



Организация  
Объединенных Наций по  
вопросам образования,  
науки и культуры

## Специальная группа экспертов (СГЭ) по подготовке проекта рекомендации об этических аспектах искусственного интеллекта

Рассылается по списку

SHS/BIO/AHEG-AI/2020/4

Париж, 7 мая 2020 г.

Оригинал: английский

### Итоговый документ:

## Первоначальный вариант проекта рекомендации об этических аспектах искусственного интеллекта

В соответствии с решением 40-й сессии Генеральной конференции ЮНЕСКО ([резолуция 40 С/37](#)) в марте 2020 г. Генеральный директор учредил Специальную группу экспертов (СГЭ) для подготовки проекта рекомендации об этических аспектах искусственного интеллекта.

С учетом сложной ситуации, вызванной пандемией коронавируса COVID-19, работа СГЭ, проходившая в период с конца марта по начало мая 2020 г., была организована в онлайн-режиме и завершилась подготовкой представленного в настоящем документе первоначального варианта проекта рекомендации об этических аспектах искусственного интеллекта.

Следует подчеркнуть, что до начала сентября 2020 г. СГЭ продолжит вносить в первоначальный проект поправки с учетом замечаний и комментариев, которые будут получены в ходе многосторонних консультаций, запланированных на июнь-июль 2020 г.

Представленный вариант не претендует на исчерпывающий характер и не обязательно отражает мнение всех государств – членов ЮНЕСКО.

## ПЕРВОНАЧАЛЬНЫЙ ВАРИАНТ ПРОЕКТА РЕКОМЕНДАЦИИ ОБ ЭТИЧЕСКИХ АСПЕКТАХ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

### ПРЕАМБУЛА

Генеральная конференция Организации Объединенных Наций по вопросам образования, науки и культуры (ЮНЕСКО) на своей 41-й сессии, проходившей в Париже с ... по ... ноября 2021 года,

**напоминая**, что в соответствии с положениями своего Устава ЮНЕСКО стремится укоренять в сознании людей идею защиты мира и ставит своей задачей содействовать сотрудничеству народов в области образования, науки, культуры, коммуникации и информации в интересах поощрения всеобщего уважения справедливости, законности и прав человека, а также основных свобод, провозглашенных для всех народов,

**размышляя** о серьезном воздействии, которое искусственный интеллект (ИИ) способен оказать на общество, среду обитания и жизнь человека, в том числе на его интеллектуальные способности, отчасти обусловленном появлением новых технологий, позволяющих воздействовать при помощи ИИ на процессы мышления и принятия человеком решений, а также на сферу образования, науки, культуры, коммуникации и информации,

**принимая во внимание**, что системы искусственного интеллекта способны принести человечеству огромную пользу, но при этом ставят перед ним фундаментальные вопросы этического порядка, касающиеся, в частности, предвзятости, которую такие системы могут порождать и усугублять, что потенциально может вести к неравенству и маргинализации и представлять угрозу с точки зрения культурного и социального разнообразия, а также гендерного равенства; необходимость обеспечения прозрачности и объяснимости работы алгоритмов и данных, на основе которых проводится обучение интеллектуальных систем; и потенциальные последствия их применения с точки зрения неприкосновенности частной жизни, свободы слова, участия в социально-экономических и политических процессах, а также окружающих условий,

**признавая**, что развитие ИИ-технологий может усугубить существующее в мире расслоение и неравенство и что никто против собственной воли не должен остаться забытым ни с точки зрения возможности получения выгод от использования ИИ, ни в плане защиты от негативных последствий его применения, и при этом признавая различие существующих в разных странах условий,

**сознавая** тот факт, что страны с низким и средним уровнем дохода, включая, в частности, страны Африки, Латинской Америки и Карибского бассейна, государства Центральной Азии, а также малые островные развивающиеся государства, имеют дело с ускорением темпов внедрения информационных технологий и ИИ и что цифровая экономика ставит перед обществами сложные задачи и открывает широкие возможности для развития их творческого потенциала, требующего учета культурного опыта, жизненных установок и знаний коренных народов в интересах развития экономик,

**соглашаясь** с тем, что ИИ-технологии способны благотворно воздействовать на природную среду посредством потенциала, которым они обладают в области экологических и климатических исследований, управления рисками стихийных бедствий и в сфере сельского хозяйства, но что для реализации этого потенциала необходим справедливый доступ к научно-техническим знаниям и что соответствующие потенциальные выгоды должны соизмеряться с воздействием на экологию в рамках всего цикла производства интеллектуальных систем и технических средств обработки информации,

**подчеркивая**, что учет рисков и этических аспектов не должен препятствовать нововведениям, а напротив, должен поощрять новый ответственный подход к научно-исследователь-

ской и инновационной деятельности, в которой процесс исследования, проектирования, создания, развертывания и применения ИИ основывается на нравственных принципах и этических соображениях,

**напоминая**, что в ноябре 2019 года Генеральная конференция ЮНЕСКО на своей 40-й сессии приняла резолюцию 40 С/37, в которой она уполномочила Генерального директора инициировать «разработку международного нормативного акта по этическим аспектам искусственного интеллекта (ИИ) в форме рекомендации», который должен быть представлен Генеральной конференции на ее 41-й сессии в 2021 году,

**будучи убеждены** в том, что настоящий нормативный документ, подготовленный на основе глобального нормотворческого подхода с акцентом на уважительном отношении к человеческому достоинству и правам человека, включая принципы разнообразия, взаимозависимости, инклюзивности и справедливости, сможет направить процесс исследования, проектирования, создания, развертывания и применения ИИ в ответственное русло,

**отмечая**, что основы правового регулирования в сфере ИИ-технологий и возможные социальные последствия их применения обусловили позиционирование ИИ на стыке этики, прав человека, международного и национального законодательства, свободы исследовательской и инновационной деятельности и благополучия человека,

**соглашаясь** с тем, что этические принципы и ценности сами по себе не обязательно являются правовыми нормами, однако могут в значительной степени влиять на выработку и осуществление мер политики и правовых норм, выступая в качестве ориентиров в тех случаях, когда сфера действия таких норм недостаточно ясна или когда такие нормы еще не приняты вследствие высоких темпов технологического развития, сочетающегося с относительно медленным реагированием на уровне мер политики,

**будучи убеждены** в том, что согласованные на глобальном уровне этические стандарты могут способствовать повсеместной гармонизации правовых норм, касающихся ИИ, а также ответственному применению норм действующего международного права, при условии, что такое применение не противоречит рамочным этическим принципам и не причиняет вреда на местном уровне,

**признавая** положения Всеобщей декларации прав человека (1948 г.), в том числе ее статьи 27, в которой особо подчеркнуто право человека участвовать в научном прогрессе и пользоваться его благами, положения международных нормативно-правовых документов в области прав человека, в том числе Конвенции Организации Объединенных Наций о ликвидации всех форм дискриминации в отношении женщин (1979 г.), Конвенции Организации Объединенных Наций о правах ребенка (1989 г.) и Конвенции Организации Объединенных Наций о правах инвалидов (2006 г.), а также положения Конвенции ЮНЕСКО об охране и поощрении разнообразия форм культурного самовыражения (2005 г.),

**напоминая** о Декларации ЮНЕСКО об ответственности нынешних поколений перед будущими поколениями (1997 г.), Декларации Организации Объединенных Наций о правах коренных народов (2007 г.), докладе Генерального секретаря Организации Объединенных Наций о последующей деятельности по итогам второй Всемирной ассамблеи по проблемам старения (A/66/173) (2011 г.), посвященном положению в области прав человека пожилых людей, докладе Специального представителя Генерального секретаря Организации Объединенных Наций по вопросу о правах человека и транснациональных корпорациях и других предприятиях (A/HRC/17/31) (2011 г.), в котором кратко представлен документ «Руководящие принципы предпринимательской деятельности в аспекте прав человека: осуществление рамок Организации Объединенных Наций в отношении «защиты, соблюдения и средств правовой защиты», о принятой 26 сентября 2019 г. Советом по правам человека резолюции по вопросу о праве на неприкосновенность частной жизни в эпоху цифровых технологий (A/HRC/RES/42/15), о

Рекомендации ЮНЕСКО в отношении научной деятельности и научных работников (исследователей) (2017 г.), о показателях ЮНЕСКО, касающихся универсальности интернета (2019 г.), в том числе о принципах П.О.Д.М., о докладе под названием «Век цифровой взаимозависимости» (2019 г.), подготовленном созданной Генеральным секретарем Организации Объединенных Наций Группой высокого уровня по цифровому сотрудничеству, а также о программных материалах и докладах проводимых МСЭ Всемирных саммитов «Искусственный интеллект во благо»,

