Перевод

Следующий документ является переводом на русский язык документа “AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defense”, изначально опубликованного на английском языке Советом по оборонным инновациям (СпОИ) Министерства обороны США (МО) 31 октября 2019 г. Центр безопасности и новых технологий (CSET) заказал этот перевод на русский язык, чтобы привлечь внимание мировой аудитории к ИИ-принципам, сформулированным советом СпОИ. CSET – это непартийный аналитический центр, базирующийся в Джорджтаунском университете и занимающийся изучением воздействия новейших технологий на безопасность. CSET не входит в структуру СпОИ или МО. Перевод, представленный CSET, не является официальным правительственным переводом ИИ-принципов СпОИ. Перевод ИИ-принципов СпОИ, представленный CSET, не означает одобрения этих принципов.

Оригинальная версия этого документа на английском языке доступна по следующей ссылке: https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENT.PDF

Принципы ИИ:

Рекомендации по этичному использованию искусственного интеллекта Министерством обороны

Совет по оборонным инновациям

I. Цель

Руководство Министерства обороны (МО) обратилось в Совет по оборонным инновациям (Defense Innovation Board - СпОИ) с просьбой разработать этические принципы для искусственного интеллекта (ИИ) при проектировании, разработке и развертывании ИИ для боевых и небоевых целей. Основываясь на фундаменте существующих этических, правовых и политических концепций и отвечая на сложности быстро развивающейся сферы ИИ, Совет стремился разработать принципы, согласующиеся с миссией Министерства сдерживать развязывание войны и обеспечивать безопасность страны. В данном документе дается сводка проекта СпОИ, и он включает в себя краткую подоплеку, основные положения существующих этических принципов МО, выходящих за пределы ИИ, набор предложенных этических принципов ИИ и набор рекомендаций для содействия принятию этих принципов и продвижения более широкой цели безопасности, защиты и надежности ИИ. Полный отчет СпОИ включает в себя детальные разъяснения и обращается к более широкому историческому, политическому и теоретическому контексту для дачи этих рекомендаций. Он доступен по адресу innovation.defense.gov/ai.

СпОИ является независимым федеральным консультативным советом, предоставляющим советы и рекомендации высшему командованию МО; он не говорит от лица МО. Цель данного отчета – ранняя попытка предоставить почву для наводящего на размышления диалога внутри Министерства и внешне в нашем более широком обществе. Министерство несет полную ответственность за определение того, как лучше всего отнестись к рекомендациям из данного отчета.

II. Подоплека

Почему МО должно уделить первостепенное внимание этике ИИ? ИИ трансформирует наше общество и влияет на способы ведения бизнеса, социального взаимодействия и ведения войны.[1] Во многих отношениях поле ИИ находится на этапе становления. Недавние быстрые продвижение в сфере вычислительной техники сделали возможным прогресс применений ИИ, которые десятилетиями оставались теоретическими. Тем не менее, практические применения ИИ зачастую оказываются шаткими, и дисциплина разработки ИИ развивается, оставляя в зачаточном состоянии нормы использования ИИ. В целом государственный сектор, частный бизнес, академические организации и гражданское общество вовлечены в дебаты, ведущиеся по поводу обещания, опасности и надлежащего использования ИИ. Национальная безопасность – важнейшая грань этих дебатов. Теперь пришло время на этой ранней стадии возрождения интереса к ИИ провести серьезные обсуждения норм разработки ИИ и использования в военном контексте – задолго *до* возможного инцидента.

Наши противники и соперники распознали преобразующий потенциал ИИ и интенсивно инвестируют в него, модернизируя свои силы в ходе активного вовлечения в провокационные действия по всему земному шару. Китай решительно и открыто настроен

---

[1] См. Сводку по стратегии Министерства обороны относительно искусственного интеллекта в 2018 году: Использование ИИ для улучшения нашей безопасности и процветания и Стратегия национальной безопасности Соединенных Штатов Америки, декабрь 2017 г.

