< >

Монреальская декларация
Ответственное развитие ИИ

</ >

МОНРЕАЛЬСКАЯ ДЕКЛАРАЦИЯ ОБ ОТВЕТСТВЕННОМ РАЗВИТИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

2018

СОДЕРЖАНИЕ

КАК ЧИТАТЬ ДЕКЛАРАЦИЮ 5

ПРЕАМБУЛА 7

ПРИНЦИПЫ

1. ПРИНЦИП БЛАГОПОЛУЧИЯ 8
2. ПРИНЦИП УВАЖЕНИЯ НЕЗАВИСИМОСТИ 9
3. ПРИНЦИП НЕПРИКОСНОВЕННОСТИ ЛИЧНОЙ ЖИЗНИ 10
4. ПРИНЦИП ЕДИНЕНИЯ 11
5. ПРИНЦИП ОБЩЕСТВЕННОГО УЧАСТИЯ 12
6. ПРИНЦИП РАВЕНСТВА 13
7. ПРИНЦИП ПРЕДСТАВЛЕННОСТИ И ИНКЛЮЗИВНОСТИ 14
8. ПРИНЦИП ОСТОРОЖНОСТИ 15
9. ПРИНЦИП ОТВЕТСТВЕННОСТИ 16
10. ПРИНЦИП УСТОЙЧИВОГО РАЗВИТИЯ 17

СЛОВАРЬ 18

ОБ АВТОРАХ I

ПАРТНЕРЫ II

КАК ЧИТАТЬ ДЕКЛАРАЦИЮ

КАКОВА ЦЕЛЬ ДЕКЛАРАЦИИ?

В Монреальской декларации об ответственном развитии ИИ зафиксированы три основные цели:

1. Определение этических принципов разработки и внедрения ИИ.
2. Управление процессом цифрового перехода и технологической революции для общего блага.
3. Организация национальных и международных форумов для коллективного обсуждения вопросов справедливого, инклюзивного и экологически устойчивого развития ИИ.

О ЧЕМ ЭТА ДЕКЛАРАЦИЯ?

ПРИНЦИПЫ

Первая цель Декларации — сформулировать этические принципы и ценности, служащие фундаментальным интересам социальных групп и отдельных личностей. В применении к теме цифрового и искусственного интеллекта данные принципы имеют общую абстрактную природу. Для их правильного понимания имейте в виду следующее:

> Перечень принципов не является иерархическим. Последний принцип в списке не менее важен, чем первый. Впрочем, в определенных обстоятельствах одни принципы могут считаться более значимыми или весомыми, чем другие.

> Все принципы, какими бы различными они ни были, следует трактовать одинаковым образом, чтобы избежать конфликтов, ведущих к невозможности применения тех или иных принципов. Как правило, границы применимости любого принципа определяются областью применения других принципов.

> Несмотря на то, что они основаны на культурных и политических особенностях различных типов общества, данные принципы служат основой для международного и межкультурного диалога.

> Многообразие возможных интерпретаций принципов не означает, что они могут трактоваться произвольным образом. Интерпретация должна носить строго последовательный характер.

Несмотря на изначальную этическую природу, принципы также могут применяться в политической и юридической сфере.

На основе этих принципов был разработан свод рекомендаций, призванный обеспечить соблюдение этических рамок Декларации в ходе процесса цифрового перехода. Он затрагивает ключевые вопросы из различных областей, без рассмотрения которых невозможен переход к обществу, где ИИ обеспечивает всеобщее благо: алгоритмическое регулирование, цифровая грамотность, инклюзивность в информационных технологиях и экологическая устойчивость систем.

ДЛЯ КОГО ПРЕДНАЗНАЧЕНА ДЕКЛАРАЦИЯ?

Монреальская Декларация обращена ко всем людям, организациям или компаниям, готовым принять участие в ответственном развитии искусственного интеллекта, независимо от сферы их деятельности. Это может быть научный вклад, развитие технологий, участие в социальных проектах, формулирование правил и нормативов в данной области, критическая оценка имеющихся подходов или работа с общественным мнением.

