Методические материалы по практическим работам

С.С. Жумажанова

Этика и правовые проблемы искусственного интеллекта

**ПРАКТИЧЕСКИЕ (СЕМИНАРСКИЕ) ЗАНЯТИЯ**

1. На практических (семинарских) занятиях преподавателем проводится устный опрос студентов по темам пройденных лекций:

* Национальный Кодекс этики в сфере искусственного интеллекта.
* Принципы ИИ.
* Вопросы социальной ответственности, связанные с участием и развитием общества при внедрении систем ИИ.
* Создание и использование этичного и социально приемлемого ИИ в современном мире.
* Технологии ИИ в задачах защиты информации.
* Правовые аспекты участия человека в естественно-научных (биомедицинских) экспериментах для обучения ИИ.
* Методы оценки рисков, возникающих из-за внедрения ИИ, а также передовые методы обнаружения и снижения воздействия ИИ.
* Вопросы доверия к ИИ.

Активность студентов при обсуждении, дискуссиях и решении задач по темам лекций учитывается преподавателем, как один из способов текущего контроля на практических занятиях.

1. Студенты также проходят промежуточные тестирования, по результатам которого выставляется оценка по пятибалльной системе. Студенты, получившие за оба теста оценку не менее «удовлетворительно», допускаются к экзамену. Тест состоит из 10 тестовых заданий (для получения оценки «удовлетворительно» необходимо дать правильные ответы на 60% или более тестовых вопросов, «хорошо» - на 75% или более, «отлично» - на 90% или более).
2. На практических (семинарских) занятиях преподавателем проводится контроль выполнения студентами индивидуального домашнего задания (ИДЗ). Студенты задают вопросы преподавателю и получают рекомендации либо указания по выполнению ИДЗ.

**ИНДИВИДУАЛЬНОЕ ДОМАШНЕЕ ЗАДАНИЕ**

***Темы ИДЗ:***

1. Создание репрезентативной базы общедоступных (статистических) сведений о субъектах для обучения систем ИИ с учетом законодательства РФ о защите ПДн.

2. Сбор репрезентативной базы «динамических» биометрических данных пользователей для обучения систем ИИ с учетом этических принципов и законодательства РФ о защите ПДн.

3. Сбор репрезентативной базы «статических» биометрических данных пользователей для обучения систем ИИ с учетом этических принципов и законодательства РФ о защите ПДн.

***Требования к выполнению ИДЗ:***

* Сбор наборов данных людей (пользователей, субъектов) проводится с соблюдением этических принципов (компетентность, порядочность, профессиональная ответственность, исключение дискриминации для различных социальных групп, уважение к правам человека и др.). Эти данные на выбор студента или группы студентов, которая будет выполнять задание, могут включать как общедоступные сведения, содержащие, например, статистические данные, обезличенные данные, так и самостоятельно собранные данные. Полученная выборка должна быть репрезентативной, и пригодной для ее использования в целях обучения систем (методов и алгоритмов) ИИ.
* При сборе данных необходимо разработать протокол проведения опытов, отражающий ход экспериментов. Форма протокола должна учитывать особенности эксперимента и цели сбора данных, а также обеспечивать простоту и удобство проведения опытов и обработки результатов.
* В случае сбора данных с участием испытуемых необходимо разработать форму согласия на добровольное участие в эксперименте и дальнейшую обработку данных участников. Участие испытуемого в опытах возможно только после получения от него письменного согласия на добровольное участие и дальнейшую обработку его персональных данных.
* После выполнения ИДЗ студентом готовится доклад, отражающий ход и результат выполнения работы, а также презентация с основными результатами работы. К докладу прилагается собранная студентом база данных, которая передается преподавателю на машинном носителе (или иным доступным студенту и преподавателю способом передачи).

***Доклад***

По результатам индивидуальных домашних заданий необходимо подготовить доклад (краткий отчет и выступление с презентацией).

**Требования к отчету.** Объем отчета в электронном виде (форматы – .doc, docx, pdf) должен быть от 15 до 30 стр., оформленный в соответствии с ГОСТ 7.32-2017.

Доклад должен состоять из следующих пунктов:

* титульный лист;
* содержание;
* введение;
* основная часть;
* заключение;
* список используемой литературы;
* приложения.

