

Этические аспекты использования систем искусственного интеллекта

Б.С. Горячкин, Ю.П. Королёва

Московский государственный технический университет им. Н.Э. Баумана

Аннотация. В современном обществе все чаще возникают проблемы, связанные с этикой искусственного интеллекта (ИИ). ИИ используется повсеместно, и отсутствие этических стандартов и кодекса вызывает необходимость их создания для обеспечения безопасности и комфорта пользователей. Цель работы - провести анализ подходов к этике искусственного интеллекта и выделить параметры оценки подходов для создания систем, соответствующих этическим стандартам и удовлетворяющим потребности пользователей. Рассмотрены подходы к этике искусственного интеллекта. Выделены параметры для оценки подходов. Для каждого параметра выделены основные характеристики. Описанные в работе параметры помогут достичь более высоких результатов при создании стандартов для разработки более безопасных и комфортных для пользователя систем.

Ключевые слова: кодекс, параметры, показатели, характеристики, этика, искусственный интеллект.

Введение

Сегодня искусственный интеллект является одной из самых стремительно развивающихся областей в сфере информационных технологий. Для развития общества он имеет большое значение. Но из-за быстрого развития возникло и множество вопросов, связанных с моралью и этикой. Не существует определенных стандартов и общепризнанных кодексов, что приводит к необходимости создания таких систем, которые будут обеспечивать безопасность и комфорт пользователей.

Машинная этика — это область моральной науки, занимающаяся разработкой подходов к установлению механизма принятия этических решений интеллектуальными машинами и системами [1]. Существует множество примеров важности машинной этики. Например, если беспилотный автомобиль окажется в ситуации, где нужно будет выбрать, кому уступать дорогу в случае неизбежной аварии. Отсутствие этического кодекса может привести к негативным и даже опасным последствиям для людей [2]. Именно поэтому машинная этика очень важна для развития искусственного интеллекта и играет большую роль для развития технологий.

Таким образом, целью данной работы можно назвать анализ подходов к этике искусственного интеллекта и выделение характеристик оценки подходов для создания систем, которые будут соответствовать этическим стандартам и удовлетворять потребности пользователей [3].

Для достижения этой цели были поставлены следующие задачи:

- определить подходы к этике искусственного интеллекта;
- определить характеристики для оценки;
- выделить параметры для каждой характеристики.

Подходы к этике искусственного интеллекта

Рассмотрим шесть наиболее известных подходов к этике в системах управления. Первые три относятся к нисходящим подходам.

Первый подход, который мы рассмотрим, *декларативный*, или, другими словами, догматический подход. Данный подход программируется на основе определенных этических принципов и стандартов. Проблема может заключаться в том, что система будет принимать решения, основываясь лишь на тех принципах, которые были конкретно ей предоставлены. Как пример такого подхода, можно привести утилитаризм [4], три закона робототехники [5], различные религии [6]. Более конкретно можно рассмотреть кодекс искусственного интеллекта от Google [7], содержащий 7 основных принципов этики, которые чем-то напоминают законы робототехники Азимова.

Второй рассмотренный подход - *процессуальный*, который также является нисходящим и включает в себя набор правил, которые определяют выполнение этических стандартов. Отличие от предыдущего подхода заключается в том, что эти правила подвергаются дополнительной проверке на этичность и соблюдение стандартов, а также проводится оценка последствий возможных действий, которые могли бы быть предприняты [6].

Например, дополнительная система защищенности ИИ-продуктов, которая предоставляет возможность тестирования, исправления и проверки модели. Она работает со многими крупными клиентами [8].

Еще один подход – *нормативный* или правовой. Нормативная этика сосредоточена на изучении норм морали [9]. Данный подход основан на внедрении принятых законов и постановлений в отношении этики ИИ.

Примеров законов множество. В 2016 году был опубликован «Общий регламент по защите данных» (GDPR) [10], представляющий собой положение о защите данных и конфиденциальности в Европейском Союзе. Также в 2021 году Европейская комиссия выпустила «Закон об искусственном интеллекте» [11]. В этом законе установлены подходы регулирования использования систем ИИ в Европейском союзе.