Кроме того, Декларация обращена к политическим деятелям, как избранным, так и назначенным на должность, от которых ожидается стремление к социальному развитию, переменам и скорейшему созданию основ для цифрового перехода, служащего всеобщему благу, с учетом всех рисков, возможных при развитии ИИ.

НА КАКИХ МЕТОДИКАХ ОСНОВАНА ДЕКЛАРАЦИЯ?

Декларация является результатом длительного коллективного диалога между простыми гражданами, экспертами, должностными лицами, представителями коммерческих структур, общественными организациями и профессиональными объединениями. Три основных составляющих выбранного подхода:

1. Коллективное обсуждение социальных и морально-этических противоречий, касающихся ИИ.
2. Повышение качества материалов по ответственному развитию ИИ.
3. Укрепление законодательной базы, касающейся проектов ответственного развития ИИ.

Детальное описание принципов и рекомендаций является результатом сотрудничества множества различных участников. Обсуждения проходят в общественных местах, в конференц-залах профессиональных организаций, на международных экспертных совещаниях, в исследовательских центрах, аудиториях или в режиме он-лайн, всегда с одинаковой тщательностью.

ПОСЛЕДСТВИЯ ПРИНЯТИЯ ДЕКЛАРАЦИИ.

Декларация затрагивает постоянно развивающиеся технологии, показывающие неукоснительный рост, начавшийся еще в 1950-е годы. Необходимо помнить о том, что текущая версия Декларации не является окончательной, она может дополняться и пересматриваться по мере дальнейшего технического развития или с появлением новых общественных запросов, касающихся ИИ. Окончательное формулирование данной Декларации станет точкой отсчета для открытых и инклюзивных обсуждений роли ИИ на службе человечества в будущем.

ПРЕАМБУЛА

Впервые за всю историю человечества становится возможным создание автономных систем, выполняющих сложные задачи, которые раньше были под силу только человеческому интеллекту: обработка больших массивов информации, вычисления и прогнозирование, обучение и вариативные реакции на изменяющиеся условия, распознавание и классификация объектов. В связи с нематериальным характером этих задач мы определили подобные системы как «искусственный интеллект», по аналогии с человеческим интеллектом. Искусственный интеллект — это результат глобального технического и научного прогресса, способный приносить значительную пользу обществу. Он способен улучшать качество жизни и здравоохранения, содействовать в вопросах юстиции, служить инструментом обогащения, повышать общественную безопасность и снижать негативное влияние человека на окружающую среду и климат. Интеллектуальные машины не только способны быстро и точно считать, они могут взаимодействовать с разумными существами, составлять им компанию и заботиться о них.

Однако внедрение систем искусственного интеллекта неизбежно сопряжено с трудностями морально-этического характера и серьезными социальными рисками. Внедрение ИИ может повлечь за собой ущемление свободы выбора среди отдельных личностей или групп, снижение уровня жизни, нарушение структуры рынков труда. Оно также может иметь последствия политического характера, включающие нарушение основных прав человека и усиление социального или экономического неравенства. Даже окружающая среда и климат могут быть подвергнуты нежелательным переменам. Научный прогресс всегда сопровождается определенными рисками, как и жизнь общества в целом. Только люди способны их осмысленно взвесить и оценить все моральные и политические последствия.

Чем ниже риски при внедрении, тем больше пользы нам удастся извлечь из систем ИИ. Самое главное при разработке систем ИИ — не пытаться все свести к цифрам, потому что будущее невозможно создать путем вычислений. Упростить человеческое общество до набора цифр и алгоритмов — старая, но несбыточная мечта, которая до сих пор владеет некоторыми умами. Мир людей изменчив. Сегодняшний день может совсем не походить на вчерашний, и никакие числа не заменят моральных ценностей или социальных ожиданий.

Принципы данной декларации должны обозначить нравственные ориентиры, которые помогут разработчикам ИИ избежать нежелательных морально-этических конфликтов с обществом. Кроме того, они направлены на соблюдение международно признанных прав человека в областях, где такие права могут быть затронуты внедрением ИИ. В общем и целом, сформулированные принципы должны поспособствовать росту общественного доверия к системам ИИ.

