



Организация  
Объединенных Наций по  
вопросам образования,  
науки и культуры

# Исполнительный совет 206 EX/42

Двести шестая сессия

ПАРИЖ, 27 марта 2019 г.  
Оригинал: английский

Пункт 42 предварительной повестки дня

## Предварительное исследование технических и правовых аспектов целесообразности принятия нормативного документа по этике искусственного интеллекта

### РЕЗЮМЕ

Расширенная рабочая группа по этике искусственного интеллекта (ИИ) Всемирной комиссии по этике научных знаний и технологий (КОМЕСТ) подготовила предварительное исследование по этике искусственного интеллекта, которое включает адресованную ЮНЕСКО рекомендацию изучить возможность принятия нормативного документа в этой области.

В соответствии с Правилами процедуры, касающимися рекомендаций государствам-членам и международных конвенций, подпадающих под действие пункта 4 статьи IV Устава, Генеральный директор представляет Исполнительному совету результаты этого исследования, включая содержащийся в приложении полный текст исследования расширенной рабочей группы КОМЕСТ. Это позволит Исполнительному совету сформулировать соответствующие замечания и принять решение с целью включения данного пункта в предварительную повестку дня 40-й сессии Генеральной конференции. Расходы, связанные с финансовыми и административными последствиями, будут по мере необходимости покрываться за счет добровольных взносов.

Ожидаемые меры со стороны Исполнительного совета: предлагаемое решение в пункте 18.



Job: 1902517

## I. ЮНЕСКО – ЛАБОРАТОРИЯ ИДЕЙ

1. Искусственный интеллект<sup>1</sup> (ИИ) является одним из ключевых элементов эпохи конвергенции технологий, который сопряжен с глубокими последствиями для человека, культуры, общества и окружающей среды. Вероятно, ИИ изменит будущее образования, науки, культуры и коммуникации, то есть всех тех областей, которые включены в мандат ЮНЕСКО.

2. Хотя ИИ и обладает потенциалом для того, чтобы изменить будущее человечества в лучшую сторону и в интересах устойчивого развития, всё больше людей осознают связанные с ИИ риски и проблемы, особенно в плане усугубления существующего неравенства и диспропорций, а также последствия его использования для сферы прав человека. Для того чтобы наметить возможные сценарии и задействовать потенциал ИИ для реализации возможностей в сфере развития при сохранении контроля над рисками, важно выработать более всестороннее понимание того, как общество меняется под воздействием таких революционных технологий, как ИИ.

3. Такая работа должна сопровождаться этическим анализом ввиду того, что технологии ИИ не являются нейтральными, а отличаются внутренне присущей предвзятостью, обусловленной данными, которые используются в ходе их обучения, и решениями, которые принимаются в процессе обучения с применением этих данных, а также ввиду того, что решения ИИ не могут быть в полной мере предсказуемы. Кроме того, поскольку ИИ представляет собой распределенную технологию, управление которой сегодня на практике осуществляется многочисленными учреждениями, организациями и компаниями, для анализа вопросов ответственного управления ИИ необходим плюралистический, междисциплинарный, поликультурный и многосторонний подход, который бы стимулировал обсуждение вопросов о том, к какому будущему для человечества мы стремимся. В ходе этого анализа необходимо рассмотреть основные проблемы в области разработки технологий ИИ, касающиеся заложенных в алгоритмах предубеждений, включая гендерные предрассудки, защиты неприкосновенности частной жизни и персональных данных, рисков создания новых форм социальной изоляции и неравенства, а также вопросов справедливого распределения выгод и рисков, подконтрольности, ответственности, влияния на занятость, эволюции труда, человеческого достоинства, прав человека, безопасности и рисков двойного применения.

4. В стремлении использовать потенциал ИИ на благо человечества ЮНЕСКО провела в Париже в 2018 г. ряд совещаний, включая [семинар по искусственному интеллекту](#), совместно организованный постоянными представительствами Бразилии и Турции при ЮНЕСКО (июнь 2018 г.), открытое обсуждение на тему [«Мобилизация искусственного интеллекта для укрепления основанных на знаниях обществ и ответственного управления»](#), состоявшееся в фонде Mozilla Foundation (ноябрь 2018 г.), и дискуссию на тему [«ИИ на службе прав человека и ЦУР: содействие многосторонним, инклюзивным и открытым подходам»](#), состоявшуюся в рамках [Форума по управлению интернетом](#) (ноябрь 2018 г.). В сотрудничестве с партнерами ЮНЕСКО на местах в различных регионах было проведено несколько мероприятий, включая [Первый форум по ИИ и интернету вещей в «умных» устойчивых городах](#) (Буэнос-Айрес, Аргентина, май 2018 г.). Еще одним проявлением заинтересованности ЮНЕСКО в анализе глобального диалога по этике ИИ, нацеленного на нормы и стандарты, стал прошедший в декабре 2018 г. в Бен-Герире, Марокко, [Форум по ИИ в Африке](#), участники которого, в частности, заявили о необходимости разработки стратегии ИИ для Африки и создания систем ИИ, ориентированных на человеческое измерение. Общий план действий ЮНЕСКО в сфере ИИ был представлен государствам-членам на [информационном совещании](#) 22 января 2019 г., за которым последовала дискуссия с участием экспертов из разных стран на тему [«Будущее технологий: надежда или страх?»](#). Этот анализ получил дальнейшее развитие 4 марта 2019 г. в Штаб-квартире ЮНЕСКО, где прошла глобальная конференция [«Принципы искусственного](#)

<sup>1</sup> Хотя единого определения «искусственного интеллекта» (ИИ) не существует, речь, как правило, идет о машинах, способных имитировать определенные функции человеческого интеллекта, включая такие его свойства, как восприятие, обучаемость и логическое мышление, а также способности к решению задач, языковому взаимодействию и даже художественному творчеству.

[интеллекта: на пути к гуманистическому подходу?»](#), а также в ходе организованной ЮНЕСКО [Недели мобильного обучения](#) 2019 г. (4-8 марта), которая была посвящена ИИ и устойчивому развитию. В мае 2019 г. в Пекине, Китай, состоится [международная конференция по ИИ и образованию](#).

## II. ОПЫТ РАБОТЫ ЮНЕСКО В ОБЛАСТИ ЭТИКИ НАУКИ И ТЕХНОЛОГИЙ И В СФЕРЕ БИОЭТИКИ

5. В среднесрочной стратегии ЮНЕСКО на 2014-2021 гг. (37 С/4) подчеркивается, что установление норм и стандартов является одной из ключевых функций Организации. ЮНЕСКО обладает многолетним опытом нормотворческой деятельности на глобальном уровне и проводит анализ вопросов биоэтики и этики науки и технологий силами Всемирной комиссии по этике научных знаний и технологий (КОМЕСТ) и Международного комитета по биоэтике (МКБ). ЮНЕСКО играет ведущую роль в выработке международных стандартов и в сотрудничестве в данной сфере. В рамках этой деятельности вот уже 25 лет ведется диалог с участием многих заинтересованных сторон и разрабатываются нормативные рекомендации по этике науки и технологий, такие как Всеобщая декларация о геноме человека и правах человека (1997 г.), Международная декларация о генетических данных человека (2003 г.), Всеобщая декларация о биоэтике и правах человека (2005 г.), Декларация этических принципов в связи с изменением климата (2017 г.) и Рекомендация в отношении научной деятельности и научных работников (исследователей) (2017 г.). КОМЕСТ и МКБ также рассматривали связанные с ИИ вопросы. В 2017 г. МКБ утвердил доклад о больших массивах данных и здоровье, а КОМЕСТ – доклад об этике робототехники. Кроме того, в конце 2017 г. КОМЕСТ приступила к анализу этических аспектов интернета вещей.

6. Ввиду масштабности социальных последствий ИИ многие организации и правительства обеспокоены этическими аспектами его использования. В настоящее время разрабатываются национальные, региональные и другие стратегии и рамочные документы в сфере ИИ. Растет число докладов и руководящих документов по ИИ и этике, включая подготовленные Советом Европы, Европейской комиссией, Институтом инженеров по электротехнике и электронике (IEEE), Организацией экономического сотрудничества и развития (ОЭСР), Международным союзом электросвязи (МСЭ) и Всемирной организацией здравоохранения (ВОЗ), в которых определены важные принципы разработки, развития и внедрения ИИ. ЮНЕСКО внимательно следит за обсуждением этих документов, являясь, например, наблюдателем в группе экспертов высокого уровня по ИИ Европейской комиссии, членом группы экспертов по ИИ ОЭСР (AIGO), партнером Всемирного саммита МСЭ «Искусственный интеллект во благо» и участником других межправительственных форумов.

7. ЮНЕСКО может внести в эту дискуссию уникальный вклад, опираясь на имеющиеся у нее значительные сравнительные преимущества, которые обусловлены универсальностью ее членского состава и опытом междисциплинарной деятельности. В частности, Организация может предоставить поистине глобальную платформу для диалога по этике ИИ, которая позволит объединить усилия развитых и развивающихся стран, сблизить различающиеся в культурном и этическом отношении подходы и наладить взаимодействие между различными заинтересованными сторонами, представляющими как государственный, так и частный сектор. Поэтому ЮНЕСКО способна дополнить многочисленные руководящие принципы и указания по этике, которые в настоящее время разрабатываются государственными органами, компаниями и общественными организациями, собственным вкладом в развитие ИИ на благо всего человечества и во имя мира и устойчивого развития. Преследуя эту цель, ЮНЕСКО выступает в качестве связующего звена между государствами-членами, которые неоднократно подчеркивали свою поддержку планируемой деятельности ЮНЕСКО в области ИИ, гражданским обществом, техническим сообществом, научными кругами и частным сектором, в том числе предприятиями культурно-творческой сферы, опираясь на свой опыт проведения консультаций с участием многих заинтересованных сторон и формирования нейтрального консенсуса.

8. В развитие результатов работы КОМЕСТ Генеральный директор в августе 2018 г. обратилась к КОМЕСТ с просьбой подготовить предварительное исследование по этике ИИ, которое могло бы внести вклад в проводимый ЮНЕСКО в этой области анализ. Была создана расширенная рабочая группа КОМЕСТ по этике ИИ с участием трех внешних экспертов, которой было поручено подготовить предварительное исследование по этике искусственного интеллекта, полный текст которого приводится в приложении к настоящему документу.

9. В исследовании расширенной рабочей группы КОМЕСТ по этике ИИ рассматриваются этические вопросы, связанные с мандатом ЮНЕСКО. С точки зрения ЮНЕСКО, ИИ ставит множество вопросов, касающихся роли образования в обществе. Нынешние представления об «обучении на протяжении всей жизни», возможно, потребуют трансформации в модель непрерывного образования, предполагающую разработку новых видов академических квалификаций. В связи с ИИ возникает потребность в том, чтобы образование способствовало повышению уровня грамотности в области ИИ, критического мышления, адаптируемости к запросам рынка труда и подготовки инженеров в вопросах этики. В сфере естественных и общественных наук, а также наук о жизни и окружающей среде в контексте ИИ появляются фундаментальные вопросы, касающиеся наших представлений о научном понимании и объяснении. Это также влияет на то, каким образом мы применяем научные знания в социальном контексте. Системы ИИ требуют ответственного внедрения в научную практику и в процесс принятия решений, которые должны оцениваться и контролироваться человеком и не приводить к обострению структурного неравенства. Хотя ИИ и является мощным инструментом творчества, в связи с ним возникают важные вопросы о будущем искусства, о правах и вознаграждении творческих работников и о целостности цепочек создания художественных ценностей. Системы ИИ должны содействовать культурному разнообразию, инклюзивности и обогащению человеческого опыта, препятствуя углублению цифрового разрыва. В дополнение к этому следует поощрять многоязычный подход. Искусственный интеллект играет все более важную роль в обработке, структурировании и предоставлении информации, в связи с чем следует уделять внимание новым цифровым разрывам между странами и внутри различных социальных групп. Системы ИИ должны способствовать реализации свободы выражения мнений, обеспечению всеобщего доступа к информации, повышению качества журналистики и укреплению свободных, независимых и плюралистических средств массовой информации, не допуская при этом распространения дезинформации. Следует поощрять регулирование интернета с участием многих заинтересованных сторон.

10. В предварительном исследовании также рассматриваются глобальные в этическом отношении аспекты укрепления мира, культурного разнообразия, гендерного равенства и устойчивого развития в увязке с деятельностью ЮНЕСКО. В интересах поддержания мира ИИ может использоваться для получения информации о движущих силах конфликтов и ни при каких обстоятельствах не должен выполнять каких-либо действий вне контроля со стороны человека. Нарождающаяся цифровая экономика ставит перед африканскими и другими развивающимися странами важные социальные задачи и открывает перед ними широкие возможности. С этической точки зрения в целях развития экономики африканских стран ИИ должен быть интегрирован в национальную политику и стратегии развития с опорой на местные культурные традиции, ценности и знания. Следует избегать гендерной предвзятости при разработке алгоритмов, при формировании используемых для их обучения массивов данных и при применении алгоритмов в процессе принятия решений. Системы ИИ должны разрабатываться на устойчивой основе с учетом всего производственного цикла систем на базе ИИ и информационных технологий (ИТ). ИИ может использоваться для экологического мониторинга и управления рисками, а также для предотвращения и смягчения последствий экологических кризисов.

### **III. ЦЕЛЕСООБРАЗНОСТЬ ИЛИ НЕОБХОДИМОСТЬ НОРМАТИВНОГО ДОКУМЕНТА**

11. Как явствует из разворачивающихся сегодня дискуссий, на глобальном уровне существует потребность в общих универсальных этических принципах, отражающих ключевые

ценности, которые должны лежать в основе разработки систем ИИ. Ввиду своего транснационального характера долгосрочные решения могут быть найдены только на глобальном уровне. В рамках своего нормотворческого мандата ЮНЕСКО следует предпринять шаги к разработке нормативного документа по этике ИИ, что будет способствовать повышению осведомленности различных заинтересованных сторон об этических последствиях ИИ для социальных, культурных и научных аспектов жизни общества. Такой документ должен предусматривать глобальный механизм документирования социокультурных изменений, вызванных стремительным нелинейным развитием ИИ, и сопутствующих этических вопросов. Он также должен способствовать интеграции универсальных ценностей в системы ИИ, которые необходимо привести в соответствие с согласованными на международном уровне документами и стандартами в области прав человека и увязать с ориентированным на человека подходом.

12. Учитывая, что число деклараций национального и регионального уровня об этических принципах ИИ в 2018 г. увеличилось, представляется своевременным, чтобы ЮНЕСКО рассмотрела вопрос о подготовке глобального нормативного документа об этике ИИ с использованием глобального, плюралистического, междисциплинарного, поликультурного и многостороннего подхода, охватывающего все сферы компетенции Организации и опирающегося на ее разнообразные сетевые структуры.

13. Кроме того, в основу нового нормативного документа по этике ИИ также могла бы лечь важная работа ЮНЕСКО по осуществлению решений Всемирной встречи на высшем уровне по вопросам информационного общества (ВВИО). Начиная с 1997 г., ЮНЕСКО выступает с инициативами в отношении этических аспектов информационного общества, что является одним из направлений деятельности в рамках Плана действий ВВИО, ответственность за которое несет ЮНЕСКО. Еще одной отправной точкой для создания нового нормативного документа является обширный опыт ЮНЕСКО по внедрению в область цифровых технологий ориентированного на человека подхода, свойственного основанному на знаниях обществам. Было принято решение о том, что работа в области этики станет одним из приоритетов межправительственной программы ЮНЕСКО «Информация для всех» (ПИДВ). Принципы универсальности интернета R.O.A.M. (отражающие концепцию основанного на правах человека, открытого и доступного интернета, регулируемого с участием многих заинтересованных сторон), которые были утверждены 38-й сессией Генеральной конференции, включены в руководящие положения ЮНЕСКО, касающиеся поощрения разнообразия форм культурного самовыражения в цифровой среде. В совокупности с соответствующими показателями принципы R.O.A.M. представляют собой согласованный механизм, на который ЮНЕСКО и другие заинтересованные стороны могут опираться в определении подходов к прикладному применению, регулированию и этике ИИ.