Вторая группа подходов, которую мы рассмотрим, восходящие. Одним из них является *технологический* подход. Здесь поведение системы изменяется, в зависимости от результатов. Этические принципы в данном подходе изучаются на основе опыта и изменяются, после обнаружения ошибок, дискриминации, негативных результатов. Данный подход можно сравнить с жизненным опытом человека, который с начала и до конца жизни учится различать «плохо» и «хорошо». Использование данного подхода можно показать на примере алгоритма Casuist BDI-Agent, который рассуждает на основании разных ситуаций [12].

Восходящие подходы используют знания и опыт людей для формирования этического видения и оценки моральности своих действий. В связи с этим они требуют больше данных и наблюдений, что непросто выполнить на практике [6].

Одним из гибридных подходов является *культурный*. Переводчики переводят тексты в зависимости от сферы, языка и культуры входящих данных. Данный подход основан на исследовании интесекциональности и

учете различных культур [13]. Он учитывает различные этические правила, опираясь на них и подстраиваясь под определенную культуру.

Еще один гибридный подход – *интегративный*, он объединяет в себе нисходящие подходы, и подходы восходящие, тем самым учитывает больше этических правил и поведений. Данный подход основывается на том, что «важна не только природа, но и воспитание» [14]. Он использует уже установленные правила и обучается дополнительно на основе наблюдений. Например, использование NEAT, разработанного в 2002 году Кеннетом О. Стэнли и Ристо Мииккулайненом позволяет адаптировать, улучшать и расширять свои возможности в динамичной среде [15].

Для оценки каждого подхода необходимо выявить более значимые характеристики. Таковыми были выбраны следующие:

□ *Адаптивность*

Условия, ситуация, контекст могут стремительно меняться, и это приводит к необходимости учета адаптации под них. Системы должны учитывать расовые, гендерные и культурные различия, иначе это может привести к нарушению прав человека и норм морали. Необходимо оценивать, насколько подход может изменяться в зависимости от ситуации. [16]

□ *Безопасность*

В 2019 году Указом Президента (№ 490 от 10.10.2019 г.) были сформированы базовые принципы развития ИИ в России, и одним из принципов стала безопасность ИИ-продуктов. С развитием искусственного интеллекта возникает все больше проблем с конфиденциальностью данных. В связи с этим необходимо оценивать, насколько безопасна система и насколько защищен человек. Игнорирование данной характеристики может приводить к распространению личных данных и тому, что будут нарушены личные границы человека, что противоречит определению этических норм.

□ *Понятность*

В вышеупомянутом указе также утверждалось, что ИИ-продукты должны быть понятными с точки зрения принятия решений. Доверие пользователей очень важно при разработке любой системы. Представим ситуацию, когда при использовании компании автоматической системы найма сотрудников человеку приходит отказ, несмотря на его анкету с большими достижениями. Он может решить, что это произошло из-за расовых или гендерных соображений и из-за отсутствия понимания, как работает алгоритм.

□ *Эффективность*

Наличие дезинформации, отсутствие ответов на вопросы, связанных с сокращением рабочих мест при замещении сотрудников ИИ, возможность появления оружия — все это приводит к необходимости создания не только работающих, но и эффективно решающих данные проблемы, систем. Данная характеристика позволит оценить, как использование того или иного подхода может помочь улучшить систему и превратить ее в этически стабильную.

□ *Практичность*

В Бразилии государственными органами используются [17]:

- Модели естественного языка для определения стандартного потребительского поведения в области телекоммуникаций;
- ИИ для прогнозирования среднесуточного трафика на федеральных трассах в сфере транспорта;
- Чат-боты для обслуживания клиентов в банках;
- Искусственный интеллект для предсказания, какой овощ лучше всего выращивать в сельском хозяйстве;
- ИИ для автоматического сканирования апелляций, подаваемых в суд.

А также во множестве других сфер. Как видно области использования ИИ многообразны, в связи с этим требуется оценка, насколько практичны подходы для беспроблемной интеграции этики в любую сферу.

