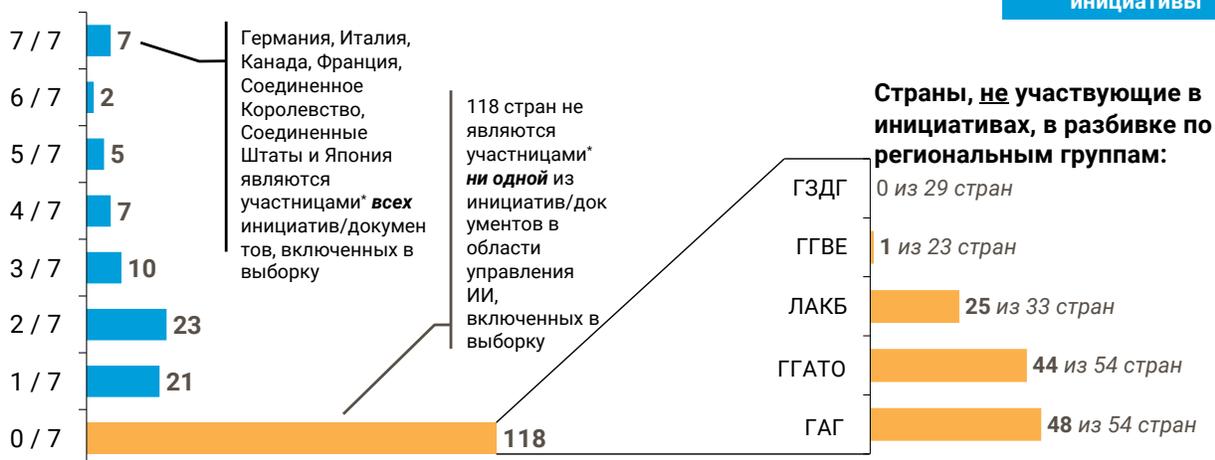
2 Руководящий принцип 1: управление ИИ должно быть инклюзивным, осуществляться всеми и на благо всех; руководящий принцип 2: управление ИИ должно осуществляться в интересах общества; руководящий принцип 3: управление ИИ должно строиться параллельно управлению данными и поощрению общедоступности данных; руководящий принцип 4: управление ИИ должно быть универсальным, сетевым и основываться на адаптивном сотрудничестве многих заинтересованных сторон; руководящий принцип 5: управление ИИ должно опираться на Устав Организации Объединенных Наций, международные нормы в области прав человека и другие согласованные международные обязательства, такие как ЦУР.

3 Исключая Рекомендацию Организации Объединенных Наций по вопросам образования, науки и культуры (ЮНЕСКО) по этике использования искусственного интеллекта (2021 год), а также две резолюции Генеральной Ассамблеи по вопросу об ИИ от 2024 года: «Использование возможностей безопасных, защищенных и надежных систем искусственного интеллекта в целях устойчивого развития» (78/265) и «Укрепление международного сотрудничества в деле наращивания потенциала в области искусственного интеллекта» (78/311).

## Рисунок (а): Участие в семи международных инициативах по управлению ИИ, реализуемых вне системы Организации Объединенных Наций

Выборка: Принципы ИИ ОЭСР (2019 год), Принципы ИИ Г-20 (2019 год), Группа Совета Европы по разработке конвенции по ИИ (2022–2024 годы), Заявление министров ГПИИ (2022 год), Декларация Блетчли (2023 год) и Сеульское заявление министров (2024 год).

Выборка включает исключительно межрегиональные (не региональные) инициативы



\* По признаку подтверждения соответствующих межправительственных документов. Страны не считаются участниками той или иной многосторонней инициативы исключительно в силу их членства в Европейском союзе или Африканском союзе. Сокращения: ГАГ – Группа африканских государств; ГГАТО – Группа государств Азии и Тихого океана; ГГВЕ – Группа государств Восточной Европы; Г-20 – Группа 20; Г-7 – Группа семи; ГПИИ – Глобальное партнерство по искусственному интеллекту; ЛАКБ – Латинская Америка и Карибский бассейн; ОЭСР – Организация экономического сотрудничества и развития; ГЗДГ – Группа западноевропейских и других государств.

**xvi** Для того, чтобы режимы управления ИИ были эффективными, они должны охватывать весь мир: эффективно предотвращать «гонку вооружений ИИ» или «гонку на понижение» в области безопасности и прав, выявлять инциденты, возникающие в результате принятия решений на протяжении жизненного цикла ИИ, которые охватывают множество юрисдикций, и реагировать на них, стимулировать обучение, поощрять взаимную совместимость и обеспечивать совместное использование преимуществ ИИ. Технология не имеет границ, и по мере ее распространения иллюзия того, что какое-то одно государство или группа государств могут (или должны) ее контролировать, будет ослабевать.

**xvii** Пробелы в координации между инициативами и учреждениями могут привести к тому, что мир будет разделен на зоны, где действуют разрозненные и не совместимые между собой режимы управления ИИ. Координация

отсутствует и в системе Организации Объединенных Наций. Хотя многие структуры Организации Объединенных Наций затрагивают вопросы управления ИИ, их конкретные мандаты означают, что ни одна из них не занимается этим комплексно.

**xviii** Однако представительства и координации действий недостаточно. Подотчетность требует практической реализации, с тем чтобы обязательства по глобальному управлению ИИ воплотились в осязаемые результаты на практике, в том числе в области развития потенциала и поддержки малых и средних предприятий с целью обеспечить общность возможностей. Многие из этого будут происходить на национальном и региональном уровнях, однако для устранения рисков и использования преимуществ необходимо также более активно действовать на глобальном уровне.

### 3. Укрепление глобального сотрудничества

**xix** Наши рекомендации продвигают целостную концепцию глобального сетевого, гибкого и динамичного подхода к управлению ИИ в интересах человечества, предполагающего наличие общего понимания, общих позиций и общих выгод. Только такой инклюзивный и всеобъемлющий подход к управлению ИИ может обеспечить принятие мер в связи с многогранными и меняющимися проблемами и возможностями, которые порождает ИИ в глобальном масштабе, способствуя при этом международной стабильности и справедливому развитию.

**xx** Руководствуясь принципами, изложенными в нашем промежуточном докладе, мы подготовили предложения, направленные на устранение пробелов и обеспечение согласованности в быстро формирующейся экосистеме международных мер и инициатив по управлению ИИ, что поможет избежать фрагментации усилий и упущенных возможностей. Для эффективной поддержки этих мер и налаживания партнерских отношений с другими учреждениями мы предлагаем создать облегченную и гибкую структуру, которая станет олицетворением согласованных усилий: подразделение по вопросам ИИ в составе Секретариата Организации Объединенных Наций, близкое к Генеральному секретарю, которое будет работать как «клей», эффективно и устойчиво объединяющий предложенные здесь инициативы.

#### А. Общее понимание

**xxi** Глобальный подход к управлению ИИ начинается с общего понимания его потенциала, возможностей, рисков и неопределенностей. Необходимы своевременные, беспристрастные и

достоверные научные знания и информация об ИИ, с тем чтобы государства–члены могли сформировать общее фундаментальное понимание во всем мире и сбалансировать информационную асимметрию между компаниями, в которых расположены дорогостоящие лаборатории ИИ, и остальным миром (в том числе путем обмена информацией между компаниями, развивающими ИИ, и более широким сообществом ИИ).

**xxii** Объединение научных знаний наиболее эффективно работает на глобальном уровне, позволяя осуществлять совместные инвестиции в глобальное общественное благо и сотрудничать в интересах общества, сводя воедино разрозненные и дублирующие друг друга усилия.

#### Международная научная группа по вопросам ИИ

**xxiii** Опираясь на такие прецеденты, как Межправительственная группа экспертов по изменению климата (МГЭИК) и Научный комитет Организации Объединенных Наций по действию атомной радиации, международная междисциплинарная научная группа по вопросам ИИ могла бы обобщать и стимулировать передовые исследования для информирования ученых, политиков, государств–членов и других заинтересованных сторон, желающих ознакомиться с научными взглядами на технологию ИИ или ее применение из беспристрастного, заслуживающего доверия источника.

**xxiv** Научная группа под эгидой Организации Объединенных Наций могла бы собирать экспертную информацию о возможностях, связанных с ИИ. Это может включать содействие «глубоким погружениям» в прикладные области ЦУР, такие как здравоохранение, энергетика, образование, финансы, сельское хозяйство, климат, торговля и занятость.

**xxv** Оценки рисков могут также опираться на работу других исследовательских инициатив в области ИИ, а Организация Объединенных Наций предлагает уникальную надежную «безопасную гавань» для обмена идеями о современном положении дел в данной области. Благодаря объединению разрозненных знаний, имеющихся в распоряжении стран и компаний, которые в противном случае могут не принять участие в обсуждениях, группа экспертов под эгидой Организацией Объединенных Наций может помочь исправить ошибочные представления и укрепить доверие во всем мире.

**xxvi** Такая группа должна работать независимо, при поддержке коллектива специалистов системы Организации Объединенных Наций, сформированного из представителей предлагаемого ниже подразделения по вопросам ИИ и соответствующих учреждений Организации Объединенных Наций, таких как Международный союз электросвязи (МСЭ) и Организация Объединенных Наций по вопросам образования, науки и культуры (ЮНЕСКО). Она должна поддерживать партнерские отношения с инициаторами исследовательских усилий, возглавляемых другими международными учреждениями, такими как Организация экономического сотрудничества и развития (ОЭСР) и Глобальное партнерство по искусственному интеллекту.

## Рекомендация 1

# международная научная группа по вопросам ИИ

Мы рекомендуем создать независимую международную научную группу по вопросам ИИ, которая состояла бы из различных междисциплинарных экспертов в этой области, работающих в своем личном качестве на добровольной основе. При поддержке предлагаемого подразделения Организации Объединенных Наций по вопросам ИИ и других соответствующих учреждений системы Организации Объединенных Наций, в партнерстве с другими соответствующими международными организациями, она бы руководствовалась мандатом, включающим в себя:

- a) публикацию ежегодного доклада с обзором возможностей, рисков и неопределенностей, связанных с ИИ, и с выявлением областей научного консенсуса в отношении технологических тенденций и сфер, где необходимы дополнительные исследования;
- b) подготовку ежеквартальных тематических исследовательских обзоров, посвященных областям, в которых ИИ может помочь в достижении ЦУР, с упором на области, представляющие общественный интерес, которые могут быть недостаточно охвачены; и
- c) выпуск специальных отчетов по возникающим вопросам, в том числе о появлении новых рисков или существенных пробелов в механизмах управления.

## В. Общие позиции

**xxvii** Наряду с общим пониманием ИИ нужны общие позиции для выработки совместимых подходов к управлению, основанных на глобальных нормах и принципах в интересах всех стран. Этого необходимо добиться на глобальном уровне, с тем чтобы предотвратить «гонку на понижение» в вопросах регулирования и уменьшить трансграничные регулятивные трения; максимально расширить возможности обучения и технической совместимости; и эффективно реагировать на вызовы, возникающие в связи с трансграничным характером ИИ.

### Политический диалог по вопросам управления ИИ

**xxviii** Необходим инклюзивный политический форум, с тем чтобы все государства-члены, опираясь на опыт заинтересованных сторон, могли обмениваться передовой практикой, которая была бы основана на правах человека и благоприятствовала развитию, способствовала бы взаимодействию подходов к управлению и учитывала бы трансграничные проблемы, требующие дальнейшего политического рассмотрения. Это не означает глобального управления всеми аспектами ИИ, но может заложить основу для международного сотрудничества и более эффективно увязать усилия отрасли и стран с глобальными нормами и принципами.

**xxix** Институционализация таких многосторонних обменов под эгидой Организации Объединенных Наций может обеспечить

надежную и инклюзивную площадку для обсуждения новых методов управления и соответствующих политических мер. Стимулируя выход за пределы зон комфорта, диалог между странами, позиции которых не совпадают, а также между государствами и заинтересованными сторонами может стимулировать обучение и заложить основы для более широкого сотрудничества, например, в области стандартов безопасности и прав, а также на период глобального кризиса. Участие Организации Объединенных Наций необходимо для того, чтобы закрепить эти усилия в виде как можно более широкого набора общих норм.

**xxx** В сочетании с наращиванием потенциала (см. рекомендации 4 и 5) такой инклюзивный диалог о подходах к управлению может помочь государствам и компаниям обновить свои нормативные подходы и методологии, с тем чтобы реагировать на ускоряющиеся темпы развития ИИ. Связи с международной научной группой могли бы усилить эту динамику, сравнимую с отношениями между МГЭИК и Конференцией Организации Объединенных Наций по изменению климата.

**xxxi** Политический диалог мог бы начаться на полях уже проводимых заседаний в Нью-Йорке (включая, например, Генеральную Ассамблею<sup>4</sup>) и Женеве. На одной из встреч, проводимых два раза в год, можно было бы уделять больше внимания возможностям в различных секторах, а на второй – рискам<sup>5</sup>. В будущем подобная встреча могла бы стать подходящим форумом для обмена информацией об инцидентах, связанных с ИИ, таких как те, которые исчерпывают или превышают возможности существующих учреждений.

4 По аналогии с политическим форумом высокого уровня в контексте ЦУР, который проводится под эгидой Экономического и Социального Совета.

5 Для выявления возможностей и рисков можно было бы привлечь к работе соответствующие подразделения системы Организации Объединенных Наций, в том числе МСЭ – по стандартам ИИ; МСЭ, Конференцию Организации Объединенных Наций по торговле и развитию (ЮНКТАД), Программу развития Организации Объединенных Наций (ПРООН) и Управление по координации деятельности в целях развития – по применению ИИ для достижения ЦУР; ЮНЕСКО – по этике и управленческому потенциал; Управление Верховного комиссара Организации Объединенных Наций по правам человека (УВКПЧ) – по вопросам подотчетности на основе существующих норм и стандартов; Управление по вопросам разоружения – по регулированию применения ИИ в системах военного назначения; ПРООН – по поддержке национального потенциала в целях развития; Форум по вопросам управления Интернетом – по многостороннему взаимодействию и диалогу; Всемирную организацию интеллектуальной собственности (ВОИС), Международную организацию труда (МОТ), Всемирную организацию здравоохранения (ВОЗ), Продовольственную и сельскохозяйственную организацию Объединенных Наций (ФАО), Всемирную продовольственную программу, Управление Верховного комиссара Организации Объединенных Наций по делам беженцев (УВКБ), ЮНЕСКО, Детский фонд Организации Объединенных Наций, Всемирную метеорологическую организацию и другие организации – по отраслевому применению и управлению.

**xxxii** Одна часть каждого раунда диалога могла бы быть посвящена национальным подходам под руководством государств-членов, а вторая – привлечению опыта и вкладов ключевых заинтересованных сторон, в частности технологических компаний и представителей гражданского общества. Помимо официальных заседаний в рамках диалога, взаимодействие многих заинтересованных сторон по вопросам

политики в области ИИ может опираться на другие существующие механизмы более специализированного характера, такие как совещания МСЭ по теме «ИИ во благо», ежегодные заседания Форума по вопросам управления Интернетом, Глобальный форум ЮНЕСКО по этике ИИ и «электронная неделя» Конференции Организации Объединенных Наций по торговле и развитию (ЮНКТАД).

*Рекомендация 2:*

## **политический диалог по вопросам управления ИИ**

Мы рекомендуем организовать межправительственный и многосторонний политический диалог по управлению ИИ на полях существующих заседаний в рамках Организации Объединенных Наций и проводить его раунды два раза в год. Его целью будет:

- a)** обмен передовым опытом в области управления ИИ, способствующим развитию при соблюдении, защите и реализации всех прав человека, включая использование возможностей, а также управление рисками;
- b)** содействие достижению общего понимания в отношении реализации мер по управлению ИИ разработчиками и пользователями из частного и государственного секторов в целях повышения международной совместимости механизмов управления ИИ;
- c)** добровольный обмен информацией о серьезных инцидентах, связанных с ИИ, которые создают значительную нагрузку на государственные учреждения или превышают их возможности в плане реагирования на них; и
- d)** по мере необходимости – обсуждение докладов международной научной группы по вопросам ИИ.

## Обмен стандартами в области ИИ

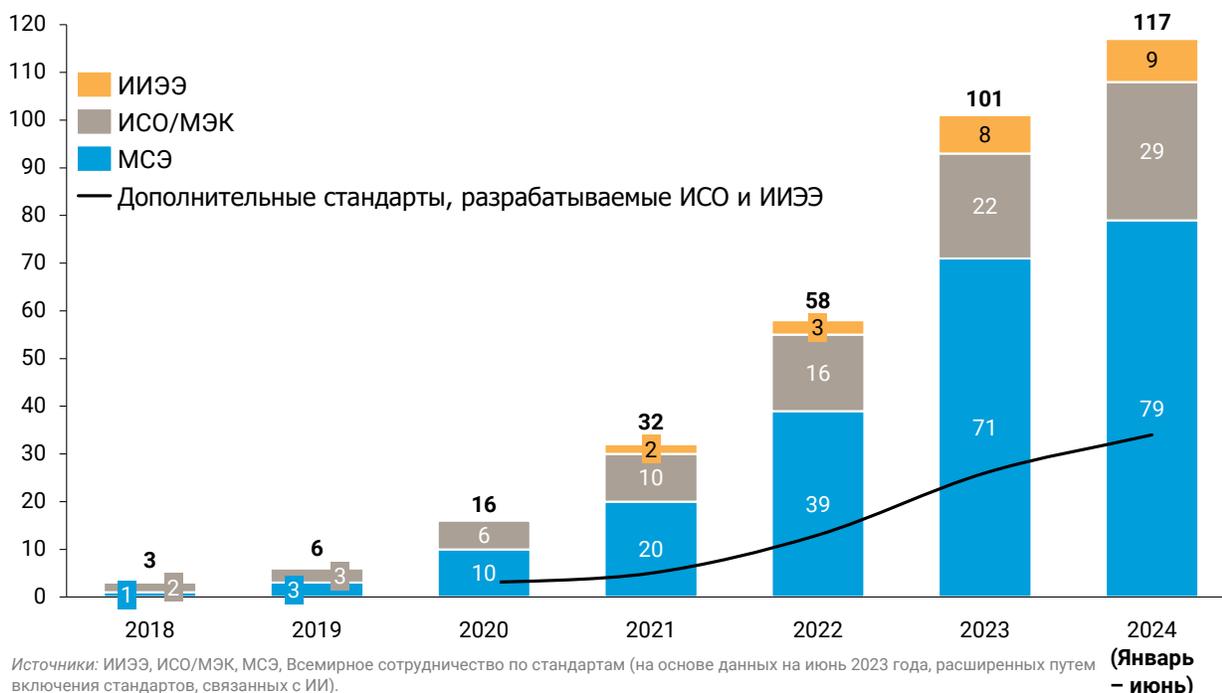
**xxxiii** На начальном этапе изучения систем ИИ существовало мало стандартов, помогающих ориентироваться в этой новой сфере и анализировать ее. В последнее время количество стандартов увеличилось. Рисунок б) иллюстрирует растущее число стандартов, принимаемых МСЭ, Международной организацией стандартизации (ИСО), the Международной электротехнической комиссией (МЭК) и Институтом инженеров по электротехнике и электронике (ИИЭЭ).

**xxxiv** У органов по стандартизации нет единой терминологии, и многие понятия, обычно используемые в отношении ИИ – справедливость, безопасность, прозрачность – не имеют согласованных определений.

Кроме того, существуют различия между стандартами, которые были приняты в узких технических целях или для внутренней валидации, и стандартами, которые призваны включить в себя более широкие этические принципы. На сегодняшний день мы имеем формирующийся набор стандартов, которые не основаны на общем понимании смысла или оторваны от ценностей, которые они призваны поддерживать.

**xxxv** Опираясь на опыт международной научной группы и привлекая членов различных национальных и международных организаций, которые внесли свой вклад в разработку стандартов, а также представителей технологических компаний и гражданского общества, система Организации Объединенных Наций могла бы выполнять роль информационного центра по стандартам ИИ, которые будут применяться во всем мире.

### Рисунок (б): Количество стандартов, связанных с ИИ



## Рекомендация 3

# обмен стандартами в области ИИ

Мы рекомендуем создать центр обмена стандартами ИИ, который объединял бы представителей национальных и международных организаций, занимающихся разработкой стандартов, технологических компаний, гражданского общества и представителей международной научной группы. Круг задач центра будет включать в себя:

- a) разработку и ведение реестра определений и применимых стандартов для анализа и оценки систем ИИ;
- b) обсуждение и оценку стандартов и процедур их разработки; и
- c) выявление пробелов, требующих разработки новых стандартов.

## С. Общие выгоды

**xxxvi** Повестка дня в области устойчивого развития на период до 2030 года с ее 17 ЦУР может придать четкость целям разработки, внедрения и использования ИИ, направив инвестиции на решение глобальных проблем развития. Без всеобъемлющего и инклюзивного подхода к управлению ИИ потенциал ИИ в плане позитивного вклада в достижение ЦУР может быть упущен, а его внедрение может непреднамеренно усилить или усугубить неравенство и предубеждения.

**xxxvii** ИИ не является панацеей для решения проблем устойчивого развития; это один из компонентов более широкого комплекса решений. Чтобы по-настоящему раскрыть потенциал ИИ в плане решения социальных проблем, необходимо сотрудничество между правительствами, научными кругами, представителями отрасли и гражданским обществом, с тем чтобы решения с использованием ИИ были инклюзивными и справедливыми.

**xxxviii** Во многом это зависит от доступа к талантам, вычислительным мощностям (или «вычислительным ресурсам») и данным таким образом, чтобы способствовать процветанию культурного и языкового разнообразия. Базовая инфраструктура и ресурсы для ее поддержания также являются необходимыми условиями.

**xxxix** Что касается талантов, то не каждому обществу нужны кадры специалистов по информатике для создания собственных моделей. Однако независимо от того, покупаются, заимствуются или создаются технологии, необходим базовый социотехнический потенциал для понимания возможностей и ограничений ИИ, а также для надлежащего использования возможностей ИИ с учетом специфических рисков.

**xi** Отсутствие вычислительных ресурсов – одно из самых больших препятствий для входа в сферу ИИ. Из 100 крупнейших высокопроизводительных вычислительных кластеров в мире, способных обучать

большие модели ИИ, ни один не расположен в развивающейся стране<sup>6</sup>. Нереально обещать доступ к вычислительным ресурсам, которые с трудом приобретают даже самые богатые страны и компании. Скорее мы стремимся заложить основу для преодоления разрыва в области ИИ для тех, кто не может получить необходимые средства обеспечения другими способами, в том числе путем поддержки инициатив по созданию распределенных и объединенных моделей разработки ИИ.

**xli** Что касается данных, то в контексте ИИ принято говорить о ненадлежащем использовании данных (например, о нарушении неприкосновенности частной жизни) или об упущениях в использовании данных (неспособность использовать существующие наборы данных). Однако есть и другая проблема – дефицит данных, с которой сталкивается большая часть земного шара, где ощущается их нехватка. Невозможность в полной мере отразить языковое и культурное разнообразие мира связывают с предвзятостью в системах ИИ, однако при этом она также может быть упущенной возможностью для этих сообществ получить доступ к преимуществам ИИ.

**xlii** Для поддержки инклюзивного и эффективного участия всех государств-членов в экосистеме ИИ необходим набор общих ресурсов, включая открытые модели, и здесь глобальные подходы имеют явные преимущества.

## Сеть по наращиванию потенциала

**xliii** Растущий государственный и частный спрос на человеческий и иной потенциал ИИ совпадает с появлением национальных, региональных и государственно-частных центров передового опыта в области ИИ, которые выполняют международные функции по наращиванию потенциала. Глобальная сеть может служить параллельной платформой, которая расширяет круг возможных партнеров и повышает операционную совместимость подходов к наращиванию потенциала.

**xliv** Начиная с целей в области развития, сформулированных в Декларации тысячелетия, и заканчивая ЦУР, Организация Объединенных Наций уже давно занимается наращиванием потенциала отдельных лиц и учреждений<sup>7</sup>. Сеть учреждений, связанных с Организацией Объединенных Наций, могла бы расширить возможности стран, стремящихся к партнерству в области наращивания потенциала. Она также могла бы стать катализатором создания новых национальных центров передового опыта для стимулирования развития местных инновационных экосистем ИИ на основе совместимых подходов, соответствующих нормативным обязательствам Организации Объединенных Наций.

6 Прокси-индикатор, поскольку большинство высокопроизводительных вычислительных кластеров не имеют графических процессоров (ГП) и малоприспособлены для продвинутого ИИ.

7 Благодаря деятельности ЮНЕСКО, ВОИС и других учреждений Организация Объединенных Наций помогает поддерживать богатое разнообразие культур и традиций получения знаний во всем мире. Университет Организации Объединенных Наций уже давно занимается наращиванием потенциала посредством проектов в сфере высшего образования и научных исследований, а Учебный и научно-исследовательский институт Организации Объединенных Наций помогает готовить должностных лиц в областях, имеющих решающее значение для устойчивого развития. Методология оценки готовности ЮНЕСКО является одним из ключевых инструментов для поддержки государств-членов в реализации Рекомендации ЮНЕСКО по этике искусственного интеллекта. Среди других примеров подобного рода – Академия ВОЗ в Лионе (Франция), Виртуальный институт ЮНКТАД, Программа стипендий Организации Объединенных Наций по разоружению под руководством Управления по вопросам разоружения, а также программы наращивания потенциала, реализуемые МСЭ и ПРООН.

Такая сеть способствовала бы развитию альтернативной парадигмы разработки технологий ИИ: организованной по принципу «снизу вверх», многопрофильной, открытой и основанной на сотрудничестве. В рамках усилий в национальном масштабе можно

было бы продолжать использовать инструменты диагностики, такие как Методология оценки готовности к ИИ ЮНЕСКО, для выявления недостатков на национальном уровне, а международная сеть помогала бы их устранять.

## Рекомендация 4:

### сеть по наращиванию потенциала

Мы рекомендуем создать сеть для наращивания потенциала в области ИИ, которая объединила бы ряд сотрудничающих центров наращивания потенциала, связанных с Организацией Объединенных Наций, и предоставила бы ключевым участникам экспертные знания, вычислительную технику и данные для обучения ИИ. Цель такой сети заключалась бы в следующем:

- a) стимулирование и координация региональных и глобальных усилий по наращиванию потенциала в области ИИ путем поддержки налаживания взаимосвязей между ними;
- b) укрепление возможностей государственных должностных лиц в области управления ИИ для содействия развитию при одновременном обеспечении уважения, защиты и реализации всех прав человека;
- c) предоставление услуг экспертов, вычислительных ресурсов и данных для обучения ИИ в нескольких центрах исследователям и социальным предпринимателям, стремящимся применять ИИ для решения местных задач в интересах общества, в том числе при помощи следующих инструментов:
  - i) протоколы, позволяющие междисциплинарным исследовательским группам и предпринимателям в условиях нехватки компьютеров получать доступ к вычислительным ресурсам для обучения/настройки и применения своих моделей в соответствии с местными условиями;
  - ii) учебно-экспериментальные центры для тестирования потенциальных решений в сфере ИИ и практического обучения;
  - iii) пакет онлайн-образовательных возможностей по ИИ, предназначенный для студентов университетов, молодых исследователей, социальных предпринимателей и должностных лиц государственного сектора; и
  - iv) программа стипендий для перспективных специалистов, которые смогут провести время в академических институтах или технологических компаниях.

## Глобальный фонд для ИИ

**xlvi** Многие страны сталкиваются с дефицитом финансов и ресурсов, который ограничивает их возможности в плане надлежащего и эффективного использования ИИ. Несмотря на все усилия по наращиванию потенциала (рекомендация 4), некоторые страны могут по-прежнему не иметь возможности получить доступ к обучению, вычислениям, моделям и учебным данным без международной поддержки. Другие усилия по финансированию также могут оказаться неэффективными без этой поддержки.

**xlvii** Предлагая фонд, мы не говорим о том, чтобы гарантировать доступ к передовым вычислительным ресурсам и возможностям. Ответ не всегда заключается в наращивании вычислительных мощностей. Нам также нужны более эффективные способы объединения талантов, вычислений и данных. Цель фонда – устранить основные пробелы в потенциале и сотрудничестве для тех, кто не может получить доступ к необходимым средствам, с тем чтобы:

- a. нуждающиеся страны смогли получить доступ к средствам поддержки ИИ, что позволило бы преодолеть разрыв в области ИИ;
- b. сотрудничество в области наращивания потенциала ИИ способствовало формированию привычек сотрудничества и смягчало геополитическую конкуренцию;
- c. страны с различными подходами к регулированию имели стимулы к разработке общих шаблонов для регулирования данных, моделей и приложений в целях решения возникающих на уровне общества проблем, связанных с ЦУР и научными прорывами.

**xlviii** Такая ориентация на общественные интересы делает фонд дополняющим предложением о создании сети для наращивания потенциала ИИ, в которую фонд также будет направлять ресурсы. Фонд обеспечил бы наличие независимых возможностей для мониторинга воздействия и смог бы привлекать и объединять неденежные взносы, в том

числе от организаций частного сектора, для предоставления учебных программ по ИИ, времени, вычислительной техники, моделей и наборов данных по ценам ниже рыночных. Таким образом мы сможем добиться того, чтобы огромные районы мира не остались в стороне, а получили возможность использовать ИИ для достижения ЦУР в различных условиях.

**xliv** В интересах всех было бы обеспечить в цифровом мире такой же уровень сотрудничества, как и в физическом. Можно провести аналогию с усилиями по борьбе с изменением климата, где затраты на переходный период, смягчение последствий или адаптацию не распределяются равномерно, и международная помощь имеет ключевое значение для оказания содействия странам с ограниченными ресурсами, с тем чтобы они могли присоединиться к глобальным усилиям по решению проблемы планетарного масштаба.

## Рекомендация 5

# глобальный фонд для ИИ

Мы рекомендуем создать глобальный фонд для ИИ, с тем чтобы преодолеть разрыв в области ИИ. Работая под управлением независимой руководящей структуры, фонд будет получать финансовые и неденежные взносы из государственных и частных источников и распределять их, в том числе через сеть для наращивания потенциала, в целях облегчения доступа к средствам ИИ, способствующим расширению возможностей на местах для достижения ЦУР, включая:

- a) совместное использование вычислительных ресурсов для обучения и отладки моделей разработчиками ИИ из стран, не имеющих достаточного местного потенциала или средств для его приобретения;
- b) учебно-экспериментальные центры и инструменты бенчмаркинга и тестирования для внедрения передового опыта разработки безопасных и надежных моделей и управления данными;
- c) решения в области управления, безопасности и операционной совместимости, имеющие глобальное применение;
- d) наборы данных и исследования способов объединения данных и моделей для реализации проектов, связанных с ЦУР; и
- e) хранилище моделей ИИ и специально подобранных наборов данных для ЦУР.

## Глобальная система данных ИИ

**i** Доступ к данным для обучения ИИ через рыночные или иные механизмы является важнейшим фактором, способствующим процветанию местных экосистем ИИ, особенно в странах, сообществах, регионах и демографических группах с «недостающими» данными (см. раздел «Общие выгоды» выше).

**ii** Только глобальные коллективные действия могут стимулировать взаимную совместимость, управление, сохранение конфиденциальности, расширение прав и возможностей и укрепление прав таким образом, чтобы способствовать «гонке на повышение» между юрисдикциями в направлении защиты прав человека и других

согласованных обязательств, доступности данных и справедливой компенсации субъектам данных при управлении сбором, созданием, использованием и монетизацией данных для обучения ИИ. Руководствуясь этой целью, мы предлагаем создать глобальную систему данных ИИ.

**iii** Такая система не приведет к созданию новых прав, связанных с данными. Скорее, она будет решать вопросы доступности, совместимости и использования данных для обучения ИИ. Она поможет выработать общее понимание того, как согласовать различные национальные и региональные системы защиты данных. Это также может способствовать процветанию местных экосистем ИИ, поддерживая культурное и языковое разнообразие, а также ограничивая дальнейшую экономическую концентрацию.

**liii** Эти меры могут быть дополнены продвижением идеи общности данных и положений о хостинге трастов данных в областях, имеющих отношение к ЦУР, на основе шаблонов соглашений о хранении данных и обмене ими на справедливой, безопасной и равноправной основе. Разработка этих шаблонов, а также фактическое хранение и анализ данных, хранящихся в общих базах или трастах, могут быть поддержаны по линии предлагаемой сети для наращивания потенциала и глобального фонда для ИИ (рекомендации 4 и 5).

**liv** Организация Объединенных Наций имеет уникальную возможность поддержать разработку глобальных принципов и практических механизмов управления данными для обучения ИИ и их использования в соответствии с согласованными международными обязательствами в области прав человека, интеллектуальной

собственности и устойчивого развития, опираясь на многолетнюю работу сообщества специалистов по данным и интегрируя ее с последними достижениями в области этики и управления ИИ. Это аналогично роли Комиссии Организации Объединенных Наций по праву международной торговли в развитии международной торговли путем разработки трансграничных рамочных механизмов, как связанных, так и не связанных с нормативно-правовыми аспектами.

**lv** Аналогичным образом, Комиссия по науке и технике в целях развития и Статистическая комиссия занимаются вопросами, которые связаны с данными для развития и данными по ЦУР. Всемирная организация интеллектуальной собственности (ВОИС) также рассматривает важные вопросы контента, авторского права и охраны знаний и выражений культуры коренных народов.

## Рекомендация 6

# глобальная система данных ИИ

Мы рекомендуем создать глобальную систему данных ИИ, сформировав ее в рамках процесса, инициированного соответствующим учреждением, таким как Комиссия Организации Объединенных Наций по праву международной торговли, и с учетом работы других международных организаций, для следующих целей:

- a) изложение связанных с данными определений и принципов глобального управления данными для обучения ИИ, в том числе на основе существующих передовых видов практики, и содействие культурному и языковому разнообразию;
- b) установление общих стандартов проверки и использования данных для обучения ИИ в целях обеспечения прозрачной и основанной на правах человека подотчетности в разных юрисдикциях; и
- c) создание формирующих рынок механизмов управления данными и обмена ими для обеспечения процветания местных экосистем ИИ в глобальном масштабе, включая, в частности:
  - i) трасты данных;
  - ii) эффективно управляемые глобальные рынки для обмена анонимными данными в целях обучения моделей ИИ; и
  - iii) типовые соглашения для облегчения международного доступа к данным и глобальной операционной совместимости, потенциально в качестве технико-юридических протоколов к рамочной программе.

## D. Согласованность усилий

**lvi** Вышеизложенные предложения направлены на устранение пробелов в представительстве, координации и реализации, выявленных в формирующемся международном режиме управления ИИ. Эти пробелы можно устранить с помощью партнерства и сотрудничества с существующими институтами и механизмами, с тем чтобы способствовать общности понимания, позиций и выгод.

**lvii** Тем не менее без специального координационного центра в Организации Объединенных Наций для поддержки и обеспечения согласования этих и других усилий мир не будет иметь инклюзивного сетевого, гибкого и согласованного подхода, необходимого для эффективного и справедливого управления ИИ как трансграничной, быстро меняющейся и универсальной технологией.

**lviii** Разрозненные нормы и институты, описанные в разделе «Глобальные пробелы в управлении ИИ» выше, отражают широко распространенное признание того, что управление ИИ является глобальной необходимостью. Неравномерность этой реакции требует определенных согласованных усилий.

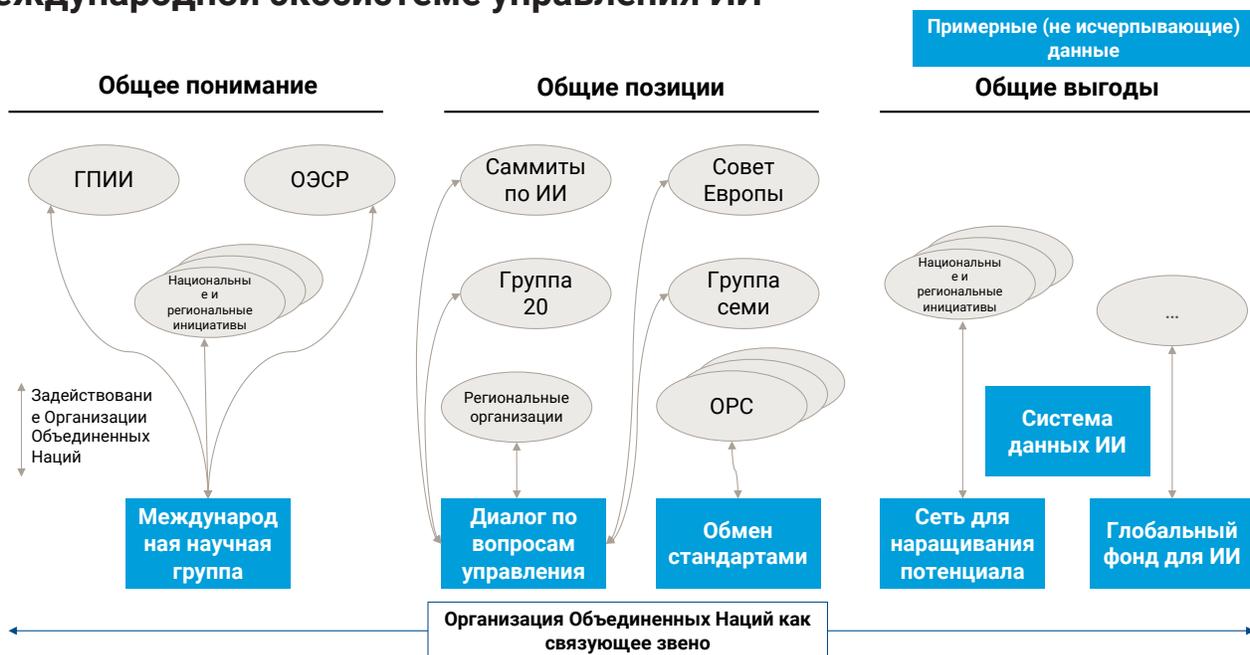
## Подразделение по вопросам ИИ в рамках Секретариата Организации Объединенных Наций

**lix** Поэтому мы предлагаем создать облегченный механизм, который будет служить «клеем», поддерживающим и стимулирующим предложения, содержащиеся в настоящем докладе, в том числе с помощью партнерств, а также позволит системе Организации Объединенных Наций выступать единым фронтом в развивающейся экосистеме управления ИИ.

**lx** Эта небольшая и гибкая структура в виде подразделения по вопросам ИИ в рамках Секретариата Организации Объединенных Наций будет подотчетна Генеральному секретарю, что, в свою очередь, откроет возможность для использования взаимосвязей в рамках всей системы Организации Объединенных Наций, не привязываясь к какой-либо ее части. Это важно ввиду неопределенности будущего ИИ и большой вероятности того, что он проникнет во все аспекты человеческой деятельности.

**lxi** Такой орган должен быть гибким, продвигать идею инклюзивности и оперативно налаживать партнерские отношения для ускорения координации и осуществления, используя в первую очередь существующие ресурсы и функции в рамках системы Организации Объединенных Наций. Основное внимание следует уделять применению ИИ в гражданских целях.

## Рисунок (с): Предлагаемая роль Организации Объединенных Наций в международной экосистеме управления ИИ



Сокращения: ГПИИ – Глобальное партнерство по искусственному интеллекту; ОЭСР – Организация экономического сотрудничества и развития; ОРС – организации по разработке стандартов.

**lxii** Этот орган мог бы быть частично укомплектован сотрудниками Организации Объединенных Наций, прикомандированными из специализированных учреждений и других подразделений системы Организации Объединенных Наций, таких как МСЭ, ЮНЕСКО, Управление Верховного комиссара Организации Объединенных Наций по правам человека (УВКПЧ), ЮНКТАД, Университет Организации Объединенных Наций и Программа развития Организации Объединенных Наций (ПРООН). Ему следует привлекать к работе различные заинтересованные стороны, включая компании, гражданское общество и научные круги, и работать в партнерстве с ведущими организациями, не входящими в систему

Организации Объединенных Наций (см. рис. с). Это позволит Организации Объединенных Наций наладить связи для достижения общности понимания, позиций и выгод в международной экосистеме управления ИИ.

**lxiii** Рекомендация 7 сформулирована на основе четкой оценки того, где Организация Объединенных Наций может принести пользу, и в том числе тех областей, где она может играть ведущую роль, где она может содействовать координации и где ей следует отойти в сторону. Она также учитывает преимущества существующих организационных механизмов, включая заранее согласованное финансирование и административные процессы, которые хорошо налажены и понятны.

## Рекомендация 7

# подразделение по вопросам ИИ в рамках Секретариата

Мы рекомендуем создать в рамках Секретариата подразделение по вопросам ИИ, подотчетное Генеральному секретарю. Оно должно быть облегченным и гибким по своей организационной структуре, по возможности опираясь на уже имеющиеся профильные структуры Организации Объединенных Наций. Выполняя роль «клея», скрепляющего и катализирующего предложения, которые содержатся в настоящем докладе, поддерживая партнерские отношения и взаимодействуя с другими процессами и институтами, это подразделение должно руководствоваться мандатом, включающим в себя:

- a)** поддержку предлагаемой международной научной группы, диалога по вопросам политики, обмена стандартами, сети для наращивания потенциала и, в случае необходимости, глобального фонда и глобальной системы данных ИИ;
- b)** проведение информационно-разъяснительной работы с различными заинтересованными сторонами, включая технологические компании, гражданское общество и научные круги, по возникающим проблемам ИИ; и
- c)** консультирование Генерального секретаря по вопросам, связанным с ИИ, в координации с другими соответствующими подразделениями системы Организации Объединенных Наций в целях обеспечения комплексного реагирования Организации Объединенных Наций.

## Е. Размышления на тему организационных моделей

**Ixiv** Дискуссии об ИИ часто переходят в крайности. В ходе наших консультаций по всему миру мы общались с теми, кто видит будущее безграничных благ, предоставляемых все более дешевыми и все более полезными системами ИИ. Мы также общались с теми, кто опасается мрачного будущего, разделения, безработицы и даже вымирания<sup>8</sup>.

**Ixv** Мы не знаем, какое будущее более вероятно – утопическое или антиутопическое. В то же время мы понимаем, что технология может пойти в направлении, которое устранил эту двойственность. В данном отчете рассматриваются возможности и риски ближайшего будущего, основанные на научных данных и фактах.

**Ixvi** Семь вышеизложенных рекомендаций позволяют нам с наибольшей вероятностью рассчитывать на то, что мы сможем воспользоваться преимуществами ИИ и одновременно минимизировать и смягчить имеющиеся риски, поскольку ИИ продолжает развиваться. Мы также помним о практических проблемах, связанных с созданием международных институтов в более широком масштабе. Именно поэтому мы предлагаем сетевой институциональный подход со структурно легкой и динамичной поддержкой. В том случае, если риски увеличатся, а цена возможностей возрастет, такие расчеты могут измениться.

**Ixvii** Мировые войны привели к созданию современной международной системы; разработка все более мощного химического, биологического и ядерного оружия привела к установлению режимов, ограничивающих его распространение и способствующих применению лежащих в его основе технологий в мирных целях. Развивающееся понимание

нашей общей человеческой природы привело к формированию современной системы прав человека и нашей неизменной приверженности делу достижения ЦУР в интересах всего человечества. Изменение климата превратилось из частной проблемы в глобальный вызов.

**Ixviii** Аналогичным образом, ИИ может подняться на уровень, требующий более значительных ресурсов и более широких полномочий, нежели предлагается в вышеупомянутых рекомендациях, с более сложными функциями разработки норм, осуществления, мониторинга, проверки и подтверждения, правоприменения, подотчетности, средств защиты от вреда и чрезвычайных мер. Поэтому имеет смысл поразмышлять над такими институциональными моделями. Заключительный раздел настоящего доклада призван внести свой вклад в эту работу.

## 4. Призыв к действиям

**Ixix** Мы по-прежнему с оптимизмом смотрим на будущее ИИ и его созидательный потенциал. Однако этот оптимизм находится в зависимости от реалистичного подхода к вопросу о рисках и неадекватности существующих структур и стимулов. Данная технология слишком важна, а ставки слишком высоки для того, чтобы полагаться только на рыночные силы и фрагментарную систему национальных и многосторонних действий.

