ЦЕНТРАЛЬНЫЙ УНИВЕРСИТЕТ

УТВЕРЖДЕНА

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

Рабочая программа дисциплины (модуля)
«Избранные темы исследований в ИИ»

Направление подготовки: *02.04.01 Математика и компьютерные науки*

Направленность (профиль) подготовки: *Продуктовый менеджмент*

Квалификация (степень) выпускника: *магистр*

Форма обучения: *очная*

Срок освоения программы: *2 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)

Рабочая программа дисциплины (модуля) «Избранные темы исследований в ИИ» составлена в соответствии с федеральным государственным образовательным стандартом высшего образования – магистратура по специальности 02.04.01 Математика и компьютерные науки, профиль Продуктовый менеджмент, утвержденный приказом Министерства науки и высшего образования Российской Федерации № 810 от 23.08.2017 года.

Изучение дисциплины (модуля) «Избранные темы исследований в ИИ» позволяет студентам быть в курсе последних достижений в области ИИ, что критически важно для их профессиональной подготовки и конкурентоспособности на рынке труда. Кроме того, оно способствует развитию критического мышления и инновационного подхода к решению сложных проблем, что является ключевым для успешной карьеры в быстро меняющемся технологическом мире.

Место дисциплины (модуля) в структуре образовательной программы

Настоящая дисциплина (модуль) включена в учебный план по программе подготовки магистратуры по направлению 02.04.01 Математика и компьютерные науки, профиль Продуктовый менеджмент и входит в вариативную часть Блока 1, формируемую участниками образовательных отношений, как дисциплина по выбору.

Дисциплина (модуль) изучается на 2 курсе в 3 семестре, доступна для прохождения при условии успешного завершения дисциплин (модулей) «Machine Learning (Машинное обучение)», «Natural Language Processing (Обработка естественного языка)» и общеуниверситетского факультатива «Deep Learning (Глубокое обучение)».

Цель изучения дисциплины (модуля): углубление знаний студентов о современных методах и технологиях искусственного интеллекта, развитие навыков их применения для решения актуальных задач.

Задачи изучения дисциплины (модуля):

- изучить теоретические основы формализации проблем в областях выравнивания ИИ, механистической интерпретируемости и мультимодальных больших языковых моделей;
- проанализировать текущие вызовы и будущие направления развития в указанных областях искусственного интеллекта;
- освоить методы реализации и валидации результатов исследований в зависимости от поставленных целей;
- разработать навыки программирования на Python для экспериментов с генеративными моделями ИИ, включая обучение и проверку;
- научиться критически оценивать алгоритмы и методы, выбирать их обоснованно и следить за новыми тенденциями в генеративном ИИ.

В результате освоения дисциплины (модуля) обучающийся должен:

знать:

- формализацию и математику современных методов в AI Alignment, Mechanistic Interpretability, Multimodal LLMs;
- какие проблемы существуют в областях AI Alignment, Mechanistic Interpretability, Multimodal LLMs сегодня, и к каким проблемам области идут в будущем;
- как валидировать полученные результаты в зависимости от целей исследований.

уметь:

- реализовывать методы AI Alignment, Mechanistic Interpretability, Multimodal LLMs;
- искать решения для реализации;

— писать код для экспериментов – как для обучения моделей, так и для их валидации.

владеть:

— языком программирования Python и навыком работы с библиотеками для генеративного AI;

— навыком критического анализа результатов и обоснования выбора алгоритмов и методов, используемых в проекте;

— навыком самообразования (следить за новыми исследованиями и тенденциями в области генеративного AI).

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-6.	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	УК-6.1.	Знает основные методы самооценки и анализа своей деятельности, а также принципы управления временем и целеполагания
		УК-6.2.	Умеет ставить реалистичные и достижимые цели, определять приоритеты в своей деятельности, а также разрабатывать и внедрять планы по совершенствованию своих навыков и компетенций на основе полученной самооценки
		УК-6.3.	Имеет практический опыт применения методов самооценки в своей профессиональной деятельности, включая участие в тренингах, семинарах и проектах, направленных на развитие личной эффективности и профессионального роста
ОПК-2.	Способен создавать и исследовать новые математические модели в естественных науках, совершенствовать и разрабатывать концепции, теории и методы	ОПК-2.1.	Знает основные математические модели и методы, используемые в естественных науках, включая статистическое моделирование, дифференциальные уравнения и численные методы, а также современные подходы к исследованию и анализу данных
		ОПК-2.2.	Умеет разрабатывать и адаптировать математические модели для решения конкретных проблем в естественных науках, проводить их анализ и верификацию, а также интерпретировать полученные результаты в контексте научных исследований
		ОПК-2.3.	Имеет практический опыт создания и исследования математических моделей в