Сопровождаемость

Сопровождаемость ПО - процесс улучшения, оптимизации и устранения дефектов программного обеспечения (ПО) после передачи в эксплуатацию [18]. Как и ПО, подходы этики необходимо оценивать с точки зрения наличия дефектов, наличия обратной связи, соответствия этическому взгляду общества. Если система работает, не учитывая запросы пользователей, рано или поздно это приведет к последствиям, нарушающим этические нормы.

Надежность

На данный момент есть множества негативных примеров, когда система, имеющая ограниченные стандарты, нарушала этические рамки людей. Случаи расизма, гендерного неравенства, социальной несправедливости, грубых высказываний. Оценка надежности подходов к этике искусственного интеллекта может дать понимание, насколько часто происходят ошибки, что на это влияет и как можно уменьшить частоту некорректных результатов. Важно еще отметить, что игнорирование данной характеристики в случае сбоя приведет к тому, что все стандарты, опыт и наблюдения могут потеряться.

Легкость использования

При использовании определенного подхода необходимо понимать, насколько его легко воспринять как тому, кто будет использовать его при разработке, так и пользователю. Представим такую ситуацию, что система со временем начала выдавать неверные результаты, если алгоритмы и логика подхода непонятны, потребуется много времени для решения проблемы, что негативно скажется как на отношении пользователей к данной системе, так и на опыте разработчика.

□ *Универсальность*

Из-за различий культур, религий, государственных законов и политик организаций, этические стандарты могут отличаться. Единый этический стандарт труднодостижим. Поэтому универсальность даст понимание, насколько повсеместно подход может использоваться. Это может пригодиться в международных организациях.

□ *Корректность*

Для оценки моральной компетентности системы проводилось множество экспертных или неэкспертных тестов, где результаты работы системы сравнивались с истиной от экспертов и истиной, которые предоставляли простые граждане. Стандартом экспертного теста были стандарты экспертов по нормативной этике для оценки моральности агентов ИИ. Неэкспертные тесты брали за эталон народную мораль. При этом люди могли также участвовать в оценке этики ИИ на основании своих собственных принципов [6]. Нельзя исключать возможность того, что могут появиться неверные или предвзятые данные. Именно для этого необходимо оценивать такую характеристику, как корректность.

Существующие подходы можно связать с характеристиками посредством того, насколько данный подход нуждается в оценке той или иной характеристики (рис.1, таблица 1).

Таблица №1

Связь подходов и характеристик

Подходы	Характеристики
1	2
Декларативный	<ul style="list-style-type: none">• Адаптивность• Понятность• Сопровождаемость,• Универсальность

Таблица №1, продолжение

1	2
---	---

Процессуальный	<ul style="list-style-type: none">• Безопасность• Эффективность• Надежность• Корректность
Нормативный	<ul style="list-style-type: none">• Понятность• Эффективность• Корректность
Культурный	<ul style="list-style-type: none">• Практичность• Легкость и удобство использования• Универсальность
Технологический	<ul style="list-style-type: none">• Адаптивность• Эффективность• Надежность
Интегративный	<ul style="list-style-type: none">• Сопровождаемость• Надежность• Универсальность