**Ixx** Организация Объединенных Наций может стать каналом для нового общественного договора в сфере ИИ, который позволит добиться глобального согласия на режим управления, защищающий и расширяющий возможности всех нас. Такой общественный договор обеспечит справедливое распределение возможностей, а риски не будут перекладываться на плечи наиболее уязвимых слоев населения или же на будущие поколения, как это, к сожалению, происходит в случае с изменением климата.

8 См. <https://safe.ai/work/statement-on-ai-risk>.

**lxxi** Как группа и как отдельные люди, представляющие многие области знаний, организации и регионы мира, мы с нетерпением ждем продолжения этого важного разговора. Вместе со многими другими людьми, с которыми мы общались на этом

пути, и глобальным сообществом, которое они представляют, мы надеемся, что этот доклад внесет вклад в наши совместные усилия по управлению ИИ в интересах человечества.

## Рисунок (d): Заседание Консультативного органа высокого уровня по искусственному интеллекту в Сингапуре, 29 мая 2024 года



# 1. Введение

- 1 Консультативный орган высокого уровня Генерального секретаря по искусственному интеллекту был создан для анализа и выработки рекомендаций в отношении международного управления искусственным интеллектом (ИИ). Для членского состава Консультативного органа характерно широкое разнообразие с точки зрения географии, пола, специализации и возраста; мы привлекаем специалистов из правительств, гражданского общества, частного сектора и научных кругов. Интенсивные и многоплановые дискуссии позволили нам прийти к широкому согласию (которое нашло свое отражение в нашем промежуточном докладе<sup>1</sup>) относительно наличия глобального дефицита механизмов управления в сфере ИИ. В том докладе мы сформулировали руководящие принципы в отношении этой роли и тех функций, которые, возможно, потребуются выполнять на международном уровне.
- 2 В последующие месяцы полезный вклад в нашу работу внесли многочисленные отзывы и консультации. Они включали 18 «глубоких погружений» в конкретные проблемные области с участием более 500 экспертов, свыше 250 письменных материалов, поступивших от более чем 150 организаций и 100 частных лиц из всех регионов, оценку рисков, связанных с ИИ, с участием около 350 экспертов из всех регионов, выявление возможностей с участием около 120 экспертов из всех регионов, а также регулярные консультации и брифинги с государствами-членами, подразделениями Организации Объединенных Наций и другими группами заинтересованных сторон в рамках более чем 40 мероприятий во всех регионах<sup>2</sup>. Члены Консультативного органа также активно
- участвовали в форумах по всему миру, провели более сотни виртуальных дискуссий и три пленарных очных заседания – в Нью-Йорке, Женеве и Сингапуре.
- 3 Таким образом, у настоящего заключительного доклада много авторов. Хотя он не может отразить все богатство и разнообразие высказанных мнений, он свидетельствует о нашей общей приверженности обеспечению того, чтобы разработка, внедрение и использование ИИ осуществлялись на благо всего человечества, а также усилиям, направленным на эффективное и инклюзивное управление ИИ на международном уровне.
- 4 В этом докладе подтверждаются выводы промежуточного доклада Консультативного органа о возможностях и факторах, рисках и проблемах; в нем также подчеркивается необходимость глобального управления ИИ и излагаются семь рекомендаций.
- 5 В их число входит рекомендация относительно создания научной группы, призванная содействовать общему пониманию потенциала, возможностей, рисков и неопределенностей ИИ. На основе этого общего понимания нам нужны механизмы для выработки общих позиций относительно того, как ИИ должен регулироваться на международном уровне. Достижение этой цели зависит от регулярного диалога и разработки приемлемых для всех и общеприменимых стандартов.
- 6 В докладе также содержатся рекомендации относительно общих выгод, призванных обеспечить справедливое распределение преимуществ, связанных с ИИ, которое, в

1 См. <https://un.org/ai-advisory-body>.

2 Обзорную информацию о консультациях см. в приложении С.

свою очередь, может зависеть от доступа к моделям или возможностям, таким как таланты, вычислительные мощности (или «вычислительные ресурсы») и данные. Речь идет о сети для наращивания потенциала, глобальном фонде для ИИ и глобальной системе данных ИИ.

**7** Для того, чтобы активизировать эти усилия, наладить партнерство с другими инициативами и учреждениями в решении проблем и использовании возможностей ИИ и обеспечить единое мнение системы Организации Объединенных Наций по вопросам ИИ, мы предлагаем создать в рамках Секретариата Организации Объединенных Наций подразделение по вопросам ИИ.

**8** Хотя мы и рассматривали возможность рекомендовать создание международного агентства по ИИ, в настоящее время мы не рекомендуем предпринимать такой шаг; вместе с тем мы признаем необходимость того, чтобы управление шло в ногу с развитием технологий.

**9** Помимо непосредственных многосторонних обсуждений и процессов с участием правительств, наш доклад также предназначен для гражданского общества и частного сектора, исследователей и заинтересованных лиц во всем мире. Мы прекрасно понимаем, что достижение намеченных нами амбициозных целей возможно только при многостороннем глобальном участии.

**10** В целом мы считаем, что будущее этой технологии по-прежнему не предопределено. Такая точка зрения подтвердилась в ходе нашего «глубокого погружения» в тему, связанную с вектором развития данной технологии, и дебатов между сторонниками «открытого» и «закрытого» подходов к ее развитию (см. вставку 9). Более крупные и мощные модели, разрабатываемые все меньшим числом корпораций, – это один из вариантов будущего. Альтернативным

вариантом может стать более разнообразный глобальный инновационный ландшафт, в котором доминируют взаимодействующие между собой малые и средние модели ИИ, предполагающие возможность многопланового применения в социальной и экономической областях. Наши рекомендации направлены на то, чтобы сделать последнее более вероятным, признавая при этом имеющиеся риски.

**11** С момента своего основания Организация Объединенных Наций была привержена делу содействия экономическому и социальному прогрессу всех народов<sup>3</sup>. Цели в области развития, сформулированные в Декларации тысячелетия, ставили амбициозные задачи по обеспечению экономических возможностей для всех людей в мире, а принятые позднее цели в области устойчивого развития (ЦУР) были призваны примирить необходимость развития с экологическими ограничениями нашей планеты. Расширение разработки, внедрения и использования инструментов и систем искусственного интеллекта ставит перед нами следующую серьезную задачу: обеспечить, чтобы мы вместе встретили наше цифровое будущее, а не увеличили существующий цифровой разрыв.

**12** Инклюзивное управление ИИ – это, пожалуй, одна из самых сложных задач в области управления, с которыми будет сталкиваться Организация Объединенных Наций. Существует несоответствие между доминирующей ролью частного сектора в ИИ и Вестфальской системой международной политики. В период острой геополитической конкуренции государства стремятся воспользоваться потенциалом ИИ для достижения могущества и процветания. Многие общества все еще находятся на периферии разработки, внедрения и использования ИИ, в то время как некоторые из них охвачены волнением, смешанным с беспокойством по поводу повсеместного воздействия ИИ.

3 Включая торговлю, прямые иностранные инвестиции и передачу технологий в качестве необходимых предпосылок долгосрочного развития.

**13** Несмотря на возникающие трудности, никто не может оставаться в стороне. Ставки просто слишком высоки для Организации Объединенных Наций, ее государств-членов и более широкого сообщества, чьи чаяния представляет Организация Объединенных Наций. Мы надеемся, что этот доклад в определенной мере послужит ориентиром в наших совместных усилиях по управлению ИИ в интересах всего человечества.

## А. Возможности и предпосылки

**14** ИИ преобразует тот мир, в котором мы живем. Этот комплекс технологий<sup>4</sup> обладает огромным потенциалом для достижения благих целей: от открытия новых областей научных

исследований (см. вставку 1) и оптимизации энергосетей до улучшения здоровья населения и сельского хозяйства<sup>5</sup>. Если будут реализованы потенциальные возможности, открывающиеся благодаря использованию инструментов ИИ для отдельных людей, секторов экономики, научных исследований и других областей, представляющих общественный интерес, это может сыграть важную роль в стимулировании экономики наших стран (см. вставку 2), а также изменить наши общества к лучшему. Использование ИИ в общественных интересах – например, для прогнозирования и решения проблем пандемий, наводнений, лесных пожаров и отсутствия продовольственной безопасности – может даже способствовать прогрессу в достижении ЦУР.

### Вставка 1: Потенциал ИИ в плане развития науки

Опираясь на преобразующее наследие Интернета, ИИ вполне может стать следующим крупным прорывом в сфере научного прогресса. Всемирная паутина облегчила ученым обмен огромными объемами экспериментальных данных, научных работ и документации. ИИ опирается на этот фундамент, позволяя анализировать обширные массивы данных, выявлять скрытые закономерности, строить новые гипотезы и ассоциации и ускорять темпы открытий, в том числе путем проведения масштабных экспериментов с помощью автоматизированной робототехники.

Влияние ИИ на науку распространяется на все основные дисциплины. От биологии до физики, от экологии до общественных наук, ИИ внедряется в исследовательские процессы и ускоряет генерирование научных знаний. Некоторые из тех заявлений, которые делаются сегодня, возможно, являются преувеличенными, тогда как другие уже подтвердились на практике, и их долгосрочный потенциал представляется многообещающим<sup>а</sup>.

<sup>а</sup> См. John Jumper and others, «Highly accurate protein structure prediction with AlphaFold», *Nature*, vol. 596 (July 2021), pp. 583–589; см. Также Josh Abramson and others, «Accurate structure prediction of biomolecular interactions with AlphaFold 3», *Nature*, vol. 630, pp. 493–500 (May 2024).

<sup>4</sup> По данным Организации экономического сотрудничества и развития (ОЭСР), «система ИИ – это машинная система, которая для достижения явных или неявных целей на основе полученных данных определяет, как генерировать выходные данные, такие как прогнозы, контент, рекомендации или решения, которые могут влиять на физическую или виртуальную среду. Различные системы ИИ отличаются друг от друга по уровню автономности и адаптивности после развертывания» (см. <https://oecd.ai/en/wonk/ai-system-definition-update>).

<sup>5</sup> Вместе с тем мы считаем, что для оценки заявлений о пользе ИИ необходим тщательный анализ, проведенный экспертами в данной области. Стремление использовать ИИ во благо должно основываться на научных данных и тщательной оценке компромиссов и альтернатив. В дополнение к научным исследованиям трансформируются и общественные науки.

## Вставка 1: Потенциал ИИ в плане развития науки (продолжение)

Например, в биологии с помощью ИИ была решена существующая уже 50 лет проблема сворачивания белков и прогнозирования их структуры. В частности, спрогнозирована структура более 200 млн различных белков, а полученной базой данных с открытым доступом на момент написания статьи пользуются свыше 2 млн ученых более чем в 190 странах, многие из которых работают над забытыми болезнями. С тех пор эта технология была распространена и на другие биомолекулы жизни – ДНК, РНК, лиганды и их взаимодействия.

Для болезни Альцгеймера, болезни Паркинсона и бокового амиотрофического склероза (БАС) эксперты с помощью искусственного интеллекта определяют биомаркеры заболевания и прогнозируют реакцию на лечение, значительно повышая точность и скорость диагностики и составления плана лечения<sup>b</sup>. В целом ИИ помогает развивать точную медицину (например, при нейродегенеративных заболеваниях), подбирая лечение на основе генетических и клинических профилей. Технологии ИИ также помогают ускорить открытие и разработку новых химических соединений<sup>c</sup>.

В радиоастрономии скорость и масштаб данных, собираемых современными приборами, такими как квадратно-километровая решетка, могут превзойти традиционные методы. ИИ может изменить ситуацию, в том числе помочь выбрать, на какой части данных сосредоточиться для получения новых знаний. Благодаря «неконтролируемой кластеризации» ИИ может выявлять закономерности в данных без указания того, что именно нужно искать<sup>d</sup>. Применение ИИ в исследованиях в области общественных наук также может дать глубокое понимание сложной человеческой динамики, расширяя наши представления о социальных тенденциях и экономическом развитии.

Со временем, благодаря беспрецедентному уровню междисциплинарности, ИИ может разрабатываться и внедряться для создания новых научных областей, подобно тому, как биоинформатика и нейроинформатика возникли в результате интеграции вычислительных методов с исследованиями в сфере биологии и неврологии. Способность ИИ интегрировать и анализировать разнообразные массивы данных из таких областей, как изменение климата, продовольственная безопасность и здравоохранение, при ответственном подходе может открыть возможности для исследований, которые соединят эти традиционно разделенные области.

ИИ также может усилить воздействие научных исследований на государственную политику, позволяя подтверждать сложные гипотезы, например, объединяя климатические модели с данными о сельском хозяйстве для прогнозирования рисков продовольственной безопасности и увязывая эти данные с результатами усилий в сфере здравоохранения. Еще одна перспектива – активизация гражданской науки и использование местных знаний и данных для решения глобальных проблем.

b Isaias Ghebrehiwet and others, «Revolutionizing personalized medicine with generative AI: a systematic review», *Artificial Intelligence Review*, vol. 57, No. 127 (April 2024).

c Amil Merchant and others, «Scaling deep learning for materials discovery», *Nature*, vol. 624, pp. 80–85 (November 2023).

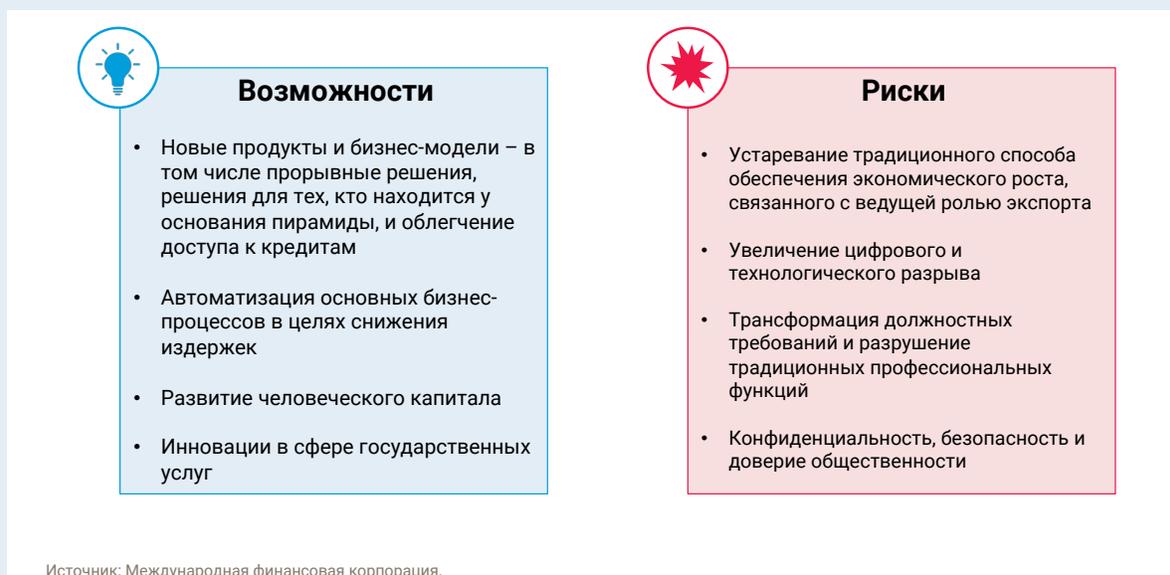
d Zack Savitsky, «Astronomers are enlisting AI to prepare for a data downpour», *MIT Technology Review*, 20 May 2024.

## Вставка 2: Экономические возможности ИИ

Со времен промышленной революции всего несколько инноваций резко ускорили темпы экономического прогресса. Эти внедренные на ранних этапах «технологии общего назначения» изменили множество секторов и отраслей. Последние значительные изменения произошли с появлением компьютеров и наступлением цифровой эпохи. Эти технологии изменили экономику и повысили производительность труда во всем мире, но в полной мере их воздействие проявилось только через несколько десятилетий.

Генеративный ИИ ломает тенденцию медленного внедрения. Эксперты считают, что его преобразующий эффект будет заметен уже в этом десятилетии. Такая стремительная интеграция означает, что новые разработки в области ИИ могут быстро перестроить отрасли, изменить рабочие процессы и повысить производительность. Таким образом, оперативное внедрение ИИ может беспрецедентно изменить нашу экономику и общество.

### Рисунок 1: Некоторые возможности и риски для развития, связанные с ИИ, в условиях формирующихся рынков



Экономические выгоды от использования ИИ могут быть значительными. Хотя трудно предсказать все последствия воздействия ИИ на сложную экономику наших стран, прогнозы показывают, что ИИ может ощутимо увеличить мировой валовой внутренний продукт, что окажет соответствующее влияние практически на все отрасли. Предприятиям, особенно микро- и малым и средним, ИИ может открыть доступ к передовым инструментам аналитики и автоматизации, которые раньше были доступны только крупным корпорациям. Широкая применимость ИИ позволяет предположить, что ИИ может стать технологией общего назначения. Таким образом, ИИ может повысить производительность труда отдельных людей, малых и крупных предприятий и других организаций в таких разнообразных секторах, как розничная торговля, производство и операционная деятельность, здравоохранение и государственный сектор, в развитых и развивающихся странах<sup>a</sup>.

<sup>a</sup> James Manyika and Michael Spence, «The coming AI economic revolution: can artificial intelligence reverse the productivity slowdown?», Foreign Affairs, 24 October 2023.

## Вставка 2: Экономические возможности ИИ (продолжение)

Это потребует широкого внедрения ИИ как на внутриотраслевом, так и на межотраслевом уровнях; его применения в целях повышения производительности; а также разработки моделей ИИ, повышающих производительность труда и открывающих новые возможности для широкомасштабной экономической деятельности. Кроме того, это потребует инвестиций и повышения капиталовооруженности, совместных инноваций, изменений в том, что касается процессов и организации, а также готовности трудовых ресурсов и стимулирующей политики.

Тем не менее, хотя ИИ способен повысить производительность труда, стимулировать международную торговлю и увеличить доходы, ожидается, что он также повлияет на работу. Исследования показывают, что в одних случаях ИИ может оказать помощь работникам, а в других – вытеснить их с рабочих мест<sup>b</sup>. Исследования, в том числе проведенные Международной организацией труда (МОТ), показывают, что в обозримом будущем ИИ, скорее всего, будет в большей степени помогать работникам, а не вытеснять их<sup>c</sup>.

Исследования также показали, что вытеснение рабочих мест, когда оно происходит, по всей вероятности, будет происходить по-разному в экономиках, находящихся на разных стадиях развития<sup>d</sup>. Хотя страны с развитой экономикой подвержены большему риску, они также лучше подготовлены к использованию ИИ и дополнению своих трудовых ресурсов. Страны с низким и средним уровнем дохода могут иметь меньше возможностей для использования этой технологии. Кроме того, интеграция ИИ в сферу занятости может непропорционально повлиять на определенные демографические группы: в некоторых секторах женщины могут столкнуться с более высоким риском вытеснения рабочих мест.

Без целенаправленных и скоординированных усилий по преодолению цифрового разрыва потенциальная способность ИИ использоваться для поддержки устойчивого развития и борьбы с нищетой не будет реализована, в результате чего значительные слои населения планеты останутся в неблагоприятном положении в стремительно меняющейся технологической среде, что приведет к усугублению существующего неравенства.

Для успешной интеграции ИИ в мировую экономику нам необходимо эффективное управление, которое позволит управлять рисками и обеспечит справедливые результаты. Это, в частности, подразумевает создание учебно-экспериментальных центров нормативной направленности для тестирования систем ИИ, развитие международного сотрудничества по стандартам и создание механизмов для постоянной оценки влияния ИИ на рынки труда и общество. Помимо продуманных национальных стратегий в области ИИ и международной поддержки, необходимо предпринять усилия по следующим направлениям.

- Развитие навыков: реализация учебных и образовательных программ, направленных на развитие навыков работы с ИИ у всего трудового коллектива – от базовой цифровой грамотности до передовых технических знаний – для подготовки работников к будущему, дополненному ИИ.
- Цифровая инфраструктура: значительные инвестиции в цифровую инфраструктуру, особенно в развивающихся странах, для преодоления разрыва в области ИИ и содействия широкому внедрению ИИ.

b Erik Brynjolfsson and others, «Generative AI at work», National Bureau of Economic Research, working paper 31161, 2023; см. также Shakked Noy and Whitney Zhang, «Experimental evidence on the productivity effects of generative artificial intelligence», Science, vol. 381, No. 6654, pp. 187–192 (July 2023).

c Pawel Gmyrek and others, Generative AI and Jobs: A Global Analysis of Potential Effects on Job Quantity and Quality (Geneva: ILO, 2023).

d Mauro Cazzaniga and others, «Gen-AI: artificial intelligence and the future of work», staff discussion note SDN2024/001 (Washington, D.C.: International Monetary Fund, 2024).

## Вставка 2: Экономические возможности ИИ (продолжение)

- Интеграция на рабочих местах: использование общественного диалога и государственно-частных партнерств для управления интеграцией ИИ на рабочих местах, обеспечения участия работников в этом процессе и защиты трудовых прав.
- Соображения, связанные с цепочкой создания стоимости: обеспечение достойных условий труда по всей цепочке создания стоимости в сфере ИИ, включая часто упускаемые из виду области, такие как аннотирование данных и модерация контента, для справедливого развития ИИ.

## В. Основные предпосылки использования ИИ в интересах человечества

**15** Потенциальные возможности, возникающие в результате развития и применения ИИ, не обязательно будут освоены или реализованы на справедливой основе. В мае 2024 года анализ финансирования проектов в области ИИ для содействия прогрессу в достижении ЦУР показал, что только 10 процентов выделенных субсидий получили организации, базирующиеся в странах с низким или средним уровнем дохода; для частного капитала этот показатель составил 25 процентов (более 90 процентов из них – в Китае)<sup>6</sup>.

## С. Управление как одна из ключевых предпосылок

**16** Если мы стремимся к тому, чтобы преимущества ИИ были полностью реализованы и получены не только несколькими людьми в небольшом количестве стран, необходимо создать соответствующие механизмы на глобальном уровне. Для того, чтобы ИИ использовался на общее благо, а его возможности распределялись справедливо, потребуются правительственные и межправительственные меры по стимулированию участия частного сектора, научных кругов и гражданского общества. Любая система управления должна формировать стимулы на глобальном уровне, с тем чтобы способствовать достижению более масштабных и всеохватывающих целей, а также помогать выявлять параметры компромиссов и действовать соответствующим образом.

<sup>6</sup> «Анализ информации о местонахождении штаб-квартир получателей субсидий из базы данных фондов, расположенных в Соединенных Штатах, показывает, что с 2018 по 2023 годы только 10 процентов субсидий, которые выделялись на инициативы ИИ, направленные на достижение одной или нескольких ЦУР, получили организации, базирующиеся в странах с низким или средним уровнем дохода... Анализ данных, касающихся частного капитала, показывает, что 36 процентов из 9000 компаний, занимающихся ЦУР, имеют штаб-квартиры в Соединенных Штатах, но при этом такие компании получили 54 процента общего объема финансирования. Мы также обнаружили, что, хотя 20 процентов из 9000 компаний, занимающихся вопросами ЦУР, имеют штаб-квартиры в странах с низким или средним уровнем дохода, они получили более значительную долю (25 процентов) общего объема финансирования. Одна из причин этого заключается в том, что существенную долю инвестиций получают китайские компании ... Остальные развивающиеся страны в выборке получили только 3 процента финансирования, при этом составляя 7 процентов выборки» (Medha Bankhwal and others, «AI for social good: improving lives and protecting the planet», McKinsey & Company, May 2024).

## D. Риски и вызовы

- 17** Разработка, внедрение и использование ИИ влекут за собой риски, которые могут охватывать одновременно множество областей. Мы концептуализируем риски, связанные с ИИ, через призму уязвимостей; это предлагает способ формирования политических повесток, основанный на учете уязвимостей.
- 18** Вызовы для традиционных систем регулирования возникают из-за скорости, непрозрачности и автономности ИИ. Ускорение технической разработки и внедрения ИИ также повышает ставки для международного управления, поскольку его универсальный характер имеет трансграничные последствия для нескольких областей одновременно.

## E. Риски, связанные с ИИ

- 19** Такие проблемы, как предвзятость в системах ИИ и неправомерное наблюдение с помощью ИИ, все чаще фиксируются в документах. Использование передовых моделей ИИ сопряжено с другими рисками, такими как конфабляция больших языковых моделей, интенсивное потребление ресурсов и риски для мира и безопасности. Дезинформация, генерируемая ИИ, угрожает демократическим институтам.
- 20** Составление всеобъемлющего списка рисков ИИ на все времена не имеет большого смысла, если учесть повсеместный и быстро развивающийся характер ИИ и его применения; мы считаем, что полезнее рассматривать риски с точки зрения уязвимых сообществ, а также общего достояния (см. пункты 26–28 ниже).

**21** О текущем восприятии рисков экспертами можно судить по результатам обзорного исследования, проведенного в связи с нашей работой («Глобальная оценка рисков, связанных с ИИ»; см. приложение E), в ходе которого мнения о тенденциях и рисках, связанных с ИИ, были получены от 348 экспертов по ИИ в различных дисциплинах и 68 странах во всех регионах<sup>7</sup>. В целом 7 из 10 опрошенных экспертов обеспокоены или очень обеспокоены тем, что вред (существующий или новый) от ИИ станет значительно более серьезным и/или широко распространенным в ближайшие 18 месяцев (см. приложение E).

- 22** Из списка примеров областей риска, связанных с ИИ<sup>8</sup>, большинство экспертов были обеспокоены или очень обеспокоены вредом, связанным со следующими факторами (см. также рис. 2):
- социальными последствиями ИИ:  
**78 процентов** – ущерб целостности информации [вопрос j], **74 процента** – неравенство, включая, в частности, концентрацию богатства и власти в одних руках [вопрос h], **67 процентов** – дискриминация / лишение прав, особенно среди маргинализированных сообществ [вопрос i];
  - преднамеренным использованием ИИ во вред другим: **75 процентов** – использование государственными субъектами в вооруженных конфликтах [вопрос b], **72 процента** – злонамеренное использование негосударственными субъектами [вопрос a] и **65 процентов** – использование государственными субъектами в ущерб отдельным лицам [вопрос c].

<sup>7</sup> Список респондентов был составлен на основе данных Канцелярии Посланника Генерального секретаря по вопросам технологий (КПГСТ) и сетей Консультативного органа, включая участников «глубоких погружений». В период проведения исследования регулярно приглашались дополнительные эксперты для улучшения представительности выборки. Итоговая выборка n=348 представляет собой сильную, сбалансированную глобальную выборку респондентов, обладающих соответствующим опытом и знаниями, которые позволяют составить обоснованное мнение о рисках ИИ (методологию см. в приложении E).

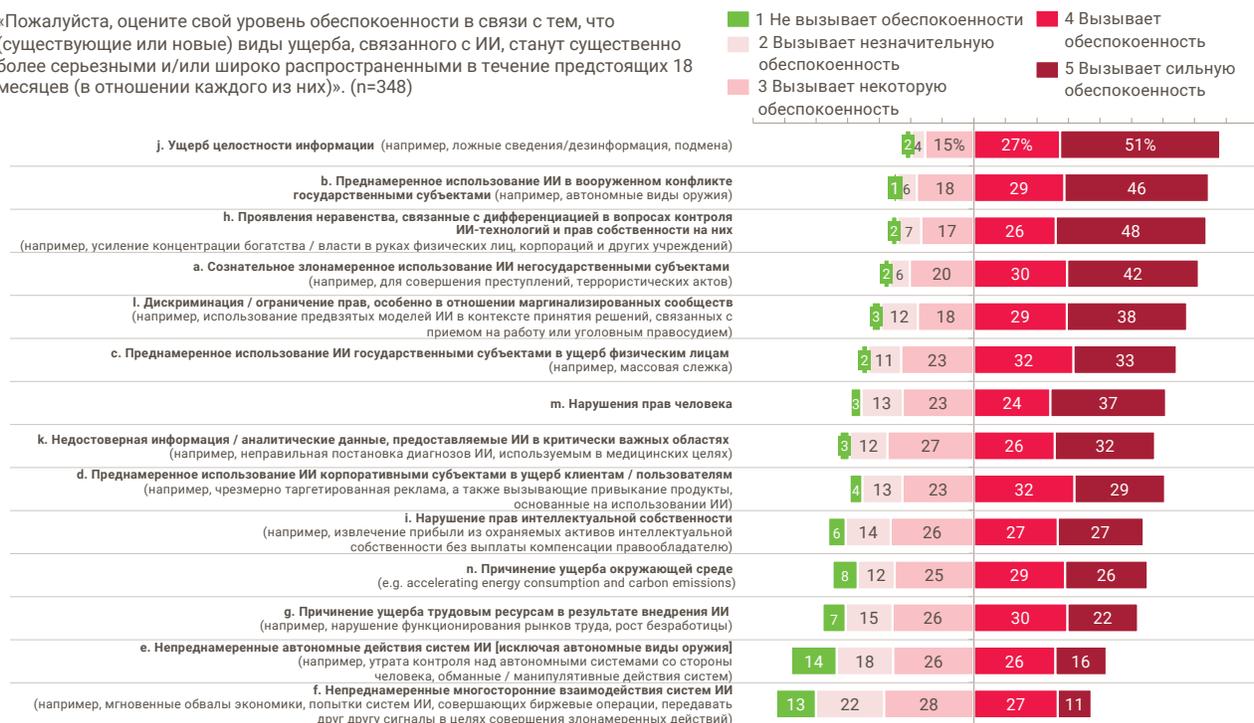
<sup>8</sup> Список был построен с учетом классификации рисков на основе уязвимостей, приведенной во вставке 4, более ранняя версия которой была представлена в нашем промежуточном докладе. представлена в нашем промежуточном докладе.

**23** Во всех областях риска, кроме двух, большинство опрошенных экспертов по ИИ были обеспокоены или очень обеспокоены возможностью реального нанесения ущерба. Хотя менее половины экспертов выразили такую обеспокоенность в отношении непреднамеренного ущерба от ИИ [вопросы

е и f], каждый шестой из тех, кто был очень обеспокоен непреднамеренным ущербом от ИИ, отметил, что, по их мнению, к 2025 году самое неожиданное или значительное влияние на риски, связанные с ИИ, окажут автономные системы.<sup>9</sup>

## Рисунок 2: Уровень обеспокоенности экспертов по поводу рисков, связанных с ИИ, в различных областях

«Пожалуйста, оцените свой уровень обеспокоенности в связи с тем, что (существующие или новые) виды ущерба, связанного с ИИ, станут существенно более серьезными и/или широко распространенными в течение предстоящих 18 месяцев (в отношении каждого из них)». (n=348)



Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

<sup>9</sup> Вопрос: «Как вы думаете, какие новые тенденции сегодня могут оказать наиболее неожиданное и/или значительное влияние на риски, связанные с ИИ, в течение следующих 18 месяцев?»

**24** Мнения экспертов разошлись, в том числе в зависимости от того региона, который они представляли, и их пола (более подробные результаты см. в приложении E). Это указывает на важное значение инклюзивного представительства при проведении мероприятий по выявлению общих рисков. Несмотря на различия, результаты показали обеспокоенность по поводу ущерба, который может быть нанесен ИИ в ближайший год, что свидетельствует о том, что эксперты считают необходимым в ближайшем будущем заняться рисками в различных областях и уязвимых местах.

**25** Кроме того, использование автономного оружия в контексте вооруженных конфликтов, преступности или терроризма, и в частности использование ИИ для обеспечения общественной безопасности, вызывает серьезные вопросы, связанные с правовыми и гуманитарными моментами, а также с безопасностью (см. вставку 3)<sup>10</sup>.

**26** Однако управление рисками не ограничивается их перечислением или определением приоритетов. Подход к рискам на основе уязвимостей может сместить акцент в политических программах с «чего» каждого риска (например, «риск для безопасности») на то, «кто» подвергается риску и «где», а также на то, кто должен нести ответственность в каждом случае.

### **Вставка 3: ИИ и вопросы национальной и международной безопасности**

Многие технологии ИИ не просто имеют двойное назначение, но и по своей сути являются «многоцелевыми». Применение ИИ в правоохранительной деятельности и пограничном контроле становится все более распространенным и вызывает озабоченность в отношении надлежащей правовой процедуры, слежки и отсутствия подотчетности в отношении обязательств государств по соблюдению норм в области прав человека, закрепленных во Всеобщей декларации прав человека и других документах.

Среди проблем использования ИИ в военной сфере – новая гонка вооружений, снижение порога конфликта, размывание границ между войной и миром, распространение соответствующих возможностей среди негосударственных субъектов и отступление от давно устоявшихся принципов международного гуманитарного права, таких как военная необходимость, разграничение, пропорциональность и ограничение ненужных страданий. По юридическим и моральным соображениям решения об убийстве не должны автоматизироваться с помощью искусственного интеллекта. Государства должны взять на себя обязательство воздерживаться от развертывания и использования военных систем ИИ в вооруженных конфликтах способами, которые не будут в полной мере соответствовать международному праву, включая международное гуманитарное право и международные нормы в области прав человека.

В настоящее время 120 государств-членов поддерживают новый договор по автономному оружию, и как Генеральный секретарь, так и президент Международного комитета Красного Креста призвали завершить переговоры по такому договору к 2026 году. Консультативный орган настоятельно призывает государства-члены последовать этому призыву.

<sup>10</sup> Этот список носит исключительно иллюстративный характер и затрагивает лишь некоторые из рисков, с которыми сталкиваются люди и общества.

## Вставка 3: ИИ и вопросы национальной и международной безопасности (продолжение)

Консультативный орган считает необходимым определить четкие «красные линии», определяющие случаи незаконного использования, включая использование ИИ для автономного выбора и поражения целей. Опираясь на существующие обязательства по оценке оружейных систем в международном гуманитарном праве, государства должны требовать от производителей оружия, чтобы они, используя контрактные обязательства и другие средства, проводили правовые и технические обзоры для предотвращения неэтичного проектирования и разработки военных систем ИИ. Кроме того, государствам следует разработать порядок проведения правовых и технических обзоров использования ИИ, а также оружия и средств ведения войны и обмениваться соответствующим передовым опытом.

Государствам также необходимо выработать общее понимание в отношении механизмов тестирования, оценки, проверки и валидации систем ИИ в сфере безопасности и военной сфере. Им следует сотрудничать в целях наращивания потенциала и распространения знаний путем обмена передовым опытом и содействия ответственному управлению жизненным циклом систем ИИ в сфере безопасности и военной области. Для предотвращения приобретения мощных и потенциально автономных систем ИИ опасными негосударственными субъектами, такими как преступные или террористические группировки, государствам следует создать соответствующие механизмы контроля и процессы на протяжении всего жизненного цикла систем ИИ, включая управление процессами завершения жизненного цикла (т. е. вывода из эксплуатации) военных систем ИИ.

Для целей обеспечения транспарентности можно было бы создать «консультативные советы» в целях предоставления независимых экспертных консультаций и осуществления контроля на протяжении всего жизненного цикла применения ИИ в сфере безопасности и военных целях. Представители отрасли и другие субъекты должны рассмотреть механизмы предотвращения неправомерного использования технологий ИИ в злонамеренных целях, а также их непреднамеренного использования в военных целях.

- 27** Это весьма важно, поскольку меняющиеся риски по-разному проявляются для разных людей и обществ. Подход, основанный на уязвимостях, также предложенный в нашем промежуточном докладе, предлагает открытую основу для сосредоточения внимания на тех, кому может навредить ИИ, что может стать основой для динамичного управления рисками (см. вставку 4).
- 28** Политическая значимость рассмотрения рисков, связанных с ИИ, через призму уязвимостей иллюстрируется анализом вопросов управления ИИ с точки зрения конкретной уязвимой группы, например детей (см. вставку 5).

## Вставка 4: Классификация рисков, связанных с ИИ, на основе существующих или потенциальных уязвимостей

---

### Физические лица

- Человеческое достоинство, ценность или возможности (например, манипулирование, обман, подталкивание, вынесение приговоров, эксплуатация, дискриминация, равное обращение, судебное преследование, слежка, потеря самостоятельности человека и целеуказание с помощью ИИ).
- Физическая и психическая целостность, здоровье, безопасность и охрана (например, подталкивание, одиночество и изоляция, нейротехнологии, смертоносное автономное оружие, автономные автомобили, медицинская диагностика, доступ к медицинскому обслуживанию, взаимодействие с химическими, биологическими, радиологическими и ядерными системами).
- Жизненные возможности (например, образование, работа и жилье).
- (Другие) права человека и гражданские свободы, такие как право на презумпцию невиновности (например, предиктивная работа полиции), право на справедливый суд (например, прогнозирование рецидивизма, виновность, рецидивизм, прогнозирование и автономные судебные процессы), свобода выражения мнений и информации (например, подталкивание, персонализированная информация, информационные пузыри), частная жизнь (например, технология распознавания лиц), свобода собраний и передвижения (например, технология отслеживания в общественных местах).

### Политика и общество

- Дискриминация и несправедливое отношение к группам, в том числе на основе индивидуальных или групповых признаков (таких, как пол), групповая изоляция и маргинализация.
- Различные последствия для детей, пожилых людей, лиц с инвалидностью и уязвимых групп.
- Международная и национальная безопасность (например, автономное оружие, охрана порядка и пограничный контроль в отношении мигрантов и беженцев, организованная преступность, терроризм, распространение и эскалация конфликтов).
- Демократия (например, выборы и доверие).
- Целостность информации (например, дезинформация или введение в заблуждение, дипфейки и персонализированные новости).
- Верховенство закона (например, функционирование и доверие к институтам, правоохранительным и судебным органам).
- Культурное разнообразие и изменения в человеческих отношениях (например, универсализация и фальшивые друзья).
- Социальная сплоченность (например, «информационные пузыри», снижение доверия к институтам и источникам информации).
- Ценности и нормы (например, этические, моральные, культурные и правовые).

### Экономика

- Концентрация власти.
- Зависимость от технологий.
- Неравенство экономических возможностей, условий доступа к рынкам, а также распределения и выделения ресурсов).

## Вставка 4: Классификация рисков, связанных с ИИ, на основе существующих или потенциальных уязвимостей (продолжение)

- Недостаточное использование ИИ.
- Чрезмерно активное использование ИИ, или слишком большой упор на технологические решения.
- Стабильность финансовых систем, критической инфраструктуры и институтов.
- Охрана интеллектуальной собственности.

### Окружающая среда

- Чрезмерное потребление энергии, воды и материальных ресурсов (включая редкие полезные ископаемые и другие природные ресурсы).

## Вставка 5: Сосредоточение внимания на детях в контексте управления ИИ

Для того, чтобы компании и школы учитывали потребности и права детей, необходим комплексный подход к управлению, учитывающий их уникальные особенности. Дети генерируют треть всех данных и вырастут, когда экономика и мир привыкнут к использованию ИИ. В этой вставке обобщены некоторые меры, связанные с данной темой, которые обсуждались в ходе наших «глубоких погружений».

### Уделение первоочередного внимания правам и мнениям детей:

Управление ИИ должно признавать детей в качестве одной из приоритетных заинтересованных сторон, подчеркивая их право на развитие без привыкания к технологиям и их право на отказ от них. В отличие от общих подходов, ориентированных на человеческий фактор, управление, ориентированное на детей, должно учитывать долгосрочные последствия для их мировоззрения, самовосприятия, жизненного выбора и возможностей. Вовлечение детей в процессы разработки и управления имеет решающее значение для обеспечения безопасности и пригодности систем ИИ для их использования детьми.

### Научные исследования и разработка политики:

Нам нужны обширные исследования, чтобы понять, как ИИ влияет на социальное, когнитивное и эмоциональное развитие детей с течением времени. Эти исследования должны стать основой для обсуждения политики и принятия защитных мер в разных странах.

### Защита и неприкосновенность частной жизни:

Дети не должны использоваться в качестве объектов для экспериментов с ИИ. Защита частной жизни детей имеет первостепенное значение. Технологии ИИ должны включать в себя строгие протоколы защиты данных и предоставлять контент, соответствующий возрасту.

### Оценка воздействия на детей и проектирование с учетом интересов детей:

Обязательная оценка систем ИИ на предмет их воздействия на детей необходима для обеспечения их пригодности и безопасности. Системы ИИ должны разрабатываться с учетом потребностей детей и с самого начала включать в себя функции безопасности и ограничения. В принятии решений относительно конструктивных особенностей таких систем должны принимать участие сами дети.



**30** Эти результаты иллюстрируют важность инклюзивного представительства при достижении общего понимания рисков ИИ и общих позиций по политическим программам, как это предусмотрено рекомендациями 1 и 2. Без такого представительства политические программы управления ИИ могут быть сформулированы таким образом, что в них не будут учтены интересы части человечества, которая тем не менее будет затронута.

## Ф. Вызовы, требующие реагирования

**31** Помимо рисков и вреда, которые могут стать реальностью в ближайшем будущем, эволюция разработки, внедрения и использования ИИ также создает проблемы в контексте существующих институтов, что, в свою очередь, влияет на стратегии управления ИИ. Технологические темпы развития передового ИИ и его универсальный характер подвергают

еще более серьезному испытанию способность человечества принимать своевременные меры реагирования.

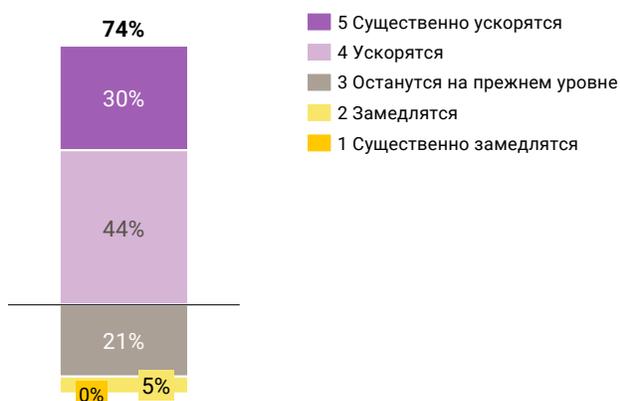
**32** Гонка в борьбе за лидирующие позиции в сфере разработки и внедрения систем ИИ идет вразрез с традиционными системами регулирования и режимами управления. Большинство экспертов, опрошенных в рамках Глобальной оценки рисков, связанных с ИИ, ожидают ускорения темпов развития ИИ в ближайшие 18 месяцев как в области разработки (74 процента), так и в области внедрения и применения (89 процентов) (см. рис. 4).

**33** Как было отмечено в пункте 23 выше, некоторые эксперты ожидают развертывания автономных систем в 2025 году. Кроме того, ведущие технические эксперты признают, что многие модели ИИ остаются непрозрачными, а их результаты не вполне предсказуемы или контролируемы, хотя негативные побочные эффекты могут повлиять на других людей в глобальном масштабе.

