

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Суворов Антон Дмитриевич
Должность: Ректор
Дата подписания: 13.06.2025 20:30:47
Уникальный программный ключ:
a39bdb15d680d5b065bfcad0af5c1efb14747dc0

СКОЛКОВСКИЙ ИНСТИТУТ НАУКИ И ТЕХНОЛОГИЙ (Сколтех)

Рабочая программа дисциплины	Современные методы и алгоритмы генеративного искусственного интеллекта
------------------------------	---

Преподаватель	Кузнецов Андрей
---------------	-----------------

Аннотация

Описание курса

В этом курсе мы обсудим современные подходы и проблемы в области генеративного ИИ. Во-первых, мы углубимся в фундаментальные модели и мультимодальность, которые включают в себя большие языковые модели в качестве базового представления знаний и различные подходы к связыванию модальностей. Далее мы познакомимся в мультимодальными архитектурами, такими как Fromage, Kosmos, LLaVA, OmniFusion, Visual Mamba и т. д., чтобы получить полное представление о том, как работают современные модели. Говоря о модальностях, мы будем рассматривать изображения, видео, аудио как основные нетекстовые источники данных в наши дни. Затем мы перейдем к генеративным моделям для синтеза и редактирования изображений, видео и 3D. Сюда войдут модели на базе трансформеров (DALL-E, eDiff-I) и диффузии (Imagen, Stable Diffusion, Kandinsky и т.д.), а также современные архитектуры диффузионных трансформеров (Stable Diffusion 3, Snap Video и др. В целом вы получите исчерпывающие знания о текущем состоянии GenAI и будете готовы углубиться в любую из изученных областей генеративного ИИ.

1. Основная информация

Академический уровень курса	Магистратура
Количество кредитов	3

Предварительные требования к курсу / рекомендации

Студенты должны быть знакомы с архитектурой transformer. Было бы полезно использовать диффузию для генерации данных.

Тип оценки - дифференцированная

Отображение оценок в процентах

A:	86
B:	76
C:	66

D:	56
E:	46
F:	0

2. Содержание курса

Тема	Краткое содержание	Лекции (час)	Семинары (час)	Лабораторные занятия (час)	Самостоятельная работа (час)
1. Генеративный искусственный интеллект	Введение в генеративный искусственный интеллект. Существующие области исследований. Направления исследований	1,5	1,5		6
2. Большие языковые модели и приложения	Трансформеры и механизм привлечения внимания. Существующие подходы и решения. Простые трансформеры, генерация с расширенным поиском, подход, основанный на участии экспертов	1,5	1,5		6
3. Базовые модели	Определение базовой модели. Языковая модель как базовая модель. Задачи и вызовы.	1,5	1,5		6
4. Мультимодальность Восприятие Часть I	Типы слияния: раннее, среднее и позднее. Модели визуального языка. Типы кодеров изображений и видео. Слияние кодеров и LLM. Обзор подходов, основанных на transformer, и современных архитектур. Задачи, вызовы и ориентиры.	1,5	1,5		6
5. Мультимодальность Восприятие Часть II	Модели пространства состояний. плюсы и минусы.	1,5	1,5		6
6. Длительный контекстный анализ	Способы восприятия видео: от коротких видеороликов до потокового понимания.	1,5	1,5		6
7. Приложения, специфичные для домена	Мультимодальность в робототехнике, самоуправляемых автомобилях, медицине и других областях исследований и промышленности.	1,5	1,5		6
8. Создание мультимедийного контента	Модели генерации изображений на основе трансформаторов. Подходы DALL-E, Parti и другие.	1,5	1,5		6
9. Модели, основанные на диффузии	Диффузия для генерации контента. Генерация изображений с использованием диффузии. Существующие модели и подходы: Стабильны Диффузия, Kandinsky 2.1, 3.0 и т.д. Диффузионные	1,5	1,5		6

	преобразователи, методы согласования потоков для генерации видео.				
--	---	--	--	--	--

3. Результаты обучения

Результаты обучения в Сколтехе указаны в соответствии со структурой результатов обучения в Сколтехе