#### **IV. ФОРМА ДОКУМЕНТА**

14. Характер будущего нормативного документа (декларация, рекомендация или конвенция) должен быть определен Генеральной конференцией в соответствии с Правилами процедуры, касающимися рекомендаций государствам-членам и международных конвенций, подпадающих под действие пункта 4 статьи IV Устава. Авторы различных докладов, указанных в предварительном исследовании расширенной рабочей группы КОМЕСТ, сходятся во мнении о том, что было бы не только желательно, но и необходимо в срочном порядке принять меры к разработке не имеющего обязательной силы глобального инструмента в форме рекомендации. Рекомендация – с учетом ее необязательного характера и ориентации на принципы и нормы международного регулирования того или иного конкретного вопроса – представляется более гибким инструментом, который в большей степени соответствует сложности этических вопросов, возникающих в связи с ИИ.

#### **V. ПРЕДЛАГАЕМЫЙ МЕТОД И ГРАФИК РАБОТЫ**

15. В соответствии с Правилами процедуры, касающимися рекомендаций государствам-членам и международных конвенций, подпадающих под действие пункта 4 статьи IV Устава,

Исполнительному совету предлагается рассмотреть данное предварительное исследование и принять решение о том, следует ли включить его в предварительную повестку дня 40-й сессии Генеральной конференции. Если Исполнительный совет даст данному исследованию положительную оценку, Генеральный директор в соответствии с вышеупомянутыми Правилами процедуры направит государствам-членам текст предварительного исследования, а также касающиеся его замечания и решения Исполнительного совета не позднее чем за 70 дней до открытия 40-й сессии Генеральной конференции, т. е. до конца августа 2019 г.

16. Во исполнение статьи 6 вышеупомянутых Правил процедуры Генеральной конференции будет предложено после рассмотрения данного исследования и касающихся его замечаний Исполнительного совета принять решение о дальнейших действиях.

## VI. ФИНАНСОВЫЕ АСПЕКТЫ

17. Секретариат выполнил тщательную оценку последствий и расходов, которые будут связаны с подготовкой нового нормативного документа в случае принятия Исполнительным советом решения о включении этого пункта в предварительную повестку дня 40-й сессии Генеральной конференции. Учитывая финансовые ограничения, часть соответствующих расходов можно было бы покрыть за счет добровольных взносов в финансовой и натуральной форме.

## VII. ПРЕДЛАГАЕМЫЙ ПРОЕКТ РЕШЕНИЯ

18. С учетом вышеизложенного Исполнительному совету предлагается принять следующее решение:

Исполнительный совет,

1. рассмотрев документ 206 EX/42 и приложение к нему,
2. приветствует предварительное исследование технических и правовых аспектов целесообразности принятия нормативного акта об этике искусственного интеллекта;
3. постановляет включить данный пункт в предварительную повестку дня 40-й сессии Генеральной конференции;
4. предлагает Генеральному директору представить Генеральной конференции на ее 40-й сессии предварительное исследование технических и правовых аспектов целесообразности принятия нормативного документа по этике искусственного интеллекта, содержащееся в документе 206 EX/42, а также соответствующие замечания и решения Исполнительного совета по данному вопросу;
5. рекомендует Генеральной конференции на ее 40-й сессии предложить Генеральному директору представить проект текста нового нормативного документа по этике искусственного интеллекта в форме рекомендации для рассмотрения Генеральной конференцией на ее 41-й сессии.



SHS/COMEST/EXTWG-ETHICS-AI/2019/1  
Париж, 26 февраля 2019 г.  
Оригинал: английский

## **ПРЕДВАРИТЕЛЬНОЕ ИССЛЕДОВАНИЕ ПО ЭТИКЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

Настоящее предварительное исследование подготовлено расширенной рабочей группой КОМЕСТ по этике искусственного интеллекта на основе результатов работы КОМЕСТ в области этики робототехники (2017 г.) и этических аспектов интернета вещей (IoT).

Данный документ не имеет исчерпывающего характера и не обязательно отражает мнения государств – членов ЮНЕСКО.

# ПРЕДВАРИТЕЛЬНОЕ ИССЛЕДОВАНИЕ ПО ЭТИКЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

## СОДЕРЖАНИЕ

### ВВЕДЕНИЕ

#### I. ЧТО ТАКОЕ ИИ?

- I.1 Определение
- I.2 Как работает ИИ?
- I.3 Чем ИИ отличается от других технологий?

#### II. ЭТИЧЕСКИЕ СООБРАЖЕНИЯ

##### II.1 Образование

- II.1.1 Роль образования в обществе
- II.1.2 ИИ в преподавании и обучении
- II.1.3 Подготовка инженеров в сфере ИИ

##### II.2 ИИ и научное знание

- II.2.1 ИИ и научное объяснение
- II.2.2 ИИ, науки о жизни и здравоохранение
- II.2.3 ИИ и экология
- II.2.4 ИИ и общественные науки
- II.2.5 Принятие решений с помощью ИИ

##### II.3 Культура и культурное разнообразие

- II.3.1 Творчество
- II.3.2 Культурное разнообразие
- II.3.3 Язык

##### II.4 Коммуникация и информация

- II.4.1 Дезинформация
- II.4.2 Журналистика данных и автоматизированная журналистика

##### II.5 ИИ в укреплении мира и безопасности

##### II.6 ИИ и гендерное равенство

##### II.7 Африка и задачи в области ИИ

#### III. НОРМАТИВНЫЙ ДОКУМЕНТ

- III.1 Декларация или рекомендация
- III.2 Предложения по содержанию нормативного документа



## ПРЕДВАРИТЕЛЬНОЕ ИССЛЕДОВАНИЕ ПО ЭТИКЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

### ВВЕДЕНИЕ

1. Мир переживает период стремительного развития «искусственного интеллекта» (ИИ). Достижения в этой области приводят к появлению машин, способных учиться и выполнять когнитивные задачи, с которыми раньше могли справиться только люди. Такое технологическое развитие, по-видимому, будет иметь далеко идущие социальные и культурные последствия. Поскольку ИИ является когнитивной технологией, различные ее аспекты неразрывно связаны с основными сферами деятельности ЮНЕСКО: образованием, наукой, культурой и коммуникацией. Алгоритмы стали играть решающую роль в выборе информации и новостей, которые люди читают, музыки, которую люди слушают, и решений, которые люди принимают. За консультациями к системам ИИ все чаще обращаются врачи, ученые и судьи. В области научных исследований ИИ стал играть немаловажную роль в анализе и интерпретации данных. В дополнение к этому продолжающееся замещение труда людей интеллектуальными технологиями требует новых инструментов обеспечения жизнеспособности и гибкости трудовых ресурсов. Влиятельные мыслители, такие как Стивен Хокинг, даже высказывали опасения о том, что ИИ может создать угрозу для существования человечества, поскольку он способен взять под контроль многие аспекты нашей повседневной жизни и организации общественной сферы.

2. В 1950-е годы термин «искусственный интеллект» был введен для обозначения машин, способных выполнять нестандартные задачи. По мере увеличения вычислительных мощностей этот термин стал применяться к машинам, которые обладают способностью к обучению. Хотя единого определения ИИ не существует, общепризнанно, что системы, основанные на ИИ или на «когнитивных вычислениях», потенциально способны имитировать или даже превосходить человеческие когнитивные способности, такие как чувственное восприятие, языковое взаимодействие, логическое мышление и анализ, решение задач и даже творчество. Более того, такие «интеллектуальные машины» могут демонстрировать сходные с человеческими способности к обучению, подкрепленные механизмами самосоотнесения и самокоррекции и основанные на алгоритмах «машинного» или даже «глубинного» обучения с использованием «нейронных сетей», имитирующих функционирование человеческого мозга.

3. В последнее время крупные транснациональные технологические компании во многих регионах мира начали вкладывать значительные средства в интеграцию ИИ в свою продукцию. Вычислительная техника стала достаточно мощной для того, чтобы выполнять очень сложные алгоритмы и работать с «большими данными» – огромными массивами информации, которую можно использовать для машинного обучения. Эти компании имеют доступ к практически неограниченным вычислительным мощностям, а также к данным, которые собраны с миллиардов людей и которые можно загружать в системы ИИ в качестве исходного материала для их обучения. Кроме того, благодаря продукции этих компаний стремительно растет влияние ИИ на повседневную жизнь людей и на такие профессиональные сферы, как здравоохранение, образование, научные исследования, коммуникации, транспорт, безопасность и искусство.

4. Столь масштабное влияние ИИ вызывает опасения, которые могут сказаться на доверии людей к этим технологиям. Эти опасения связаны с целым рядом проблем в диапазоне от возможных преступных действий, мошенничества и кражи личных данных до домогательств и сексуальных надругательств, от человеконенавистнических высказываний и дискриминации до распространения дезинформации и в более общем плане от прозрачности алгоритмов до оправданности доверия к системам ИИ. Поскольку многие из этих проблем не могут быть решены только путем регулирования, ЮНЕСКО предлагает создать механизм многостороннего управления, который мог бы стать оптимальным инструментом привлечения различных заинтересованных сторон к разработке и осуществлению нормативных, этических и политических принципов, а также способствовать расширению прав и возможностей пользователей.

5. Ввиду масштабности социальных последствий многие организации и правительства обеспокоены этическими аспектами использования ИИ. Европейская комиссия сформировала группу экспертов высокого уровня по ИИ, включающую представителей научных кругов, гражданского общества, промышленности, а также Европейского альянса по ИИ – форума, в рамках которого проводится широкое и открытое обсуждение всех аспектов развития ИИ и его последствий. Европейская группа по этике в науке и новых технологиях выпустила *Заявление об искусственном интеллекте, робототехнике и автономных системах* (EGE, 2018). Европейская комиссия опубликовала информационное сообщение *Искусственный интеллект для Европы* (ЕС, 2018), а Совет Европы подготовил целый ряд докладов об ИИ и сформировал комитет экспертов для работы над документом *Правозащитные аспекты автоматизированной обработки данных и различных форм искусственного интеллекта*. Организация IEEE выступила с Глобальной инициативой по этике автономных и интеллектуальных систем. ОЭСР инициировала проект «Переход к цифровым технологиям», который направлен на то, чтобы помочь органам, ответственным за разработку политики во всех соответствующих областях, лучше разобраться в цифровой революции, происходящей в различных секторах экономики и в обществе в целом. ОЭСР также создала экспертную группу (AIGO) для выработки рекомендаций в отношении принципов определения сферы применения искусственного интеллекта в обществе. МСЭ и ВОЗ создали целевую группу под названием «Искусственный интеллект на службе здоровья». Кроме того, многие страны приступили к анализу своих этических и политических подходов к ИИ, примерами чего могут служить доклад Виллани во Франции (Villani et al., 2018), доклад Палаты лордов в Великобритании (House of Lords, 2017) и доклад Исполнительного офиса Президента США (2016).

6. ЮНЕСКО имеет возможность внести уникальный вклад в эту дискуссию. ИИ оказывает влияние на основные сферы деятельности ЮНЕСКО. В связи с этим ЮНЕСКО может дополнить многочисленные руководящие принципы и указания по этике, которые в настоящее время разрабатываются государственными органами, компаниями и общественными организациями, междисциплинарным, универсальным и целостным подходом к развитию ИИ на благо человечества и во имя мира и устойчивого развития.

7. В данной области уже есть несколько рамочных документов и инициатив, которые можно взять за основу. Во-первых, существует правозащитный рамочный документ, который лег в основу Женевской декларации, принятой в рамках Всемирной встречи на высшем уровне по вопросам информационного общества (ВВИО) 2003 г., и в которой говорится, что «при использовании ИКТ и в ходе создания контента следует уважать права человека и основные свободы других лиц, включая право на неприкосновенность частной жизни и право на свободу мысли, совести и религии, согласно положениям соответствующих международных документов» (WSIS, 2003). ВВИО (2005) предлагает применять многосторонний подход, предусматривающий эффективное взаимодействие между всеми вовлеченными сторонами, включая правительства, частный сектор, гражданское общество, международные организации и научно-техническую общественность. В рамках деятельности по итогам ВВИО ЮНЕСКО учла этот многосторонний подход и взяла на себя ответственность за реализацию направлений деятельности, связанных с вопросами доступа (С3), электронным обучением (С7), культурным разнообразием (С8), средствами массовой информации (С9) и этическими аспектами информационного общества (С10).

8. Во-вторых, существует рамочная концепция *универсальности интернета* и связанные с ней *принципы R.O.A.M.*, утвержденные Генеральной конференцией на ее 38-й сессии в 2015 г. (UNESCO, 2015b). Эти принципы, охватывающие вопросы прав человека, открытости, доступности и участия многих заинтересованных сторон, стали результатом исследования ЮНЕСКО «Основные аспекты укрепления инклюзивных обществ знаний», проведенного для 38-й сессии Генеральной конференции (UNESCO, 2015a). В итоговом документе конференции под названием «Объединяя усилия: варианты будущих действий» ЮНЕСКО взяла на себя обязательство содействовать на основе правозащитного подхода этическому анализу, исследованиям и общественному диалогу по вопросу о последствиях внедрения новых и перспективных технологий и их потенциальном воздействии на общество. В дополнение к этому

Межправительственный совет программы ЮНЕСКО «Информация для всех» (ПИДВ) на своей 18-й сессии рассмотрел и утвердил Кодекс этики для информационного общества, который был разработан рабочей группой ПИДВ по информационной этике.

9. В целях анализа этических аспектов ИИ в первой части настоящего исследования разъясняются вопросы о том, что такое искусственный интеллект, как он работает и чем отличается от других технологий. Во втором разделе рассматриваются этические аспекты ИИ применительно к основным направлениям работы ЮНЕСКО (образование, наука, культура и коммуникация), а также глобальные этические вопросы, касающиеся укрепления мира, культурного разнообразия, гендерного равенства и устойчивого развития. Настоящее исследование следует рассматривать не как результат всеобъемлющего анализа, а как попытку наметить пути для изучения широкого круга тем в диапазоне от культурного разнообразия до доверия к науке, от художественного творчества до критического мышления и от принятия решений с использованием ИИ до роли ИИ в развивающихся странах. В третьем разделе настоящего предварительного исследования в общем виде описаны основные элементы, которые, с точки зрения ЮНЕСКО, должны быть отражены в надлежащем этическом анализе ИИ.

## I. ЧТО ТАКОЕ ИИ?

### I.1 Определение

10. Идея «искусственного интеллекта» (ИИ) – в виде представлений об «искусственно созданных» «умных» существах, машинах или инструментах – проходит пунктиром через всю историю человечества. В различных формах она представлена как в западных, так и в незападных религиях, мифологии, литературе и философских традициях. Уже сам этот факт свидетельствует об извечном пытливом интересе человечества к такого рода сущностям, и несмотря на то, что проявления этого интереса принимали различные в культурном отношении формы, это явление носит всеобщий или межкультурный характер. Сегодня увлеченность ИИ, и в том числе его этическими аспектами, усиливается благодаря его развитию и использованию в реальных условиях.

11. Любое рассмотрение этических аспектов ИИ требует уточнения возможных значений этого понятия. Этот термин был введен в 1955 году Джоном Маккарти, Марвином Л. Мински, Натаниэлем Рочестером и Клодом Е. Шенноном. Предполагалось, что «исследования в области искусственного интеллекта» будут «проводиться исходя из предположения о том, что любой аспект познания или любое другое свойство интеллекта в принципе могут быть настолько точно описаны, что для их имитации может быть создана машина» (McCarthy et al., 2006 [1955], с. 12). По мере развития и диверсификации этой области знания в последующие десятилетия число значений термина «искусственный интеллект» росло, и на сегодняшний день какого-либо общепризнанного его определения не существует. Различные определения ИИ связаны с различными подходами, принятыми в рамках отдельных дисциплин, таких как информатика, электротехника, робототехника, психология и философия.

12. Несмотря на многочисленность и разнообразие определений ИИ, на самом общем уровне существует определенный консенсус в отношении возможности разграничения двух концепций ИИ, одна из которых обычно называется «теоретической», или «научной», а другая – «прагматической», или «технологической».