на то, чтобы стать мировым лидером в сфере ИИ к 2030 году, и тратит миллиарды долларов на то, чтобы обрести преимущество.[2] Россия тоже интенсивно инвестирует в применения ИИ и тестирует свои системы в реальных боевых сценариях использования.[3] Сокрушительная реальность будущего МО ясна для генерал-лейтенанта Джека Шэнэхэна (Jack Shanahan), директора объединенного центра ИИ (Joint AI Center - ОЦИИ): "Я не хочу видеть будущего, в котором наши потенциальные противники имеют военную силу с поддержкой ИИ, а мы нет... Я не могут тратить часы или дни на принятие решений. Могут оставаться лишь секунды и микросекунды, когда может использоваться ИИ."[4]

Стратегия национальной обороны 2018 года (National Defense Strategy - СНО) призывает к более крупным инвестициям в ИИ и автономные системы, чтобы обеспечить США конкурентоспособными военными преимуществами.[5]ИИ-стратегия Министерства обороны, согласованная с СНО, утверждает, что МО полно решимости использовать потенциал ИИ, “чтобы решительно трансформировать все функции Министерства, тем самым поддерживая и защищая военнослужащих, обеспечивая безопасность жителей США, защищая союзников и партнеров, а также улучшая экономичность, эффективность и скорость наших операций.”[6] ИИ-стратегия далее отмечает, что она “будет излагать свое видение и руководящие принципы для использования ИИ правомерным и этическим образом, чтобы продвигать наши ценности.”[7] Напирая на необходимость увеличенного взаимодействия с академическими организациями, частным бизнесом и международным сообществом с целью “продвижения этики использования ИИ и безопасности в военном контексте”, Министерство подчеркнуло свою приверженность к этичному и ответственному развитию и развертыванию ИИ. *“Лидерство в военной этике и безопасности ИИ”* – это действительно один из пяти столпов стратегии.

МО не является первой организацией, распознавшей важность создания этических принципов для разработки и использования ИИ, но СпОИ заметил, что многие существующие наборы таких принципов вызывают больше вопросов, чем дают ответов о пределах допустимого использования ИИ. Важно отметить, что в крайне важной области национальной безопасности США оказываются в состоянии технологического соперничества с авторитарными державами, продвигающими использование ИИ способами, несовместимыми с правовыми, этическими и моральными нормами демократических стран. Наша цель – заложить принципы, предложенные здесь, в долгосрочную этическую концепцию МО – ту, что выдержит приход и развертывание появляющихся военных технологий или технологий двойного использования в течение десятилетий, и отразит наши демократические нормы и ценности.

Однако мы признаем, что уникальные характеристики и уязвимости ИИ требуют новых

---

[2] Savage, Luiza Ch. и Nancy Scola. “‘Нас превзошли по расходам. Нас опередили’: Уступает ли Америка будущее ИИ Китаю?” Politico. 18 июля 2019 г.
[3] Konaev Margarita, и Samuel Bendett. “Российский бой с поддержкой ИИ: приходит в городе недалеко от вас?” War on the Rocks. 31 июля 2019 г.
[4] См. “Пресс-брифинг Джека Шэнэхэна в Министерстве обороны по поводу инициатив, связанных с ИИ.” Министерство обороны, 30 августа 2019 г.
[5] Стратегия национальной обороны Соединенных Штатов Америки, 2018.
[6] Министерство обороны, Сводка по стратегии Министерства обороны относительно искусственного интеллекта в 2018 году: Использование ИИ для улучшения нашей безопасности и процветания. стр. 4. (Далее “ИИ-стратегия”).
[7] Министерство обороны (n 6) 8.

способов справляться с потенциальными неумышленными негативными последствиями.[8] В области национальной безопасности анализ непредвиденного поведения является ключом при рассмотрении вопроса, можно ли принять на вооружение появившуюся технологию. Неопределенность, связанная с непреднамеренными последствиями, не уникальная для ИИ; она всегда была свойственна всем инженерным применениям. Например, люди строили мосты и здания, а также манипулировали энергией и физическими материалами еще до того, как соответствующие области гражданского строительства и химической инженерии сформировались в виде формальных дисциплин, что приводило ко многим непредвиденным происшествиям.[9] Сегодня, несмотря на недостаток согласованных способов использования ИИ, которые доводили бы до максимума социальную пользу и сокращали бы непреднамеренные последствия, “люди приступают к созданию систем общественного масштаба для вмешательства и принятия решений с участием машин, людей и окружающей среды.”[10] Поэтому этические принципы ИИ должны обогатить дискуссии о том, как развивать все еще зарождающуюся область ИИ безопасными и ответственными способами. Здесь есть параллель с тем, как гражданское строительство и химическая инженерия развивали свои культуры этического поведения. Они делали это, определяя и поддерживая обязательства технического совершенства со стороны исполнителей.