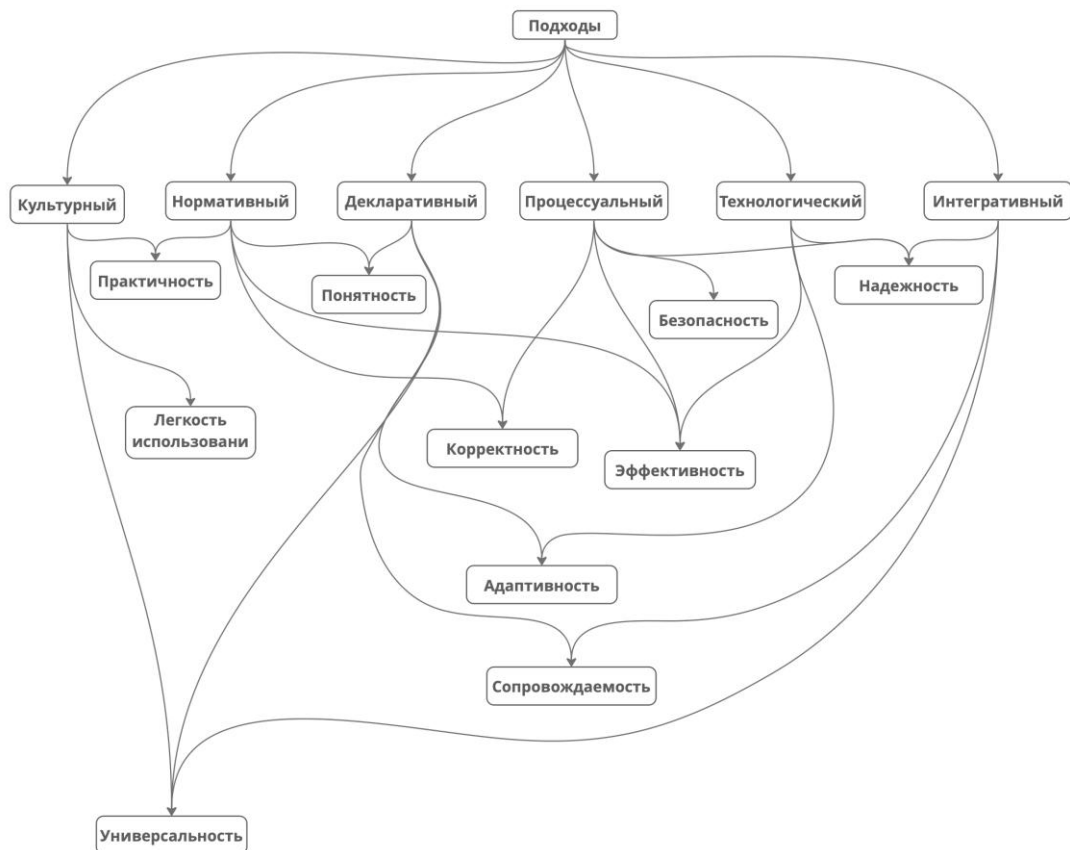


Рис. 1. – Связь подходов и характеристик

Основные параметры характеристик для оценки этики ИИ

Чтобы оценить выбранные характеристики, необходим набор параметров, по которым будет производиться оценка. Параметры в приведенных ниже таблицах были подобраны на основе определений характеристик, анализа различных ГОСТов и статей. Предлагаемые параметры позволят более объективно оценить подходы и дать независимую экспертную оценку (см. табл. 2-11).

Таблица №2

Параметры для характеристики «Адаптивность»

Параметр	Критерии	Оценка
Гибкость модели обозначает способность модели адаптироваться, развиваться и учиться на вводе данных [19]	Скорость адаптации к новым сценариям	Экспертная оценка 1/0
	Способность учиться	Количество успешных результатов при изменении данных/ Общее количество изменений данных
	Количество культурных групп, удовлетворенных результатами ГОСТ	Экспертная оценка 1/0
Возможность изменения алгоритма при анализе обратной связи пользователей	Наличие учета обратной связи	Экспертная оценка 1/0
	Эффективность изменений	Количество негативных результатов – Количество негативных результатов после изменения данных
Наличие динамической смены методов обучения [18]	Скорость обучения	
	Устойчивость функционирования	Количество положительных результатов при изменении контекста / Общее количество изменений контекста
Автономность системы - способность системы работать без вмешательства человека или прямого контроля [6]	Уровень автономии системы (ГОСТ 28195-89 Оценка качества программных средств)	Экспертная оценка 1/0
	Результат автономии	Количество негативных исходов, которые возникли из-за автономности системы

Таблица №3

Параметры для характеристики «Безопасность»