**напоминая также** о разработанных другими межправительственными организациями рамочных положениях, регламентирующих этические аспекты ИИ, в частности о соответствующих документах по правам человека и других правовых актах, принятых Советом Европы, а также о деятельности Специального комитета СЕ по искусственному интеллекту (СКИИ), о связанной с ИИ деятельностью Европейского союза и об учреждении Европейской комиссией Группы экспертов высокого уровня по искусственному интеллекту, в том числе о подготовленном ею документе «Руководящие принципы создания заслуживающего доверия, этичного искусственного интеллекта», о работе Группы экспертов ОЭСР по искусственному интеллекту (ГЭО-ИИ) и принятой Советом ОЭСР Рекомендации по искусственному интеллекту, о разработанных на основе указанной рекомендации и изложенных в Заявлении министров торговли и цифровой экономики «Группы двадцати» «Принципах «Группы двадцати» в области развития искусственного интеллекта», о принятом на саммите «Группы семи» в Шарлевуа заявлении под названием «Общее видение будущего искусственного интеллекта», о деятельности Рабочей группы Африканского союза по искусственному интеллекту и Рабочей группы Лиги арабских государств по искусственному интеллекту,

**подчеркивая** необходимость уделения особого внимания странам с низким и средним уровнем дохода, включая, в частности, страны Африки, Латинской Америки и Карибского бассейна, государства Центральной Азии, а также малые островные развивающиеся государства, с учетом их недостаточной представленности в дискуссиях, посвященных этическим аспектам ИИ, что дает повод для упреков в игнорировании местного опыта и знаний, принципов культурного и этического плюрализма, систем ценностей и требований глобальной справедливости,

**будучи в курсе** разработки многими государствами национальных рамок, касающихся этических аспектов и регулирования связанной с ИИ деятельности,

**будучи также в курсе** многочисленных инициатив и рамочных руководящих документов по этическим аспектам искусственного интеллекта, разработанных частными структурами, профессиональными и неправительственными организациями, в частности таких, как Глобальная инициатива ИИЭР по этике автономных интеллектуальных систем и ее «Рекомендации по этически обоснованному проектированию», экспертный доклад Всемирного экономического форума по теме «Глобальное управление в сфере технологий: многосторонний подход», разработанный Международной сетью профессиональных союзов документ «Главные десять принципов этичного применения искусственного интеллекта», Монреальская декларация об ответственном развитии искусственного интеллекта, инициатива «Принципы гармоничного использования искусственного интеллекта» (ПГИИ) и инициатива «Принципы партнерства в области искусственного интеллекта»,

**будучи убеждены** в том, что применение ИИ может принести значительные выгоды, но также и в том, что получение этих выходов может быть отягощено задолженностью в сфере инновационной деятельности и сопряжено с перекосами в доступе к научным знаниям, сложностями с реализацией права на получение информации и дефицитом потенциала креативности в рамках циклов разработки, ограниченностью кадровых и институциональных возможностей, препятствиями в плане доступа к технологическим инновациям и отсутствием адекватной материально-технической и нормативной базы для работы с данными,

**отмечая** экономическую конкуренцию на внутри- и межстрановом уровнях, а также на уровне транснациональных компаний, обуславливающую нацеленность стратегий и нормативно-правовой базы в области ИИ на обеспечение национальных и коммерческих интересов, в то время как для решения сложных задач, которые ИИ выносит на повестку дня в условиях многообразия и взаимосвязанности культур и этических систем, а также для минимизации последствий потенциального ненадлежащего использования ИИ-технологий необходимо сотрудничество в глобальном масштабе,

**в полной мере учитывая**, что стремительное развитие интеллектуальных систем наталкивается на сложности на этапе их осмысления и внедрения, что обусловлено разнообразием существующих в мире нравственных установок и культур, недостаточно быстрой адаптацией правовых систем к техническому прогрессу и реалиям информационного общества, а также потенциальной угрозой слома в результате применения ИИ-технологий существующих на региональном и местном уровнях нравственных устоев и ценностных ориентиров,

1. **принимает** настоящую Рекомендацию об этических аспектах искусственного интеллекта;
2. **рекомендует** государствам-членам применять на практике положения настоящей Рекомендации путем принятия надлежащих мер, в том числе, при необходимости, законодательного или иного характера, направленных на претворение в жизнь в пределах своей юрисдикции сформулированных в Рекомендации принципов и норм, с учетом конституционной практики и особенностей структуры управления каждого государства;
3. **рекомендует также** государствам-членам довести настоящую Рекомендацию до сведения органов власти и управления, государственных, коммерческих и некоммерческих учреждений и организаций, занимающихся исследованием, проектированием, созданием, развертыванием и использованием систем искусственного интеллекта.

## I. СФЕРА ПРИМЕНЕНИЯ

1. В настоящей Рекомендации рассматриваются этические аспекты развития искусственного интеллекта. Используемый подход предлагает рассматривать этическую составляющую ИИ в качестве целостной системы взаимозависимых ценностных установок, принципов и процедур, способных направлять организованные сообщества на всех этапах жизненного цикла интеллектуальной системы, рассматривая человеческое достоинство и благополучие человека в качестве целевых ориентиров для ответственного учета возможных известных и неизвестных последствий применения ИИ-систем в процессе их взаимодействия с людьми и внешней средой. Под жизненным циклом интеллектуальной системы понимаются этапы соответственно исследования, проектирования, создания, развертывания и применения ИИ-системы, при этом термин «применение» может означать техническое обслуживание, эксплуатацию, вывод из эксплуатации и демонтаж интеллектуальной системы. При разработке данного документа задача предложить единое определение искусственного интеллекта не ставилась, поскольку по мере технического прогресса время от времени потребовалось бы внесение в такое определение изменений. Цель, скорее, заключалась в том, чтобы учесть особенности ИИ-систем, имеющие важную этическую составляющую, и в отношении которых существует широкий международный консенсус. В контексте настоящей Рекомендации системы искусственного интеллекта рассматриваются как технические системы, способные обрабатывать информацию способом, напоминающим разумное поведение, которое обычно включает в себя такие аспекты, как обучение, распознавание, прогнозирование, планирование или контроль. Данная Рекомендация рассматривает интеллектуальные системы следующим образом:

- (a) Во-первых, ИИ-системы характеризуются использованием моделей и алгоритмов, обеспечивающих способность обучения и решения когнитивных задач, таких как формулирование рекомендаций или принятие решений в реальной и виртуальной

среде. Интеллектуальные системы способны функционировать с различной степенью автономности путем моделирования знаний и представления полученных результатов, а также использования данных и расчета корреляционных зависимостей. В ИИ-системах используется несколько подходов и технологий, в том числе такие, как:

- (i) самообучение системы, включающее глубокое обучение и обучение с подкреплением;
  - (ii) автоматизированное рассуждение, которое включает планирование, диспетчеризацию, представление знаний, поиск и оптимизацию;
  - (iii) кибер-физические системы, в том числе интернет вещей и робототехника, объединяющие в себе функции контроля, распознавания, обработки данных, собранных датчиками, а также работу исполнительных элементов в среде функционирования ИИ-систем.
- (b) Во-вторых, наряду с аспектами этического характера, аналогичными тем, что возникают в рамках применения любой технологии, системы искусственного интеллекта ставят перед человечеством вопросы нового типа. Некоторые из них обусловлены тем, что интеллектуальные системы оказались способны делать то, что раньше было под силу только живым существам, а в ряде случаев только человеку. Эти особенности позволили возложить на ИИ-системы новые серьезные задачи по обеспечению жизнедеятельности человека и общества. Если попытаться заглянуть дальше, то в долгосрочной перспективе системы искусственного интеллекта смогут соперничать с человеком с точки зрения своей способности осмысления человеческого опыта и моделирования человеческого сознания, что порождает дополнительные вопросы в отношении реальной самостоятельности, уникальности и величия человека, хотя на сегодняшний день вопрос пока так не ставится.
- (c) В-третьих, несмотря на то, что связанные с ИИ вопросы этического характера касаются в основном конкретного воздействия ИИ-систем на человека и человеческое общество, существует и другой комплекс этических вопросов, относящихся к взаимодействию человека с интеллектуальными системами, а также к возможным последствиям, которые такое взаимодействие может оказать на восприятие нами как технологий, так и самого человека. В Рекомендации признается, что оба вида этических вопросов тесно взаимосвязаны и являются неотъемлемыми компонентами этического подхода к ИИ.

2. Отдельное внимание в настоящей Рекомендации уделено более широким по охвату этическим последствиям применения ИИ в основных областях деятельности ЮНЕСКО, а именно в образовании, науке, культуре, коммуникации и информации, которые были проанализированы в подготовленном в 2019 году Всемирной комиссией ЮНЕСКО по этике научных знаний и технологий (КОМЕСТ) Предварительном исследовании этических аспектов искусственного интеллекта:

- (a) Системы искусственного интеллекта связаны с образованием в целом ряде аспектов: в силу последствий, которые их развитие может иметь для ситуации на рынке труда и возможностей трудоустройства, они подрывают социально значимую роль образования; они способны повлиять на традиционные методы обучения; они базируются на том предположении, что подготовка специалистов по искусственному интеллекту и программистов обеспечивает осведомленность о социальных и этических последствиях развития ИИ.

