ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

УТВЕРЖДЕНА

Решением Ученого совета
АНО ВО «Центральный университет»
«24» июня 2025 г.
Протокол № 2

Рабочая программа дисциплины (модуля)
«Черные ящики и цифровые агенты: социология искусственного интеллекта»

Направление подготовки: *02.03.01 Математика и компьютерные науки*

Направленность (профиль) подготовки: *Разработка*

Квалификация (степень) выпускника: *бакалавр*

Форма обучения: *очная*

Срок освоения программы: *4 года*

Год набора: *2025*

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 0215B0C000D1B3FEB44A37EE631E0C27D4
Владелец: Ивашкевич Евгений Васильевич
Действителен: с 13.01.2026 до 13.04.2027

Электронный документ

Москва
2025

Содержание

1. **Краткая характеристика дисциплины (модуля)** 3
2. **Перечень планируемых результатов обучения** 5
3. **Тематический план** 7
4. **Содержание дисциплины (модуля)** 7
5. **Учебно-методическое обеспечение** 9
6. **Материально-техническое обеспечение** 9
7. **Методические и оценочные материалы** 11

1. Краткая характеристика дисциплины (модуля)

Рабочая программа дисциплины (модуля) «Черные ящики и цифровые агенты: социология искусственного интеллекта» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – бакалавриат по направлению 02.03.01 «Математика и компьютерные науки», профиль «Разработка», утвержденный приказом Министерства науки и высшего образования Российской Федерации № 807 от 23.08.2017 года.

Изучение дисциплины (модуля) «Черные ящики и цифровые агенты: социология искусственного интеллекта» формирует способность обучающихся видеть искусственный интеллект не как автономный инструмент, а как социотехнический фактор, встроенный в сложные сети взаимодействий между людьми, алгоритмами, инфраструктурами и институтами.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки бакалавриата по направлению 02.03.01 «Математика и компьютерные науки», профиль «Разработка» и входит в вариативную часть Блока 1 формируемую участниками образовательных отношений как дисциплина по выбору.

Дисциплина (модуль) доступна к изучению на 2, 3 или 4 курсе с 3 по 8 семестры на выбор.

Цель изучения дисциплины (модуля): сформировать у обучающихся способность анализировать искусственный интеллект не как автономную технологию, а как социотехнический актор, встроенный в сети взаимодействия людей, инструментов и инфраструктур, и научить проектировать его внедрение с учётом операционных, организационных, культурных и властных последствий.

Задачи изучения дисциплины (модуля):

— освоить ключевые концепты социологии ИИ — «чёрный ящик», актор, сеть, двойной интерфейс — и применять их для анализа цифровых систем;

— научиться описывать цифровые приложения как гетерогенные сети, включающие как людей, так и нечеловеческих акторов (алгоритмы, API, файлы, протоколы);

— овладеть методикой анализа поведения ИИ-агента по логам и записям сессий, выявляя логику его действий и точки нештатного поведения;

— развить навык проектирования «пространства для агента» с учётом двойного интерфейса, фонового режима и границ взаимодействия;

— научиться оценивать и аргументировать социальный эффект от внедрения ИИ на четырёх уровнях: операционном, организационном, культурном и властном.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

— базовые понятия социологии ИИ: актор/актант, сеть, «чёрный ящик» (blackboxing), антропоморфизм, ситуативное действие;

— из чего складывается работа ИИ-агента: цикл «запрос → рассуждение → использование инструмента → действие», роль инфраструктуры (API, права доступа, файлы);

— понятие двойного интерфейса (вывод для человека и для машины) и принципы «тихого» ИИ (human-out-of-the-loop);

— понятие калиброванного доверия к алгоритмам и причины, по которым прозрачность не всегда повышает доверие;

— четыре уровня социального эффекта (ROI) от внедрения ИИ: операционный, организационный, культурный, властный.