Параметр	Критерии	Оценка
Защита конфиденциальности данных пользователя	Количество способов защиты личных данных	Экспертная оценка 1/0
	Доступность конфиденциальных данных	Количество утечки данных
Предотвращение угроз	Обнаружение аномалий при внесении в существующий подход изменений	Успешно обнаруженные случаи/ Общее количество случаев
	Наличие мер по защите информации при возникновении угроз [21]	Экспертная оценка 1/0
Обеспечение возможности контроля рисков	Наличие процедур обнаружения рисков	Экспертная оценка 1/0
	Возможность обновления на основании выявленных рисков [21]	Экспертная оценка 1/0
Соблюдение законодательства	Соответствие законам и нормативам	Количество принятых политик, соответствующих законам / Общее число нормативных политик для конкретной области
	Включение законодательных политик	Количество нормативных политик в подходе
Безопасность при обучении	Защищенность алгоритма	Количество ситуаций, когда система подвергалась атакам
	Количество уязвимых мест в алгоритме	Количество уязвимых мест в алгоритме

Таблица №4

Параметры для характеристики «Понятность»

Параметр	Критерии	Оценка
Понятность решений [6]	Способность переобучения на негативных результатах	Экспертная оценка 1/0
	Количество случаев, когда пользователю непонятно решение	Обратная связь от пользователей
	Количество случаев, когда разработчику непонятно решение	Обратная связь от разработчика
Понятность алгоритмов	Уровень понимания логики	Экспертная оценка 1/0
	Понимание принципов (пользователи)	Процент людей, понимающих введенный этический стандарт
	Понимание принципов (разработчики)	Процент людей, понимающих введенный этический стандарт
Доступность информации	Открытый доступ к информации	Экспертная оценка 1/0
	Наличие сбоев системы (подразумевает случаи, когда доступность информации привела к негативным последствиям)	Наличие сбоев системы
Ясность стандартов	Наличие четких стандартов	Экспертная оценка 1/0
	Соответствие законодательству	Процентное соотношение пунктов, существующих в законодательных документах и не существующих

Таблица №5

Параметры для характеристики «Эффективность»

Параметр	Критерии	Оценка
Ресурсоемкость (ГОСТ 28195-89 Оценка качества программных средств)	Минимально необходимые ресурсы	Экспертная оценка 1/0
	Масштабируемость при обработке данных	Максимальный объем памяти
	Требуемый объем памяти	Требуемый объем памяти
Выполнение различных кейсов	Временная эффективность	Количество времени на выполнение непредвиденных сценариев
	Достоверность результатов на реальных кейсах	Процент верных решений
	Количество негативных результатов	Количество примеров
	Негативные отзывы	Количество отзывов

Таблица №6

Параметры для характеристики «Практичность»

Параметр	Критерии	Оценка
Обучаемость	Скорость обучения при вводе новых данных	Скорость обучения при вводе новых данных
	Возможность самообучения на основе новых данных	Экспертная оценка 1/0
Возможность включения в различные области и сферы	Количество областей, куда может быть интегрирован один подход	Экспертная оценка 1/0
	Специфичность подхода	Экспертная оценка – насколько интернациональным и межкультурным является подход

Таблица №7

Параметры для характеристики «Сопровождаемость»

Параметр	Критерии	Оценка
Структурность (ГОСТ 28195-89 Оценка качества программных средств)	Наличие службы тех. поддержки	Экспертная оценка 1/0
	Связь этических принципов	Экспертная оценка 1/0
	Наличие документов, необходимых для понимания всех технических решений	Экспертная оценка 1/0
Повторяемость	Степень использования типовых решений на специфические ситуации	Процент использования типовых решений
Обратная связь	Наличие обратной связи с пользователями и операторами.	Экспертная оценка 1/0
	Возможность самообучения на основе обратной связи	Экспертная оценка 1/0
	Участие в принятии решений	Процент учтенных замечаний, полученных на основании обратной связи

Таблица №8

Параметры для характеристики «Надежность»

Параметр	Критерии	Оценка
Устойчивость функционирования (ГОСТ 28195-89 Оценка качества программных средств)	Наличие тестов для проверки работоспособности	Экспертная оценка 1/0
	Возможность обработки ошибок [23]	Экспертная оценка 1/0
Работоспособность [23]	Обработка заданного объема информации	Скорость обработки информации
	Время генерации результата	Время генерации результата
	Стабильность	Количество сбоев
	Восстанавливаемость [24]	Время восстановления работоспособности после сбоя

