

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Галунин Сергей Александрович
Должность: проректор по учебной работе
Дата подписания: 26.11.2024 14:26:37
Уникальный программный ключ:
08ef34338325bdb0ac5a47baa5472ce36cc3fc3b

Приложение к ОПОП
«Организация и программирова-
ние интеллектуальных систем»



СПбГЭТУ «ЛЭТИ»
ПЕРВЫЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ

МИНОБРНАУКИ РОССИИ

федеральное государственное автономное образовательное учреждение высшего образования
**«Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И.Ульянова (Ленина)»
(СПбГЭТУ «ЛЭТИ»)**

РАБОЧАЯ ПРОГРАММА

дисциплины

«ЭТИКА И БЕЗОПАСНОСТЬ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»

для подготовки бакалавров

по направлению

09.03.01 «Информатика и вычислительная техника»

по профилю

«Организация и программирование интеллектуальных систем»

Санкт-Петербург

2024

ЛИСТ СОГЛАСОВАНИЯ

Разработчики:

профессор, д.м.н. Мелдо А.А.

Рабочая программа рассмотрена и одобрена на заседании кафедры ВТ
09.01.2024, протокол № 1

Рабочая программа рассмотрена и одобрена учебно-методической комиссией
ФКТИ, 24.01.2024, протокол № 1

Согласовано в ИС ИОТ

Начальник ОМОЛА Загороднюк О.В.

1 СТРУКТУРА ДИСЦИПЛИНЫ

Обеспечивающий факультет	ФКТИ
Обеспечивающая кафедра	ВТ
Общая трудоемкость (ЗЕТ)	3
Курс	4
Семестр	8

Виды занятий

Лекции (академ. часов)	16
Практические занятия (академ. часов)	16
Иная контактная работа (академ. часов)	1
Все контактные часы (академ. часов)	33
Самостоятельная работа, включая часы на контроль (академ. часов)	75
Всего (академ. часов)	120

Вид промежуточной аттестации

Лифф зачет (курс) 1

2 АННОТАЦИЯ ДИСЦИПЛИНЫ

«ЭТИКА И БЕЗОПАСНОСТЬ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»

Исторически сложилось, что развитие технологий происходит не планомерно, а экспоненциально. Примерами могут служить четыре промышленных революции, под влиянием которых происходила перестройка общества в целом. Каждая из них порождалась экономической целесообразностью и привлекательностью повышения уровня и качества жизни социума. XXI век ознаменован четвертой промышленной революцией, которая подразумевает широкое внедрение киберфизических систем в производство, в том числе технологий на основе искусственного интеллекта (ИИ). Футурологические исследования ИИ основаны на двух основных сценариях, воспринимаемых с одной стороны, как мощный скачок производственных процессов, трансформацию и оптимизацию всех сфер жизни человека, с другой стороны, как повышение риска непрогнозируемых последствий, связанных с сегрегацией общества, кибербезопасностью, юридической ответственностью, расшатыванием политических систем. С этих позиций высока актуальность интеграции нормативной и прикладной этики в технически трансформирующемся обществе. Этические вопросы, связанные с ИИ, затрагивают всех участников жизненного цикла этих систем, от разработчиков до конечных пользователей, существует потребность юридического, технического, социального регулирования с целью сохранения подконтрольности этого нового явления.

Данная дисциплина направлена на изучение исторических аспектов нормативной и прикладной этики, отечественного и международного правового регулирования по вопросам обеспечения и сохранения прав и свобод человека в условиях использования систем ИИ, применение и адаптацию традиционных этических норм к условиям глобальной цифровизации.