уметь:

— описывать цифровое приложение как сеть людей и не-людей (находить не менее 4 нечеловеческих акторов);

— анализировать запись сессии ИИ-агента: восстанавливать логику, находить точку сбоя, отличать «план» от ситуативного поведения;

— распознавать антропоморфные ловушки в интерфейсах и предлагать неантропоморфные решения;

— оценивать выгоду от внедрения ИИ по четырём уровням, а не только в сэкономленных часах;

— аргументированно защищать проектные решения, в том числе созданные с помощью ИИ.

владеть:

— навыками «вскрытия чёрного ящика» — реконструкции работы ИИ-системы по следам (логам, скриншотам, распечаткам);

— методикой проектирования «пространства для агента» (вольера): двойной интерфейс, границы, фоновый режим работы;

— инструментами социотехнического анализа бизнес-процесса для внедрения ИИ-агента;

— навыком подготовки и публичной защиты концепта (Pitch Deck и аналитическая записка).

2. Перечень планируемых результатов обучения

Компетенции, формируемые в результате освоения дисциплины (модуля) при проведении учебных занятий в форме контактной работы обучающихся с педагогическими работниками Университета и в форме самостоятельной работы обучающихся:

Компетенция	Содержание компетенции	Индикатор компетенции	Перечень планируемых результатов обучения по дисциплине (модулю)
УК-1.	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.1.	Знает методы поиска и анализа информации в области разработки, основные принципы критической оценки источников информации и их релевантности
		УК-1.2.	Умеет критически оценивать источники информации и синтезировать данные из различных источников для решения задач, применять системный подход к анализу и решению комплексных проблем
		УК-1.3.	Имеет практический опыт работы с современными инструментами и технологиями для обработки информации, формулировании и структурировании задач на основе полученной информации
УК-2.	Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.1.	Знает действующие правовые нормы, регулирующие деятельность в области решения задач, основные методы и подходы к определению круга задач
		УК-2.2.	Умеет определять круг задач в рамках поставленной цели, выбирать оптимальные способы решения задач, учитывая имеющиеся ресурсы и ограничения
		УК-2.3.	Имеет практический опыт применения знаний о правовых нормах и ресурсах в реальных ситуациях, разработки и реализации решений в соответствии с установленными ограничениями
ОПК-1.	Способен консультировать и использовать фундаментальные знания в области математического анализа, комплексного и функционального анализа алгебры, аналитической	ОПК-1.1.	Знает основные концепции и теории в области математического анализа и смежных дисциплин; методы и подходы, используемые в различных областях математики

	геометрии, дифференциальной геометрии и топологии, дифференциальных уравнений, дискретной математики и математической логики, теории вероятностей, математической статистики и случайных процессов, численных методов, теоретической механики в профессиональной деятельности	ОПК-1.2.	Умеет применять математические методы для решения профессиональных задач
		ОПК-1.3.	Имеет практический опыт разработки и реализация математических моделей в профессиональной деятельности
ПК-1.	Способен формулировать задачи с математической точностью, обосновывать утверждения строго и анализировать полученные результаты в области математики и компьютерных наук	ПК-1.1.	Обладает базовыми знаниями, полученными в области математических наук, программирования и информационных технологий
		ПК-1.2.	Умеет находить, формулировать и решать стандартные задачи в собственной научно-исследовательской деятельности в математике и информатике
		ПК-1.3.	Имеет опыт работы с задачами в области математики и компьютерных наук, включая применение математических методов для решения практических задач

3. Тематический план

№ п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы				ТКУ (текущий контроль успеваемости)
		Очная форма				
		Аудиторная работа		Контроль	Самостоятельная работа	
		Семинары	Консультации			
1	Конец антропоморфизма: ИИ как нечеловеческий актор	6			16	Аудиторная работа
2	Пространство для агентов: как спроектировать «вольер»	6			17	Аудиторная работа
3	Вскрытие чёрного ящика: анатомия цифровой агентности (практикум)	6			17	Аудиторная работа, Коллоквиум
4	Бизнес-кейс: Оцифровка организации	6			17	Аудиторная работа, Защита кейсов
5	Человек — Машина — Организация: синтез и защита проектов	6			17	Аудиторная работа, Проект
	Зачет с оценкой					
	Итого:	***30***			***84***	
Объем дисциплины (модуля) (в ак. ч.)		***114***				
Объем дисциплины (модуля) (в зач. ед.)		***3***				