### Рисунок 4: Ожидания экспертов в отношении технического развития ИИ

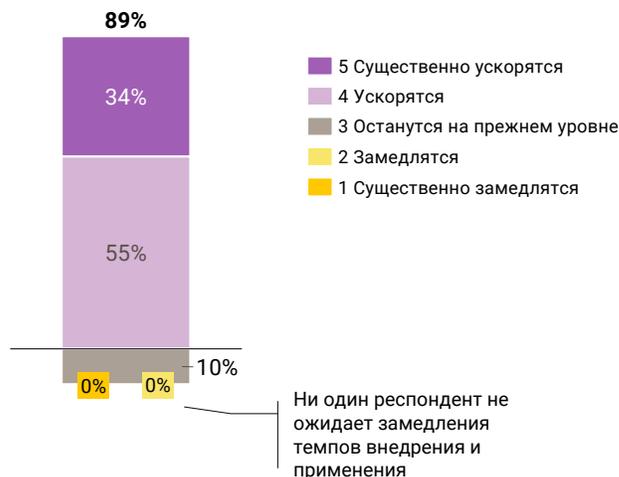
**74% респондентов ожидают ускорения темпов технических изменений (30% ожидают их существенного ускорения)**

"Ожидаете ли вы, что в предстоящие 18 месяцев темпы технических изменений в сфере ИИ (включая, например, разработку / внедрение новых моделей) по сравнению с предшествующими 3 месяцами..." (n=348)



**89% респондентов ожидают, что темпы внедрения и применения ускорятся (34% – существенно ускорятся)**

"Ожидаете ли вы, что в предстоящие 18 месяцев темпы внедрения и применения ИИ (включая, например, новые виды использования ИИ коммерческими компаниями / правительственными ведомствами) по сравнению с предшествующими 3 месяцами..." (n=348)



Примечание: общая сумма может не быть равна 100% по причине округления. Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

- 34** Растущая зависимость от автоматизированного принятия решений и создание контента непрозрачными алгоритмами может иметь пагубные последствия для справедливого обращения и безопасности. Хотя люди часто продолжают нести юридическую ответственность за решения об автоматизации процессов, которые влияют на других людей, механизмы подотчетности могут развиваться недостаточно быстро для того, чтобы механизмы такой подотчетности работали оперативно и эффективно.
- 35** Таким образом, в масштабах всего общества возникает риск того, что все меньшее число людей будет нести ответственность за вред, причиненный в результате их решений об автоматизации процессов с помощью ИИ, несмотря на то, что в мире появляются все более мощные системы. Это требует гибкого управления для обеспечения того, чтобы механизмы подотчетности не отставали от ускоряющегося развития ИИ.
- 36** Если исходить из того, что темпы разработки и внедрения ИИ бросают вызов существующим институтам, то аналогичные последствия будут иметь и масштабность соответствующих процессов. Являясь технологией общего назначения с глобальным охватом, передовой ИИ может быть развернут в различных областях, оказывая многообразное влияние на общество и приводя к широким политическим последствиям.
- 37** Влияние и потенциальное воздействие ИИ, пересекающегося с многочисленными областями, включая финансы, рынки труда, образование и политические системы, предвещают широкие последствия, требующие подхода, охватывающего все общество (см. примеры во вставке 6). Существующие институты должны разработать комплексные межотраслевые меры реагирования, направленные на принятие мер в связи с широкомасштабным воздействием ИИ на общество.
- 38** Темпы, масштабы и неопределенность разработки, внедрения и использования ИИ подчеркивают ценность комплексного, сквозного и гибкого подхода к ИИ. На международном уровне комплексная перспектива должна отражаться в сетевом институциональном подходе к управлению ИИ на межотраслевой и трансграничной основе, который вовлекал бы заинтересованные стороны, но при этом не оказывался бы в их власти.
- 39** Что касается изменения климата, то мир лишь с запозданием осознал, что необходим комплексный подход к глобальным коллективным действиям. В сфере ИИ есть возможность изначально заложить этот принцип в конструкцию системы.
- 40** Вышеперечисленные проблемы усугубляются концентрацией богатства и процесса принятия решений среди горстки частных компаний, занимающихся разработкой и внедрением ИИ, включая, в частности, транснациональные корпорации. В связи с этим возникает еще один вопрос о том, как заинтересованные стороны могут участвовать в управлении ИИ, не подрывая при этом общественные интересы.

## Вставка 6: Последствия для общества, связанные с ИИ

В рамках проведенных Консультативным органом мероприятий с более широкой степенью охвата его члены провели консультации с рядом заинтересованных сторон, с тем чтобы обсудить последствия ИИ для общества. В этой вставке обобщены основные проблемы и потенциальные инициативы, выдвинутые в ходе «глубоких погружений» в данную тему.

### **Общественные и психологические последствия, а также последствия для сообществ:**

По мере того, как ИИ становится все более мощным и широко распространенным, его разработка, внедрение и применение будут становиться все более персонализированными, что может привести к отчуждению и зависимости. По мнению некоторых членов Консультативного органа, ИИ, обученный на данных отдельного человека, и его последующая роль в качестве основного собеседника и посредника могут отражать переломный момент для человечества, чреватый возникновением новых острых социальных проблем и усугублением существующих.

Например, будущие системы ИИ смогут генерировать бесконечный поток высококачественного видеоконтента с учетом личных предпочтений человека. Усиление социальной изоляции, отчуждение, проблемы с психическим здоровьем, потеря человеком самостоятельности, влияние на эмоциональный интеллект и социальное развитие – вот лишь некоторые из возможных последствий.

Эти вопросы недостаточно изучены политиками в контексте таких технологий, как интеллектуальные устройства и Интернет; они практически совсем не изучены в контексте ИИ, поскольку существующие системы управления ставят во главу угла риски для отдельных людей, а не для общества в целом.

Когда политики рассматривают будущие меры, связанные с ИИ, они должны учитывать и эти факторы, а также разрабатывать политику, способствующую благополучию общества, особенно молодежи. Правительственные меры могут способствовать формированию среды, в которой первоочередное внимание уделяется личному общению между людьми, сделать более доступной помощь в области психического здоровья, а также увеличить инвестиции в спортивные сооружения, публичные библиотеки и искусство.

Тем не менее профилактика лучше лечения: разработчики должны создавать продукты без вызывающих привыкание персонализированных функций, следить за тем, чтобы они не вредили психическому здоровью и укрепляли (а не подрывали) чувство общей принадлежности к обществу. Технологические компании должны разработать политику управления социальными рисками наравне с другими рисками в рамках усилий по выявлению и снижению рисков на протяжении всего жизненного цикла продуктов ИИ.

## Вставка 6: Последствия для общества, связанные с ИИ (продолжение)

---

### Дезинформация и доверие:

Дипфейки, голосовые клоны и автоматизированные кампании по дезинформации представляют собой конкретную и серьезную угрозу демократическим институтам и процессам, таким как выборы, а также демократическим обществам и социальному доверию в целом, в том числе посредством иностранного информационного манипулирования и вмешательства (ИИМВ). Развитие замкнутых информационных экосистем, усиленных ИИ и использующих персональные данные, может оказать глубокое воздействие на общество, потенциально делая его более восприимчивым к идеям нетерпимости и насилия по отношению к другим.

Для защиты целостности представительных государственных учреждений и процессов необходимы надежные системы проверки и обнаружения дипфейков, а также процедуры быстрого уведомления и удаления контента, который может ввести в заблуждение таким образом, что это повлечет за собой вред или внесет раскол в общество, или способствовать пропаганде войны и конфликтов и разжиганию ненависти. Лица, не являющиеся публичными фигурами, должны иметь защиту от создания другими лицами дипфейков с использованием их образа в мошеннических, клеветнических или иных неправомерных целях. Сексуализированные дипфейки представляют особую опасность для женщин и девочек и могут быть формой гендерного насилия.

Добровольные обязательства со стороны частного сектора, такие как маркировка дипфейков или предоставление пользователям возможности отмечать и затем удалять дипфейки, созданные или распространяемые со злым умыслом, – это важные первые шаги. Однако они не позволяют в достаточной степени снизить общественные риски. Вместо этого необходим глобальный подход с участием многих заинтересованных сторон, а также жесткие обязательства. Общие стандарты аутентификации контента и его цифрового происхождения позволят выработать признанный во всем мире подход к идентификации искусственных и модифицированных ИИ изображений, а также видео- и аудиоматериалов.

Кроме того, обмен информацией между государственными и частными структурами в режиме реального времени, основанный на международных стандартах, позволил бы создать возможности быстрого реагирования для немедленного удаления вводящего в заблуждение контента или ИИМВ до того, как он успеет стать вирусным. Тем не менее эти процессы должны включать в себя гарантии того, что ими не будут манипулировать или злоупотреблять для оправдания цензуры.

Эти действия должны сопровождаться профилактическими мерами, направленными на повышение устойчивости общества к дезинформации и пропаганде с использованием ИИ, такими как кампании по расширению информированности общественности о потенциале ИИ, способном подорвать целостность информации. Государствам-членам следует также проводить кампании по повышению медиа- и цифровой грамотности, поддерживать инициативы по проверке фактов и вкладывать средства в укрепление потенциала сообщества борцов с ИИМВ.

## 2. Необходимость глобального управления

- 41** Сегодня существует дефицит глобального управления в отношении ИИ. Несмотря на активное обсуждение этических норм и принципов, система норм, институтов и инициатив еще только зарождается и полна пробелов. Подотчетность и средства защиты от ущерба часто бывают заметны прежде всего своим отсутствием. Соблюдение норм опирается на добровольность. Существует фундаментальное несоответствие между риторикой на высоком уровне, разрабатываемыми, внедряемыми и используемыми системами и условиями, необходимыми для обеспечения безопасности и инклюзивности. Как мы отмечали в нашем промежуточном докладе, управление ИИ имеет решающее значение не только для решения проблем и устранения рисков, но и для обеспечения того, чтобы мы использовали его потенциал таким образом, чтобы никто не остался за бортом<sup>11</sup>.
- 42** В частности, не вызывает никаких сомнений настоятельная необходимость глобального управления. Сырье для ИИ – от важнейших полезных ископаемых до обучающих данных – добывается в глобальном масштабе. ИИ общего назначения, развернутый на трансграничном уровне, порождает множество способов применения во всем мире. Ускоренное развитие ИИ концентрирует власть и богатство в глобальном масштабе, что влечет за собой геополитические и геоэкономические последствия. Более того, в настоящее время никто не понимает всех внутренних механизмов ИИ настолько, чтобы полностью контролировать результаты его применения или прогнозировать его развитие. Лица, принимающие решения, также не могут нести ответственность за разработку, развертывание или использование систем, которые они не понимают. Между тем негативные побочные эффекты и последствия таких решений для смежных областей также могут быть глобальными.
- 43** Несмотря на глобальный охват ИИ, национальные и региональные институциональные структуры и нормативные акты заканчиваются на физических границах государств. Это снижает способность любой отдельной страны регулировать последующее применение ИИ, которое приводит к трансграничному ущербу, или решать проблемы в сложных трансграничных цепочках поставок элементов вычислительной инфраструктуры, потоков обучающих данных и источников энергии, которые лежат в основе разработки и использования ИИ. Ведущие компании, занимающиеся разработкой ИИ, зачастую оказывают более значительное прямое влияние на последующее применение ИИ (через снижение рисков на начальном этапе), чем большинство стран, действующих в одиночку.
- 44** Разработка, внедрение и использование такой технологии не могут быть предоставлены только прихотям рынков. Национальные правительства и региональные организации будут играть решающую роль. Однако в дополнение к факторам, связанным с равенством, справедливостью, предотвращением ущерба и компенсацией за него, сама природа данной технологии – трансграничной по структуре и применению – требует глобального многоотраслевого подхода. Без глобальной всеохватывающей системы, в которой участвуют заинтересованные

11 См. <https://un.org/ai-advisory-body>.

стороны, и с учетом динамики конкуренции как правительства, так и компании могут поддаться искушению срезать углы или поставить на первое место собственные интересы.

**45** Таким образом, ИИ создает глобальные проблемы и возможности, которые требуют комплексного и глобального подхода, охватывающего все политические, экономические, социальные, этические, правозащитные, технические, экологические и другие области. Такой подход может превратить множество развивающихся инициатив в единое и взаимоувязанное целое, основанное на международном праве и адаптируемое в разных условиях и во времени.

**46** Необходимость глобального управления ИИ возникает на фоне геополитической и геоэкономической конкуренции за влияние и рынки. Однако устранение рисков, связанных с ИИ, при одновременном обеспечении справедливого использования возможностей требует согласованных глобальных действий. Увеличение цифрового разрыва может ограничить преимущества ИИ горсткой государств и отдельных лиц, в то время как риски и вред будут затрагивать многие, особенно уязвимые, группы населения.

## А. Руководящие принципы и функции в контексте международного управления ИИ

**47** В нашем промежуточном докладе мы изложили пять принципов, которые должны лечь в основу новых международных институтов управления ИИ:

- **Руководящий принцип 1:** управление ИИ должно осуществляться инклюзивно, с участием всех и на благо всех

- **Руководящий принцип 2:** управление ИИ должно осуществляться в общественных интересах
- **Руководящий принцип 3:** механизмы управления ИИ должны формироваться параллельно управлению данными и поощрению общего пользования данными
- **Руководящий принцип 4:** управление ИИ должно быть универсальным, сетевым и основываться на адаптивном сотрудничестве многих заинтересованных сторон
- **Руководящий принцип 5:** управление ИИ должно основываться на Уставе Организации Объединенных Наций, международных нормах в области прав человека и других согласованных международных обязательствах, таких как ЦУР.

**48** Во вставке 7 обобщены отзывы об этих принципах, в которых подчеркивается важность прав человека и необходимость большей ясности в отношении эффективной реализации руководящих принципов, в том числе в плане управления данными. В них нам предлагается решить проблему обеспечения того, чтобы поддержка инклюзивности подкреплялась действиями, а маргинализированные группы не остались в стороне.

**49** В нашем промежуточном докладе мы также предложили несколько организационных функций, которые можно было бы выполнять на международном уровне (см. рис. 5). Авторы отзывов в основном подтвердили необходимость этих функций на глобальном уровне, но при этом призвали к созданию дополнительных функций вспомогательного характера, связанных с управлением данными и ИИ, с тем чтобы реализовать на практике руководящий принцип 3 (механизмы управления ИИ должны формироваться параллельно управлению данными и поощрению общего пользования данными).

## Рисунок 5: Функции по управлению ИИ, которые предлагается выполнять на международном уровне



### Вставка 7: Отзывы о руководящих принципах

#### Упор на управление ИИ на основе прав человека:

По итогам обширных консультаций, проведенных Консультативным органом высокого уровня после публикации его промежуточного доклада, руководящий принцип 5 (управление ИИ должно основываться на Уставе Организации Объединенных Наций, международных нормах в области прав человека и других согласованных международных обязательствах, таких как ЦУР) получил самую широкую поддержку во всех категориях заинтересованных сторон, включая правительства, гражданское общество, сообщество технических специалистов, научные круги и частный сектор. В частности, речь идет об уважении, поощрении и соблюдении прав человека, о преследовании за их нарушения, а также о резолюции 78/265 Генеральной Ассамблеи об использовании возможностей безопасных, защищенных и надежных систем искусственного интеллекта в целях устойчивого развития, единогласно принятой в марте 2024 года.

В ходе обсуждений Консультативный орган пришел к убеждению, что для снижения рисков и вреда от ИИ, решения вопросов, связанных с новыми видами использования, и обеспечения того, чтобы ИИ действительно приносил пользу всему человечеству и никого не оставлял без внимания, права человека должны быть в центре управления ИИ, обеспечивая основанную на правах человека подотчетность во всех юрисдикциях. Это основополагающее обязательство в отношении прав человека носит сквозной характер и относится ко всем рекомендациям, представленным в настоящем заключительном докладе.

#### Конкретные механизмы реализации и ясность в отношении руководящих установок:

Многие заинтересованные стороны подчеркнули необходимость разработки подробных планов действий и четких руководящих установок для обеспечения эффективной реализации руководящих принципов

## **Вставка 7: Отзывы о руководящих принципах (продолжение)**

Консультативного органа по международному управлению ИИ. Правительственные ведомства предложили разработать четкие рекомендации по определению и обеспечению общественных интересов, а также механизмы участия общественности и надзорной деятельности. Представители частного сектора нередко подчеркивали необходимость разработки четкой политики и использования существующей нормативно-правовой базы для поддержания конкурентных и инновационных рынков ИИ. Многие международные организации и организации гражданского общества также призвали к созданию гибких систем управления, способных своевременно реагировать на развивающиеся технологии. Некоторые из них конкретно просили создать новую структуру, обладающую «мускулами и зубами» и не ограничивающуюся простой координацией.

### **Механизмы обеспечения ответственности ключевых субъектов:**

Многие затрагивали проблему ответственности за дискриминационные, предвзятые и другие вредоносные системы ИИ; звучали предложения относительно создания механизмов, обеспечивающих ответственность за причиненный вред и наличие средств правовой защиты, а также решения проблемы концентрации технологического потенциала и рыночного влияния. Многие организации подчеркивали необходимость борьбы с бесконтрольной властью, а также обеспечения прав потребителей и честной конкуренции. Академические учреждения признали сильные стороны руководящих принципов, заключающиеся в их универсальности и инклюзивности, но при этом предложили улучшить взаимодействие с заинтересованными сторонами. Представители частного сектора подчеркнули важность ответственного использования ИИ, а также устранения барьеров, препятствующих доступу к нему.

### **Более конкретные функции, относящиеся к управлению данными ИИ:**

В ходе многих консультаций упоминалось отсутствие систем управления данными, причем заинтересованные стороны указывали, что Организация Объединенных Наций является естественной площадкой для диалога по вопросам управления данными. Правительства подчеркивали необходимость создания надежных систем управления данными, в которых первоочередное внимание уделялось бы конфиденциальности, защите данных и их справедливому использованию, и выступали за разработку международных руководящих принципов для управления сложными данными в процессе разработки ИИ. Было предложено разработать эти рамки на основе прозрачного и инклюзивного процесса с учетом таких этических аспектов, как согласие и неприкосновенность частной жизни.

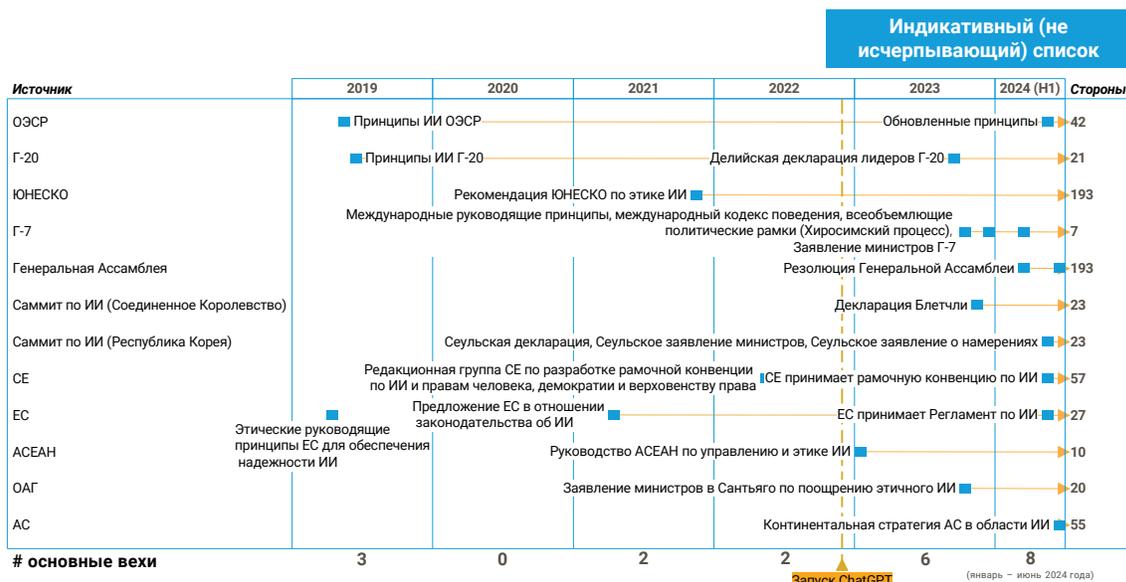
Представители научных кругов подчеркнули, что в краткосрочной перспективе управление данными должно стать приоритетной задачей. Представители частного сектора отметили, что меры по управлению данными должны дополнять меры по управлению ИИ, делая акцент на всеобъемлющих законах о конфиденциальности и ответственном использовании ИИ. Международные организации и организации гражданского общества указали, что управление данными для обучения ИИ должно защищать права потребителей и поддерживать честную конкуренцию между разработчиками ИИ путем предоставления неисключительного доступа к данным для обучения ИИ, обратив особое внимание на призыв к принятию конкретных и практических мер по управлению данными. Организация Объединенных Наций была названа ключевой площадкой для решения этих проблем управления и преодоления диспропорций в том, что касается ресурсов.

- 50** Что касается институционально более «жестких» функций управления ИИ, таких как мониторинг, проверка, отчетность, соблюдение, стабилизация положения с подотчетностью, реагирование и обеспечение соблюдения соответствующих норм, то в отзывах отмечается, что для институционализации таких функций сначала потребуются международные договорные обязательства и что доводы в пользу институционализации таких функций в управлении ИИ как технологией пока не прозвучали.
- 51** Не все функции должны выполняться исключительно Организацией Объединенных Наций. Однако для того, чтобы превратить лоскутное одеяло норм и институтов в страховочную сетку, обеспечивающую и поддерживающую устойчивые инновации на благо всего человечества, необходимо общее понимание научных и общих фундаментальных положений, лежащих в основе правил и стандартов, по которым мы оцениваем, достигают ли механизмы управления своих целей.
- 52** В ходе наших консультаций мы услышали призывы к проведению более подробного обзорного анализа существующих и новых усилий по международному регулированию ИИ, а также пробелов, которые необходимо устранить для справедливого, эффективного и действенного международного управления ИИ.

## **В. Формирующийся ландшафт международного управления ИИ**

- 53** Очевидно, что в настоящее время нет недостатка в документах и диалогах, посвященных управлению ИИ. Правительствами, компаниями и консорциумами, а также региональными и международными организациями были разработаны сотни руководств, рамочных документов и принципов. Десятки форумов собирают различных участников, представляющих устоявшиеся межправительственные механизмы, экспертные органы и специальные инициативы с участием многих заинтересованных сторон. Они дополняются уже действующими и разрабатываемыми нормативными актами на национальном и региональном уровнях.
- 54** Растет число международных инициатив, выдвигаемых правительствами (см. рис. 6). Эти новые инициативы все чаще основываются на сквозном подходе к управлению ИИ на международном уровне, объединяющем принципы, декларации, заявления и другие документы, которые касаются ИИ не в каких-то конкретных аспектах, а комплексно. Их реализация кардинально ускорилась начиная с 2023 года, что было подстегнуто выпуском множества крупных языковых моделей ИИ общего назначения после выхода ChatGPT в ноябре 2022 года.

## Рисунок 6: Межрегиональные и региональные инициативы в области управления ИИ, основные этапы, 2019–2024 годы (H1)



Сокращения: АСЕАН – Ассоциация государств Юго-Восточной Азии; АС – Африканский союз; СЕ – Совет Европы; ЕС – Европейский союз; Г-20 – Группа 20; Г-7 – Группа семи; ГП ИИ – Глобальное партнерство по искусственному интеллекту; ОАГ – Организация американских государств; ОЭСР – Организация экономического сотрудничества и развития; ЮНЕСКО – Организация Объединенных Наций по вопросам образования, науки и культуры.

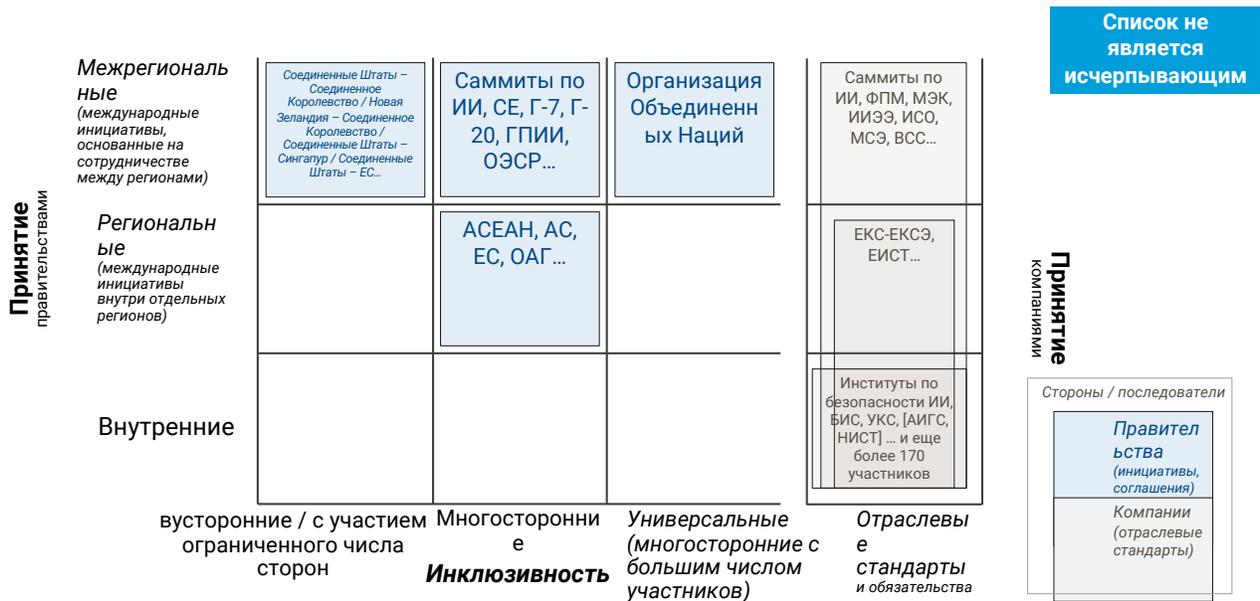
**55** Параллельно с этим были разработаны и опубликованы отраслевые стандарты по ИИ для принятия на международном уровне. Другие инициативы с участием многих заинтересованных сторон также направлены на преодоление разрыва между государственным и частным секторами, в том числе на таких дискуссионных площадках, как Форум по вопросам управления Интернетом.

**56** На рисунке 7 представлен обзор некоторых авторов инициатив по управлению ИИ и отраслевых стандартов с разбивкой по географическому диапазону и охвату (говоря об этой недавней работе, мы отмечаем многолетние усилия ученых, гражданского общества и профессиональных организаций).

**57** Примерами соответствующих региональных и межрегиональных многосторонних инициатив являются инициативы, осуществляемые Африканским союзом, различными принимающими сторонами саммитов по ИИ, Ассоциацией государств Юго-Восточной Азии, Советом Европы, Европейским союзом, Группой семи (Г-7), Группой 20 (Г-20), Глобальным партнерством по искусственному интеллекту, Организацией американских государств и Организацией экономического сотрудничества и развития (ОЭСР) и др.

**58** Наш анализ существующих механизмов управления, скорее всего, устареет в течение нескольких месяцев. Тем не менее он может помочь проиллюстрировать, как текущие и возникающие международные инициативы по управлению ИИ соотносятся с нашими руководящими принципами формирования новых глобальных институтов управления ИИ, включая принцип 1 (управление ИИ должно осуществляться инклюзивно, с участием всех и на благо всех).

## Рисунок 7: Авторы инициатив, касающихся управления, которые конкретно сосредоточены на ИИ



Список не является исчерпывающим

Сокращения: АИГС – Американский институт государственных стандартов; АСЕАН – Ассоциация государств Юго-Восточной Азии; АС – Африканский союз; БИС – Британский институт стандартов; ЕКС – Европейский комитет по стандартизации; ЕКСЭ – Европейский комитет по стандартизации в электротехнике; СЕ – Совет Европы; ЕИСТ – Европейский институт стандартов по телекоммуникациям; ЕС – Европейский союз; ФПМ – Форум передовых моделей; Г-20 – Группа 20; Г-7 – Группа семи; ГПКИ – Глобальное партнерство по искусственному интеллекту; МЭК – Международная электротехническая комиссия; ИИЭЭ – Институт инженеров по электронике и электротехнике; ИСО – Международная организация по стандартизации; МСЭ – Международный союз электросвязи; НИСТ – Национальный институт стандартов и технологий; ОАГ – Организация американских государств; ОЭСР – Организация экономического сотрудничества и развития; УКС – Управление Китая по стандартизации; ВСС – Всемирное сотрудничество по стандартам.

## 3. Пробелы в глобальном управлении ИИ

- 59** Многочисленные национальные, региональные, многосторонние и другие инициативы, упомянутые выше, принесли значимые результаты и стали основой для нашей работы; многие из их представителей внесли письменный вклад в наши обсуждения или приняли участие в наших консультациях.
- 60** Тем не менее, за исключением нескольких инициатив, выдвинутых Организацией Объединенных Наций<sup>12</sup>, ни одна из них не может считаться по-настоящему глобальной по своему охвату. Эти пробелы в управлении ИИ на международном уровне являются проблемой, поскольку данная технология является глобальной и будет иметь всеохватывающее воздействие.
- 61** Отдельные пробелы в координации между инициативами и институтами рискуют расколоть мир на разрозненные и несовместимые режимы управления ИИ.
- 62** Кроме того, пробелы в реализации и подотчетности снижают способность государств, частного сектора, гражданского общества, научных кругов и технического сообщества воплощать обязательства, какой бы широкой поддержкой они ни пользовались, в ощутимые результаты.

### А. Пробелы в представительстве

- 63** Проведенный нами анализ различных инициатив по управлению ИИ, не относящихся к Организации Объединенных Наций и охватывающих различные регионы, показывает, что большинство инициатив не являются полностью репрезентативными в своих межправительственных аспектах.
- 64** Многие из них исключают целые регионы мира. Как показано на рисунке 8, если рассматривать семь многосторонних межрегиональных инициатив в области ИИ, не относящихся к системе Организации Объединенных Наций, с частично совпадающим членством, то семь стран являются участниками всех из них, в то время как 118 стран не являются участниками ни одной из них (в основном на глобальном Юге, с неравномерным представительством даже тех стран, которые являются лидерами в сфере ИИ; см. рис. 12).
- 65** Избирательность вполне объяснима на ранней стадии формирования механизмов управления, когда присутствует определенная доля эксперимента, конкуренции вокруг норм и различных уровней комфорта в отношении новых технологий. Однако по мере становления международного управления ИИ глобальное представительство становится все более важным с точки зрения справедливости и эффективности.

<sup>12</sup> Рекомендация Организации Объединенных Наций по вопросам образования, науки и культуры (ЮНЕСКО) об этике искусственного интеллекта (2021 год) и две резолюции Генеральной Ассамблеи по ИИ.

## Рисунок 8: Участие в семи международных инициативах по управлению ИИ, реализуемых вне системы Организации Объединенных Наций

Выборка: Принципы ИИ ОЭСР (2019 год), Принципы ИИ Г-20 (2019 год), Группа Совета Европы по разработке конвенции по ИИ (2022–2024 годы), Заявление министров ГПИИ (2022 год), Декларация Блетчли (2023 год) и Сеульское заявление министров (2024 год).

Выборка включает исключительно межрегиональные (не региональные) инициативы



\* По признаку подтверждения соответствующих межправительственных документов. Страны не считаются участниками той или иной многосторонней инициативы исключительно в силу их членства в Европейском союзе или Африканском союзе. Сокращения: ГАГ – Группа африканских государств; ГГАТО – Группа государств Азии и Тихого океана; ГГВЕ – Группа государств Восточной Европы; Г-20 – Группа 20; Г-7 – Группа семи; ГПИИ – Глобальное партнерство по искусственному интеллекту; ЛАКБ – Латинская Америка и Карибский бассейн; ОЭСР – Организация экономического сотрудничества и развития; ГЗДГ – Группа западноевропейских и других государств.

**66** Помимо неинклюзивности существующих усилий, в национальных и региональных инициативах, направленных на достижение общего научного понимания ИИ, также существуют пробелы в плане представленности. Эти пробелы в представленности могут проявляться в процессах принятия решений относительно того, как определять масштабы, выделять ресурсы и проводить оценки.

**67** Равенство требует, чтобы большее число людей играло значимую роль в принятии решений о том, как управлять технологиями, которые влияют на всех нас, а также признания того, что многие сообщества исторически были исключены из этих обсуждений. Сравнительно ограниченное присутствие соответствующих тем в повестках дня крупномасштабных инициатив, которые являются

приоритетными для определенных регионов, свидетельствует о дисбалансе, обусловленном недопредставленностью<sup>13</sup>.

**68** Для того, чтобы режимы управления ИИ были эффективными, они должны охватывать весь мир: эффективно предотвращать «гонку вооружений ИИ» или «гонку на понижение» в области безопасности и прав, обеспечивать эффективное реагирование на вызовы, обусловленные трансграничным характером ИИ<sup>14</sup>, а также стимулировать обучение, поощрять взаимную совместимость и способствовать совместному использованию преимуществ ИИ. Кроме того, различные мнения, в том числе и не совпадающие друг с другом, позволяют прогнозировать угрозы и выверять такие ответные меры, которые станут творческими и адаптируемыми.

<sup>13</sup> Например, управление наборами данных для обучения ИИ, доступ к вычислительным ресурсам, развитие потенциала ИИ, связанные с ИИ риски, касающиеся дискриминации маргинализированных групп и использования ИИ в вооруженных конфликтах (см. результаты Глобальной оценки рисков, связанных с ИИ, в приложении Е, которые показывают разное восприятие рисков респондентами из Группы западноевропейских и других стран по сравнению с прочими странами). Многие государства и маргинализированные сообщества также лишены возможности пользоваться благами ИИ или могут несоразмерно страдать от его вреда. Справедливость требует разнообразного и инклюзивного подхода, учитывающего мнения всех регионов и равномерно распределяющего возможности при одновременном снижении рисков.

<sup>14</sup> В тех случаях, когда будут прочерчены «красные линии» – возможно, по аналогии с запретом на клонирование человека, – они будут соблюдаться только при условии глобального принятия соответствующих норм, а также контроля за их соблюдением. Данное утверждение не теряет своей актуальности несмотря на то, что, как это ни парадоксально, в нынешней парадигме на фоне снижения стоимости отдельных систем ИИ стоимость продвинутых систем ИИ (возможно, наиболее важных с точки зрения контроля) растет.

**69** Ограничивая круг стран, участвующих в ключевых процессах формирования повестки дня, налаживания отношений и обмена информацией, избирательная многосторонность может ограничить достижение тех самых целей, которые перед ней ставятся. К ним относятся взаимная совместимость новых подходов к управлению ИИ, глобальная безопасность ИИ и общее понимание научных аспектов ИИ на глобальном уровне (см. рекомендации 1, 2 и 3 относительно того, что делает глобальный подход особенно эффективным).

**70** Две резолюции Генеральной Ассамблеи по ИИ, принятые в 2024 году<sup>15</sup>, свидетельствуют о признании ведущими странами ИИ необходимости устранять пробелы в представительстве в контексте международного управления ИИ, и Организация Объединенных Наций могла бы стать форумом для объединения усилий всего мира в этом направлении.

**71** Глобальный цифровой договор в сентябре 2024 года и Форум Всемирной встречи на высшем уровне по вопросам информационного общества в 2025 году предлагают две дополнительных политических возможности для институционализации глобального репрезентативного набора процессов управления ИИ в целях устранения пробелов в представительстве<sup>16</sup>.

## В. Пробелы в координации

**72** Постоянно возникающие и развивающиеся инициативы по управлению ИИ не гарантируют эффективного взаимодействия в интересах человечества. Напротив, возникли пробелы в координации. Эффективное взаимодействие между инициативами, основанными на избирательной многосторонности (см. рис. 8), и другими региональными инициативами не гарантировано, что чревато несовместимостью между регионами.

**73** Кроме того, не существует глобальных механизмов, позволяющих всем международным организациям по разработке стандартов (см. рис. 7), международным научно-исследовательским инициативам или инициативам по наращиванию потенциала ИИ координировать свои действия друг с другом, что подрывает операционную совместимость подходов и приводит к фрагментации. Возникающие в результате этого пробелы в координации между различными инициативами, не являющимися глобальными по своему охвату, в некоторых случаях лучше всего устранять на глобальном уровне.

**74** Отдельные пробелы в координации возникают в системе Организации Объединенных Наций, что находит свое отражение в массиве разнообразных документов и инициатив Организации Объединенных Наций, касающихся ИИ. На рисунке 9 показаны 26 связанных с Организацией Объединенных Наций документов в конкретных областях, которые могут

<sup>15</sup> Резолюции 78/265 («Использование возможностей безопасных, защищенных и надежных систем искусственного интеллекта в целях устойчивого развития») и 78/311 («Укрепление международного сотрудничества в деле наращивания потенциала в области искусственного интеллекта»).

<sup>16</sup> Различные многосторонние инициативы, включая Принципы ИИ ОЭСР, Хиросимский процесс Группы семи по ИИ и Рамочную конвенцию Совета Европы по искусственному интеллекту, открыты для сторонников и приверженцев за пределами стран-инициаторов. Однако такая открытость может не обеспечить представительства и легитимности в той скорости и масштабах, которые необходимы для того, чтобы идти в ногу с ускоряющимся распространением ИИ на глобальном уровне. Тем временем в международных процессах управления ИИ сохраняются пробелы в представительстве, а принятие решений сосредоточено в руках нескольких стран и компаний.

относиться к ИИ, причем 25 из них имеют обязательную силу и потребуют толкования применительно к ИИ. Еще 32 документа Организации Объединенных Наций и связанных с ней организаций на уровне отдельных областей посвящены непосредственно ИИ, но ни один из них не имеет обязательной силы<sup>17</sup>. В некоторых случаях они позволяют устранить риски, связанные с ИИ, и использовать преимущества ИИ в конкретных областях.

**75** Уровень активности свидетельствует о важности ИИ для программ Организации Объединенных Наций. По мере распространения ИИ, затрагивающего все более широкие аспекты жизни общества, к различным подразделениям системы Организации Объединенных Наций будет поступать все больше призывов к действиям, в том числе посредством разработки обязательных к исполнению норм. Это также свидетельствует о несистематическом характере мер реагирования, которые в основном естественным образом развивались

## Рисунок 9: Отдельные документы Организации Объединенных Наций и связанных с ней организаций, касающиеся управления ИИ

| Список не является исчерпывающим   |  |  |  |   |  |
|--|--|--|--|---|--|
| <p><b>Этика и политика</b></p> <p><b>ЮНЕСКО</b></p> <ul style="list-style-type: none"> <li>Рекомендация по этике искусственного интеллекта</li> </ul> <p><b>ВОЗ</b></p> <ul style="list-style-type: none"> <li>Руководство по этике искусственного интеллекта и управлению им в здравоохранении</li> </ul> <p><b>Детский фонд Организации Объединенных Наций (ЮНИСЕФ)</b></p> <ul style="list-style-type: none"> <li>Политическое руководство в отношении использования ИИ в интересах детей</li> <li>Аргументы в пользу более эффективного управления данными детей: манифест</li> <li>Ответственность в вопросах данных для детей (d4c.org)</li> </ul> <p><b>Программа Организации Объединенных Наций по населенным пунктам (ООН-Хабитат)</b></p> <ul style="list-style-type: none"> <li>Руководство по соблюдению прав человека в контексте цифровой трансформации городов</li> <li>Политические рамки усилий по обеспечению интересов населения, инклюзивности и прав человека в рамках развития "умных городов"</li> </ul> <p><b>Структура "ООН-Женщины"</b></p> <ul style="list-style-type: none"> <li>КПЖ67 ("Согласованные выводы")</li> </ul> <p><b>Фонд Организации Объединенных Наций в области народонаселения</b></p> <ul style="list-style-type: none"> <li>Программа действий для международной конференции по народонаселению и развитию: рамочная программа, сосредоточенная на проблемах народонаселения и основанная на защите прав человека</li> </ul> | <p><b>Права человека</b></p> <p><b>УВКПЧ</b></p> <ul style="list-style-type: none"> <li>Международная конвенция о ликвидации всех форм расовой дискриминации*</li> <li>Международный пакт о гражданских и политических правах*</li> <li>Международный пакт об экономических, социальных и культурных правах*</li> <li>Конвенция о ликвидации всех форм дискриминации в отношении женщин*</li> <li>Конвенция против пыток и других жестоких, бесчеловечных или унижающих достоинство видов обращения и наказания*</li> </ul> <p><b>MOT</b></p> <ul style="list-style-type: none"> <li>Конвенция о наилучших формах детского труда, 1999 год (№ 182)*</li> <li>Конвенция о безопасности и гигиене труда, 1981 год (№ 155)*</li> <li>Конвенция об основах, содействующих безопасности и гигиене труда, 2006 год (№ 187)*</li> <li>Конвенция 1958 года о дискриминации в области труда и занятий, 1958 год (№ 111)*</li> <li>Конвенция о защите прав представителей работников, 1971 год (№ 135)*</li> <li>Конвенция о политике в области занятости, 1964 год (№ 122)*</li> <li>Свод правил MOT по защите личных данных работников</li> </ul> <p><b>ЮНИСЕФ</b></p> <ul style="list-style-type: none"> <li>Конвенция о правах ребенка*</li> </ul> | <p><b>Технические стандарты</b></p> <p><b>МСЭ</b></p> <ul style="list-style-type: none"> <li>ИИ в сфере операций и управления телекоммуникациями*</li> <li>ИИ в «умных» системах и городах*</li> <li>ИИ в сфере управления сетями и сетевых услуг, ИИ в контексте конкретных технологий или приложений</li> </ul> <p><b>ПРООН</b></p> <ul style="list-style-type: none"> <li>Стандарт цифрового общественного достояния для систем ИИ (разработано совместно с АЦОД)</li> </ul> <p><b>ИКАО</b></p> <ul style="list-style-type: none"> <li>Приложения к Чикагской конвенции*</li> </ul> <p><b>Мир и безопасность</b></p> <p><b>УВРООН</b></p> <ul style="list-style-type: none"> <li>Статья 36 Дополнительного протокола I к Женевским конвенциям*</li> <li>Конвенция о биологическом оружии*</li> <li>Конвенция о химическом оружии</li> </ul> <p><b>Здравоохранение</b></p> <p><b>ВОЗ</b></p> <ul style="list-style-type: none"> <li>Соборования, касающиеся нормативных аспектов использования искусственного интеллекта в сфере здравоохранения</li> <li>Генерирование данных для медицинских устройств на основе искусственного интеллекта: рамочный документ по вопросам валидации и оценки качества обучения</li> <li>Руководящие принципы по этике и управлению использованием искусственного интеллекта в здравоохранении</li> </ul> | <p><b>Коммуникации</b></p> <p><b>Департамент глобальных коммуникаций</b></p> <ul style="list-style-type: none"> <li>Продолжающаяся работа в отношении принципов целостности информации</li> </ul> <p><b>ЮНЕСКО</b></p> <ul style="list-style-type: none"> <li>Руководящие принципы управления цифровыми платформами</li> </ul> <p><b>Торговля</b></p> <p><b>ВТО</b></p> <ul style="list-style-type: none"> <li>Генеральное соглашение по торговле услугами*</li> <li>Технические барьеры в торговле*</li> <li>Соглашение по информационным технологиям*</li> <li>Торговые аспекты прав интеллектуальной собственности*</li> <li>Соглашение об упрощении процедур торговли</li> </ul> <p><b>ЮНСИТРАЛ</b></p> <ul style="list-style-type: none"> <li>Проект положений по автоматизированному заключению договоров</li> </ul> <p><b>Интеллектуальная собственность</b></p> <p><b>ВОИС</b></p> <ul style="list-style-type: none"> <li>Римская конвенция об охране прав артистов-исполнителей, производителей фонограмм и вещателей организаций*</li> <li>Бернская конвенция по охране литературных и художественных произведений*</li> <li>Пекинский договор по аудиовизуальным исполнениям*</li> <li>Договор о патентной кооперации*</li> </ul> | <p><b>Наркотики и преступность</b></p> <p><b>Управление Организации Объединенных Наций по наркотикам и преступности</b></p> <ul style="list-style-type: none"> <li>Киотская декларация</li> </ul> <p><b>ЮНИКРИ</b></p> <ul style="list-style-type: none"> <li>Политические рамки для ответственного установления ограничений в отношении распознавания лиц. Пример использования: расследования, проводимые правоохранительными органами.</li> <li>Руководство по ответственному использованию инноваций, связанных с ИИ, в правоприменении</li> </ul> <p><b>КТУ ООН</b></p> <ul style="list-style-type: none"> <li>Глобальная контртеррористическая стратегия Организации Объединенных Наций: восьмой обзор (A/RES/77/298)</li> </ul> <p><b>Образование</b></p> <p><b>ЮНЕСКО</b></p> <ul style="list-style-type: none"> <li>Руководство по использованию генеративного ИИ в сфере образования и исследований</li> <li>Проект моделей компетенций, касающихся ИИ, для студентов и преподавателей</li> <li>Компетенции в области ИИ и цифровой трансформации для гражданских служащих</li> </ul> <p><b>Прочее</b></p> <p><b>ООН-Хабитат</b></p> <ul style="list-style-type: none"> <li>Рамочные параметры оценки рисков, связанных с ИИ</li> <li>Международные руководящие принципы человекоцентричных «умных» городов</li> </ul> <p><b>Организация Объединенных Наций по промышленному развитию</b></p> <ul style="list-style-type: none"> <li>Декларация Абу-Даби, ЮНИДО ГК.18</li> </ul> <p><b>Управление Организации Объединенных Наций по снижению риска бедствий</b></p> <ul style="list-style-type: none"> <li>Сендайская рамочная программа по снижению риска бедствий*</li> </ul> |  |
| <p><b>ОБОЗНАЧЕНИЯ:</b></p> <p><b>Применимо к ИИ</b></p> <p>Может быть применимо к ИИ</p> <p>* Документ, обязательный для исполнения.</p>   |  |  |  |   |  |

Источник: «Информационный документ системы Организации Объединенных Наций по управлению ИИ: анализ существующих в системе Организации Объединенных Наций институциональных моделей, функций и международных нормативных рамок, применимых к управлению ИИ», 28 февраля 2024 года.