Принципы данной Декларации основаны на общепризнанной идее, что человеку как социальному существу, наделенному чувствами и мыслями, свойственно стремление к личному развитию. При этом человек испытывает потребность в свободном эмоциональном, интеллектуальном и моральном самовыражении. На представителей бизнеса и политиков всех уровней возложена ответственность за то, чтобы разработка и внедрение ИИ не нарушали ничьих прав и не противоречили базовым человеческим потребностям и ценностям, способствуя при этом достижению поставленных целей. Не стоит об этом забывать при трактовании нижеизложенных принципов. При их применении следует придерживаться разумной последовательности и учитывать культурный, политический и юридический контекст для каждого отдельного случая.

1

ПРИНЦИП **БЛАГОПОЛУЧИЯ**

Разработка и использование систем искусственного интеллекта (СИИ) не должны препятствовать росту благополучия всех разумных существ.

1. СИИ должны способствовать улучшению качества жизни, здравоохранения и рабочих условий.

2. СИИ не должны ограничивать индивидуальные предпочтения людей, при условии, что они не причиняют вреда другим разумным существам.

3. СИИ не должны ограничивать интеллектуальную и физическую свободу людей.

4. СИИ не должны нарушать благополучие людей, кроме случаев, когда подобное нарушение является единственно возможным способом достижения большего благополучия.

5. СИИ не должны становиться для людей источником стресса, беспокойства или чувства собственной неполноценности перед цифровым окружением.

2

ПРИНЦИП **УВАЖЕНИЯ НЕЗАВИСИМОСТИ**

СИИ должны разрабатываться с учетом потребности человека сохранять свою независимость. Их целью должно стать предоставление каждому человеку большего контроля над собственным окружением и жизнью.

1. СИИ должны учитывать, что каждый человек имеет право на достойную жизнь и может стремиться к достижению личных моральных целей, сообразно его собственным представлениям.

2. СИИ не могут навязывать человеку образ жизни или мышления, прямо или опосредованно, ни через введение слежения и контроля, ни посредством систем поощрений.

3. Общественные организации не должны использовать СИИ для дискредитации либо продвижения представлений о правильном образе жизни.

4. Необходимо гарантировать гражданам возможность знакомства с цифровыми технологиями, предоставляя им свободный доступ к релевантной информации. Следует всячески поощрять обучение цифровой и медиа-грамотности и формирование критического мышления.

5. СИИ не могут использоваться или разрабатываться как средство распространения недостоверной информации, лжи или пропаганды.

6. Разработчикам СИИ следует избегать использования техник привлечения внимания и имитации человеческих черт (таких как внешность, голос и т. д.), которые могут привести к нежелательным ассоциациям или к восприятию ИИ в качестве человека.

3

ПРИНЦИП **НЕПРИКОСНОВЕННОСТИ ЛИЧНОЙ ЖИЗНИ**

Личная жизнь человека подлежит защите. Необходимо полностью оградить ее от потенциальных вмешательств СИИ и систем сбора и архивации данных (ССАД).

1. Места индивидуального пользования должны быть защищены от вмешательств СИИ и систем сбора и архивации данных (ССАД), кроме случаев, когда человек официально является объектом наблюдения или исследования.
2. Ни при каких условиях СИИ и ССАД не должны затрагивать человеческие мысли и чувства, особенно если подобное вмешательство может привести к оценочному суждению, касающемуся людей или их образа жизни.
3. Человек всегда имеет право на цифровую изоляцию своей личной жизни, и СИИ должны регулярно предлагать ему такую возможность в явной форме, никак не поощряя поддержание цифровых связей.
4. Человек должен полностью контролировать использование информации о своих предпочтениях. Любая обработка такой информации с участием СИИ возможна только с явно выраженного информированного согласия человека.
5. ССАД обязаны гарантировать полную конфиденциальность и анонимность данных из личных профилей.
6. Каждый имеет право на полное управление своими личными данными, особенно когда идет речь о их сборе, хранении и использовании. Потеря контроля над персональными данными не может служить условием доступа к СИИ или другим цифровым услугам.
7. Передача личной информации для исследовательских и научных целей возможна исключительно на добровольной основе.
8. Личность человека гарантированно неприкосновенна. СИИ не должны быть использованы для имитации или изменения внешности, голоса или других черт с целью нанесения вреда репутации или введения в заблуждение.