13. «Теоретический», или «научный», подход к ИИ предполагает «использование понятийного аппарата и моделей ИИ для поиска ответов на вопросы о людях и других живых существах» (Boden, 2016, с. 2). Таким образом, «теоретическая», или «научная», концепция ИИ естественным образом связана с такими дисциплинами, как философия, логика, лингвистика, психология и когнитология. Она имеет дело с вопросами следующего рода: Что такое «интеллект» и чем отличается «естественный» интеллект от «искусственного»? Нужен ли символический язык для мыслительных процессов? Можно ли создать «сильный ИИ» (*подлинный ин-*

теллект того же типа и уровня универсальности, что и человеческий интеллект) в противоположность «слабому ИИ» (интеллект, который лишь *имитирует* человеческий интеллект и способен выполнять ограниченное число узко определенных задач)? Хотя подобные вопросы носят теоретический или научный характер, они связаны с рядом метафизических или духовных проблем (в том числе касающихся уникальности человека или свободы воли), которые сами по себе имеют косвенные, но тем не менее серьезные этические последствия.

14. «Прагматическая», или «технологическая», концепция ИИ ориентирована на инженерную сторону дела. Она направлена на то, чтобы на основе различных компонентов искусственного интеллекта – хрестоматийными примерами в этом отношении являются обработка естественного языка, представление знаний, машинная логика, машинное обучение, глубинное обучение, компьютерное зрение и робототехника (Russell and Norvig, 2016, стр. 2-3) – создавать машины или программы, способные самостоятельно выполнять задачи, которые в ином случае требуют человеческого интеллекта и участия. «Практический», или «технологический», подход к ИИ стал чрезвычайно успешным за счет того, что его начали применять в сочетании с ИКТ (информационно-коммуникационными технологиями). Инновации, связанные с ИИ, сегодня используются во многих областях современной жизни, таких как транспорт, медицина, коммуникации, образование, наука, финансы, право, военное дело, маркетинг, обслуживание клиентов и организация досуга. Эти инновации напрямую связаны с целым рядом этических проблем, начиная с ликвидации традиционных рабочих мест и ответственности за возможный физический или психологический ущерб для человека и заканчивая общей дегуманизацией отношений между людьми и общества в целом. В настоящее время ни одна из систем ИИ не может рассматриваться в качестве разумного субъекта действия универсального назначения, способного хорошо функционировать в широком диапазоне условий внешней среды, что является неотъемлемым свойством человеческого интеллекта.

15. Одна из особенностей ИИ заключается в том, что он является «чужеродным» по отношению к нам, людям, в том смысле, что то, как этот интеллект работает, кажется нам странным и загадочным. В основе этой «чужеродности» лежит то, что можно назвать «неосознанной результативностью». Высокофункциональные системы ИИ, такие как AlphaGo или Watson, могут демонстрировать впечатляющие результаты, даже не зная, что именно они делают. Программа AlphaGo победила нескольких мастеров го, даже не подозревая, что она играет в человеческую игру под названием «го». Система Watson отвечала на дьявольски трудные вопросы настолько быстро, что большинству людей в отведенное время было трудно эти вопросы даже понять. И все же Watson не «отвечает» на вопросы в человеческом смысле этого слова, система скорее «рассчитывает» вероятность правильности нескольких возможных ответов на основе автоматизированного анализа имеющейся базы данных. Системы AlphaGo и Watson показывают блестящие результаты, не ведая, что творят.

16. Безусловно, в связи с этим возникают важные философские вопросы о том, являются ли «игра» AlphaGo и «ответы» Watson «подлинными» или нет. Однако еще более значимым с этической точки зрения представляется тот факт, что мы, люди, просто не привыкли к этому виду интеллекта. Всякий раз, когда мы сталкиваемся с запоминающимися произведениями искусства и литературы или яркими научными работами, мы естественным образом предполагаем, что за ними стоит «сознательный» интеллект. Мы отдаем себе отчет в том, что за 9-й симфонией Бетховена стоит уникальная личность композитора, а за теоремой Гёделя о неполноте – пылливый ум ее автора. Тот простой факт, что мы не должны применять это привычное эмпирическое правило к блестящим результатам, демонстрируемым высокофункциональными системами на основе ИИ, создает серьезные социальные и этические проблемы. Поскольку мы привыкли взаимодействовать с разумными в поведенческом плане субъектами на эмоциональном и социальном уровне, мы естественным образом начинаем эмоционально и социально взаимодействовать и с «высокофункциональным искусственным интеллектом, не обладающим сознанием», например, с так называемыми «эмоциональными» или «социальными» роботами, такими как «умные помощники» Алекса, Сири или Google-ассистент. На современном этапе технологического развития высокофункциональный ИИ, не обладающий сознанием, не может должным образом отвечать на сложные эмоциональные и

социальные ожидания людей, но его внешнее поведение в сочетании с человеческим воображением может породить «нереалистичную» надежду на подлинное взаимодействие с людьми. Важно помнить, что кажущийся «эмоциональным» разум ИИ в гораздо большей степени плод нашего воображения, чем реальность. Все согласны с тем, что системы на основе искусственного интеллекта не обладают сознанием, понимаемым как человеческое переживание, даже если они могут отвечать на вопросы о контексте своих действий. Важно не отождествлять переживание с интеллектом, даже если, по мнению некоторых экспертов, последние разработки в сфере ИИ могут послужить причиной для пересмотра важности наличия такого переживания или осознания для того, чтобы быть человеком. Если в основе того, что делает человека человеком лежит переживание, с этической точки зрения необходимо, чтобы использование ИИ приводило к его сохранению и обогащению, а не к вытеснению или выхолащиванию. Однако может случиться так, что наш опыт взаимодействия с высокофункциональными системами ИИ, не обладающими сознанием, будет влиять на наше взаимодействие с обычными людьми, сознанием наделенными.

## **I.2 Как работает ИИ?**

17. Для того чтобы система ИИ могла решать задачи, посильные разуму человека, она должна воспринимать окружающую среду, собирать данные в динамическом режиме, быстро их обрабатывать и реагировать на них, опираясь на свой прошлый «опыт», заложенные в систему принципы принятия решений и предвидение будущего. Однако технология, лежащая в основе ИИ, является обычной ИКТ: она основана на сборе/получении данных, их хранении, обработке и передаче. Уникальные характеристики когнитивных машин обусловлены количеством, которое переходит в качество. В основе технологии ИИ лежат следующие элементы:

- (a) *Динамические данные.* На систему должны воздействовать изменяющиеся условия внешней среды, и она должна получать все необходимые данные с помощью различных датчиков, классифицировать и хранить их, а также быть в состоянии оперативно их обрабатывать.
- (b) *Оперативная обработка.* Когнитивные машины должны реагировать быстро. Поэтому ИИ должен обладать надежными, быстродействующими и мощными вычислительными и коммуникационными средствами.
- (c) *Принципы принятия решений.* ИИ принимает решения на основе алгоритмов машинного обучения. Поэтому его реакция на конкретную задачу зависит от «опыта», которым располагает система, то есть от тех данных, которые она получала. Алгоритмы, лежащие в основе принимаемых когнитивными машинами решений, базируются на некоторых общих принципах, которые алгоритм соблюдает и пытается оптимизировать с учетом поступающих данных.

Имеющаяся в настоящее время возможность эффективным образом интегрировать алгоритмы динамического сбора данных и машинного обучения в целях оперативного принятия решений и позволяет создавать «когнитивные машины».

## **I.3 Чем ИИ отличается от других технологий?**

18. Большинство технологий 20-го века основаны на моделях. Это означает, что ученые изучают природу и предлагают научные модели для ее описания, на основе которых и происходит развитие технологий. Например, понимание принципов распространения электромагнитных волн является основой технологии беспроводной связи. Моделирование человеческого мозга, однако, представляет собой задачу, которая все еще далека от той стадии, когда когнитивные машины будут основаны на моделях. Поэтому ИИ реализован на базе иного подхода – подхода, основанного на данных.

19. Подход, основанный на данных, формирует ядро машинного обучения, которое, как правило, реализуется на базе «искусственных нейронных сетей» (ИНС). ИНС состоят из серии узлов, концептуально схожих с нейронами мозга и соединенных между собой связями, проходящими через несколько слоев. Узлы входного слоя получают информацию из внешней среды, и в каждом таком узле происходит нелинейное преобразование данных. Такие системы «учатся» выполнять задачи, анализируя примеры (маркированные данные), и обычно не имеют в своей программе каких-либо правил или моделей, специфичных для конкретной задачи. В конечном счете, глубинное обучение строится на многослойных ИНС, что позволяет машине распознавать сложные концепты, такие как человеческие лица и тела, понимать устную речь и классифицировать изображения любого типа.

20. Ключевым аспектом способности ИИ демонстрировать возможности, сходные с человеческими, является масштабируемость. Эффективность систем на основе ИИ зависит от данных, которые они получают, и для обеспечения максимальной эффективности доступ к соответствующим данным должен быть неограниченным. Могут существовать технические ограничения на доступ к данным, но способ отбора и классификации данных является также и социокультурным вопросом (Crawford, 2017). Классификация зависит от культуры и складывается в результате исторического развития, что может приводить к тому, что решения, принимаемые алгоритмом, будут предвзятыми. Если та же система получит доступ к другому массиву данных, степень этой предвзятости может быть снижена, но она не может быть устранена полностью (Executive Office of the President, 2016). Важно отметить, что для выполнения требований статьи 27 Всеобщей декларации прав человека, в которой говорится, что каждый человек имеет право пользоваться благами научного прогресса, и для обеспечения разнообразия доступных ИИ наборов данных необходимо содействовать наращиванию потенциала государств как в области подготовки кадров, так и в сфере инфраструктуры.

21. ИИ достиг стадии технологической зрелости во многом благодаря энергичным действиям транснациональных компаний, которые не связаны местными и национальными ограничениями. Кроме того, для обеспечения оперативности обработки данных и надежности систем вычислительные процессы разнесены по разным точкам, а местонахождение системы ИИ не определяется местом ее эксплуатации. На практике ИИ опирается на облачные технологии, а это значит, что устройства хранения и обработки данных могут находиться где угодно. Технология ИИ характеризуется следующими признаками:

- (a) Хотя ИИ широко применяется в публичной сфере, технологии ИИ разрабатываются и продвигаются транснациональными компаниями, большинство из которых работают в частном секторе, вследствие чего общественное благо не является их главным приоритетом.
- (b) ИИ не привязан к какому-либо физическому местоположению. Это создает проблемы для регулирования технологий ИИ на национальном и международном уровнях.
- (c) Технология основана на доступности как персональных, так и публичных данных.
- (d) Технологии ИИ не являются нейтральными, им внутренне присуща предвзятость, обусловленная данными, используемыми в ходе их обучения, и решениями, которые принимаются в процессе обучения с применением этих данных.
- (e) Решения ИИ и когнитивных машин не могут быть в полной мере предсказуемыми или объяснимыми. ИИ не сводится к механическим или детерминированным процессам, его программа совершенствуется, обучаясь по мере поступления динамических данных и интегрируя полученный на практике опыт в процесс принятия решений.

## **II. ЭТИЧЕСКИЕ СООБРАЖЕНИЯ**

22. Искусственный интеллект сопряжен с существенными социальными и культурными последствиями. Как и многие другие информационные технологии, ИИ ставит вопросы свободы выражения мнений, неприкосновенности частной жизни и слежки, прав собственности на данные, предвзятости и дискриминации, манипулирования информацией и доверия, властных отношений и воздействия на окружающую среду вследствие потребления энергии. Кроме того, ИИ создает новые проблемы особого рода, связанные с его взаимодействием с когнитивными способностями человека. Системы, основанные на ИИ, чреватые определенными последствиями с точки зрения человеческого понимания и опыта. Алгоритмы социальных сетей и новостных сайтов могут способствовать распространению дезинформации и воздействовать на восприятие смысла понятий «факты» и «истина», а также сказываться на политическом взаимодействии и активности. Машинное обучение может способствовать укоренению и усугублению предвзятости, которая, в свою очередь, может повлечь за собой углубление неравенства и отчуждения и создать угрозу для культурного разнообразия. Масштабы и мощность технологий ИИ усиливают асимметрию между отдельными людьми, их группами и целыми странами, проявляющуюся, в частности, в возникновении так называемого «цифрового разрыва» внутри стран и между ними. Этот разрыв может усугубляться отсутствием доступа к таким ключевым элементам, как алгоритмы обучения и классификации, данные, используемые для «тренировки» и оценки алгоритмов, людские ресурсы, необходимые для разработки и настройки программного обеспечения и для подготовки данных, а также вычислительные ресурсы для хранения и обработки данных.

23. Как следствие, искусственный интеллект требует тщательного анализа. С точки зрения ЮНЕСКО, важнейшие этические вопросы, возникающие в связи с искусственным интеллектом, связаны с его влиянием на культуру и культурное разнообразие, образование, научное знание, коммуникацию и информацию. В дополнение к этому, учитывая глобальную ориентацию ЮНЕСКО, особого внимания заслуживают такие глобальные этические темы, как укрепление мира, устойчивое развитие, гендерное равенство и специфические проблемы Африки.

### **II.1 Образование**

24. Искусственный интеллект ставит множество вопросов, касающихся роли образования в обществе. Во-первых, ИИ требует переосмысления социальных задач, стоящих перед образованием. Вытеснение рабочей силы, обусловленное внедрением ИИ в ряде областей, требует среди прочего переподготовки работников и выработки нового подхода к определению содержательного наполнения квалификаций, присваиваемых по итогам учебных программ. Кроме того, в эпоху ИИ образование должно давать гражданам возможность осваивать новые формы критического мышления, частью которых является «алгоритмическая грамотность» и способность анализировать воздействие ИИ на информацию, знания и принятие решений. Еще одна группа этических вопросов, касающихся ИИ и образования, связана с ролью ИИ в учебном процессе как таковом в качестве элемента цифровой обучающей среды, образовательной робототехники и систем «аналитики обучения», с учетом того, что все эти сферы требуют ответственного подхода к разработке и внедрению. И наконец, инженеры и разработчики ПО должны пройти соответствующую подготовку, что позволит обеспечить ответственный подход к проектированию и реализации систем ИИ.

#### **II.1.1 Роль образования в обществе**

25. Одной из основных социальных проблем, связанных с ИИ, является вытеснение рабочей силы. Скорость изменений, обусловленных внедрением ИИ, создает беспрецедентные по масштабам проблемы (Illanes et al., 2018). В ближайшем будущем это повлечет за собой необходимость переподготовки большого числа работников и будет иметь серьезные последствия для карьерных перспектив учащихся. По данным экспертного исследования McKinsey, проведенного в 2017 г., «руководители все чаще рассматривают инвестиции в переподготовку и

повышение квалификации работающих сотрудников в качестве неотложной задачи, стоящей перед их компаниями» (Illanes et al., 2018).

26. Вследствие этого ИИ будет подталкивать общество к переосмыслению образования и его социальной роли. Предлагаемое университетами традиционное «формальное» образование может оказаться недостаточным в период ускоренного развития цифровой экономики и приложений на базе ИИ. До настоящего времени стандартная модель образования, как правило, сводилась к передаче «базовых знаний» (Oppenheimer, 2018) и была нацелена на обеспечение функциональной грамотности в таких областях, как чтение, письмо и математика. В 21-м веке информация и знания присутствуют повсеместно, в результате чего возникает потребность не только в «грамотности в обращении с данными», которая позволяет учащимся воспринимать и анализировать эту информацию и эффективно управлять ею, но и в «ИИ-грамотности», необходимой для критического осмысления того, каким образом интеллектуальные компьютерные системы участвуют в выявлении информационных потребностей, а также в выборе, интерпретации, хранении и представлении данных.