В отчете необходимо отразить:

* в зависимости от способа сбора данных, разработанный протокол проведения экспериментов, в котором описана процедура сбора данных, возможные последствия при сборе данных для исследователей и испытуемых, возможное воздействие (степень возможного воздействия) раздражителя (если такой необходим в ходе экспериментов) на испытуемого (при необходимости описать возможные варианты воздействия раздражителя в зависимости от личных физиологических данных человека);
* разработанную студентом форму согласия испытуемого на участие в экспериментах по сбору данных. Согласие должно соответствовать современных требованиям по защите персональных данных (ФЗ 152), Декларации о защите прав человека, Конституции РФ и иных нормативно-правовых актов РФ в сфере защиты прав человека;
* обоснование выбранных типов данных, возможные сферы их применения, в частности при развитии систем ИИ кроме предложенных преподавателем данной дисциплины, их качественный и количественный (объем выборки) состав, описание структур данных.

**Требования к презентации.** По основным результатам выполненной работы готовится презентация. Общий объем презентации – не менее 10 слайдов, форматы презентации – pptx, ppt, pdf.

Презентация должна содержать следующие обязательные разделы:

* титульная страница (первый слайд);
* введение;
* основная часть презентации;
* заключение.

Выступление по результатам работы должно занимать от 10 до 15 минут (+ 5-10 минут на вопросы/ответы).

4. Описание практических занятий.

**Практика № 1.**

*1) Опрос студентов по следующим вопросам:*

* Термины и определения в области ИИ.
* Сферы применения ИИ (медицина, финансы, промышленность и др.).

*2) Тестовые задания:*

1. *Кто является одним из наиболее известных основателей теории искусственного интеллекта, опубликовавший в 1950 году статью «Вычислительные машины и разум»?*

Джон Серл

Чарлз Бэббидж

Алан Тьюринг

*2. Чтобы наделить машину способностью воспринимать видео- и звуковую информацию достаточно:*

Установить камеру

Установить микрофон

Установить камеру, микрофон и применить алгоритм интерпретации

*3. Система, представляющая собой ИИ, должна:*

Вести себя также как человек

Вести себя не как человек

Пока неизвестно

*4. Первый разработанный чат-бот, имитирующий чат с женщиной?*

A.L.I.C.E.

Mobilemonkey

Manychat

*5. Является ли ИИ все еще простым инструментом*

Да

Нет

Пока неизвестно

*6. Может ли ИИ «обладать» интуицией, при этом не «обладать» здравым смыслом*

Да

Нет

Пока неизвестно

*7. Сложно ли ИИ принимать «приемлемые» решения этического характера?*

Да

Нет

Пока неизвестно

*8. Так ли важны реальные данные для ИИ, т.е. У кого больше данных, тот «в плюсе»?*

Да

Нет

Пока неизвестно

*9. Верно ли утверждение, что ИИ является больше дополненным интеллектом, т.е. дополняющим человека?*

Да

Нет

Пока неизвестно

*10. В каких существующих ныне играх ИИ еще не превзошел человека?* Покер

Шахматы

Шашки

*3) Формирование подгрупп студентов для выполнения каждой подгруппой одной из трех работ ИДЗ по темам на выбор (одна тема ИДЗ может выполняться одной и более подгруппами):*

* Создание репрезентативной базы общедоступных (статистических) сведений о субъектах для обучения систем ИИ с учетом законодательства РФ о защите ПДн.
* Сбор репрезентативной базы «динамических» биометрических данных пользователей для обучения систем ИИ с учетом этических принципов и законодательства РФ о защите ПДн.
* Сбор репрезентативной базы «статических» биометрических данных пользователей для обучения систем ИИ с учетом этических принципов и законодательства РФ о защите ПДн.

Ответы на вопросы студентов по выбранной теме и выполнению ИДЗ.

**Практика № 2.**

*1) Опрос студентов по следующим вопросам:*

* Необходимость создания Кодекса этики в сфере ИИ.
* Содержание Кодекса.
* Акторы в сфере ИИ.
* Общие этические принципы/ценности поведения Акторов ИИ.
* Этические стандарты поведения Акторов в сфере ИИ.
* Механизмы реализации Кодекса этики в сфере ИИ.
* Документы нормативно-технического регулирования в сфере ИИ.