SUBJECT SUMMARY

«ETHICAL AND SECURITY OF ARTIFICIAL INTELLIGENCE»

Historically, the development of technology does not occur systematically, but exponentially. Examples are four industrial revolutions, which determined global restructuring of society. Each of them was caused with the economic profitability and improvement of society life level and quality. The XXI century is marked by the fourth industrial revolution, which implies the widespread introduction of cyber-physical systems, including technologies based on artificial intelligence (AI) into industry. Futurological researches of AI is based on two main scripts. On the one hand it is a powerful move of industry processes, transformation and optimization of all human life spheres, on the other hand, it is an increasing of the risk of unpredictable consequences which are segregation of society, cybersecurity, legal responsibility, and destabilisation of political systems. So, the relevance of integrating normative and applied ethics into a technically transforming society is high. Ethical issues related to AI affect all participants in the life cycle of these systems, from developers to end users, there is a need for legal, technical, and social regulation in order to maintain the control of this new phenomenon.

This discipline is aimed at studying the historical aspects of normative and applied ethics, domestic and international legal regulation on ensuring and preserving human rights and freedoms in the conditions of using AI systems, the application and adaptation of traditional ethical norms to the of global digitalization condition.

3 ОБЩИЕ ПОЛОЖЕНИЯ

3.1 Цели и задачи дисциплины

1. Цели дисциплины:

- освоение принципов мультидисциплинарного взаимодействия при создании систем на основе ИИ;
- разработка и применение концептуальных подходов взаимодействия человека и искусственного интеллекта с сохранением ценности прав и свобод человека;
- формирование навыков разработки методов и алгоритмов машинного обучения для различных сфер деятельности с учетом принципов этики.

2. Задачи изучения дисциплины:

- изучение тенденций развития ИИ в России и в мире наряду с оценкой потенциальных рисков применения технологий ИИ;
- изучение нормативноправовой базы России и мирового опыта нормативного регулирования в области информационных технологий;
- изучение и разработка способов использования систем ИИ, гарантирующих их безопасность и надежность для человека

3. Знания нормативно-правовой базы в области разработки и использования ИИ, защиты результатов интеллектуальной деятельности.

4. Умения:

- анализировать и оценивать этические проблемы, связанные с разработкой систем ИИ
- применять принципы этики при создании систем ИИ;
- критически оценивать риски и последствия внедрения систем ИИ;
- понимать социальные, правовые и экономические аспекты при внедрении систем ИИ;
- развивать критическое мышление и способность анализировать сложные эти-

ческие проблемы в области ИИ.

5. Приобретаемые навыки:

- способность к мультидисциплинарному взаимодействию;
- принятие решений для устранения ограничений прав и свобод и угроз безопасности человека при применении систем ИИ;
- применение этических норм и норм права при создании и внедрении систем ИИ.

3.2 Место дисциплины в структуре ОПОП

Дисциплина изучается на основе ранее освоенных дисциплин учебного плана:

1. «Основы машинного обучения»

и обеспечивает подготовку выпускной квалификационной работы.

3.3 Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

В результате освоения образовательной программы обучающийся должен достичь следующие результаты обучения по дисциплине:

Код компетенции/ индикатора компетенции	Наименование компетенции/индикатора компетенции
ПК-0	Способен разрабатывать информационные модели и применять их для решения задач профессиональной деятельности
ПК-0.1	<i>Знает современные виды информационных моделей, применяемых при решении задач профессиональной деятельности</i>
ПК-0.2	<i>Создает и модифицирует информационные модели для решения задач профессиональной деятельности</i>
ПК-0.3	<i>Применяет информационные модели для решения задач профессиональной деятельности</i>

4 СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1 Содержание разделов дисциплины