27. Кроме того, в условиях постоянно развивающегося рынка труда система образования больше не может быть нацелена на обучение людей какой-то одной конкретной профессии. Образование должно делать людей разносторонними и адаптируемыми, готовыми к жизни в мире, в котором технологии формируют динамичный рынок труда и в котором работники должны самостоятельно переучиваться на регулярной основе. Нынешние представления об «обучении на протяжении всей жизни», возможно, потребуют трансформации в модель непрерывного образования, предполагающую разработку новых видов академических квалификаций.

### II.1.2 ИИ в преподавании и обучении

28. Открытые образовательные ресурсы (ООР) стали важным элементом образовательного ландшафта благодаря бесплатному доступу к первоклассным лекциям и другим учебным ресурсам, размещенным в интернете. Потенциал воздействия ООР на образование во всем мире беспрецедентен, но еще не в полной мере реализован, о чем свидетельствует невысокий процент учащихся, успешно оканчивающих массовые открытые онлайн-курсы (МООК). Большое разнообразие и охват имеющихся ресурсов породили две проблемы. Во-первых, стало непросто отыскать ресурс, подходящий конкретному учащемуся или учителю, желающему использовать этот ресурс при подготовке собственных учебных материалов. Это привело ко второй проблеме, заключающейся в уменьшении разнообразия вследствие того, что некоторые ресурсы стали очень популярными в ущерб другим материалам, возможно, более релевантным, но менее доступным.

29. В качестве примера можно привести проект «X5GON» программы «Горизонт-2020» (Cross Modal, Cross Cultural, Cross Lingual, Cross Domain and Cross Site Global OER Network: <https://www.x5gon.org/>). В рамках этого проекта, финансируемого Европейским союзом, разрабатываются методы, основанные на искусственном интеллекте и позволяющие как учащимся, так и преподавателям находить ресурсы, отвечающие стоящим перед ними учебным целям, с учетом особенностей их ситуации. Например, учителю из Африки могут быть рекомендованы лекции, в ходе которых та или иная тема раскрывается на материале, знакомом местному или коренному населению, с учетом соответствующих культурных особенностей и местных условий, но которые в то же время позволяют учащимся из других стран, интересующимся специфическими для Африки проблемами, подыскать тематические ресурсы, основанные на африканском материале, возможно, в переводе с местного языка.

30. Таким образом, ИИ потенциально способен решить обе указанные выше проблемы. Первая проблема решается путем оказания помощи в подборе ресурсов, которые в наибольшей степени соответствуют потребностям учащегося или преподавателя, на основе моделирования их интересов и целей, а также путем использования усовершенствованных интерфейсов доступа к огромным хранилищам ООР, существующим сегодня в мире. Адаптируя



рекомендации к нуждам отдельного учащегося или учителя, ИИ решает и вторую проблему, поскольку в этом случае рекомендации не направляют пользователя автоматически к наиболее популярному ресурсу, посвященному той или иной теме. Кроме того, здесь есть потенциал для налаживания связей между учащимися, представляющими различные культуры, что способствует межкультурному обмену идеями, а значит и углублению взаимопонимания и взаимного уважения.

### II.1.3 Подготовка инженеров в сфере ИИ

31. Разработка технологий будущего находится в руках специалистов технического профиля. Традиционно инженеров учат создавать продукты, обеспечивающие оптимальную производительность с минимальными затратами (в отношении энергии, спектра, пространства, веса и т. д.) в условиях заданных внешних ограничений. В последние десятилетия в сфере инженерной этики были разработаны различные методы интеграции элементов этического анализа, ответственности и обоснования в процесс проектирования. В контексте ИИ для обозначения методов проектирования, недвусмысленным образом ориентированных на человеческие ценности, был введен термин «этически обоснованное проектирование» (ЭОП) (IEEE, 2018).

32. Методы этически обоснованного проектирования особенно важно применять при разработке систем ИИ и других автономных интеллектуальных систем (АИС), поскольку это позволяет решать этические проблемы на том этапе, когда технологию еще можно адаптировать. Хорошей иллюстрацией такого подхода является концепция «встроенной конфиденциальности». Нарушений права на неприкосновенность частной жизни будет меньше, если сохранять не все данные, а только те, которые необходимы для решения конкретной задачи. Примером может служить подсчет участников массовых мероприятий, когда людей в толпе считают по фотографиям. В этом случае, если фотография предварительно обработана так, что видны только контуры (очертания) фигур, люди останутся неузнанными и алгоритм подсчета выполнит свою задачу без вмешательства в частную жизнь участников. Разработчики ИИ могут сходным образом учесть и другие этические вопросы, такие как предотвращение алгоритмической предвзятости и обратной отслеживаемости, сведение к минимуму злоупотреблений технологиями и обеспечение объяснимости алгоритмических решений.

33. Инженерное образование в сегодняшнем мире по большей части сводится к естественнонаучным и технологическим курсам, напрямую не связанным с анализом человеческих ценностей, который был бы очевидным образом нацелен на реальный рост благополучия людей и экологического благосостояния. Очень важно изменить это положение вещей и начать обучать будущих инженеров и программистов методам этически обоснованного проектирования систем ИИ. Для этого нужно ясно сознавать потенциальные социальные и этические последствия применения разрабатываемых технологий и возможных злоупотреблений ими. IEEE (всемирная организация, объединяющая более 400 000 инженеров-электротехников) уже проводит работу в данном направлении в рамках глобальной инициативы по этике автономных и интеллектуальных систем (<https://ethicsinaction.ieee.org/>). Решение этой проблемы также требует активных усилий по обеспечению гендерной интеграции и социокультурного многообразия инженерного сообщества, а также всестороннего учета социальных и этических последствий разработки систем ИИ. Следует поощрять диалог между инженерным корпусом и общественностью, направленный на содействие обмену информацией о потребностях и задачах общества, а также о методах, которые инженеры применяют в своей повседневной работе и исследовательской деятельности.

## II.2 ИИ и научное знание

34. Судя по всему, применение ИИ окажет значительное воздействие на практику научной работы. В сфере естественных и общественных наук, а также наук о жизни и окружающей среде в связи с ИИ возникают фундаментальные вопросы, касающиеся наших представлений

о научном понимании и объяснении. Это также влияет на то, каким образом мы применяем научные знания в социальном контексте.

### **II.2.1 ИИ и научное объяснение**

35. Вследствие применения все более мощных форм машинного и глубинного обучения ИИ ставит под сомнение существующие представления о том, что должно считаться удовлетворительным научным объяснением и чего следует естественным образом ожидать от предсказуемо успешных научных теорий. С точки зрения общепринятой в науке модели, получившей название дедуктивно-номологической, надлежащее научное объяснение должно давать возможность достоверного прогнозирования конкретных явлений на основе научных законов, теорий и наблюдений. Так, мы можем с полным правом сказать, что мы способны объяснить движение Луны вокруг Земли с точки зрения ньютоновской механики, только если можем на основе той же ньютоновской механики методом дедукции предсказать дальнейшее движение Луны по орбите. Такие прогнозы, как правило, основываются на понимании причинно-следственных связей или на унифицирующей интерпретации кажущихся разнородными явлений.

36. ИИ же, напротив, может стабильно выдавать впечатляюще точные прогнозы на основе наборов данных, не предъявляя нам каких-либо объяснений, основанных на причинно-следственных связях или носящих унифицирующий характер. Алгоритмы ИИ не работают с теми семантическими понятиями, которые используются людьми в процессе формирования научного понимания того или иного явления. Этот разрыв между успешными прогнозами с одной стороны и удовлетворительным научным объяснением с другой, вероятно, будет играть ключевую роль в научной практике, а также в принятии решений с помощью ИИ.

37. Все это может сказаться на доверии к науке, обычно основанном на научном методе, который систематически и прозрачным образом объясняет различные явления и позволяет делать осмысленные прогнозы, базирующиеся на фактах. Очевидный успех алгоритмов машинного обучения, выражающийся в том, что они демонстрируют сопоставимые результаты без привлечения такой научно обоснованной модели, может иметь последствия для восприятия и оценки науки и научных исследований обществом.

38. В дополнение к этому исследования показывают, что качество машинного обучения в значительной степени зависит от данных, которые используются для «тренировки» алгоритмов. Однако поскольку большинство приложений на базе ИИ разрабатываются частными компаниями, эти данные не всегда в достаточной мере прозрачны, что отличает их от традиционного научного метода, который гарантирует достоверность результатов, требуя их воспроизводимости, то есть возможности повторения при проведении тех же экспериментов.

### **II.2.2 ИИ, науки о жизни и здравоохранение**

39. Развитие технологий ИИ, применяемых в области наук о жизни, и в частности в медицине, с годами значительно изменило ландшафт здравоохранения и биоэтики. Эти технологии могут принести положительные результаты, например, в плане повышения точности роботизированной хирургии и улучшения ухода за детьми-аутистами, но в то же время они вызывают вопросы этического характера, которые, в частности, обусловлены сопряженными с ними расходами в контексте нехватки ресурсов в системе здравоохранения и прозрачностью, которую они должны обеспечить в целях соблюдения права пациентов на принятие самостоятельных решений.

40. Отдельным людям, не являющимся профессионалами, ИИ дает возможность по-новому подойти к решению вопросов, связанных со здоровьем и медициной. Использование интернет-сайтов и рост числа мобильных приложений для самодиагностики позволяют людям ставить себе диагноз без обращения к врачу. Такая ситуация может сказаться как на авторитете медицины, так и на восприятии приемлемости самолечения со всеми сопутствующими рис-

ками. Она также изменяет отношения между врачом и пациентом и свидетельствует о необходимости введения той или иной формы регулирования, не препятствующей инновациям и не ограничивающей автономии пациента.

41. Технологии ИИ могут помочь медицинским работникам высвободить дополнительное время для работы с пациентами, например, за счет облегчения ввода данных и канцелярской работы, однако в то же время эти технологии могут подменить собой целостный и человеческий подход к пациенту. Известная технология Watson for Oncology компании IBM является прорывом в лечении рака, но при этом она поднимает важные вопросы о характере компетенций и образования медицинских работников и связанных с ними ожиданий, а также об ответственности врачей, работающих с данной системой. Аналогичные опасения вызывает разработка автоматических программ-собеседников для общения с людьми, обращающимися за психологической помощью и консультациями, приложений, предназначенных для выявления эпизодов психических заболеваний на ранних стадиях, и систем ИИ для постановки психиатрических диагнозов на основе информации о поведении людей в социальных сетях и интернете, что, естественно, также вызывает важные вопросы, касающиеся конфиденциальности. Кроме того, для пожилых людей внедряются такие ИИ-технологии, как социальные роботы-помощники, которые с медицинской точки зрения могут быть полезны, например, пациентам с деменцией, но которые одновременно вызывают беспокойство по поводу уменьшения роли людей в уходе за больными и, как следствие, углубления социальной изоляции.

42. ИИ также привносит новый ракурс в продолжающуюся дискуссию о том, где закладывается «лечение» и начинается «усовершенствование» человека. Существуют проекты по интеграции ИИ с человеческим мозгом с помощью «нейронного интерфейса» - ячеистой структуры, которая растет вместе с мозгом и выполняет роль бесшовного интерфейса между мозгом и компьютером, действуя через систему кровообращения хозяина (Hinchliffe, 2018). Эти технологические разработки напрямую связаны с вопросами о том, что значит быть человеком и что такое «нормальное» функционирование человека.

### **II.2.3 ИИ и экология**

43. ИИ может принести пользу и экологической науке, в которой он может найти разнообразное применение. Его можно использовать для обработки и интерпретации данных, относящихся к сферам экологии, системной биологии, биоинформатики, а также космических и климатических исследований, что будет способствовать углублению научного понимания различных процессов и механизмов. Совершенствование методов переработки отходов, экологического мониторинга и восстановления окружающей среды, а также более эффективное энергопотребление могут приводить к прямым экологическим выгодам. В сфере сельского хозяйства ИИ может способствовать повышению урожайности (например, за счет автоматизированного внесения удобрений и орошения), улучшению условий жизни животных, а также снижению рисков, связанных с болезнями, вредителями и неблагоприятными погодными условиями. В то же время ИИ способен изменить восприятие природы человеком – либо в позитивном ключе, обогащая представления людей о красоте и независимости, либо в негативном – за счет усиления «инструментализации» природы или отчуждения человека от животного мира и окружающей среды в целом.

44. Для каких целей бы ни применялся ИИ, потенциальные выгоды всегда следует соразмерять с воздействием на окружающую среду всего производственного цикла ИИ и информационных технологий (ИТ). В частности, следует принимать в расчет добычу редкоземельных элементов и других полезных ископаемых, производство энергии, необходимой для изготовления и питания соответствующих систем, а также образование отходов в процессе их производства и по окончании жизненного цикла. Распространение систем ИИ, по-видимому, усилит и без того растущую озабоченность по поводу увеличения объемов электронных отходов и повышения спроса на редкоземельные элементы со стороны производителей вычислительной техники. Помимо воздействия на окружающую среду и здоровье, электронные отходы

имеют важные социально-политические последствия, связанные, в первую очередь, с вывозом этих отходов в развивающиеся страны и к местам проживания уязвимых групп населения (Heacock et al., 2015).

45. В сфере управления рисками стихийных бедствий ИИ может использоваться для прогнозирования таких природных катаклизмов, как цунами, землетрясения, торнадо и ураганы, и для принятия соответствующих мер реагирования. В качестве конкретного примера можно привести геосервер ЮНЕСКО Г-ВАДИ (система информации о водных ресурсах и развитии), который используется для информационной поддержки планирования действий в чрезвычайных ситуациях и управления гидрологическими рисками, связанными, в частности, с наводнениями, засухами и экстремальными погодными условиями. Его вспомогательная система PERSIANN (оценка осадков на основе данных дистанционного зондирования с использованием искусственных нейронных сетей) представляет собой спутниковый алгоритм поиска данных об осадках, обеспечивающий получение информации в близком к реальному масштабе времени. Алгоритм классификации облаков CCS системы PERSIANN (доступен по адресу: <http://hydiss.eng.uci.edu/>) был оптимизирован для наблюдения за экстремальными осадками в очень высоком пространственном разрешении и широко используется во всем мире для отслеживания ураганов. В составе системы также есть мобильное приложение iRain (<http://en.unesco.org/news/irain-new-mobile-app-promote-citizen-science-and-support-water-management>), которое предоставляет натуралистам-любителям возможность участвовать в сборе данных на основе совместного финансирования с участием многих сторон.

46. Интересно, что к усилиям по противодействию стихийным бедствиям в последнее время подключились даже частные компании. Одним из примеров такого рода является проект Google по прогнозированию наводнений с использованием ИИ (<https://www.blog.google/products/search/helping-keep-people-safe-ai-enabled-flood-forecasting/>). С этой точки зрения развитие технологий ИИ, потенциально способных принести пользу в сфере предупреждения и ликвидации последствий чрезвычайных ситуаций, следует поощрять.

#### **II.2.4 ИИ и общественные науки**

47. В широком плане исследования в области общественных наук направлены на выявление причинно-следственных связей, лежащих в основе межличностного и социального взаимодействия. Поскольку на большинство социальных явлений по-разному влияют многочисленные причинные факторы, для выявления главных из них и определения силы их воздействия социологи обычно прибегают к статистическому анализу релевантных эмпирических данных. В его ходе крайне важно проводить различие между простыми статистическими корреляциями и подлинными причинно-следственными связями. Не вызывает сомнений, что ИИ способен помочь социологам ориентироваться в огромных массивах данных, выявлять вероятные причинно-следственные механизмы и проверять достоверность предполагаемых взаимозависимостей. В то же время ИИ может «подгонять» данные под себя и «обнаруживать» мнимые причинно-следственные связи там, где их нет. Такая возможность может привести к социальным противоречиям, особенно если предполагаемые причинно-следственные связи являются чувствительными в этическом отношении и касаются, например, предположений о расовых различиях в уровне интеллекта. В этом случае мы опять же не должны автоматически принимать на веру «выводы» ИИ без их оценки человеком.