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	Конец антропоморфизма: ИИ как нечеловеческий актор	ИИ — не «умная программа»: знакомство с курсом и взглядом социолога. Бруно Латур и идея «актора»: почему дверь — тоже немного «учёный». «Чёрный ящик»: почему ИИ ломается там, где человек бы импровизировал.
2	Пространство для агентов: как спроектировать «вольер»	Где «живёт» ИИ-агент: невидимая инфраструктура вокруг нас Один сайт — два читателя: что такое «двойной интерфейс». Тихий ИИ: когда лучше, чтобы агент не разговаривал. Защита кейсов.
3	Вскрытие чёрного ящика: анатомия цифровой агентности (практикум)	Что происходит, когда ChatGPT «думает»: цикл ИИ-агента. Практикум «Аутопсия агента»: разбираем сессию ИИ как исследователи. Когда мы разрешаем ИИ действовать: вопрос ответственности. Калиброванное доверие к ИИ. Коллоквиум по Блокам 1–3.
4	Бизнес-кейс: Оцифровка организации	Как ИИ зашёл в обычную канцелярию большой компании.

		Воркшоп «Де-антропоморфизация»: убираем «человечность» из чат-ботов. Ролевая игра «Комитет по внедрению» + как считать выгоду по 4 уровням.
5	Человек — Машина — Организация: синтез и защита проектов	Гибридный коллектив. Финальная подготовка проекта.

5. Учебно-методическое обеспечение

Университет располагает полным набором лицензионного и свободно распространяемого программного обеспечения, включая продукты отечественного производства.

Каждый студент в течение всего периода обучения получает индивидуальный неограниченный доступ к электронно-библиотечной системе и электронной информационно-образовательной среде университета. Эти системы предоставляют возможность доступа к ресурсам из любой точки, где есть подключение к сети Интернет, как на территории университета, так и за его пределами.

Студентам обеспечен удаленный доступ к современным профессиональным базам данных и информационным справочным системам.

Основная литература:

1. Бессмертный, И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — 2-е изд. — Москва : Издательство Юрайт, 2026. — 250 с. — (Высшее образование). — ISBN 978-5-534-20734-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/583859.

2. Форд, М. Архитекторы интеллекта: вся правда об искусственном интеллекте от его создателей : практическое руководство / М. Форд. - Санкт-Петербург : Питер, 2020. - 416 с. - (Серия «Библиотека программиста»). - ISBN 978-5-4461-1254-8. - Текст : электронный. - URL: https://znanium.com/catalog/product/1733710.

Дополнительная литература:

1. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : учебник / П. Флах. - 2-е изд. - Москва.:ДМК Пресс, 2023. - 401 с. - ISBN 978-5-89818-300-4. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2085038.

2. Мэрфи, К. П. Вероятностное машинное обучение: введение : практическое руководство / К. П. Мэрфи ; пер. с англ. А. А. Слинкина. - Москва : ДМК Пресс, 2023. - 990 с. - ISBN 978-5-93700-119-1. - Текст : электронный. - URL: https://znanium.com/catalog/product/2109489.

6. Материально-техническое обеспечение

Университет располагает материально-технической базой, соответствующей действующим противопожарным правилам и нормам и обеспечивающей проведение всех видов дисциплинарной и междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Помещения, которые представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского (практического) типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования. Помещения укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Изучение дисциплины (модуля) обеспечивается в учебных аудиториях, оснащенных:

- столами и стульями;
- компьютерной техникой;
- механическими калькуляторами;

— специализированным оборудованием, включая демонстрационное оборудование.

Помещения для самостоятельной работы обучающихся, в том числе приспособленные для использования инвалидами и лицами с ограниченными возможностями здоровья, оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду Университета.