Таблица №9

Параметры для характеристики «Легкость использования»

Параметр	Критерии	Оценка
Легкость освоения	Полнота документации	Экспертная оценка 1/0
	Доступность программных документов	Полнота описания взаимодействия между системой и пользователями (ГОСТ 28195-89 Оценка качества программных средств)
Доступность	Количество социальных групп, которым доступна система	Экспертная оценка 1/0
	Доступ для групп с ограниченными возможностями	Экспертная оценка 1/0
	Наличие перевода на различные языки	Количество языков
Удобство эксплуатации и обслуживания	Наличие кодекса этики	Экспертная оценка 1/0
	Точность	Процент соответствия обработки данных и предоставления результатов характеру решаемых задач (ГОСТ 28195-89 Оценка качества программных средств)
	Легкость обслуживания	Уровень необходимых квалификаций обслуживающего персонала (ГОСТ 28195-89 Оценка качества программных средств)
	Управление данными	Экспертная оценка 1/0

Таблица №10

Параметры для характеристики «Универсальность»

Параметр	Критерии	Оценка
Гибкость – возможность использования в различных областях применения (ГОСТ 28195-89 Оценка качества программных средств)	Количество областей применения	<i>Количество областей применения</i>
	Многофункциональность	<i>Наличие возможности интегрировать подход в любую сферу</i>
Модифицируемость – простота внесения изменений	Степень сложности внесения изменений	<i>Количество времени, необходимого для переобучения после внесения изменений</i>
	Возможность внесения изменений	<i>Экспертная оценка 1/0</i>
Мобильность – возможность применения без дополнительных трудозатрат	Широта охвата функций	<i>Количество поддерживаемых функций на различных устройствах / Общее количество функций</i>
	Наличие дополнительных трудозатрат после интеграции подхода в систему	<i>Экспертная оценка</i>

Таблица №11

Параметры для характеристики «Корректность»

Параметр	Критерии	Оценка
Логическая корректность и согласованность (ГОСТ 28195-89 Оценка качества программных средств)	Отсутствие противоречий в пунктах стандарта	Процент случаев, выявляющих противоречия
	Тестируемость	Количество успешно проведенных тестов/ Общее количество тестов
	Выполнение всех требований для системы	Экспертная оценка
	Соответствие стандартам	Процент соответствия этических стандартам
	Защищенность	Процент предотвращения несанкционированного доступа
Проверенность	Полнота проверки данных	Экспертная оценка 1/0
	Функции для проверки стандартов	Наличие проверки корректности кодекса

Более детальная оценка подходов должна проводиться отдельно для каждой области применения, так как значимость характеристик будет определяться областью использования. Например, в системах кредитования большую роль будут играть такие характеристики, как прозрачность (для понимания пользователем, как работает система), безопасность для защиты данных, а также адаптивность, так как наличие тех или иных факторов может существенно повлиять на исход одобрения и отказа при получении кредита.

Заключение

В заключение хотелось бы отметить, что развитие этики искусственного интеллекта имеет огромное значение для повышения эффективности систем управления. Оценка подходов даст возможность развития системам, которые будут удовлетворять нормам этики и морали в обществе, что предоставит пользователям больший комфорт и безопасность. Также понятно, что благодаря данной оценке появится возможность более детально изучить каждый подход и сферу этики и, возможно, появятся решения для некоторых этических вопросов и идеи для улучшения подходов.

Но, для определения весовых коэффициентов и проведения оценки подходов, необходимо рассматривать не только сам подход, но и область его применения.