4.1.1 Наименование тем и часы на все виды нагрузки

№ п/п	Наименование темы дисциплины	Лек, ач	Пр, ач	ИКР, ач	СР, ач
1	Введение. Этика в ходе истории технологий.	2	2		6
2	Основные принципы этики в области ИИ. Кодекс этики в сфере ИИ.	2	2		15
3	Правовые аспекты разработки и применения систем ИИ.	2	2		8
4	Принцип ответственности и социальной справедливости в разработке и использовании систем ИИИ	2	2		8
5	Этические проблемы, связанные с разработкой и использованием систем ИИ в медицине.	2	2	1	14
6	Этические проблемы, связанные с разработкой и использованием систем ИИ в образовании.	2	2		7
7	Вопросы безопасности при использовании систем ИИ.	2	2		10
8	Роль междисциплинарного взаимодействия при разработке систем ИИ. Заключение	2	2		7
	Итого, ач	16	16	1	75
	Из них ач на контроль	0	0	0	0
	Общая трудоемкость освоения, ач/зе				108/3

4.1.2 Содержание

№ п/п	Наименование темы дисциплины	Содержание
1	Введение. Этика в ходе истории технологий.	Предмет дисциплины, ее задачи. Этика в человеческом обществе, исторические примеры этических проблем, связанных с развитием технологий
2	Основные принципы этики в области ИИ. Кодекс этики в сфере ИИ.	Предпосылки создания кодекса этики в сфере ИИ. Акторы ИИ и их ответственность. Вопросы назначения технологий ИИ, польза и риски применения.
3	Правовые аспекты разработки и применения систем ИИ.	Вопросы контроля над технологиями ИИ со стороны человека. Нормативно-правовое регулирование разработок ИИ в РФ и мире
4	Принцип ответственности и социальной справедливости в разработке и использовании систем ИИИ	Оценка качества систем ИИ. Прикладные задачи ИИ. Обзор социальных рисков, связанных с использованием ИИ: влияние на занятость населения, появление новых профессий, массовые коммуникации и социальные сети, влияние ИИ на культуру

№ п/п	Наименование темы дисциплины	Содержание
5	Этические проблемы, связанные с разработкой и использованием систем ИИ в медицине.	Доверие к новым технологиям со стороны пациента и врача. Вопросы защиты персональных данных пациента. Факторы, ограничивающие применение систем ИИ в медицине, вопросы внедрения технологий ИИ в медицине. Оценка пользы и рисков применения технологий ИИ в медицине
6	Этические проблемы, связанные с разработкой и использованием систем ИИ в образовании.	Трансформация системы образования под влиянием новых технологий. Этические аспекты отношений «ученик-учитель» в условиях развития технологий ИИ. Оценка пользы и рисков применения технологий ИИ в сфере образования
7	Вопросы безопасности при использовании систем ИИ.	Этические аспекты в задачах обеспечения безопасности использования ИИ. 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. Обезличивание персональных данных. Эффективность использования физиологических и поведенческих признаков человека в задачах идентификации и аутентификации личности.
8	Роль междисциплинарного взаимодействия при разработке систем ИИ. Заключение	Определение междисциплинарного взаимодействия. Вклад знаний специалистов различных сфер в разработку систем ИИ. Развитие новых компетенций под влиянием развития ИИ. Сотрудничество и конкуренция в вопросах разработки систем ИИ. Итоги курса. Перспективы развития вопросов этики в условиях четвертой технологической революции

4.2 Перечень лабораторных работ

Лабораторные работы не предусмотрены.

4.3 Перечень практических занятий

Наименование практических занятий	Количество ауд. часов
1. Методологические основы и принципы ИИ	2
2. Создание и использование этичного и социально приемлемого ИИ	2
3. Подходы к измерению функциональных характеристик прикладных систем ИИ	4
4. Методы оценки рисков, возникающих из-за внедрения ИИ	4
5. Процедуры подтверждения безопасности функционирования системы ИИ	4
Итого	16

4.4 Курсовое проектирование

Курсовая работа (проект) не предусмотрены.

4.5 Реферат

Реферат не предусмотрен.

4.6 Индивидуальное домашнее задание

Индивидуальное домашнее задание не предусмотрено.

4.7 Доклад

Доклад не предусмотрен.

4.8 Кейс

Кейс не предусмотрен.