- (b) Во всех областях естественных, социальных и гуманитарных наук применение ИИ-технологий влияет на общие подходы в вопросах научного понимания и объяснения, а также на способы использования научных знаний в качестве основы для принятия решений.
- (c) Развитие искусственного интеллекта влияет на культурную самобытность и разнообразие культур. ИИ способен позитивно воздействовать на индустрию культуры и творчества, но, с другой стороны, может стать фактором концентрации производства культурного контента, данных и, следовательно, доходов в руках весьма ограниченного круга участников рынка, что потенциально негативно отразится на разнообразии форм культурного самовыражения и обеспечении принципа равенства.
- (d) В сфере коммуникации и информации все более значимую роль, с большой долей вероятности, будет играть машинный перевод. Это может существенным образом повлиять на язык как средство человеческого самовыражения во всех сферах жизни, что налагает на всех обязанность внимательного отношения к языкам и проблематике сохранения их разнообразия. Кроме того, ИИ-технологии бросают вызов традиционным методам журналистики и подрывают социально значимую роль журналистов, работников СМИ и занимающихся журналистикой авторов материалов в социальных сетях; при этом интеллектуальные системы используются как для распространения недостоверной информации и введения в заблуждение, так и для выявления подобной практики.

3. Настоящая Рекомендация адресована государствам. В то же время в случае необходимости и актуальности она может служить методологической основой для принятия решений или практических мер на уровне отдельных лиц или групп лиц, местных властей, учреждений и предприятий государственного и частного секторов, в особенности заинтересованных в развитии ИИ сторон, под которыми понимаются хозяйствующие субъекты, имеющие непосредственное отношение к жизненному циклу конкретных интеллектуальных систем, в том числе организации и частные лица, занимающиеся исследованием, проектированием, созданием, развертыванием и применением ИИ-технологий.

## **II. ЦЕЛИ И ЗАДАЧИ**

4. Цель настоящей Рекомендации – предложить ценностные ориентиры, принципы и стратегические рекомендации в отношении этических аспектов исследования, проектирования, создания, развертывания и применения технологий на основе искусственного интеллекта, которые гарантировали бы использование ИИ-систем во благо всего человечества, отдельного человека, общественных систем и окружающей природной среды.

5. Сложный характер связанных с ИИ этических проблем требует принятия адекватных комплексных ответных мер на основе сотрудничества большого числа заинтересованных сторон, представляющих различные уровни и направления деятельности сообществ в глобальном, региональном и национальном масштабе.

6. Будучи адресована главным образом директивным органам государств, входящих и не входящих в состав ЮНЕСКО, данная Рекомендация преследует также цель предложить международным организациям, национальным и транснациональным компаниям, неправительственным организациям, инженерам-разработчикам и ученым, в том числе представителям гуманитарных, естественных и социальных наук, религиозным и общественным организациям, руководящие принципы, которые поощряли бы многосторонний подход на основе общепризнанных этических норм и позволяли заинтересованным сторонам взаимодействовать и нести общую ответственность на основе глобального межкультурного диалога.

### III. ЦЕННОСТИ И ПРИНЦИПЫ

7. Как подчеркнуто в преамбуле данной Рекомендации, этические принципы и ценности сами по себе не обязательно являются правовыми нормами. Они выполняют важную функцию на этапе формулирования мер стратегического характера и правовых норм, поскольку нравственные ценности являются воплощением согласованных на международном уровне ожиданий в отношении того, что является благом и что надлежит поощрять и поддерживать. В этом смысле ценности лежат в основе принципов.

8. Этические ценностные установки поощряют благое нравственное поведение, соотносящееся с тем, как это понимается международным сообществом. Они являются для принципов фундаментом, в то время как принципы «разъясняют» и конкретизируют лежащие в их основе ценностные установки, что упрощает реализацию последних в форме директивных положений и практических мер.

#### III.1. ЦЕННОСТИ

##### Человеческое достоинство

9. Исследование, проектирование, создание, развертывание и применение систем искусственного интеллекта должно осуществляться на основе принципа уважения и защиты человеческого достоинства. Достоинство каждой человеческой личности является ценностью, лежащей в основе всех прав человека и основных свобод, в связи с чем при разработке и адаптации ИИ-систем этому аспекту должно уделяться самое пристальное внимание. Человеческое достоинство связано с признанием объективно присущей каждому отдельному человеку ценности и по этой причине не зависит от национальной принадлежности, правового статуса, социально-экономического положения, пола и сексуальной ориентации, религии, языка, этнического происхождения, политических или иных убеждений.

10. Уважительное отношение к человеческому достоинству должно обеспечиваться в первую очередь всеми субъектами, вовлеченными в процесс исследования, проектирования, создания, развертывания и применения интеллектуальных систем, а во-вторых, поощряться путем разработки новых законов, реализации инициатив в области управления, демонстрации положительного опыта совместной разработки и использования ИИ-технологий, а также распространения по мере развития таких технологий национальных и международных технических и методологических руководств, издаваемых правительственными структурами.

##### Права человека и основные свободы

11. Применительно к ИИ смысл ценности уважения, защиты и поощрения прав человека и основных свобод заключается в том, что процесс исследования, проектирования, создания, развертывания и применения ИИ-систем должен отвечать требованиям и осуществляться с соблюдением международных норм, принципов и стандартов в области прав человека.

##### «Никого не оставить без внимания»

12. Крайне важно, чтобы методы исследования, проектирования, создания, развертывания и применения систем искусственного интеллекта обеспечивали уважительное отношение ко всем категориям населения и поощряли творчество во всем его многообразии. Проблемы дискриминации и предвзятости, цифрового отставания и разрыва в уровне знаний, а также глобальные проявления неравенства должны оставаться в поле зрения на всех этапах жизненного цикла любой ИИ-системы.

13. В связи с этим деятельность, связанная с исследованием, проектированием, созданием, развертыванием и применением интеллектуальных систем, должна способствовать расширению прав и возможностей всех людей и вестись с учетом потребностей, характерных для раз-



личных возрастов, культурных систем, лиц с ограниченными возможностями, женщин и девочек, малообеспеченных, социально незащищенных и уязвимых категорий населения, и не должна использоваться для ограничения возможностей выбора человеком подходящего ему образа жизни или жизненного опыта, в том числе предусматривать факультативное использование ИИ-систем. Кроме того, необходимо предпринять усилия по преодолению нехватки соответствующей технической инфраструктуры, образовательных возможностей и практических навыков, а также нормативно-правовой базы, в особенности в странах с низким и средним уровнем дохода.

### **Жизнь в гармонии**

14. Жизнь в гармонии как ценность заключается в учете в процессе исследования, проектирования, создания, развертывания и применения систем искусственного интеллекта того факта, что все люди взаимосвязаны. В основе понятия взаимосвязанности лежит понимание, что каждый человек является частью большего целого, и что это целое становится слабее, когда тем или иным образом ослабляются составляющие его части.

15. Эта ценность подразумевает, что деятельность, связанная с исследованием, проектированием, созданием, развертыванием и применением ИИ-систем не должна порождать конфликты и насилие, вести к дискриминации, обезличению или созданию небезопасных для людей условий, к порождению разногласий между отдельными людьми и группами лиц и настраиванию их друг против друга, а также ставить под угрозу гармоничное сосуществование человека с окружающей его средой, поскольку это имело бы для человечества негативные последствия. Предназначение данной ценностной установки заключается в признании того, что заинтересованные в развитии ИИ стороны содействуют достижению цели жизни в гармонии, а именно обеспечению будущего общего блага.

### **Доверие**

16. Системы искусственного интеллекта должны вызывать доверие. В основе такого доверия лежит социально-технологический подход, означающий что методы исследования, проектирования, создания, развертывания и применения ИИ-систем должны стимулировать, а не подрывать доверие людей друг к другу и к интеллектуальным системам.

17. В каждом конкретном случае применения ИИ доверие необходимо заслужить: в широком смысле доверие является критерием принятия ИИ-систем обществом. В связи с этим люди должны иметь объективные основания полагать, что ИИ-технологии приносят пользу и что в интересах минимизации потенциальных негативных последствий принимаются адекватные меры.

### **Защита окружающей природной среды**

18. Цель данной ценностной установки – обеспечить, чтобы в деятельности, связанной с исследованием, проектированием, созданием, развертыванием и применением систем искусственного интеллекта во внимание принимались аспекты, касающиеся поощрения экологического благополучия. Все вовлеченные на различных этапах жизненного цикла ИИ-систем субъекты должны соблюдать соответствующие международные и национальные нормы в области охраны окружающей среды и обеспечения устойчивого развития с целью минимизации потенциального воздействия факторов, связанных с изменением климата, в том числе выбросов углерода системами на базе ИИ, и предотвращения моделей эксплуатации и истощения запасов природных ресурсов, ведущих к ухудшению состояния окружающей среды.