*2) Тестовые задания:*

*1.* *На основе чего создан национальный кодекс этики в сфере искусственного интеллекта?*

Федеральных законов, регулирующих отношения в сфере использования ИИ

Национальной стратегии развития ИИ на период до 2030 года

Гражданского кодекса РФ

*2. Что устанавливает кодекс этики в сфере искусственного интеллекта?*

Общие этические принципы и стандарты поведения, которыми следует руководствоваться Акторам ИИ

Основание и принципы уголовной ответственности в сфере искусственного интеллекта

Устанавливает нормы административного права сфере искусственного интеллекта

*3.* *Допустима ли эксплуатация, а также проектирование, разработка, тестирование и СИИ, способных целенаправленно причинять вред?*

Да

Допускается, если СИИ имеет такую возможность

Нет

*4. Актор ИИ – это …*

Программное обеспечение для создания СИИ

Лицо, принимающее участие в жизненном цикле СИИ

Сфера жизнедеятельности человека, в которую внедряется СИИ

*5. Являются ли Акторами ИИ пользователи СИИ?*

Да

Нет

Неизвестно

*6. Как должна оцениваться ответственность Акторов ИИ?*

Ответственность должна соответствовать ущербу, который может быть причинён в процессе использования

Ответственность устанавливается в процессе проектирования СИИ

Акторы ИИ не несут ответственность за ущерб, который может быть причинён в процессе использования

*7. Каким документом руководствуются Акторы ИИ при оценке категории моральной неприемлемости последствий?*

Уголовный кодекс РФ

Кодекс этики в сфере искусственного интеллекта

Кодекс РФ об административных правонарушениях

*8. Обязаны ли Акторы ИИ соблюдать законодательство РФ в области персональных данных?*

Нет

Зависит от ситуации

Да

*9. На каких этапах Акторы ИИ должны знать и соблюдать положения законодательства Российской Федерации?*

Создания и внедрения технологий ИИ

Использования технологий ИИ

На всех этапах

*10. Что не входит в стандарты поведения Акторов ИИ согласно кодексу этики в сфере искусственного интеллекта?*

Действовать ответственно и добросовестно

Применять технологии ИИ только в случае невозможности использования физических возможностей человека

Содействовать доверию общества к технологиям ИИ в интересах устойчивого развития

*3) Контроль выполнения ИДЗ, обсуждение со студентами хода выполнения выбранных тем ИДЗ, ответы на вопросы студентов по выполнению ИДЗ.*

**Практика № 3.**

*1) Опрос студентов по следующим вопросам:*

* Ответственность частных лиц и организаций при негативных последствиях, возникающих в результате проектирования, разработки и использования ими систем на базе ИИ.
* Справедливость и исключение дискриминации при использовании технологий ИИ для различных социальных групп.
* Прозрачность и объяснимость работы ИИ, направленная на то, чтобы пользователи понимали, как ИИ принимает решения, как система ИИ была разработана и протестирована, чтобы гарантировать, что она работает так, как изначально задумано.
* Профессиональная ответственность проектировщиков и разработчиков систем ИИ.
* Использование ИИ с целью пропаганды человеческих ценностей и принесения максимальной пользы обществу.
* Использование систем ИИ с учетом обеспечения защиты сведений ограниченного доступа.
* Соблюдение принятых законов и нормативно-правовых актов, международных норм поведения.

*2) Тестовые задания:*

*1. Что такое «КОМЕСТ»?*

Комиссия стандартизации принципов разработки СИИ

Кодекс, устанавливающий правила использования СИИ

Всемирная комиссия по этике научных знаний и технологий

*2. Что не умеет ИИ?*

Обучаться

Создавать себе подобных

Прогнозировать

*3. Что является целью «Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники до 2024 года»?*

Массовое внедрение СИИ в различные сферы жизнедеятельности человека

Создание СИИ способного решить любую задачу, поставленную человеком

Определение основных подходов к трансформации системы нормативного регулирования в РФ для обеспечения возможности создания и применения таких технологий в различных сферах экономики

*4. Технологии искусственного интеллекта – это …*

Программные средства, предназначенные для создания ИИ

Технологии, основанные на использовании ИИ

Аппаратные средства, предназначенные для создания ИИ

*5. Что не является принципом развития и использования технологий ИИ?*

Прозрачность

Защита прав и свобод человек

Недопущение конкуренции

*6. Возможно ли страхование СИИ?*

Да

Нет

Возможно только в некоторых странах

*7. Для чего внедряется ИИ в сферу транспорта?*

Для создания и эксплуатации автоматизированного транспорта

Для регистрации пассажиров общественного транспорта без участия человека

Для фиксации фактов нарушения ПДД пешеходами

*8. Использование технологий ИИ в государственном (муниципальном) управлении …*

Недопустимо

Позволяет сократить количество рутинных операций и повысить скорость принятия решений