Принимая во внимание ведущиеся глобальные дебаты о том, когда и при каких обстоятельствах подходит применение ИИ в контексте национальной безопасности, важно отметить, что МО имеет обязательство перед американским народом и его союзниками сохранять свое стратегическое и технологическое преимущество над соперниками и противниками, которые могут использовать ИИ для целей, несовместимых с ценностями Министерства. Мы предполагаем, чтобы появляющиеся принципы служили руководством в этом направлении, тогда как МО будет продолжать придерживаться правомерного и ответственного поведения; будет использовать существующий этический фундамент, переводя и адаптируя этику в области ИИ; будет помогать формировать новые международные нормы по использованию ИИ; и будет обеспечивать, чтобы мы одновременно опирались на технологические преимущества, смягчая при этом ее потенциальный вред.

Разнообразие точек зрения. Чтобы помочь Министерству в этом нелегком деле, высшее командование МО попросило СпОИ вовлечь широкую аудиторию для предоставления рекомендаций по возможным этическим принципам ИИ и по тому, как эти принципы могут быть интегрированы в существующую этическую концепцию, в рамках которой Министерство осуществляет свою миссию.

СпОИ провел 15-меясчное исследование, задуманное быть надежным, инклюзивным и прозрачным. Процесс включал в себя сбор публичных комментариев как через Интернет, так и лично; проведение двух публичных слушаний в крупных университетах; а также проведение трех экспертных обсуждений в формате "круглого стола" с участием десятков

---

[8] См. сопроводительный документ для детального описания того, как ИИ представляет разного рода этические проблемы по сравнению с другими технологиями, а также для дальнейшего более глубокого обсуждения принципов.

[9] См. взгляд д-ра Майкла Джордана (Dr. Michael Jordan) на ИИ как на потенциально новую инженерную дисциплину.

[10] Там же.

экспертов из академических организаций, промышленности, гражданского общества и Министерства. Среди участников "круглого стола" были исследователи в области ИИ, получившие премию Тьюринга, отставные генерал-полковники, правозащитники, политологи, активисты по ограничению вооружений, технологические предприниматели и другие представители общественности.[11] Кроме того, Министерство сформировало рабочую группу по этике и принципам МО, включив в нее государственных должностных лиц из близких государств-партнеров, чтобы помочь СпОИ в сборе информации и продвижению сотрудничества. СпОИ также провел командно-штабные учения, чтобы испытать принципы в реалистических политических сценариях, принимая во внимание возможные применения ИИ в боевых действиях. После взвешенного рассмотрения предложений более 100 внутренних и внешних экспертов, отражающих широкий спектр точек зрения, и почти 200 страниц публичных комментариев,[12] СпОИ разработал этот предложенный набор этических принципов ИИ и сопроводительные рекомендации для рассмотрения министром обороны. Эти принципы – специфичные для ИИ – встроены в контекст существующих этических, правовых и политических концепций, которые Министерство использует для руководства в своей деятельности.

Определение ИИ. Искусственный интеллект представляет собой очень широкую дисциплину, определяемую многими различными способами для многих различных целей.[13] Для ясности и направления данного проекта мы используем этот термин в следующем смысле: *множество методов и технологий обработки информации для выполнения целенаправленных задач и средств рассуждений в процессе выполнения задач.* Говоря о более широком спектре, мы используем термин искусственного интеллекта (ИИ); однако, когда мы конкретно адресуемся к система машинного обучения (МО), мы используем термин МО. Более того, мы используем термин “система ИИ”, когда подразумеваем системы, имеющие компонент ИИ в рамках полной системы или системы систем.[14]

Мы используем это определение ИИ, поскольку оно согласуется с тем, как МО смотрела на ИИ-системы и разрабатывала и развертывала их на протяжении последних 40 лет. Оно дает нам возможность проводить тонкое различие между унаследованными системами и новыми системами, такими, как те, что используют машинное обучение. Использование этого термина дает нам возможность подтвердить, что более ранняя и важная работа, достигнутая МО, происходила в рамках существующей этической концепции МО, обрисованной ниже.