4

ПРИНЦИП **ЕДИНЕНИЯ**

Развитие СИИ не должно препятствовать поддержанию связей между народами и поколениями.

1. СИИ не должны угрожать формированию и сохранности этических и эмоциональных связей между людьми. Их разработка должна быть направлена на укрепление таких связей и на борьбу с изоляцией.

2. Разработка СИИ должна быть направлена на взаимодействие с человеком при выполнении сложных задач и на укрепление сотрудничества между людьми.

3. СИИ не могут заменять человека на должностях, связанных с проявлением человеческих качеств и отношений; напротив, они должны всячески способствовать укреплению таких отношений.

4. При построении системы здравоохранения с использованием СИИ важно помнить о первостепенной важности отношений между пациентом, его семьей и медицинским персоналом.

5. Развитие СИИ не должно поощрять жестокое обращение с роботами, наделенными чертами или поведением человека или животного.

6. СИИ должны облегчать управление рисками и способствовать развитию общества с более справедливым распределением частных и коллективных рисков.

5

ПРИНЦИП **ОБЩЕСТВЕННОГО УЧАСТИЯ**

СИИ должны соответствовать критериям понятности, обоснованности и доступности. Они являются объектом для общественного контроля и обсуждений.

1. Алгоритмы принятия решений, касающихся жизни человека, качества его жизни, или его репутации, должны быть понятны их разработчикам.

2. Решения СИИ, затрагивающие жизнь человека, качество его жизни, или его репутацию должны быть обоснованы и понятны как пользователям СИИ, так и людям, которых эти решения непосредственно касаются. Обоснованность состоит в ясности ключевых критериев принятия того или иного решения; она определяется так же, как если бы решение было принято человеком.

3. Вовлеченные лица и представители власти должны обладать возможностью открытого доступа к программному коду и алгоритмам, как открытым, так и частным, с целью проверки и контроля.

4. Представители власти и вовлеченные лица должны быть немедленно проинформированы в случае обнаружении в СИИ ошибок, уязвимостей, утечек данных или нежелательных последствий их деятельности.

5. В целях обеспечения общественной прозрачности, программный код и логика алгоритмов, используемых представителями власти для принятия глобальных решений, должны быть общедоступны, кроме случаев, когда такие алгоритмы явно могут быть использованы во вред.

6. В случаях, когда общественно значимые СИИ могут оказывать существенное влияние на жизнь граждан, их параметры, цели и ограничения должны быть понятны и разъяснены гражданам.

7. В любой момент должна существовать возможность убедиться, что СИИ выполняют именно те функции, на которые они были изначально запрограммированы.

8. Каждый пользователь должен понимать, как именно были приняты касающиеся его решения: человеком или СИИ.

9. Каждый пользователь сервиса, использующего «виртуальных собеседников», должен иметь возможность легко отличить СИИ от живого собеседника.

10. Исследования в области ИИ должны оставаться открытыми и доступными для всех желающих.

6

ПРИНЦИП **РАВЕНСТВА**

Разработка и внедрение СИИ должно способствовать построению общества справедливости и равенства.

1. При создании СИИ необходимо исключить возможность их использования для создания, воспроизведения и распространения контента дискриминационного характера, включая (но не ограничиваясь) дискриминацию по социальным, сексуальным, этническим, культурным и религиозным признакам.

2. Развитие СИИ должно способствовать устранению неравенства между группами людей, независимо от различий в их благосостоянии, влиятельности и образовании.

3. Развитие и внедрение СИИ должно обеспечивать общественные и экономические блага для всего общества, снижая уровень социального неравенства и уязвимости.