Тормозит выполнение рутинных операций и понижает скорость принятия решений

*9. Что такое кодекс этики ИИ?*

Свод этических норм, образующих систему нравственных правил, которая основана на их добровольном принятии и выполнении

Такого не существует на данный момент

Перечень правил по написанию ИИ

*10. Важнейшее свойство ИИ?*

Автономность

Самообучение

Простота использования

*3) Контроль выполнения ИДЗ, обсуждение со студентами хода выполнения выбранных тем ИДЗ, ответы на вопросы студентов по выполнению ИДЗ.*

**Практика № 4.**

*1) Опрос студентов по следующим вопросам:*

* Выявление групп заинтересованных сторон, на которые влияют системы ИИ, и решение их проблем.
* Минимизация воздействий на окружающую среду при использовании технологий ИИ.
* Контроль над технологиями ИИ со стороны человека и общества.

*2) Тестовые задания:*

*1. В чём заключается задача согласования?*

Создание ИИ, которые остаются безопасными, даже когда они действуют автономно в больших масштабах

Создание ИИ, которые смогут самостоятельно исключать возможность нанесения вреда человеком ИИ

Согласование аппаратных средств с возможностями СИИ

*2. Проблема контроля искусственного интеллекта – это …*

Задача, которая состоит в том, чтобы создать инструмент контроля над искусственным сверхразумом, если он выйдет из-под контроля

Задача, которая состоит в том, чтобы создать искусственный сверхразум, который будет полезен людям, и при этом избежать непреднамеренного создания сверхразума, который нанесёт вред

Невозможность использования технологий ИИ пользователем

*3. Почему целесообразно применять технологии ИИ в здравоохранении?*

Обладая доступом к научной литературе и миллионам историй болезней, ИИ может быстро классифицировать случай, соотнести его с похожими и сформулировать предложения по плану лечения

Низкая квалификация врачей во всём мире

Упразднение учебных заведений, обучающих будущий медицинский персонал

*4. Одной из целей применения ИИ в сфере охраны окружающей среды является …*

Ограничение деятельности человека

Предсказание развития событий в крайне сложных системах для своевременного выявления возможных рисков наступления негативных последствий

Изменение окружающей среды так под нужды человека

*5. Для улучшения экологической обстановки в мире ИИ …*

Может принудительно ограничить использование людьми ресурсов планеты

Может остановить все производства, существенно влияющие на экологию

Может более точно рассчитать количество используемых энергоресурсов

*6. Искусственный интеллект может поможет в открытии новых материалов?*

Нет

Да

Неизвестно

*7. Что не относится к методам охраны окружающей среды искусственным интеллектом?*

Ограничение

Прогнозирование

Моделирование

*8. «Умный город» — это …*

Программное обеспечение для проектирования СИИ

Концепция интеграции нескольких информационных и коммуникационных технологий для управления городским имуществом

Организация, осуществляющее внедрение СИИ во все сферы жизнедеятельности человека

*9. Может ли ИИ использоваться для повышения продуктивности труда в сельском хозяйстве?*

Да

Нет

Может только в странах с развитым сельским хозяйством

*10. Какой подход, отраженный в Кодексе этики ИИ рассматривает права и свободы человека как наивысшую ценность?*

Недискриминация

Уважение автономии и свободы воли человека

Человеко-ориентированный и гуманистический подход

*3) Контроль выполнения ИДЗ, обсуждение со студентами хода выполнения выбранных тем ИДЗ, ответы на вопросы студентов по выполнению ИДЗ.*

**Практика № 5.**

*1) Опрос студентов по следующим вопросам:*

* Опыт применения ИИ в социальном обеспечении.
* Рынок ИИ в России и мире.
* Использование ИИ в госуправлении, производстве, медицине, образовании и культуре.
* Влияние использования ИИ на занятость населения за счет повышения автоматизации бизнес-процессов, создания рабочих мест.