Мы также делаем четкое различие, что *ИИ – не то же самое, что принцип автономии.*

---

[11] Полный список участников (давших согласие опубликовать свои имена) приведен в приложении соответствующего сопроводительного документа и на веб-сайте СпОИ.

[12] См. также ссылки и видео публичных комментариев на веб-сайте СпОИ.

[13] Стратегия МО по ИИ в 2018 году определяет ИИ следующим образом: “способность машин выполнять задачи, которые обычно требуют человеческого интеллекта – например, распознавание образов, обучение на опыте, формулирование выводов, построение прогнозов и принятие мер – будь то в цифровом виде или в качестве интеллектуального программного обеспечения в составе автономных физических систем.” Наше определение не препятствует подходу ИИ-стратегии МО, а допускает более широкий спектр применений ИИ, не требующих человеческого интеллекта. Расширенное обсуждение различных определений ИИ приведено в нескольких программных документах в сопроводительной документации.

[14] В некоторых документах предпочтение отдается фразе “системы с поддержкой ИИ”, но для наших целей нет разницы между ИИ-системами и системами с поддержкой ИИ.

Хотя некоторые автономные системы могут использовать ИИ в своей программной архитектуре, это не всегда так. Например, Директива МО 3000.09 относится к автономии в системах вооружения, но она не относится ни к ИИ как к таковому, ни к возможностям ИИ, не принадлежащим системам вооружений.[15]

Наконец, ИИ сам по себе не является ни позитивным, ни негативным.[16] Это только техническое средство, наделяющее возможностями, сходное в этом отношении с электричеством, двигателем внутреннего сгорания и компьют ерами, и именно решения людей определяют, будет ли ИИ способствовать или подрывать наши усилия сделать мир более безопасным и более процветающим.

III. Существующие этические ценности и концепции МО

ИИ есть и будет существенным техническим средством по всему МО в небоевых и боевых функциях. В действительности, ИИ составляет лишь одну из многих технологий, используемых Министерством, и преподносит проблемы тестирования и эксплуатации, аналогичные тем, что возникают с другими крупными, технически сложными системами, которые МО успешно и безопасно развернуло. Во всех этих случаях основанная на ценностях концепция, в рамках которой МО и вооруженные силы осуществляют свои операции[17], а также правовые конструкции, в рамках которой действуют МО и гражданское общество США, включая Конституцию США, Раздел 10 Кодекса США и другие действующие законы, составляют фундамент, на котором должны функционировать этические принципы ИИ. Предложенные специфичные для ИИ принципы, обрисованные ниже, возникают из *существующих и широко принятых* этических и правовых обязательств.

Эта твердо установленная этическая концепция и сопровождающие ее ценности направляют МО в том, как оно принимает и исполняет решения. Это находит свое доказательство в различных заявлениях, программных документах и в существующих правовых обязательствах. Формальные договоренности включают в себя Право вооруженных конфликтов и существующие международные договоры, тогда как многочисленные меморандумы министра обороны подчеркивают важность этичного поведения среди военнослужащих. По отдельности и вместе, эта совокупность доказательств показывает, что этическая концепция МО отражает ценности и принципы

---

[15] См. Директиву МО 3000.09, которая определяет автономную систему вооружения как “система вооружения, которая, будучи однажды активированной, может выбирать цели и открывать огонь по ним без дальнейшего вмешательства со стороны человека-оператора. Это относится и к автономным системам вооружения, действующим под присмотром людей и рассчитанным на то, чтобы позволять человеку-оператору отменять операцию системы вооружений, но способным выбирать цели и открывать по ним огонь без дальнейшего вмешательства человека после своей активации.”