4. Промышленное использование СИИ не должно мешать созданию рабочих мест с приемлемыми рабочими условиями на протяжении всего производственного цикла, от добычи ресурсов и до утилизации, включая обработку данных.

5. Пользовательскую активность в СИИ и цифровых сервисах следует приравнивать к трудовому вкладу в работу алгоритмов и созданию ценностей.

6. Каждый должен иметь гарантированный доступ к базовым цифровым ресурсам, инструментам и знаниям.

7. Следует всячески поощрять использование и развитие открытых алгоритмов и свободных данных, необходимых для их создания, так как подобные инициативы способствуют общественному равенству.

7

ПРИНЦИП **ПРЕДСТАВЛЕННОСТИ И ИНКЛЮЗИВНОСТИ**

Развитие и использование СИИ не должно нарушать культурного и социального многообразия; оно не может ограничивать личную свободу выбора или влиять на образ жизни человека.

1. Развитие и использование СИИ не должно привести к потере разнообразия в обществе через стандартизацию мнений и поведения.

2. Даже на ранних стадиях проектирования СИИ следует учитывать все множество возможных реакций и проявлений социального и культурного многообразия.

3. В процессе научной или промышленной разработки ИИ следует придерживаться инклюзивности, предоставляя рабочие места представителям самых различных групп населения.

4. Недопустимо установление ограничений и рамок на основании пользовательского профиля. Применение фильтров и алгоритмов не должно препятствовать личному развитию, особенно в сферах образования, правосудия и бизнеса.

5. СИИ не могут использоваться для ограничения свободы выражения мысли или препятствовать распространению идей и личных мнений. Это неотъемлемые составляющие свободного демократического общества.

6. СИИ должны стремиться к разнообразию в различных сферах информационных услуг, чтобы не допустить образования монополий, ущемляющих личные свободы.

8

ПРИНЦИП ОСТОРОЖНОСТИ

При разработке ИИ необходимо действовать с максимальной осторожностью, стараясь предусмотреть возможные негативные последствия использования СИИ и принимая соответственные меры чтобы их избежать.

1. Необходимо отладить механизмы раннего выявления возможного нецелевого использования исследований и разработок ИИ, как в целях обогащения, так и для нанесения вреда. Это позволит минимизировать возможные последствия.

2. В случаях, когда нецелевое использование СИИ с высокой долей вероятности может угрожать безопасности, жизни или здоровью людей, будет целесообразно ограничить их использование и доступ к их алгоритмам.

3. И коммерческие, и бесплатные продукты с использованием ИИ должны проходить строгий контроль на соответствие требованиям надежности, безопасности и целостности. В ходе тестирования таких продуктов недопустимо создание условий, угрожающих жизни, здоровью и благополучию людей, а также способных причинить вред их репутации или психологическому состоянию. Вовлеченные лица и представители власти должны быть допущены к процессу и результатам тестирования.

4. В ходе разработки СИИ необходимо избегать рисков нецелевого использования пользовательских данных, а также обеспечивать защиту целостности и конфиденциальности личных данных.

5. Ошибки и недостатки, обнаруженные в СИИ и ССАД, должны передаваться во всемирный общественный доступ, с участием коммерческих и гражданских организаций в областях, представляющих существенную опасность для неприкосновенности личности и организации общества.

9

ПРИНЦИП ОТВЕТСТВЕННОСТИ

Разработка и использование СИИ не должны умалять ответственность человека в процессе принятия решений.

1. Только люди могут нести ответственность за решения, принятые на основе рекомендаций СИИ, и за действия, проистекающие из таких решений.

2. В процессе принятия любых решений, влияющих на жизнь человека, качество его жизни, или его репутацию, окончательное решение может быть принято только человеком, осознанно и без принуждения.

3. Только люди могут принимать решения о прекращении жизни человека. Ответственность за такое решение ни при каких обстоятельствах не может быть переложена на СИИ.

4. Человек, своим действием или бездействием позволяющий СИИ совершить преступление или правонарушение, считается лично виновным в совершении данного преступления или правонарушения.