*2) Тестовые задания:*

*1. Подразделение корпорации IBM под названием «IBM Watson Health» создано для …*

Внедрения технологий ИИ в образование

Внедрения технологий ИИ в здравоохранение

Внедрения технологий ИИ в сферу государственного управления

*2. Какое утверждение не относится к умным чат-ботам?*

Круглосуточный и ежедневный доступ к различным видам услуг и помощи

Самостоятельные ответы на вопросы пользователей

Переадресация вопросов пользователя сотрудникам организации

*3. «Адаптивное обучение» — это …*

Процесс модернизации учебного процесса с учётом предпочтений обучающегося

Метод обучения ИИ

Образовательный метод, который использует компьютерные алгоритмы, чтобы организовать взаимодействие с обучаемым

*4. Что относится к недостаткам адаптивного обучения?*

Полное включение обучающегося в процесс обучения

Отсутствие реакции на психологическое состояние обучающегося

Адаптация программы обучения в соответствии с индивидуальными способностями обучающегося

*5. Что относится к преимуществам адаптивного обучения?*

Адаптация программы обучения в соответствии с индивидуальными способностями обучающегося

Контроль третьего лица за правильностью принятия решений ИИ

Отсутствие реакции на психологическое состояние обучающегося

*6. Может ли ИИ самостоятельно принимать решение о наличии или отсутствии вины подсудимого в судебной деятельности?*

Да

Нет

Зависит от законодательства страны

*7. Что является целью использования СИИ в следственной деятельности?*

Расследование преступления без участия человека

Определение меры наказания преступника

Анализ больших объемов информации с целью выявления сведений, имеющих значение для расследования преступлений

*8. Проект национальной программы, направленный на полный переход к электронному взаимодействию граждан и государства, называется …*

Информатизация экономики

Цифровая экономика

Автоматизация экономики

*9. Какая из СИИ относится к сфере образования?*

Querium

Scanderm

Инсилико

*10. «Технет» — это …*

Организация, регулирующая деятельность ИИ в России

План развития передовых производственных технологий в России

Система ИИ в сфере гос. Управления

*3) Контроль выполнения ИДЗ, обсуждение со студентами хода выполнения выбранных тем ИДЗ, ответы на вопросы студентов по выполнению ИДЗ.*

**Практика № 6.**

*1) Опрос студентов по следующим вопросам:*

* Этические аспекты использования ИИ в задачах обеспечения информационной безопасности.
* Проблема защиты персональных данных человека при обработке в системах ИИ.
* Проблема отчуждения аутентификаторов (ключей и паролей) от личности человека и ее решение с помощью биометрических систем.
* Проблема защиты биометрических персональных данных человека.
* Обезличивание персональных данных.
* Технологии компьютерного зрения и распознавания образов на базе ИИ.

*2) Тестовые задания:*

*1. Может ли ИИ обойти проверку CAPTCHA?*

Да

Нет

Такой проверки не существует

*2. «EDR» — это …*

Устройства и аналитические платформы, которые обнаруживают атаки на сетевом уровне и позволяют оперативно на них реагировать

Платформы обнаружения атак на рабочих станциях, серверах, любых компьютерных устройствах и оперативного реагирования на них

Платформы раннего детектирования угроз и реагирования на них, действующие на основе большого количества различных данных (Data Lake) и индикаторов компрометации (ioc)

*3. «NDR» — это …*

Системы поведенческого анализа пользователей и информационных сущностей, которые обнаруживают случаи необычного поведения и используют их для детектирования внутренних и внешних угроз

Платформы раннего детектирования угроз и реагирования на них, действующие на основе большого количества различных данных (Data Lake) и индикаторов компрометации (ioc)

Устройства и аналитические платформы, которые обнаруживают атаки на сетевом уровне и позволяют оперативно на них реагировать

*4. «UEBA» — это …*

Системы поведенческого анализа пользователей и информационных сущностей, которые обнаруживают случаи необычного поведения и используют их для детектирования внутренних и внешних угроз

Платформы обнаружения атак на рабочих станциях, серверах, любых компьютерных устройствах и оперативного реагирования на них

Устройства и аналитические платформы, которые обнаруживают атаки на сетевом уровне и позволяют оперативно на них реагировать

*5. «TIP» — это …*

Платформы обнаружения атак на рабочих станциях, серверах, любых компьютерных устройствах и оперативного реагирования на них

Платформы раннего детектирования угроз и реагирования на них, действующие на основе большого количества различных данных (Data Lake) и индикаторов компрометации (ioc)

Системы поведенческого анализа пользователей и информационных сущностей, которые обнаруживают случаи необычного поведения и используют их для детектирования внутренних и внешних угроз

*6. В чём заключается отличие «SOAR» от «SIEM»? ИИ помогает не только проводить анализ, но и автоматически реагировать надлежащим образом на выявленные угрозы*