Обучающимся предоставляется доступ (в том числе удаленный) к ресурсам информационно-телекоммуникационной сети «Интернет», электронным ресурсам (в том числе электронным библиотечным системам, современным профессиональным базам данных и информационным справочным системам):

№	Наименование портала (издания, курса, документа)	Ссылка
1.	Научная электронная библиотека elibrary.ru библиотека	https://elibrary.ru/defaultx.asp
2.	База данных для IT-специалистов	https://habr.com
3.	База данных ScienceDirect	https://www.sciencedirect.com
4.	Официальный сайт Министерства науки и высшего образования Российской Федерации	https://minobrnauki.gov.ru/
5.	Федеральный портал «Российское образование"	https://www.edu.ru/
6.	Информационная система "Единое окно доступа к образовательным ресурсам"	http://window.edu.ru/
7.	Единая коллекция цифровых образовательных ресурсов	http://school-collection.edu.ru/
8.	Федеральный центр информационно - образовательных ресурсов	http://fcior.edu.ru/

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), в том числе комплект лицензионного программного обеспечения, современные профессиональные базы данных и информационные справочные системы:

Наименование ПО	Производство	Лицензионное / свободно распространяемое
Операционные системы:		
Microsoft Imagine (Windows Client, Server)	зарубежное	лицензионное
Браузеры:		
Яндекс.Браузер	отечественное	свободно распространяемое
Google Chrome	зарубежное	свободно распространяемое
Офисные приложения:		
Microsoft Imagine (Visio, OneNote)	зарубежное	лицензионное
TeXstudio	зарубежное	свободно распространяемое
Adobe Acrobat Reader	зарубежное	свободно распространяемое
Программное обеспечение для планирования и учета времени:		
Toggle app	зарубежное	свободно распространяемое
Системы управления проектами:		
Microsoft Imagine (Project)	зарубежное	лицензионное
Системы управления базами данных:		
Microsoft Imagine (SQL Server)	зарубежное	лицензионное
Системы резервного копирования (backup):		
Acronis Backup Advanced for HyperV	зарубежное	лицензионное
Справочно-правовые системы:		
КонсультантПлюс: справочно-правовая система	отечественное	лицензионное

Средства антивирусной защиты:		
Kaspersky Endpoint Security для бизнеса Стандартный Russian Edition	отечественное	лицензионное
Среды разработки:		
Visual Studio Code	зарубежное	свободно распространяемое
Bash (Unix shell)	зарубежное	свободно распространяемое
Anaconda	зарубежное	свободно распространяемое
Robotic Operating System	зарубежное	свободно распространяемое
CopelliaSim	зарубежное	свободно распространяемое
Google Colaboratory	зарубежное	свободно распространяемое
Пакеты программных средств и библиотек:		
AutoPsy	зарубежное	свободно распространяемое
Interactive Disassembler (IDA)	зарубежное	свободно распространяемое
Системы управления библиографической информацией:		
Zotero	зарубежное	свободно распространяемое
Сервисы и службы:		
Bind	зарубежное	свободно распространяемое
Docker	зарубежное	свободно распространяемое

7. Методические и оценочные материалы

Методические указания для обучающихся по освоению дисциплины (модуля)

В процессе изучения дисциплины (модуля) «Черные ящики и цифровые агенты: социология искусственного интеллекта» в рамках текущего контроля успеваемости используются такие виды учебной работы, как семинары, аудиторная работа, кейсы, коллоквиумы, проект, а также различные виды самостоятельной работы обучающихся по заданию преподавателя, направленные на развитие навыков профессиональной лексики, закрепление практических профессиональных компетенций, поощрение инициатив.

Семинар – это форма учебной деятельности, проводимая в учебном заведении под руководством преподавателя, где студенты активно участвуют в обсуждениях, практических заданиях и других формах взаимодействия.

Для успешной подготовки к семинару рекомендуется заранее ознакомиться с темой занятия и основными материалами, чтобы иметь возможность активно участвовать в обсуждении. Также полезно подготовить вопросы и идеи для обсуждения, что поможет глубже понять материал и продемонстрировать заинтересованность.