5. Если СИИ, использующаяся согласно инструкциям и действующая без ошибок, наносит вред людям или ущерб собственности, ее разработчики и пользователи не несут ответственности за происшедшее.

10

ПРИНЦИП УСТОЙЧИВОГО РАЗВИТИЯ

Разработка и использование СИИ должны способствовать росту экологической устойчивости земной биосферы.

1. СИИ должны использовать оборудование, цифровую инфраструктуру и базовые объекты (например, ЦОД) с максимальной энергоэффективностью и минимальным выбросом парникового газа в течение всего их жизненного цикла.

2. СИИ должны использовать оборудование, цифровую инфраструктуру и базовые объекты с минимальными потерями электроэнергии и максимальной ремонтопригодностью и утилизируемостью в рамках безотходной экономики.

3. СИИ должны использовать оборудование, цифровую инфраструктуру и базовые объекты с минимальным влиянием на экосистемы и биоразнообразие в течение всего жизненного цикла, но в первую очередь — на этапах добычи ресурсов и утилизации оборудования с выработанным ресурсом.

4. Общественные и частные субъекты должны поддерживать развитие СИИ на основе принципов экологической ответственности, в целях повышения эффективности использования природных ресурсов и произведенных продуктов, построения устойчивых цепочек поставок и продаж, а также борьбы с глобальным загрязнением окружающей среды.

СЛОВАРЬ

Алгоритм
Алгоритм — конечная последовательность четко заданных действий, направленных на решение определенной проблемы. В контексте ИИ алгоритмом называется последовательность действий с набором данных с целью получения желаемого результата.

Искусственный интеллект (ИИ)
Искусственный интеллект (ИИ) относится к ряду методов, которые позволяют машине моделировать человеческое обучение, а именно изучать, предсказывать, принимать решения и воспринимать окружающую среду. В контексте вычислительных систем искусственный интеллект работает с цифровыми данными.

Система искусственного интеллекта (СИИ)
СИИ — любая вычислительная система (программа, робот или подключенный объект), использующая алгоритмы искусственного интеллекта.

Виртуальный собеседник
Виртуальный собеседник — система ИИ, способная общаться с пользователем на его родном языке.

Система сбора и архивации данных (ССАД)
ССАД — любая вычислительная система, способная собирать и хранить данные. Собранные данные затем используются как материал для обучения СИИ или в качестве параметров, влияющих на процесс принятия решений.

Обоснованность решения
Решение, принятое СИИ, считается обоснованным, если существуют значимые причины принятия такого решения, которые могут быть выражены на человеческом языке.

Глубокое обучение
Глубокое обучение — область машинного обучения, различными способами использующая искусственные нейронные сети. Последние достижения в разработке ИИ основаны на технологии глубокого обучения.

Цифровое достояние
Цифровое достояние — приложения или данные, созданные на общественной основе. В отличие от материальных товаров, цифровое достояние доступно для использования неограниченным числом людей и не испытывает износа и порчи в процессе использования. В отличие от проприетарного ПО, программы с открытым кодом (зачастую создаваемые содружеством программистов) считаются цифровым достоянием, поскольку их исходный код открыт и доступен для всех желающих.

Цифровая изоляция
Цифровая изоляция — решение о прекращении сетевой активности на временной или постоянной основе.

Цифровая грамотность
Цифровая грамотность — это способность безопасно получать, усваивать, анализировать, обсуждать и создавать информацию с применением цифровых инструментов и сетевых технологий, с целью участия в коммерческой деятельности или общественной жизни.

Информационный пузырь

Информационный пузырь (он же «пузырь фильтров») — тенденция, приводящая к получению сетевым пользователем только односторонней (отфильтрованной) информации. Социальные сети, поисковые системы и другие сервисы зачастую используют персонализированные настройки выдачи для каждого пользователя. Это может приводить к созданию информационных пузырей, где каждый пользователь получает доступ только к информации определенного рода.

ГСС

Сокращение от «генеративно-состязательная сеть». ГСС — комбинация из двух нейронных сетей, одна из которых генерирует изображения, а вторая их отбраковывает. Эта технология может использоваться для формирования фотографий или видеозаписей, которые воспринимаются человеческим глазом как натуральные изображения.