19. При этом необходимо задействовать интеллектуальные системы для выработки решений, направленных на защиту природной среды и сохранение нашей планеты, путем поощрения подходов и методов, используемых в экономике замкнутого цикла.

## III.2. ПРИНЦИПЫ

20. С учетом того, что любая искусственная интеллектуальная система характеризуется набором важных ситуативных особенностей, постоянно меняющихся под влиянием антропогенных и технологических факторов, принципы подразделены на две группы.

21. В первую группу включены принципы, отражающие отличительные особенности человеко-машинного интерфейса, т.е. взаимодействия человека с ИИ-системами. Следует отметить, что исследование, проектирование, создание, развертывание и применение искусственных интеллектуальных систем влияет на возможности принятия человеком решений в двух аспектах: во-первых, с точки зрения расширения сферы автономной деятельности системы и принимаемых ею решений, а во-вторых, посредством воздействия на возможность человека влиять как положительным, так и отрицательным образом.

22. Вторая группа объединяет принципы, которые отражают параметры, связанные с отличительными особенностями самих ИИ-систем. Они важны с точки зрения обеспечения соответствия методов исследования, проектирования, создания, развертывания и применения интеллектуальных систем согласованному на международном уровне видению в отношении этического поведения.

### ПЕРВАЯ ГРУППА ПРИНЦИПОВ

#### Служение интересам человека и процветанию

23. Исследование, проектирование, создание, развертывание и применение систем искусственного интеллекта должно способствовать благополучию и процветанию человека и среды его обитания. На всех этапах жизненного цикла ИИ-систем необходимо заботиться о повышении качества жизни всех людей и поощрять осуществление каждым человеком своих прав в полном объеме; при этом любой человек или группа лиц вправе свободно трактовать понятие «качество жизни» в той мере, в какой это не наносит людям физического или психического ущерба и не умаляет их достоинства с точки зрения существующего определения.

24. Исследование, проектирование, создание, развертывание и применение ИИ-систем может проводиться в целях содействия общению уязвимых категорий населения, включая, в том числе, детей, пожилых людей или лиц с нарушениями здоровья, однако не должно ни при каких условиях вести к унижению человека или унижению человеческого достоинства, а также к нарушению или попранию прав человека.

#### Соответствие

25. Исследование, проектирование, создание, развертывание и применение систем искусственного интеллекта не должно выходить за рамки того, что необходимо для достижения законных целей или выполнения намеченных задач, и при этом должно соответствовать конкретным условиям.

26. Метод реализации конкретной ИИ-системы должен определяться на основе следующих соображений: (а) выбранный метод должен быть оптимальным и отвечающим поставленной цели; (b) выбранный метод не должен явным образом противоречить основополагающим ценностным установкам, сформулированным в настоящем документе; (c) выбранный метод должен соответствовать конкретным условиям.

#### Подконтрольность и подчиненность человеку

27. Во всех случаях должна существовать возможность возложения как правовой, так и этической ответственности, связанной с исследованием, проектированием, созданием, развертыванием и применением систем искусственного интеллекта, на конкретное физическое либо

существующее юридическое лицо. В этом смысле под термином «подконтрольность человеку» подразумевается не только контроль со стороны конкретного физического лица, но и общественный контроль.

28. Возможны ситуации, когда по соображениям эффективности человек может осуществлять контроль совместно с ИИ-системами. Тем не менее такое решение о делегировании в ограниченном числе случаев части функций контроля все равно принимается человеком, поскольку цель исследования, проектирования, создания, развертывания и применения ИИ-систем состоит в том, чтобы помочь людям принимать решения и выполнять определенные функции, но не заменить человека, на котором всегда будет лежать конечная ответственность.

### **Устойчивость**

29. В интересах содействия формированию устойчивых обществ заинтересованные в развитии ИИ стороны должны принимать во внимание социальные, экономические и экологические аспекты устойчивого развития всего человечества, а также проявлять заботу об окружающей среде. Исследование, проектирование, создание, развертывание и применение ИИ-систем должно преследовать цель содействия обеспечению устойчивости в соответствии с согласованными на глобальном уровне рамочными принципами, в частности целями в области устойчивого развития.

### **Разнообразие и инклюзивность**

30. При проведении исследований, проектировании, создании, развертывании и применении систем искусственного интеллекта необходимо обеспечить уважение и поощрение разнообразия и инклюзивности, как минимум, соответствие международным нормам, стандартам и принципам в области прав человека, в том числе демографическое, культурное и социальное разнообразие и инклюзивность.

### **Неприкосновенность частной жизни**

31. При проведении исследований, проектировании, создании, развертывании и применении систем искусственного интеллекта должны быть обеспечены уважение, защита и поощрение неприкосновенности частной жизни – права, имеющего важнейшее значение с точки зрения защиты человеческого достоинства и свободы выбора. На всех этапах жизненного цикла ИИ-систем, в том числе применительно к сбору данных, контролю использования данных на основе информированного согласия и разрешений, а также к раскрытию информации о применении и использовании данных и обеспечению личных прав на владение данными и доступ к ним, должны обеспечиваться надлежащие механизмы управления данными.

### **Осведомленность и грамотность**

32. Для обеспечения эффективного общественного участия и возможности принятия гражданами обоснованных решений, касающихся использования ими систем искусственного интеллекта, необходимо содействовать повышению уровня осведомленности и знаний населения в отношении ИИ-технологий и ценности данных посредством осуществления просветительской деятельности, организации пропагандистских кампаний и профессиональной подготовки. Необходимо, чтобы дети были защищены от обоснованно прогнозируемых негативных последствий применения ИИ-систем, имели доступ к таким системам в рамках получения образования и профессиональной подготовки, а также не утратили естественных способностей в результате использования систем на основе ИИ.

## **Многостороннее и адаптивное управление**

33. Система управления в сфере ИИ должна гибко реагировать на происходящие технологические изменения и связанное с этим изменение в моделях хозяйствования, быть инклюзивной (предусматривать многостороннее участие заинтересованных сторон), иметь по возможности распределенную между различными уровнями структуру и обеспечивать оптимальный процесс принятия управленческих решений на основе межотраслевого системного подхода.

34. Методы управления должны предусматривать многовариантность управленческих механизмов, начиная с систем «мягкого» управления, саморегулирования и сертификации, и заканчивая «жестким» управлением на основе национального законодательства, а там, где это возможно и необходимо, – на основе норм международного права. Во избежание негативных последствий и непреднамеренного ущерба система управления должна включать компоненты прогнозирования, защиты, мониторинга воздействия, обеспечения выполнения и возмещения ущерба.

## **ВТОРАЯ ГРУППА ПРИНЦИПОВ**

### **Справедливость**

35. Заинтересованные в развитии ИИ стороны должны соблюдать принцип справедливости, равноправия и недискриминационности, а также прилагать все возможные усилия к тому, чтобы минимизировать последствия и избежать усиления или закрепления на любом из этапов жизненного цикла ИИ-системы социально-технологических предубеждений, в том числе на почве расовой, национальной или гендерной принадлежности, возраста или культуры.

### **Прозрачность и объяснимость**

36. Несмотря на необходимость, в принципе, прилагать все возможные усилия для повышения прозрачности и объяснимости работы систем искусственного интеллекта в интересах обеспечения доверия со стороны людей, степень такой прозрачности и объяснимости должна всегда учитывать контекст применения ИИ в целях сбалансированной увязки принципа прозрачности и объяснимости с другими принципами, в частности, с принципами обеспечения безопасности и защиты данных.

37. С учетом контекста применения и степени конфиденциальности используемых данных прозрачность ИИ-системы означает открытость для общественности информации, касающейся исследования, проектирования, создания, развертывания и применения системы. Такая информация может также содержать анализ факторов, воздействующих на конкретный прогноз или решение, однако, как правило, не предполагает раскрытия конкретных алгоритмов или наборов данных. В этом смысле прозрачность представляет собой социально-технологический аспект и имеет целью заручиться доверием людей в отношении ИИ-систем.

38. Термин «объяснимость» касается обеспечения понимания и общего представления в отношении результатов работы ИИ-систем. Под объяснимостью моделей ИИ имеется в виду также доступность для понимания исходных данных, непосредственных результатов и поведения каждой алгоритмической структуры и того, как все это влияет на результаты моделирования. Как следствие, «объяснимость» тесно связана с «прозрачностью», поскольку результаты и ведущие к их получению подпроцессы, должны быть понятными и прослеживаемыми с учетом контекста применения.