Литература

1. Наумов В.Б., Тютюк Е.В. Правовые проблемы машинного обучения // Образование и право. 2020. №6. С. 219-231.
2. Бажанов Н.Н. Искусственный интеллект: опасность завышенных ожиданий // Инженерный вестник Дона, 2023, №9. URL: ivdon.ru/ru/magazine/archive/n9y2023/8677.
3. Тимохин М.Ю., Шаранин В.Ю. Искусственный интеллект и теория принятия решений: современные тенденции // Инженерный вестник Дона, 2023, №10. URL: ivdon.ru/ru/magazine/archive/n10y2023/8746.
4. Резаев А.В., Трегубова Н.Д. "Эмоциональный утилитаризм" и пределы развития искусственного интеллекта // Мониторинг общественного мнения: экономические и социальные перемены. 2022. №2(168). С. 4-23.
5. Баранов П.П. Правовое регулирование робототехники и искусственного интеллекта в России: некоторые подходы к решению проблемы // Северо-Кавказский юридический вестник. 2018. №1. С. 39-45.



6. Huang C., Zhang Z., Mao B., Yao X. An overview of artificial intelligence ethics // IEEE Transactions on Artificial Intelligence. 2022. V. 4. №4. pp. 799-819.
 7. Google опубликовал 7 принципов этики ИИ // Хабр. URL: habr.com/ru/articles/413637/ (дата обращения: 18.03.2024).
 8. Система проверки защищенности ИИ продуктов // URL: anti-malware.ru/news/2023-10-03-114534/42036 (дата обращения: 18.03.2024).
 9. Кашкин С.Ю. Искусственный интеллект и робототехника: возможность вторжения в права человека и правовое регулирование этих процессов в ЕС и мире // Lex Russica. 2019. №7(152). С. 151-159.
 10. Regulation (EU) 2016/679 of the European Parliament and of the Council // URL: eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434 (дата обращения: 18.03.2024).
 11. Regulation of the European Parliament and of the Council // URL: eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206 (дата обращения: 18.03.2024).
 12. Honarvar A.R., Ghasem-Aghaee N. Casuist BDI-agent: a new extended BDI architecture with the capability of ethical reasoning // Artificial Intelligence and Computational Intelligence: International Conference, AICI 2009, Shanghai, China, November 7-8, 2009. Proceedings 1. Springer Berlin Heidelberg, 2009. pp. 86-95.
 13. Roche C., Wall P.J., Lewis D. Ethics and diversity in artificial intelligence policies, strategies and initiatives // AI and Ethics. 2023. V. 3. №4. pp. 1095-1115.
 14. Gigerenzer G. Moral satisficing: Rethinking moral behavior as bounded rationality // Topics in cognitive science. 2010. V. 2. №3. pp. 528-554.
-

15. Lang S., Reggelin T., Schmidt J., Müller M., Nahhas A. NeuroEvolution of augmenting topologies for solving a two-stage hybrid flow shop scheduling problem: A comparison of different solution strategies // Expert Systems with Applications. 2021. V. 172. pp. 114666.
16. Инновационные технологии устойчивого развития, адаптивный искусственный интеллект // URL: cdo2day.ru/analytics/innovacionnye-tehnologii-ustojchivogo-razvitija-adaptivnyj-iskusstvennyj-intellekt-prognoz-ot-gartner/?ysclid=lttiyvu54i787164616 (дата обращения: 18.03.2024).
17. Corrêa N.K., De Oliveira N., Massmann D. On the Efficiency of Ethics as a Governing Tool for Artificial Intelligence. URL: <https://arxiv.org/abs/2210.15289>.
18. Взгляд тестировщика на сопровождаемость ПО // URL: habr.com/ru/articles/431732/ (дата обращения: 18.03.2024).
19. Колесников Д.О., Константинова Л.И. Динамическая смена методов обучения искусственных нейронных сетей // Радиоэлектроника и информатика. 2004. №4(29). pp. 50-51.
20. Бегаев А.Н., Кашин С.В., Маркевич Н.А., Марченко А.А., Павлов Д.Д., Сертификация программного обеспечения по требованиям доверия. СПб: Университет ИТМО, 2020. 40 с.
21. Надежность программного обеспечения // URL: areliability.com/nadyozhnost-programmnogo-obespecheniya/ (дата обращения: 18.03.2024).
22. Горбаченко И.М. Оценка качества программного обеспечения для создания систем тестирования // Фундаментальные исследования. 2013. №6-4. С. 823-827.