Нет отличий

ИИ только проводит анализ и не реагирует надлежащим образом на выявленные угрозы

*7. «Antifraud» — это … Объединение EDR и NDR в одну систему*

Антивирус, работающий на основе использования ИИ

Системы, позволяющие выявлять угрозы в бизнес-процессах и предотвращать мошеннические операции в режиме реального времени

*8. Что НЕ относится к преимуществам использования ИИ в сфере ИБ?* Сокращение времени выявления проблем и реагирования на инциденты

Увеличение штата сотрудников, обслуживающих СИИ

Сокращение затрат на организацию безопасности

*9. «lockbit» — это …*

Вредоносное ПО, блокирующее доступ к компьютерным системам и требующее от пользователя выкуп за восстановление данных

Антивирус, работающий на основе использования ИИ

Вредоносное ПО, внедряющее в систему ИИ для похищения данных пользователя

*10. «Cylane Protect» — это …*

Вредоносное ПО, блокирующее доступ к компьютерным системам и требующее от пользователя выкуп за восстановление данных

Антивирус, работающий на основе использования ИИ

Вредоносное ПО, внедряющее в систему ИИ для похищения данных пользователя

*3) Контроль выполнения ИДЗ, обсуждение со студентами хода выполнения выбранных тем ИДЗ, ответы на вопросы студентов по выполнению ИДЗ.*

**Практика № 7.**

*1) Опрос студентов по следующим вопросам:*

* Формирование базы медицинских (биомедицинских, биометрических) образов для обучения ИИ.
* Добровольное информированное согласие на участие в научных экспериментах.
* Возрастные ограничения участия в экспериментах.
* Принципы и задачи биоэтики.
* Обеспечение безопасности медицинских (биомедицинских, биометрических) данных при их получении, хранении и обработке в процессе обучения ИИ и проведения вычислительных экспериментов.

*2) Тестовые задания:*

*1. Необходимым условием работы систем с ИИ является?*

Анализ больших объемов данных

Участие оператора

Усиление безопасности

*2. Что такое биоэтика?*

Область междисциплинарных исследований этических, философских и антропологических проблем

Список норм при проведении научных исследований

Этика в биологии

3. Какая основная задача биоэтики? Заложить и воспитывать в человеке морально-нравственные, философские основы концепции мирного гармоничного соразвития человечества и его биоокружения

Регулировать исследования

Обезопасить всех участников исследований

*4. Могут ли данные, собранные в результате медицинского эксперимента, идентифицировать человека?*

Да

Нет

В совокупности с другими данными

*5. Каким должен быть источник данных?*

Открытым для общества

Закрытым для общества

Открытым для ограниченного круга лиц

*6. Во всех случаях информация, поступающая ИИ, должна быть деперсонифицирована?*

Да

Нет

В особых случаях

*7. В каком году был утвержден Нюрнбергский кодекс?*

1945

1947

1964

*8. От чего должна исходить задача ИИ в реальном времени?*

От реальной проблемы

От технологии

От желания разработчика

*9. Чем можно повысить эффективность ИИ в медицине?*

Заострением внимания на одном методе ИИ

Комбинированием различных методов ИИ

Повысить эффективность невозможно

*10. Собираемые данные должны быть пригодны к компьютерной обработке?*

Да

Нет

В некоторых случаях

*3) Контроль выполнения ИДЗ, обсуждение со студентами хода выполнения выбранных тем ИДЗ, ответы на вопросы студентов по выполнению ИДЗ.*

**Практика № 8.**

*1) Опрос студентов по следующим вопросам:*

* Характер рисков использования систем ИИ.
* Между­народные стандарты, содержащие рекомендации по управлению рисками, с которыми сталкиваются организации при разработке и применении методов и систем ИИ.
* Оценка свойств устойчивости систем ИИ.