[16] Но это не означает, что технология, включая ИИ, является нейтральной по отношению к ценностям. Технологические артефакты, как и ИИ-системы, отражают ценности проектировщиков, разработчиков и пользователей, а также обществ, в которых они пребывают и принимают решения.

[17] См. Главные ценности Министерства обороны, Главные ценности военно-воздушных сил США, Главные ценности сухопутных войск США, Главные ценности военно-морских сил и морской пехоты США и Главные ценности береговой охраны США.

американского народа и Конституции США.[18] [19]

Особую важность представляет собой обязательство МО поддерживать Право вооруженных конфликтов, поскольку оно является международно признанным правовым ориентиром для поведения всех вооруженных сил.[20] Для США эта совокупность правовых норм и принципов включает в себя договоры, принятые США, такие как Женевские конвенции 1949 года; международное обычное право, следующее из общей и последовательной практики США, исходящей из ощущения правовой обязанности; и руководство МО по праву вооруженных конфликтов.[21]

Существующие правила права вооруженных конфликтов могут применяться, когда в вооруженном конфликте используются новые технологии.[22] Например, в руководстве МО по праву вооруженных конфликтов 2015 года отражается работа, проделанная в 2012 году в связи с Директивой МО 3000.09, чтобы установить, как право вооруженных конфликтов применяется к использованию автономных функций в системах вооружений.[23] Фундаментальные принципы права вооруженных конфликтов предоставляют общее руководство для поведения во время войны, где не применяются конкретные правила, и таким образом обеспечивается концепция для рассмотрения новых правовых и этических вопросов, ставящихся появляющимися технологиями, подобными ИИ. Например, если ИИ был добавлен к вооружению, такое оружие следует рассмотреть на предмет соответствия существующим законным требованиям, таким как требование, чтобы оружие не было рассчитано на причинение излишних страданий, или не было бы по сути неизбирательным. Кроме того, в рамках права вооруженных конфликтов командиры и другие лица, принимающие решения, должны принимать решения из лучших побуждений и на основе доступной им информации и обстоятельств, наложенных на них в то время. Использование ИИ для поддержки принятия командных решений согласуется с обязательствами права вооруженных конфликтов, включая обязанность принимать реально осуществимые меры предосторожности, чтобы снизить риск причинения вреда гражданскому населению и другим покровительствуемым лицам и объектам.[24]

МО имеет надежные процессы внедрения права вооруженных конфликтов, включая обучение, нормативы и процедуры, доклад об инцидентах с заявленными нарушениями, расследования и рассмотрения инцидентов, а также должные корректирующие действия.[25] Чтобы выполнять и облегчать эти действия, МО инвестировало за последние полвека сотни миллиардов долларов, чтобы обеспечить безопасность и надежность своих систем и

---

[18] Право вооруженных конфликтов представляет собой совокупность международно-правовых норм и принципов, принятых по отношению к боевым действиям и устанавливающим законность поведения в рамках вооруженного конфликта.

[19] См. меморандум министра обороны Марка Эспера и меморандум бывшего министра обороны Джеймса Мэттиса.

[20] Хотя право вооруженных конфликтов является важным ориентиром для МО, оно не применяется ко всем ситуациям, в которых Министерство могло применить ИИ. Мы более глубоко освещаем эти ситуации в сопроводительной документации.

[21] См. руководство МО по праву вооруженных конфликтов.

[22] Эти правила основаны на пяти фундаментальных принципах, служащих в качестве основания права вооруженных конфликтов: военная необходимость, гуманность, соразмерность, различение и честь.

[23] МО (n 18) 395.

[24] МО (n 18) § 5.2.3.2 и 5.3.

[25] См.Директива МО 2311.01E.

платформ вооружений, чтобы создавать более точное оружие, уменьшающее потери среди гражданского населения и защищающее гражданскую инфраструктуру при достижении военных целей. Кроме того, МО постоянно поощряет изменения в том, как его персонал поддерживает эти стандарты и ответственно использует эти инструменты.