Аудиторная работа – активная работа студента на семинаре, его ответы на вопросы преподавателя и участие в дискуссии.

Для успешного участия в семинаре студентам рекомендуется заранее ознакомиться с темой обсуждения, прочитать необходимые материалы и подготовить вопросы. Важно активно слушать и вовлекаться в дискуссию, высказывая свои мнения и аргументируя их. При ответах на вопросы преподавателя стоит быть уверенным, четким и логичным, опираясь на изученный материал. Также полезно поддерживать диалог с однокурсниками, чтобы обогатить обсуждение и расширить свои знания.

Проект – исследовательская работа по курсу и презентация результатов.

Для успешной подготовки к проекту: четко определите цели и задачи проекта, распределите роли и обязанности между участниками, а также установите сроки выполнения каждой части работы. Регулярно проводите встречи для обсуждения прогресса и решения возникающих вопросов.

Кейс – практическая работа студентов над реальными или смоделированными

задачами, что позволяет студенту применять теоретические знания на практике.

Студент самостоятельно разрабатывает стратегию решения поставленной задачи, что способствует развитию навыков критического мышления и самостоятельного принятия решений. Такой подход помогает подготовить будущих специалистов к реальным вызовам в их профессиональной деятельности.

Коллоквиум – устные ответы на вопросы, список которых известен студенту заранее.

В процессе подготовки к коллоквиуму необходимо проанализировать учебные материалы, ознакомившись с лекциями, учебниками и дополнительными источниками, акцентируя внимание на ключевых темах. Рекомендуется создать структурированные конспекты, выделяя основные идеи, термины и формулы.

Самостоятельная работа – работа студентов, направленная на углубленное изучение отдельных тем и вопросов учебной дисциплины (модуля).

В процессе самостоятельной работы студенты взаимодействуют с рекомендованными материалами при минимальном участии преподавателя. Задачи студента включают работу с конспектами лекций (обработка текста), повторное изучение учебных материалов планов и тезисов ответов, изучение дополнительных тем, выполнение учебно-исследовательских заданий и другое.

Система оценивания результатов обучения по дисциплине (модулю)

Критерии получения уровня и оценивания сформированности компетенций по дисциплине (модулю) «Черные ящики и цифровые агенты: социология искусственного интеллекта»

Оценивание уровня учебных достижений, обучающихся по дисциплине (модулю), осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Промежуточная аттестация по дисциплине (модулю) осуществляется в форме ***зачета с оценкой***, при этом проводится оценка компетенций, сформированных по дисциплине.

Для оценивания текущего контроля успеваемости и промежуточной аттестации используется десятибалльная шкала оценивания, которая соотносится с традиционной пятибалльной шкалой следующим образом, если иное не указано в силлабусе дисциплины:

Десятибалльная оценка	**Пятибалльная оценка**	**Оценка за зачет**	**Общая характеристика результата обучения по дисциплине (модулю)**
10	Отлично	Зачтено	Студент полностью владеет знаниями, изложенными в рабочей программе, и глубоко осмысляет дисциплину. Он самостоятельно и логически последовательно отвечает на все вопросы, акцентируя внимание на наиболее важном. Умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя ключевые моменты и устанавливая
9	Отлично	Зачтено	
8	Отлично	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			причинно-следственные связи. Четко формулирует ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты дисциплины (модуля) с практическими задачами.
7	Хорошо	Зачтено	Студент обладает знаниями предмета почти в полном объеме рабочей программы и самостоятельно, логически последовательно и всесторонне отвечает на все вопросы, акцентируя внимание на наиболее значимых моментах. Он умеет анализировать, сравнивать, классифицировать, обобщать, конкретизировать и систематизировать изученный материал, выделяя его ключевые аспекты и устанавливая причинно-следственные связи. Формулирует свои ответы, уверенно интерпретирует результаты анализов и других исследований, а также решает сложные ситуационные задачи. Студент хорошо знаком с методами исследования, необходимыми для практической деятельности, и умеет связывать теоретические аспекты предмета с практическими задачами.
6	Хорошо	Зачтено	
5	Удовлетворительно	Зачтено	Студент обладает базовыми знаниями по дисциплине (модулю), но испытывает трудности при самостоятельных ответах и использует неточные формулировки. В ходе ответов он допускает ошибки, касающиеся сути вопросов. Студент способен
4	Удовлетворительно	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
3	Не сдан	Не зачтено	Студент не овладел обязательным минимумом знаний по предмету и не может ответить на вопросы, даже если преподаватель задает дополнительные наводящие вопросы.
2	Не сдан	Не зачтено	
1	Не сдан	Не зачтено	