### **Безопасность и защищенность**

39. В деятельности, связанной с исследованием, проектированием, созданием, развертыванием и применением искусственных интеллектуальных систем следует остерегаться непреднамеренного причинения вреда (потенциальная угроза безопасности) и уязвимости к

атакам (задачи по защите), что позволит обеспечить безопасность и защищенность конкретной ИИ-системы на всех этапах ее жизненного цикла.

40. Органы государственного управления призваны играть первостепенную роль в обеспечении безопасности и защищенности ИИ-систем, в том числе путем разработки на национальном и международном уровнях стандартов и норм, согласующихся с международными нормами, стандартами и принципами в области прав человека. В целях недопущения масштабного ущерба необходимо последовательно поддерживать проведение системных исследований в отношении потенциальных угроз безопасности и защищенности, связанных с различиями в подходах к реализации долгосрочных проектов в области ИИ.

#### **Ответственность и подотчетность**

41. В соответствии с действующим международным законодательством в области прав человека и руководящими документами по вопросам этики вовлеченные в создание ИИ стороны должны нести моральную и юридическую ответственность на всех этапах жизненного цикла ИИ-систем. Ответственность за решения и меры, принятые с задействованием в той или иной степени системы искусственного интеллекта, во всех случаях должна в конечном счете возлагаться на причастные к ИИ-системам стороны.

42. Для обеспечения подотчетности в отношении связанной с ИИ-системами деятельности и получаемых результатов необходимо разработать соответствующие механизмы. Для обеспечения возможности проверки и отслеживания (работоспособности) ИИ-систем внимание необходимо уделять как техническим, так и организационным аспектам.

### **IV. ОБЛАСТИ, ТРЕБУЮЩИЕ ПРИНЯТИЯ ДИРЕКТИВНЫХ МЕР**

#### **ЦЕЛЬ ДЕЯТЕЛЬНОСТИ I: ЭТИЧНОЕ УПРАВЛЕНИЕ**

43. Обеспечение соответствия методов исследования, проектирования, создания, развертывания и применения ИИ основополагающим этическим ценностям, таким как права человека, разнообразие, инклюзивность и т.д.

#### **Стратегическая мера 1: Поощрение разнообразия и инклюзивности**

44. Государствам-членам следует взаимодействовать с международными организациями в интересах активного вовлечения всех государств-членов, в особенности стран с низким и средним уровнем дохода, в международные дискуссии по проблематике ИИ. Это может быть сделано посредством предоставления финансовых средств для обеспечения равного регионального представительства или с использованием любых других механизмов.

45. Государствам-членам следует требовать от занимающихся развитием ИИ-систем заинтересованных сторон предавать гласности и пресекать любые выявленные в рамках функционирования ИИ-систем культурные и социальные стереотипы, будь то обусловленные алгоритмами этих систем или человеческой халатностью, а также обеспечить, чтобы используемые для обучения ИИ-систем наборы данных не усугубляли культурное и социальное неравенство. Необходимо предусмотреть механизмы, которые позволят конечным пользователям сообщать о выявлении подобных проявлений неравенства, предубеждения и стереотипных установок.

46. Государствам-членам следует обеспечить, чтобы при проведении исследований, проектировании, создании, развертывании и применении ИИ-систем занимающиеся их развитием заинтересованные стороны демонстрировали знание и уважение существующих социально-культурных различий, в том числе местных обычаев и религиозных традиций, а также соблюдение международных стандартов и норм в области прав человека.

47. Государствам-членам следует предпринять шаги по устранению характерного в настоящее время для разработки ИИ-систем дисбаланса в вопросах разнообразия, причем как с точки зрения разнообразия обучающих наборов данных, так и в плане разнообразия самих участвующих в развитии ИИ заинтересованных сторон. Государствам-членам следует взаимодействовать со всеми секторами, международными и региональными организациями и другими структурами в целях поощрения женщин и девочек к участию во всех этапах жизненного цикла ИИ-систем путем создания стимулов, доступа к помощи наставников, демонстрации примеров для подражания, а также обеспечения защиты от домогательств. Необходимо также работать над повышением доступности сферы ИИ для лиц различного этнического происхождения и инвалидов. Кроме того, следует поощрять равные возможности доступа к благам, которые открывают ИИ-технологии, особенно для социально незащищенных групп населения.

48. Государствам-членам следует сотрудничать с международными организациями в целях всестороннего учета этических аспектов ИИ посредством включения дискуссий по этим вопросам в программы соответствующих международных, межправительственных и многосторонних форумов.

## **ЦЕЛЬ ДЕЯТЕЛЬНОСТИ II: ОЦЕНКА ВОЗДЕЙСТВИЯ**

49. Создание потенциала превентивного наблюдения для своевременного реагирования на негативные или иные непредвиденные последствия, связанные с применением систем искусственного интеллекта.

### **Стратегическая мера 2: Реагирование, связанное с изменениями на рынке труда**

50. Государствам-членам следует предпринимать шаги по оценке и учету воздействия, оказываемого ИИ-технологиями на рынки труда, а также его последствий с точки зрения требований, предъявляемых к образованию. Эти усилия могут включать изучение более широкого спектра «основных навыков» на всех ступенях образования, что обеспечит новым поколениям реальный шанс находить работу в условиях стремительно меняющегося рынка труда, а также повысит осведомленность молодежи в отношении этических аспектов искусственного интеллекта. Такие навыки, как умение учиться, умение общаться, умение работать в команде, умение соперничать и делиться своими знаниями в различных областях, должны стать предметом изучения наряду со специальными и техническими навыками. Ключевое значение имеет прозрачность в отношении востребованных навыков, а также учет этой информации в школьных программах.

51. Государствам-членам следует сотрудничать с частными структурами, неправительственными организациями и другими заинтересованными сторонами на предмет организации справедливых условий переобучения сотрудников, рискующих потерять работу. Эти усилия могут включать реализацию программ повышения квалификации и профессиональной переподготовки, рассмотрение нестандартных вариантов сохранения за сотрудниками их рабочих мест на период переобучения, а также изучение возможностей включения в программы социальной помощи тех, кто не сможет сохранить работу.

52. Государствам-членам следует поощрять аналитические исследования о воздействии ИИ-технологий на состояние местного рынка труда, что позволит предвидеть характер будущих изменений и проблем в этой области. Такие исследования должны показывать, каких именно социально-экономических секторов и географических территорий в наибольшей степени коснется массовое применение ИИ-технологий.

53. Государствам-членам необходимо разработать стратегии управления трудовыми ресурсами, ориентированные на оказание поддержки женщинам и недопредставленным на рынке труда группам населения, что позволит не оставить никого на обочине цифровой эко-

номики, в основе которой лежат ИИ-технологии. Следует также изучить возможность целевого финансирования программ для женщин и недопредставленных групп населения, направленных на повышение уровня профессиональной подготовки, расширение возможностей трудоустройства, продвижение по службе и профессиональный рост.

### **Стратегическая мера 3: Учет социально-экономического воздействия ИИ**

54. Государствам-членам следует разработать механизмы для предотвращения ведущего к неравенству монопольного применения ИИ, будь то в сфере обработки данных, научных исследований, технологий, рыночного хозяйства и т.д.

55. Государствам-членам необходимо сотрудничать с международными организациями, частными и неправительственными структурами в целях обеспечения адекватного распространения ИИ-грамотности среди населения, особенно в странах с низким и средним уровнем дохода, что позволит сократить цифровое отставание и неравенство в доступе к цифровым технологиям в результате широкого внедрения ИИ-систем.

56. Государствам-членам следует создать механизмы мониторинга и оценки инициатив и политики, касающихся этических аспектов искусственного интеллекта. В число возможных механизмов входит, в частности, создание информационного архива, в котором будут обобщены инициативы по соблюдению этических норм во всех областях компетенции ЮНЕСКО; механизм обмена опытом, позволяющий государствам-членам получать замечания и комментарии от других государств-членов в отношении своей политики и инициатив; а также руководство для разработчиков ИИ-систем, позволяющее оценить соблюдение ими стратегических рекомендаций, сформулированных в настоящем документе.

57. Государствам-членам рекомендуется рассмотреть вопрос о создании механизма сертификации систем искусственного интеллекта, аналогичного тем, что используются для сертификации медицинского оборудования. Речь может идти о различных классах сертификации в зависимости от уровня «чувствительности» области применения и ожидаемого воздействия на жизнь человека, окружающую среду и этические аспекты, в том числе такие, как равенство, разнообразие и культурные ценности. Такой механизм мог бы предусматривать различные уровни экспертизы систем, данных и соблюдения этических норм. В то же время он не должен препятствовать инновационной деятельности или ставить в невыгодное положение небольшие предприятия или начинающие компании, требуя от них большого объема бумажной работы. Указанные механизмы, при необходимости подтверждаемые путем повторной сертификации, будут также включать стандартный компонент мониторинга для обеспечения надежности, неизменной добросовестности и соблюдения этических требований на всех этапах жизненного цикла ИИ-системы.