#### **II.2.5 Принятие решений с помощью ИИ**

48. Методы, основанные на ИИ, потенциально могут оказать огромное воздействие на широкий спектр областей, начиная с юриспруденции и судопроизводства и заканчивая оказанием помощи в принятии решений законодательным и административным государственным органам. Например, они могут повысить эффективность и четкость работы адвокатов как при предоставлении консультаций, так и в ходе судебных разбирательств, что пойдет на пользу самим адвокатам, их клиентам и обществу в целом. Существующие пакеты программного

обеспечения для судей могут быть дополнены и обогащены инструментами на базе ИИ для содействия в подготовке новых решений (СЕРЕУ, 2018).

49. Ключевым вопросом при использовании ИИ в этих целях является характер и интерпретация результатов работы алгоритмов, которые не всегда понятны человеку<sup>2</sup>. Этот вопрос касается и более широкой области принятия решений на основе данных. Поскольку система ИИ в состоянии анализировать, обрабатывать и классифицировать очень большие объемы весьма разнородных данных, которые к тому же могут очень быстро изменяться, такая система считается способной предлагать, а если ей это будет позволено, то и принимать решения в сложных ситуациях. Примеры такого использования ИИ, рассматриваемые в настоящем докладе, включают экологический мониторинг, прогнозирование стихийных бедствий и реагирование на них, предупреждение общественных беспорядков и боевое планирование.

50. Вместе с тем в вопросе о правильности предлагаемого ИИ решения следует проявлять осторожность. Такое решение не обязательно является беспристрастным, справедливым, точным или уместным. На него влияют неточности, избирательность результатов, заложенная или привнесенная в систему предвзятость, а также ограничения, свойственные процессу обучения. Люди не только имеют гораздо более широкое «мировоззрение», но и обладают неявными знаниями, которые позволяют им демонстрировать лучшие по сравнению с ИИ результаты в критических или сложных ситуациях, например, при принятии решений в условиях боя. В идеале оптимальным решением является то, которое мог бы принять человек, если бы он или она были в состоянии обработать огромный массив данных в приемлемые сроки. Однако люди располагают иными возможностями и принимают решения на основе принципиально иной архитектуры, обладая при этом восприимчивостью к возможным проявлениям предвзятости.

51. Весьма сомнительно, что ИИ – по крайней мере, в ближайшем будущем – будет в состоянии справляться с неоднозначными и быстро меняющимися данными или интерпретировать и осуществлять намерения человека в том виде, который они могли бы принять, если бы человек был способен обрабатывать сложные массивы разнородных данных. Даже если решения, вырабатываемые машиной, будет контролировать включенный в процесс управления человек, этого может оказаться недостаточно для принятия «правильного» решения: поскольку когнитивный ИИ принимает решения не так, как это делают люди, человек не обладает знаниями и информацией, необходимыми для того, чтобы определить, соответствует ли решение, выработанное на основе данных, его намерениям. Более того, стохастическое поведение когнитивного ИИ наряду с обусловленной им неспособностью человека понять, почему система сделала тот или иной выбор, приводят к снижению уровня доверия к выбранному решению.

52. Поучительным примером, иллюстрирующим некоторые проблемы использования ИИ для поддержки принятия решений в социальном контексте, является инструмент скрининга семей AFST, представляющий собой прогностическую модель, используемую для предупреждения случаев жестокого обращения с детьми и оставления их без попечения в округе Аллегени, штат Пенсильвания (см. <https://www.alleghenycountyanalytics.us/wp-content/uploads/2017/07/AFST-Frequently-Asked-Questions.pdf>). Этот инструмент был разработан в надежде на то, что решения, основанные на фактических данных, будут объективными и беспристрастными и позволят решить проблемы, возникающие перед органом государственного управления в условиях ограниченности ресурсов. Орган власти, внедривший этот инструмент, как

---

<sup>2</sup> К.Д. Эшли пишет: «Поскольку алгоритм машинного обучения (МО) усваивает правила на основе статистических закономерностей, которые могут вызывать у людей удивление, такие правила не всегда представляются людям обоснованными. [...] Хотя выведенные машиной правила могут приводить к точным прогнозам, они не отталкиваются от человеческого опыта и могут быть менее понятны для людей, чем правила, сформулированные экспертом «вручную». Поскольку правила [...], выведенные алгоритмом, не обязательно отражают эксплицитные юридические знания или опыт, они могут не соответствовать критериям обоснованности, используемым экспертами-людьми». (Ashley, 2017, стр. 111).

представляется, действовал с благими намерениями. Однако недавние исследования показали, что использование инструмента AFST приводит к неблагоприятным последствиям для населения, чьи интересы он был призван защищать (Eubanks, 2018b, стр. 190; Eubanks, 2018a). Система завышает число бедных семей в выборке и для интерпретации и прогнозирования жестокого обращения с детьми использует косвенные показатели, причем таким образом, что это изначально ставит малообеспеченные работающие семьи в невыгодное положение. В результате система усугубляет существующую структурную дискриминацию бедных слоев и оказывает непропорционально сильное негативное воздействие на уязвимые группы населения.

53. В некоторых случаях использование ИИ в качестве субъекта принятия решений (с помощью человека или в полностью автономном режиме) может даже рассматриваться как договор с дьяволом: за скорость обработки и сортировки больших массивов данных, которая присуща системам на основе ИИ, мы должны заплатить отказом от возможности влиять на принимаемые ими решения. А ведь такие решения могут иметь далеко идущие последствия, особенно в конфликтных ситуациях.

### **II.3 Культура и культурное разнообразие**

54. Внедрение ИИ, вероятно, будет иметь существенные последствия для культуры и художественного творчества. Мы уже начинаем наблюдать первые примеры зарождающегося художественного взаимодействия между интеллектуальными алгоритмами и людьми творческих профессий, которое в конечном итоге может создать серьезные проблемы для охраны прав творческих работников, для предприятий культурно-креативного сектора и для судьбы культурного наследия. В то же время применение алгоритмов в сфере потоковых онлайн-медиа и машинного перевода, по-видимому, будет иметь последствия для культурного разнообразия и языка.

#### **II.3.1 Творчество**

55. Искусственный интеллект все шире используется в творчестве и художественной практике человека – в диапазоне от программ «автоподстройки», автоматически корректирующих высоту нот, взятых певцом, до алгоритмов, помогающих создавать произведения изобразительного искусства, сочинять музыку и писать романы или стихи. Творчество, под которым понимается способность создавать новый, оригинальный контент, задействуя воображение или изобретательность, играет центральную роль в открытом, инклюзивном и плюралистическом обществе. По этой причине воздействие ИИ на творчество человека заслуживает особого внимания. Хотя ИИ и является мощным инструментом творчества, в связи с ним возникают важные вопросы о будущем искусства, о правах и вознаграждении творческих работников и о целостности цепочек создания художественных ценностей.

56. Хорошей иллюстрацией в этом отношении является проект «Еще один Рембрандт», в ходе которого совершенно новая картина Рембрандта была создана с помощью ИИ и трехмерного принтера (Microsoft Europe, 2016). Такие произведения искусства требуют нового определения понятия «авторство», которое позволило бы справедливо оценить творческий вклад как «оригинального» автора, так и алгоритмов и технологий, создавших само это произведение. В связи с этим возникают и другие вопросы: как поступать в случае, если ИИ создает произведения искусства самостоятельно? Если живого автора заменяют машины и алгоритмы, в какой степени вообще может идти речь о принадлежности авторских прав? Может и должен ли алгоритм признаваться в качестве автора и пользоваться теми же правами, что и художник?

57. Хотя очевидно, что ИИ способен создавать «оригинальные» художественные произведения, в разработке технологий и алгоритмов ИИ, а зачастую и в создании произведений искусства, которые служат источником вдохновения для художественных работ ИИ, всегда

участвуют люди. С этой точки зрения ИИ можно рассматривать в качестве новой художественной техники, использование которой приводит к созданию произведений искусства нового типа. Если мы действительно хотим сохранить понятие авторства применительно к произведениям, созданным ИИ, потребуется провести анализ вклада различных авторов, «стоящих за» каждым таким произведением, и взаимосвязей между ними. Из этого следует, что мы должны будем разработать новые механизмы, которые позволят провести различие между пиратством и плагиатом с одной стороны и оригинальностью и творчеством с другой, а также дадут возможность по достоинству оценить творческий вклад людей в результат нашего взаимодействия с ИИ. Эти механизмы нужны для того, чтобы предотвратить сознательную эксплуатацию труда и творческого потенциала людей, обеспечить адекватное вознаграждение и признание вклада творческих работников, защитить целостность цепочек создания художественных ценностей и сделать так, чтобы культурный сектор сохранил способность создавать достойные рабочие места.

### II.3.2 Культурное разнообразие

58. ИИ также тесно связан с культурным разнообразием. Хотя он и может оказать позитивное воздействие на культурно-творческий сектор, не все творческие работники и предприниматели обладают навыками и ресурсами, необходимыми для использования технологий ИИ при создании и распространении своих произведений. Коммерческая логика крупных платформ может привести к повышению концентрации предложения товаров и услуг, данных и доходов в области культуры в руках лишь небольшого числа участников рынка, что может иметь негативные последствия для разнообразия форм культурного самовыражения в целом, в том числе опасность создания нового, теперь уже креативного разрыва, а также к усилению маргинализации развивающихся стран.

59. По мере превращения этих платформ в доминирующие механизмы распространения произведений искусства крайне важно будет обеспечить разнообразие и справедливый доступ к этим платформам для творческих работников всех жанров и направлений. В этом контексте особого внимания требуют представители творческих профессий из развивающихся стран. Творческие работники и предприниматели, работающие в сфере культуры, должны иметь доступ к программам профессиональной подготовки, финансовым услугам, инфраструктуре и оборудованию, необходимым для участия в работе этого нового культурного сектора и рынка.

60. Кроме того, алгоритмы, используемые компаниями, которые занимаются потоковой передачей медийной продукции, такими как Spotify и Netflix, оказывают значительное влияние на выбор музыки и фильмов пользователями. Поскольку эти платформы не только делают произведения искусства доступными, но и *предлагают* пользователям остановить свой выбор на том или ином продукте, важно, чтобы их алгоритмы были разработаны так, чтобы не ставить в привилегированное положение одни произведения искусства в ущерб другим в результате ограничения рекомендаций наиболее востребованными произведениями определенного жанра или набором, пользующимся наибольшей популярностью у клиентов с похожими интересами. Аналогичные опасения уже высказывали представители других организаций (ARCEP, 2018). Прозрачность и подконтрольность этих алгоритмов имеют решающее значение для обеспечения доступа к различным формам художественного самовыражения и создания возможностей для активного участия в культурной жизни.

61. ИИ может сыграть важную роль и в сохранении культурного наследия. Он может использоваться, например, для мониторинга и анализа изменений на объектах культурного наследия, вызванных проблемами развития, изменением климата, стихийными бедствиями и вооруженными конфликтами. ИИ может также использоваться для отслеживания незаконного оборота культурных ценностей и разрушения культурных объектов, а также для поддержки сбора данных в целях проведения работ по их восстановлению и реконструкции.

### II.3.3 Язык

62. В нашем все более глобальном мире растущую роль, по-видимому, будет играть машинный перевод. Вследствие этого ИИ будет существенным образом влиять на язык и средства выражения человеческой мысли во всех сферах жизни. Это влечет за собой ответственность за бережное отношение к «естественным» языкам (которые следует отличать от искусственных языков и компьютерного кода) и к их разнообразию. Язык, в конечном счете, является основой человеческой идентичности, социальной гармонии, образования и развития человеческого потенциала. С момента своего основания ЮНЕСКО признает важность языка для расширения доступа к качественному образованию, построения инклюзивного общества, основанного на знаниях, и сохранения культурного наследия и форм культурного самовыражения (UNESCO, 2002).

63. Центральным элементом сложной взаимосвязи между ИИ и языком является посредническая роль «формальных языков» (языков, в которых слова строятся на основе алфавита). Технологии ИИ во многих случаях требуют, чтобы слова и предложения, сформулированные на любом из многочисленных используемых в мире естественных языков, были переведены на один из формальных языков, который может быть воспринят компьютером. Во многих случаях перевод с естественного языка на формальный не является нейтральным процессом, поскольку любой перевод такого рода приводит к смысловым «потерям», ведь отнюдь не все языковые оттенки и особенности могут быть в полной мере формализованы.

64. Еще одним элементом является перевод с одного естественного языка на другой, который осуществляется при посредстве этих формальных языков. Есть несколько объективных сложностей, с которыми сталкивается машинный перевод: в разных языках слова могут иметь разные значения, а лингвистические или понятийные соответствия между языками могут отсутствовать. В этих случаях перевод становится очень сложной, а то и технически невыполнимой задачей. Кроме того, не всегда в полной мере поддаются переводу контекстуальные и культурные коннотации слов и выражений. Несмотря на весьма значительный прогресс, достигнутый в последние годы, по крайней мере в отношении наиболее распространенных языков, автоматический или машинный перевод зачастую остается слишком ненадежным, для того чтобы его можно было использовать, например, в технических областях, где решающее значение имеет лексическая и понятийная точность, а также в культурной сфере и в литературе.

65. Эти два аспекта машинного перевода имеют важное значение не только с точки зрения качества перевода и риска непонимания при межъязыковой коммуникации, но и с точки зрения языкового разнообразия. Вполне вероятно, что машинный перевод, по крайней мере в краткосрочной перспективе, будет развиваться в первую очередь в приложении к основным мировым языкам, особенно английскому. Эта технология требует больших массивов данных, сформированных на основе переводов, выполненных людьми. Для значительного числа менее распространенных языков таких массивов данных зачастую не имеется. Вместе с тем эта технология может сыграть положительную роль, позволяя людям выражать свои мысли на менее распространенных языках.

66. Аналогичные процессы уже происходят в сфере радиовещания. В то время как коммерческие радиокomпании в основном производят передачи на широко распространенных языках, тем самым укрепляя позиции воплощенной в этих языках культуры, общинные радиостанции часто создают материалы на местных языках, тем самым стимулируя плюрализм и разнообразие в медийной сфере. Как отмечается в Руководстве ЮНЕСКО по общинным средствам массовой информации, «[общинные СМИ] получили распространение во всех регионах мира в результате стремления общественных движений и общинных организаций получить инструмент для обсуждения их проблем и забот и выражения их культурной и языковой самобытности» (UNESCO, 2013, стр. 7). Таким образом, средства массовой информации могут реально способствовать сохранению языков и культурного разнообразия.



67. Схожим образом машинный перевод уже используется в качестве инструмента поощрения разнообразия и защиты языков коренных народов. Так, в Австралии исследователем *Центра передового опыта в области языковой динамики ARC* было записано почти 50 000 часов устной речи. В ходе обработки этих записей лингвистам предстояло прослушать и расшифровать их, а также выбрать короткие фрагменты, которые могли содержать ключевые элементы грамматики и словаря. Без ИИ это заняло бы около 2 миллионов часов. К настоящему времени благодаря использованию ИИ выполнено моделирование 12 языков коренных народов Австралии, включая языки кунвок, криол, мангарайи, наккара, питьянтъятхара, варлпири и вубуй (O'Brien, 2018).

68. Эти примеры свидетельствуют о том, что ИИ, как и любая другая технология, должен развиваться и использоваться таким образом, чтобы не угрожать культурному разнообразию, а способствовать его сохранению. Если мы хотим сберечь многоязычие и функциональную совместимость различных языков, на решение этой задачи должны быть выделены соответствующие технические и финансовые ресурсы (Palfrey and Gasser, 2012; Santosuosso and Malerba, 2015).

## **II.4 Коммуникация и информация**

69. Искусственный интеллект играет все более важную роль в обработке, структурировании и предоставлении информации. Автоматизированная журналистика и алгоритмические новостные ленты в социальных сетях – лишь некоторые из примеров этой тенденции, в связи с которой возникают вопросы доступа к информации, дезинформации, дискриминации, свободы выражения мнений, неприкосновенности частной жизни, а также медийной и информационной грамотности. Внимания также требуют новые цифровые разрывы между странами и внутри различных социальных групп.