			рамках научных проектов или исследований, включая участие в публикациях, конференциях или коллаборациях, где были разработаны и апробированы новые концепции и методы
ПК-3.	Способен решать задачи профессиональной деятельности в области продуктового менеджмента, формулировать результаты анализа и выявлять последствия полученных данных для принятия обоснованных решений и оптимизации продуктов	ПК-3.1.	Знает методы и инструменты продуктового менеджмента
		ПК-3.2.	Умеет применять аналитические инструменты и программное обеспечение для обработки и визуализации данных, а также формулировать выводы на основе проведенного анализа
		ПК-3.3.	Имеет опыт работы над реальными проектами в области продуктового менеджмента, включая анализ пользовательского поведения и оптимизацию продуктов на основе полученных данных
ПК-4.	Способен публично представлять собственные и известные научные результаты	ПК-4.1.	Знает основные принципы эффективного публичного выступления, методы визуализации данных и основные требования к научным презентациям, включая структуру и содержание
		ПК-4.2.	Умеет четко и логично формулировать свои научные результаты, адаптируя их для различных аудиторий, а также использовать визуальные средства для улучшения восприятия информации
		ПК-4.3.	Имеет практический опыт участия в научных конференциях, семинарах или других мероприятиях, где успешно представлял свои и известные научные результаты, получая обратную связь и взаимодействуя с аудиторией

3. Тематический план

№ п/п	Наименование раздела дисциплины (модуля)	Трудоемкость, академические часы			ТКУ (текущий контроль успеваемости)
		Очная форма			
		Аудиторная работа	Контроль	Самостоятельная работа	
		Семинары (практические занятия)			
1	RL for LLMs	10		51	Домашние задания, Коллоквиум
2	Mechanistic Interpretability	10		51	Домашние задания, Коллоквиум
3	Multimodal LLMs	10		52	Домашние задания, Коллоквиум
	Зачет		6		
	Итого:	***30***	***6***	***154***	
Объем дисциплины (модуля) (в ак. ч.)		***190***			
Объем дисциплины (модуля) (в зач. ед.)		***5***			

4. Содержание дисциплины (модуля)

№п/п	Наименование раздела дисциплины (модуля)	Содержание дисциплины (модуля) по темам
1	RL for LLMs	Основы RL-методов Интро в GRPO Advanced research beyond GRPO Агентность Q&A
2	Mechanistic Interpretability	Основы In-context Learning Анализ репрезентаций в трансформерах Каузальный анализ трансформера и circuit discovery Hyper Networks Q&A
3	Multimodal LLMs	Multimodal VLMs (basics) Agents and multi-step training World Models Vision-Language-Action models Q&A

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. Коул, А. Искусственный интеллект и компьютерное зрение. Реальные проекты на Python, Keras и TensorFlow : практическое руководство / А. Коул, С. Ганджу, М. Казам. - Санкт-Петербург : Питер, 2023. - 624 с. - (Бестселлеры O'Reilly). - ISBN 978-5-4461-1840-3. - Текст : электронный. - URL: https://znanium.com/catalog/product/2123884.

2. Мишра, П. Объяснимые модели искусственного интеллекта на Python. Модель искусственного интеллекта. Объяснения с использованием библиотек, расширений и фреймворков на основе языка Python : практическое руководство / П. Мишра ; пер. с англ. С. В. Минца. - Москва : ДМК Пресс, 2022. - 298 с. - ISBN 978-5-93700-124-5. - Текст : электронный. - URL: https://znanium.com/catalog/product/2109490.

Дополнительная литература:

1. Аймен Эль Амри, GPT-4. Руководство по использованию API Open AI : практическое руководство / Аймен Эль Амри ; пер. с англ. В. С. Яценкова. – Москва : ДМК Пресс, 2024. - 276 с. – ISBN 978-5-93700-299-0. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2205080.

2. Барретт, С. Ф. Arduino: искусственный интеллект и машинное обучение : практическое руководство / С. Ф. Барретт ; с англ. Ю. В. Ревича. – Москва : ДМК Пресс, 2024. - 244 с. – ISBN 978-5-93700-276-1. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2205065.

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

Электронный документ

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) «Избранные темы исследований в ИИ» в рамках текущего контроля успеваемости используются такие виды учебной работы, как семинары, домашние задания, коллоквиумы, 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.

Домашнее задание – набор задач по темам недели.