58. Государствам-членам необходимо поощрять частные компании к тому, чтобы привлекать различные заинтересованные стороны к вопросам методического руководства их ИИ-системами, а также рассмотреть возможность введения должности сотрудника по этическим аспектам ИИ-систем либо использовать какой-либо иной механизм в целях контроля проведения оценки воздействия, экспертизы и постоянного мониторинга, а также для обеспечения соблюдения установленных в отношении ИИ-систем этических требований.

59. Государствам-членам следует разработать стратегии управления данными, обеспечивающие непрерывную оценку качества данных, используемых для обучения ИИ-систем, в том числе адекватности процессов сбора и осуществления выборки данных, надежности мер по обеспечению безопасности и защиты данных, а также создать механизмы обратной связи для анализа и учета допущенных ошибок и обмена успешным опытом со всеми вовлеченными в развитие ИИ заинтересованными сторонами. Предметом первостепенного внимания в подобных стратегиях должно стать достижение баланса между использованием метаданных и защитой конфиденциальности личной информации.

#### **Стратегическая мера 4: Воздействие применения ИИ на культуру и окружающую среду**

60. Государствам-членам рекомендуется в соответствующих случаях использовать ИИ-системы в рамках деятельности, связанной с сохранением, развитием и изучением культурного наследия в его материальной и нематериальной форме, включая редкие языки, например, посредством разработки или актуализации учебно-просветительских программ, связанных с применением систем ИИ в этих областях и ориентированных на интересы учреждений и граждан.

61. Государствам-членам рекомендуется исследовать и учитывать воздействие применения ИИ-систем, в особенности прикладных программ обработки естественного языка, таких как автоматизированный перевод и голосовые помощники, на нюансы человеческого языка. Такие исследования могут касаться максимального задействования потенциала этих систем в интересах устранения культурных пробелов и углубления человеческого понимания, а также таких негативных последствий, как уменьшение распространенности редких языков, местных диалектов, тональных и культурных вариаций, связанных с человеческим языком и речью.

62. Государствам-членам следует поощрять и поддерживать совместные исследования, посвященные последствиям долгосрочного взаимодействия человека с системами искусственного интеллекта. Такая деятельность должна осуществляться с использованием широкого разнообразия норм, принципов, протоколов и отраслевых подходов и на основе анализа изменения привычек, а также тщательной оценки косвенного воздействия, оказываемого ИИ на культуру и общество.

63. С учетом использования ИИ для создания, производства, распространения и трансляции огромного разнообразия товаров и услуг культурного назначения и принимая во внимание важность сохранения культурного наследия и культурного разнообразия, государствам-членам следует содействовать развитию обучения творческих работников и специалистов использованию ИИ-технологий в целях оценки ими целесообразности применения этих технологий в своей профессиональной деятельности.

64. Государствам-членам следует содействовать осведомленности местных индустрий культуры и начинающих предприятий, работающих в области культуры, в отношении программных средств на основе ИИ, а также оценке ими этих средств в целях предотвращения риска их высокой концентрации на рынке товаров и услуг культурного назначения.

65. Государствам-членам следует принять меры по оценке и снижению экологического воздействия в результате применения систем ИИ, включая, в том числе, их углеродного следа. Кроме того, им необходимо стимулировать продвижение этически обоснованных и экологических решений на основе ИИ, а также способствовать внедрению таких решений в различных контекстах. В порядке примера можно упомянуть применение ИИ в следующих целях:

- (a) повышение эффективности охраны, мониторинга состояния и рациональности использования природных ресурсов;
- (b) содействие в предотвращении, регулировании и решении проблем, связанных с изменением климата;
- (c) содействие повышению эффективности и устойчивости продовольственной экосистемы;
- (d) ускорение перехода к экологически безопасной энергетике и широкое внедрение соответствующих технологий.



### **ЦЕЛЬ ДЕЯТЕЛЬНОСТИ III: УКРЕПЛЕНИЕ ПОТЕНЦИАЛА В ВОПРОСАХ ЭТИЧНОГО ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

66. Создание кадрового и организационного потенциала для оценки этического воздействия, надзора за соблюдением этических норм и методического руководства этой деятельностью.

#### **Стратегическая задача 5: Содействие распространению знаний и повышению осведомленности об этических аспектах ИИ**

67. Государствам-членам следует, сообразуясь с национальными планами развития образования и традициями, включать этические аспекты ИИ в учебные программы всех ступеней школьного и университетского образования и поощрять формирование междисциплинарных навыков на стыке технических и социально-гуманитарных дисциплин. Онлайн-учебные курсы и цифровые ресурсы должны разрабатываться в том числе на местных языках и в доступных для лиц с ограниченными возможностями форматах.

68. Государствам-членам следует поощрять обучение «требуемым навыкам» использования ИИ-технологий, в частности базовым навыкам работы с данными, а также умению понимать и писать простые коды, что особенно важно для стран, существенно отстающих в вопросах обучения таким навыкам.

69. С учетом быстрых темпов развития систем на основе ИИ государствам-членам необходимо обеспечить гибкий характер университетских учебных программ и упростить процедуру их обновления. Кроме того, в целях содействия быстрому и гибкому обновлению учебных программ следует изучить возможность интеграции онлайн- и послешкольного образования, а также стекирования документов об образовании.

70. Государствам-членам следует поощрять осуществление программ повышения общей осведомленности об ИИ-технологиях и открытый доступ к информации о потенциальных возможностях и проблемах, связанных с развитием ИИ. Такая информация должна быть доступна как для подготовленной, так и для неподготовленной аудитории с уделением особого внимания недопредставленным в сфере ИИ группам населения.

71. Государствам-членам следует поощрять исследовательские инициативы, касающиеся применения ИИ в числе прочего в сфере преподавания, подготовки преподавательских кадров и электронного обучения, что призвано расширить потенциальные возможности и минимизировать потенциальные проблемы и риски, связанные с ИИ-технологиями. Подобные усилия должны всегда сопровождаться адекватной оценкой воздействия применения ИИ на качество образования, на обучаемых и их преподавателей, а также обеспечивать, чтобы обучение на основе ИИ-технологий расширяло возможности и обогащало опыт обеих этих групп.

72. Государствам-членам следует поддерживать заключение между академическими учреждениями и профессиональным сообществом соглашений о взаимодействии в целях преодоления дефицита требуемых навыков, а также поощрять сотрудничество между производственным сектором, академическим сообществом, общественными организациями и органами государственного управления в целях приведения предлагаемых образовательными учреждениями программ и стратегий в области профессионального образования в соответствие с потребностями экономики. Необходимо поощрять проектно-ориентированные подходы к обучению в области искусственного интеллекта, позволяющие устанавливать партнерские отношения между компаниями, университетами и исследовательскими центрами.

73. Государствам-членам следует, в особенности, содействовать участию женщин, представителей различных рас и культур, а также инвалидов в учебных программах по тематике ИИ, начиная с общей средней школы и заканчивая высшим образованием, а также содействовать выявлению передового опыта и обмену им с другими государствами-членами.

### **Стратегическая задача 6: Поощрение исследования этических аспектов искусственного интеллекта**

74. Государствам-членам следует поощрять исследование этических аспектов ИИ путем непосредственного финансирования данного направления деятельности, либо стимулирования государственного и частного секторов к инвестированию средств в эту сферу.

75. Государствам-членам необходимо предусмотреть, чтобы занимающиеся проблематикой ИИ исследователи имели подготовку в вопросах учета этических аспектов при проведении научных исследований, а также требовать от них включения этического компонента в планы проведения и результаты исследований, в частности анализа используемых наборов данных и методов их аннотирования, а также определения качества и диапазона полученных результатов.

76. Государствам-членам и частным компаниям следует по возможности содействовать доступу научного сообщества к исследовательским данным на национальном уровне в целях укрепления потенциала научного сообщества, в особенности в развивающихся странах. Такой доступ не должен, тем не менее, осуществляться в нарушение принципа неприкосновенности личной информации граждан.

77. Государствам-членам следует поощрять гендерное разнообразие при проведении научным и профессиональным сообществами исследований в области ИИ путем стимулирования женщин к участию в такой деятельности, создания механизмов противодействия стереотипным представлениям о роли женщин и домогательствам в исследовательской среде, а также поощрения научных и частных организаций к обмену позитивным опытом в вопросах содействия гендерному разнообразию.

78. Государствам-членам и финансирующим органам следует поощрять междисциплинарные исследования в области ИИ путем включения в них помимо науки, техники, инженерии и математики (ЕНТИМ) других дисциплин, таких как право, международные отношения, политические науки, образование, философия, культура, лингвистика и т.д., что обеспечит критический подход к исследованиям в области ИИ и своевременное выявление потенциального ненадлежащего использования или негативных последствий применения ИИ-технологий.