17

В ходе опроса, проведенного Координационным советом руководителей системы Организации Объединенных Наций в феврале 2024 года среди 57 структур Организации Объединенных Наций, было получено 50 документов, касающихся управления ИИ; ответы представили 44 из 57 структур, в том числе Экономическая комиссия для Латинской Америки и Карибского бассейна; Экономическая и социальная комиссия для Азии и Тихого океана; Экономическая и социальная комиссия для Западной Азии; Продовольственная и сельскохозяйственная организация Объединенных Наций (ФАО); Международное агентство по атомной энергии (МАГАТЭ); Международная организация гражданской авиации (ИКАО); Международный фонд сельскохозяйственного развития; МОТ; Международный валютный фонд; Международная организация по миграции; Международный торговый центр; Международный союз электросвязи (МСЭ); Структура Организации Объединенных Наций по вопросам гендерного равенства и расширения прав и возможностей женщин (Структура «ООН-женщины»); Объединенная программа Организации Объединенных Наций по ВИЧ/СПИДУ (ЮНЭЙДС); Конференция Организации Объединенных Наций по торговле и развитию (ЮНКТАД); Департамент по экономическим и социальным вопросам; Департамент глобальных коммуникаций; Канцелярия Генерального секретаря; Управление по координации гуманитарных вопросов; Управление Верховного комиссара Организации Объединенных Наций по правам человека; Контртеррористическое управление; Управление по вопросам разоружения; Управление информационно-коммуникационных технологий; КПГСТ; Программа развития Организации Объединенных Наций (ПРООН); Управление Организации Объединенных Наций по снижению риска бедствий; Программа Организации Объединенных Наций по окружающей среде; Канцелярия Посланника Генерального секретаря по вопросам технологий; ЮНЕСКО; Рамочная конвенция Организации Объединенных Наций об изменении климата; Фонд Организации Объединенных Наций в области народонаселения; Управление Верховного комиссара Организации Объединенных Наций по делам беженцев (УВКБ ООН); Детский фонд Организации Объединенных Наций; Межрегиональный научно-исследовательский институт Организации Объединенных Наций по вопросам преступности и правосудия; Организация Объединенных Наций по промышленному развитию; Управление Организации Объединенных Наций по наркотикам и преступности/Отделение Организации Объединенных Наций в Вене; Управление Организации Объединенных Наций по обслуживанию проектов; Ближневосточное агентство Организации Объединенных Наций для помощи палестинским беженцам и организации работ; Университет Организации Объединенных Наций; Добровольцы Организации Объединенных Наций; Всемирная торговая организация; Всемирный почтовый союз; Группа Всемирного банка; Всемирная продовольственная программа; Всемирная организация здравоохранения (ВОЗ); и Всемирная организация интеллектуальной собственности (ВОИС). См. «United Nations system white paper on AI governance: an analysis of the UN system's institutional models, functions, and existing international normative frameworks applicable to AI governance» (URL: <https://unsceb.org/united-nations-system-white-paper-ai-governance>).

в конкретных областях и без всеобъемлющей стратегии. Возникающие в результате этого пробелы в координации приводят к дублированию и препятствуют совместимости и результативности предпринимаемых усилий.

- 76** Количество и разнообразие подходов свидетельствуют о том, что система Организации Объединенных Наций реагирует на возникающую проблему. При надлежащей организации и в сочетании с процессами, использующими комплексный подход, эти усилия могут обеспечить эффективный и устойчивый путь к инклюзивному международному управлению ИИ в конкретных областях. Это позволило бы оказывать ощутимое, согласованное и скоординированное воздействие на такие сферы, как здравоохранение, образование, технические стандарты и этика, а не просто способствовать росту числа инициатив и институтов в этой растущей области. Международное право, включая международные нормы в области прав человека, обеспечивает общую нормативную основу для всех усилий, связанных с ИИ, тем самым способствуя координации и согласованности.
- 77** Хотя деятельность многих структур Организации Объединенных Наций затрагивает вопросы управления ИИ, их конкретные мандаты означают, что ни одна из них не делает этого всеобъемлющим образом, а назначенные ими правительственные координационные центры также специализированы<sup>18</sup>. Это ограничивает способность существующих структур Организации Объединенных Наций самостоятельно решать многогранные проблемы, связанные с ИИ, в глобальном масштабе. На национальном и региональном уровнях эти пробелы устраняются новыми учреждениями<sup>19</sup>, такими как институты безопасности ИИ или ведомства по вопросам ИИ, которые обеспечивают надлежащий сквозной подход.

## С. Пробелы в реализации

- 78** Вместе с тем представительства и координации недостаточно. Необходимы действия и последующие шаги для обеспечения того, чтобы обязательства в отношении эффективного управления воплощались в ощутимые результаты на практике. Нужны дополнительные усилия для обеспечения подотчетности. Давление со стороны коллег и взаимное обучение – вот два элемента, которые могут способствовать подотчетности.
- 79** Взаимодействие с частным сектором будет иметь не меньшее значение для обеспечения реальной подотчетности и возмещения ущерба. Организация Объединенных Наций имеет опыт такого взаимодействия в рамках Руководящих принципов предпринимательской деятельности в аспекте прав человека. В не меньшей степени нам потребуется активное участие гражданского общества и научных экспертов, с тем чтобы правительства и частные компании были добросовестны в своих обязательствах и заявлениях.
- 80** Отсутствие механизмов, способствующих использованию преимуществ ИИ для общественного блага как внутри стран, так и в отношениях между ними, является одним из основных пробелов в реализации. Многие страны разработали национальные стратегии по развитию инфраструктуры и кадровых ресурсов, связанных с ИИ, и появилось несколько инициатив по оказанию международной помощи<sup>20</sup>. Однако такие инициативы недостаточно связаны между собой и страдают от нехватки ресурсов.

18 Например, министерства образования, науки и культуры (ЮНЕСКО); телекоммуникации или ИКТ (МСЭ); промышленность (Организация Объединенных Наций по промышленному развитию); и вопросы занятости (МОТ).

19 Включая учреждения, созданные в Канаде, Республике Корея, Сингапуре, Соединенном Королевстве, Соединенных Штатах, Японии и Европейском союзе.

20 В рамках усилий на национальном уровне можно продолжать использовать инструменты диагностики, такие как Методология оценки готовности к ИИ ЮНЕСКО, для выявления недостатков на уровне стран, а международная сеть поможет их устранить.

**81** На глобальном уровне взаимная увязка национальных и региональных инициатив по развитию потенциала и объединение ресурсов для поддержки тех стран, которые не участвуют в таких усилиях, может помочь обеспечить, чтобы ни одна страна не осталась

в стороне от использования возможностей, связанных с ИИ. Еще одним важным пробелом в реализации является отсутствие специального фонда для наращивания потенциала в области ИИ, несмотря на существование некоторых механизмов финансирования цифрового потенциала (вставка 8).

## Вставка 8: Пробелы в глобальном финансировании потенциала

Насколько известно Консультативному органу, в настоящее время не существует глобальных фондов для наращивания потенциала в области ИИ, обладающих достаточным масштабом и мандатом для финансирования значительных инвестиций, необходимых для преодоления разрыва в области ИИ.

По ориентировочным оценкам, необходимая сумма составляет от 350 млн до 1 млрд долл. США в год<sup>а</sup>, включая неденежные взносы частного сектора, которые должны способствовать наращиванию потенциала ИИ по всем направлениям, включая кадры, вычислительные ресурсы, данные для обучения, разработку моделей и междисциплинарное сотрудничество в области применения ИИ. Примерами существующих многосторонних механизмов финансирования являются:

### а) Совместный фонд для ЦУР

Этот фонд является широким и охватывает все ЦУР, а также меры реагирования на чрезвычайные ситуации. Он поддерживает инициативы на страновом уровне по комплексной политике Организации Объединенных Наций и стратегической финансовой поддержке стран для продвижения ЦУР. Фонд помогает Организации Объединенных Наций осуществлять и стимулировать финансирование и программирование деятельности по достижению ЦУР. Начиная с 2017 года 30 участвующих организаций системы Организации Объединенных Наций получили в общей сложности 223 млн долл. США. Фонд не финансирует напрямую национальные правительства, сообщества или организации, а также не финансирует трансграничные инициативы.

В 2023 году у фонда было около 16 доноров, которые внесли в него в общей сложности 57,7 млн долл. США, а в 2024 году – 58,8 млн долл. США. С 2017 года частный сектор внес 83 155 долл. США, а в 2023 и 2024 годах взносы со стороны частного сектора отсутствовали.

Большая часть средств фонда (60 процентов) направляется на реализацию мероприятий по линии пяти ЦУР: цели 2 (ликвидация голода), цели 5 (гендерное равенство), цели 7 (недорогостоящая и чистая энергия), цели 9 (индустриализация, инновации и инфраструктура) и цели 17 (партнерство).

<sup>а</sup> Менее одного процента предполагаемой годовой суммы инвестиций частного сектора в ИИ в 2023 году.

## Вставка 8: Пробелы в глобальном финансировании потенциала ИИ (продолжение)

---

В рамках направления «Цифровая трансформация политики» (запущенного в 2023 году) фонд профинансировал один проект стоимостью 250 000 долл. США, который был поровну распределен между Международным союзом электросвязи (МСЭ) и Программой развития Организации Объединенных Наций (ПРООН). По состоянию на конец 2023 финансового года коэффициент освоения средств составил 2,27 процента. Мероприятия по цифровой трансформации составляют небольшую часть деятельности фонда и, как правило, связаны с другими ЦУР (включая, в частности, обеспечение связи и цифровой инфраструктуры для поддержки предоставления услуг, например, в малых островных развивающихся государствах).

### **б) Всемирный банк, Партнерство в целях развития цифровых технологий**

Этот фонд оказывает поддержку странам в разработке и осуществлении цифровой трансформации, уделяя особое внимание инфраструктуре широкополосной связи, доступу и использованию, цифровой общественной инфраструктуре, а также генерированию, доступности и использованию данных. К концу 2022 года фонд инвестировал 10,7 млрд долл. США более чем в 80 стран.

Партнерство включает в себя целевой фонд, связанный с кибербезопасностью, созданный несколькими донорами (Германия, Королевство Нидерландов, Эстония и Япония) для поддержки развития национального потенциала в области кибербезопасности.

## 4. Укрепление глобального сотрудничества

**82** Очертив проблему глобального дефицита управления, мы переходим к рекомендациям, направленным на то, чтобы устранить основные пробелы в ближайшем будущем.

**83** Наши рекомендации продвигают целостную концепцию глобального сетевого, гибкого и динамичного подхода к управлению ИИ в интересах человечества, предполагающего наличие общего понимания, общей основы и общих выгод в целях расширения представительства, обеспечения координации и укрепления реализации (см. рис. 10). Только такой инклюзивный и всеобъемлющий подход к управлению ИИ может решить многогранные и меняющиеся задачи, а также реализовать

возможности, которые предоставляет ИИ в глобальном масштабе, способствуя при этом международной стабильности и справедливому развитию.

**84** Руководствуясь принципами, изложенными в нашем промежуточном докладе (см. пункт 47), мы подготовили предложения, направленные на устранение пробелов и обеспечение согласованности в быстро формирующейся экосистеме международных мер и инициатив по управлению ИИ, что поможет избежать фрагментации и упущенных возможностей. Для эффективной поддержки этих мер и налаживания партнерских отношений с другими учреждениями мы предлагаем

**Рисунок 10: Обзор рекомендаций и их роли в устранении пробелов в глобальном управлении ИИ**

| Цель  | Расширение представленности   | Содействие координации | Укрепление реализации |
|---|---|------------------------|-----------------------|
| <b>Общее понимание</b><br>Международная научная группа по вопросам ИИ   | ✓   | ✓                      |                       |
| <b>Общие позиции</b><br>Политический диалог по вопросам управления ИИ<br>Обмен стандартами в сфере ИИ         | ✓   | ✓                      | (✓)                   |
| <b>Общие выгоды</b><br>Сеть для наращивания потенциала<br>Глобальный фонд для ИИ<br>Глобальная сеть данных ИИ | ✓   | ✓                      | ✓                     |
| <b>Согласованные усилия</b><br>Подразделение по вопросам ИИ в составе Секретариата                            | Консультирование Генерального секретаря по вопросам, связанным с ИИ, работа по достижению единства мнений в рамках системы Организации Объединенных Наций, привлечение государств и заинтересованных сторон, установление партнерских отношений и взаимодействия с другими процессами и учреждениями, а также поддержка других предложений по мере необходимости. |                        |                       |

создать легкую и гибкую структуру, которая станет олицетворением согласованных усилий: подразделение по вопросам ИИ в составе Секретариата Организации Объединенных Наций, близкое к Генеральному секретарю, которое будет работать как «клей», объединяющий предложенные здесь инициативы.

- 85** Организация Объединенных Наций далека от совершенства. Тем не менее легитимность, вытекающая из ее уникальной инклюзивности в сочетании с ее укорененностью в обязательных для исполнения нормах международного права, включая международные нормы в области прав человека, дает надежду на управление ИИ на благо и в защиту интересов человечества справедливым, эффективным и действенным образом<sup>21</sup>.

## А. Общее понимание

- 86** Глобальный подход к управлению ИИ начинается с общего понимания его потенциала, возможностей, рисков и неопределенностей.
- 87** Сфера ИИ быстро развивается, порождая чрезмерное количество информации, вследствие чего становится всё труднее отличать шумиху от реальности. Это может привести к путанице, помешать достижению общего понимания и дать преимущества крупным компаниям, занимающимся разработкой ИИ, в ущерб политикам, гражданскому обществу и населению.
- 88** Кроме того, недостаток международного научного сотрудничества и обмена информацией может породить глобальные недоразумения и подорвать международное доверие.
- 89** Государствам-членам необходимы своевременные, беспристрастные и достоверные научные знания и информация об ИИ для формирования

общего фундаментального понимания во всем мире, а также для уравнивания информационной асимметрии между компаниями, располагающими дорогостоящими лабораториями ИИ, и остальным миром, в том числе посредством обмена информацией между компаниями ИИ и более широким сообществом ИИ.

- 90** Такие меры наиболее эффективно работают на всемирном уровне, позволяющем осуществлять совместные инвестиции в глобальное общественное благо и сотрудничать в интересах общества, сводя в единое целое разрозненные и дублирующие друг друга усилия.

## Международная научная группа по вопросам ИИ

### Рекомендация 1: международная научная группа по вопросам ИИ

Мы рекомендуем создать независимую международную научную группу по вопросам ИИ, которая состояла бы из различных междисциплинарных экспертов в этой области, работающих в своем личном качестве на добровольной основе. При поддержке предлагаемого подразделения Организации Объединенных Наций по вопросам ИИ и других соответствующих учреждений системы Организации Объединенных Наций, в партнерстве с другими соответствующими международными организациями, ей был бы вверен мандат, включающий в себя:

- a. публикацию ежегодного доклада с обзором возможностей, рисков и неопределенностей, связанных с ИИ, и с выявлением областей научного консенсуса в отношении технологических тенденций и сфер, где необходимы дополнительные исследования;
- b. подготовку ежеквартальных тематических исследовательских обзоров, посвященных

<sup>21</sup> Такое управление также должно быть инклюзивным и последовательным, способствуя укреплению глобального мира и безопасности.

областям, в которых ИИ может помочь в достижении ЦУР, с упором на области, представляющие общественный интерес, которые могут быть недостаточно охвачены;  
и

с. выпуск специальных докладов по возникающим вопросам, в том числе о появлении новых рисков или существенных пробелов в механизмах управления.

- 91** Прецедент создания подобных учреждений имеется. В качестве примеров можно привести Научный комитет Организации Объединенных Наций по действию атомной радиации, Межправительственную научно-политическую платформу по биоразнообразию и экосистемным услугам (МПБЭУ), Научный комитет по антарктическим исследованиям и Межправительственную группу экспертов по изменению климата (МГЭИК).
- 92** Эти структуры известны своими системными подходами к сложным, повсеместным проблемам, затрагивающим различные отрасли и группы населения планеты. Однако несмотря на то, что они могут служить источниками вдохновения, ни одна из них не подходит для оценки технологий ИИ и не должна воспроизводиться напрямую. Вместо этого необходим индивидуальный подход.
- 93** Опираясь на такие прецеденты, независимая, международная и междисциплинарная научная группа по вопросам ИИ могла бы систематизировать и продвигать передовые исследования, с тем чтобы информировать тех, кто ищет данные о научных взглядах на технологии ИИ или их применение из беспристрастного, заслуживающего доверия источника. Примером одного из видов вопросов, в решение которых группа могла бы внести свой вклад, является продолжающаяся дискуссия об открытых и закрытых системах ИИ, о которой говорится во вставке 9.
- 94** Деятельность научной группы под эгидой Организации Объединенных Наций имела бы широкую сферу охвата, распространяясь на весь спектр имеющихся приоритетов на инклюзивной и комплексной основе. Это могло бы включать в себя поиск экспертных знаний о возможностях, связанных с ИИ, и содействие «глубокому погружению» в прикладные области ЦУР, такие как здравоохранение, энергетика, образование, финансы, сельское хозяйство, климат, торговля и занятость.
- 95** При оценке рисков можно также использовать результаты работы других научно-исследовательских инициатив в области ИИ, причем Организация Объединенных Наций может служить уникальной и надежной «тихой гаванью» для обмена идеями о современном состоянии дел. Международное право, и в том числе международные нормы в области прав человека, станет компасом для выявления имеющихся рисков. Благодаря объединению знаний в странах и компаниях, которые в противном случае не смогли бы принять участие в обсуждении, группа экспертов под эгидой Организации Объединенных Наций может помочь исправить ошибочные представления и укрепить доверие во всем мире.
- 96** Такая научная группа не обязательно будет проводить собственные исследования, но станет катализатором сетевых действий.<sup>22</sup> Она могла бы обобщать, анализировать и переводить разработки в области ИИ для своей аудитории, освещая потенциальные варианты использования. Это уменьшит информационную асимметрию, поможет избежать неправильного направления инвестиций и обеспечит поток информации через глобальную сеть экспертов.
- 97** Работа группы была бы ориентирована на три основные целевые аудитории.
- а. Первая из них – глобальное научное сообщество<sup>23</sup>. Переход фундаментальных исследований в области ИИ в руки частных корпораций, отчасти обусловленный

<sup>22</sup> В частности, это могло бы делаться на основе деятельности уже существующих отраслевых и региональных групп.

<sup>23</sup> Группа могла бы также вести информационно-разъяснительную работу, ориентированную на более широкую аудиторию, включая гражданское общество и широкую общественность.

стоимостью вычислительных мощностей, вызвал опасения, что такие исследования могут быть чрезмерно зависимы от финансовых интересов. Научная группа могла бы способствовать расширению исследований в государственных учреждениях по всему миру, ориентированных на общественное благо.

- b. Во-вторых, регулярные независимые оценки позволят получать информацию, имеющую полезное значение для государств-членов, политиков и участников других процессов, рекомендованных в настоящем докладе. Ежегодное исследование рисков, проводимое мировыми экспертами, поможет сформировать повестку дня диалогов по управлению ИИ, предложенных в рекомендации 2. Доклад о современном положении дел в отрасли послужит основой для разработки стандартов, предложенных в рекомендации 3, а также для создания сети по наращиванию потенциала, предложенной в рекомендации 4.
- c. В-третьих, благодаря своим общедоступным докладам группа могла бы служить источником объективной высококачественной информации для общественности.

**98** Глобальный охват сетей, уникально доступных через Организацию Объединенных Наций, позволит обеспечить общее понимание на самой широкой основе, делая выводы актуальными для различных социально-экономических и географических условий. Таким образом, группа может задействовать Организацию Объединенных Наций в качестве надежной платформы для достижения инклюзивного сетевого междисциплинарного понимания заинтересованных сторон.

**99** Группа может быть создана на первоначальный период в 3–5 лет (с возможностью продления при условии проведения обзора Генеральным секретарем) и могла бы функционировать на основе следующих параметров.

- a. Вначале группа могла бы состоять из 30–50 членов, назначаемых на основе сочетания предложений государств-членов и самовыдвижения, аналогично тому, как был создан Консультативный орган. Она должна быть ориентирована на научный опыт в различных дисциплинах и должна обеспечивать разнообразное представительство по регионам и полу, а также отражать междисциплинарный характер ИИ. Состав может периодически меняться в рамках общего мандата, рассчитанного на 3–5 лет.
- b. Группа будет проводить виртуальные встречи (и очные пленарные заседания, возможно, дважды в год). Заседания могут проводиться поочередно в городах, где размещаются соответствующие структуры Организации Объединенных Наций, в том числе в странах глобального Юга. Следует поощрять формирование тематических рабочих групп, добавляя, по мере необходимости, дополнительных членов и привлекая сети академических партнеров. Можно рассмотреть вопрос о приглашении к участию в этих рабочих группах соответствующих подразделений Организации Объединенных Наций<sup>24</sup>.
- c. Группа будет работать независимо, особенно в том, что касается ее выводов и заключений, при поддержке группы представителей системы Организации Объединенных Наций, сформированной из сотрудников предлагаемого подразделения по вопросам ИИ и соответствующих учреждений Организации Объединенных Наций, таких как МСЭ и Организация Объединенных Наций по вопросам образования, науки и культуры (ЮНЕСКО).
- d. Она должна сотрудничать с другими международными организациями, такими как ОЭСР и Глобальное партнерство по искусственному интеллекту, и опираться на результаты исследований, проводимых в рамках других соответствующих процессов,

таких как недавний научный доклад о рисках, связанных с продвинутыми моделями ИИ, подготовленный по заказу Соединенного Королевства<sup>25</sup>, и соответствующих региональных организаций.

- е. Руководящий комитет будет разрабатывать программу исследований, обеспечивая учет мнений и этических соображений, следить за распределением ресурсов, развивать сотрудничество с сетью академических институтов и другими заинтересованными сторонами, а также анализировать деятельность и результаты работы группы.

**100** Опираясь на уникальную объединяющую силу Организации Объединенных Наций и широкий глобальный охват групп заинтересованных сторон, международная научная группа может обеспечить надежные процессы и результаты научного сотрудничества и выровнять информационную асимметрию таким образом, чтобы устранить пробелы в представительстве и координации, указанные в пунктах 66 и 73, тем самым способствуя справедливому и эффективному международному управлению ИИ.

## Вставка 9: Открытые и закрытые системы ИИ

Среди тем, обсуждавшихся в ходе наших консультаций, были постоянные дебаты об открытых и закрытых системах ИИ. Системы ИИ, открытые в той или иной степени, часто называют «ИИ с открытыми спецификациями», но это несколько неверное название, если сравнивать его с программным обеспечением (кодом) с открытыми спецификациями. Важно понимать, что открытость систем ИИ — это скорее спектр, а не отдельный атрибут.

В одной из статей объясняется, что «полностью закрытая система ИИ доступна только определенной группе. Это может быть компания-разработчик ИИ или определенная группа внутри нее, в основном для внутренних исследований и разработок. С другой стороны, более открытые системы могут предоставлять публичный доступ или делать доступными определенные части, такие как данные, код или характеристики модели, с тем чтобы облегчить внешнюю разработку ИИ»<sup>a</sup>.

Системы ИИ с открытым исходным кодом в области генеративного ИИ несут в себе как риски, так и возможности. Компании часто ссылаются на «безопасность ИИ» как на причину отказа от раскрытия спецификаций системы, что отражает постоянные трения между открытыми и закрытыми подходами в отрасли. Дискуссии обычно разворачиваются вокруг двух крайностей: полной открытости, которая подразумевает предоставление всех компонентов модели и наборов данных, и частичной открытости, которая предполагает раскрытие только весовых коэффициентов модели.

Системы ИИ с открытыми спецификациями способствуют инновациям и часто являются обязательным условием государственного финансирования. На открытой стороне спектра, когда базовый код находится в свободном доступе, разработчики со всего мира могут экспериментировать, улучшать и создавать новые приложения. Это способствует развитию среды сотрудничества, в которой легко происходит обмен идеями и опытом. Некоторые лидеры отрасли утверждают, что такая открытость жизненно важна для инноваций и экономического роста.

<sup>a</sup> Angela Luna, «The open or closed AI dilemma», 2 May 2024. См. URL: <https://bipartisanpolicy.org/blog/the-open-or-closed-ai-dilemma/>.

<sup>25</sup> International Scientific Report on the Safety of Advanced AI: Interim Report. См. URL: [www.gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai](http://www.gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai).

## Вставка 9: Открытые и закрытые системы ИИ (продолжение)

Однако в большинстве случаев модели ИИ с открытым исходным кодом доступны в виде интерфейсов прикладного программирования. В этом случае исходный код не раскрывается, исходные веса никогда не изменяются, а обновления моделей становятся новыми моделями.

Кроме того, модели с открытыми спецификациями, как правило, меньше и прозрачнее. Такая прозрачность способствует укреплению доверия, позволяет заблаговременно решать этические вопросы, а также поддерживает валидацию и воспроизведение, поскольку пользователи могут изучить внутреннюю работу системы ИИ, понять процесс принятия решений и выявить потенциальные предубеждения.

Закрытые системы ИИ дают разработчикам больше возможностей для контроля. Кроме того, системы с закрытыми спецификациями могут быть более упорядоченными и эффективными, поскольку кодовая база не подвергается постоянному развитию за счет публичного вклада. Многие компании считают полную открытость непрактичной и пропагандируют частичную открытость как единственно возможный вариант. Однако такая точка зрения упускает из виду потенциал сбалансированного подхода, который позволяет достичь «конструктивной открытости»<sup>b</sup>.

Конструктивная открытость существует между двумя крайними точками спектра и может быть адаптирована к различным случаям использования. Этот сбалансированный метод способствует безопасному, инновационному и инклюзивному развитию ИИ, позволяя осуществлять общественный контроль и независимый аудит раскрытых данных об обучении и тонкой настройке. Поскольку открытость – это нечто большее, чем просто обмен весами моделей, она может способствовать инновациям и инклюзивности, помогая применять их в научных исследованиях и образовании.

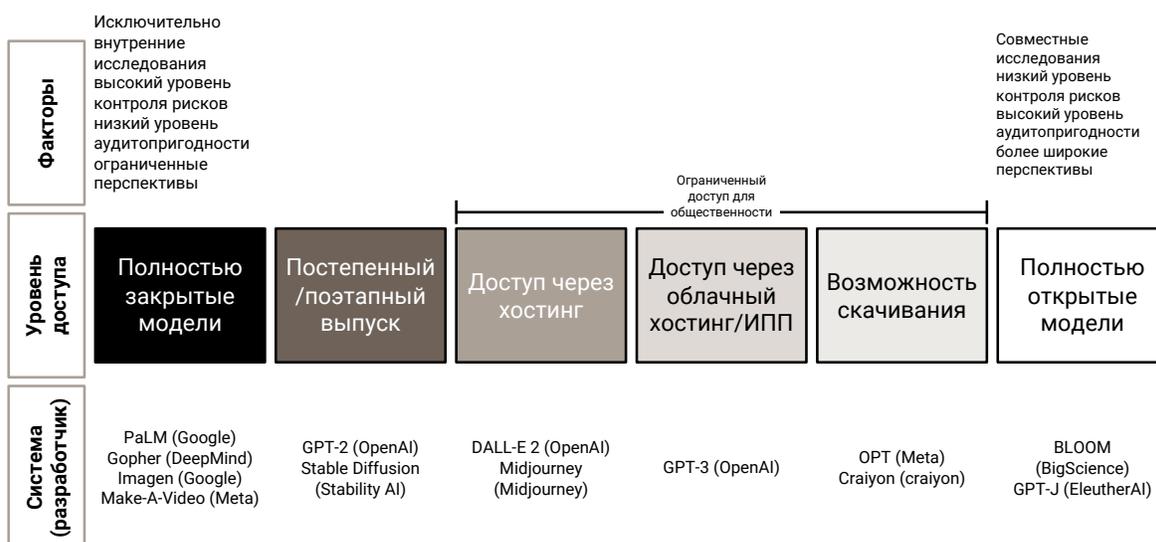
Определение «ИИ с открытыми спецификациями» находится в процессе эволюции<sup>c</sup>, и на него часто влияют корпоративные интересы, как показано на рисунке 11. Для решения этой проблемы мы рекомендуем инициировать процесс, координируемый предложенной выше международной научной группой, для разработки всестороннего и поэтапного подхода к открытости. Это позволит применять конструктивные, основанные на фактах подходы к открытости, помогая пользователям и политикам делать осознанный выбор моделей и архитектур ИИ.

b Inspired by Andreas Liesenfeld and Mark Dingemane, «Rethinking open source generative AI: open-washing and the EU AI Act», The 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24) (June 2024).

c The Open Source AI Definition – draft v. 0.0.3. См. URL: <https://opensource.org/deepdive/drafts/the-open-source-ai-definition-draft-v-0-0-3>.

## Вставка 9: Открытые и закрытые системы ИИ (продолжение)

Рисунок 11: Корпоративные интересы и открытость



Источник: Irene Solaiman, "The gradient of generative AI release: methods and considerations", *Proceedings of the 2023 Association for Computing Machinery (ACM) Conference on Fairness, Accountability, and Transparency* (June 2023), pp. 111–122.

Раскрытие данных, даже если оно ограничено ключевыми элементами, необходимо для понимания эффективности модели, обеспечения воспроизводимости и оценки юридических рисков. Разъяснение разных уровней открытости может помочь противостоять мнимой корпоративной «открытости» и способствовать развитию прозрачной технологической экосистемы.

Также важно, чтобы по мере развития технологии мы рассматривали режимы управления для применения как открытых, так и закрытых систем ИИ. Нам необходимо разработать руководящие принципы ответственного отношения к ИИ, обязательные нормы и измеримые стандарты для разработчиков и проектировщиков продуктов и услуг, в которых используются технологии ИИ, а также для их пользователей и всех участников, вовлеченных в их жизненный цикл.

## В. Общие позиции

**101** Наряду с общим пониманием ИИ необходимо выработать подходы к управлению, которые были бы совместимы в разных юрисдикциях и опирались на международные нормы, такие как Всеобщая декларация прав человека (см. принцип 5 выше).

**102** Это необходимо на глобальном уровне не только для справедливого представительства, но и для предотвращения «гонки на понижение» в сфере регулирования, уменьшения межгосударственных трений, обеспечения максимальной технической и онтологической совместимости, а также обнаружения и реагирования на инциденты, возникающие в результате принятия решений на протяжении жизненного цикла ИИ, которые охватывают множество юрисдикций.

## Политический диалог по вопросам управления ИИ

### Рекомендация 2: политический диалог по вопросам управления ИИ

Мы рекомендуем организовать межправительственный и многосторонний политический диалог по управлению ИИ на полях существующих заседаний в рамках Организации Объединенных Наций и проводить его раунды два раза в год. Его целью будет:

- a. обмен передовым опытом в области управления ИИ, способствующим развитию при соблюдении, защите и реализации всех прав человека, включая использование возможностей, а также управление рисками;
- b. содействие достижению общего понимания в отношении реализации мер по управлению ИИ разработчиками и пользователями из частного и государственного секторов в целях повышения международной совместимости механизмов управления ИИ;
- c. добровольный обмен информацией о серьезных инцидентах, связанных с ИИ, которые создают значительную нагрузку на государственные учреждения или превышают их возможности в плане реагирования на них; и
- d. по мере необходимости – обсуждение докладов международной научной группы по вопросам ИИ.

- 103** В настоящее время международное управление ИИ представляет собой в лучшем случае фрагментарную картину. Сто восемнадцать стран не являются участниками ни одной из семи последних известных межправительственных инициатив по управлению ИИ, не относящихся к системе

Организации Объединенных Наций<sup>26</sup> (см. рис. 8). Пробелы в представительстве наблюдаются даже среди 60 стран, обладающих наибольшим потенциалом в области ИИ, что подчеркивает избирательность современного международного управления ИИ (см. рис. 12).

- 104** Необходимо всеохватывающий политический форум, с тем чтобы все государства-члены, опираясь на опыт заинтересованных сторон, могли обмениваться передовым опытом, способствующим развитию при соблюдении, защите и осуществлении всех прав человека, продвигать взаимно совместимые подходы к управлению и отслеживать общие риски, требующие дальнейших мер политического характера.
- 105** Это не означает глобального управления всеми аспектами ИИ (что невозможно и нежелательно, учитывая расхождения в интересах и приоритетах государств). Тем не менее обмен мнениями о развитии ИИ и мерах реагирования в области политики может заложить основу для международного сотрудничества.
- 106** Организация Объединенных Наций обладает уникальными возможностями для содействия проведению таких диалогов на инклюзивной основе таким образом, чтобы помочь государствам-членам эффективно сотрудничать. Существующий и формирующийся комплекс норм системы Организации Объединенных Наций может заложить прочную нормативную основу для согласованных действий, опираясь на Устав Организации Объединенных Наций, права человека и другие нормы международного права, включая экологическое право и международное гуманитарное право, а также ЦУР и другие международные обязательства<sup>27</sup>.

<sup>26</sup> Эти инициативы не всегда поддаются прямому сравнению. Некоторые из них отражают работу существующих международных или региональных организаций, в то время как другие основаны на специальных приглашениях стран-единомышленников.

<sup>27</sup> Например, Устав Организации Объединенных Наций (преамбула, цели и принципы, а также статьи 13, 55, 58 и 59). См. также основные международные документы по правам человека (Всеобщая декларация прав человека; Международный пакт о гражданских и политических правах; Международный пакт об экономических, социальных и культурных правах; Международная конвенция о ликвидации всех форм расовой дискриминации; Конвенция о правах ребенка; Конвенция о ликвидации всех форм дискриминации в отношении женщин; Конвенция против пыток; Конвенция о правах инвалидов; Конвенция о правах мигрантов; Международная конвенция для защиты всех лиц от насильственных исчезновений); международно-правовые документы в области прав человека (Женевские конвенции; Конвенция о конкретных видах обычного оружия; Конвенция о геноциде; Гаагская конвенция); документы по смежным принципам, таким как различие, пропорциональность и предосторожность, а также 11 принципов по системам смертоносного автономного оружия, принятых в рамках Конвенции о конкретных видах обычного оружия); документы по разоружению и контролю над вооружениями в части запрета на оружие массового уничтожения (Договор о нераспространении ядерного оружия; Конвенция о химическом оружии; Конвенция о биологическом оружии); документы по экологическому праву (Рамочная конвенция Организации Объединенных Наций об изменении климата; Конвенция о запрещении военного или любого иного враждебного использования средств воздействия на природную среду); Парижское соглашение и связанные с ним принципы, такие как принцип предосторожности, принцип интеграции и участие общественности; а также не имеющие обязательной силы обязательства по Повестке дня в области устойчивого развития на период до 2030 года, гендерным вопросам и этике, например, Рекомендация ЮНЕСКО по этике искусственного интеллекта.

# Рисунок 12: 60 стран-лидеров в сфере ИИ (индекс Tortoise за 2023 год) в контексте выборки из наиболее важных многосторонних инициатив в области управления ИИ с межправительственным компонентом

■ участник

| Страна / сторона*   | Глобальный индекс Tortoise в сфере ИИ (2023 год) | Количество инициатив, в которых участвует страна/сторона | В хронологическом порядке → |                             |                                   |                                     |                               |   |  |
|---|--|--|-----------------------------|-----------------------------|-----------------------------------|-------------------------------------|-------------------------------|---|--|
|   |  |  | Принципы ИИ ОЭСР (2019 год) | Принципы ИИ Г-20 (2019 год) | Редакционная группа CE (2022 год) | Заявление министров ГПНИ (2022 год) | Декларация Блетчли (2023 год) | Заявление министров Г-7 по Хиросимском у процессу в области ИИ (2023 год) | Сеульское заявление министров (2024 год) |
| Соединенные Штаты Америки   | 1  | 7  |                             |                             |                                   |                                     |                               |   |  |
| Китай   | 2  | 2  |                             |                             |                                   |                                     |                               |   |  |
| Сингапур  | 3  | 4  |                             |                             |                                   |                                     |                               |   |  |
| Соединенное Королевство Великобритании и Северной Ирландии          | 4  | 7  |                             |                             |                                   |                                     |                               |   |  |
| Канада  | 5  | 7  |                             |                             |                                   |                                     |                               |   |  |
| Республика Корея  | 6  | 5  |                             |                             |                                   |                                     |                               |   |  |
| Израиль   | 7  | 5  |                             |                             |                                   |                                     |                               |   |  |
| Германия  | 8  | 7  |                             |                             |                                   |                                     |                               |   |  |
| Швейцария   | 9  | 4  |                             |                             |                                   |                                     |                               |   |  |
| Финляндия   | 10   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Королевство Нидерландов   | 11   | 5  |                             |                             |                                   |                                     |                               |   |  |
| Япония  | 12   | 7  |                             |                             |                                   |                                     |                               |   |  |
| Франция   | 13   | 7  |                             |                             |                                   |                                     |                               |   |  |
| Индия   | 14   | 4  |                             |                             |                                   |                                     |                               |   |  |
| Австралия   | 15   | 6  |                             |                             |                                   |                                     |                               |   |  |
| Дания   | 16   | 3  |                             |                             |                                   |                                     |                               |   |  |
| Швеция  | 17   | 3  |                             |                             |                                   |                                     |                               |   |  |
| Люксембург  | 18   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Ирландия  | 19   | 4  |                             |                             |                                   |                                     |                               |   |  |
| Австрия   | 20   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Испания   | 21   | 5  |                             |                             |                                   |                                     |                               |   |  |
| Бельгия   | 22   | 3  |                             |                             |                                   |                                     |                               |   |  |
| Италия  | 23   | 7  |                             |                             |                                   |                                     |                               |   |  |
| Норвегия  | 24   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Эстония   | 25   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Объединенные Арабские Эмираты                                       | 27   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Португалия  | 28   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Российская Федерация  | 29   | 1  |                             |                             |                                   |                                     |                               |   |  |
| Саудовская Аравия   | 30   | 3  |                             |                             |                                   |                                     |                               |   |  |
| Мальта  | 31   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Бразилия  | 33   | 4  |                             |                             |                                   |                                     |                               |   |  |
| Новая Зеландия  | 34   | 3  |                             |                             |                                   |                                     |                               |   |  |
| Словения  | 35   | 3  |                             |                             |                                   |                                     |                               |   |  |
| Венгрия   | 36   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Турция  | 37   | 6  |                             |                             |                                   |                                     |                               |   |  |
| Исландия  | 38   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Чили  | 39   | 3  |                             |                             |                                   |                                     |                               |   |  |
| Катар   | 40   | 0  |                             |                             |                                   |                                     |                               |   |  |
| Литва   | 41   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Малайзия  | 42   | 0  |                             |                             |                                   |                                     |                               |   |  |
| Греция  | 43   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Индонезия   | 44   | 3  |                             |                             |                                   |                                     |                               |   |  |
| Вьетнам   | 45   | 0  |                             |                             |                                   |                                     |                               |   |  |
| Колумбия  | 46   | 1  |                             |                             |                                   |                                     |                               |   |  |
| Аргентина   | 47   | 4  |                             |                             |                                   |                                     |                               |   |  |
| Словакия  | 48   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Мексика   | 49   | 5  |                             |                             |                                   |                                     |                               |   |  |
| Египет  | 50   | 1  |                             |                             |                                   |                                     |                               |   |  |
| Уругвай   | 51   | 1  |                             |                             |                                   |                                     |                               |   |  |
| Армения   | 52   | 1  |                             |                             |                                   |                                     |                               |   |  |
| Южная Африка  | 53   | 1  |                             |                             |                                   |                                     |                               |   |  |
| Тунис   | 54   | 0  |                             |                             |                                   |                                     |                               |   |  |
| Марокко   | 55   | 0  |                             |                             |                                   |                                     |                               |   |  |
| Бахрейн   | 56   | 0  |                             |                             |                                   |                                     |                               |   |  |
| Пакистан  | 57   | 0  |                             |                             |                                   |                                     |                               |   |  |
| Шри-Ланка   | 58   | 0  |                             |                             |                                   |                                     |                               |   |  |
| Нигерия   | 59   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Кения   | 60   | 2  |                             |                             |                                   |                                     |                               |   |  |
| Европейский союз  | n/a  | 5  |                             |                             |                                   |                                     |                               |   |  |
| <b>Всего (включая те инициативы, которые не показаны в таблице)</b> |  |  | <b>47</b>                   | <b>20</b>                   | <b>58</b>                         | <b>29</b>                           | <b>29</b>                     | <b>7</b>  | <b>28</b>                                |

\* Включая такие юрисдикции, как Святой Престол и Европейский союз.

Источники:  
 • ОЭСР, Рекомендация Совета по искусственному интеллекту (принята 21 мая 2019 года), см. URL: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>.  
 • G20, Принципы ИИ (июнь 2019 года), см. URL: [https://www.mofa.go.jp/policy/economy/g20\\_summit/osaka19/pdf/documents/en/annex\\_08.pdf](https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/pdf/documents/en/annex_08.pdf).  
 • ГПНИ, 2022 год, заявление министров (22 ноября 2022 года), см. URL: [https://one.oecd.org/document/GPAI/C\(2022\)7/FINAL/en/pdf](https://one.oecd.org/document/GPAI/C(2022)7/FINAL/en/pdf).  
 • Декларация Блетчли (1 ноября 2023 года), см. URL: <https://gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>.  
 • Г-7, Хиросимский процесс F7 в области ИИ, Заявление министров цифровизации и технологий (1 декабря 2023 года), см. URL: [https://www.soumu.go.jp/hiroshimaiprocess/pdf/document02\\_en.pdf](https://www.soumu.go.jp/hiroshimaiprocess/pdf/document02_en.pdf).  
 • Совет Европы, Рамочная конвенция по искусственному интеллекту и правам человека, демократии и верховенству права (принята 17 мая 2024 года), см. URL: <https://coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence>.  
 • Сеульское заявление министров по укреплению безопасности, инновационности и инклюзивности в области ИИ, Сеульский саммит по ИИ (22 мая 2024 года).  
 • Tortoise Media, Глобальный индекс в сфере ИИ (2023 год), URL: <https://tortoisemedia.com/intelligence/global-ai/#rankings>.

**107** В сочетании с опытом международной научной группы и наращиванием потенциала (см. рекомендации 1, 4 и 5) инклюзивный диалог в Организации Объединенных Наций может помочь государствам и компаниям обновить свои подходы и методологии регулирования, с тем чтобы идти в ногу с ускоряющимся развитием ИИ на основе взаимодействия, способствующего достижению общего согласия. В данной связи могут быть востребованы некоторые отличительные особенности Организации Объединенных Наций, о которых говорится ниже.

- a. Инклюзивный диалог с опорой на комплекс норм Организации Объединенных Наций, включая Устав Организации Объединенных Наций, права человека и международное право, может способствовать «гонке на повышение» в подходах к управлению. И наоборот, в отсутствие универсального всемирного членства, присущего Организации Объединенных Наций, международные коллективные действия могут оказаться под большим прессингом, что приведет к «гонке на понижение» между юрисдикциями в области регулирования безопасности и сферы применения ИИ.
- b. Кроме того, всемирный характер членского состава Организации Объединенных Наций может способствовать координации между существующими инициативами, которые не являются глобальными по своему характеру, для обеспечения их большей совместимости. Многие участники проведенных нами консультаций призывали к тому, чтобы Организация Объединенных Наций стала ключевой площадкой для обеспечения мягкой координации между существующими региональными и многосторонними инициативами с учетом различных ценностей, присущих разным культурам, языкам и регионам.
- c. Предсказуемые, прозрачные, основанные на правилах и обоснованные процедуры

Организации могут обеспечить постоянное политическое взаимодействие, с тем чтобы сблизить страны, придерживающиеся различных позиций, и смягчить опасные разногласия. Помимо укрепления доверия, отношений и каналов коммуникации в кризисные времена, надежные инклюзивные диалоги могут способствовать формированию новых норм, корпуса обычного права, а также соглашений, которые укрепляют сотрудничество между государствами.