Дисциплина (модуль) «Черные ящики и цифровые агенты: социология искусственного интеллекта» оценивается следующим образом:

Активность	Вес	Количество	Описание
Накопительная оценка			
Аудиторная работа	20%	15	Активное участие в семинарах: ответы на вопросы преподавателя и участие в дискуссии
Коллоквиум	25%	1	Устные ответы на вопросы, список которых известен студенту заранее
Защита кейса	25%	1	Анализ реальных ситуаций и разработка практических решений на основе теоретических знаний
Проект	30%	1	Проект - презентация результатов исследовательской работы по дисциплине (модулю)

Итоговая оценка по дисциплине (модулю) «Черные ящики и цифровые агенты: социология искусственного интеллекта» выставляется по накопительной оценке: «0,2 × аудиторная работа + 0,25 × коллоквиум + 0,25 × защита кейса + 0,3 × проект».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные вопросы для подготовки к семинарам

Семинар 1.

1. Почему называть ИИ «умным», «решающим» или «думающим» — это не просто метафора, а когнитивная ловушка, искажающая понимание его работы?
2. Как концепция актора в теории акторно-сетевого подхода (ANT) помогает описать ИИ не как имитацию человека, а как элемент асимметричной сети?
3. Приведите пример, когда антропоморфизация ИИ приводит к ошибкам в проектировании интерфейса или управлении ожиданиями пользователей.
4. Какие признаки поведения ИИ-агента свидетельствуют не о «интеллекте», а о калиброванной реакции на структуру данных и инфраструктуру?
5. В чём разница между «ситуативным действием» ИИ и человеческим решением, и как это влияет на ответственность за результат?

Семинар 2.

1. Что входит в понятие «вольер для агента»: какие границы, ресурсы и правила необходимы для устойчивой работы ИИ-агента?

2. Как двойной интерфейс (для человека и для машины) влияет на проектирование взаимодействия между пользователем и ИИ-агентом?

3. Почему важно выделять фоновый режим работы ИИ и как он отличается от интерактивного?

4. Какие инфраструктурные компоненты (API, файлы, права доступа) следует учитывать при создании «пространства» для цифрового агента?

5. Приведите пример провального внедрения ИИ из-за непродуманного «вольера»: каких элементов не хватало, и к чему это привело?

Семинар 3.

1. Что такое «чёрный ящик» в социологии ИИ, и какие методы позволяют его «вскрыть» без доступа к исходному коду?

2. Как по логам, скриншотам или распечаткам сессии можно восстановить цикл «запрос → рассуждение → использование инструмента → действие»?

3. Как отличить запланированное поведение ИИ от ситуативного, вызванного сбоем или неожиданным вводом?

4. Какие нечеловеческие акторы (кроме самого ИИ) участвуют в типичной сессии агента, и как их можно идентифицировать по следам?

5. Почему прозрачность алгоритма не всегда повышает доверие к ИИ, и какие альтернативы «открытому коду» существуют для построения калиброванного доверия?

Примерное описание формата и критерии оценивания к защите кейсов

Тема занятия: решение и защита кейса в формате ролевой игры.

Описание задания: вы — междисциплинарная команда консультантов, приглашённая в крупную компанию для разработки и обоснования внедрения ИИ-агента в работу юридического отдела. Отдел ежедневно обрабатывает сотни запросов от сотрудников: по трудовому кодексу, внутренним регламентам, договорным шаблонам. Сейчас вся нагрузка ложится на трёх юристов, что приводит к задержкам и усталости.