Стоит упомянуть дополнительный пример: С момента их спуска на воду военные корабли США с атомными установками безопасно плавали более пятидесяти лет без единого инцидента с атомным реактором и радиоактивного выброса, который сказался бы на здоровье людей или морской флоре и фауне. За более чем 162 миллиона миль атомные реакторы безопасно вырабатывали энергию, насчитав более 6900 реакторо-лет безопасной работы.[26]

Мы привели этот пример не как случай, в котором МО следует применить ИИ для своего ядерного арсенала. Этим примером мы подчеркиваем усилия для создания культуры безопасности и точности, которая полностью представляет стандарт, установленный МО для разработки и управления сложными системами. Этот важнейший фундамент МО для расширения его этической культуры на новые технически сложные начинания, такие как разработка и развертывание ИИ.[27]

IV. Этические принципы ИИ для МО

Мы снова заявляем, что использование ИИ должно происходить а контексте существующей концепции МО. Опираясь на этот фундамент, вы предлагаем следующие принципы, более специфичные для ИИ, и отмечаем, что они применимы как боевым, так и не к боевым системам. ИИ является быстро развивающейся сферой, и ни одна организация, разрабатывающая или размещающая сейчас ИИ-системы, либо поддерживающая этические принципы ИИ, не может заявить, что она решила все проблемы, заложенные в следующих принципах. Однако Министерству следует поставить своей целью, что его использование ИИ отвечает следующим принципам:

1. **Ответственность.** Люди должны проявлять мудрость и оставаться ответственными за разработку, развертывание, использование и результаты действий ИИ-систем МО.
2. **Объективность.** МО следует предпринять взвешенные шаги к тому, чтобы избежать непреднамеренной предвзятости при разработке и развертывании боевых и небоевых ИИ-систем, что может неумышленно привести к причинению вреда людям.
3. **Прослеживаемость.** Инженерная дисциплина ИИ в МО должна быть достаточно отработанной, чтобы технические эксперты обладали должным пониманием технологии, процесса разработки и способов работы своих ИИ-систем, включая прозрачные и подконтрольные методологии, источники данных, методику

---

[26] Информационный бюллетень министерства военно-морского флота США и национального управления по ядерной безопасности США “Программа разработки ядреных силовых установок для ВМС.” Сентябрь 2017 г.

[27] За более глубоким описание всех аспектов существующей этической концепции МО обращайтесь к сопроводительной документации.

проектирования и документацию.

4. **Надежность.** ИИ-системы МО должны иметь ясно очерченную и определенную область применения, и безопасность, защищенность и надежность таких систем должна быть проверена и гарантирована в течение всего их срока эксплуатации в пределах области применения.
5. **Управляемость.** ИИ-системы МО должны быть спроектированы и сконструированы так, чтобы выполнять назначенную функцию, обладая при этом возможностью обнаруживать и избегать нанесения непреднамеренного вреда или нарушения, а также возможностью ручного или автоматического отключения или деактивации развернутых систем, которые демонстрируют непреднамеренное угрожающее или прочее поведение.

V. Рекомендации

В ходе разработки предложенных этических принципов ИИ, СпОИ определил полезные действия, которые могут помочь при формулировке и внедрению этих принципов. В конечном счете МО определит, какие принципы будут им приняты, но независимо от точной природы одобренных принципов, следующие двенадцать рекомендаций послужат поддержкой в этой работе:

1. **Формализовать эти принципы по официальным каналам МО.** Объединенному центру ИИ следует рекомендовать Министру обороны выпустить надлежащие сообщения и политики, чтобы обеспечить долговечность этих этических принципов ИИ.
2. **Учредить руководящий комитет по ИИ в МО.** Заместителю Министра обороны следует учредить комитет на высшем уровне, подотчетный ему/ей и отвечающий за обеспечение того, что надзор и исполнение ИИ-стратегии МО, а также ИИ-проекты МО согласуются с этическими принципами ИИ, принятыми в МО. Поддержание этических принципов ИИ требует того, чтобы МО интегрировало их во многие основополагающие аспекты принятия решений, начиная с концептуального уровня, такого как DOTMLPF[28], и заканчивая более осязаемыми связанными с ИИ областями, такими как обмен данными, облачные вычисления, человеческие ресурсы и ИТ-политики.
3. **Культивировать и взращивать поле ИИ-инжиниринга.** Отделу заместителя МО по НИОКР и специализированным лабораториям следует поддерживать рост и становление дисциплины ИИ-инжиниринга, основываясь на прочных инженерных практиках, давно культивированных Министерством, более значительно вовлекая более широкое сообщество исследования ИИ, предоставляя специфические возможности раннего карьерного роста и адаптируя наследие МО в области безопасности и ответственности к полю ИИ, чтобы интегрировать технологию ИИ в более сложные инженерные системы.