**108** С оперативной точки зрения:

- a. Политический диалог может начаться на полях существующих встреч в Нью-Йорке, таких как Генеральная Ассамблея<sup>28</sup>, в Женеве и городах глобального Юга.
- b. Одна часть каждого раунда диалога может быть посвящена национальным подходам под руководством государств-членов, а вторая – привлечению опыта и вклада ключевых заинтересованных сторон, в частности технологических компаний и представителей гражданского общества.
- c. Участие правительств может быть открыто для всех государств-членов или регионально сбалансированной группы (для более целенаправленного обсуждения среди формируемой на основе ротации и представительной заинтересованной подгруппы) или же сочетания того и другого, калибруемого по мере необходимости для различных пунктов или сегментов повестки дня с течением времени, по мере развития технологии и появления или усиления глобальных проблем. Фиксированная конфигурация может оказаться бесполезной, учитывая динамичный характер технологии и политического контекста.
- d. Помимо заседаний в рамках официального диалога, многостороннее взаимодействие по вопросам политики в области ИИ может

---

28 По аналогии с политическим форумом высокого уровня в контексте ЦУР, который работает под эгидой Экономического и Социального Совета.

- также использовать другие существующие механизмы, такие как форум МСЭ «ИИ во благо», ежегодное совещание Форума по управлению Интернетом, форум ЮНЕСКО по этике ИИ и электронная неделя Конференции Организации Объединенных Наций по торговле и развитию (ЮНКТАД), открытая для участия представителей всех государств-членов на добровольной основе.
- e. В соответствии с инклюзивным характером диалога повестка дня обсуждений может быть широкой, с тем чтобы охватить различные точки зрения и проблемы. Например, на встречах, проводимых два раза в год, можно было бы уделять больше внимания возможностям в различных секторах на одной встрече и тенденциям риска на другой<sup>29</sup>. Сюда можно отнести использование ИИ для достижения ЦУР, способы защиты детей, минимизацию воздействия климата, а также обмен мнениями о подходах к управлению рисками. На заседаниях также можно обсудить определения терминов, используемых в управлении ИИ, и технические стандарты ИИ, а также, при необходимости, доклады международной научной группы.
- f. Кроме того, различные заинтересованные стороны, в частности технологические компании и представители гражданского общества, могут быть приглашены к сотрудничеству через существующие институты, о которых говорится ниже, а также к проведению семинаров по

вопросам политики, посвященных конкретным аспектам управления ИИ, таким как ограничения (если таковые имеются) открытых подходов к наиболее продвинутым формам ИИ, пороговые значения для отслеживания и сообщения об инцидентах с ИИ, применение законодательства о правах человека к новым видам применения или же использование конкурентного права/ антимонопольного законодательства для решения проблемы концентрации власти среди технологических компаний<sup>30</sup>.

g. Сотрудничая с существующими структурами, такими как ОЭСР, предлагаемое подразделение по вопросам ИИ могло бы также создать базу данных с примерами управления ИИ, включая законодательство, политику и институты со всего мира, для рассмотрения в рамках политического диалога.

**109** Несмотря на то, что в 2024 году Генеральная Ассамблея приняла две резолюции по ИИ, в настоящее время в Организации Объединенных Наций не существует официального институционального диалога по вопросам управления ИИ, который четко соответствовал бы инклюзивному характеру данной рекомендации. Подобные процессы существуют на международном уровне, но в основном в региональных или многосторонних объединениях (пункт 57), которые не являются последовательно инклюзивными и глобальными.

29 Для выявления возможностей и рисков можно привлечь соответствующие подразделения системы Организации Объединенных Наций, включая МСЭ – по стандартам ИИ; МСЭ, ЮНКТАД, ПРООН и Управление по координации развития – по применению ИИ для достижения ЦУР; ЮНЕСКО – по этике и потенциалу управления; Управление Верховного комиссара Организации Объединенных Наций по правам человека (УВКПЧ) – по подотчетности в области прав человека на основе существующих норм и механизмов; Управление по вопросам разоружения – по регулированию использования ИИ в военных системах; ПРООН – по поддержке национального потенциала в области развития; Форум по вопросам управления Интернетом – по взаимодействию и диалогу с участием различных заинтересованных сторон; ВОИС, МОТ, ВОЗ, ФАО, Всемирную продовольственную программу, УВКБ ООН, ЮНЕСКО, Детский фонд Организации Объединенных Наций, Всемирную метеорологическую организацию и др. – по секторальному применению и управлению.

30 Такой формат встреч также мог бы предоставить возможность для обсуждения с участием многих заинтересованных сторон любых ужесточений механизмов глобального управления ИИ. Это может включать, например, запреты на разработку неконтролируемых или неуправляемых систем ИИ или же требования о том, чтобы все системы ИИ были достаточно прозрачными, с тем чтобы их последствия можно было отследить до юридического субъекта, который может взять на себя ответственность за них.

**110** Дополняя динамичный процесс многосторонних и региональных саммитов по ИИ<sup>31</sup>, Организация Объединенных Наций может предложить стабильную площадку для диалога по вопросам управления ИИ. Инклюзивность как принцип – важнейшее требование для того, чтобы играть стабилизирующую роль в геополитически сложные времена, – может также устранить пробелы в представительстве и координации, перечисленные в пунктах 64 и 72, способствуя более эффективным коллективным действиям по управлению ИИ в общих интересах всех стран.

## Обмен стандартами в области ИИ

### Рекомендация 3: обмен стандартами в области ИИ

Мы рекомендуем создать центр обмена стандартами ИИ, который объединял бы представителей национальных и международных организаций, занимающихся разработкой стандартов, технологических компаний, гражданского общества и представителей международной научной группы. Круг задач центра будет включать в себя:

- a. разработку и ведение реестра определений и применимых стандартов для анализа и оценки систем ИИ;
- b. обсуждение и оценку стандартов и процедур их разработки; и
- c. выявление пробелов, требующих разработки новых стандартов.

**111** Когда изучение систем ИИ только начиналось, существовало мало стандартов, помогающих ориентироваться в этом новом ландшафте или анализировать его. Тест Тьюринга – может ли машина продемонстрировать поведение, эквивалентное человеческому (или

неотличимое от него), – захватил воображение широкой общественности, но он имеет скорее культурное, нежели научное значение. Весьма показательным, что некоторые из величайших достижений вычислительных систем измерялись их успехами в играх, например, когда компьютер смог обыграть человека в шахматы, го, покер или «Джеопарди». Такие показатели были легко понятны неспециалистам, но не были ни строгими, ни особенно научными.

**112** В последнее время наблюдается увеличение числа стандартов. На рис. 13 показано растущее число соответствующих стандартов, принятых МСЭ, Международной организацией по стандартизации (ИСО), Международной электротехнической комиссией (МЭК) и Институтом инженеров по электротехнике и электронике (ИИЭЭ)<sup>32</sup>.

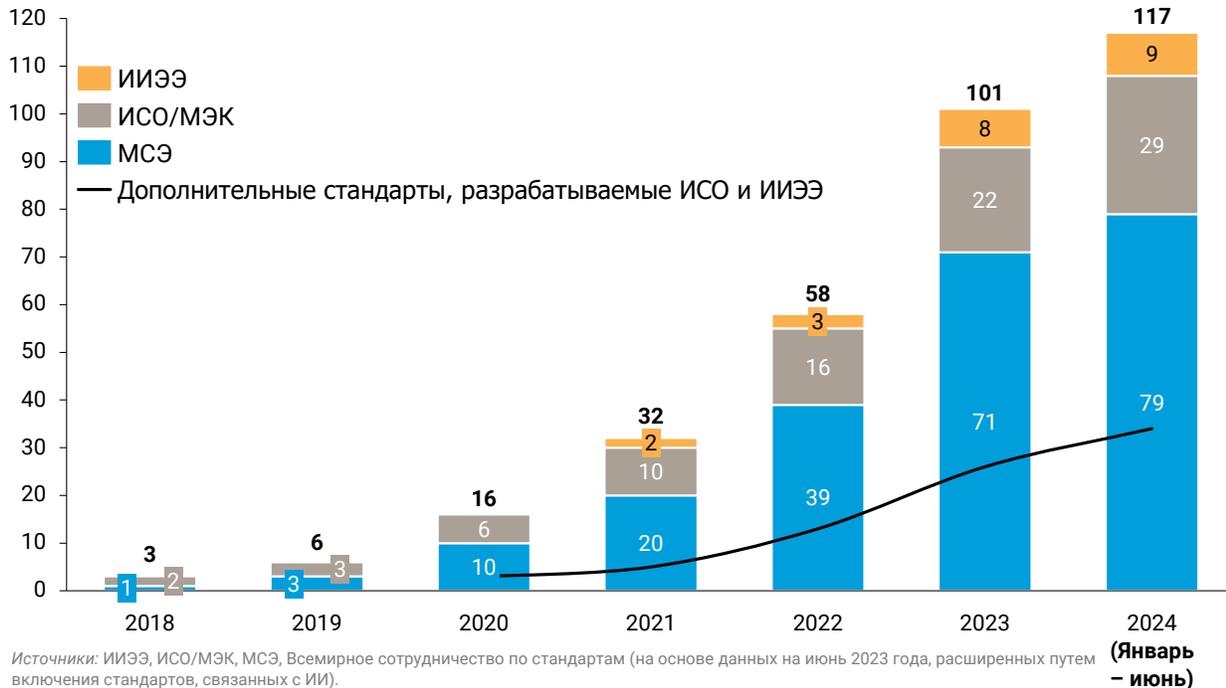
**113** В данной сфере выделяются две тенденции. Во-первых, эти стандарты были разработаны в основном для решения конкретных вопросов. Общая терминология не проработана, и многие термины, которые обычно используются в отношении ИИ – справедливость, безопасность, прозрачность – не имеют согласованных определений и не поддаются измерению (несмотря на недавнюю работу ОЭСР и Национального института стандартов и технологий, принявших новый подход для динамических систем, таких как ИИ).

**114** Во-вторых, существует разрыв между теми стандартами, которые были приняты в узкотехнических целях или для внутренней проверки, и теми, которые призваны учитывать более широкие этические принципы. Ученые-компьютерщики и социологи часто по-разному интерпретируют одну и ту же концепцию, и объединенная парадигма социотехнических стандартов является многообещающей, но пока еще не стала реальностью (см. вставку 10).

31 Хотя многочисленные саммиты по ИИ помогли подгруппе из 20–30 стран объединить усилия по вопросам безопасности ИИ, участие в них было непоследовательным: Бразилия, Китай и Ирландия одобрили Декларацию Блетчли в ноябре 2023 года, но не одобрили Сеульское заявление министров шесть месяцев спустя (см. рис. 12). И наоборот, Мексика и Новая Зеландия поддержали Сеульское заявление министров, но не поддержали Декларацию Блетчли.

32 На национальном и многонациональном уровнях также появляется множество новых стандартов, таких как Добровольные обязательства Белого дома в области ИИ (Соединенные Штаты) и кодексы практики Европейского союза для Закона об ИИ.

## Рисунок 13: Количество стандартов, касающихся ИИ



### Вставка 10: Стандарты, касающиеся безопасности ИИ

Комплексный подход к обеспечению безопасности ИИ предполагает понимание возможностей передовых моделей ИИ, принятие стандартов безопасного проектирования и внедрения, а также оценку как самих систем, так и их более широкого воздействия.

В прошлом стандарты ИИ были ориентированы в основном на технические спецификации, в которых подробно описывалось, как должны создаваться и эксплуатироваться системы. Однако, поскольку технологии ИИ все больше влияют на общество, необходимо перейти к социотехнической парадигме. Этот переход признает, что системы ИИ не существуют в вакууме; они взаимодействуют с пользователями-людьми и влияют на общественные структуры. Современные стандарты ИИ могут включать в себя этические, культурные и общественные соображения наряду с техническими требованиями. Применительно к безопасности это включает обеспечение надежности и интерпретируемости, а также оценку и смягчение рисков для индивидуальных и коллективных прав<sup>а</sup>, национальной и международной безопасности и общественной безопасности в различных контекстах.

Одна из главных целей недавно созданных национальных институтов по безопасности ИИ – обеспечить последовательные и эффективные подходы к вопросам безопасности ИИ. Согласование таких подходов позволит системам ИИ соответствовать высоким стандартам безопасности на международном уровне, обеспечивая трансграничные инновации и торговлю при соблюдении строгих протоколов безопасности.

Поскольку понятие «безопасность» зависит от контекста, привлечение представителей различных заинтересованных сторон и культур к созданию таких стандартов повышает их актуальность и эффективность, а также способствует единообразному пониманию определений и концепций. Благодаря учету различных точек зрения протоколы могут более полно отображать широкий спектр потенциальных рисков и выгод, связанных с технологиями ИИ.

<sup>а</sup> См. документ A/HRC/53/42 («Права человека и процессы установления технических стандартов для новых и появляющихся цифровых технологий: доклад Управления Верховного комиссара Организации Объединенных Наций по правам человека») и резолюцию 53/29 Совета по правам человека («Новые и формирующиеся цифровые технологии и права человека»).

- 115** В результате мы имеем формирующийся набор стандартов, которые не основаны на общем понимании смыслов или оторваны от ценностей, которые они призваны поддерживать. Что особенно важно, существует мало согласованных стандартов, касающихся энергопотребления и ИИ. Отсутствие интеграции соображений прав человека в процессы разработки стандартов – еще один пробел, который необходимо устранить<sup>33</sup>.
- 116** Это влечет за собой реальные издержки. Помимо озабоченности государств-членов и различных лиц, многие наши консультации выявили обеспокоенность бизнеса (включая малые и средние предприятия в развивающихся странах) тем, что фрагментарное управление и непоследовательные стандарты повышают стоимость ведения бизнеса в условиях продолжающейся глобализации.
- 117** В настоящем отчете не предлагается, чтобы Организация Объединенных Наций вносила свой вклад в увеличение количества стандартов. Вместо этого, опираясь на опыт международной научной группы (предложенной в рекомендации 1) и привлекая членов различных организаций, которые внесли свой вклад в разработку стандартов, а также представителей технологических компаний и гражданского общества, система Организации Объединенных Наций могла бы стать центром обмена информацией по стандартам ИИ, которые будут применяться во всем мире<sup>34</sup>.
- 118** Дополнительная ценность Организации будет заключаться в содействии обмену мнениями между широким кругом организаций, занимающихся разработкой стандартов, для обеспечения максимальной глобальной совместимости технических стандартов при одновременном включении новых данных о разработке социотехнических стандартов в обсуждение стандартов ИИ.
- 119** Сбор и распространение информации о стандартах ИИ с опорой на существующие инициативы, такие как Центр стандартов ИИ<sup>35</sup>, и во взаимодействии с ними позволит участникам из разных организаций, занимающихся разработкой стандартов, прийти к общему языку в ключевых областях.
- 120** При поддержке предлагаемого подразделения по вопросам ИИ обмен стандартами также выиграет от тесных связей с международной научной группой по техническим вопросам и практическим диалогом по моральным, этическим, нормативным, правовым и политическим вопросам.
- 121** При наличии соответствующего согласия МСЭ, ИСО/МЭК и ИИЭЭ могли бы совместно возглавить проведение первоначального саммита по стандартам ИИ с ежегодным принятием последующих мер для сохранения актуальности и динамики. Для того, чтобы заложить основы социотехнического подхода, включающего экономические, этические и правозащитные соображения, следует также привлечь ОЭСР, Всемирную организацию интеллектуальной собственности (ВОИС), Всемирную торговую организацию, Управление Верховного комиссара Организации Объединенных Наций по правам человека (УВКПЧ), МОТ, ЮНЕСКО и другие соответствующие структуры Организации Объединенных Наций<sup>36</sup>.

33 См. документ A/HRC/53/42 (Права человека и процессы установления технических стандартов для новых и появляющихся цифровых технологий: доклад Управления Верховного комиссара Организации Объединенных Наций по правам человека) и резолюцию 53/29 Совета по правам человека (Новые и формирующиеся цифровые технологии и права человека).

34 Даже это может показаться сложной задачей, однако прогресс на пути к заключению глобального соглашения о минимальном налоге демонстрирует возможность коллективных действий даже в экономически и политически сложных областях.

35 См. <https://aistandardshub.org>.

36 Сюда могли бы входить соответствующие отраслевые, национальные и региональные организации, занимающиеся стандартами.

**122** Кроме того, обмен данными о стандартах следует использовать в качестве основы для усилий по наращиванию потенциала, предусмотренных рекомендацией 4, обеспечивая поддержку стандартами практики на местах. В связи с этим можно было бы обмениваться информацией о разработанных на национальном или региональном уровнях инструментах, позволяющих проводить самостоятельную оценку соблюдения стандартов.

**123** В настоящее время в докладе не предлагается, чтобы Организация Объединенных Наций делала нечто большее, нежели просто выступала в качестве форума для обсуждения и согласования стандартов. Если со временем стандарты безопасности будут формализованы, они смогут послужить основой для мониторинга и проверки со стороны профильной структуры, которая может быть в итоге создана.

## С. Общие выгоды

**124** Повестка дня на период до 2030 года с ее 17 ЦУР может придать ИИ уникальную целенаправленность, направив вектор инвестиций в сторону от расточительного и вредного использования и на решение глобальных проблем развития. В противном случае инвестиции будут гнаться за прибылью даже ценой негативных внешних последствий для других сторон. Еще один важный вклад, который может внести Организация Объединенных Наций, – это увязка конструктивного применения ИИ с обеспечением справедливого распределения его возможностей (вставка 11).

**125** Как мы утверждали в нашем промежуточном докладе, это в значительной степени зависит от наличия доступа к кадровым ресурсам, вычислительным мощностям и данным, причем таким образом, чтобы способствовать процветанию культурного и языкового разнообразия. Управление само по себе может стать ключевым фактором, способствующим согласованию стимулов, формированию доверия и устойчивой практики, а также развитию взаимодействия между странами и предметными областями. Без комплексного и инклюзивного подхода к управлению ИИ потенциал ИИ для позитивного вклада в достижение ЦУР может быть упущен, а его внедрение может непреднамеренно усилить существующие проявления неравенства и предубеждений.

**126** В ходе обширных консультаций, проведенных Консультативным органом по таким темам, как образование, здравоохранение, данные, гендерные вопросы, дети, мир и безопасность, творческие отрасли и занятость, стало очевидно, что ИИ обладает значительным потенциалом для существенного ускорения прогресса в достижении ЦУР благодаря своим возможностям стимулировать инновации и реализацию мероприятий в различных важнейших областях.

**127** Вместе с тем ИИ не является панацеей, позволяющей решить все проблемы развития; он представляет собой один из компонентов более широкого комплекса решений и может даже усугубить некоторые из этих проблем, включая, например, проблему изменения климата. Для того, чтобы реально раскрыть потенциал ИИ в плане решения социальных проблем, крайне важно наладить сотрудничество между правительствами, научными кругами, отраслью и гражданским обществом.

## Вставка 11: ИИ и ЦУР

Потенциал ИИ в развитии науки (вставка 1) и создании экономических возможностей (вставка 2) лежит в основе надежды на то, что ИИ может ускорить прогресс в достижении ЦУР. В обзоре соответствующих данных, проведенном в 2023 году, утверждается, что ИИ может способствовать выполнению 134 задач (79%) по всем ЦУР, как правило, благодаря технологическому совершенствованию, которое может позволить преодолеть некоторые из существующих ограничений<sup>а</sup>.

Общее представление о нынешних мнениях экспертов дают результаты исследования возможностей, заказанного для нашей работы, в ходе которого было опрошено более 120 экспертов из 38 стран об их ожиданиях относительно положительного влияния ИИ на научные прорывы, экономическую деятельность и ЦУР. В ходе опроса задавались вопросы только о возможных положительных последствиях ИИ.

В целом у экспертов были различные ожидания относительно того, насколько скоро ИИ может привести к ощутимым положительным последствиям (см. также рис. 14):

- Наиболее оптимистично они оценивают **ускорение темпов научных открытий**: 7 из 10 опрошенных заявили, что ИИ, скорее всего, окажет значительное положительное влияние в ближайшие три года или раньше в странах с высоким/выше среднего уровнем дохода, а 28 процентов прогнозируют то же самое для стран с уровнем дохода ниже среднего/низким уровнем дохода.
- Примерно 5 из 10 ожидали, что в ближайшие три года или раньше в странах с высоким/выше среднего уровнем дохода будет отмечаться значительное положительное влияние на **рост экономической активности**, а 32 процента ожидали того же в странах с уровнем дохода ниже среднего/низким уровнем дохода.
- В общей сложности 46 процентов экспертов предполагали, что в странах с высоким/выше среднего уровнем дохода значительное положительное влияние на **прогресс в достижении ЦУР** может быть оказано в ближайшие три года или раньше. Однако только 21 процент из них ожидал этого в странах с доходами ниже среднего/низкими доходами, а 4 из 10 экспертов полагали, что такое ощутимое положительное воздействие на достижение ЦУР в этих странах может проявиться не ранее чем через 10 лет.

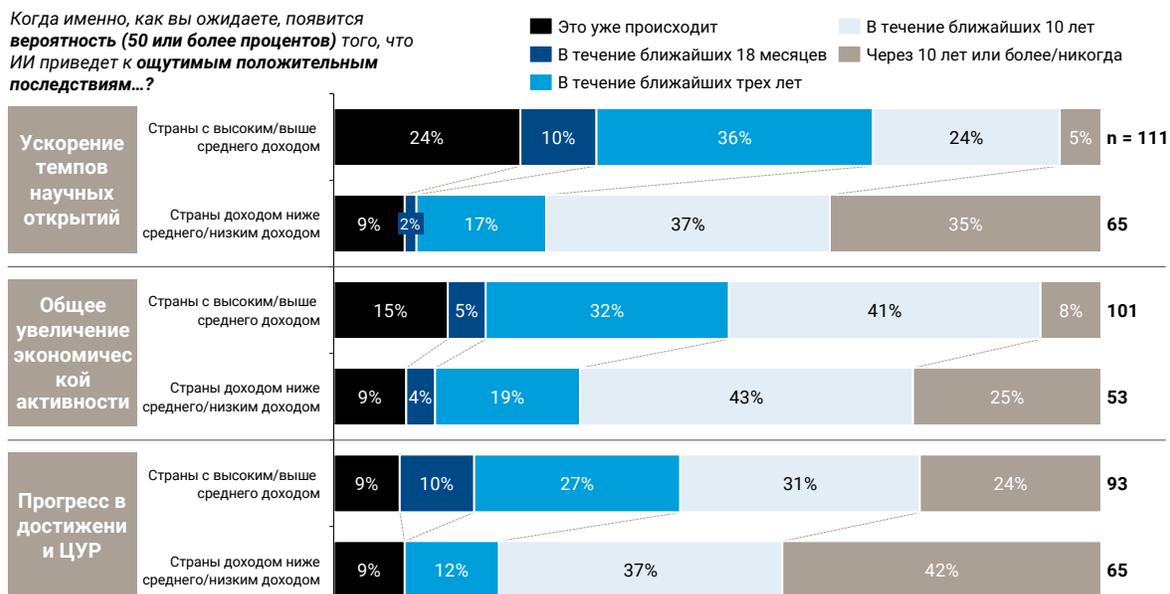
Эксперты ожидали, что в ближайшие три года в странах с высоким уровнем дохода ИИ окажет более значительное положительное воздействие во всех областях, включая ускорение темпов научных открытий, повышение экономической активности<sup>б</sup> и 14 областей ЦУР, о которых шла речь (см. рис. 15). Наиболее оптимистично эксперты оценивали положительное влияние ИИ на здравоохранение и образование (ЦУР 3 и 4), где 20–25 процентов экспертов ожидали значительного или преобразующего положительного влияния ИИ в ближайшие три года в странах с высоким/выше среднего уровнем дохода. Наименее оптимистично они оценили положительное влияние ИИ на гендерное равенство и неравенство (ЦУР 5 и 10), причем две трети из них ожидают, что ИИ не окажет положительного влияния на сокращение неравенства внутри стран или между странами в странах как с высоким, так и с низким уровнем дохода.

a См. Ricardo Vinuesa and others, "The role of artificial intelligence in achieving the Sustainable Development Goals". Nature Communication, vol. 11, No. 233 (January 2020). В этом исследовании также указывалось, что ИИ может оказать негативное влияние на выполнение 59 задач (35 процентов, также по всему спектру ЦУР).

b Доля экспертов, ожидающих «значительного положительного влияния» на рост экономической активности и ускорение научных открытий в течение трех лет, на первом графике выше, чем на втором. Это может быть связано с тем, что в ответах на вопрос, представленных на первой диаграмме, присутствует уточнение «когда именно, как вы ожидаете, появится вероятность (50 или более процентов) того, что ИИ приведет к ощутимым положительным последствиям» (выделено автором), которое отсутствует на второй диаграмме.

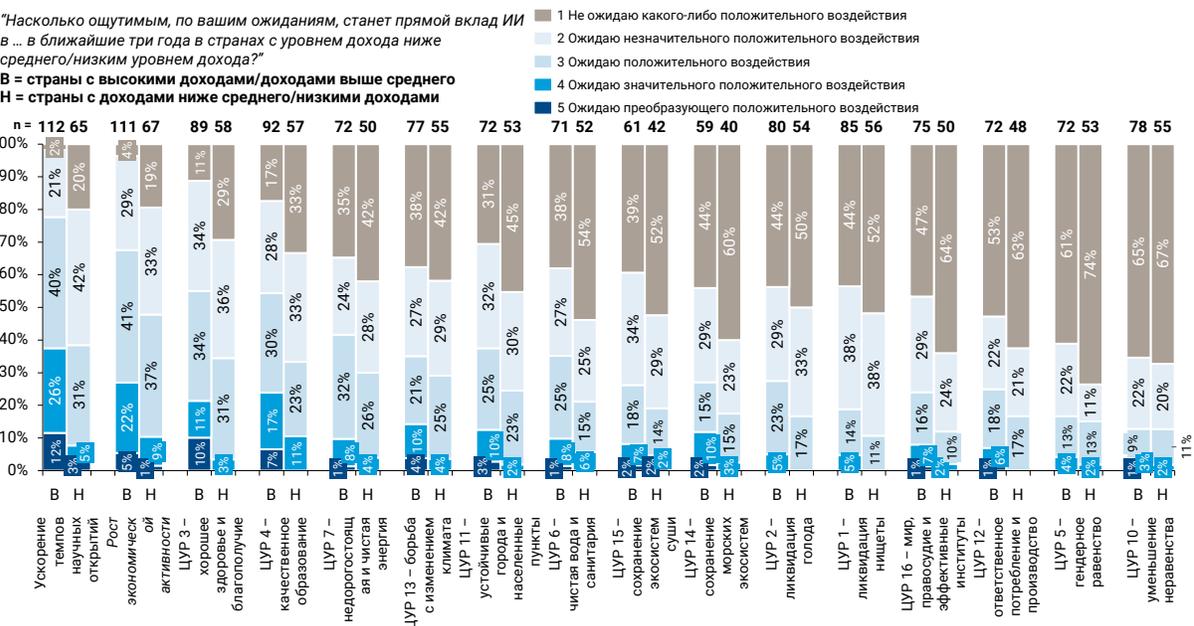
## Вставка 11: ИИ и ЦУР (продолжение)

### Рисунок 14: Ожидания экспертов в отношении сроков, которые потребуются для того, чтобы проявились ощутимые положительные последствия ИИ, в разбивке по областям



Примечание: исключая ответы "Не знаю" / "Затрудняюсь ответить" и незаполненные ответы. Конкретные величины n указаны в отношении каждого вопроса. Источник: Оценка рисков, связанных с ИИ, по состоянию на 21 августа 2024 года (n=121, ответы на вопросы, касающихся стран с уровнем дохода ниже среднего/низким уровнем дохода, предоставили 78 респондентов).

### Рисунок 15: Ожидания экспертов в отношении ощутимого положительного воздействия ИИ в ближайшие три года в разбивке по областям и ЦУР

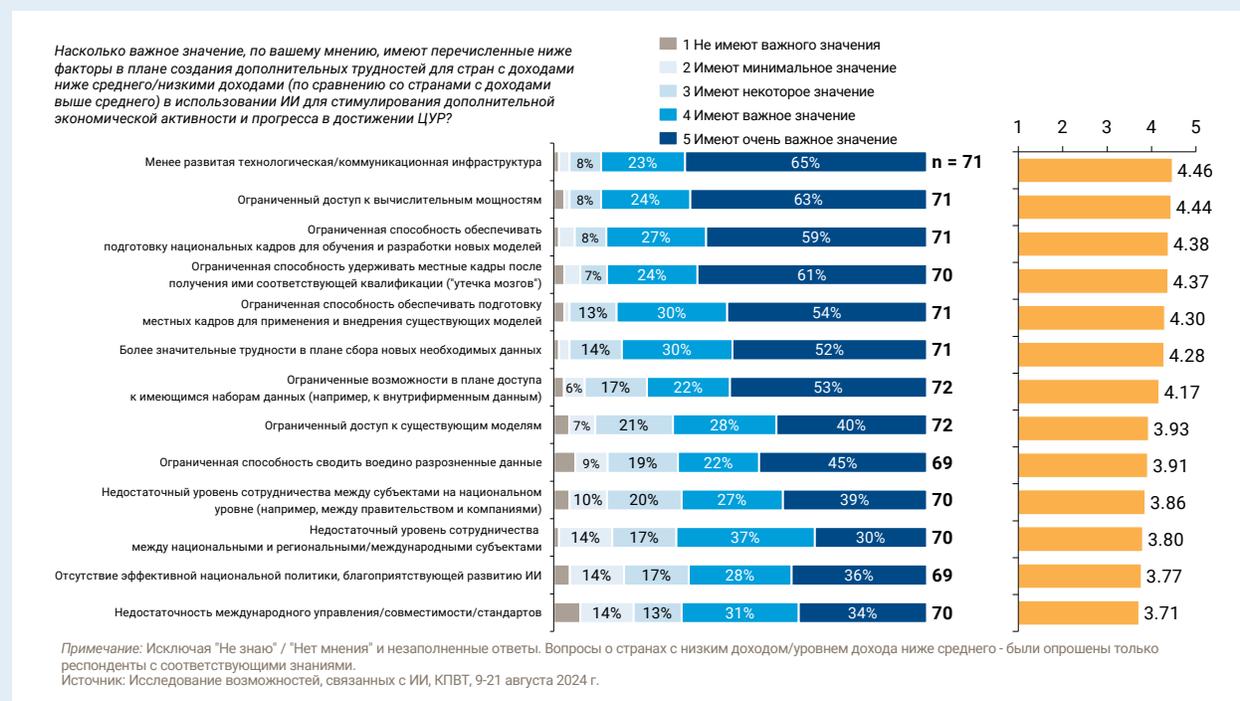


Примечание: исключая ответы "Не знаю" / "Затрудняюсь ответить" и незаполненные ответы. Конкретные величины n указаны в отношении каждого вопроса. Вопросы не затрагивали ЦУР 8, 9 и 17. Источник: Оценка рисков, связанных с ИИ, по состоянию на 21 августа 2024 года (n=121, ответы на вопросы, касающихся стран с уровнем дохода ниже среднего/низким уровнем дохода, предоставили 78 респондентов).

## Вставка 11: ИИ и ЦУР (продолжение)

Можно ожидать, что ИИ окажет более раннее и значительное воздействие на страны с более высоким уровнем дохода, что отчасти объясняется препятствиями, сдерживающими развитие стран с уровнем дохода ниже среднего и с низким уровнем дохода (см. рис. 16). Более половины респондентов назвали отсутствующие предпосылки – от неразвитой инфраструктуры до отсутствия соответствующей внутренней политики и международного управления – важными факторами, создающими дополнительные трудности для стран с низким уровнем дохода в использовании ИИ в интересах активизации экономической деятельности и достижения ЦУР.

## Рисунок 16: Оценка экспертами препятствий, мешающих использованию ИИ для стимулирования дополнительной экономической активности и прогресса в достижении ЦУР в странах с уровнем дохода ниже среднего/низким уровнем дохода



Эти результаты указывают на то, что возможный вклад ИИ в достижение ЦУР весьма сомнителен и что он по-прежнему сильно зависит от недостающих факторов. Это особенно актуально для менее развитых стран, которым уже не хватает многого из того, что есть у более развитых стран: от инфраструктуры до политики. Без сотрудничества, направленного на наращивание потенциала и облегчение доступа к ключевым благоприятствующим факторам, существующие разрывы в области ИИ могут еще больше расшириться и укорениться, ограничивая способность ИИ внести конструктивный вклад в научный прогресс, развитие экономики и прогресс в достижении ЦУР до 2030 года.

- 128** Эффективность решений в сфере ИИ зависит от качества и доступности данных, а качество и репрезентативность наборов данных, связанных с ЦУР, вызывают серьезную озабоченность, поскольку могут не отражать реальное положение дел в определенных группах населения. Кроме того, решения в сфере ИИ, разработанные экспертами по ИИ без всесторонней осведомленности о пересекающихся областях применения, часто работают на экспериментальной основе и не являются достаточно надежными и эффективными в реальных условиях развития. Именно поэтому решения в области ИИ должны разрабатываться совместно и внедряться с глубоким пониманием их социального, экономического и культурного контекста. Они должны вписываться в более широкие местные и национальные стратегии по цифровой трансформации и устранению цифрового неравенства.
- 129** Так, например, возможности ИИ в странах с низким уровнем дохода и уровнем дохода ниже среднего не могут быть реализованы без обеспечения надежного электроснабжения и подключения к Интернету для функционирования центров обработки данных, поддержания стабильной работы компьютеров, доступа к глобальным массивам данных, участия в международном научном сотрудничестве и использования облачных инструментов ИИ. Поэтому мы присоединяемся к призывам инвестировать в базовую цифровую инфраструктуру, которая является необходимым условием для участия развивающихся стран в развитии ИИ и получения от этого выгоды.
- 130** Нарращивание потенциала ИИ имеет ключевое значение для того, чтобы люди во всем мире, независимо от стадии развития их региона, могли пользоваться достижениями ИИ. Стратегическое наращивание потенциала, подкрепленное адекватным финансированием, также необходимо для того, чтобы сделать технологии ИИ эффективными, устойчивыми и отвечающими общественным интересам, что, в свою очередь, имеет важнейшее значение

для глобальных усилий в области развития. Ниже мы рассмотрим три важнейших фактора, способствующих наращиванию национального потенциала в области ИИ: наличие технических специалистов, доступ к вычислениям и наличие качественных данных. Затем мы порекомендуем конкретные действия.

## Кадровые ресурсы

- 131** Способность обществ всего мира участвовать в использовании положительных результатов внедрения ИИ зависит прежде всего от людей. Важно признать, что не каждому обществу нужны кадры ученых-компьютерщиков для создания собственных моделей. Однако независимо от того, покупаются, заимствуются или создаются технологии, для того чтобы понимать возможности и ограничения ИИ и надлежащим образом использовать возможности ИИ, необходимы человеческие ресурсы.
- 132** Такой потенциал – в первую очередь в государственном секторе, а также в научных кругах, бизнесе и гражданском обществе – повысит эффективность стратегий в области ИИ и качество их реализации в различных секторах. Развитие человеческого потенциала, связанного с ИИ, также будет иметь жизненно важное значение для сохранения мирового культурного и языкового разнообразия и создания высококачественных массивов данных для будущих разработок ИИ. В сущности, это наращивание потенциала для ИИ в интересах общества.
- 133** Развитие людских ресурсов в различных регионах с преобладанием молодежи в составе населения, таких как Африка (в первой половине этого века треть мировой рабочей силы будет состоять из африканцев), также будет иметь жизненно важное значение для будущего глобального кадрового резерва. Укрепление потенциала женщин в сфере технологий должно быть направлено, с одной стороны, на устранение существующего гендерного разрыва, а с другой – на предотвращение гендерного

разрыва в области ИИ. Сектору ИИ также необходимо больше женщин на руководящих должностях, с тем чтобы отразить гендерные аспекты в сфере управления ИИ. Первый шаг в этом процессе – расширение возможностей девочек в плане развития талантов в области ИИ.

## Вычислительные ресурсы

**134** Несмотря на предпринимаемые усилия по разработке менее требовательных к вычислительным ресурсам подходов к ИИ, в контексте обучения качественных моделей ИИ по-прежнему остро стоит вопрос доступа к недорогим вычислительным машинам<sup>37</sup>. Это один из самых серьезных барьеров для входа в сферу ИИ для компаний глобального Юга, а также для многих стартапов и малых и средних предприятий глобального Севера. Из 100 крупнейших высокопроизводительных вычислительных кластеров в мире, способных обучать большие модели ИИ, ни один не расположен в развивающейся стране<sup>38</sup>. В топ-300 входит только одна африканская страна. На две страны приходится половина всех сверхмощных центров обработки данных в мире<sup>39</sup>.

**135** Большинство разработчиков получают доступ к вычислительной инфраструктуре через облачные сервисы; многие из них предпочитают сотрудничать с крупными облачными компаниями, с тем чтобы обеспечить себе надежный доступ к вычислительным мощностям. Возможно, со временем проблемы с цепочкой поставок будут решены, а конкуренция приведет к появлению более разнообразных источников аппаратного обеспечения, включая высокопроизводительные чипы для обучения моделей и чипы ускорителей ИИ для установки на мобильные устройства. Однако

в обозримом будущем это ограничение будет оставаться серьезным препятствием на пути к созданию более глобальной и инклюзивной инновационной экосистемы ИИ.

**136** Как это ни странно, вычислительные мощности могут простаивать или быстро устаревать. Потенциальная выгода заключается в стопроцентном использовании таких мощностей в течение всех циклов амортизации. Однако при этом необходимо преодолеть трудности, связанные с совместимостью различных аппаратных конфигураций и планированием сложных задач, сохраняя приоритет критически важного использования (например, для метеорологических прогнозов).

**137** Кроме того, в отсутствие кадровых ресурсов и данных вычислительные ресурсы сами по себе не имеют ценности. В предлагаемом глобальном фонде для ИИ мы рассматриваем вопрос о том, как решить все три проблемы путем сочетания финансовой и нефинансовой поддержки.

## Данные

**138** Хотя в рамках многих дискуссий об экономике ИИ основное внимание уделяется «войне за таланты» и конкуренции за аппаратные средства, такие как графические процессоры (ГП), данные имеют не менее важное значение. Облегчение доступа к качественным обучающим данным в масштабе для обучения моделей ИИ силами стартапов и малых и средних предприятий, а также механизмы компенсации держателям данных и создателям обучающих данных способами, обеспечивающими соблюдение соответствующих прав, могут стать важнейшим фактором, способствующим процветанию экономики ИИ. Объединение данных в интересах общества для достижения конкретных ЦУР является одним из ключевых аспектов (см. вставку 12), но этого недостаточно.

37 Консультативный орган знает о недавнем случае, когда компания, расположенная в одной из стран глобального Юга, потратила 70 млн долл. США на трехмесячное обучение большой языковой модели. Собственные графические процессоры (ГП) вместо их аренды у поставщиков облачных услуг обошлись бы во много раз дешевле.

38 См. <https://top500.org/statistics/sublist>; прокси-индикатор, поскольку большинство высокопроизводительных вычислительных кластеров не имеют ГП и малопригодны для продвинутого ИИ.

39 UNCTAD, *Digital Economy Report 2021* (Geneva, 2021).

## Вставка 12: Объединение данных, относящихся к достижению ЦУР, в интересах общества

Объединение данных и создание совместных механизмов разработки ИИ, в рамках которых общие модели проходят перекрестное обучение на объединенных данных, могут сыграть ключевую роль в продвижении общественных интересов там, где в противном случае данные отсутствуют или слишком скудны для использования преимуществ ИИ. Многофункциональные и многопрофильные пулы данных могут позволить создавать междисциплинарные массивы данных, охватывающие различные области ЦУР и получаемые из различных источников.

В качестве примера можно рассмотреть сложный вопрос оценки воздействия изменения климата на здоровье. Для эффективного решения этой задачи необходим междисциплинарный подход, объединяющий эпидемиологические данные о распространенности заболеваний с метеорологическими данными, отслеживающими колебания климата. Объединив эти разные типы данных из различных стран мира при сохранении должного уровня конфиденциальности, исследователи смогут использовать ИИ для выявления закономерностей и корреляций, которые не очевидны из индивидуальных наборов данных.

Учет данных из всех стран обеспечивает всеобъемлющий охват, отражающий глобальный характер изменения климата, а также различные экологические последствия и состояние здоровья населения в разных регионах. Междисциплинарное происхождение данных повышает точность работы моделей, направленных на прогнозирование будущих кризисов в области общественного здравоохранения или стихийных бедствий, вызванных изменением климата.

- 139** В контексте ИИ принято говорить о «ненадлежащем» использовании данных (например, нарушении неприкосновенности частной жизни) или об «упущенных возможностях» использования данных (неспособности использовать имеющиеся наборы данных), но смежной проблемой являются «отсутствие» данных, которое характерно для значительных частей земного шара, испытывающих нехватку данных. Одним из примеров подобного рода является здравоохранение, где примерно половина самых крупных наборов данных может быть отнесена к десятку организаций, из которых одна находится в Европе, одна – в Азии, а остальные – в Северной Америке<sup>40</sup>.
- 140** Другой пример – сельское хозяйство, где для создания эффективных моделей ИИ требуются данные по сложной взаимосвязи факторов (таких как климат, почва и методы управления посевами). Сельское хозяйство также часто страдает от нехватки данных и инфраструктуры сбора данных. Необходимы целенаправленные усилия по формированию наборов сельскохозяйственных данных, особенно в контексте устойчивости продовольственных систем к изменению климата.
- 141** Аналогично проблеме неформального капитала, те, чьи данные – от записей о рождении до финансовых операций – не фиксируются, могут оказаться не в состоянии воспользоваться преимуществами экономики ИИ, получить государственные пособия или доступ к кредитам. Использование смоделированных данных может лишь частично компенсировать потребность в новых наборах данных.
- 142** В отзывах на подготовленный нами промежуточный доклад отмечалось, что в нем недостаточно четко прописано, каким образом

40 См. <https://2022.internethhealthreport.org/facts>.

существующая практика поиска, использования и неразглашения данных для обучения ИИ, охватывающая сразу несколько юрисдикций, угрожает правам и приводит к экономической концентрации. Было рекомендовано рассмотреть вопрос о том, как международное управление ИИ может обеспечить и стимулировать более разнообразное участие в использовании данных для ИИ.

### **Формирование на международной основе основополагающего общественного потенциала ИИ в общих интересах**

**143** Если говорить о трех вышеперечисленных факторах, то страны с развитой экономикой имеют возможность и обязаны содействовать наращиванию потенциала ИИ на основе международного сотрудничества. В свою очередь, они получают выгоду от более широкой цифровой экономики, а также от притока качественных кадровых ресурсов и данных. Важно отметить, что от внедрения эффективных механизмов управления ИИ в рамках такого сотрудничества выиграют все.

**144** Сотрудничество должно быть направлено на повышение квалификации специалистов в области ИИ, рост грамотности населения в области ИИ, укрепление потенциала управления ИИ, расширение доступа к инфраструктуре ИИ, продвижение платформ данных и знаний, отвечающих различным культурным и региональным потребностям, а также на более широкое использование приложений и возможностей сервисов ИИ. Только такой комплексный подход может обеспечить равный доступ к преимуществам ИИ, с тем чтобы ни одна страна не осталась за бортом.

**145** Многие заинтересованные стороны, с которыми мы консультировались, подчеркнули, что необходимо разработать подробные стратегии по объединению глобальных ресурсов для наращивания потенциала, стимулирования коллективных действий по справедливому распределению возможностей и ликвидации цифрового разрыва.

## **Сеть для наращивания потенциала**

### **Рекомендация 4: сеть для наращивания потенциала**

Мы рекомендуем создать сеть для наращивания потенциала в области ИИ, которая объединила бы ряд сотрудничающих центров наращивания потенциала, связанных с Организацией Объединенных Наций, и предоставила бы ключевым участникам экспертные знания, вычислительные мощности и данные для обучения ИИ. Цель такой сети заключалась бы в следующем:

- a. стимулирование и координация региональных и глобальных усилий по наращиванию потенциала в области ИИ путем поддержки налаживания взаимосвязей между ними;
- b. укрепление возможностей государственных служащих в области управления ИИ для содействия развитию при одновременном обеспечении уважения, защиты и реализации всех прав человека;
- c. предоставление услуг экспертов, вычислительных ресурсов и данных для обучения ИИ в нескольких центрах исследователям и социальным предпринимателям, стремящимся применять ИИ для решения местных задач в интересах общества, в том числе при помощи следующих инструментов:
  - i. протоколы, позволяющие междисциплинарным исследовательским группам и предпринимателям в условиях нехватки компьютеров получать доступ к вычислительным ресурсам для обучения/настройки и применения своих моделей в соответствии с местными условиями;
  - ii. учебно-экспериментальные центры для тестирования потенциальных решений в сфере ИИ и практического обучения;

- iii. пакет онлайн-образовательных возможностей по ИИ, предназначенный для студентов университетов, молодых исследователей, социальных предпринимателей и должностных лиц государственного сектора; и
- iv. программа стипендий для перспективных специалистов, которые смогут провести время в академических институтах или технологических компаниях.