Ваша задача — предложить цифрового агента, который будет работать в фоновом режиме, взаимодействовать с базой знаний, корпоративными API и людьми, но не имитировать человека. Вам нужно:

Спроектировать «вольер» для агента: определить его границы, доступы, инструменты (API, файлы), двойной интерфейс (для юристов и для системы).

Провести де-антропоморфизацию: убрать из интерфейса и поведения все признаки «человечности» (имя, эмодзи, фразы вроде «я думаю», «сейчас подумаю»).

Описать агента как актора в сети: выделить не менее 4 нечеловеческих акторов, участвующих в его работе (например: база знаний, шлюз аутентификации, система логирования).

Оценить выгоду по 4 уровням:

– операционный (сэкономленное время),

– организационный (изменение роли юристов),

– культурный (новые ожидания от автоматизации),

– властный (перераспределение полномочий, контроль за данными).

Подготовить короткую защиту (5–7 минут) в формате Pitch: представить решение перед «комитетом по внедрению» (преподаватель и студенты в ролях топ-менеджеров, ИТ-директора, юристов).

Критерии оценивания:

— Глубина социотехнического анализа: умение описать ИИ не как инструмент, а как актора в гетерогенной сети взаимодействий.

— Качество «вскрытия чёрного ящика» и проектирования вольера: чёткое определение границ, доступов, инфраструктуры, двойного интерфейса.

— Последовательность де-антропоморфизации: отказ от человеческих метафор, осознанное проектирование неантропоморфного поведения и интерфейса.

— Полнота оценки по 4 уровням: аргументированная интерпретация выгод на операционном, организационном, культурном и властном уровнях.

— Убедительность и структура защиты: логичность, ясность, способность отвечать на вопросы и аргументировать проектные решения в условиях ролевой игры.

Примерное описание задания и критерии оценивания к проекту

Проект: «Проектирование ИИ-агента для гибридного коллектива»

Цель проекта: сформировать у студента способность проектировать взаимодействие между людьми и искусственным интеллектом как социотехническую систему, в которой ИИ понимается не как «умная копия человека», а как нечеловеческий актор, требующий особой инфраструктуры, границ и режимов взаимодействия.

Описание задания: студенту (или команде) необходимо разработать концепт ИИ-агента для реальной или условной организации (например: отдел кадров, служба поддержки, научная лаборатория) и защитить его. Проект должен демонстрировать понимание ключевых концептов: антропоморфизация, «чёрный ящик», двойной интерфейс, вольер, калиброванное доверие и четырёхуровневая оценка эффекта.

Задачи и шаги выполнения:

1. Выберите контекст внедрения — опишите организацию, её задачи, болевые точки, где человек тратит время на рутину, подвержен ошибкам или перегрузке.
2. Проведите анализ «чёрного ящика» — на основе реальных или гипотетических логов, сессий или интерфейсов восстановите цикл «запрос → рассуждение → использование инструмента → действие» будущего агента.
3. Спроектируйте «вольер» для агента :

– определите его границы, доступы к API, файлам, базам данных;

– спроектируйте двойной интерфейс: один — для пользователей, другой — для администраторов/систем;

– обоснуйте режим работы: фоновый (тихий ИИ) или интерактивный.

4. Проведите де-антропоморфизацию:

– уберите из поведения и интерфейса все признаки человеческого присутствия (имя, эмодзи, фразы вроде «я думаю»);

– предложите альтернативные формы коммуникации (статусы, сигналы, метки).

5. Оцените эффект по 4 уровням:

– операционный (экономия времени),

– организационный (изменение ролей),

– культурный (новые нормы взаимодействия),

– властный (распределение контроля и ответственности).

6. Подготовьте защиту в формате Pitch (5–7 минут):

– представьте проблему,

– опишите агента как актора в сети,

– покажите «вольер» и интерфейсы,

– обоснуйте выгоду и доверие.

7. Оформите аналитическую записку (1–2 страницы) или слайд-дек (8–10 слайдов) с ключевыми аргументами.