[28] DOTMLPF расшифровывается как "доктрины, методы организации, формы подготовки, обеспечения, методы управления, образование личного состава и учебно-материальная база для развития штабов".

4. **Усилить обучение в МО и расширить учебные программы.** Каждая воинская служба, оперативное командование, подразделение аппарата МО, оборонное агентство и оборонные войсковые объекты должны ввести программы для обучения и подготовки, отвечающие связанным с ИИ навыкам и знанию соответствующего персонала МО.[29] Следует сделать широко доступными различные программы обучения и подготовки в сфере ИИ, начиная с младшего персонала и заканчивая ИИ-инженерами и старшим руководством; в них следует использовать существующий цифровой контент в сочетании со специальными инструкциями, написанными руководителями и экспертами.[30] Необходимо, чтобы младшие офицеры, сержантский и рядовой состав, а также вольнонаемные прошли обучение и подготовку по ИИ в начале своей карьеры, и чтобы МО предоставляло возможности для продолжения обучения на протяжении всей их карьеры посредством системы официального профессионального военного образования и практического применения.
5. **Инвестировать в исследования по новым аспектам безопасности ИИ.** Отделу заместителя МО по политике и Управлению общих оценок следует инвестировать в понимание новых подходов к соревнованию и сдерживанию в век ИИ, особенно когда есть связь с другими областями, такими как кибербезопасность, квантовые вычисления, информационные операции и биотехнология, Зоны повышенного внимания включают в себя соревновательную и обостряющуюся динамику ИИ, избегание опасного распространения, воздействие на стратегическую стабильность, возможности сдерживания путем устрашения и возможности взаимовыгодных обязательств между странами.
6. **Инвестировать в исследования, направленные на улучшение воспроизводимости**. Отделу заместителя МО по НИОКР следует инвестировать в исследования, улучшающие воспроизводимость ИИ-систем. Трудности, которые ИИ-сообщество переживает в этой области, предоставляют возможность для МО внести свой вклад в понимание того, как работают сложные ИИ-модели.[31] Это также поможет разобраться с проблемой “черного ящика” применительно к ИИ.[32]
7. **Определить эталоны надежности.** Отделу заместителя МО по НИОКР следует изучить, как лучше всего создать соответствующие эталоны для измерения производительности ИИ-систем, включая относящиеся к человеческой производительности.
8. **Укрепить методики оценки и тестирования ИИ.** Под руководством Отдела оценки и доводочных испытаний МО следует использовать или улучшить существующие процедуры тестирования, оценки, проверки и подтверждения для

---

[29] См. ИИ-стратегия МО, стр. 14 (“Предоставление исчерпывающего обучения и подготовки по ИИ и взращивание высококвалифицированных кадров”).
[30] Там же.
[31] Многочисленные выдающиеся ученые в области ИИ, включая связанных с NeurIPS, самой знаменитой конференцией сообщества, недавно начали биться с техническими и финансовыми препятствиями, возникающими при решении проблемы воспроизводимости ИИ-систем.
[32] Проблема “черного ящика” означает неспособность людей понять, как ИИ-системы приходят к конкретным выводам, из-за многих скрытых и необъяснимых путей, которыми алгоритмы оценивают различные входные данные, что часто вдет к утрате доверия к ИИ-системам.