4.9 Организация и учебно-методическое обеспечение самостоятельной работы

Изучение дисциплины сопровождается самостоятельной работой студентов с рекомендованными преподавателем литературными источниками и информационными ресурсами сети Интернет.

Планирование времени для изучения дисциплины осуществляется на весь период обучения, предусматривая при этом регулярное повторение пройденного материала. Обучающимся, в рамках внеаудиторной самостоятельной работы, необходимо регулярно дополнять сведениями из литературных источников материал, законспектированный на лекциях. При этом на основе изучения рекомендованной литературы целесообразно составить конспект основных положений, терминов и определений, необходимых для освоения разделов учебной

дисциплины.

Особое место уделяется консультированию, как одной из форм обучения и контроля самостоятельной работы. Консультирование предполагает особым образом организованное взаимодействие между преподавателем и студентами, при этом предполагается, что консультант либо знает готовое решение, которое он может предписать консультируемому, либо он владеет способами деятельности, которые указывают путь решения проблемы.

Текущая СРС	Примерная трудоемкость, ач
Работа с лекционным материалом, с учебной литературой	20
Опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	0
Самостоятельное изучение разделов дисциплины	0
Выполнение домашних заданий, домашних контрольных работ	0
Подготовка к лабораторным работам, к практическим и семинарским занятиям	25
Подготовка к контрольным работам, коллоквиумам	0
Выполнение расчетно-графических работ	0
Выполнение курсового проекта или курсовой работы	0
Поиск, изучение и презентация информации по заданной проблеме, анализ научных публикаций по заданной теме	0
Работа над междисциплинарным проектом	0
Анализ данных по заданной теме, выполнение расчетов, составление схем и моделей, на основе собранных данных	0
Подготовка к зачету, дифференциированному зачету, экзамену	30
ИТОГО СРС	75

5 Учебно-методическое обеспечение дисциплины

5.1 Перечень основной и дополнительной литературы, необходимой для освоения дисциплины

№ п/п	Название, библиографическое описание	К-во экз. в библ.
Основная литература		
1	Искусственный интеллект [Текст] : Справ.: В 3 кн. / Под ред. Э.В.Попова. Кн. 2 : Модели и методы : справ. изд. / Под ред. Д.А.Поспелова, 1990. - 303, [1] с.	46
2	Душкин Р. В. Искусственный интеллект [Электронный ресурс] : научное издание / Р. В. Душкин, 2019. -280 с.	неогр.
3	Хултен Д. Разработка интеллектуальных систем [Электронный ресурс] : руководство / Д. Хултен, 2019. -284 с.	неогр.
4	Никольский С. Н. Автоматизация информационного поведения и искусственный интеллект [Электронный ресурс] : учебное пособие, 2020. -95 с.	неогр.
Дополнительная литература		
1	Родичев Ю. А. Информационная безопасность. Национальные стандарты Российской Федерации. 2-е изд. Учебное пособие [Электронный ресурс] / Ю. А. Родичев, 2019. -304 с.	неогр.

5.2 Перечень ресурсов информационно-телеkomмуникационной сети «Интернет», используемых при освоении дисциплины

№ п/п	Электронный адрес
1	ISO/IEC/IEEE 12207:2017, Systems and software engineering —Software life cycle processeshttps://www.iso.org/standard/63712.html
2	ISO/IEC 247752:2021, Information technology —Storage management —Part 2: Common Architecturehttps://www.iso.org/standard/81142.html
3	ISO/IEC CD 23894, Artificial Intelligence—Risk Managementhttps://www.iso.org/standard/77304.html
4	Технический Комитет 164https://www.tc164.ru/

5.3 Адрес сайта курса

Адрес сайта курса: <https://vec.etu.ru/moodle/course/view.php?id=14845>

6 Критерии оценивания и оценочные материалы

6.1 Критерии оценивания

Для дисциплины «Этика и безопасность искусственного интеллекта» предусмотрены следующие формы промежуточной аттестации: зачет с оценкой.