Критерии оценивания:

• Глубина социотехнического анализа: умение описать ИИ-агента как элемент сети, включающей людей, API, протоколы, файлы и институциональные нормы.

• Качество проектирования «вольера» и двойного интерфейса: чёткость границ, обоснованность доступов, различие между машинным и человеческим интерфейсом.

• Последовательность де-антропоморфизации: отказ от человеческих метафор и осознанное проектирование неантропоморфного поведения.

• Полнота оценки по 4 уровням: аргументированная интерпретация последствий внедрения на операционном, организационном, культурном и властном уровнях.

• Убедительность и структура защиты: логичность повествования, способность отвечать на вопросы, аргументировать выбор, избегать технооптимизма.

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1.	Назовите социологический подход, рассматривающий не только людей, но и объекты (например, двери, ИИ) как действующие силы в сети.	Актерно-сетевой подход (ANT) / Бруно Латур	УК-1
2.	Укажите термин, описывающий ситуацию, когда внутреннее устройство системы игнорируется, а внимание фокусируется только на входе и выходе.	Чёрный ящик (Black Box)	УК-1
3.	Назовите ошибку восприятия, при которой пользователю приписывают ИИ человеческие качества (эмоции, намерения).	Антропоморфизм / Антропоморфная ловушка	УК-2
4.	Укажите логическое условие, отличающее нечеловеческого актора от человеческого в рамках симметричной антропологии.	Отсутствие интенциональности / Действие через посредников	ОПК-1
5.	Назовите термин, обозначающий ограниченное пространство с правилами и ресурсами, в котором действует ИИ-агент.	Вольер / Пространство для агента	УК-2
6.	Укажите принцип проектирования интерфейса, предполагающий разделение потоков информации для человека и для машины.	Двойной интерфейс / Два читателя	ПК-1
7.	Назовите режим работы ИИ, при котором он выполняет задачи автономно, без диалога с пользователем.	Тихий ИИ / Фоновый режим	УК-2
8.	Укажите структурный элемент инфраструктуры, необходимый агенту для совершения действия в цифровой среде.	API / Права доступа	ОПК-1
9.	Назовите цикл взаимодействия ИИ-агента, включающий запрос, рассуждение и использование инструмента.	Цикл агента / Запрос → Рассуждение → Действие	УК-1

10.	Укажите метод анализа поведения ИИ путем реконструкции его действий по логам и следам без доступа к коду.	Аутопсия агента / Вскрытие чёрного ящика	УК-1
11.	Назовите тип доверия к ИИ, основанный на предсказуемости и контексте, а не на полной прозрачности алгоритма.	Калиброванное доверие	УК-2
12.	Укажите этическую категорию, которая становится проблематичной при делегировании действий автономному агенту.	Ответственность / Подотчётность	ПК-1
13.	Назовите уровень эффекта от внедрения ИИ, связанный с изменением распределения полномочий в организации.	Властный уровень / Уровень власти	УК-2
14.	Укажите процесс устранения человеческих метафор из интерфейсов и поведения ИИ-систем.	Де-антропоморфизация	УК-2
15.	Назовите метод оценки выгоды от внедрения ИИ, включающий 4 уровня (технический, организационный, властный, культурный).	Модель 4 уровней эффекта	ПК-1
16.	Укажите логическую структуру организации, рассматриваемую как сеть людей, алгоритмов и протоколов.	Гетерогенная сеть / Социотехническая система	ОПК-1
17.	Назовите термин, описывающий коллектив, состоящий из людей и автоматизированных систем, работающих совместно.	Гибридный коллектив	УК-1
18.	Укажите уровень социального эффекта, на котором оцениваются изменения в корпоративной культуре и восприятии технологий.	Культурный уровень / Культурный эффект	УК-2
19.	Назовите задачу финального проекта курса, связанную с синтезом знаний о человеке и машине.	Проектирование гибридного коллектива / Защита проекта	ПК-1
20.	Укажите принцип системного подхода, применяемый для анализа взаимодействия человека и ИИ в организации.	Системный анализ / Холистический подход	ОПК-1