### **ЦЕЛЬ ДЕЯТЕЛЬНОСТИ IV: ПРОЦЕССЫ РАЗВИТИЯ И МЕЖДУНАРОДНОЕ СОТРУДНИЧЕСТВО**

79. Обеспечение совместного этического подхода в вопросах использования ИИ-технологий в целях разработки прикладных программ в сфере развития с учетом огромных возможностей, которые эти технологии открывают в плане ускорения процессов развития.

### **Стратегическая задача 7: Поощрение этичного применения искусственного интеллекта в рамках процессов развития**

80. Государствам-членам следует поощрять этичное использование потенциала ИИ в связанных с развитием областях, таких как здравоохранение, сельское хозяйство/продовольственное обеспечение, образование, культура, окружающая среда, управление водными ресурсами, управление инфраструктурой, экономическое планирование и развитие и других областях.

81. Государствам-членам и международным организациям необходимо стремиться заложить основы для международного сотрудничества в вопросах использования ИИ в интересах развития, в том числе путем предоставления квалифицированных специалистов, финансовой поддержки, данных, экспертных знаний в предметных областях и технических возможностей, а также содействия проведению рабочих совещаний с участием технических специалистов и

экономистов в целях преодоления сложных проблем в области развития, в частности, стоящих перед странами с низким и средним уровнем дохода и НРС.

82. Государствам-членам следует поощрять международное сотрудничество в вопросах исследования ИИ с участием, в том числе, исследовательских центров и сетей, что будет способствовать более широкому вовлечению в это сотрудничество исследователей из стран с низким и средним уровнем дохода, а также из других развивающихся регионов.

#### **Стратегическая задача 8: Содействие международному сотрудничеству по вопросам этики искусственного интеллекта**

83. Государствам-членам следует взаимодействовать в рамках международных организаций и по линии научно-исследовательских учреждений в вопросах проведения исследований по этическим аспектам ИИ. Государственные и частные организации должны обеспечить справедливое и беспристрастное применение алгоритмов и данных, используемых в широком спектре подобластей искусственного интеллекта, – от поддержания правопорядка и уголовного правосудия до обеспечения занятости, здравоохранения и образования, – в том числе исследование особенностей понимания справедливости и беспристрастности в различных культурах и контекстах, а также изучение того, как совместить эти особенности с осуществимыми с технической точки зрения решениями.

84. Государствам-членам следует поощрять международное сотрудничество в вопросах развития и применения ИИ в интересах преодоления географических и технологических барьеров. Эта работа потребует объединения усилий большого числа заинтересованных сторон на национальном, региональном и международном уровнях. Обмен научно-техническими знаниями и обсуждение вопросов развития технологий должны проводиться между государствами-членами и их населением, между государственным и частнохозяйственным секторами, а также между государствами-членами и в самих государствах.

#### **ЦЕЛЬ ДЕЯТЕЛЬНОСТИ V: ОРГАНИЗАЦИЯ УПРАВЛЕНИЯ В ИНТЕРЕСАХ ЭТИЧНОГО ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

85. Поощрение и направление деятельности по включению этического компонента в механизмы управления ИИ-системами.

#### **Стратегическая задача 9: Создание механизмов управления в интересах этического применения искусственного интеллекта**

86. Государствам-членам необходимо обеспечить, чтобы любой механизм управления в сфере ИИ отвечал следующим критериям:

- (a) Инклюзивность: допускает и поощряет участие представителей коренного населения, женщин, молодежи и пожилых людей, лиц с ограниченными возможностями, а также представителей других меньшинств и слабо представленных групп населения.
- (b) Прозрачность: допускает надзор со стороны соответствующих национальных структур или доверенных третьих сторон. Применительно к СМИ речь могла бы идти о межведомственной целевой группе, уполномоченной проверять источники информации; в отношении технологических компаний могла бы проводиться внешняя проверка этапов разработки и внедрения, а также результатов внутренней проверки; в отношении государств-членов могла бы быть использована процедура обзора в рамках форумов по правам человека.
- (c) Многодисциплинарность: любой проблемный аспект должен рассматриваться всеобъемлющим образом, а не только с технической точки зрения.

- (d) Многостороннее участие: необходимо заключение международных соглашений в целях смягчения последствий и возмещения любого ущерба, который может возникнуть в определенной стране по вине компании или пользователя, размещающихся в другой стране. Это не отменяет того, что страны и регионы разрабатывают собственные регламентирующие положения, отражающие их культурные особенности.

87. Государствам-членам следует содействовать развитию цифровой экосистемы ИИ, отвечающей этическим требованиям, а также обеспечению доступа к ней. Такая экосистема объединяет в себе, среди прочего, цифровые технические средства и поддерживающую инфраструктуру, а также, при необходимости, механизмы обмена ИИ-знаниями. В связи с этим государствам-членам следует проанализировать возможность пересмотра своей политики и нормативно-правовой базы, в том числе в вопросах доступа к информации и открытого государственного управления, с целью отражения в них конкретных требований в отношении ИИ, а также содействовать развитию таких структур, как фонды данных, в целях обеспечения в числе прочего безопасного, добросовестного, законного и этически корректного обмена данными.

88. Государствам-членам следует поощрять разработку и применение на глобальном и региональном уровнях сопоставимых руководящих принципов в области ИИ, включая этические аспекты, а также обеспечить сбор данных, необходимых для оценки, мониторинга и контроля прогресса в развитии этически ориентированных ИИ-систем.

89. Государствам-членам следует рассмотреть возможность разработки и применения международного рамочного нормативного документа по поощрению развития сотрудничества между государствами и другими заинтересованными сторонами.

#### **Стратегическая задача 10: Обеспечение надежности систем искусственного интеллекта**

90. Государствам-членам и частным компаниям следует принять надлежащие меры в целях обеспечения мониторинга всех этапов жизненного цикла любой системы искусственного интеллекта, в том числе поведения алгоритмов, отвечающих за принятие решений, данных, а также вовлеченных в создание ИИ сторон, особенно в сфере государственных услуг и там, где необходимо прямое взаимодействие с конечным пользователем.

91. Государствам-членам следует обеспечить разработку четких требований в отношении прозрачности и объяснимости ИИ-систем с отражением следующих аспектов:

- (a) Область применения: некоторые направления деятельности, такие как правоохранительная сфера, безопасность, образование и здравоохранение, вероятнее всего, будут иметь более высокую в сравнении с другими потребность в прозрачности и объяснимости.
- (b) Целевая аудитория: характер информации о примененных в ИИ-системе алгоритмах и результатах ее работы, а также форма требуемых пояснений могут отличаться в зависимости от того, кем запрашиваются разъяснения (это могут быть пользователи, эксперты в предметной области, разработчики и т.д.).
- (c) Осуществимость: многие из алгоритмов ИИ пока не поддаются объяснению; в отношении других приведение в удобочитаемую форму существенно повышает накладные расходы на внедрение системы. В связи с этим до создания технических возможностей полной расшифровки с минимальными последствиями для функциональности будет существовать компромисс между точностью/качеством работы системы и степенью объяснимости ее алгоритмов.

92. Государствам-членам следует поощрять проведение исследований по тематике прозрачности и объяснимости ИИ посредством дополнительного финансирования этого направления работы для различных областей и на различных уровнях (технический, естественный язык и т.д.).

93. Государствам-членам и международным организациям следует рассмотреть вопрос о разработке международных стандартов, описывающих поддающиеся измерению и проверке уровни прозрачности, что позволит объективно оценивать ИИ-системы и определять степень их соответствия требованиям.

### **Стратегическая задача 11: Обеспечение ответственности, подотчетности и защиты личной информации**

94. Государствам-членам следует пересмотреть и в необходимых случаях адаптировать нормативно-правовую базу в целях обеспечения подотчетности и ответственности за содержание и результаты работы ИИ-систем на различных этапах их жизненного цикла. Правительствам следует утвердить рамочные принципы ответственности либо уточнить толкование уже существующих рамок, что позволит устанавливать лиц, ответственных за принимаемые ИИ-системами решения и их поведение. При разработке нормативно-правовой базы правительствам следует, в частности, учитывать, что функции, связанные с ответственностью и подотчетностью, должны во всех случаях возлагаться на физическое или юридическое лицо: полномочия не могут быть делегированы системе искусственного интеллекта; аналогичным образом ИИ-система не может быть наделена статусом юридического лица.