Зачет с оценкой

Оценка	Описание
Неудовлетворительно	Курс не освоен. Студент испытывает серьезные трудности при ответе на ключевые вопросы дисциплины
Удовлетворительно	Студент в целом овладел курсом, но некоторые разделы освоены на уровне определений и формулировок
Хорошо	Студент овладел курсом, но в отдельных вопросах испытывает затруднения. Умеет решать задачи
Отлично	Студент демонстрирует полное овладение курсом, способен применять полученные знания при решении конкретных задач

Особенности допуска

В соответствии с текущим контролем успеваемости и посещения лекций и занятий студенты допускаются к дифф. зачету. Дифф. зачет проводится в очном формате в отведенное время. В соответствии с изученными темами, на основе перечня контрольных вопросов студенту предлагается взять один билет, состоящий из двух вопросов. Время на подготовку 20 минут. Оценка выставляется по результату устного ответа студента. Оценивается раскрытие темы вопроса , умение подать материал. При необходимости студенту предлагается дополнительный вопрос.

6.2 Оценочные материалы для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине

Вопросы к дифф.зачету

№ п/п	Описание
1	ИИ и современный мир (основные направления изменения общественной жизни).
2	Тест Тьюринга (преимущества и недостатки)
3	ИИ как субъект права
4	Риски нарушения прав и свобод человека при использовании ИИ
5	Какими должны быть обязанности разработчика ИИ?
6	Междисциплинарное взаимодействие при разработке и использовании систем ИИ
7	Задачи обеспечения информационной безопасности при использовании ИИ
8	Качество наборов данных при создании систем ИИ
9	Предпосылки создания Кодекса в сфере ИИ
10	Этика в информационном обществе
11	Что первично: этика или технологический прогресс
12	Использование систем ИИ в здравоохранении, преимущества и недостатки
13	Использование систем ИИ с системе образования, преимущества и недостатки
14	ИИ, авторское право, право на интеллектуальную собственность
15	Можно ли запретить некоторые виды технологий ИИ? «За» и «против» ограничений применения технологий ИИ
16	ИИ и рынок труда (основные тренды)
17	Машинное обучение с учителем. Роль эксперта в создании алгоритма ИИ
18	Этика использования технологий ИИ при итоговой аттестации. Перспективы и риски использования ГПТ.
19	Особенности этических аспектов ИИ в медицине. Значение доказательности и презентативности результатов, полученных с помощью ИИ.

Форма билета

Министерство науки и высшего образования Российской Федерации
ФГАОУ ВО «Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» имени В.И. Ульянова (Ленина)»

БИЛЕТ № 1

Дисциплина Этика и безопасность искусственного интеллекта ФКТИ

1. Искусственный интеллект и рынок труда (основные тренды).
2. Качество наборов данных при создании систем ИИ.

УТВЕРЖДАЮ

Заведующий кафедрой

М.С. Куприянов

Образцы задач (заданий) для контрольных (проверочных) работ

Примеры вопросов к коллоквиуму:

Какие философские и этические концепции оказали наибольшее влияние на формирование этики технологий?

Какие вызовы и проблемы возникают при автоматизированном принятии решений в диагностике и лечении?

Как влияет использование ГПТ на когнитивные способности человека?

Доступ к информации в XX и в XXI веке, особенности, задачи, требующие решения.