*2) Тестовые задания:*

*1. «Объяснимый ИИ» — это …*

Модель ИИ, которая может предоставить общую информацию о том, как программа принимает решение, раскрывает сильные и слабые стороны программы, критерии принятия решения

Модель ИИ, которая может предоставить человеку информацию обо всём

Модель ИИ, которая применяется в адаптивном обучении

*2. «Технологическая сингулярность» — это …*

Точка во времени, когда ИИ выйдет из-под контроля человека

Момент времени, когда появился первый ИИ

Точка во времени, с которой машины начинают совершенствовать сами себя, без помощи кого-либо

*3. Кто сформулировал закон ускоряющейся отдачи?*

А. Тьюринг

Р. Курцвейл

К. Шеннон

*4. В каком стандарте рассматриваются вопросы обеспечения доверия к системам искусственного интеллекта?*

ГОСТ Р 53622—2009

ГОСТ 34.601

ГОСТ Р 59276—2020

*5. «Доверенная СИИ» — это …*

Уверенность потребителя, и при необходимости. Организаций, ответственных за регулирование вопросов создания и применения СИИ

СИИ, в отношении которой потребитель и. При необходимости, организации, ответственные за регулирование вопросов создания и применения СИИ, проявляют доверие

Минимально необходимая и достаточная совокупность характеристик СИИ, позволяющая потребителю. Организациям, ответственным за регулирование вопросов создания и применения СИИ, достоверно оценивать качество системы при решении конкретной прикладной задачи

*6. «Доверие к СИИ» — это …*

Уверенность потребителя, и при необходимости. Организаций, ответственных за регулирование вопросов создания и применения СИИ

СИИ, в отношении которой потребитель и. При необходимости, организации, ответственные за регулирование вопросов создания и применения СИИ, проявляют доверие

Минимально необходимая и достаточная совокупность характеристик СИИ, позволяющая потребителю. Организациям, ответственным за регулирование вопросов создания и применения СИИ, достоверно оценивать качество системы при решении конкретной прикладной задачи

*7. «Предвзятость» — это …*

Свойство СИИ, заключающееся в возможности представления причин, приводящих к тому или иному решению системы, в виде, понятном человеку

Свойство СИИ, заключающееся в способности принимать решения ожидаемым для человека способом

Свойство СИИ, заключающееся в принятии ошибочных решений, связанных со статистической смещенностью обучающей выборки исходных данных

*8. «Предсказуемость» — это …*

Свойство СИИ, заключающееся в возможности представления причин, приводящих к тому или иному решению системы, в виде, понятном человеку

Свойство СИИ, заключающееся в способности принимать решения ожидаемым для человека способом

Свойство СИИ, заключающееся в принятии ошибочных решений, связанных со статистической смещенностью обучающей выборки исходных данных

*9. Чем обеспечивается проверка доверия к СИИ?*

Подтверждением специально назначаемой комиссии о доверии СИИ

Соответствием характеристик СИИ международным стандартам в области регулирования деятельности, связанной с ИИ

Подтверждением соответствия представительного набора существенных характеристик системы ИИ требованиям, установленным разработчиком СИИ, потребителем СИИ и организацией, ответственной за регулирование вопросов создания и применения СИИ

*10. К чему ведёт быстрая масштабируемость ИИ?*

Неспособности контролировать

Повышает риск этических ошибок при разработке алгоритмов

Увеличению стоимости

*3) Контроль выполнения ИДЗ, обсуждение со студентами хода выполнения выбранных тем ИДЗ, ответы на вопросы студентов по выполнению ИДЗ.*

**Практика № 9.**

*1) Опрос студентов по следующим вопросам:*

* Понятие доверенного ИИ.
* Этические аспекты доверия к ИИ.
* Российские и мировые стандарты, связанные с доверенными системами ИИ.

*2) Тестовые задания:*

*1. Что НЕ является целью Рекомендации об этических аспектах ИИ?*

Стимулировать использование систем на основе ИИ в мирных целях

Использовать системы на основе ИИ для решения военных конфликтов

Заложить основу, которая позволит использовать ИИ на благо всего человечества, отдельного человека, обществ, окружающей среды и экосистем и не допустить причинения им вреда

*2. Согласно Рекомендации об этических аспектах ИИ «жизненный цикл ИИ» — это …*

Совокупность этапов развёртывания и использования без этапов проектирования и создания

Совокупность этапов проектирования и создания без этапов развёртывания и использования

Совокупность этапов, начиная с этапов исследования, проектирования и создания и заканчивая этапами развертывания и использования

*3. Что НЕ является ценностной установкой Рекомендации об этических аспектах ИИ?*

Содействие в военных конфликтах

Благополучие окружающей среды и экосистем

Обеспечение разнообразия и инклюзивности

*4. Решение об использовании ИИ-системы и выбор конкретного метода ее реализации должны основываться на следующих критериях:*

(а) благополучие окружающей среды и экосистем;

(b) обеспечение разнообразия и инклюзивности;