При работе над домашними заданиями важно внимательно ознакомиться с требованиями и сроками выполнения. Рекомендуется разбивать задания на этапы, чтобы избежать перегрузки и лучше усвоить материал. Использовать различные источники информации, включая учебники и онлайн-ресурсы, для более глубокого понимания темы.

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) «Избранные темы исследований в ИИ»

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***, 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:

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 (модуль). 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 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.
9	Отлично	Зачтено	
8	Отлично	Зачтено	
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
6	Хорошо	Зачтено	

Десятибалльная оценка	Пятибалльная оценка	Оценка за зачет	Общая характеристика результата обучения по дисциплине (модулю)
			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.
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 решать только самые простые задачи и владеет лишь минимальным набором методов исследования.
4	Удовлетворительно	Зачтено	
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) «Избранные темы исследований в ИИ» оценивается следующим образом:

Активность	Вес	Описание
Домашние задания	30%	Набор задач по темам недели
Коллоквиумы	70%	Устные ответы на вопросы, список которых известен студенту заранее

Формула расчёта итоговой оценки по дисциплине (модулю) «Избранные темы исследований в ИИ»: «0,3 × среднее за домашние задания + 0,7 × среднее за коллоквиумы».

Текущий контроль успеваемости обучающихся по дисциплине (модулю)

Примерные домашние задания

Домашнее задание 1.

Задание:

1. Кратко описать различия между SFT, RLHF и GRPO (цели, оптимизируемые функции, требования к данным).

2. Реализовать упрощённый RL-пайплайн (например, PPO или GRPO-подобную схему) для небольшой языковой модели или симулированной reward-модели.

3. Проанализировать влияние reward-функции на поведение модели (reward hacking, стабильность обучения).

4. Предложить расширение за пределами GRPO (например, multi-objective reward, self-play, RLAIF).

5. Кратко описать, как RL влияет на агентность LLM.

Домашнее задание 2.

Задание:

1. Провести анализ in-context learning на небольшой модели (например, сравнить поведение с разным числом примеров в prompt).
2. Исследовать внутренние представления:
   - извлечь hidden states;
   - визуализировать их (PCA/UMAP);
   - сравнить слои.
3. Провести простой каузальный анализ:
   - ablation отдельных attention-голов или нейронов;
   - оценить влияние на предсказание.
4. Описать концепцию circuit discovery.
5. Кратко объяснить роль hyper networks в модификации весов.

Домашнее задание 3.

Задание:

1. Описать архитектуру VLM (fusion-подход, cross-attention, alignment).
2. Реализовать простой мультимодальный пайплайн (например, CLIP-подобное выравнивание или VLM-инференс).
3. Объяснить multi-step training (pretraining → alignment → instruction tuning).
4. Рассмотреть концепцию World Models и их роль в планировании.
5. Описать Vision-Language-Action модель и пример её применения (роботика или embodied AI).

Примерные вопросы к коллоквиумам

Коллоквиум 1.

1. В чём различие между Supervised Fine-Tuning (SFT) и RLHF?
2. Почему SFT недостаточно для выравнивания модели под человеческие предпочтения?
3. Как формализуется задача RL для языковой модели (состояние, действие, награда)?
4. В чём различие между policy-based и value-based методами в контексте LLM?
5. Как работает PPO и почему он стал стандартом в RLHF?
6. Какие проблемы PPO проявляются при обучении больших языковых моделей?
7. В чём ключевая идея GRPO и какие ограничения PPO он пытается преодолеть?
8. Как строится reward model и какие источники данных для неё используются?
9. Какие риски связаны с reward hacking?
10. Что такое KL-регуляризация и зачем она нужна в RLHF?
11. Почему обучение RL для LLM может быть нестабильным?
12. Что такое off-policy и on-policy обучение, и что чаще используется в RLHF?
13. Как multi-objective reward влияет на поведение модели?
14. В чём различие между RLHF и RLAIF?
15. Как RL усиливает агентность модели?
16. Какие ограничения RL-подходов при обучении LLM существуют сегодня?
17. Как масштаб модели влияет на эффективность RL?
18. Какие существуют альтернативы RL для alignment?

Коллоквиум 2.