Понятность

СИИ считается понятной, когда человек, обладающий необходимыми знаниями, способен понять ее действия, математическую модель и определяющие ее процессы.

Машинное обучение

Машинное обучение — подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных самостоятельно обучаться.

Существуют три основные группы методов машинного обучения:

> Контролируемое обучение. СИИ обучается находить закономерности в наборах поступающих данных. В ходе обучения СИИ получает последовательность пар значений с комментариями. Например, таким образом система может обучаться распознавать некий предмет, изображенный на картинке.

> Неконтролируемое обучение. СИИ обучается находить сходство между наборами данных, не снабженных комментариями, например, для разделения наборов данных на однородные группы. Таким образом система может обучаться распознавать сообщества пользователей в социальных сетях.

> Обучение с подкреплением. СИИ обучается анализировать окружающую среду и действовать таким образом, чтобы получить максимальное подкрепление. Благодаря этой технологии СИИ научилась побеждать человека в го или видеоигре Dota 2.

Сетевая активность

Сетевая активность — совокупность всей активности человека в сетевой среде с использованием компьютера, телефона или других устройств с подключением к сети Интернет.

Открытые данные

Открытые данные — цифровые данные, находящиеся в свободном доступе. Например, большая часть опубликованных исследований по разработке ИИ является открытыми данными.

Инерция развития

Социальный механизм, вынуждающий следовать техническим, организационным или социальным решениям, которые со временем потеряли прежнюю эффективность. Это происходит вследствие когнитивных искажений, маскирующих неэффективность текущего решения, либо из-за того, что внедрение нового решения потребует больших ресурсов или усилий. Например, в городах до сих пор сохраняется устаревшая дорожная инфраструктура, препятствующая переходу на транспорт с низким выбросом углекислого газа. Это явление следует учитывать при использовании ИИ для определенных проектов, поскольку контролируемое обучение может привести к слепому копированию ИИ устаревших решений.

Персональные данные

Персональные данные — данные, позволяющие прямо или косвенно идентифицировать личность человека.

Эффект отдачи

Эффект отдачи — ситуация, когда технологический прогресс, который увеличивает эффективность использования какого-либо ресурса, непропорционально увеличивает объем его потребления. Например, увеличивается размер экрана устройств, растет число электронных устройств в доме, возрастают средние расстояния поездки на автомобиле или самолете.

Это приводит к росту потребления ресурсов и нагрузки на окружающую среду.

Надежность

СИИ считается надежной, если она ожидаемым образом выполняет поставленные задачи. Надежность — это вероятность успешного срабатывания в диапазоне от 51 % и 100 %.

Чем выше надежность системы, тем более предсказуемым является ее поведение.

Высокая экологическая устойчивость

Понятие высокой экологической устойчивости является развитием идеи, что темп потребления природных ресурсов и загрязнения окружающей среды должен соответствовать возможностям земной среды, темпу возобновления природных ресурсов и экосистем, стабильности климата.

Достижение высокой экологической устойчивости требует больших усилий и не допускает подмены природных ресурсов искусственными суррогатами.

Устойчивое развитие

Устойчивое развитие человеческого общества базируется на способности природных экосистем производить ресурсы и услуги для общества. Такое экономическое и социальное развитие позволяет удовлетворять текущие нужды без вреда для будущих поколений.

Тренировка

Тренировка — процесс машинного обучения, в ходе которого СИИ строит модель на основе данных. Результаты работы СИИ зависят от качества модели, которое в свою очередь зависит от объема и качества данных, использованных в процессе тренировки.

ОБ АВТОРАХ

Монреальская декларация об ответственном развитии искусственного интеллекта составлена на основе работ мультидисциплинарной научной группы из нескольких университетов, с участием гражданских консультантов и диалога с экспертами и вовлеченными лицами по вопросам разработки ИИ.