(с) жизнь в мирных, справедливых и взаимосвязанных обществах

(а) выбранный метод ИИ является подходящим и пропорциональным для достижения данной законной цели;

(b) выбранный метод ИИ не противоречит базовым ценностным установкам;

(с) выбранный метод ИИ подходит для конкретных условий и основывается на подтвержденных результатах научных исследований

(а) безопасность и защищенность;

(b) справедливость и отказ от дискриминации;

(с) устойчивость

*5. «Прозрачность работы СИИ» — это …*

Предоставление должной информации соответствующим адресатам для обеспечения понимания и повышения доверия

Доступность для понимания исходных данных, непосредственных результатов и функционирования каждой алгоритмической структуры и того, как все это влияет на выданный системой результат

Не только контроль со стороны отдельного человека, но и, в необходимых случаях, инклюзивный контроль со стороны общества

*6. «Объяснимость работы СИИ» — это …*

Предоставление должной информации соответствующим адресатам для обеспечения понимания и повышения доверия

Доступность для понимания исходных данных, непосредственных результатов и функционирования каждой алгоритмической структуры и того, как все это влияет на выданный системой результат

Не только контроль со стороны отдельного человека, но и, в необходимых случаях, инклюзивный контроль со стороны общества

*7. «Подконтрольность работы СИИ» — это …*

Предоставление должной информации соответствующим адресатам для обеспечения понимания и повышения доверия

Доступность для понимания исходных данных, непосредственных результатов и функционирования каждой алгоритмической структуры и того, как все это влияет на выданный системой результат

Не только контроль со стороны отдельного человека, но и, в необходимых случаях, инклюзивный контроль со стороны общества

*8. Какой организацией приняты Рекомендации об этических аспектах ИИ?*

ВОЗ

ЮНЕСКО

МККК

*9. «МКБ» — это …*

Международный комитет по биоэтике

Межправительственный комитет по биоэтике

Международная комиссия по этике научных знаний и технологий

*10. «МПКБ» — это …*

Международный комитет по биоэтике

Межправительственный комитет по биоэтике

Международная комиссия по этике научных знаний и технологий

*3) Контроль выполнения ИДЗ, обсуждение со студентами хода выполнения выбранных тем ИДЗ, ответы на вопросы студентов по выполнению ИДЗ.*

**Практика № 10.**

*1) Опрос студентов по следующим вопросам:*

* Итоги курса.
* Перспективы и направления развития этического ИИ в России и мире.
* Сильный и слабый ИИ.
* Тест Тьюринга.

*2) Тестовые задания:*

1. *Если ИИ способен собирать кубик Рубика, читать вслух, управлять автомобилем, то это -*

Сильный ИИ

Слабый ИИ

Супер ИИ

*2. Модель зрительного восприятия предложена*

Тьюринг

Розенблатт

Шенк

*3. При эволюционном подходе к построению систем ИИ*

Исследуется черный ящик

Идет наблюдение за величинами, которые поступают к нам на вход и за величинами, которые выходят от нас

Внимание уделяется построению начальной модели и правилам, по которым она может изменяться

*4. Функция вида Ф(x,w), которая возвращает предполагаемый номер класса объекта, характеризуемого вектором признаков x называется*

Тестировщиком

Учителем

Классификатором

*5. Является ли тест Тьюринга одним из лучших методов для оценки искусственного интеллекта?*

Да

Нет

Пока не доказано

*6. В чем состоит цель теста Тьюринга?*

В определении наличия ИИ у исследуемого объекта

В определении наличия интеллекта у исследуемого объекта

В определении уровня интеллекта у исследуемого объекта

*7. Кто задает вопросы исследуемому объекту в ходе теста Тьюринга?*

Множество компьютеров

Множество людей

Никто

*8. Кто предложил концепцию «китайской комнаты» для доказательства отсутствия понимания у программы?*

Алан Тьюринг

Роджер Шенк

Джон Серл

*9. Эксперимент «Китайская комната» демонстрирует, что*

Алгоритм ИИ полностью понимает выполняемые задачи

Алгоритм ИИ не имеет понимания выполняемых задач

Выполнение алгоритма ИИ говорит о понимании

*10. Что является имитацией осмысления алгоритма?*

Пока неизвестно

Решение с сохранением алгоритма

Решение без понимания алгоритма

*3) Выступление студентов с докладами (презентациями) по результатам выполнения выбранной темы ИДЗ. Ответы студентов на вопросы преподавателя по теме ИДЗ.*