References

1. Naumov V.B., Tytyuk E.V. Obrazovanie i pravo. 2020. №6. pp. 219-231.
-



2. Bazhanov N.N. Inzhenernyj vestnik Dona, 2023, №9 URL: ivdon.ru/ru/magazine/archive/n9y2023/8677.
 3. Timokhin M.Yu., Sharanin V.Yu. Inzhenernyj vestnik Dona, 2023, №10 URL: ivdon.ru/ru/magazine/archive/n10y2023/8746.
 4. Rezaev A.V., Tregubova N.D. Monitoring obshchestvennogo mneniya: ekonomicheskie i sotsial'nye peremeny. 2022. №2(168). pp. 4-23.
 5. Baranov P.P. Severo-Kavkazskiy yuridicheskiy vestnik. 2018. №1. pp. 39-45.
 6. Huang C., Zhang Z., Mao B., Yao X. IEEE Transactions on Artificial Intelligence. 2022. V. 4. №4. pp. 799-819.
 7. Google opublikoval 7 printsipov etiki II [Google has published 7 principles of AI ethics]. Habr. URL: habr.com/ru/articles/413637/ (accessed: 18.03.2024).
 8. Sistema proverki zashchishchennosti II produktov [Security and product verification system] URL: anti-malware.ru/news/2023-10-03-114534/42036 (accessed: 18.03.2024).
 9. Kashkin S.Yu. Lex Russica. 2019. №7(152). pp. 151-159.
 10. Regulation (EU) 2016/679 of the European Parliament and pf the Council URL: eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434 (accessed: 18.03.2024).
 11. Regulation of the European Parliament and pf the Council URL: eur-lex.europa.eu/legal-content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206 (accessed: 18.03.2024).
 12. Honarvar A.R., Ghasem-Aghae N. Artificial Intelligence and Computational Intelligence: International Conference, AICI 2009, Shanghai, China, November 7-8, 2009. Proceedings 1. Springer Berlin Heidelberg, 2009. pp. 86-95.
-

13. Roche C., Wall P.J., Lewis D. AI and Ethics. 2023. V. 3. №4. pp. 1095-1115.
14. Gigerenzer G. Topics in cognitive science. 2010. V. 2. №3. pp. 528-554.
15. Lang S., Reggelin T., Schmidt J., Müller M., Nahhas A. Expert Systems with Applications. 2021. V. 172. pp. 114666.
16. Innovatsionnye tekhnologii ustoychivogo razvitiya, adaptivnyy iskusstvennyy intellekt [Innovative technologies of sustainable development, adaptive artificial intelligence] URL: cdo2day.ru/analytics/innovatsionnye-tekhnologii-ustoychivogo-razvitiya-adaptivnyj-iskusstvennyj-intellekt-prognoz-ot-gartner/?ysclid=lttiyyvu54i787164616 (accessed: 18.03.2024).
17. Corrêa N.K., De Oliveira N., Massmann D. On the Efficiency of Ethics as a Governing Tool for Artificial Intelligence URL: <https://arxiv.org/abs/2210.15289>.
18. Vzglyad testirovshchika na soprovozhdaemost' PO [The tester's view on software maintainability] URL: habr.com/ru/articles/431732/ (accessed: 18.03.2024).
19. Machine Learning: Flexible and Inflexible Models. URL: baeldung.com/cs/ml-flexible-and-inflexible-models (accessed: 18.03.2024).
20. Begaev A.N., Kashin S.V., Markevich N.A., Marchenko A.A., Pavlov D.D., Sertifikatsiya programmnoy obespecheniya po trebovaniyam doveriya [Certification of software according to the requirements of trust]. SPb: Universitet ITMO, 2020. 40 p.
21. Nadezhnost' programmnoy obespecheniya [Software reliability]. URL: areliability.com/nadyozhnost-programmnogo-obespecheniya/. (accessed: 18.03.2024).
22. Gorbachenko I.M. Fundamental'nye issledovaniya. 2013. №6-4. pp. 823-827

Дата поступления: 20.02.2024

Дата публикации: 4.04.2024