95. Государствам-членам предлагается ввести механизм оценки воздействия ИИ-систем для выявления и анализа преимуществ и рисков, связанных с их применением, а также для принятия мер по предотвращению, смягчению последствий и мониторингу рисков. В соответствии с изложенными в настоящей Рекомендации принципами в ходе оценки рисков должны выявляться последствия для прав человека и окружающей среды, а также этические и социальные последствия. Правительствам следует утвердить нормативно-правовую базу, устанавливающую порядок проведения государственными органами оценок воздействия на ИИ-системы, приобретенные, разработанные и/или внедренные этими органами в целях прогнозирования последствий, снижения рисков, недопущения вредного воздействия, облегчения участия граждан и решения стоящих перед обществом проблем. В рамках оценки воздействия следует требовать от государственных органов проведения самооценки в отношении имеющихся и предлагаемых к внедрению систем искусственного интеллекта, в ходе которой в числе прочего должна определяться целесообразность применения ИИ-систем в конкретной области государственного сектора, а также наиболее подходящий для этого метод. Кроме того, в рамках оценки должны создаваться соответствующие механизмы контроля, в том числе обеспечивающие возможность проведения проверок и функции отслеживания и объяснения, которые позволяют оценивать алгоритмы, данные и методы проектирования. Оценка должна также предусматривать проведение внешней проверки ИИ-систем. При этом проводимая оценка должна носить многопрофильный, многосторонний, мультикультурный, плюралистический и инклюзивный характер.

96. Государствам-членам следует вовлекать все охваченные экосистемой ИИ заинтересованные стороны (включая, в том числе, представителей гражданского общества, правоохранительные органы, страховщиков, инвесторов, производителей, инженеров-разработчиков, юристов и пользователей) в процесс выработки норм в случае отсутствия таковых. Такие нормы могут впоследствии трансформироваться в передовой опыт и законы. Наряду с этим, для ускорения разработки законов и стратегий в контексте стремительного развития новых технологий и в целях их апробирования в безопасных условиях до официального вступления в силу государствам-членам предлагается использовать механизм так называемых регуляторных «песочниц».

97. Государствам-членам следует обеспечить возможность расследования, наказания виновных и возмещения причиненного пользователям ИИ-систем ущерба, в том числе путем поощрения предприятий частного хозяйственного сектора к созданию механизмов устранения нарушений. С этой целью следует содействовать обеспечению проверяемости и отслеживаемости соблюдения предъявляемых к ИИ-системам, в особенности к автономным интеллектуальным системам, требований.

98. Государствам-членам следует обеспечить соответствующие гарантии соблюдения основополагающего права на неприкосновенность частной жизни, в том числе путем принятия или применения правовых норм, обеспечивающих надлежащую защиту в соответствии с международным правом. В отсутствие таких закрепленных законом положений государствам-членам следует настоятельно рекомендовать всем занимающимся ИИ заинтересованным сторонам, в том числе частным компаниям, разрабатывающим и эксплуатирующим ИИ-системы, обеспечить соблюдение в рамках этих систем принципа неприкосновенности личной информации.

99. Государствам-членам необходимо предоставить гражданам возможность контроля за использованием их личной информации/данных, в частности сохранить за ними право доступа к касающимся их данным, а также «право быть забытым».

100. Государствам-членам следует обеспечить повышенный уровень защиты идентифицирующих персональных данных или данных, которые, в случае их разглашения, могут причинить человеку значительный ущерб, физический вред или страдания. В качестве примера можно упомянуть данные о совершенных правонарушениях, уголовных разбирательствах и обвинительных приговорах, а также связанных с этим мерах безопасности; биометрические данные; персональные данные, касающиеся расовой или этнической принадлежности, политических убеждений, членства в профсоюзах, религиозных или иных верований, здоровья или сексуальной жизни.

101. Государствам-членам следует стремиться закрепить в отношении данных принцип «общего достояния», что будет способствовать функциональной совместимости наборов данных и одновременно обеспечит их защищенность, а также максимально бдительный контроль при их сборе и использовании. Такие усилия, при наличии возможностей и в случае практической осуществимости, могли бы включать финансирование создания наборов данных «золотой стандарт», в том числе открытых и надежных наборов данных по различным предметным областям, создаваемых с согласия субъектов данных, в случаях, когда получение согласия регламентировано законом, а также поощрять этические подходы в области развития технологий на основе обмена качественными данными, хранящимися в общем надежном и защищенном пространстве данных.

## **V. МОНИТОРИНГ И ОЦЕНКА**

102. Государствам-членам следует с учетом конкретных условий, структур управления и конституционных положений проводить мониторинг и оценку политики, программ и механизмов, касающихся этических аспектов ИИ, используя сообразно обстоятельствам комбинацию количественного и качественного подходов. Государствам-членам предлагается рассмотреть возможность принятия следующих мер:

- (a) внедрение соответствующих исследовательских механизмов для оценки ответственности и эффективности этической составляющей политики и стимулов в области ИИ с учетом поставленных задач;
- (b) сбор и распространение при поддержке ЮНЕСКО и международного сообщества специалистов по вопросам этики ИИ отчетов о достигнутых результатах, передовом опыте, инновациях и исследованиях, касающихся этических аспектов ИИ и результатов его применения.

103. Возможные механизмы мониторинга и оценки могли бы включать обсерваторию по вопросам соблюдения установленных в отношении ИИ-систем этических норм во всех областях компетенции ЮНЕСКО; механизм обмена опытом, позволяющий государствам-членам обеспечивать обратную связь в отношении своих инициатив; и «измеритель соответствия» – руководство для разработчиков ИИ-систем, позволяющее оценить соблюдение ими стратегических рекомендаций, сформулированных в настоящем документе.

104. Для оценки действенности и эффективности этической составляющей политики в области ИИ на основе сопоставления с согласованными стандартами, приоритетами и целевыми показателями, включая отдельные целевые показатели для находящихся в неблагоприятном положении и уязвимых групп населения, необходимо разработать соответствующие инструменты и показатели. Эта деятельность может включать проведение оценок в государственных и частных учреждениях, оценку поставщиков услуг и программ, в том числе проведение самооценки, а также поисковые исследования и разработку наборов показателей. Сбор и обработка данных должны осуществляться в соответствии с законодательством о защите данных.

105. Методы мониторинга и оценки должны обеспечивать широкое участие соответствующих заинтересованных сторон, включая, в том числе, лиц разных возрастных категорий, инвалидов, женщин и девочек, представителей малообеспеченных, социально незащищенных и уязвимых групп населения, а также соблюдение принципа социального и культурного разнообразия в интересах повышения эффективности методов обучения и укрепления взаимосвязи между полученными результатами, принимаемыми решениями, обеспечением прозрачности и ответственностью за выполнение.

## **VI. ИСПОЛЬЗОВАНИЕ И ПРИМЕНЕНИЕ ПОЛОЖЕНИЙ НАСТОЯЩЕЙ РЕКОМЕНДАЦИИ**

106. Государствам-членам следует стремиться расширять и дополнять собственные меры в отношении настоящей Рекомендации путем сотрудничества со всеми национальными и международными правительственными и неправительственными организациями, чья деятельность подпадает под действие положений и соответствует целям настоящей Рекомендации.

107. Государствам-членам и указанным в настоящей Рекомендации заинтересованным сторонам следует принять все возможные меры для соблюдения изложенных выше положений в целях реализации на практике основополагающих ценностных установок, принципов и мер, сформулированных в настоящей Рекомендации.

## **VII. СОДЕЙСТВИЕ ОСУЩЕСТВЛЕНИЮ НАСТОЯЩЕЙ РЕКОМЕНДАЦИИ**

108. ЮНЕСКО призвана стать главным учреждением Организации Объединенных Наций в вопросах популяризации и распространения настоящей Рекомендации и, соответственно, будет работать во взаимодействии с другими органами ООН, включая, в том числе, Группу высокого уровня Генерального секретаря Организации Объединенных Наций по цифровому сотрудничеству, КОМЕСТ, Международный комитет по биоэтике (МКБ), Межправительственный комитет по биоэтике (МПКБ), Международный союз электросвязи (МСЭ) и другие соответствующие органы Организации Объединенных Наций, занимающиеся вопросами этики искусственного интеллекта.

109. Кроме того, ЮНЕСКО будет взаимодействовать с другими международными организациями, включая, в том числе, Африканский союз (АС), Ассоциацию государств Юго-Восточной Азии (АСЕАН), Совет Европы (СЕ), Евразийский экономический союз (ЕАЭС), Европейский союз (ЕС), Организацию экономического сотрудничества и развития (ОЭСР) и Организацию по безопасности и сотрудничеству в Европе (ОБСЕ), а также Институт инженеров по электротехнике и радиоэлектронике (ИИЭР) и Международную организацию по стандартизации (ИСО).

## **VIII. ЗАКЛЮЧИТЕЛЬНЫЕ ПОЛОЖЕНИЯ**

110. Настоящую Рекомендацию следует рассматривать как единое целое, а сформулированные в ней основополагающие ценностные установки и принципы – как взаимодополняющие и взаимосвязанные. Каждый из принципов следует рассматривать с учетом основополагающих ценностных установок.

111. Ничто в настоящей Рекомендации не может быть истолковано как дающее тому или иному государству, другому социальному субъекту, группе или отдельному лицу какие-либо основания заниматься любой деятельностью или совершать любые действия в нарушение прав человека, основных свобод, человеческого достоинства и принципов уважения к жизни на Земле и за ее пределами.