1. ФУНДАМЕНТАЛЬНЫЕ ЗНАНИЯ
 - 1.2. Знание прикладной науки и техники, науки, в том числе современные методы и инструменты
 - 2.1. ПОЗНАНИЕ И СПОСОБЫ РАССУЖДЕНИЯ
 - 2.1.1. Аналитическое мышление и решение проблем
 - 2.1.2. Системное мышление
 - 2.1.3. Творческое мышление
 - 3.1. КОММУНИКАЦИЯ В МЕЖДУНАРОДНОЙ СРЕДЕ
 - 3.1.3. Устная презентация и обсуждение
 - 3.1.4. Вопросы, слушание и диалог
 - 3.1.5. Общение на английском языке в научной, деловой и общественной среде
 - 3.2. КОМАНДНАЯ РАБОТА И ЛИДЕРСТВО
 - 3.2.1. Формирование эффективных команд
 - 4.2. ДАЛЬНОВИДНОСТЬ — ИЗОБРЕТЕНИЕ НОВЫХ ТЕХНОЛОГИЙ ПОСРЕДСТВОМ ИССЛЕДОВАНИЙ
 - 4.2.1. Процесс исследования — гипотеза, доказательства и защита
 - 4.2.3. Исследования, направленные на разработку новых технологий
 - 4.2.4. Представление о пользе новой науки и техники

4. Задания и выставление оценок

Требование к физической посещаемости (% от числа занятий)	80
---	----

К требованиям о личном посещении:

студентам разрешается пропускать занятия только по уважительной причине, имеющей обоснование, например, медицинское заключение

Тип назначения	Краткое содержание задания	% от итоговой оценки за курс
Домашние задания	Домашнее задание 1 — Задание по точной настройке мультимодальной модели. Домашнее задание 2 — Задание по преобразованию текста в изображение/Тонкая настройка модели из текста в видео	50
Финальный проект	В ходе курса студент должен представить презентацию и отчет о проекте по генеративному ИИ	50

5. Критерии оценки

<u>Задание 1 Типа</u>	Домашние задания
------------------------------	------------------

Пример задания 1

Имея предварительно подготовленную модель визуального языка, разработайте ее подходящую, точно настроенную версию, чтобы улучшить навыки модели в распознавании рукописного текста.

Критерии оценки для задания 1

- 1) учащийся должен предоставить полную информацию о процессе разработки модели, начиная со сбора и обработки набора данных, заканчивая методом точной настройки для задачи и, наконец, предоставлением выбранного значения показателя
- 2) исходный код модели не должен быть простым форком решения с открытым исходным кодом, а должен представлять собой отдельное репозиторий на GitHub, организованное в хорошо структурированном виде
- 3) задание должно сопровождаться презентацией метода и всеми деталями его реализации

<u>Задание 2 Типа</u>	Финальный проект
------------------------------	------------------

Пример задания 2

Разработайте проектное предложение по использованию мультимодальных моделей для анализа данных дистанционного зондирования земли

- 1) идея реализации проекта должна быть реалистичной и простой для понимания
- 2) в качестве базовых следует использовать архитектуры SoTA, а выбор должен быть объяснен
- 3) должны быть описаны обучающие данные или должна быть инструкция по сбору данных
- 4) преимуществом будет реализованный PoC проекта на github
- 5) проектное предложение должно включать описание всех этапов исследования для реализации (документ и/или презентацию).
- 6) команда по разработке проекта может включать от 1 до 3 человек

6. Учебники и интернет-ресурсы

Необходимые учебники	ISBN-13 (or ISBN-10)
Mastering PyTorch: Create and deploy deep learning models from CNNs to multimodal models, LLMs, and beyond, Ashish Ranjan Jha, 2024	978-1801074308
Generative Artificial Intelligence. Exploring the Power and Potential of Generative AI, Shivam Solanki, Drupad Khublani, 2024	979-8868804038

Документы	DOI or URL
Y. Jin, et al. Efficient Multimodal Large Language Models: A Survey, 2024.	https://arxiv.org/abs/2405.10739v2
J. Cho, et al. Sora as an AGI World Model? A Complete Survey on Text-to-Video Generation, 2024	https://arxiv.org/abs/2403.05131v2
C. Zhang, et al. Text-to-image Diffusion Models in Generative AI: A Survey, 2023	https://arxiv.org/abs/2303.07909v2

Веб-ресурсы (ссылки)	Описание
https://github.com/BradyFU/Awesome-Multimodal-Large-Language-Models	
https://github.com/jianzhnie/awesometext-to-video	
https://github.com/Yutong-Zhoucv/Awesome-Text-to-Image	

7. Оборудование

Программное обеспечение
Google Colab