Весь комплект контрольно-измерительных материалов для проверки сформированности компетенции (индикатора компетенции) размещен в закрытой части по адресу, указанному в п. 5.3

6.3 График текущего контроля успеваемости

Неделя	Темы занятий	Вид контроля
1	Принципы ИИ. Национальный Кодекс этики в сфере искусственного интеллекта.	Коллоквиум
2	Вопросы социальной ответственности, связанные с участием и развитием общества при внедрении систем ИИ	Коллоквиум
3	Методы оценки рисков, возникающих из-за внедрения ИИ. Передовые методы обнаружения и снижения воздействия ИИ	Коллоквиум
4	Обеспечение безопасности при применении технологии ИИ	Коллоквиум
5	Этические аспекты доверия к ИИ	
6		Коллоквиум

6.4 Методика текущего контроля

на лекционных занятиях

Текущий контроль включает в себя:

- контроль посещаемости (не менее 80 % занятий);

на практических (семинарских) занятиях

Практические работы защищаются студентами индивидуально. Каждый студент получает вопрос по теоретической и практической части после чего ему предоставляется время для подготовки ответа. При обсуждении ответа преподаватель может задать несколько уточняющих вопросов. В случае если студент демонстрирует достаточное знание вопроса, работа считается защищенной.

На защите практической работы студент должен показать: понимание и умение объяснять особенности применяемых методов, возможные области их применения и т.д., прогнозировать реакции исследуемого объекта на различные воздействия, навыки и умения, приобретенные при выполнении практических работы.

Критерием оценки работы на коллоквиумах является оценка, выставляемая по 5-ти балльной шкале в соответствии со следующими критериями:

оценка в 5 баллов ("отлично") выставляется при отличном выполнении задания, то есть при наличии полных (с детальными пояснениями и выкладками), оригинальных и правильных решений задач, дополненных при необходимости документами, полученными в результате реализации (проверки) решения, верных ответов и высококачественного оформления работы;

оценка в 4 баллов ("хорошо") выставляется при правильном выполнении задания, то есть при наличии полных (с пояснениями и выкладками), оригинальных и правильных решений задач, дополненных при необходимости документами, полученными в результате реализации (проверки) решения, верных ответов;

оценка в 3 баллов ("удовлетворительно") выставляется при наличии отдельных неточностей в ответах (включая грамматические ошибки) или неточностях в решении задач непринципиального характера (описки и случайные ошибки арифметического характера);

оценка в 2 и ниже баллов ("неудовлетворительно") выставляется в случаях, когда в ответах и в решениях задач имеются неточности и ошибки, свидетельствующие о недостаточном понимании вопросов и требующие дополнительного обращения к тематическим материалам.

Текущий контроль включает в себя выполнение, сдачу в срок отчетов и их защиту по всем практическим работам, по результатам которой студент получает допуск к зачету.

самостоятельной работы студентов

Контроль самостоятельной работы студентов осуществляется на практических занятиях студентов по методикам, описанным выше.

7 Описание информационных технологий и материально-технической базы

Тип занятий	Тип помещения	Требования к помещению	Требования к программному обеспечению
Лекция	Лекционная аудитория	1) Количество посадочных мест – в соответствии с контингентом, 2) рабочее место преподавателя, персональный компьютер (IBM, совместимый Pentium или выше), проектор, экран/интерактивная панель, меловая/маркерная доска.	1) Windows XP и выше; 2) Microsoft Office 2007 и выше
Практические занятия	Аудитория	1) Количество посадочных мест – в соответствии с контингентом; 2) рабочее место преподавателя, персональный компьютер (IBM, совместимый Pentium или выше), проектор, экран/интерактивная панель, меловая/маркерная доска.	1) Windows XP и выше; 2) Microsoft Office 2007 и выше
Самостоятельная работа	Помещение для самостоятельной работы	Оснащено компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.	1) Windows XP и выше; 2) Microsoft Office 2007 и выше

8 Адаптация рабочей программы для лиц с ОВЗ

Адаптированная программа разрабатывается при наличии заявления со стороны обучающегося (родителей, законных представителей) и медицинских показаний (рекомендациями психолого-медико-педагогической комиссии). Для инвалидов адаптированная образовательная программа разрабатывается в соответствии с индивидуальной программой реабилитации.

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ

№ п/п	Дата	Изменение	Дата и номер протокола заседания УМК	Автор	Начальник ОМОЛА