**146** От целей в области развития, сформулированных в Декларации тысячелетия, до ЦУР Организация Объединенных Наций уже в течение долгого времени вносит свой вклад в наращивание потенциала как отдельных лиц, так и организаций<sup>41</sup>. Благодаря деятельности ЮНЕСКО, ВОИС и других структур Организация Объединенных Наций помогает поддерживать богатое разнообразие культур и традиций генерирования знаний по всему миру.

**147** В то же время развитие потенциала для ИИ потребует нового подхода, в частности междисциплинарной подготовки для формирования нового поколения междисциплинарных экспертов в таких областях, как общественное здравоохранение и ИИ или продовольственные и энергетические системы и ИИ.

**148** Потенциал также должен быть увязан с результатами посредством практического обучения в учебно-экспериментальных центрах<sup>42</sup> и реализации совместных проектов, объединяющих данные и вычислительные мощности для решения общих проблем. В эту инфраструктуру совместного обучения должны быть встроены оценка рисков, тестирование на безопасность и другие методологии управления.

**149** Учитывая срочность и масштаб проблемы, мы предлагаем использовать стратегический подход, который предусматривает объединение и посредничество в доступе к вычислительным ресурсам через сеть высокопроизводительных вычислительных узлов, стимулирование разработки критически важных наборов данных в областях, связанных с ЦУР, поощрение обмена моделями ИИ, внедрение передового опыта управления ИИ и наращивание многопрофильного кадрового потенциала для ИИ в интересах общества, что, в свою очередь, обеспечило бы междисциплинарную интеграцию экспертных знаний в области прав человека.

**150** Иными словами, вместо того, чтобы по одному гнаться за важнейшими факторами, способствующими достижению целей, через разрозненные проекты, мы предлагаем **комплексную стратегию, реализуемую через сеть сотрудничающих центров**. Появляющиеся инициативы по развитию потенциала и ИИ для ЦУР, такие как инициатива Международной сети вычислений и ИИ (МСВИИ), запущенная Швейцарией, могут помочь создать первоначальную критическую массу для этой стратегии.

**151** В идеале в каждом регионе мира должно быть не менее одного-двух вычислительных узлов. Два экспертных центра, участвующих в Глобальном партнерстве по искусственному интеллекту, могли бы присоединиться к Организации Объединенных Наций в поддержке сети для наращивания потенциала. Академические учреждения и частный сектор, вносящие свой вклад в наращивание потенциала, могли бы подключаться к сети через ближайший региональный узел или через одну из международных организаций, поддерживающих сеть.

<sup>41</sup> Университет Организации Объединенных Наций уже давно занимается наращиванием потенциала с помощью высшего образования и научных исследований, а Учебный и научно-исследовательский институт Организации Объединенных Наций помогает готовить должностных лиц в областях, имеющих решающее значение для устойчивого развития. Методология оценки готовности ЮНЕСКО является ключевым инструментом для оказания поддержки государствам-членам в выполнении Рекомендации ЮНЕСКО по этике искусственного интеллекта. Среди других примеров - Академия ВОЗ в Лионе, Виртуальный институт ЮНКТАД, Программа стипендий Организации Объединенных Наций по разоружению, реализуемая Управлением по вопросам разоружения, а также программы наращивания потенциала, осуществляемые МСЭ и ПРООН.

<sup>42</sup> Учебно-экспериментальные центры создавались различными национальными учреждениями, включая финансовые и медицинские структуры, такие как Управление по развитию инфокоммуникаций Сингапура.

- 152** Нас особенно обнадеживает перспектива сотрудничества между странами, например, на основе объединенного доступа к вычислительным мощностям и смежной инфраструктуре. Как отмечалось в нашем промежуточном докладе, полезные уроки можно извлечь из опыта Европейской организации ядерных исследований (ЦЕРН). «Распределенная структура», подобная той, которая существует в ЦЕРН, переосмысленная для ИИ и объединенная в сеть с участием различных государств и регионов, могла бы способствовать расширению возможностей для более широкого доступа к инструментам и экспертным знаниям в области ИИ.
- 153** Мы представляем себе сеть для наращивания потенциала как катализатор национальных и региональных возможностей, а не как центр сосредоточения оборудования, талантов и данных. Ускоряя процесс обучения, она могла бы стимулировать национальные центры передового опыта к развитию местных инновационных экосистем ИИ, устраняя основные пробелы в координации и реализации, упомянутые в пунктах 73, 80 и 81. В рамках усилий на национальном уровне можно продолжать использовать такие инструменты диагностики, как Методология оценки готовности к ИИ ЮНЕСКО, предназначенная для того, чтобы помочь оценить начальную зрелость стран, выявить пробелы и определить, как «дорожные карты» по наращиванию потенциала могут быть адаптированы для каждой страны и региона, а международная сеть поможет устранить эти пробелы.
- 154** Предлагаемое подразделение по вопросам ИИ, возможно, лучше всего подходит для того, чтобы сосредоточиться на стратегии, партнерстве и аффилированности и обеспечить связь узлов с сетью, выполняя задачи, связанные с объединением, а не с реорганизацией. Оно также может помочь в организации доступа к вычислительным мощностям в рамках сети. Узел или узлы сети могут выступать в качестве ведущих элементов по конкретным аспектам обучения, размещать у себя учебно-экспериментальные центры или

высокопроизводительные вычислительные кластеры для разработки моделей ИИ. Узлы могут сотрудничать в рамках исследовательских программ по таким темам, как использование данных с сохранением конфиденциальности, новые методы соединения различных типов оборудования или наборов данных для обучения моделей, а также способы использования моделей ИИ в сочетании друг с другом.

- 155** Мы надеемся на то, что такая сеть также будет способствовать развитию альтернативной парадигмы разработки технологий ИИ: организованной по принципу «снизу вверх», межотраслевой, межрегиональной, открытой и совместной. Учитывая растущие энергетические и другие затраты на обучение и внедрение моделей ИИ, а также перспективу неэффективного использования вычислительных ресурсов, имеет смысл объединять вычислительные мощности для доступа к ним на условиях совместного использования времени и использовать такой доступ для подготовки многопрофильных специалистов, а также генерирования данных и создания моделей ИИ для ЦУР.

## Глобальный фонд для ИИ

### Рекомендация 5: глобальный фонд для ИИ

Мы рекомендуем создать глобальный фонд для ИИ, с тем чтобы преодолеть разрыв в области ИИ. Работая под управлением независимой руководящей структуры, фонд будет получать финансовые и неденежные взносы из государственных и частных источников и распределять их, в том числе через сеть для наращивания потенциала, в целях облегчения доступа к средствам ИИ, способствующим расширению возможностей на местах для достижения ЦУР, включая:

- a. совместное использование вычислительных ресурсов для обучения и отладки моделей разработчиками ИИ из стран, не имеющих достаточного местного потенциала или средств для его приобретения;

- b. учебно-экспериментальные центры и инструменты бенчмаркинга и тестирования для внедрения передового опыта разработки безопасных и надежных моделей и управления данными;
- c. решения в области управления, безопасности и операционной совместимости, имеющие глобальное применение;
- d. наборы данных и исследования способов объединения данных и моделей для реализации проектов, связанных с ЦУР; и
- e. хранилище моделей ИИ и специально подобранных наборов данных для ЦУР.

**156** Предлагаемая здесь модель разработки и использования ИИ аналогична первоначальному видению Интернета: распределенная, но сопряженная инфраструктура, способная к взаимодействию и расширению возможностей. Общественным интересам будет лучше отвечать рынок, на котором модели ИИ, а также инфраструктура и данные, на которые они опираются, являются совместимыми, эффективно управляемыми и заслуживающими доверия. Этого нельзя достичь автоматически. Необходимы целенаправленные усилия, подкрепленные достаточными ресурсами.

**157** Мы подходим к этой рекомендации реалистично, отдавая себе отчет в наличии мощных рыночных сил, определяющих порядок доступа к кадровым ресурсам и вычислительным мощностям, а также в геополитической конкуренции, препятствующей сотрудничеству в области науки и техники. К сожалению, многие страны могут оказаться не в состоянии получить доступ к обучению, вычислительным ресурсам, моделям и учебным данным без международной поддержки. Существующие усилия по финансированию также не смогут расширяться без такой поддержки.

**158** Выравнивание игрового поля – это отчасти вопрос справедливости. В наших коллективных интересах также создать такой мир, в котором все вносят свой вклад в общую экосистему и

получают от нее пользу. Это касается не только государств. Обеспечение доступа различных заинтересованных сторон к инфраструктуре разработки и тестирования моделей ИИ также поможет снять озабоченность по поводу концентрации непропорционально большой власти в руках горстки технологических компаний.

## Цели и задачи фонда

**159** Предлагаая создать фонд, мы не ставим целью гарантировать доступ к вычислительным ресурсам и возможностям, которые с трудом приобретают даже самые богатые страны и компании. Ответ не всегда заключается в увеличении вычислительных мощностей. Нам также могут понадобиться другие способы использования существующих высокопроизводительных вычислительных ресурсов, которые созданы для пикового использования и не обязательно предназначены для ИИ. Возможно, существуют более эффективные способы объединения талантов, вычислительных ресурсов и данных.

**160** Таким образом, цель состоит в том, чтобы устранить основные пробелы в координации и реализации, указанные в пунктах 73, 80 и 81, для тех, кто не может получить доступ к необходимым средствам поддержки другими способами, с тем чтобы добиться прогресса на следующих направлениях:

- a. нуждающиеся страны смогут получить доступ к средствам поддержки ИИ, что позволит преодолеть разрыв в области ИИ;
- b. сотрудничество в области наращивания потенциала ИИ ведет к формированию навыков сотрудничества и позволяет смягчить геополитическую конкуренцию;
- c. страны с различными подходами к регулированию имеют стимулы к разработке общих шаблонов для регулирования данных, моделей и приложений в целях решения на уровне общества в целом тех проблем, которые связаны с ЦУР и научными прорывами.

- 161** Потенциал, создаваемый за счет ресурсов глобального фонда, будет ориентирован на ЦУР и совместное глобальное управление ИИ (вставка 13). Например, он может включать в себя «механизм управления» для тестирования безопасности и надежности. Это помогло бы внедрить передовой опыт во всей пользовательской базе, снизив при этом бремя валидации для мелких пользователей.
- 162** Такая ориентация на общественные интересы делает глобальный фонд дополнением к предложению о создании сети для наращивания потенциала ИИ, в которую фонд будет направлять ресурсы. Фонд также обеспечит наличие независимого потенциала для мониторинга последствий. Действуя таким образом, мы добьемся того, чтобы обширные районы мира не остались за бортом, а получили возможность использовать ИИ для достижения ЦУР в различных условиях.
- 163** Все заинтересованы в том, чтобы в цифровом мире, как и в физическом, существовало сотрудничество. Можно провести аналогию с усилиями по борьбе с изменением климата, где затраты на переходный период, смягчение последствий или адаптацию не распределяются равномерно, и международная помощь необходима для оказания содействия странам с ограниченными ресурсами, с тем чтобы они могли присоединиться к глобальным усилиям по решению проблемы планетарного масштаба.
- 164** В данной связи основное внимание уделяется использованию финансирования для обеспечения создания минимального потенциала в странах различных регионов для понимания потенциала ИИ для устойчивого развития, адаптации и создания моделей с учетом местных потребностей, а также присоединения к международным совместным усилиям в области ИИ.

### **Вставка 13: Глобальный фонд для ИИ: примеры возможных инвестиций**

Относительно скромный по своим размерам фонд может помочь создать минимально необходимую общую вычислительную инфраструктуру для обучения малых и средних моделей. Такие модели обладают важным потенциалом для достижения ЦУР, включая, например, обучение фермеров на их местном языке.

Эти инвестиции также позволят создать учебно-экспериментальную инфраструктуру для разработчиков, которые смогут дорабатывать существующие модели с открытыми спецификациями, используя свои собственные контекстные и высококачественные данные. Доступ к вычислительной инфраструктуре и учебно-экспериментальной инфраструктуре может быть предоставлен на условиях временного пользования, а разумная плата за пользование будет способствовать покрытию эксплуатационных расходов.

Третьим направлением использования средств может стать помощь в формировании эталонных наборов данных для отдельных ЦУР, где отсутствуют коммерческие стимулы. Усилия по разработке и тестированию моделей и сбору данных могут быть стратегически объединены в мощный практический подход к расширению возможностей ИИ, связанный с конкретными результатами.

Наконец, фонд мог бы стимулировать исследования и разработки не только в области применения ИИ в контексте развития и достижения ЦУР, но и в области увязки вычислительных ресурсов и моделей, а также анализа новых механизмов управления.

## Управление фондом

**165** Фонд будет заниматься сбором и консолидацией взносов в неденежной форме, в том числе от организаций частного сектора. Координация финансовых и неденежных взносов требует соответствующего уровня независимого надзора и подотчетности. Механизмы управления должны быть инклюзивными, а членами правления должны быть филантропы, а также представители правительств, частного сектора, гражданского общества и учреждений Организации Объединенных Наций. Они должны учитывать вклад ученых и экспертов, направляемый, например, через предлагаемую международную научную группу, а также обеспечивать нейтралитет и доверие для сотрудничества в области данных и разработки моделей.

## Деятельность фонда

**166** Модель работы фонда должна учитывать уроки, извлеченные из опыта работы совместных международных механизмов по проведению исследований и разработок, таких как ЦЕРН и ГАВИ (Альянс по вакцинам)<sup>43</sup>, а также опыт коммерческих платформ для совместного использования инфраструктуры. Она также должна основываться на опыте деятельности таких организаций, как Глобальный фонд (созданный в 2002 году для объединения ресурсов в целях борьбы с ВИЧ, туберкулезом и малярией) и Фонд анализа комплексных рисков (объединяющий данные в поддержку всех заинтересованных сторон в прогнозировании, предотвращении и реагировании на кризисы).

## Глобальная система данных ИИ

### Рекомендация 6: глобальная система данных ИИ

Мы рекомендуем создать глобальную систему данных ИИ, сформировав ее в рамках процесса, инициированного соответствующим учреждением, таким как Комиссия Организации Объединенных Наций по праву международной торговли, и с учетом работы других международных организаций, для следующих целей:

- a. изложение связанных с данными определений и принципов глобального управления данными для обучения ИИ, в том числе на основе существующих передовых практик, и содействие культурному и языковому разнообразию;
- b. установление общих стандартов проверки и использования данных для обучения ИИ в целях обеспечения прозрачной и основанной на правах человека подотчетности в разных юрисдикциях; и
- c. создание формирующих рынок механизмов управления данными и обмена ими для обеспечения процветания местных экосистем ИИ в глобальном масштабе, включая, в частности:
  - i. трасты данных;
  - ii. эффективно управляемые глобальные рынки для обмена анонимными данными в целях обучения моделей ИИ; и
  - iii. типовые соглашения для облегчения международного доступа к данным и глобальной операционной совместимости, потенциально в качестве технико-юридических протоколов к рамочной программе.

43 См. <https://www.theglobalfund.org/en/about-the-global-fund>.

- 167** В ходе наших консультаций мы узнали, что, несмотря на множество предложений по расширению доступа к данным и механизмов обмена данными для создания более разнообразных экосистем ИИ, до сих пор не многие из них были реализованы. Это критический пробел в развитии инклюзивных и динамичных экосистем ИИ.
- 168** Отчасти ответ на этот вопрос кроется в прозрачности культурных, языковых и других характеристик обучающих данных ИИ. Также полезно выявлять недопредставленные или «недостающие» данные. С этим связаны усилия по содействию созданию «общих массивов данных», которые стимулируют сбор учебных данных для нужд большого числа участников. Такие инициативы могут способствовать формированию передовых видов практики, демонстрируя, как дизайн может встраиваться в технико-юридические рамки для обеспечения конфиденциальности, защиты данных, совместимости и справедливого использования данных, а также соблюдения прав человека.
- 169** Рынки данных для ИИ сегодня представляют собой некое подобие «дикого Запада». Одним из принципов работы является идея «бери все, что можешь, и прячь в непрозрачных алгоритмах»; другим - эксклюзивные договорные соглашения о доступе к фирменным данным, соблюдение которых может быть обеспечено в отдельных юрисдикциях. Такие эксклюзивные отношения являются причиной обеспокоенности Управления по конкуренции и рынкам Великобритании по поводу того, что «сектор [продвинутых моделей] развивается таким образом, что это чревато негативными последствиями для рынка»<sup>44</sup>.
- 170** Поэтому мы считаем жизненно важным запустить глобальный процесс, в котором примут участие различные участники, включая страны, находящиеся на разных уровнях развития, при поддержке соответствующих международных организаций из системы Организации Объединенных Наций и из-за ее пределов (ОЭСР, ВОИС и Всемирная торговая организация), с тем чтобы создать «страховочные сетки» и «общие рельсы» для процветания экосистем обучающих данных ИИ. Результатом этого процесса должны стать не обязательные к исполнению законодательные нормы, а типовые контракты и технико-юридические соглашения. Эти механизмы содействия могут разрабатываться по одному, как протоколы к системе принципов и определений.
- 171** Хотя всестороннее обсуждение данной темы и выходит за рамки нашей работы, ключевые принципы глобальной системы данных ИИ включали бы в себя совместимость, разумное руководство, сохранение конфиденциальности, расширение прав и возможностей, укрепление прав и создание экосистемы ИИ.
- 172** Мы помним о том, что антимонопольная и конкурентная политика остается прерогативой национальных и региональных властей. Однако международные коллективные действия могут облегчить трансграничный доступ к учебным данным для местных ИИ-стартапов, которых нет внутри страны.
- 173** Организация Объединенных Наций располагает уникальными возможностями для поддержки разработки глобальных принципов и практических механизмов управления и использования данных для обучения ИИ, опираясь на многолетнюю работу сообщества специалистов по данным и интегрируя ее с последними достижениями в области этики и управления ИИ. Это аналогично усилиям Комиссии Организации Объединенных Наций по праву международной торговли в области международной торговли, в том числе в отношении юридических и неюридических трансграничных рамок и создания условий для цифровой торговли и инвестиций при помощи типовых законов об электронной торговле, облачных вычислениях и управлении идентификационными данными.

- 174** Аналогичным образом, Комиссия по науке и технике в целях развития и Статистическая комиссия включают в свою повестку дня данные, связанные с развитием, и данные по ЦУР. Кроме того, ВОИС рассматривает важные вопросы контента, авторского права и охраны знаний и выражений культуры коренных народов.
- 175** Предлагаемые здесь рамки не наносят ущерба национальным или региональным системам защиты данных и не создают новых прав, связанных с данными, а также не предписывают, как существующие права применяются на международном уровне, но должны быть разработаны таким образом, чтобы предотвратить захват коммерческими или иными интересами, которые могут подорвать защиту прав или воспрепятствовать ей. Скорее глобальная система данных ИИ будет решать сквозные вопросы доступности, совместимости и использования данных для обучения ИИ. Она поможет выработать общее понимание того, как согласовать различные национальные и региональные системы защиты данных.
- 176** Меры по решению этих проблем на национальном и региональном уровнях являются многообещающими; при этом государственный и частный сектор уделяют больше внимания передовому опыту. Однако в отсутствие глобальной системы, регулирующей наборы данных для обучения ИИ, коммерческая конкуренция порождает «гонку на понижение» между юрисдикциями в отношении требований к доступу и использованию, что затрудняет управление цепочкой создания стоимости ИИ на международном уровне. Только глобальные коллективные действия могут способствовать «гонке на повышение» в сфере управления сбором, созданием, использованием и монетизацией обучающих данных ИИ таким образом, чтобы обеспечить взаимодействие, управление, сохранение конфиденциальности, а также расширение возможностей и прав.
- 177** Кроме того, такие меры необходимы для содействия процветанию местных экосистем ИИ и ограничения дальнейшей экономической концентрации. Эти меры могут быть дополнены продвижением идеи объединения массивов данных и положений о размещении трастов данных в областях, имеющих отношение к ЦУР (см. вставку 14). Разработка этих шаблонов, а также фактическое хранение и анализ данных, которые находятся в общих пулах или трастах, могут быть поддержаны сетью для наращивания потенциала и глобальным фондом для ИИ.

## **Вставка 14: Получение данных для обучения моделей ИИ: расширение прав и возможностей использования данных, трасты данных и механизмы трансграничных потоков данных**

---

Существует множество обстоятельств, при которых данные необходимо защищать (в том числе в целях обеспечения конфиденциальности, коммерческой тайны, охраны интеллектуальной собственности, безопасности), но при этом предоставление их для обучения моделей ИИ также принесет пользу людям и обществу.

Права на данные в законодательстве – это, как правило, права на предотвращение тех или иных действий в отношении данных. Права на конфиденциальность данных также являются личными правами человека. Институализация прав на данные может затруднить гибкое осуществление таких прав, которое позволяет использовать данные в некоторых целях без потери прав, и делать это коллективно, как группа. Даже если есть возможность гибко и конструктивно управлять разрешениями, это, как правило, требует более значительных временных затрат, технических знаний и уверенности, чем есть у большинства людей и организаций.

Механизмы, позволяющие владельцам и субъектам данных разрешать безопасное и ограниченное использование своих данных, сохраняя при этом свои права, можно назвать средствами расширения прав и возможностей использования данных. Расширение прав и возможностей использования данных может сделать гораздо больше людей и групп в обществе активными партнерами и заинтересованными сторонами в ИИ, а не только субъектами данных. Уже разрабатываются инструменты для безопасного управления доступом, включая трасты данных и приложения, защищающие конфиденциальность, для управления трансграничными потоками данных.

Трасты данных – это механизмы, позволяющие частным лицам и организациям предоставлять коллективный доступ к своим данным под контролем доверенных лиц. Владельцы данных могут устанавливать условия доступа, использования и целеполагания, выполнение которых обеспечивают доверенные лица. Владельцы и субъекты данных сохраняют свои законные права, внося свой вклад в достижение общих целей. Можно ожидать, что модель ИИ, обученная на этих данных, будет работать более точно, чем та, которая не имеет такого специфического вклада, и, таким образом, будет лучше служить благополучию этой конкретной группы или общества в целом.

Механизмы управления доступом и использованием, и в частности трансграничным доступом, опираются на специальную правовую базу. Использование этих механизмов на практике также требует адаптации к потребностям и условиям отраслей и сообществ. Необходимо выявлять и устранять пробелы в управлении данными. Успешное и широкое использование этих механизмов в будущем будет зависеть от технического обеспечения и сохранения доверия со стороны тех, кто предоставляет данные.

Поэтому мы предлагаем оказывать больше поддержки дальнейшему развитию этих инструментов и выявлению областей, в которых их использование для обучения ИИ может принести наибольшую общественную пользу.

## D. Согласованность усилий

**178** Содействуя общности понимания, принципов и выгод, изложенные выше предложения направлены на устранение пробелов, выявленных в формирующемся международном режиме управления ИИ. Пробелы в области представительства, координации и реализации могут быть устранены за счет партнерства и сотрудничества с существующими институтами и механизмами.

**179** Вместе с тем без специального координационного центра в Организации Объединенных Наций, который бы поддерживал мягкую координацию этих и других усилий и благоприятствовал ей, а также обеспечивал, чтобы система Организации Объединенных Наций выступала с единых позиций в отношении ИИ, в мире не будет инклюзивного, сетевого, гибкого и согласованного подхода, необходимого для эффективного и справедливого управления ИИ.

**180** По вышеизложенным причинам мы предлагаем создать небольшую и динамичную структуру в форме подразделения по вопросам ИИ в рамках Секретариата Организации Объединенных Наций.

### Подразделение по вопросам ИИ в рамках Секретариата Организации Объединенных Наций

#### Рекомендация 7: подразделение по вопросам ИИ в рамках Секретариата

Мы рекомендуем создать в рамках Секретариата подразделение по вопросам ИИ, подотчетное Генеральному секретарю. Оно должно быть облегченным и гибким по своей организации, по возможности опираясь на уже имеющиеся профильные структуры Организации Объединенных Наций. Выполняя роль «клея», скрепляющего

и катализирующего предложения, которые содержатся в настоящем докладе, поддерживая партнерские отношения и взаимодействуя с другими процессами и институтами, это подразделение должно руководствоваться мандатом, включающим в себя:

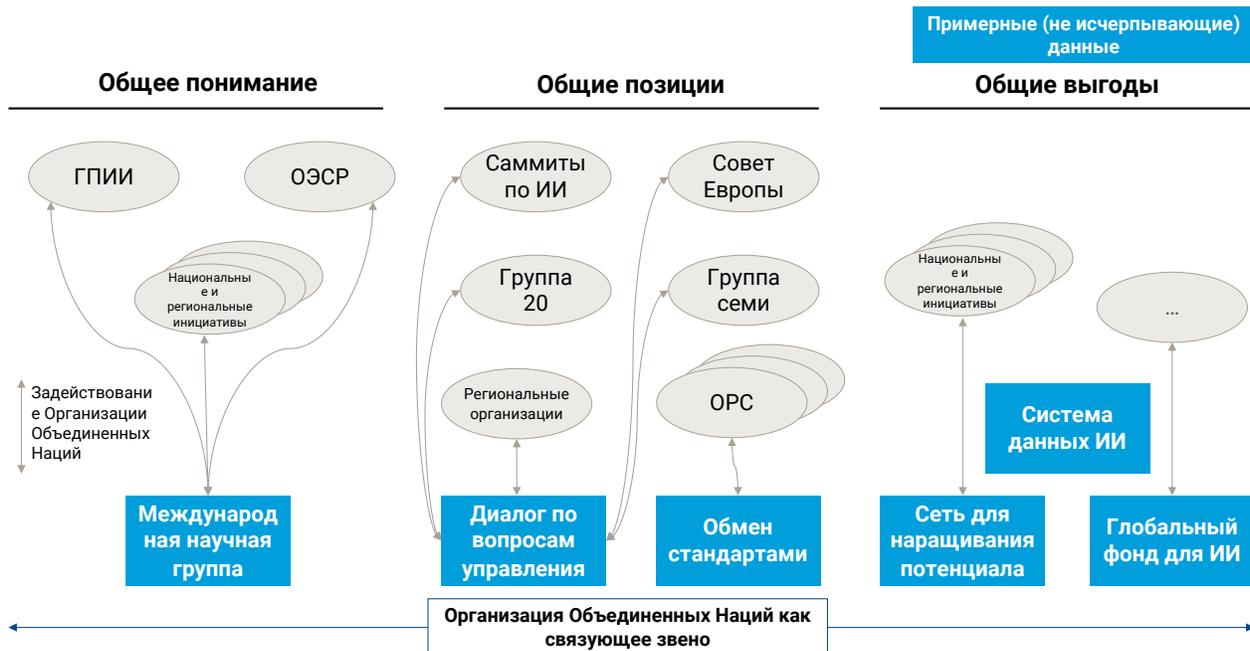
- a. поддержку предлагаемой международной научной группы, диалога по вопросам политики, обмена стандартами, сети для наращивания потенциала и, в случае необходимости, глобального фонда и глобальной системы данных ИИ;
- b. проведение информационно-разъяснительной работы с различными заинтересованными сторонами, включая технологические компании, гражданское общество и научные круги, по возникающим проблемам ИИ; и
- c. консультирование Генерального секретаря по вопросам, связанным с ИИ, в координации с другими соответствующими подразделениями системы Организации Объединенных Наций в целях обеспечения комплексного реагирования Организации Объединенных Наций.

**181** В ходе проведенных нами консультаций стало ясно, что доводы в пользу создания агентства, наделенного полномочиями по отчетности, мониторингу, проверке и правоприменению, до сих пор не были приведены, и государства-члены пока не проявляют особого интереса к созданию новой дорогостоящей организации.

**182** Поэтому мы делаем упор на тот ценный вклад, который может предложить Организация Объединенных Наций, памятуя о недостатках системы Организации Объединенных Наций, а также о том, чего реально можно достичь в течение года. В этой связи мы предлагаем создать облегченный и гибкий механизм, который будет служить «клеем», объединяющим процессы, которые способствуют достижению общности понимания, позиций и выгод, и позволит системе Организации Объединенных Наций выступать единым фронтом в развивающейся международной экосистеме управления ИИ.

- 183** Подобно тому, как страны создают специальные институты и бюро, занимающиеся вопросами национального, регионального и международного управления ИИ, мы видим необходимость в создании структуры, которая обслуживала и поддерживала бы международную научную группу по ИИ<sup>45</sup> и политический диалог по ИИ, а также стимулировала бы обмен стандартами ИИ и сеть развития потенциала – с меньшими накладными расходами и операционными издержками, чем если бы каждый из них поддерживался разными организациями.
- 184** Подразделение по вопросам ИИ в Секретариате Организации Объединенных Наций, подотчетное Генеральному секретарю, будет иметь преимущества связей в рамках всей системы Организации Объединенных Наций, не будучи привязанным к какой-то одной ее части. Это важно ввиду неопределенности будущего ИИ и большой вероятности того, что он проникнет во все сферы деятельности человека.
- 185** Небольшое и динамичное подразделение по вопросам ИИ будет иметь все возможности для налаживания связей между различными областями и организациями по вопросам управления ИИ, с тем чтобы помочь динамично устранять пробелы, работая над укреплением существующих усилий в рамках Организации Объединенных Наций и за ее пределами. Налаживая контакты и взаимосвязи с другими инициативами, например с инициативами региональных организаций и другими многосторонними инициативами, оно может способствовать снижению затрат на сотрудничество между ними.
- 186** Такой орган должен отстаивать идею инклюзивности и оперативно налаживать партнерские отношения для ускорения координации и реализации, опираясь в первую очередь на существующие ресурсы и функции в рамках системы Организации Объединенных Наций. Он может быть частично укомплектован персоналом Организации Объединенных Наций, прикомандированным из соответствующих специализированных учреждений и других подразделений системы Организации Объединенных Наций. Ей следует привлекать различные заинтересованные стороны, включая гражданское общество, представителей отрасли и научные круги, и развивать партнерские отношения с ведущими организациями, не входящими в систему Организации Объединенных Наций, такими как ОЭСР.
- 187** Подразделение по вопросам ИИ будет обеспечивать обмен информацией в рамках всей системы Организации Объединенных Наций и позволит системе выступать авторитетно и с единых позиций. Во вставке 15 перечислены возможные функции и первые результаты работы такого подразделения.
- 188** Эта рекомендация основана на четкой оценке того, где Организация Объединенных Наций может принести пользу, в том числе там, где она может играть ведущую роль, где она может заполнить пробелы, где она может содействовать координации и где она должна отойти в сторону, тесно взаимодействуя с существующими инициативами (см. рис. 17). Она также использует преимущества существующих институциональных механизмов, включая заранее согласованное финансирование и хорошо понятные административные процессы.
- 189** Необходимо учитывать меняющиеся характеристики технологии ИИ. Существует высокая вероятность технических прорывов, которые кардинально изменят существующий ландшафт моделей ИИ. Такое подразделение по вопросам ИИ должно добиваться эффективных результатов, с тем чтобы корректировать систему управления с учетом меняющегося ландшафта и реагировать на непредвиденные изменения, связанные с технологией ИИ.

## Рисунок 17: Предлагаемая роль Организации Объединенных Наций в экосистеме международного управления ИИ



Сокращения: ГПИИ – Глобальное партнерство по искусственному интеллекту; ОЭСР – Организация экономического сотрудничества и развития; ОПС – организации по разработке стандартов.

### Вставка 15: Возможные функции и результаты первого года работы подразделения по вопросам ИИ

Подразделение по вопросам ИИ должно иметь облегченную структуру и стремиться строить свою работу на принципах гибкости, надежности и сетевой организации. Там, где это необходимо, оно должно работать по принципу «спиц на втулке» для связи с другими частями системы Организации Объединенных Наций и за ее пределами.

Информационно-координационная работа может включать в себя выполнение функций ключевого узла в так называемой мягкой координационной архитектуре с участием государств-членов, многосторонних сетей, организаций гражданского общества, научных кругов и технологических компаний в рамках комплекса механизмов, которые объединяются для совместного решения проблем путем создания сетей, а также в качестве безопасного и надежного места для проведения встреч по соответствующим темам. Программа-максимум заключается в том, чтобы стать тем «клеем», который помогает удерживать вместе другие развивающиеся сети.

Поддержка различных инициатив, предложенных в данном докладе, включает в себя важную функцию обеспечения инклюзивности в контексте оперативного достижения результатов, таких как научные отчеты, диалог по вопросам управления и определение соответствующих будущих структур.

## Вставка 15: Возможные функции и результаты первого года работы подразделения по вопросам ИИ (продолжение)

---

### Общее понимание:

- способствовать комплектованию и поддержке международной научной группы.

### Общие позиции:

- обслуживание политических диалогов с участием различных заинтересованных сторон в поддержку операционной совместимости и ознакомления с соответствующими программными положениями. На первом этапе приоритетной темой является определение пороговых значений риска и рамок безопасности в разных юрисдикциях.
- Поддержка МСЭ, ИСО/МЭК и ИИЭЭ в создании механизмов обмена стандартами в области ИИ.

### Общие выгоды:

- поддержка сети для наращивания потенциала в области ИИ с упором на расширение возможностей в области ИИ среди государственных служащих и социальных предпринимателей в интересах общества. Определение первоначальной концепции сети, ожидаемых результатов, структуры управления, параметров сотрудничества и операционных механизмов.
- Определение стратегии, целей, структуры управления и операционных механизмов глобального фонда для ИИ и получение отзывов государств-членов, представителей отрасли и гражданского общества на соответствующее предложение с целью обеспечить финансирование первых проектов в течение шести месяцев после создания подразделения.
- Подготовка и публикация ежегодного списка приоритетных направлений инвестиционной деятельности в качестве ориентира как для глобального фонда для ИИ, так и для инвестиций, не относящихся к этой структуре.

### Согласованность усилий:

- создание облегченных механизмов, которые помогут государствам-членам и другим соответствующим организациям предпринимать более согласованные, скоординированные и эффективные усилия по реализации их глобальных усилий в сфере управления ИИ.
- Подготовка первоначальных рамочных параметров для руководства и мониторинга работы подразделения по вопросам ИИ, включая таксономию рисков глобального управления, обзор глобального политического ландшафта ИИ и глобальную карту заинтересованных сторон.
- Разработка и внедрение практики представления ежеквартальных отчетов и периодического проведения очных презентаций для государств-членов о ходе выполнения плана работы подразделения по вопросам ИИ, а также создание каналов обратной связи для внесения необходимых коррективов.
- Создание руководящего комитета под совместным управлением подразделения по вопросам ИИ, МСЭ, ЮНКТАД, ЮНЕСКО и других соответствующих подразделений и организаций системы Организации Объединенных Наций для ускорения работы Организации Объединенных Наций по выполнению вышеуказанных функций и проведение обзора прогресса в ускорении усилий каждые три месяца.
- Содействие совместному обучению и повышению квалификации представителей государств-членов, с тем чтобы помочь им выполнять свои обязанности по глобальному управлению ИИ, в сотрудничестве с соответствующими подразделениями и организациями системы Организации Объединенных Наций, такими как Учебный и научно-исследовательский институт Организации Объединенных Наций и Университет Организации Объединенных Наций.

## Е. Размышления на тему организационных моделей

- 190** Дискуссии на тему ИИ часто переходят в крайности. В ходе наших консультаций по всему миру мы общались с теми, кто видит будущее в безграничных возможностях, предоставляемых все более дешевыми и все более полезными системами ИИ. Мы также общались с теми, кто опасается мрачного будущего, разделения, безработицы и даже вымирания.
- 191** Мы не знаем, что может произойти в будущем. Мы понимаем, что развитие технологии может пойти в направлении, которое устранил эту двойственность. В настоящем докладе мы сосредоточились на возможностях и рисках ближайшего будущего, основываясь на научных данных. Рекомендации, изложенные в нем, дают нам надежду на то, что мы сможем воспользоваться преимуществами ИИ, сумев свести к минимуму и смягчить имеющиеся риски. Мы также помним о практических проблемах, связанных с созданием международных институтов в более широких масштабах. Именно поэтому мы предлагаем сетевой институциональный подход с облегченными и динамичными механизмами поддержки.
- 192** Однако в том случае, если риски станут более серьезными, а цена возможностей – более высокой, такие расчеты изменятся. Мировые войны привели к созданию современной международной системы; разработка все более мощного оружия привела к появлению режимов, ограничивающих его распространение и способствующих мирному использованию лежащих в его основе технологий.
- 193** Эволюция понимания нашей общей человеческой природы привела к созданию современной системы защиты прав человека и сформировала наши нынешние обязательства

по достижению ЦУР в интересах всех. Изменение климата превратилось из нишевой проблемы в глобальный вызов. Аналогичным образом, ИИ может подняться на уровень, требующий более значительных ресурсов и полномочий, чем предлагается в настоящем докладе.

- 194** В круг наших полномочий входило рассмотрение функций, форм и сроков создания нового международного агентства по ИИ. В заключительной части настоящего доклада мы приводим некоторые соображения по этому поводу, хотя в настоящее время мы не рекомендуем создавать такое агентство.

### Международное агентство по ИИ?

- 195** Если риски, связанные с ИИ, станут более серьезными и более концентрированными, государствам-членам, возможно, придется рассмотреть вопрос о создании более динамичного международного института, наделенного полномочиями по мониторингу, отчетности, проверке и правоприменению.
- 196** Подобное развитие событий не является беспрецедентным. Начиная с Гаагских конвенций 1899 и 1907 годов, Женевского протокола 1925 года и заканчивая Конвенцией о химическом оружии 1993 года, химические вещества двойного назначения уже давно ограничиваются в плане доступности, сопровождаются протоколами о хранении и использовании, а также запретом на их использование для создания оружия.
- 197** Биологическое оружие также было запрещено, наряду с периодическими ограничениями на исследования, такими как ограничения на рекомбинантную ДНК или сплайсинг генов в 1975 году. В контексте этих ограничений подчеркивалось, что сдерживание является важнейшим фактором при конструировании экспериментов, а уровень сдерживания зависит

от предполагаемого риска. Определенные классы экспериментов с высоким риском, для которых невозможно гарантировать сдерживание, были, в сущности, запрещены. Другие примеры включают исследования, угрожающие пересечь фундаментальные этические границы, такие как действующие ограничения на клонирование человека – пример «красной линии», которая может однажды понадобиться в контексте исследований ИИ, наряду с эффективным сотрудничеством в области правоприменения.

**198** Периодическая научная оценка также является характерной чертой некоторых из этих рамочных механизмов, включая, например, Научно-консультативный совет Организации по запрещению химического оружия и статью XII Конвенции о биологическом оружии.

**199** ИИ часто сравнивают с ядерной энергией. С момента расщепления атома ученым было ясно, что эта технология может быть использована во благо, даже если их исследования были направлены на создание нового страшного оружия. Показательно, что как тогда, так и сейчас ведущие ученые были среди тех, кто наиболее активно призывал ограничить эту новую технологию.

**200** Большая сделка, лежащая в основе Международного агентства по атомной энергии (МАГАТЭ), заключалась в том, что полезные свойства ядерной энергии могут быть использованы совместно – в производстве энергии, сельском хозяйстве и медицине – в обмен на гарантии того, что в дальнейшем она не будет использоваться для создания еще более мощного оружия. Как показывает режим нераспространения ядерного оружия, хорошо проработанные нормы являются необходимым, но не достаточным условием для эффективного регулирования.

**201** Пределы этой аналогии очевидны. Ядерная энергия включает в себя четко определенный набор процессов, связанных с конкретными

материалами, которые распределены неравномерно, и большая часть материалов и инфраструктуры, необходимых для создания ядерного потенциала, контролируется национальными государствами. ИИ – это аморфное понятие; его применение чрезвычайно широко, а его самые мощные возможности выходят за рамки отдельных отраслей и государств. Грандиозная сделка МАГАТЭ была направлена на оружие, которое дорого создавать и трудно скрывать; использование ИИ для производства оружия, вероятно, будет иным по своему характеру.

**202** Одна из первоначальных идей – совместное использование ядерного топлива в мирных целях – не сработала так, как планировалось. Если говорить об объединении ресурсов для совместного использования преимуществ технологии, то более подходящей для ИИ аналогией может быть ЦЕРН, который функционирует на основе объединения финансовых, людских и инфраструктурных ресурсов. Однако такое сравнение имеет свои пределы, учитывая разницу между экспериментальной фундаментальной физикой и ИИ, который требует более распределенного подхода.

**203** Еще одна несовершенная аналогия – это такие организации, как Международная организация гражданской авиации (ИКАО) и Международная морская организация (ИМО). Технологии, лежащие в основе перевозок, хорошо отработаны, и их гражданское применение можно легко отделить от военного, чего нельзя сказать об ИИ общего назначения. Сеть национальных регулирующих органов, применяющих международные нормы, разработанные в рамках ИКАО и ИМО, также хорошо развита. В центре внимания – безопасность, содействие коммерческой деятельности и оперативная совместимость. Соблюдение норм не обеспечивается по принципу «сверху вниз».

- 204** Существуют и другие подходы к соблюдению требований, которые могут послужить в качестве ориентира. Управление финансовыми рисками выигрывает от таких механизмов, как Совет по финансовой стабильности (СФС) и Группа разработки финансовых мер (ФАТФ); это делается без заключения каких-либо договоров.
- 205** В конечном счете, если регулирование ИИ потребует принятия правоприменительных мер, для придания «красным линиям» официального характера может возникнуть необходимость в том или ином механизме на глобальном уровне. Такой механизм может включать в себя официальные обязательства, подобные обязательствам ЦЕРН, по объединению ресурсов для сотрудничества в области исследований ИИ и распределению выгод в качестве части сделки.
- 206** Вместе с тем, если учитывать скорость развития, автономность и непрозрачность систем ИИ, ожидание появления угрозы может означать, что любые ответные меры будут приняты слишком поздно. Продолжение научных оценок и политического диалога позволит миру не дать застать себя врасплох. Само собой разумеется, что решение о начале официального процесса будет приниматься государствами-членами.
- 207** Возможные поводы для такого шага могут включать в себя перспективы разработки неконтролируемых или неуправляемых систем ИИ или же развертывания систем, которые невозможно отследить до человеческих, корпоративных или государственных субъектов. Они также могут включать признаки того, что системы ИИ проявляют качества, позволяющие говорить о появлении «сверхразума», хотя в современных системах ИИ этого нет.
- 208** Разумным первым шагом стало бы создание наблюдательного совета с привлечением различных авторитетных экспертов для отслеживания перспектив. Научной группе можно было бы поручить проведение исследований по этому вопросу в рамках ежеквартальной серии научных дайджестов. Со временем политический диалог мог бы стать подходящим форумом для обмена информацией об инцидентах, связанных с ИИ, таких как те, которые превышают возможности существующих структур, по аналогии с практикой МАГАТЭ по взаимным гарантиям ядерной безопасности или Всемирной организации здравоохранения (ВОЗ) – по надзору за заболеваниями.
- 209** Функции предлагаемого международного агентства по ИИ могли бы опираться на опыт соответствующих профильных агентств, таких как МАГАТЭ, Организация по запрещению химического оружия, ИКАО, ИМО, ЦЕРН и Конвенция по биологическому оружию. Они могли бы включать в себя:
- разработку и публикацию стандартов и норм в целях обеспечения безопасности ИИ;
  - мониторинг систем искусственного интеллекта, которые потенциально могут угрожать международному миру и безопасности или приводить к серьезным нарушениям прав человека или международного гуманитарного права;
  - получение и расследование сообщений об инцидентах или злоупотреблениях, а также информирование о серьезных нарушениях;
  - контроль соблюдения международных обязательств;
  - координация вопросов подотчетности, реагирования на чрезвычайные ситуации и возмещения ущерба в связи с инцидентами, касающимися безопасности ИИ;
  - поощрение международного сотрудничества по использованию ИИ в мирных целях.
- 210** Для создания любого будущего агентства по ИИ потребуется индивидуальный подход, при необходимости учитывающий опыт других учреждений (см. вставку 16).