1. Что такое mechanistic interpretability и чем она отличается от post-hoc интерпретации?
2. Почему трансформеры демонстрируют in-context learning без обновления весов?
3. Какие гипотезы объясняют природу in-context learning?
4. Как распределяются функции между слоями трансформера?
5. Что такое attention head specialization?
6. Какие методы используются для анализа внутренних представлений (hidden states)?
7. Что показывает PCA/UMAP-анализ скрытых состояний?
8. Что такое probing-классификаторы и какие у них ограничения?
9. Что такое causal tracing?
10. В чём идея ablation-анализа?
11. Что такое circuit discovery в трансформерах?
12. Какие примеры обнаруженных circuits известны в литературе?
13. Как attention влияет на формирование логитов?
14. Что такое feature superposition?
15. Как polysemantic neurons усложняют интерпретацию?
16. Что такое activation patching?
17. Как можно выявить причинно-следственные связи внутри модели?
18. Что такое hyper networks и как они модифицируют веса основной модели?
19. Какие ограничения и риски есть у mechanistic interpretability?

Коллоквиум 3.

1. Что такое мультимодальная модель и чем она отличается от unimodal?
2. Какие существуют способы объединения модальностей (early fusion, late fusion, cross-attention)?
3. Как устроена архитектура CLIP?
4. Что такое alignment в мультимодальных моделях?
5. Почему contrastive learning эффективен для vision-language задач?
6. Как происходит обучение VLM на больших масштабах данных?
7. Что такое instruction tuning в мультимодальных моделях?
8. Какие особенности возникают при multi-step training?
9. Что такое hallucinations в VLM и почему они возникают?
10. Что такое world model в контексте агентных систем?
11. Чем world model отличается от обычного энкодера среды?
12. Как world models используются для планирования?
13. Что такое embodied AI?
14. Чем Vision-Language-Action (VLA) модели отличаются от VLM?
15. Какие компоненты включает архитектура VLA?
16. Какие данные требуются для обучения VLA-моделей?
17. Какие метрики применяются для оценки мультимодальных моделей?
18. Какие основные ограничения мультимодальных LLM существуют сегодня?
19. Какие направления развития мультимодальных и агентных моделей считаются наиболее перспективными?

Задания для промежуточной аттестации по дисциплине (модулю)

№ п/п	Задание	Ответ	Компетенция
1.	Укажите инструмент для оценки эффективности	чек-лист/журнал	УК-6

Электронный документ

	работы над мультимодальными LLMs.	прогресса	
2.	Укажите тип RL-метода в AI Alignment, требующего математического моделирования.	Q-learning/Policy Gradient	ОПК-2
3.	Укажите метрику для оценки эффекта мультимодальных LLMs в продуктовой аналитике.	точность/конверсия	ПК-3
4.	Укажите метод визуализации результатов в исследованиях ИИ.	диаграммы/графики	ПК-4
5.	Назовите критерий для самооценки в разборе кейсов AI Alignment.	соответствие стандартам/анализ ошибок	УК-6
6.	Назовите метод оценки интерпретируемости в Transformer Circuits.	sparse autoencoders/attention maps	ОПК-2
7.	Назовите способ формулирования результатов анализа данных в AI Alignment.	отчет/дашборд	ПК-3
8.	Назовите инструмент для публичного разбора кейсов Multimodal LLMs.	Jupyter Notebook/видео-конференция	ПК-4
9.	Укажите метод самооценки для определения приоритетов в исследованиях AI Alignment.	рефлексия/саморефлексия	УК-6
10.	Укажите концепцию валидации мультимодальных LLMs для новой математической модели.	кросс-валидация/бейзлайн	ОПК-2
11.	Укажите последствие выявленных данных для оптимизации продукта в Mechanistic Interpretability.	улучшение UX/снижение затрат	ПК-3
12.	Укажите способ презентации альтернативных Offline Alignment методов.	кейс-стади/видео	ПК-4
13.	Назовите способ совершенствования собственной деятельности в механической интерпретируемости.	обратная связь/итеративный подход	УК-6
14.	Назовите способ масштабирования вычислений для Few-Shot Learning.	облачные сервисы/распределенные вычисления	ОПК-2
15.	Назовите задачу в исследованиях ИИ для принятия обоснованных решений.	дизайн эксперимента/прогнозирование	ПК-3
16.	Назовите элемент разбора кейсов для демонстрации научных результатов в Mechanistic Interpretability.	слайды/демо	ПК-4
17.	Укажите этап в исследованиях ИИ для реализации приоритетов на основе самооценки.	планирование/корректировка плана	УК-6
18.	Укажите метод тестирования надежности Offline Alignment методов.	unit-тесты/интеграционные тесты	ОПК-2
19.	Укажите инструмент для анализа данных в кейсах Multimodal LLMs.	Python/Pandas	ПК-3
20.	Укажите формат публичного представления результатов исследований AI Alignment.	презентация/постер	ПК-4