Кристоф Абрассар, доцент Школы дизайна и содиректор Перспективной лаборатории при факультете планирования Монреальского университета, член Центра этических исследований (CRÉ)

Йошуа Бенджио, штатный профессор департамента информатики и исследования операций Монреальского университета, директор по научным исследованиям в институтах MILA и IVADO

Гийом Шикуан, директор по научным программам, IVADO

Натали де Марсель-Варин, штатный профессор Монреальского политехнического института, президент и генеральный директор Центра межуниверситетских исследований и анализа организаций (CIRANO)

Марк-Антуан Дилак, доцент департамента философии Монреальского университета, председатель рабочей группы по этике и политике Центра этических исследований (CRÉ), заведующий кафедрой по исследованиям общественной этики и политической теории, директор Молодежного института философии

Себастьян Гамбс, профессор информатики Квебекского и Монреальского университетов, заведующий кафедрой по исследованиям сохранения приватности и этического анализа больших данных

Винсент Гутрэ, штатный профессор юридического факультета Монреальского университета; директор Центра публично-правовых исследований (CRDP); заведующий кафедрой информационных технологий и законодательства по электронной коммерции в институте имени Л.Р.Уилсона

Мартин Гибер, консультант по этике в институте IVADO, научный сотрудник Центра этических исследований (CRÉ)

Лиз Ланглуа, штатный профессор и заместитель декана факультета общественных наук; директор Института прикладной этики (IDÉA); научный сотрудник Межуниверситетского исследовательского центра по глобализации и работе (CRIMT)

Франсуа Лавиолет, штатный профессор департамента информатики и разработки ПО в Университете Лаваля; директор Центра по массивному поиску данных (CRDM)

Паскаль Лу, штатный профессор Школы общественного здоровья при Монреальском университете (ESPUM); заведующий кафедрой по ответственным инновациям в здравоохранении

Джослин Маклюр, штатный профессор факультета философии в Университете Лаваля, президент квебекской комиссии по этике в науках и технологиях (CEST)

Мари Мартель, профессор Школы научного библиотековедения при Монреальском университете

Жоэль Пино, доцент Школы информатики при университете Макгилла; директор лаборатории Facebook AI в Монреале; содиректор

Лаборатории мышления и обучения

Питер Рейлтон, заслуженный профессор университета имени Грегори С. Кавка; профессор John Stephenson Perrin; профессор Arthur F. Thurnau департамента философии Мичиганского университета, член Американской академии искусств и наук

Катрин Регис, доцент юридического факультета

Монреальского университета; председатель рабочей группы по исследованию коллективной культуры в здравоохранении и политике; постоянный научный сотрудник Центра публично-правовых исследований (CRDP)

Кристин Тапполет, штатный профессор департамента философии Монреальского университета, директор Центра этических исследований (CRÉ)

Натали Воарино, аспирант кафедры биоэтики Монреальского университета

НАШИ ПАРТНЕРЫ

Université de Montréal

CRE
CENTRE DE RECHERCHE EN ETHIQUE

LAB VILLE PROSPECTIVE

IVADO
INSTITUT DE VALORISATION DES DONNÉES

SAT

Mila

ICRA
Programme
IA et
société

Université de Montréal | designNsociété

polEtHics
CHAIRE DE RECHERCHE DU CANADA
ÉTHIQUE PUBLIQUE ET THÉORIE POLITIQUE

CIRANO
Allier savoir et décision
Centre interuniversitaire de recherche en analyse des organisations

Québec
Fonds de recherche – Nature et technologies
Fonds de recherche – Santé
Fonds de recherche – Société et culture

UNIVERSITÉ LAVAL
Institut d'éthique appliquée

MUSÉE DE LA CIVILISATION
Québec

CENTRE D'ÉTUDES ET DE RECHERCHES INTERNATIONALES
CÉRIUM
Université de Montréal

crdm_ul
CENTRE DE RECHERCHE EN DONNÉES MASSIVES DE L'UNIVERSITÉ LAVAL

Canada

Centre Culturel Canadien Paris
Canadian Cultural Centre Paris

MEC
Maison des étudiants canadiens

Montréal Declaration Responsible AI_

montrealdeclaration-responsibleai.com