### **II.4.1 Дезинформация**

70. ИИ может способствовать свободному обмену информацией и журналистской деятельности, но он может использоваться и с целью распространения дезинформации, для обозначения которой иногда применяют спорный термин «фальшивые новости». Недавние примеры, такие как дело Cambridge Analytica, показали, что алгоритмы, которые были созданы для того, чтобы избежать свойственной человеку политической предвзятости при решении вопроса о том, какой контент будет размещен в социальных сетях на наиболее заметных местах, могут использоваться для намеренного распространения сфабрикованной, манипулятивной или сеющей рознь информации среди конкретных целевых групп. В некоторых случаях такой контент может включать информацию, в обманных целях оформленную в виде новостных сообщений, а также материалы, использующиеся как средства эмоциональной пропаганды.

71. Это может негативно сказаться на нормах цивилизованной и предметной дискуссии, общественном доверии, публичных дебатах и даже демократических процессах. Существование различных, иногда поляризованных мнений является характерной особенностью любого открытого и демократического общества, в котором существует свободное и открытое общественное пространство. Однако алгоритмы социальных сетей могут усугублять поляризацию мнений, усиливая и умножая эмоциональный контент при помощи «одобрений», «пересылки ссылок», «повторов сообщений», автозаполнения поисковых запросов и других форм онлайн-рекомендаций и инструментов вовлечения, что приводит к созданию так называемых «фильтрующих пузырей» и «эхо-камер», подменяющих собой инфраструктуру для проведения дискуссий и дебатов. Пользователи, попавшие в один и тот же «пузырь», могут подвергаться воздействию потока отфильтрованной информации, вследствие чего открытые публичные пространства будут заполняться представителями все более гомогенных групп сторонников того или иного мнения, которые при этом будут занимать все более полярные позиции по отношению друг к другу.

72. Хотя некоторые крупные компании, владеющие социальными сетями, и начинают осознавать эту проблему и необходимость ее решения с участием многих заинтересованных сторон, включая гражданское общество и органы государственного регулирования, пути выхода из сложившейся ситуации, как представляется, все еще не ясны. Одним из методов поиска решений является применение принципов R.O.A.M. (права, открытость, доступность для всех, участие многих заинтересованных сторон) ЮНЕСКО, которые могут использоваться для систематического выявления областей, в которых возможны улучшения, и для анализа того, каким образом эти улучшения увязываются со всеми другими задействованными в данных процессах принципами.

73. Иногда модерация контента может быть оправдана именно как средство, позволяющее избежать распространения дезинформации и материалов, подстрекающих к насилию, ненависти и дискриминации, а также как инструмент предотвращения агрессии в межличностной коммуникации. Фильтрация может осуществляться людьми, но нередко она выполняется с помощью алгоритмов ИИ или даже самими этими алгоритмами в автоматическом режиме. Особая проблема в данном случае заключается не только в выявлении нежелательного контента, но и в обеспечении того, чтобы фильтр действовал избирательно и не навлекал обвинений в автоматизированной цензуре и неправомерных ограничениях свободы слова. Реагирование на дезинформацию и «человеконенавистнические высказывания» должно основываться на международных стандартах свободы выражения мнений и соответствовать конвенциям и декларациям ООН по данному вопросу (Статья 19, 2018а).

#### **II.4.2 Журналистика данных и автоматизированная журналистика**

74. Недавнее появление мощных в функциональном отношении систем ИИ влечет за собой различные последствия для журналистики. С одной стороны, растущие возможности использования данных и компьютерных инструментов для решения стоящих перед журналистами исследовательских задач могут сделать их работу более эффективной. С другой стороны, ИИ может также взять на себя некоторые виды выполняемых журналистами работ. Рутинные задачи, связанные с обработкой большого объема «практических данных», находятся на первом месте в списке работ, за которые может взяться ИИ, а значительная часть журналистской работы – поиск и отбор нужных данных, обобщение результатов и их ясное изложение – по сути дела, и является рутинной. ИИ уже выполняет относительно простые задания по написанию материалов в фиксированном формате на темы, требующие регулярного обновления информации, таких как обзоры рынка и отчеты о результатах спортивных соревнований. Это применение имеет еще один аспект: оно может позволить высвободить журналистов для выполнения более сложных задач, связанных с интерпретацией, анализом, проверкой и представлением новостей.

75. Автоматизированное написание новостных материалов без участия или контроля со стороны человека – это реальность, которая часто скрыта от читателя. Еще в 2006 году некоторые службы новостей (например, Thomson Financial) объявили об использовании компьютеров для создания новостных сюжетов на основе данных, чтобы ускорить донесение информации до своей аудитории. В 2014 году компания Wibbitz (Израиль) получила Гран-при Нетэкспло на Форуме ЮНЕСКО/Нетэкспло, представив приложение, которое позволяет новостным каналам без труда создавать видеоряд, отражающий основное содержание того или иного материала, с использованием текстового контента из интернета. В последнее время многие крупные традиционные СМИ начали использовать «роботизированную журналистику»: Le Monde, Press Association, Xinhua и другие агентства сообщили о применении алгоритмов для создания текстов на естественном языке на различные темы, попавшие в поле зрения журналистов.

76. В ходе создания и распространения медийного контента аналитические задачи и функции принятия решений все чаще делегируются сложным алгоритмам. СМИ все шире используют алгоритмы, которые анализируют предпочтения пользователей и модели потребления

медийной продукции (персонализация). Кроме того, в сфере журналистики алгоритмы применяются для анализа демографических, социальных и политических характеристик отдельных географических регионов, что позволяет готовить информацию, являющуюся наиболее актуальной для местных жителей, включая прогнозы погоды и спортивные отчеты. Эта практика может помочь сохранить местную журналистику и поддержать выпуск местных газет. Таким образом ИИ может способствовать повышению эффективности бизнес-моделей, используемых в журналистике.

77. В то же время журналистика на основе ИИ поднимает вопросы ответственности, прозрачности и авторского права. Проблема ответственности может возникать в тех случаях, когда сложно определить вину за подготовленные с использованием алгоритмов материалы, например, в случаях диффамации. Проблемы прозрачности и достоверности возникают, когда пользователи не понимают или не могут понять, какой именно контент создан машиной, из каких источников он поступил и насколько достоверной или ложной является та или иная информация (крайние проявления этой тенденции стали предметом продолжающейся дискуссии о подстановке лиц известных людей в не связанные с ними ролики). На очереди также проблема авторских прав, ведь создаваемый ИИ контент все меньше зависит от вклада человека, а это дает некоторым повод утверждать, что ответственность за соблюдение авторских прав должна в той или иной форме возлагаться на сами алгоритмы.

78. Многие считают, что для решения этих проблем журналисты и редакторы должны взаимодействовать с инженерами, которые разрабатывают алгоритмы. Примером в этом отношении может служить недавний запуск открытой платформы Quartz AI Studio – американского проекта по оказанию помощи журналистам в использовании машинного обучения для решения различных задач.

## II.5 ИИ в укреплении мира и безопасности

79. В соответствии с миссией и мандатом ЮНЕСКО по содействию установлению мира в настоящем исследовании также ставится задача рассмотреть роль искусственного интеллекта в вопросах миротворчества и безопасности. Тот факт, что данная тема включает потенциально возможное использование ИИ в военных целях, ни в коей мере не умаляет приверженность авторов делу мира.

80. Считается, что ИИ способен анализировать, обрабатывать и классифицировать очень большие объемы быстро меняющихся и чрезвычайно разнородных данных (Payne, 2018; Roff, 2018; Gupta, 2018). К категории «*твердых*» данных относятся спутниковые снимки и другие материалы видеонаблюдения, сигналы и данные электронной разведки, тогда как «*мягкие*» данные могут включать отчеты, документы, новостные ленты, сообщения в социальных сетях, а также политические и социологические данные. Считается, что ИИ способен упорядочивать этот огромный объем данных и на этой основе выявлять внешние и внутренние угрозы, определять цели и стратегии различных сторон, интерпретировать сложные и разноплановые намерения, лежащие в основе их действий, и выработать стратегии упреждения прогнозируемых действий и противодействия им.

81. Такой инструмент отслеживания оперативной обстановки мог бы стать мощным средством предотвращения и урегулирования конфликтов (Spiegeleire et al., 2017). Он мог бы помочь разобраться в причинах тех или иных действий людей и их результатах, а это можно было бы использовать в целях дерадикализации. Способные обучаться системы сбора «упреждающих разведанных» могут прогнозировать назревающие общественные волнения и проявления социальной напряженности, а также предлагать способы их предотвращения. Более глубокое понимание движущих сил конфликта может принудить потенциальных зачинщиков беспорядков к отказу от реализации их злонамеренных планов. Возможно, нам удастся выявить социальные аномалии на ранней стадии, выяснить, какие действия могут способствовать деэскалации угрожающей ситуации, или найти эффективные, но при этом не провоциру-

ющие участников способы противодействия попыткам разжечь межконфессиональные распри. На уровне социума ИИ, отслеживая тенденции, которые усиливают или ослабляют стабильность в обществе, и помогая нам разобраться в них, может способствовать созданию более жизнестойкого общества и помочь нам приблизиться к более мирному и свободному от конфликтов миру.

82. Обратная сторона медали заключается в том, что *ИИ изменит характер и практику конфликтов*, что будет иметь для общества последствия, выходящие далеко за рамки сугубо военной проблематики (Payne, 2018; Spiegeleire et al., 2017). ИИ не только изменит характер применения огневой мощи путем повышения эффективности развертывания оружейных систем, но и, как ожидается, позволит значительно повысить скорость и точность всех процессов, начиная с тылового обеспечения, разведки и отслеживания оперативной обстановки и заканчивая планированием и проведением операций на поле боя. Система на базе ИИ может сама рекомендовать действия, которые следует предпринять: она может подготовить пакет приказов, использующих слабые места противника, которые выявлены на основе ее собственного анализа или поиска закономерностей в действиях противника/повстанцев, и разрабатывать меры реагирования на прогнозируемые агрессивные действия. Она может также проводить свои собственные «военные игры», чтобы выяснить, какие ответные меры могут последовать за теми или иными действиями.

83. Скорость, с которой такие инструменты планирования могут выполнять поставленные задачи, повысит эффективность действий в быстро меняющихся ситуациях. Можно предположить, что будет разработан алгоритм реагирования на скоординированную атаку, например, целого роя дронов или других беспилотных боевых единиц, таких как ракеты. Скорость реагирования с применением ИИ может быть расценена как стимул к его использованию, что может иметь дестабилизирующие последствия. Эти последствия могут даже оказаться катастрофическими, о чем свидетельствуют имевшие место в прошлом случаи, когда автоматика выдавала предупреждения, но, к счастью, вмешивался командир и принимал решение не реагировать на них. Тем не менее государство, которое не пойдет по пути создания средств реагирования на основе ИИ, окажется в крайне невыгодном положении, что будет способствовать распространению этой технологии.

84. Существует возможность того, что система принятия решений на базе ИИ будет сама предпринимать наступательные действия и принимать решения об уничтожении живой силы без вмешательства человека, например, в случае использования полностью автономного оружия. Сама мысль о том, что *неодушевленное* орудие будет наделено способностью предпринимать такие действия, может радикально изменить наши представления о политике в самом широком смысле слова. Кроме того, близость возможного военного использования ИИ-систем к их гражданскому применению («легкость оружейной конверсии») означает, что их нельзя выделить в отдельную категорию, а это усложняет как этическую сторону дела, так и регулирование разработки и применения таких систем.

85. Хотя ИИ и может рассматриваться как очередная революция в военном деле, позволяющая вооруженным силам предпринимать похожие действия с помощью похожих инструментов, возможно, его реальный «революционный» потенциал (Payne, 2018; Spiegeleire et al., 2017) заключается в способности *трансформировать само понятие «вооруженные силы»*, оснастив их чем-то куда более изощренным, чем взрывные боеприпасы. Сила ИИ в условиях конфликта заключается не только в расширении возможностей технических средств, но и в переосмыслении того, что такое «вооруженные силы».

86. Мы уже наблюдаем это явление в киберпространстве, в которое ИИ привносит как оборонительные, так и атакующие средства. Сопоставление моделей, глубинное обучение и наблюдение за отклонениями от нормального функционирования позволяют обнаруживать уязвимости программного обеспечения, которые затем могут использоваться в качестве оружия в целях преодоления средств обороны. Глубокие нейронные сети способны обнаруживать и предотвращать попытки вмешательства извне. Для того чтобы быть эффективными,

средства киберобороны должны действовать быстро и, как следствие, иметь высокую степень автономности.

87. Пропаганда – еще один вид оружия, возможности которого расширились благодаря ИИ. Легкость фальсификации голосов, изображений и новостей, а также простота донесения сфабрикованной информации до целевой аудитории создает угрозы, связанные с социальной инженерией и (недобросовестным) воздействием на общественное мнение. По сути, ИИ помогает убедительно лгать и создавать более совершенные подделки. Вытекающая из этого угроза подрыва доверия к информации повышает вероятность неверной оценки намерений предполагаемого противника как в тактическом, так и в стратегическом плане.

88. ИИ также может использоваться в качестве инструмента экономического саботажа и организации сбоев в работе важнейших объектов инфраструктуры. За счет перевода средств радиоэлектронной борьбы в когнитивный режим ИИ может сыграть решающую роль в создании препятствий для доступа к электромагнитному спектру. Уже сейчас на рынке представлены системы, использующие машинное обучение, «интеллектуальные» алгоритмы и адаптивную обработку сигналов.

89. Наконец, в сфере обеспечения *внутренней* безопасности государства использование анализа массивов данных и технологий распознавания лиц означает формирование новых отношений между обществом и институтами, отвечающими за его защиту. Очевидно, что все это приводит к существенным этическим последствиям.

## II.6 ИИ и гендерное равенство

90. Системы ИИ чреваты серьезными последствиями для гендерного равенства, поскольку они могут воспроизводить существующие в обществе предубеждения, а то и усугублять их. Большинство систем на базе ИИ построено с использованием наборов данных, которые отражают реальный мир, а этот мир может быть несовершенным, несправедливым и склонным к дискриминации (Marda, 2018). Недавно используемый компанией Amazon инструмент найма на работу был признан дискриминационным по гендерному признаку, поскольку при отборе кандидатов на технические должности он отдавал предпочтение мужчинам (Reuters, 2018). Такие системы могут быть опасными не только потому, что они способствуют сохранению гендерного неравенства в обществе, но и потому, что предвзятость заложена в них неочевидным образом, а сами системы при этом провозглашаются «объективными» и «точными» (O’Neil, 2018).

91. Такое неравенство является в первую очередь результатом того, как происходит машинное обучение. Поскольку оно опирается на данные, которые получает система, особое внимание следует уделять загрузке гендерно-чувствительных данных, а также сбору данных с разбивкой по признаку пола. В случае с инструментом найма на работу компании Amazon предвзятость возникла потому, что в ходе обучения системы использовались данные ранее нанятых работников компании, которые были преимущественно мужчинами, и система «научилась» тому, что мужчинам следует отдавать предпочтение перед кандидатами женского пола (Short, 2018). Выявление уклона данных в ту или иную сторону поможет уменьшить слепую зону и сделать так, чтобы системы ИИ в максимальной степени учитывали интересы как мужчин, так и женщин. Кроме того, использование дезагрегированных по признаку пола данных в аналитике на базе ИИ дает возможность лучше понять гендерные проблемы, с которыми мы сталкиваемся сегодня.

92. Важно отметить, что вопросы гендерного неравенства возникают уже на ранних этапах концептуальной проработки и проектирования систем ИИ. Гендерные диспропорции в технических областях хорошо известны (Hicks, 2018) и очевидным образом сказываются и на заработной плате, и на перспективах продвижения по службе (Brinded, 2017). Это явление получило название «дырявой трубы»: с момента выпуска студентов технических и инженерных специальностей до их назначения на руководящие должности в компаниях доля

женщин снижается на 40% (Wheeler, 2018). Невысокий процент женщин в сфере ИИ – и в области разработки цифровых технологий в целом – означает, что женщины не на равных участвуют в процессах принятия решений, связанных с проектированием и развитием систем ИИ. В результате возникает риск того, что эти технологии будут развиваться в интересах только некоторых групп населения (Crawford, 2016).