## Вставка 16: Уроки, извлеченные из опыта ранее созданных глобальных механизмов управления

---

ИИ – это уникальный набор технологий, риски и социальные последствия которых выходят за рамки границ отдельных государств. Однако это не первый набор технологий, которые привели к созданию глобальных механизмов управления. Гражданская авиация, изменение климата, ядерная энергетика и финансирование терроризма также являются сложными и многоплановыми областями, которые потребовали глобальной реакции.

Некоторые из этих областей, такие как гражданская авиация, изменение климата и ядерная энергетика, привели к созданию новых учреждений Организации Объединенных Наций. Другие, в частности защита глобальных финансовых потоков, привели к созданию органов, не основанных на договорах, но при этом обеспечивающих надежную нормативную базу, эффективные рыночные механизмы правоприменения и прочные партнерские отношения между государственным и частным секторами.

Проводя параллели между этими мерами организационного характера и зарождающимися попытками сделать то же самое применительно к ИИ, мы не должны слишком сильно фокусироваться на том, какой именно институциональный аналог наиболее подходит для решения проблем ИИ. Наш промежуточный доклад предполагал, что нам следует в первую очередь обратить внимание на то, какие руководящие функции необходимы для эффективного и инклюзивного глобального управления ИИ, а также на то, какие уроки мы можем извлечь из прошлых попыток глобального управления.

Один из таких уроков заключается в том, что для выработки общеприемлемых политических мер необходимо достичь общего научно-технического понимания проблемы. Полезным образцом в данной связи является МГЭИК, которая продолжает рассматривать риски, связанные с изменением климата. Она служит примером того, как инклюзивный подход к подготовке докладов и выработке научного консенсуса в непрерывно эволюционирующей области может уравнивать шансы исследователей и политиков и сформировать общее понимание, необходимое для эффективной разработки политики. Процесс подготовки и распространения докладов МГЭИК и глобальных инвентаризаций, хотя и не обошелся без трудностей, сыграл центральную роль в формировании общего понимания и общей базы знаний, снижении затрат на сотрудничество и нацеленности работы Конференции сторон Рамочной конвенции Организации Объединенных Наций об изменении климата на достижение конкретных политических результатов.

По мере эволюции технологии ИИ будет не менее важно выработать в ее отношении общее научное понимание. Поскольку возможности систем ИИ продолжают развиваться, а потенциальные риски могут превышать возможности известных эффективных подходов к их снижению, международная научная группа может быть видоизменена в соответствии с возникающими потребностями.

Второй урок заключается в том, что сотрудничество между различными заинтересованными сторонами может обеспечить надежные стандарты и способствовать быстрому реагированию. Здесь ИКАО и ФАТФ предлагают полезные примеры того, как можно регламентировать узкотехнические вопросы на международном уровне. В гражданской авиации стандарты безопасности ИКАО, разработанные отраслевыми и правительственными экспертами и обеспечиваемые ограничениями доступа на рынок, гарантируют, что самолет, взлетающий, например, из Нью-Йорка, может приземлиться в Женеве, не вызывая новых проверок безопасности. Сочетание проверок безопасности, проводимых под руководством ИКАО, и проверок, проводимых государствами-членами, обеспечивает последовательное соблюдение стандартов даже на фоне дальнейшего развития технологий.

## Вставка 16: Уроки, извлеченные из опыта ранее созданных глобальных механизмов управления (продолжение)

ФАТФ, созданная G-7 в 1989 году для борьбы с отмыванием денег, является еще одним примером того, как институты «мягкого права» могут способствовать внедрению общих стандартов и их реализации. Ее система экспертных оценок для мониторинга является гибкой, а широкое признание ее рекомендаций создает репутационные издержки для тех компаний и государств-членов, которые не соблюдают их. Даже когда риски для международных финансовых потоков изменились, особенно в связи с ростом терроризма и финансирования распространения, гибкая структура и нормативная база ФАТФ позволили ей быстро реагировать на сложные вызовы и идти в ногу со временем.

Действуя в своей собственной уникальной манере, ИКАО и ФАТФ создали общепризнанные международные стандарты, внутренние механизмы оценки их соблюдения и совместимые системы реагирования на определенные категории рисков и проблем, проявляющихся в разных юрисдикциях. ИКАО обеспечивает соблюдение стандартов через стимулы и ограничения доступа на рынок, а ФАТФ создает репутационные риски за их несоблюдение. Обе организации представляют собой полезные образцы для сферы ИИ, поскольку демонстрируют, как правительства и другие заинтересованные стороны могут работать вместе, создавая сеть взаимосвязанных норм и правил и предусматривая санкции за их несоблюдение.

Третий урок заключается в том, что глобальная координация зачастую жизненно необходима для мониторинга и принятия мер в ответ на серьезные риски, способные оказать широкомасштабное воздействие. Ключевыми примерами являются модели СФС и МАГАТЭ. СФС был создан в 2009 году странами G-20 для мониторинга и предупреждения системных рисков для международной финансовой системы. Его уникальный состав, включающий представителей финансовых ведомств стран G-20 и международных финансовых организаций, а также структур, занимающихся вопросами развития, позволил ему быть динамичным, эффективным и инклюзивным в контексте координации усилий по выявлению глобальных финансовых рисков.

Подход МАГАТЭ к ядерным гарантиям предлагает иную модель. Его соглашения о всеобъемлющих гарантиях, подписанные 182 государствами, являются частью самого широкомасштабного режима Организации Объединенных Наций по обеспечению их соблюдения. Используя сочетание инспекций и мониторинга, а также угрозу принятия мер Советом Безопасности, МАГАТЭ предлагает, возможно, наиболее наглядные санкции в адрес государств-членов, не выполняющих свои обязательства.

И СФС, и МАГАТЭ демонстрируют, что международная координация является основополагающей для мониторинга серьезных рисков. По мере того как риски, связанные с ИИ, становятся все более очевидными и ярко выраженными, может возникнуть аналогичная потребность в создании нового учреждения, ориентированного на ИИ, которое будет максимально эффективно координировать усилия и отслеживать серьезные и системные риски, с тем чтобы государства-члены могли, по мере возможности, принимать упреждающие меры для нейтрализации этих рисков.

Четвертый урок заключается в том, что важно обеспечить всеобщий доступ к ресурсам, необходимым для исследований и разработок, а также к тем преимуществам, которые они дают. Поучительным является опыт ЦЕРН и МАГАТЭ. ЦЕРН объединяет ученых и физиков мирового уровня для проведения сложных исследований в области ускорителей частиц и других проектов, которые должны принести пользу человечеству. Он также предлагает услуги по обучению физиков и инженеров.

## **Вставка 16: Уроки, извлеченные из деятельности прошлых институтов глобального управления (продолжение)**

---

Аналогичным образом, МАГАТЭ облегчает доступ к технологиям, в данном случае к тем из них, которые имеют отношение к ядерной энергии и ионизирующему излучению. Основной компромисс прост: государства-члены соблюдают ядерные гарантии, а МАГАТЭ предлагает техническую помощь в использовании ядерной энергии в мирных целях. В данной связи МАГАТЭ обеспечивает инклюзивный подход к распространению преимуществ технологии среди развивающихся стран. Содействие созданию сети центров передового опыта в области ядерной безопасности со стороны МАГАТЭ аналогично нашей рекомендации о сетевом подходе к наращиванию потенциала.

Как мы уже разъясняли выше, ИИ – это комплекс технологий, выгоды от которых должны распределяться на более инклюзивной и справедливой основе, особенно среди стран глобального Юга. Именно поэтому мы и рекомендовали создать как сеть для наращивания потенциала ИИ, так и глобальный фонд для ИИ. По мере того, как мы будем узнавать все больше об ИИ благодаря работе международной научной группы, а ответственное внедрение ИИ в деятельность по поддержке ЦУР будет становиться все более актуальным, государства-члены Организации Объединенных Наций, возможно, захотят более широко институционализировать эту функцию. В этом случае им следует использовать опыт ЦЕРН и МАГАТЭ в качестве полезных образцов для поддержки более широкого доступа к ресурсам в рамках общей глобальной структуры управления ИИ.

## 5. Заключение: призыв к действию

- 211** Как эксперты мы по-прежнему с оптимизмом смотрим на будущее ИИ и его потенциал для использования во благо. Однако этот оптимизм зависит от реалистичного подхода к имеющимся рискам, а также к вопросу о неадекватности существующих структур и стимулов. Мы также должны быть реалистами в отношении взаимных международных подозрений, которые могут помешать глобальным коллективным действиям, необходимым для эффективного и справедливого управления. Эта технология слишком важна, а ставки – слишком высоки для того, чтобы полагаться только на рыночные силы и разрозненные национальные и многосторонние действия.
- 212** Мы должны быть активными и целеустремленными. За двойственностью возможностей и рисков стоит проблема быстрых и всеобъемлющих изменений. В результате воздействия ИИ мало кто останется незатронутым. Если передать управление им в руки нескольких разработчиков или тех стран, в которых они размещаются, то это приведет к глубоко несправедливой ситуации, когда последствия разработки, внедрения и использования ИИ будут навязаны большинству людей без права голоса в принятии соответствующих решений.
- 213** На человеческом уровне прошедший год, в течение которого вопросы управления ИИ стали предметом внимания и обсуждений во всем мире, вселил в нас надежду. В разных странах и отраслях наблюдаются не только расхождения, но и сильное стремление к диалогу. Привлечение различных экспертов, политиков, бизнесменов, исследователей и активистов из числа мужчин и женщин, представляющих различные регионы и дисциплины, показало нам, что разнообразие не во всех случаях порождает разногласия, а диалог способен приводить к выработке общего мнения и сотрудничеству.
- 214** Иногда мы колебались: стоит ли нам быть прагматиками и сосредоточиться на том, что кажется выполнимым? Или же нам следует стремиться к высоким целям? В конце концов мы решили сделать и то, и другое. Наши предложения отражают всеобъемлющее видение справедливого и эффективного глобального режима управления ИИ; при этом мы тщательно продумали пошаговый план его реализации.
- 215** Мы благодарны многим людям, организациям и государствам-членам, которые внесли свой вклад в наши обсуждения, включая представителей учреждений Организации Объединенных Наций и сотрудников Секретариата, которые предложили проницательные оценки возможностей и ограничений Организации Объединенных Наций в этой сложной области. Вопрос управления ИИ связан не только с управлением последствиями этой технологии. На карту также поставлено будущее многостороннего сотрудничества с участием различных заинтересованных субъектов.
- 216** Если мы оглянемся назад через пять лет, то технологический ландшафт может оказаться кардинально иным, нежели сегодня. Однако если мы будем придерживаться выбранного курса и преодолеем колебания и сомнения, то через пять лет сможем

увидеть ландшафт управления ИИ, который будет всеохватывающим и расширяющим возможности отдельных людей, сообществ и государств во всем мире. В конечном итоге важны не сами технологические изменения, а то, как человечество реагирует на них.

**217** Мы считаем, что функции и формы, рекомендованные в настоящем докладе, при добросовестной реализации могут обеспечить гибкий и адаптируемый режим, который будет идти в ногу с развитием ИИ и поможет извлечь из него выгоду и устранить риски. Они помогут нам своевременно выявлять проблемы и возможности, использовать общие принципы и рамки для согласования международных действий, развивать международное сотрудничество и наращивать потенциал отдельных лиц и институтов для адаптации к происходящим изменениям.

**218** Выполнение рекомендаций, содержащихся в настоящем докладе, может также способствовать формированию нового образа мышления: мышления, основанного

на сотрудничестве и освоении новых знаний, вовлечении многих заинтересованных сторон и широком задействовании общественности. Организация Объединенных Наций может стать проводником нового общественного договора по ИИ, который обеспечит глобальное согласие на режим управления, защищающий и расширяющий возможности всех нас. Такой договор обеспечит справедливый доступ к возможностям и их справедливое распределение, а риски не будут перекладываться на плечи наиболее уязвимых слоев населения или на будущие поколения, как мы уже видели в случае с изменением климата.

**219** В качестве коллектива и в качестве отдельных людей, представляющих многие области знаний, организации и регионы мира, мы с нетерпением ждем продолжения этого важного разговора. Вместе со многими, с кем мы общались на этом пути, и с глобальным сообществом, которое они представляют, мы надеемся, что этот доклад внесет вклад в наши совместные усилия по управлению ИИ в интересах человечества.

# Приложения

## Приложение А: члены Консультативного органа высокого уровня по искусственному интеллекту

Анна Абрамова

Омар Султан аль-Олама

Латифа аль-Абдулкарим

Эстела Аранья

Карме Артигас (Сопредседатель)

Ран Балицер

Паоло Бенанти

Абеба Бирхане

Иэн Бреммер (Сопредседатель)

Анна Кристманн

Наташа Крэмpton

Нигхат Дад

Вилас Дхар

Вирджиния Дигнум

Ариса Эма

Мохамед Фарахат

Амандип Сингх Гилл

Венди Холл

Рахаф Харфуш

Руимин Хе

Хироаки Китано

Хаксоо Ко

Андреас Краузе

Джеймс Манийка (Сопредседатель)

Мария Ванина Мартинес Поссе

Сейдина Моусса Ндиайе

Мира Мурати

Петри Миллимяки

Алондра Нельсон

Назнин Раджани

Крейг Рамлал

Эмма Рутткамп-Блоэм

Марияетье Шааке (Содокладчик)

Шарад Шарма

Яан Таллинн

Филип Тиго

Химена София Виверос Альварес

Цзэн И

Чжан Линхань

# Приложение В: круг ведения Консультативного органа высокого уровня по искусственному интеллекту

Консультативный орган высокого уровня по искусственному интеллекту, созданный Генеральным секретарем Организации Объединенных Наций, будет проводить анализ и разрабатывать рекомендации по международному управлению искусственным интеллектом. Первые доклады органа станут экспертным и независимым вкладом высокого уровня в текущие национальные, региональные и многосторонние дискуссии.

В состав органа войдут 38 членов, представляющих правительства, частный сектор, гражданское общество и научные круги, а также член-секретарь. Его состав будет сбалансирован по полу, возрасту, географическому представительству и области знаний, связанных с рисками и применением искусственного интеллекта. Члены Органа будут выступать в личном качестве.

Орган будет широко взаимодействовать и консультироваться с правительствами, частным сектором, научными кругами, гражданским обществом и международными организациями. Он будет на гибкой и инновационной основе взаимодействовать с существующими процессами и платформами, а также использовать вклад различных заинтересованных сторон. Он может создавать рабочие группы или группы по конкретным темам.

Члены Органа будут отбираться Генеральным секретарем на основе предложений государств-членов и публичного предложения к специалистам выдвигать свои кандидатуры. В его состав войдут два сопредседателя; кроме того, будет создан Исполнительный комитет. В Исполнительном комитете будут представлены все группы заинтересованных сторон.

Орган созывается на первоначальный период в один год с возможностью продления Генеральным секретарем. Он будет проводить как очные, так и онлайн-заседания.

Орган подготовит первый доклад к 31 декабря 2023 года для рассмотрения Генеральным секретарем и государствами-членами Организации Объединенных Наций. В этом первом докладе будет представлен анализ вариантов международного управления искусственным интеллектом, проведенный специалистами высокого уровня.

На основе отзывов о первом докладе Орган представит к 31 августа 2024 года второй доклад, в котором могут быть даны подробные рекомендации по функциям, форме и срокам создания нового международного агентства по управлению искусственным интеллектом.

Орган должен избегать дублирования с существующими форумами и механизмами, в рамках которых рассматриваются вопросы искусственного интеллекта. Вместо этого он должен стремиться использовать существующие платформы и партнерские организации, включая структуры Организации Объединенных Наций, работающие в смежных областях. Он должен в полной мере уважать существующие структуры Организации Объединенных Наций, а также национальные, региональные и отраслевые прерогативы в управлении искусственным интеллектом.

Работа Органа будет поддерживаться небольшим секретариатом, базирующимся в Канцелярии Посланника Генерального секретаря по вопросам технологий, и финансироваться за счет внебюджетных донорских ресурсов.

# Приложение С: список консультативных мероприятий в 2024 году

| Мероприятие   | Дата, 2024 год | Регион            |
|---|----------------|-------------------|
| ЮНЕСКО – Словения   | 5 января       | Европа            |
| Научно-консультативный совет Генерального секретаря   | 8 января       | Глобальный охват  |
| Презентация промежуточного доклада государствам-членам  | 12 января      | Глобальный охват  |
| Всемирный экономический форум в Давосе  | 24 января      | Европа            |
| Совещание старших должностных лиц членов Ассоциации государств Юго-Восточной Азии (АСЕАН), занимающихся вопросами цифровизации  | 30 января      | Азия              |
| Всемирный правительственный саммит  | 12 февраля     | Ближний Восток    |
| Монреальский институт обучающих алгоритмов (МИОА – Квебекский институт ИИ)  | 14 февраля     | Северная Америка  |
| Берлинская консультация   | 15 февраля     | Европа            |
| Евразийский форум по ИТ   | 20 февраля     | Глобальный охват  |
| Всемирный конгресс по вопросам мобильной связи  | 26 февраля     | Европа            |
| Московский государственный институт международных отношений   | 28 февраля     | Европа            |
| Семинар Королевского общества по международному управлению ИИ   | 28 февраля     | Европа            |
| Консультативная сеть министерств иностранных дел по науке и технологиям   | 28 февраля     | Глобальный охват  |
| Диалог ОЭСР и Африканского союза по вопросам ИИ   | 4 марта        | Европа            |
| Брюссельская консультация   | 5 марта        | Европа            |
| Всемирный банк, Глобальный цифровой саммит  | 5 марта        | Северная Америка  |
| Открытая наука и искусственный интеллект: вебинар по этическим вопросам   | 5 марта        | Восточная Европа  |
| Диалог ЮНЕСКО по цифровой трансформации   | 6 марта        | Европа            |
| Межпарламентский союз   | 6 марта        | Глобальный охват  |
| Сорок седьмая сессия Комитета высокого уровня по программам   | 11 марта       | Глобальный охват  |
| Глобальный молодежный саммит по цифровым правам   | 13 марта       | Латинская Америка |
| Встреча на высшем уровне Группы семи (G-7) по ИИ в Тренто, Италия   | 15 марта       | Европа            |
| Первые совещания консультативной сети, 18–19 марта  | 18 марта       | Глобальный охват  |
| Шестьдесят восьмая сессия Комиссии по положению женщин  | 21 марта       | Северная Америка  |
| Информирование Консультативным органом государств-членов о ходе работы  | 25 марта       | Глобальный охват  |
| Африканская наблюдательная организация по вопросам ответственного развития ИИ   | 25 марта       | Африка            |
| Конференция по использованию ИИ для построения устойчивого и инклюзивного будущего – Французское агентство развития   | 26 марта       | Европа            |
| Формирование глобальных норм: коллективная обратная связь   | 28 марта       | Африка            |
| Организация «Инновации в Швейцарии»   | 2 апреля       | Европа            |
| Поездка представителей КППГСТ в Китай, 9–12 апреля  | 9 апреля       | Азия              |
| Российский форум по управлению Интернетом   | 9 апреля       | Восточная Европа  |
| Дни цифровых технологий в Уортоне – финансы   | 12 апреля      | Северная Америка  |
| Поездка в Кремниевую долину   | 15 апреля      | Северная Америка  |
| Стэнфордский университет, симпозиум «ИИ и политика: глобальный анализ»  | 16 апреля      | Северная Америка  |
| Комиссия Организации Объединенных Наций по науке и технике в целях развития   | 16 апреля      | Европа            |
| Группа 20 (G-20), «Цифровая экономика», 16–18 апреля  | 17 апреля      | Латинская Америка |
| Отчет Консультативного органа перед государствами-членами о ходе работы   | 22 апреля      | Глобальный охват  |
| Университет Организации Объединенных Наций, Конференция по ИИ в Макао, 24–25 апреля   | 24 апреля      | Азия              |
| Поездка членов КППГСТ в Брюссель и Париж, 25–26 апреля  | 26 апреля      | Европа            |
| Презентация Консультативного органа для Национального консультативного комитета по ИИ (Соединенные Штаты)   | 2 мая          | Северная Америка  |
| Глобальная ассамблея по искусственному интеллекту (ГАИИ) в Эр-Рияде, совместно с Исламской всемирной организацией по вопросам образования, науки и культуры (53 страны, четыре региона) | 14 мая         | Ближний Восток    |
| ИИ в интересах устойчивого развития: вклад Казахстана в реализацию Повестки дня на период до 2030 года  | 20 мая         | Азия              |
| Группа государств Латинской Америки и Карибского бассейна   | 21 мая         | Латинская Америка |
| Академический форум БРИКС   | 22 мая         | Глобальный охват  |
| Сессия по вопросам управления ИИ в Сеуле  | 23 мая         | Азия              |
| Азиатский технологический саммит, Сингапур, 29–31 мая   | 29 мая         | Азия              |
| Глобальный саммит «ИИ во благо», 29–31 мая  | 29 мая         | Европа            |

## Приложение D: список «глубоких погружений»

| Область   | Дата (восточное летнее поясное время) |
|---|---------------------------------------|
| Образование   | 29 марта                              |
| Интеллектуальная собственность и контент                                  | 2 апреля                              |
| Дети  | 4 апреля                              |
| Мир и безопасность (1)  | 12 апреля                             |
| Мир и безопасность (2)  | 29 апреля                             |
| Сельское хозяйство (сессия 1)   | 30 апреля                             |
| Сельское хозяйство (сессия 2)   | 30 апреля                             |
| Религиозные аспекты   | 1 мая                                 |
| Направления развития общедоступного программного обеспечения и технологий | 1 мая                                 |
| Последствия для общества  | 3 мая                                 |
| Гендерные вопросы   | 7 мая                                 |
| Данные  | 13 мая                                |
| Будущее работы  | 13 мая                                |
| Стандарты (сессия 1)  | 14 мая                                |
| Стандарты (сессия 2)  | 14 мая                                |
| Мир и безопасность (3)  | 20 мая                                |
| Окружающая среда  | 20 мая                                |
| Здравоохранение   | 22 мая                                |
| Верховенство права, права человека и демократия                           | 24 мая                                |

# Приложение Е: ответы, полученные в рамках Глобальной оценки рисков, связанных с ИИ

По просьбе Консультативного органа высокого уровня по искусственному интеллекту Канцелярия Посланника Генерального секретаря по вопросам технологий (КПГСТ) провела опрос «Глобальная оценка рисков, связанных с ИИ» в рамках исследования перспектив развития ИИ в целях выяснения мнения экспертов со всего мира о рисках ИИ. Экспертам было предложено высказать свое мнение в личном качестве (не от имени своего учреждения или работодателя). Экспертам было предложено оценить, насколько, по их мнению, ускорятся или замедлятся технические изменения в сфере ИИ и (отдельно) внедрение и применение ИИ.

Их также просили оценить общий уровень обеспокоенности тем, что вред (существующий или новый), причиняемый ИИ, станет значительно более серьезным и/или широко распространенным, а также то, насколько эта обеспокоенность возросла или снизилась в последнее время. Респондентам был предложен список из 14 примерных областей возможного вреда (например, «Преднамеренное злонамеренное использование ИИ негосударственными субъектами»), с тем чтобы они могли оценить уровень своей обеспокоенности. Наконец, было предложено множество текстовых ответов, в которых экспертам предлагалось прокомментировать формирующиеся тенденции, а также высказаться по поводу отдельных лиц, групп и (эко)систем, подверженных особому риску со стороны ИИ, и подробнее остановиться на своих оценочных ответах.

Опрос проводился с 13 по 25 мая 2024 года, а список респондентов был составлен на основе данных КПГСТ и сетей Консультативного органа, включая участников «глубоких погружений» Консультативного органа. В период проведения опроса постоянно приглашались дополнительные эксперты, особенно из регионов, зачастую менее широко представленных в дискуссиях об ИИ, на основе рекомендаций первоначальных респондентов и обращения к региональным сетям. На опрос ответили более 340 респондентов, представивших богатые и разнообразные мнения (в том числе мнения мужчин и женщин из разных регионов) относительно рисков, связанных с ИИ.

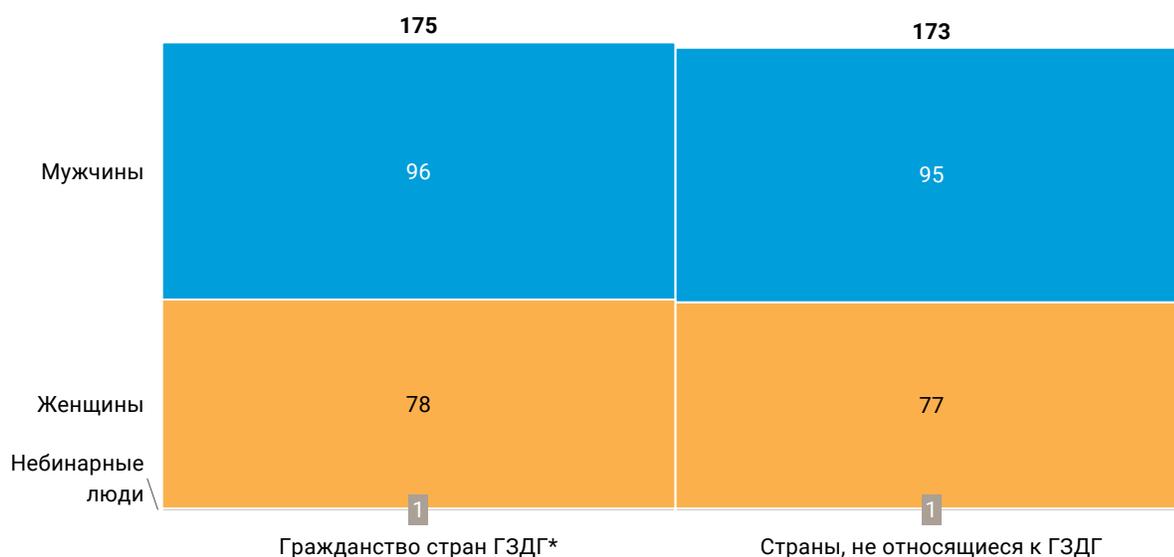
## Обзор выборки

### Разбивка по полу и региону является сбалансированной

Одномерный анализ по полу и региону не подвержен немедленной контаминации другой переменной

Респонденты в разбивке по региону, где расположена страна их гражданской принадлежности\* (n = 348)

ПО РЕГИОНУ И ПОЛУ



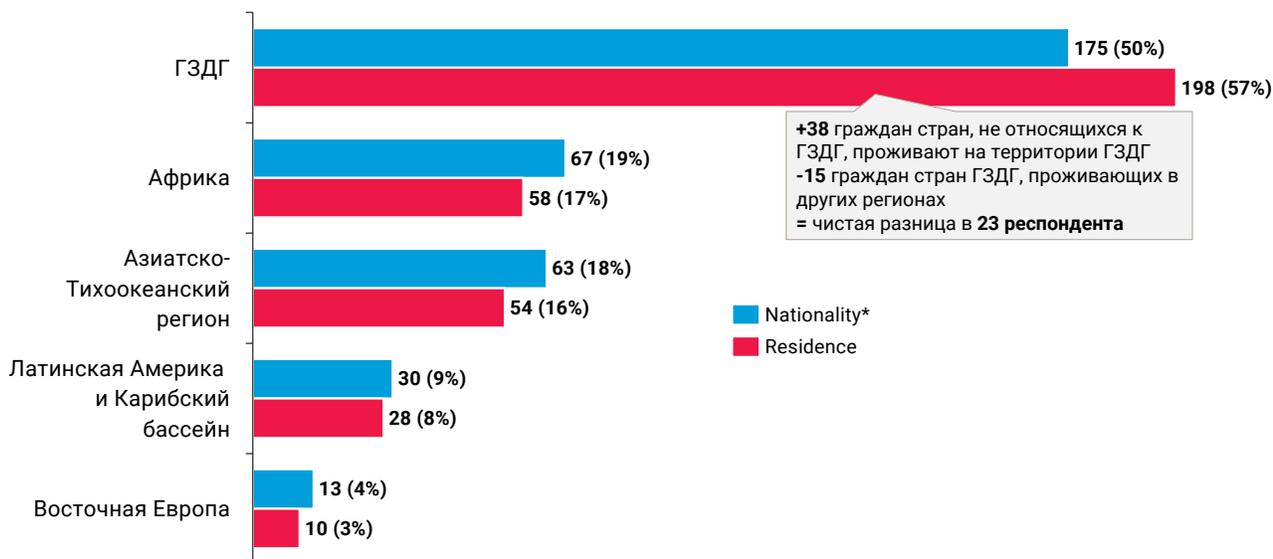
\* 43 респондента (12 процентов) указали, что являются гражданами более чем одного государства. Если страна проживания респондента совпадала с одной из стран его гражданской принадлежности, для целей анализа использовалась эта страна гражданской принадлежности (34 из 43). В остальных случаях использовалась наименее представленная гражданская принадлежность (9 из 43). Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

# выборка остается глобальной при ее рассмотрении с учетом стран проживания респондентов

84 процента респондентов проживают в регионе, где расположена страна их гражданской принадлежности

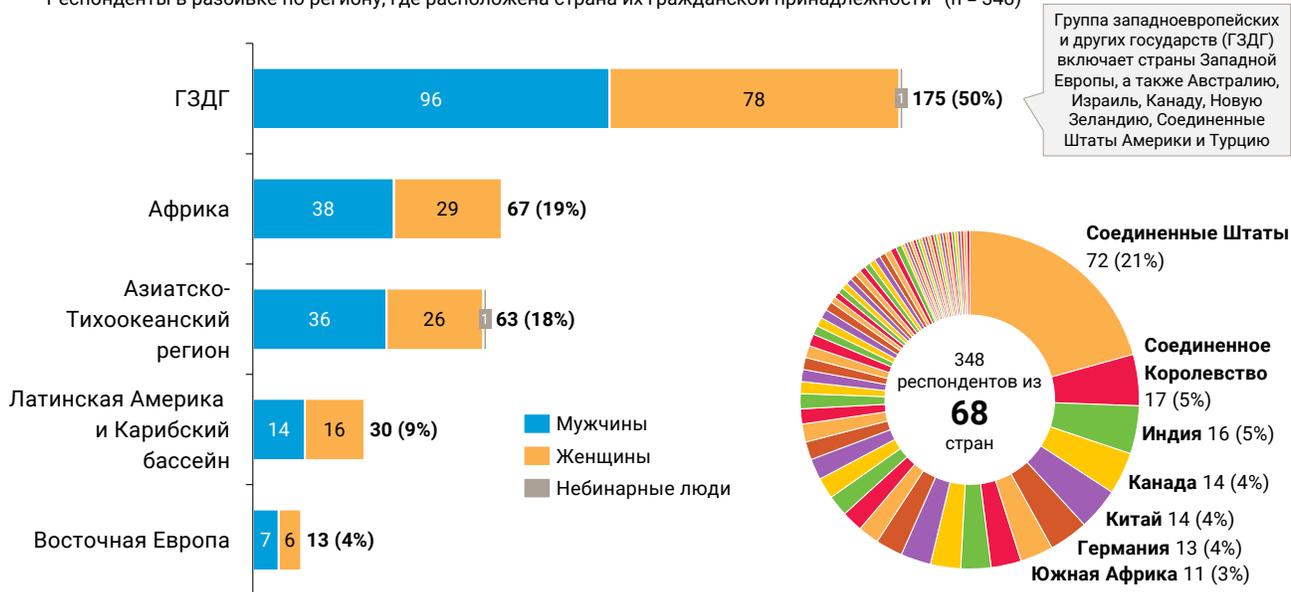
Респонденты в разбивке по региону, где расположена страна их гражданской принадлежности\* (n = 348)

**ПО РЕГИОНАМ**



\* 43 респондента (12 процентов) указали, что являются гражданами более чем одного государства. Если страна проживания респондента совпала с одной из стран его гражданской принадлежности, для целей анализа использовалась эта страна гражданской принадлежности (34 из 43). В остальных случаях использовалась наименее представленная гражданская принадлежность (9 из 43). Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

Респонденты в разбивке по региону, где расположена страна их гражданской принадлежности\* (n = 348)



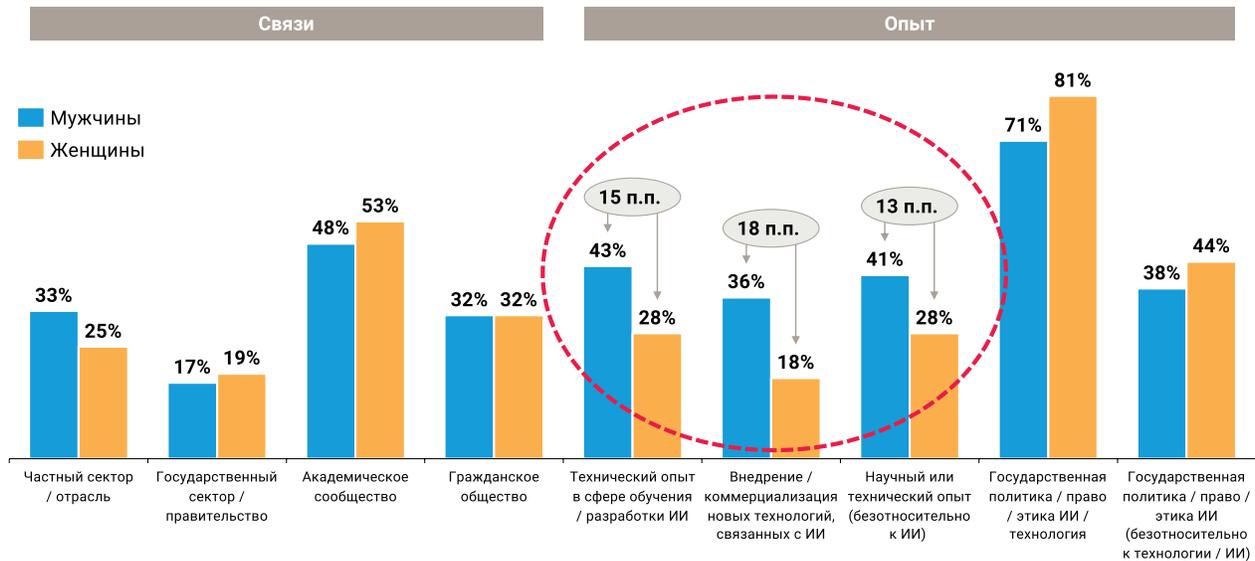
\* 43 респондента (12 процентов) указали, что являются гражданами более чем одного государства. Если страна проживания респондента совпала с одной из стран его гражданской принадлежности, для целей анализа использовалась эта страна гражданской принадлежности (34 из 43). В остальных случаях использовалась наименее представленная гражданская принадлежность (9 из 43). Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

## между характеристиками мужчин и женщин, участвовавших в опросе, имеются определенные различия

О наличии технического опыта сообщает более значительное число мужчин; о связях со сферами управления, политики, права/этики сообщает более значительное количество женщин

Процент респондентов, сообщивших о наличии связей / опыта в разбивке по гендерному признаку \* (n = 348)

ПО ГЕНДЕРНОМУ ПРИЗНАКУ



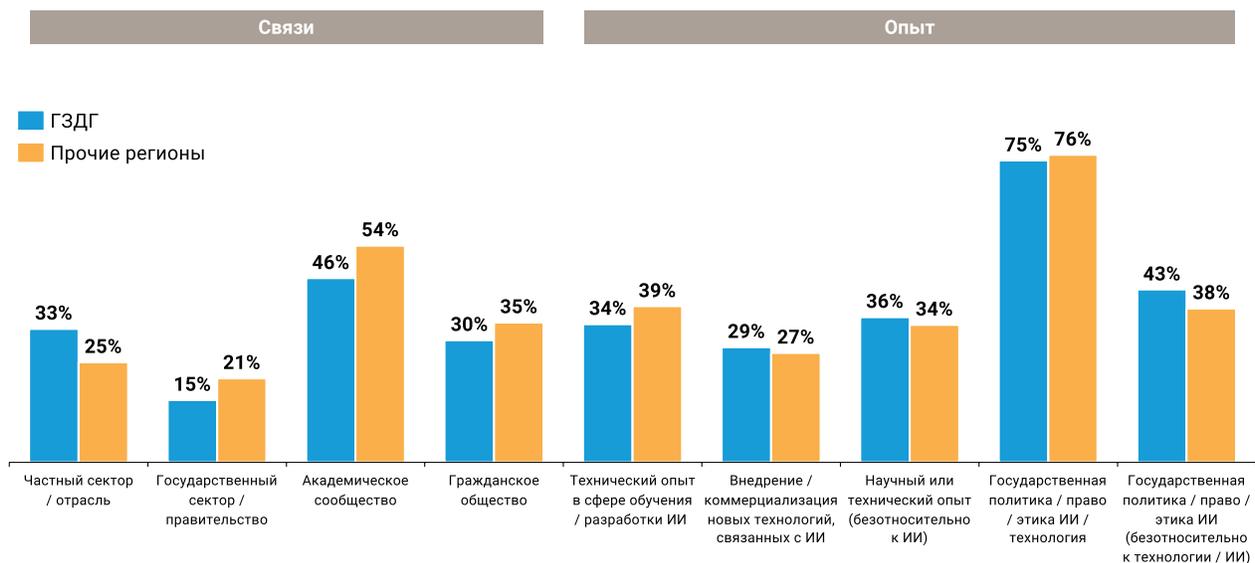
Примечание: п.п. = процентные пункты  
Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

## характеристики респондентов, представляющих страны ГЗДГ и страны, не входящие в ГЗДГ, являются достаточно схожими

Большее число респондентов из стран, не входящих в ГЗДГ, имеют отношение к государственному сектору или академическому сообществу, нежели к частному сектору или отрасли

Процент респондентов, сообщивших о наличии связей / опыта в разбивке по регионам гражданской принадлежности\* (n = 348)

ПО РЕГИОНАМ



\* 43 респондента (12 процентов) указали, что являются гражданами более чем одного государства. Если страна проживания респондента совпадала с одной из стран его гражданской принадлежности, для целей анализа использовалась эта страна гражданской принадлежности (34 из 43). В остальных случаях использовалась наименее представленная гражданская принадлежность (9 из 43). Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

# Мнения относительно ускорения развития ИИ

## 74 процента респондентов ожидают ускорения технических изменений

Среди тех, кто ожидает ускорения, процент респондентов, не являющихся гражданами стран ГЗДГ, превышает процент респондентов из стран ГЗДГ

«Ожидаете ли вы, что в предстоящие 18 месяцев темпы технических изменений в сфере ИИ (включая, например, разработку / внедрение новых моделей) по сравнению с предшествующими 3 месяцами...» (n=348)

### ПО РЕГИОНАМ И ГЕНДЕРНОМУ ПРИЗНАКУ



Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

## 89 процентов респондентов ожидают ускорения темпов внедрения и применения ИИ

Существенного ускорения ожидают несколько больше респондентов из стран, не относящихся к ГЗДГ (это особенно характерно для мужчин)

«Ожидаете ли вы, что в предстоящие 18 месяцев темпы внедрения и применения ИИ (включая, например, новые виды использования ИИ коммерческими компаниями / государственными ведомствами) по сравнению с предшествующими 3 месяцами...» (n=348)

### ПО РЕГИОНАМ И ГЕНДЕРНОМУ ПРИЗНАКУ



Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

# Ограниченное влияние технического опыта (обучение / разработка ИИ)

Респонденты настроены несколько более пессимистично в отношении технических изменений и несколько более оптимистично в отношении внедрения и применения

## ПО НАЛИЧИЮ ОПЫТА

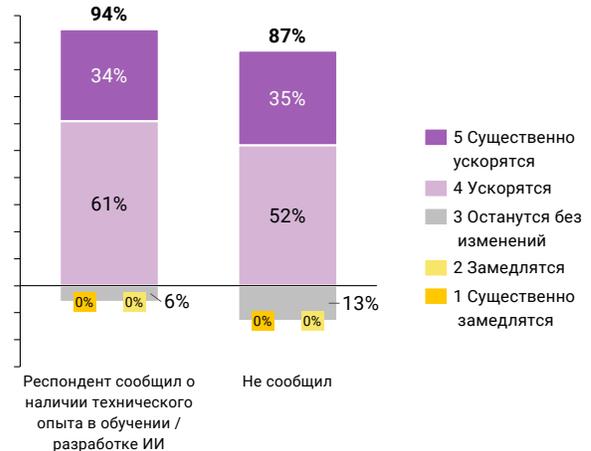
### Технические изменения

«Ожидаете ли вы, что в предстоящие 18 месяцев по сравнению с предшествующими тремя месяцами темпы технических изменений в сфере ИИ (включая, например, разработку / выпуск новых моделей)…» (n=348)



### Внедрение и применение

«Ожидаете ли вы, что в предстоящие 18 месяцев по сравнению с предшествующими тремя месяцами темпы внедрения и применения ИИ (включая, например, новые виды использования ИИ в бизнесе / работе правительств)…» (n=348)



Примечание: общая сумма может отличаться от 100 процентов по причине округления, исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

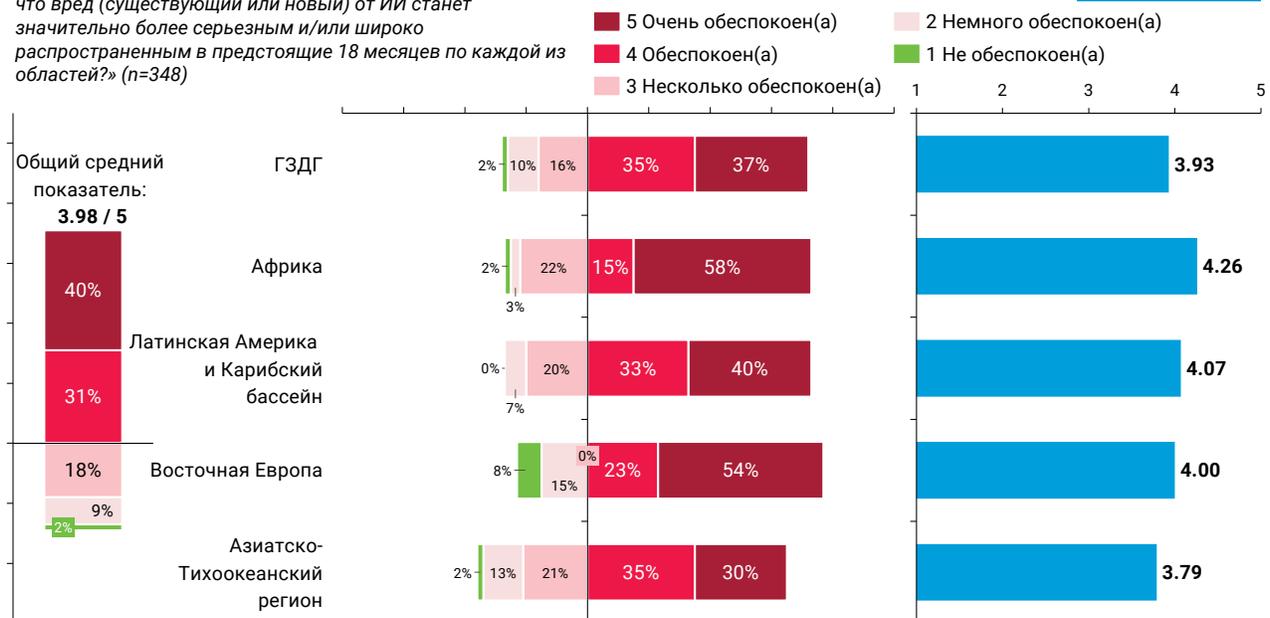
## Мнения относительно рисков, связанных с возможным вредом от ИИ, в течение предстоящих 18 месяцев (с мая 2024 года)

### 71 процент респондентов обеспокоены / очень обеспокоены возможным вредом от ИИ в предстоящие 18 месяцев

Уровень обеспокоенности респондентов из стран Африки выше, нежели у других респондентов; респонденты из Азиатско-Тихоокеанского региона обеспокоены в меньшей степени, нежели респонденты из ГЗДГ.