ИИ, и, где необходимо, создать новую инфраструктуру для ИИ-систем. Эти процедуры должны придерживаться программных рекомендаций для оценки и тестирования, подробно изложенных в Исследовании практик использования и приобретения программного обеспечения, проведенном советом СпОИ.[33] [34]

9. **Разработать методологию управления рисками.** Объединенному центру ИИ следует создать классификацию использования ИИ в МО на основе этических, защитных и юридических факторов риска.[35] Эта классификация должна поощрять и стимулировать быстрое принятие отработанных технологий в приложениях с низким уровнем риска и подчеркивать и уделять первостепенное внимание повышенным мерам предосторожности и тщательного рассмотрения в приложениях, которые менее отработаны и/или могут вести к более серьезным неблагоприятным последствиям.
10. **Обеспечить надлежащее внедрение этических принципов ИИ.** Объединенному центру ИИ следует оценить должное внедрение этих принципов и всех связанных с ними директив в рамках руководства и надзора, требуемого разделом 238 Закона о полномочиях в области национальной обороны 2019 г. и другими будущими предписаниями.
11. **Расширить исследование в понимании того, как внедрять этические принципы ИИ.** Отделу заместителя МО по НИОКР совместно с научно-исследовательскими отделами вооруженных сил следует сформировать проект в рамках Многопрофильной университетской исследовательской инициативы по вопросам безопасности, защищенности и надежности ИИ. Этот проект должен послужить стартовой точкой для непрерывных фундаментальных и академических исследований в этих областях.[36]
12. **Проводить ежегодную конференцию по безопасности, защищенности и надежности ИИ**. В свете быстро развивающейся природы области ИИ отделу заместителя МО по НИОКР следует проводить ежегодную конференцию, на которой рассматриваются этические вопросы, внедренные в безопасность, защищенность и надежность ИИ.

VI. Заключение

Данные принципы не направлены на то, чтобы разрешить спорные вопросы или

---

[33] См. Исследование практик использования и приобретения программного обеспечения.
[34] Для получения более подробных сведений и существующих возможностях оценки и тестирования ИИ в МО, а также рекомендаций по их улучшению см. Приложение IV в данном отчете.
[35] Управление перспективных исследовательских проектов МО поддержало исследование Национальной академии наук в 2014 г., по результатам которого был выпущен отчет *Появляющиеся и легкодоступные технологии и национальная безопасность: концепция для решения этических, правовых и социальных проблем*, в котором рекомендуется концепция снижения и оценки рисков для решения этических, правовых и социальных проблем, поставленных в ходе исследования появляющихся технологий в целях обеспечения национальной безопасности.
[36] См. ИИ-стратегию МО, стр. 15 (“Инвестирование в исследование и разработку отказоустойчивого, надежного и безопасного ИИ”). Национальный стратегический план НИОКР в области ИИ:Обновление 2019 г., особенно Стратегия 1 (“Осуществление долгосрочных инвестиций в исследование ИИ”) и Стратегия 4 (“Обеспечение защиты и безопасности ИИ-систем”).

ограничить возможности Министерства. Напротив, эти принципы нацелены на то, чтобы "дать зеленый свет" ИИ-системам и операциям, согласующимся с миссией МО сдерживать развязывание войны и обеспечивать безопасность нашей страны. Более того, эти принципы согласуются с существующими политическими концепциями, правом вооруженных конфликтов, национальными законами, такими как Раздел 10 Кодекса США, и неизменными этическими нормами, отражающими демократические ценности. СпОИ предлагает на рассмотрение МО эти этические принципы ИИ, включая рекомендации по их внедрению. В конце концов, этика – это не просто набор идей как таковых, а ряд целенаправленных действий и текущих вопросов.

За три года наших исследований в сферах технологии и обороны мы нашли Министерство обороны глубоко этической организацией, и не из-за опубликованных им документов, а из-за мужчин и женщин, которые взяли на себя постоянное обязательство жить и работать – а иногда сражаться и умирать – с глубоко укоренившимися убеждениями. Эти ценности должны быть предметом открытого обсуждения и критического мышления, чтобы оставаться уместными и верными. В то время как сфера ИИ развивается, верность Министерства законам США, праву вооруженных конфликтов и демократическим ценностям остается неизменной. Мы даем эти рекомендации с надеждой, что они внесут свой вклад в важное обсуждение, которое Министерство должно иметь по интерпретации существующих обязательств в контексте появляющихся технологий, таких как ИИ.