93. Кроме того, предубеждения, свойственные людям в их повседневной жизни, могут воспроизводиться и даже усиливаться в ходе разработки и использования систем ИИ. Так, наделение цифровых помощников гендерными признаками может способствовать укоренению представлений о том, что женщина должна быть услужлива и послушна. И действительно, женские голоса обычно дают персональным помощникам-роботам, которые в основном выполняют функции обслуживания клиентов, в то время как большинство роботов, предоставляющих профессиональные услуги, например, юридического или финансового характера, получают мужские голоса. Это имеет воспитательное значение в плане восприятия различий между «мужскими» и «женскими» компетенциями и понимания ролей начальника и подчиненного. В дополнение к этому концепция «гендера» в системах ИИ зачастую сводится к простому выбору из двух вариантов – мужчина или женщина. А это приводит к игнорированию и даже активному вытеснению трансгендеров, что может восприниматься ими как унижительная дискриминация (Costanza-Chock, 2018).

## **II.7 Африка и задачи ИИ**

94. Африка, как и другие развивающиеся регионы, сталкивается с ускоренным распространением информационных технологий и ИИ. Нарождающаяся цифровая экономика ставит перед креативным сообществом Африки важные социальные задачи и открывает перед ним широкие возможности.

95. В частности, в сфере сетевой инфраструктуры Африка испытывает острый дефицит и значительно отстает от других развивающихся регионов, большой проблемой остаются внутренние и региональные коммуникации, а также перебои в электроснабжении. Инфраструктурные услуги оплачиваются по высоким тарифам, даже несмотря на то что все больше и больше африканцев – в том числе живущих в городских трущобах – имеют собственные мобильные телефоны.

96. Проблемы развития, с которыми сталкиваются африканские страны, многочисленны. Система защиты прав человека и цели в области устойчивого развития (ЦУР) дают возможность применить последовательный подход к определению направлений развития ИИ. Так каким же образом можно организовать обмен технологиями и знаниями в области ИИ и нацелить его на выполнение приоритетных задач, определенных самими развивающимися странами? К их числу относятся задачи в области инфраструктуры, навыков, пробелов в знаниях, исследовательского потенциала и доступности местных данных, как об этом говорилось на Форуме ЮНЕСКО по вопросам искусственного интеллекта в Африке, состоявшемся в Политехническом университете им. Мохаммеда VI в Бен-Герире, Марокко, 12-13 декабря 2018 г.

97. Решающая роль в данной сфере принадлежит женщинам. Будучи в высшей степени активными участниками экономической деятельности в Африке, женщины выполняют большую часть сельскохозяйственных работ и владеют третьей частью всех предприятий, в некоторых странах их доля в общем числе наемных работников может достигать 70%. Они являются главной движущей силой экономики своих стран, вносят решающий вклад в обеспечение благосостояния своих семей и абсолютно незаменимы в качестве лидеров местного и национального масштаба. Выделив гендерному равенству центральное место в своей стратегии содействия развитию в Африке, Африканский банк развития признал основополагающую роль гендерного равенства в обеспечении инклюзивного роста и создании жизнеспособного общества. Доступ к образованию, распространение грамотности в области ИИ и в более глобальном плане освоение информационно-коммуникационных технологий (ИКТ) являются ключевыми

элементами расширения прав и возможностей женщин в целях предотвращения их маргинализации.

98. Уделяя особое внимание исследованиям, науке, технологиям, инженерному делу и математике, а также воспитанию гражданственности на основе ценностей, прав и обязанностей, ИИ следует интегрировать в национальную политику и стратегию развития с опорой на местные культурные традиции, ценности и знания в целях развития экономики африканских стран.

### III. НОРМАТИВНЫЙ ДОКУМЕНТ

#### III.1 Декларация или рекомендация

99. На основе содержащегося в двух первых разделах настоящего предварительного исследования по этике ИИ анализа рабочая группа внимательно рассмотрела два имеющихся в распоряжении ЮНЕСКО нормативных инструмента – декларацию и рекомендацию. Рабочая группа также опиралась на предыдущий опыт КОМЕСТ, которая выступила инициатором принятия Декларации этических принципов в связи с изменением климата 2017 г., а также участвовала в пересмотре Рекомендации в отношении научной деятельности и научных работников (исследователей) 2017 г. Рабочая группа сравнила преимущества и недостатки каждого из этих двух нормативных инструментов.

100. В отношении предложения о принятии Декларации об этике искусственного интеллекта рабочая группа отметила, что в 2018 г. число деклараций об этических принципах в области ИИ увеличилось. *Монреальская декларация об ответственном развитии ИИ* (University of Montreal, 2018), *Торонтская декларация о защите права на равенство и недискриминацию в системах машинного обучения* (Amnesty International and Access Now, 2018) и Декларация Института жизни будущего об *Асиломарских принципах ИИ* (Future of Life Institute, 2017) являются результатом различных инициатив и получают поддержку со стороны различных организаций (университетов, государственных органов, профессиональных ассоциаций, компаний, НПО). К этому списку деклараций мы должны добавить несколько документов по вопросам этики, находящихся в стадии подготовки, а именно *Руководящие принципы по этике добросовестных систем ИИ*, разрабатываемые группой экспертов высокого уровня по вопросам ИИ Европейской комиссии и основанные на правозащитном подходе, а также документ IEEE (в настоящее время находящийся в стадии согласования) под названием *Этически обоснованное проектирование: концепция приоритетности благополучия человека для автономных и интеллектуальных систем*, который адресован инженерам и направлен на интеграцию ценностных параметров в автономные интеллектуальные системы. Все эти инициативы носят позитивный характер, поскольку они способствуют обсуждению вопросов этики ИИ на различных уровнях.

101. Вместе с тем рабочая группа пришла к выводу о том, что принципы и формы реализации ценностей, отстаиваемых различными сторонами, характеризуются значительной разнородностью. Эта разнородность является следствием как того определения, которое было выбрано для ИИ, так и поставленных целей, будь то в сфере управления, подготовки инженеров или государственной политики. Вопрос заключается в следующем: позволит ли декларация ЮНЕСКО об этике ИИ преодолеть эту разнородность на основе нескольких руководящих принципов, которые будут в полной мере отражать этические аспекты ИИ и соответствовать конкретным задачам ЮНЕСКО в области образования, культуры, науки и коммуникации? Рабочая группа считает, что это возможно, но что при этом существует риск того, что в процессе подготовки декларации государства-члены в основном договорятся о некоторых общих, абстрактных и не имеющих обязательной силы принципах, поскольку речь идет именно о декларации. Будет ли в этих условиях декларация ЮНЕСКО об этике ИИ обладать какой-либо дополнительной ценностью по сравнению с другими уже существующими декларациями и инициативами? Сомнительно, что в условиях конкуренции между этическими нормами в период, когда технологии находятся на ранних этапах развития, а сферы их применения еще не устоялись, такой документ сразу же закрепится в качестве международного стандарта.

102. В связи с этим рабочая группа рассмотрела вопрос о том, насколько рекомендация является более подходящим инструментом в нынешней ситуации. В международном масштабе, на европейском уровне и в контексте национальной политики ряда стран наблюдается движение в сторону аналогичных форм регулирования цифровой экономики, в том числе с учетом отношений между двумя основными цифровыми державами – США и Китаем. Усиление критики, направленной против непрозрачности, предвзятости или действий крупных компаний, а также обусловленный кибератаками рост общественного недоверия создают новый политический климат, который оказывает влияние на развитие ИИ. Вследствие этого движение в поддержку цифрового регулирования, инициированное Европейским Союзом в области защиты персональных данных, может быть выведено на международный уровень в таких новых областях, как ИИ. Однако на этом уровне соответствующие инструменты все еще находятся на ранних стадиях разработки, хотя в стратегии ОЭСР, осуществляемой силами группы экспертов по искусственному интеллекту (AIGO), особое внимание уделяется вопросам ответственности, безопасности, прозрачности, защищенности и подконтрольности:

ОЭСР оказывает поддержку правительствам посредством анализа политики, диалога и сотрудничества, а также путем выявления образцов передовой практики. Мы проводим значительный объем работ по инвентаризации экономических и социальных последствий технологий и приложений на базе ИИ и их воздействия на политику. Эта деятельность охватывает совершенствование методов измерения ИИ и его воздействия, а также анализ таких важных политических вопросов, как изменения на рынке труда и навыки, востребованные в цифровую эпоху, неприкосновенность частной жизни, подконтрольность решений, принимаемых с использованием ИИ, и вопросы ответственности, безопасности и защищенности, которые возникают в связи с ИИ. (OECD, 2019)

103. Приоритеты ОЭСР в области публичной политики по большей части связаны с вопросами управления и надлежащей практики в сфере ИИ. Представляется, что подход ЮНЕСКО мог бы дополнить подход ОЭСР в международном плане, но с упором на те аспекты, которым обычно уделяется недостаточно внимания – культуру, образование, науку и коммуникацию. Эти аспекты непосредственно влияют на повседневную жизнь отдельных людей и населения в целом, а также на их индивидуальные устремления и общие чаяния. Подход ЮНЕСКО к подготовке рекомендации по этике ИИ будет представлен в качестве дополняющей альтернативы к стратегии в области экономического управления. В связи с этим рабочая группа считает, что выступив с инициативой принятия рекомендации, хотя она и потребует больше времени и энергии, чем декларация, ЮНЕСКО сможет внести уникальный вклад в эти усилия не только в плане их этического содержания, но и посредством конкретных предложений, адресованных государствам-членам. Одна из целей заключается в расширении возможностей и укреплении потенциала государств в сфере принятия мер в ключевых областях, затронутых развитием ИИ, таких как культура, образование, наука и коммуникация.

104. Рекомендация должна включать два элемента. Первый из них заключается в утверждении ряда основных принципов этики ИИ. Второй состоит в изложении конкретных предложений по оказанию государствам помощи в мониторинге и регулировании использования ИИ в областях, относящихся к мандату ЮНЕСКО, с помощью предусмотренного рекомендацией механизма отчетности, а также путем определения инструментов оценки этических аспектов, которые могли бы использоваться в ходе регулярного обзора политики государств в области управления процессами развития ИИ. В этом смысле ЮНЕСКО будет располагать уникальными возможностями для придания этой работе междисциплинарного характера и предоставления универсальной платформы для разработки рекомендации по этике ИИ. В частности, ЮНЕСКО могла бы объединить усилия развитых и развивающихся стран, сблизить различные в культурном и этическом отношении подходы и наладить взаимодействие между различными заинтересованными сторонами, представляющими как государственный, так и частный сектор, в рамках поистине международного процесса разработки всеобъемлющего свода принципов и предложений, касающихся этики ИИ.

105. В следующем разделе приводятся некоторые из этих предложений.



### III.2 Предложения по содержанию нормативного документа

106. На основе анализа потенциальных последствий использования искусственного интеллекта для общества рабочая группа хотела бы предложить ряд элементов, которые можно было бы включить в будущую рекомендацию по этике ИИ. Эти предложения отражают глобальное видение ЮНЕСКО и при этом касаются отдельных сфер ее компетенции.

107. Прежде всего, рабочая группа хотела бы предложить ряд общих принципов разработки, реализации и использования ИИ. К числу этих принципов относятся следующие:

- (a) **Права человека:** системы ИИ должны разрабатываться и реализовываться в соответствии с международными стандартами в области прав человека.
- (b) **Инклюзивность:** системы ИИ должны быть инклюзивными, нацеленными на недопущение предвзятости, обеспечение разнообразия и предотвращение нового цифрового разрыва.
- (c) **Благополучие:** системы ИИ должны разрабатываться в целях повышения качества жизни.
- (d) **Автономия:** системы ИИ должны уважать автономию человека, требуя постоянного контроля со стороны людей.
- (e) **Объяснимость:** системы ИИ должны быть объяснимыми, а принципы их работы понятными.
- (f) **Прозрачность:** данные, используемые для обучения систем ИИ, должны быть прозрачными.
- (g) **Осведомленность и грамотность:** для расширения прав и возможностей граждан необходима осведомленность в вопросах алгоритмов и понимание базовых принципов работы ИИ.
- (h) **Ответственность:** при разработке автономных интеллектуальных систем разработчики и компании должны принимать во внимание этические аспекты.
- (i) **Подотчетность:** необходимо разработать механизмы распределения ответственности за решения, принимаемые с помощью систем ИИ, и предпринимаемые ими действия.
- (j) **Демократия:** системы ИИ должны разрабатываться, реализовываться и использоваться в соответствии с демократическими принципами.
- (k) **Надлежащее управление:** правительства должны регулярно предоставлять отчеты об использовании ИИ в работе органов полиции, разведки и безопасности.
- (l) **Устойчивое развитие:** для каких целей бы ни применялся ИИ, потенциальные выгоды всегда нужно соразмерять с воздействием на окружающую среду всего производственного цикла систем ИИ и ИТ.

108. В более конкретном плане рабочая группа хотела бы указать на некоторые из ключевых этических проблем, касающихся основных направлений деятельности ЮНЕСКО:

- (a) **Образование:** системы ИИ требуют того, чтобы образование способствовало повышению уровня грамотности в области ИИ, критического мышления, адаптируемости к запросам рынка труда и подготовки инженеров в вопросах этики.

- (b) **Наука:** системы ИИ требуют ответственного внедрения в научную практику и в процесс принятия решений на базе ИИ; они должны оцениваться и контролироваться человеком и не допускать обострения структурного неравенства.
- (c) **Культура:** системы ИИ должны содействовать культурному разнообразию, инклюзивности и обогащению человеческого опыта, не допуская углубления цифрового разрыва. Следует поощрять многоязычный подход.
- (d) **Коммуникация и информация:** системы ИИ должны способствовать реализации свободы выражения мнений, обеспечению всеобщего доступа к информации, повышению качества журналистики и укреплению свободных, независимых и плюралистических средств массовой информации, не допуская при этом распространения дезинформации. Следует поощрять управление с участием многих заинтересованных сторон.
- (e) **Поддержание мира:** в целях содействия укреплению мира ИИ может использоваться для получения информации о движущих силах конфликта и ни при каких обстоятельствах не должен выполнять каких-либо действий вне контроля со стороны человека.
- (f) **Африка:** в целях развития экономики африканских стран ИИ должен быть интегрирован в национальную политику и стратегии развития с опорой на местные культурные традиции, ценности и знания.
- (g) **Гендерные вопросы:** следует избегать гендерной предвзятости при разработке алгоритмов, при формировании используемых для их обучения массивов данных и при применении алгоритмов в процессе принятия решений.
- (h) **Окружающая среда:** системы ИИ должны разрабатываться на устойчивой основе с учетом всего цикла производства систем ИИ и ИТ. ИИ может использоваться для экологического мониторинга и управления рисками, а также для предотвращения и смягчения последствий экологических кризисов.