«Каков ваш текущий уровень обеспокоенности в связи с тем, что вред (существующий или новый) от ИИ станет значительно более серьезным и/или широко распространенным в предстоящие 18 месяцев по каждой из областей?» (n=348)

## ПО РЕГИОНАМ



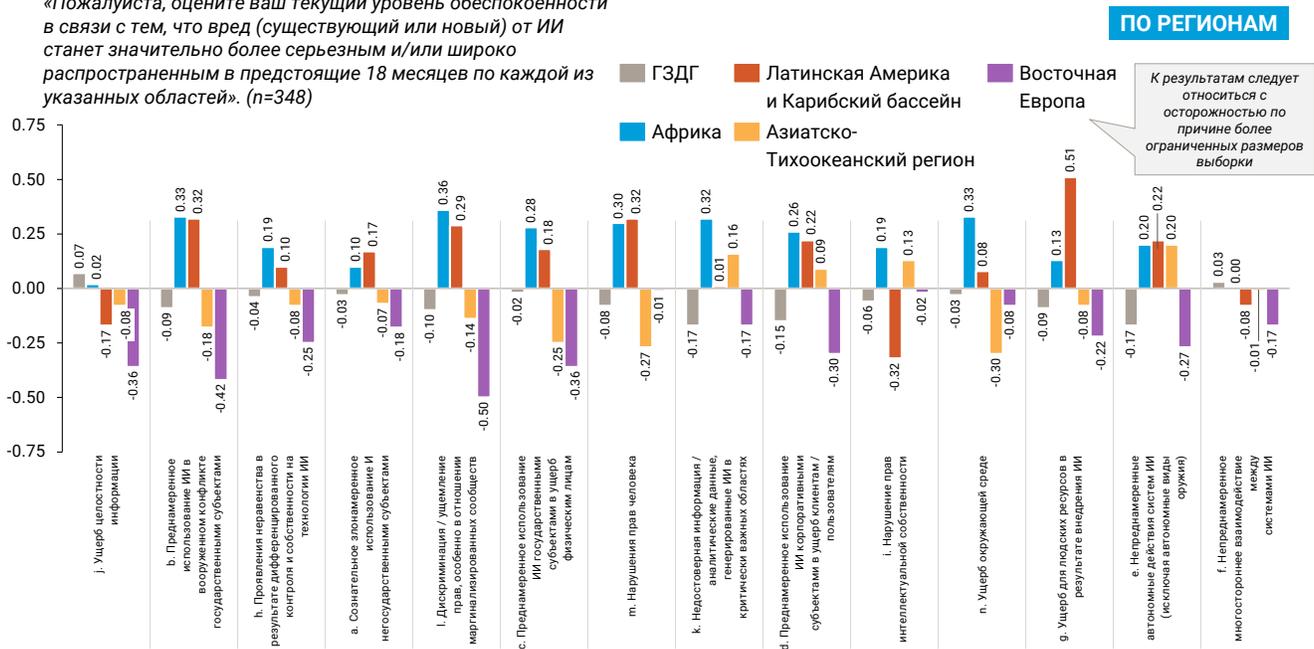
Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.



# По многим областям наибольшую обеспокоенность проявляют респонденты из Африки и из Латинской Америки и Карибского бассейна

Особенно серьезную обеспокоенность вызывает возможное использование ИИ государствами в вооруженных конфликтах, способное породить дискриминацию или нарушения прав человека

«Пожалуйста, оцените ваш текущий уровень обеспокоенности в связи с тем, что вред (существующий или новый) от ИИ станет значительно более серьезным и/или широко распространенным в предстоящие 18 месяцев по каждой из указанных областей». (n=348)



Разница между совокупным показателем (для всех регионов) и показателем соответствующего региона, которые рассчитаны на основе следующих рейтингов: 1 = Не обеспокоен(а) 2 = Немного обеспокоен(а) 3 = Несколько обеспокоен(а) 4 = Обеспокоен(а) 5 = Очень обеспокоен(а). Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

## 71 процент респондентов обеспокоены / очень обеспокоены по поводу возможного вреда от ИИ в предстоящие 18 месяцев

Женщины проявляют большую обеспокоенность, чем мужчины, особенно в странах ГЗДГ

«Каков ваш текущий уровень обеспокоенности в связи с тем, что вред (существующий или новый) от ИИ станет значительно более серьезным и/или широко распространенным в предстоящие 18 месяцев по каждой из областей?» (n=348)

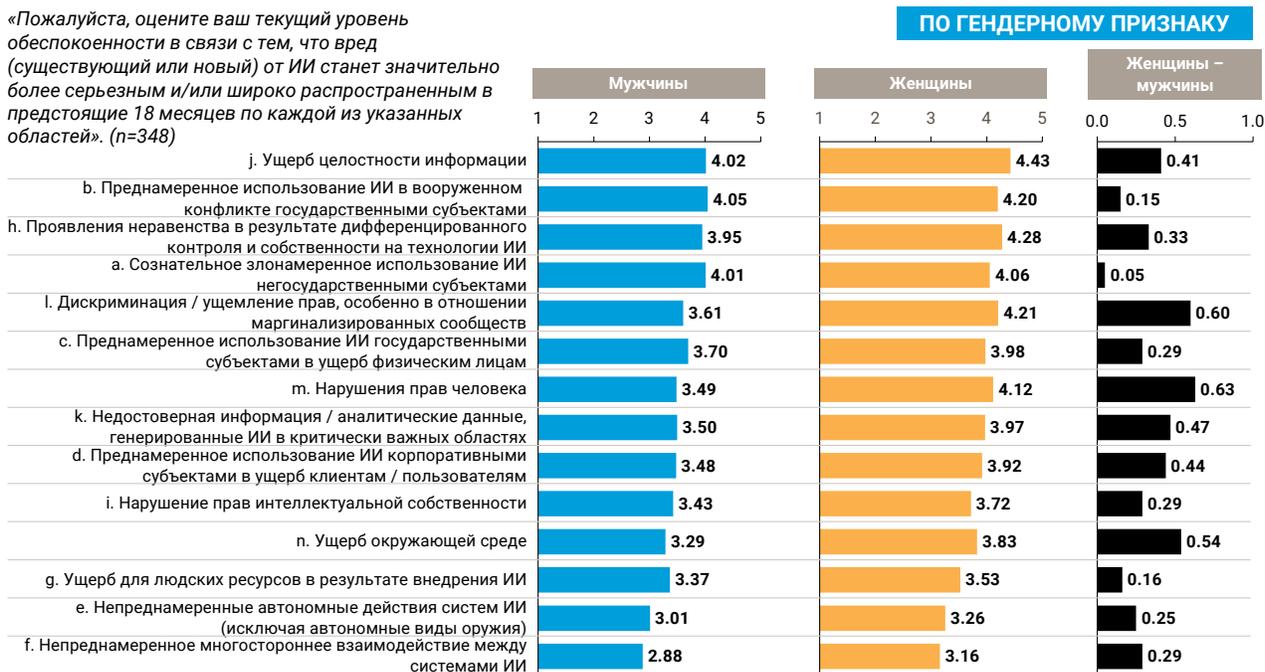


Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

# Женщины более обеспокоены, чем мужчины, по всем указанным областям

Разрыв особенно ощутим в отношении таких областей, как нарушения права человека, дискриминация и окружающая среда.

«Пожалуйста, оцените ваш текущий уровень обеспокоенности в связи с тем, что вред (существующий или новый) от ИИ станет значительно более серьезным и/или широко распространенным в предстоящие 18 месяцев по каждой из указанных областей». (n=348)

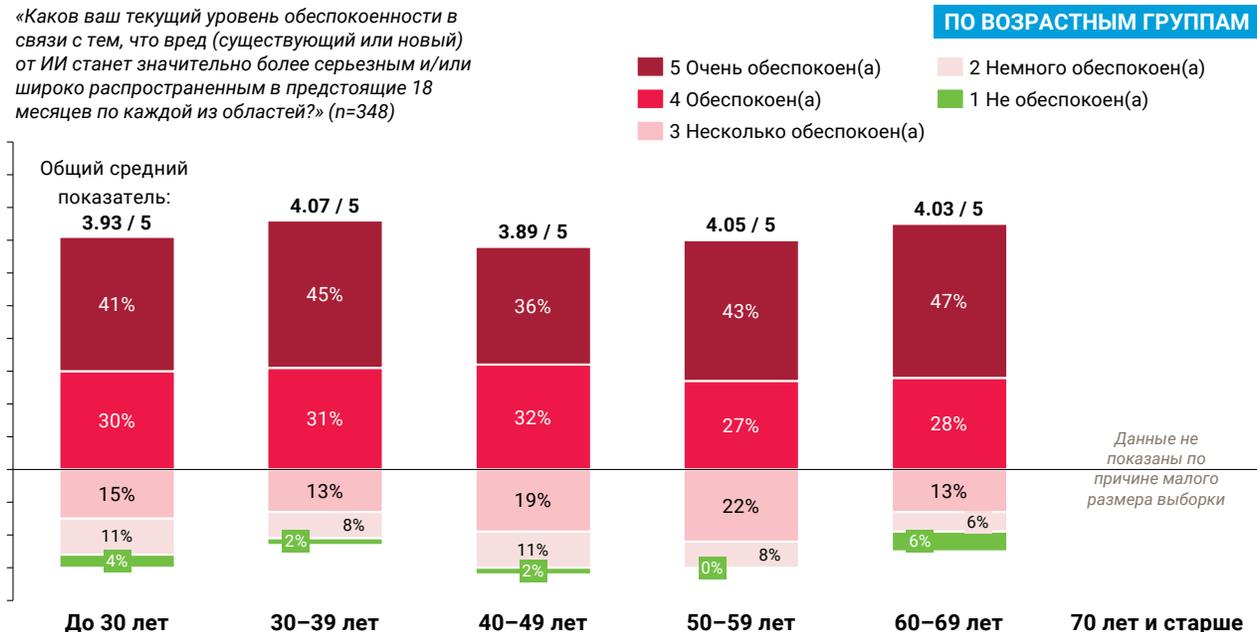


Показано: Средний показатель, рассчитанный на основе следующих рейтингов: 1 = Не обеспокоен(а) 2 = Немного обеспокоен(а) 3 = Несколько обеспокоен(а) 4 = Обеспокоен(а) 5 = Очень обеспокоен(а). Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

## 71 процентов респондентов обеспокоены / очень обеспокоены по поводу возможного вреда от ИИ в предстоящие 18 месяцев

Наблюдаются сравнительно небольшие различия в уровнях обеспокоенности в зависимости от возраста респондентов

«Каков ваш текущий уровень обеспокоенности в связи с тем, что вред (существующий или новый) от ИИ станет значительно более серьезным и/или широко распространенным в предстоящие 18 месяцев по каждой из областей?» (n=348)



Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

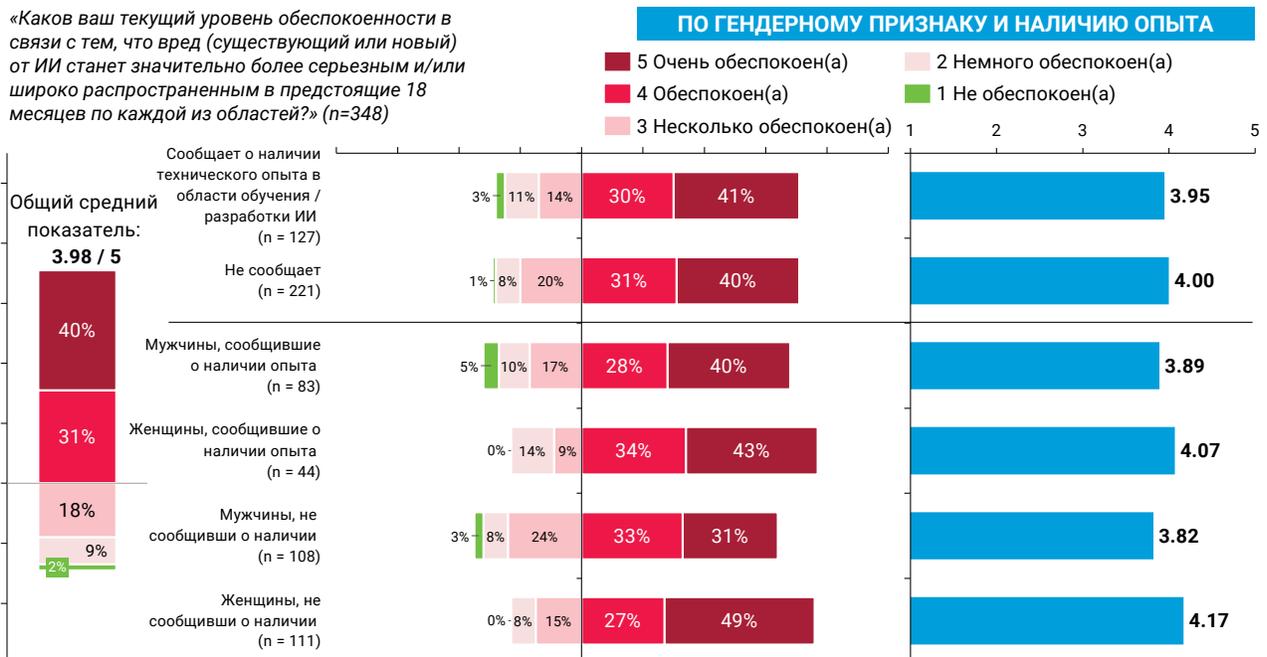
## Респонденты, сообщившие о наличии технического опыта (в сфере обучения / разработки ИИ), проявляют меньшую обеспокоенность в отношении большинства указанных областей



Показано: Разница между совокупным показателем (для всех регионов) и показателем соответствующего региона, которые рассчитаны на основе следующих рейтингов: 1 = Не обеспокоен(а) 2 = Немного обеспокоен(а) 3 = Несколько обеспокоен(а) 4 = Обеспокоен(а) 5 = Очень обеспокоен(а) *Примечание:* исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

## Технический опыт (в области обучения / разработки ИИ) имеет ограниченное влияние

Для мужчин характерна меньшая обеспокоенность, чем для женщин, вне зависимости от того, сообщили ли они о наличии соответствующего опыта.



Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

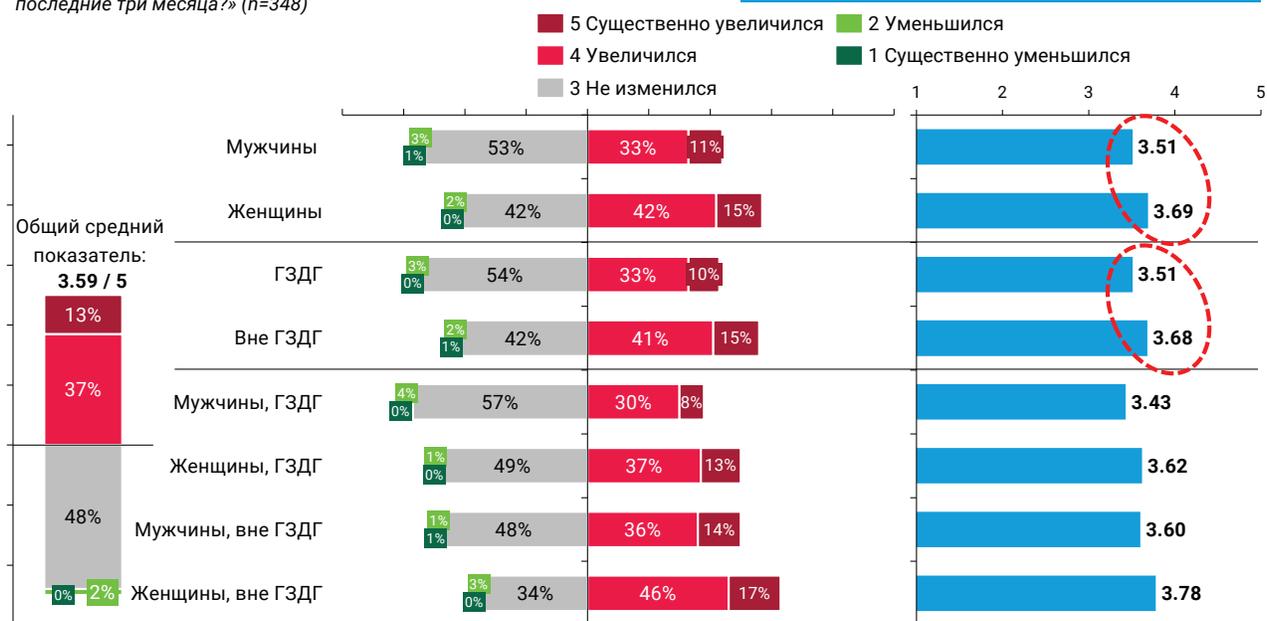
## Изменения в восприятии уровня обеспокоенности относительно рисков, связанных с возможным вредом от ИИ, за последние три месяца

### У 50 процентов респондентов уровень обеспокоенности за последние три месяца вырос; у 48 процентов – остался без изменений

Почти ни у кого из респондентов уровень обеспокоенности не снизился; более высокий уровень обеспокоенности характерен для женщин и для респондентов из стран, не входящих в ГЗДГ

«Увеличился ли ваш уровень обеспокоенности за последние три месяца?» (n=348)

ПО РЕГИОНАМ И ГЕНДЕРНОМУ ПРИЗНАКУ



Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Источник: Карта рисков ИИ, КПВТ, 13-25 Мая 2024 г.

## Приложение F: ответы на вопросы, касающиеся возможностей

По просьбе Консультативного органа высокого уровня по искусственному интеллекту Канцелярия Посланника Генерального секретаря по вопросам технологий (КПГСТ) провела глобальное исследование возможностей ИИ. Экспертам было предложено высказать свое мнение в личном качестве (не от имени своего учреждения или работодателя). Опрос был разделен на разделы, посвященные возможностям в странах с высоким уровнем дохода/уровнем дохода выше среднего и в странах с уровнем дохода ниже среднего/низким уровнем дохода, причем на эти вопросы отвечали только респонденты, сообщившие о наличии специальных знаний о ситуации в странах с уровнем дохода ниже среднего/низким уровнем дохода. В ходе опроса задавались вопросы только о возможных положительных последствиях ИИ.

Респондентам было предложено ответить на вопрос, насколько им известны конкретные примеры того, как ИИ повышает экономическую активность, ускоряет научные открытия и способствует прогрессу в достижении отдельных ЦУР<sup>1</sup>. Респондентов попросили предоставить подробную информацию, включая тематические исследования, названия организаций, данные и ссылки на соответствующие статьи/публикации/документы. Затем респондентов спросили, какого прогресса они ожидают в ближайшие три года по тем же направлениям.

В качестве дополнительного респондентам был задан вопрос о том, когда они ожидают значительного влияния ИИ по этим параметрам с 50-процентной уверенностью/вероятностью. Были заданы и другие вопросы, в том числе о том, какие субъекты участвуют в использовании определенных возможностей, какие барьеры способствуют разрыву между странами в области ИИ, сталкиваются ли конкретные группы с дополнительными ограничениями при использовании возможностей ИИ и как их можно устранить.

Опрос проводился с 9 по 21 августа 2024 года, а список респондентов был составлен на основе сетей КПГСТ и Консультативного органа, включая участников «глубоких погружений» Консультативного органа. Кроме того, для проведения опроса активно использовались как совещание Международного союза электросвязи «ИИ во благо», так и сети Конференции Организации Объединенных Наций по торговле и развитию. Всего для участия в опросе было приглашено более 1000 человек. На опрос ответили более 120 респондентов, представив богатый и разнообразный диапазон мнений (в том числе в разбивке по регионам и гендерному признаку) относительно возможностей ИИ.

---

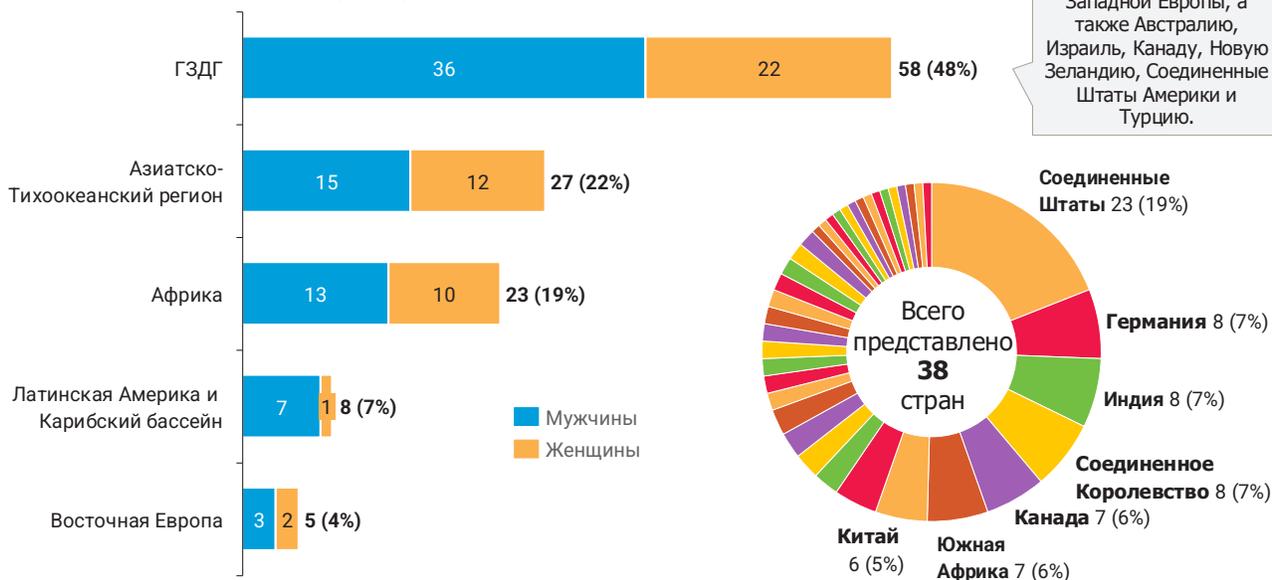
1 Отдельные вопросы в отношении ЦУР 8 («Достойная работа и экономический рост») и ЦУР 9 («Инновации, индустриализация и инфраструктура») не ставились, поскольку они тесно связаны с повышением экономической активности. Конкретно не упоминалась и ЦУР 17 («Партнерство в интересах достижения целей»).

## Обзор выборки

### Региональное представительство: активное участие респондентов со всего мира

Выборка позволяет сравнивать ответы респондентов из стран Группы западноевропейских и других государств (ГЗДГ) и из других регионов

Респонденты в разбивке по регионам и гражданской принадлежности\* (n = 121)



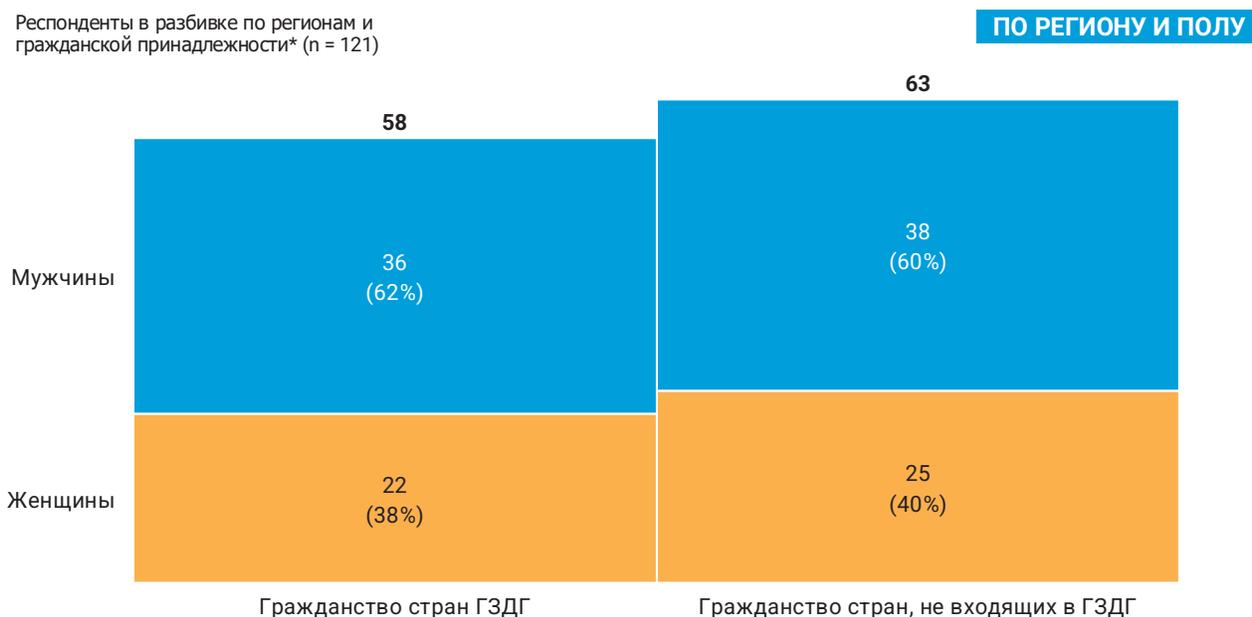
\*Девять респондентов (7 процентов) указали, что являются гражданами более чем одного государства. Если страна проживания респондента совпадала с одной из стран его гражданской принадлежности, для целей анализа использовалась эта страна гражданской принадлежности (8 из 9). В оставшемся случае использовалась наименее представленная гражданская принадлежность (1 из 9).

Источник: Исследование возможностей, связанных с ИИ, КПВТ, 9-21 августа 2024 г.

### на мужчин приходится примерно 60 процентов респондентов как в странах ГЗДГ, так и в других странах

Соответствие означает, что одномерный анализ по полу и региону не подвержен немедленной контаминации

Респонденты в разбивке по регионам и гражданской принадлежности\* (n = 121)



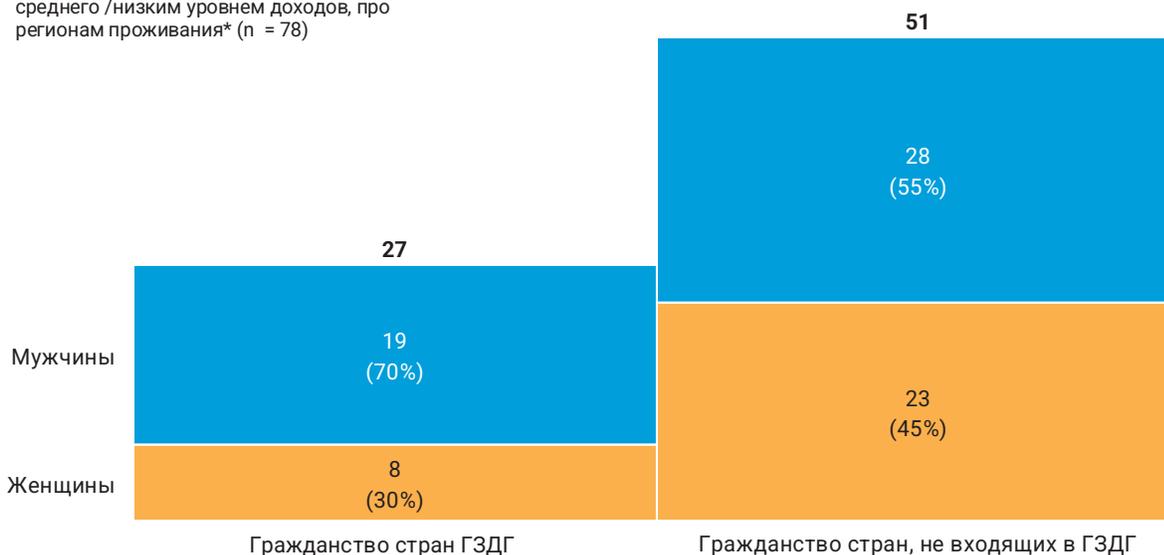
\*Девять респондентов (7 процентов) указали, что являются гражданами более чем одного государства. Если страна проживания респондента совпадала с одной из стран его гражданской принадлежности, для целей анализа использовалась эта страна гражданской принадлежности (8 из 9). В оставшемся случае использовалась наименее представленная гражданская принадлежность (1 из 9).

Источник: Исследование возможностей, связанных с ИИ, КПВТ, 9-21 августа 2024 г.

## группа респондентов, сообщивших о знании ситуации в развивающихся странах, является менее сбалансированной

Респонденты, сообщающие о конкретных знаниях о странах с уровнем дохода ниже среднего /низким уровнем доходов, про регионам проживания\* (n = 78)

ПО РЕГИОНУ И ПОЛУ



\*Девять респондентов (7 процентов) указали, что являются гражданами более чем одного государства. Если страна проживания респондента совпала с одной из стран его гражданской принадлежности, для целей анализа использовалась эта страна гражданской принадлежности (8 из 9). В оставшемся случае использовалась наименее представленная гражданская принадлежность (1 из 9). Вопросы о странах с низким доходом/уровнем дохода ниже среднего - были опрошены только респонденты с соответствующими знаниями.  
Источник: Исследование возможностей, связанных с ИИ, КПВТ, 9-21 августа 2024 г.

## Представления относительно той положительной роли, которую сыграл ИИ до настоящего времени

### К настоящему времени отмечено положительное влияние на рост и науку, тогда как в отношении большинства ЦУР оно проявляется в меньшей степени

Зафиксированные к настоящему времени последствия в странах с высоким доходом/доходом выше среднего

"Насколько много вам известно о конкретных примерах того, как ИИ сейчас или в недавнем прошлом внес прямой вклад в... в странах с высоким доходом/доходом выше среднего?"

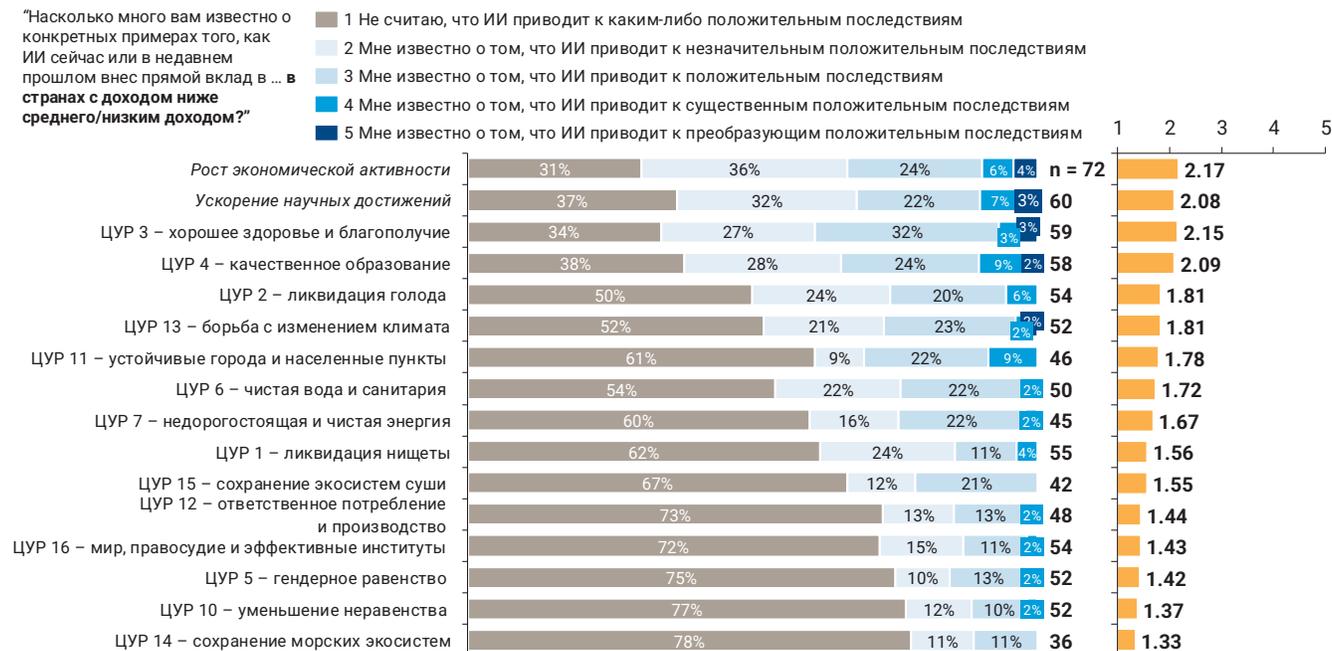
- 1 Не считаю, что ИИ приводит к каким-либо положительным последствиям
- 2 Мне известно о том, что ИИ приводит к незначительным положительным последствиям
- 3 Мне известно о том, что ИИ приводит к положительным последствиям
- 4 Мне известно о том, что ИИ приводит к существенным положительным последствиям
- 5 Мне известно о том, что ИИ приводит к преобразующим положительным последствиям



Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Вопросы в отношении ЦУР 8, 9 и 17 не задавались.  
Источник: Исследование возможностей, связанных с ИИ, КПВТ, 9-21 августа 2024 г.

## В отношении стран с более низкими доходами сообщается о менее ощутимых последствиях по всему спектру направлений

Зафиксированные к настоящему времени последствия в странах с доходом ниже среднего/низким доходом



Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Конкретные величины n указаны в отношении каждого вопроса. Вопросы в отношении ЦУР 8, 9 и 17 не задавались. Источник: Исследование возможностей, связанных с ИИ, КПВТ, 9-21 августа 2024 г.

## В отношении стран с более низкими доходами сообщается о менее ощутимых последствиях по всему спектру направлений

Наибольший разрыв наблюдается в отношении экономического роста и науки



Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Конкретные величины n указаны в отношении каждого вопроса. Вопросы в отношении ЦУР 8, 9 и 17 не задавались. Источник: Исследование возможностей, связанных с ИИ, КПВТ, 9-21 августа 2024 г.

# Представления относительно ожидаемых положительных последствий ИИ в предстоящие три года

## Ожидаются последствия для роста, науки и образования; в других областях ожидания не столь оптимистичны

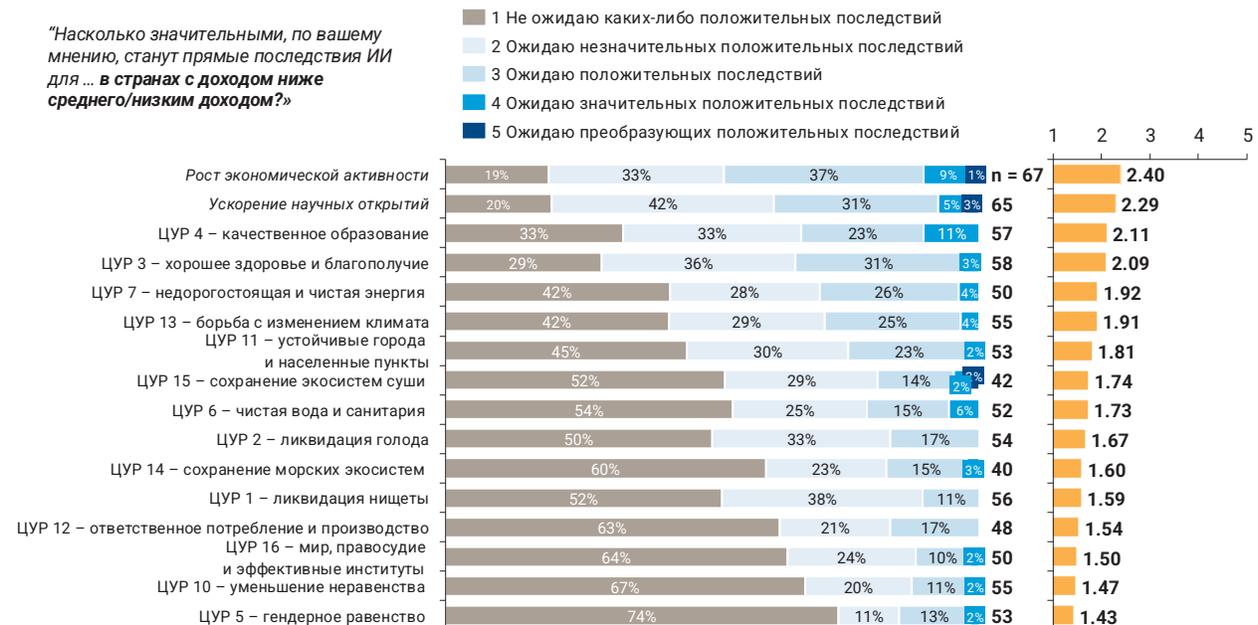
Ожидаемые в предстоящие три года последствия в странах с высоким доходом/доходом выше среднего



Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Вопросы в отношении ЦУР 8, 9 и 17 не задавались. Источник: Исследование возможностей, связанных с ИИ, КПВТ, 9-21 августа 2024 г.

## Ожидаются определенные последствия в странах с более низким доходом, но и в данном случае они более ограничены по своим масштабам

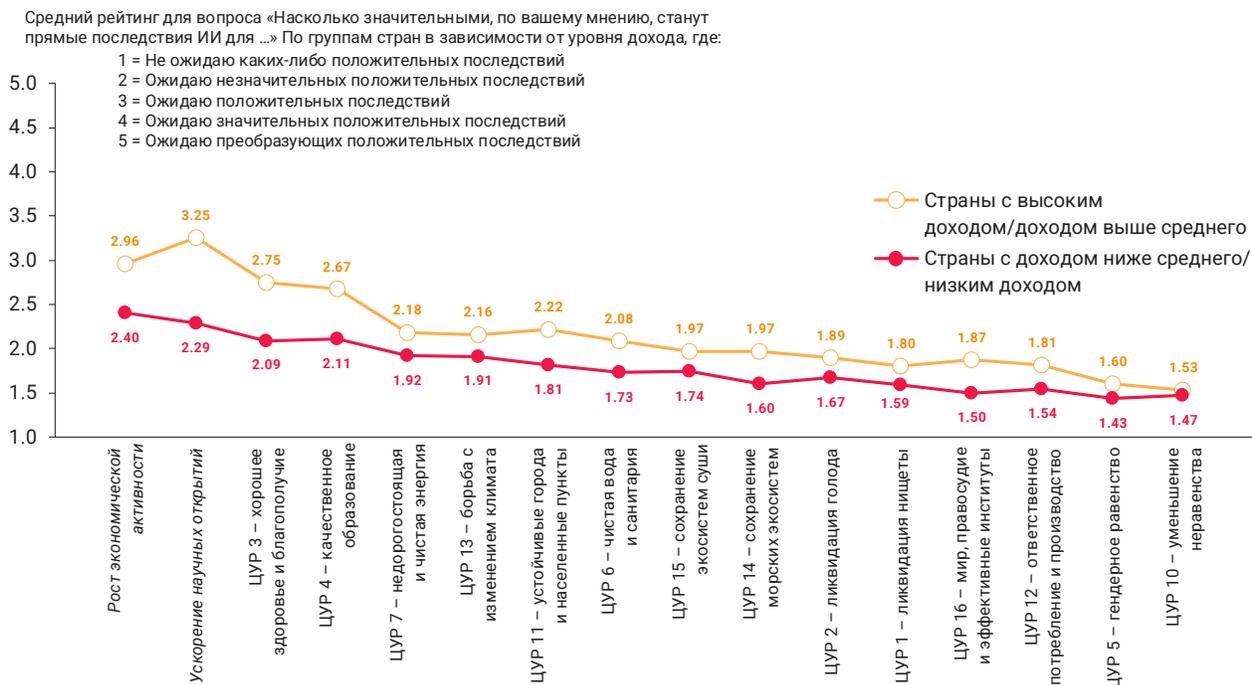
Ожидаемые в предстоящие три года последствия в странах с доходом ниже среднего/низким доходом



Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Конкретные величины n указаны в отношении каждого вопроса. Вопросы в отношении ЦУР 8, 9 и 17 не задавались. Источник: Исследование возможностей, связанных с ИИ, КПВТ, 9-21 августа 2024 г.

## В странах с более низким доходом ожидаются менее значительные последствия по всем направлениям

Разрыв наиболее очевиден по таким направлениям, как экономический рост, наука, здравоохранение и образование



Примечание: исключая ответы «Не знаю» / «Затрудняюсь ответить» и незаполненные ответы. Конкретные величины указаны в отношении каждого вопроса. Вопросы в отношении ЦУР 8, 9 и 17 не задавались. Источник: Исследование возможностей, связанных с ИИ, КПВТ, 9-21 августа 2024 г. Диаграммы, подготовленные с помощью think-cell.

## Приложение G: список сокращений

|                |   |
|----------------|---|
| <b>АСВТ</b>    | Ассоциация специалистов по вычислительной технике                                 |
| <b>ГАГ</b>     | Группа африканских государств   |
| <b>ИИ</b>      | искусственный интеллект   |
| <b>АНИС</b>    | Американский национальный институт стандартизации                                 |
| <b>ГГАТО</b>   | Группа государств Азии и Тихого океана  |
| <b>АСЕАН</b>   | Ассоциация государств Юго-Восточной Азии  |
| <b>БИС</b>     | Британский институт стандартов  |
| <b>ЕКС</b>     | Европейский комитет по стандартизации   |
| <b>ЕКСЭ</b>    | Европейский комитет по стандартизации в электротехнике                            |
| <b>ЦЕРН</b>    | Европейская организация ядерных исследований                                      |
| <b>ГГВЕ</b>    | Группа государств Восточной Европы  |
| <b>ЭТСИ</b>    | Европейский институт по стандартизации в области электросвязи                     |
| <b>ФАО</b>     | Продовольственная и сельскохозяйственная организация Объединенных Наций           |
| <b>ГРФМ</b>    | Группа разработки финансовых мер  |
| <b>ИИМВ</b>    | иностранный информационный манипулирование и вмешательство                        |
| <b>ФПМ</b>     | Форум по передовым моделям  |
| <b>СФС</b>     | Совет по финансовой стабильности  |
| <b>Г-20</b>    | Группа 20   |
| <b>Г-7</b>     | Группа семи   |
| <b>ГП</b>      | графический процессор   |
| <b>МАГАТЭ</b>  | Международное агентство по атомной энергии  |
| <b>ИКАО</b>    | Международная организация гражданской авиации                                     |
| <b>МЭК</b>     | Международная электротехническая комиссия   |
| <b>ИИЭЭ</b>    | Институт инженеров по электронике и электротехнике                                |
| <b>МОТ</b>     | Международная организация труда   |
| <b>ИМО</b>     | Международная морская организация   |
| <b>МГЭИК</b>   | Межправительственная группа экспертов по изменению климата                        |
| <b>ИСО</b>     | Международная организация стандартизации  |
| <b>МСЭ</b>     | Международный союз электросвязи   |
| <b>ЛАКБ</b>    | Латинская Америка и Карибский бассейн   |
| <b>НИСТ</b>    | Национальный институт стандартов и технологий (Соединенные Штаты)                 |
| <b>ОЭСР</b>    | Организация экономического сотрудничества и развития                              |
| <b>УВКПЧ</b>   | Управление Верховного комиссара Организации Объединенных Наций по правам человека |
| <b>КПГСТ</b>   | Канцелярия Посланника Генерального секретаря по вопросам технологий               |
| <b>УКС</b>     | Управление Китая по стандартизации  |
| <b>ЦУР</b>     | цель в области устойчивого развития   |
| <b>ЮНКТАД</b>  | Конференция Организации Объединенных Наций по торговле и развитию                 |
| <b>ПРООН</b>   | Программа развития Организации Объединенных Наций                                 |
| <b>ЮНЕСКО</b>  | Организация Объединенных Наций по вопросам образования, науки и культуры          |
| <b>УВКБ</b>    | Управление Верховного комиссара Организации Объединенных Наций по делам беженцев  |
| <b>КТУ ООН</b> | Контртеррористическое управление  |
| <b>ГЗДГ</b>    | Группа западноевропейских и других государств                                     |
| <b>ВОЗ</b>     | Всемирная организация здравоохранения   |
| <b>ВОИС</b>    | Всемирная организация интеллектуальной собственности                              |
| <b>ВСС</b>     | Всемирное сотрудничество по стандартам  |

## **Доноры**

Агентство с благодарностью признает финансовые и натуральные взносы следующих правительств и партнеров, без которых оно не смогло бы выполнять свои обязанности:

Правительство Чешской Республики  
Правительство Чешской Республики  
Правительство Финляндии  
Правительство Германии  
Правительство Италии  
Правительство Японии  
Правительство Королевства Нидерландов  
Правительство Королевства Саудовская Аравия  
Правительство Сингапура  
Правительство Швейцарии  
Правительство Объединенных Арабских Эмиратов  
Правительство Соединенного Королевства Великобритании и Северной Ирландии  
Фонд сети Омидьяра  
Международная организация франкофонии

## **Секретариат**

Саймон Честерман  
Квинтин Чу-Ламбер  
Элеонора Фурнье-Томбс  
Себастьян Франк  
Дэвид Майкл Келли  
Брайан Шунг Сын Лау  
Эндрю Морритт  
Филиппо Пьероцци  
Мехди Снен  
Изабель де Сола  
Люсия Веласко  
Ребекка Хаюнг Ву



**United  
Nations**



**AI  
Advisory  
Body**