## БИБЛИОГРАФИЯ

- AI Now. 2016. *The AI Now Report: The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term*. New York, The White House and the New York University's Information Law Institute. Available at: [https://ainowinstitute.org/AI\\_Now\\_2016\\_Report.pdf](https://ainowinstitute.org/AI_Now_2016_Report.pdf)
- Ajunwa, I., Crawford, K., and Schultz, J. 2017. Limitless Worker Surveillance. *California Law Review*. No. 735, pp. 101-142.
- Allen, G. and Chan, T. 2017. Artificial Intelligence and National Security. Harvard Kennedy School, Belfer Center for Science and International Affairs. Онлайн. Available at: <https://www.belfer-center.org/publication/artificial-intelligence-and-national-security>
- Amnesty International and Access Now. 2018. *The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems*. Toronto, RightsCon 2018. Available at: [https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration\\_ENG\\_08-2018.pdf](https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf)
- ARCEP (Autorité de régulation des communications électroniques et des postes). 2018. *Smartphones, tablets, voice assistants... Devices, the weak link in achieving an open Internet*. Paris, ARCEP. Available at: [https://www.arcep.fr/uploads/tx\\_gspublication/rapport-terminaux-fev2018-ENG.pdf](https://www.arcep.fr/uploads/tx_gspublication/rapport-terminaux-fev2018-ENG.pdf)
- Article 19. 2018a. Free speech concerns amid the “fake news” fad. Онлайн. Available at: <https://www.article19.org/resources/free-speech-concerns-amid-fake-news-fad/>
- Article 19. 2018b. Privacy and Freedom of Expression in the Age of Artificial Intelligence. Онлайн. Available at: <https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf>
- Ashley, K.D. 2017. *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. Cambridge, Cambridge University Press.
- Boden, M.A. 2016. *AI: Its Nature and Future*. Oxford, Oxford University Press.
- Brinded, L. 2017. “Robots are going to turbo charge one of society’s biggest problems”, QUARTZ (28 December 2017). Онлайн. Available: <https://qz.com/1167017/robots-automation-and-ai-in-the-workplace-will-widen-pay-gap-for-women-and-minorities/>
- Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B. and Anderson, H. 2018. *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. Available at: <https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf>
- Bunnin, N. and Yu, J. 2008. *The Blackwell dictionary of western philosophy*. John Wiley & Sons.
- Butterfield, A., Ngondi, G.E. and Kerr, A. eds. 2016. *A dictionary of Computer Science*. Oxford, Oxford University Press.
- Costanza-Chock, S. 2018. “Design justice, AI, and escape from the matrix of domination”, *Journal of Design and Science*. Онлайн. Available at: <https://jods.mitpress.mit.edu/pub/costanza-chock>
- Crawford, K. 2016. “Artificial Intelligence’s White Guy Problem”, *The New York Times* (Opinion, 25 June 2016). Онлайн. Available at: <https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html>

Crawford, K. 2017. 'The Trouble with Bias', NIPS 2017 Keynote. Available at: [https://www.youtube.com/watch?v=fMym\\_BKWQzk](https://www.youtube.com/watch?v=fMym_BKWQzk)

Cummings, M. L., Roff, H. M., Cukier, K., Patakilas, J. and Bryce, H. 2018. *Artificial Intelligence and International Affairs: Disruption Anticipated*. Chatham House Report. Available at: <https://www.chathamhouse.org/sites/default/files/publications/research/2018-06-14-artificial-intelligence-international-affairs-cummings-roff-cukier-parakilas-bryce.pdf>

Brookfield Institute and Policy Innovation Hub (Ontario). 2018. Policymakers: Understanding the Shift. Онлайн. Available at: [https://brookfieldinstitute.ca/wp-content/uploads/Brookfield-Institute\\_-The-AI-Shift.pdf](https://brookfieldinstitute.ca/wp-content/uploads/Brookfield-Institute_-The-AI-Shift.pdf)

Eubanks, V. 2018a. "A Child Abuse Prediction Model Fails Poor Families", WIRED. Онлайн. Available at: <https://www.wired.com/story/excerpt-from-automating-inequality/>

Eubanks, V. 2018b. *Automating Inequality: How high tech tools profile, police, and punish the poor*. New York, St. Martin's Press.

European Commission (EC). 2018. *Artificial Intelligence for Europe*. Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Brussels, European Commission. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN>

European Commission for the Efficiency of Justice (CEPEJ). 2018. *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*. Strasbourg, CEPEJ. Available at: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>

European Group on Ethics in Science and New Technologies (EGE). 2018. *Statement on AI, Robotics, and Autonomous System*. Brussels, European Commission. Available at: <https://publications.europa.eu/en/publication-detail/-/publication/dfbe62e-4ce9-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-78120382>

Executive Office of the President (USA). 2016. *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights*. Washington, D.C., Executive Office of the President. Available at: [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016\\_0504\\_data\\_discrimination.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf)

Frankish, K. and Ramsey, W.M. eds. 2014. *The Cambridge handbook of artificial intelligence*. Cambridge, Cambridge University Press.

Future of Life Institute. 2017. *Asilomar AI Principles*. Cambridge, Future of Life Institute. Available at: <https://futureoflife.org/ai-principles/?cn-reloaded=1>

Gupta, D.K. 2018. "Military Applications of Artificial Intelligence", Indian Defence Review (22 March 2019). Онлайн. Available at: <http://www.indiandefencereview.com/military-applications-of-artificial-intelligence/>

Heacock, M., Kelly, C.B., Asante, K.A., Birnbaum, L.S., Bergman, Å.L., Bruné, M.N., Buka, I., Carpenter, D.O., Chen, A., Huo, X. and Kamel, M. 2015. "E-waste and harm to vulnerable populations: a growing global problem", *Environmental health perspectives*, Vol. 124, No. 5, pp. 550-555.

Hicks, M. 2018. "Why tech's gender problem is nothing new", The Guardian (12 October 2018). Онлайн. Available at: [https://amp.theguardian.com/technology/2018/oct/11/tech-gender-problem-amazon-facebook-bias-women?\\_\\_twitter\\_impession=true](https://amp.theguardian.com/technology/2018/oct/11/tech-gender-problem-amazon-facebook-bias-women?__twitter_impession=true)

Hinchliffe, T. 2018. "Medicine or poison? On the ethics of AI implants in humans", The Sociable. Онлайн. Available at: <https://sociable.co/technology/ethics-ai-implants-humans/>

- House of Lords. 2017. *AI in the UK: ready, willing and able?* London, House of Lords Select Committee on Artificial Intelligence. Available at: <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>
- Illanes, P., Lund, S., Mourshed, M., Rutherford, S. and Tyreman, M. 2018. *Retraining and reskilling workers in the age of automation*. Online, McKinsey Global Institute. Available at: <https://www.mckinsey.com/featured-insights/future-of-work/retraining-and-reskilling-workers-in-the-age-of-automation>
- Institute of Electrical and Electronic Engineers (IEEE). 2018. *Ethically Aligned Design – Version 2 for Public Discussion*. New Jersey, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Available at: <https://ethicsinaction.ieee.org/>
- Laplante, P.A. 2005. *Comprehensive dictionary of electrical engineering*. Boca Raton, CRC Press.
- Latonero, M. 2018. *Governing Artificial Intelligence: Upholding Human Rights & Dignity*. Data & Society. Available at: [https://datasociety.net/wp-content/uploads/2018/10/DataSociety\\_Governing\\_Artificial\\_Intelligence\\_Upholding\\_Human\\_Rights.pdf](https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf)
- Marda, V. 2018. “Artificial Intelligence Policy in India: A Framework for Engaging the Limits of Data-Driven Decision-Making”, *Philosophical Transactions A: Mathematical, Physical and Engineering Sciences*. Онлайн. Available at: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3240384](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3240384)
- Matias, Y. 2018. Keeping people safe with AI-enabled flood forecasting. *The Keyword* (24 September 2018). Онлайн. Available at: <https://www.blog.google/products/search/helping-keep-people-safe-ai-enabled-flood-forecasting/>
- Matsumoto, D.E. 2009. *The Cambridge dictionary of psychology*. Cambridge, Cambridge University Press.
- McCarthy, J., Minsky, M. L., Rochester, N., Shannon, C. E. 2006 [1955]. “A proposal for the Dartmouth Summer Research Project on Artificial Intelligence”, *AI Magazine*, vol. 27, no. 4, pp.12-14.
- Microsoft Europe. 2016. “The Next Rembrandt”, *Microsoft News Centre Europe*. Онлайн. Available at: <https://news.microsoft.com/europe/features/next-rembrandt/>
- National Science and Technology Council (USA). 2016. *The National Artificial Intelligence Research and Development Strategic Plan*. Washington, D.C., National Science and Technology Council. Available at: [https://www.nitrd.gov/PUBS/national\\_ai\\_rd\\_strategic\\_plan.pdf](https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf)
- O'Brien, A. 2018. “How AI is helping preserve Indigenous languages”, *SBS News*. Онлайн. Available at: <https://www.sbs.com.au/news/how-ai-is-helping-preserve-indigenous-languages>
- O'Neil, C. 2018. “Amazon’s Gender-Biased Algorithm Is Not Alone”, *Bloomberg Opinion* (16 October 2018). Онлайн. Available at: <https://www.bloomberg.com/opinion/articles/2018-10-16/amazon-s-gender-biased-algorithm-is-not-alone>
- ОЭСР. 2019. *Going Digital*. Paris, OECD. Available at: <http://www.oecd.org/going-digital/ai/>
- Oppenheimer, A. 2018. *¡Sálvese quien pueda!: El futuro del trabajo en la era de la automatización*. New York, Vintage Espanol.
- Palfrey, J.G. and Gasser, U. 2012. *Interop: The Promise and Perils of Highly Interconnected Systems*. New York, Basic Books.
- Payne, K. 2018. “Artificial Intelligence: A Revolution in Strategic Affairs?”, *Survival*, Vol. 60, No. 5, pp. 7-32.

Peiser, J. 2019. “The Rise of the Robot Reporter”, The New York Times (5 February 2019). Онлайн. Available at: <https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html>

Reuters. 2018. “Amazon ditched AI recruiting tool that favored men for technical jobs”, The Guardian (11 October 2018). Онлайн. Available at: <https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>

Roff, H.M. 2018. “COMPASS: a new AI-driven situational awareness tool for the Pentagon?”, Bulletin of the Atomic Scientists (10 May 2018). Онлайн. Available at: <https://thebulletin.org/2018/05/compass-a-new-ai-driven-situational-awareness-tool-for-the-pentagon/>

Rosenberg, J.M. 1986. *Dictionary of artificial intelligence and robotics*. New York, John Wiley & Sons.

Russell, S.J. and Norvig, P. 2016. *Artificial Intelligence: A Modern Approach*, 3<sup>rd</sup> ed. Harlow, Pearson.

Santosuosso, A. and Malerba, A. 2015. “Legal Interoperability As a Comprehensive Concept in Transnational Law”, *Law, Innovation and Technology*, Vol. 5, No. 1, pp. 51-73.

Short, E. 2018. “It turns out Amazon’s AI hiring tool discriminated against women”, Siliconrepublic (11 October 2018). Онлайн. Available at: <https://www.siliconrepublic.com/careers/amazon-ai-hiring-tool-women-discrimination>

Spiegeleire, S. De, Maas, M. and Sweijs, T. 2017. *Artificial Intelligence and the Future of Defence*. The Hague, The Hague Centre for Strategic Studies.

Международная сеть профессиональных союзов. 2016. Top 10 principles for ethical artificial intelligence. Switzerland, UNI Global Union. Available at: [http://www.thefutureworldofwork.org/media/35420/uni\\_ethical\\_ai.pdf](http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf)

ЮНИСЕФ. 2017. Children in a Digital World. New York UNICEF. Available at: [https://www.unicef.org/publications/files/SOWC\\_2017\\_ENG\\_WEB.pdf](https://www.unicef.org/publications/files/SOWC_2017_ENG_WEB.pdf)

United Nations Educational, Scientific and Cultural Organization (UNESCO). 2002. *UNESCO Universal Declaration on Cultural Diversity: a vision, a conceptual platform, a pool of ideas for implementation, a new paradigm*. Paris, UNESCO. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000127162>

ЮНЕСКО. 2013. Community Media: A Good Practice Handbook. Paris, UNESCO. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000215097>

ЮНЕСКО. 2015a. Keystones to foster inclusive knowledge societies: access to information and knowledge, freedom of expression, privacy and ethics on a global internet. Paris, UNESCO. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000232563>

ЮНЕСКО. 2015b. Outcome document of the “CONNECTing the Dots: Options for Future Action” Conference. Paris, UNESCO. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000234090>

University of Montreal. 2018. *Montreal Declaration for a Responsible Development of AI*. Montreal, University of Montreal. Available at: <https://www.montrealdeclaration-responsibleai.com/>

Vernon, D. 2014. *Artificial cognitive systems: A primer*. Cambridge, MIT Press.

Villani, C., Schoenauer, M., Bonnet, Y., Berthet, C., Cornut, A.-C., Levin, F. and Rondepierre, B. 2018. *For A Meaningful Artificial Intelligence: Towards a French and European Strategy*. Paris. Available at: [https://www.aiforhumanity.fr/pdfs/MissionVillani\\_Report\\_ENG-VF.pdf](https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf)

Wheeler, T. 2018. "Leaving at Lightspeed : the number of senior women in tech is decreasing", OECD Forum (23 March 2018). Онлайн. Available: <https://www.oecd-forum.org/users/91062-tarah-wheeler/posts/31567-leaving-at-lightspeed-the-number-of-senior-women-in-tech-is-decreasing>

World Summit on the Information Society (WSIS). 2003. *Declaration of Principles. Building the Information Society: A global challenge in the new Millenium*. Geneva, WSIS. Available at: <http://www.itu.int/net/wsis/docs/geneva/official/dop.html>

ВВИО. 2005. Тунисская программа для информационного общества. Tunis, WSIS. Available at: <http://www.itu.int/net/wsis/docs2/tunis/off/6rev1.html>

**ANNEX: COMPOSITION OF THE COMEST EXTENDED WORKING GROUP  
ON ETHICS AND AI**

1. **Prof. (Mr) Peter-Paul VERBEEK (Co-Coordinator)**  
Professor of Philosophy of Technology at the University of Twente, Netherlands  
Member of COMEST (2016-2019)
2. **Prof. (Mrs) Marie-Hélène PARIZEAU (Co-Coordinator)**  
Professor, Faculty of Philosophy, Université Laval, Québec, Canada  
Member of COMEST (2012-2019)  
Chairperson (2016-2019) and Vice-Chairperson (2014-2015) of COMEST
3. **Prof. (Mr) Tomislav BRACANOVIĆ**  
Research Associate, Institute of Philosophy, Zagreb, Croatia  
Member of COMEST (2014-2021)  
Rapporteur of COMEST (2018-2019)
4. **Mr John FINNEY**  
Emeritus Professor of Physics, Department of Physics and Astronomy, London, United Kingdom  
Coordinator of the Working Group on Scientific Ethics, Pugwash Conference on Science and World Affairs  
Ex-officio Member of COMEST
5. **Mr Javier JUAREZ MOJICA**  
Commissioner, Board of the Federal Telecommunications Institute of Mexico, Mexico City, Mexico  
Member, OECD Expert Group on AI (AIGO)  
Member of COMEST (2018-2021)
6. **Mr Mark LATONERO**  
Research Lead, Data and Human Rights, Data & Society, United States of America
7. **Ms Vidushi MARDA**  
Digital Programme Officer at ARTICLE 19 (Ms Marda is based in India.)
8. **Prof. (Ms) Hagit MESSER-YARON**  
Professor of Electrical Engineering and former Vice-President for Research and Development, University of Tel Aviv, Tel Aviv, Israel  
Member, Executive Committee, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems  
Member of COMEST (2016-2019)
9. **Dr (Mr) Luka OMLADIC**  
Lecturer, University of Ljubljana, Ljubljana, Slovenia  
Member of COMEST (2012-2019)
10. **Prof. (Mrs) Deborah OUGHTON**  
Professor and Research Director, Centre for Environmental Radioactivity, Norwegian University of Life Sciences  
Member of COMEST (2014-2021)



11. **Prof. (Mr) Amedeo SANTOSUOSSO**  
Founder and Scientific Director, European Center for Law, Science and new Technologies (ECLT), University of Pavia, Pavia, Italy,  
President, First Chamber, Court of Appeal of Milan, Italy  
Member of COMEST (2018-2021)
12. **Prof. (Mr) Abdoulaye SENE**  
Environmental sociologist, Coordinator for “Ethics, Governance, Environmental and Social Responsibility”, Environmental Sciences Institute, Cheikh Anta Diop University, Dakar, Senegal  
Member of COMEST (2012-2019)  
Vice-Chairperson of COMEST (2016-2019)
13. **Prof. (Mr) John SHAWE-TAYLOR**  
UNESCO Chair in Artificial Intelligence, University College of London and Chair of the Knowledge 4 All Foundation, United Kingdom
14. **Mr Davide STORTI**  
Programme Specialist, Section for ICT in Education, Science and Culture, Communication and Information Sector, UNESCO
15. **Prof. (Mr) Sang Wook YI**  
Professor of Philosophy, Hanyang University, Seoul, Republic of Korea  
Member of COMEST (2018-2021)