

АННОТАЦИЯ рабочей программы дисциплины

Б1.В.ДВ.02.02 СОСТЯЗАТЕЛЬНЫЕ И АВТОЭНКODЕРНЫЕ МОДЕЛИ

Объем трудоемкости: 4 зач. ед. (144 часа)

Цель дисциплины:

Цель дисциплины – сформировать у студентов систематизированные знания и практические навыки в области проектирования, разработки, адаптации, оптимизации и промышленного внедрения генеративных состязательных и автоэнкодерных моделей для работы с изображениями, видео и мультимодальными данными с упором на контролируруемую генерацию изображений и видео.

Задачи дисциплины:

- Изучить математические основы, архитектуры и алгоритмы обучения современных состязательных и автоэнкодерных моделей (VAE, GAN, Diffusion Models, Autoregressive Models).
- Освоить инструментарий для работы с генеративными моделями (PyTorch, Hugging Face, Diffusers, ComfyUI) и современные техники их адаптации (Fine-Tuning, LoRA, QLoRA, Дистилляция) и оптимизации (Прунинг, Квантование), ControlNet.
- Сформировать навыки полного цикла разработки: от анализа требований и подготовки данных до обучения, оценки, оптимизации и развертывания генеративных моделей в составе информационных систем.
- Развить способность к критическому анализу научных статей, постановке и проведению экспериментов, а также к генерации идей для модификации и создания новых архитектур генеративных сетей.

Место дисциплины в структуре образовательной программы:

Дисциплина «Состязательные и автоэнкодерные модели» относится к дисциплинам по выбору, код Б1.В.ДВ.02.02.

Дисциплина в значительной степени **взаимодействует для формирования компетенций** с дисциплинами:

- Нейросетевые технологии;
- Обработка естественного языка;
- Современные методы компьютерного зрения;
- Подготовка данных машинного обучения.

Требованием к «входным» знаниям является понимание основ машинного обучения, программирования на Python, администрирования Linux.

Требования к уровню освоения дисциплины

Изучение данной учебной дисциплины направлено на формирование у обучающихся следующих компетенций:

DL-1 (Э)

Способен применять и (или) разрабатывать архитектуры глубоких нейронных сетей

DL-1.6 Способен разрабатывать, адаптировать и внедрять генеративные нейронные сети для решения практических задач, включая создание новых архитектур, оптимизацию обучения и промышленное развертывание моделей

Разрабатывает новые архитектуры генеративных сетей, адаптивно применяет архитектуру VAE+GAN; разрабатывает капсульные сети

DL-1.7 Способен разрабатывать, оптимизировать и применять автоэнкодеры (AE) и вариационные автоэнкодеры (VAE) для решения задач снижения размерности, генерации данных и обнаружения

Применяет математические основы формирования пространства скрытых эмбедингов; знает вероятностный характер и отличия естественного и искусственного генеративного процессов; Знает математические основы функционирования вероятностного

аномалий, включая создание архитектур, обучение моделей и их внедрение в продуктивную среду
DL-1.11 Способен применять, адаптировать и разрабатывать методы сжатия нейронных сетей для оптимизации производительности моделей, включая квантование, прунинг, дистилляцию и другие техники, с учетом требований к качеству и вычислительной эффективности.

DL-1.12 Способен применять, адаптировать и разрабатывать методы дообучения нейронных сетей для эффективной адаптации моделей к новым задачам и доменам.

DL-2 (Э)

Способен применять и (или) разрабатывать современные архитектуры генеративных глубоких сетей

DL-2.1 Применяет известные архитектуры генеративных глубоких нейронных сетей для решения прикладной задачи (генерация текста, генерация изображений по тексту, синтез речи и т.д.), при необходимости проводя дообучение на наборах данных

LLM-1 (П)

Способен применять и (или) разрабатывать генеративные модели и БЯМ

LLM-1.1 Знает архитектуры генеративных моделей

LLM-1.4 Понимает принципы генерации в мультимодальных моделях

LLM-2 (П)

Способен дообучать, адаптировать и оптимизировать генеративные модели под специфические задачи и условия применения

LLM-2.1 Понимает принципы fine-tune

FC-1 (Б)

Способен проводить фронтальные исследования в области архитектур, алгоритмов МО, оптимизации и математики

FC-1.2 Разрабатывает новые архитектуры глубоких нейросетей

FC-2 (Б)

Способен проводить фронтальные исследования в области фундаментальных и генеративных моделей

FC-2.3 Исследует и создает мульти-

автокодировщика; обосновывает применение дивергенции Кульбака-Лейблера через основное тождество автокодировщиков
Владеет аппаратом структурированного и неструктурированного прунинга, знает стратегии прореживания. Разрабатывает новые методы сжатия.

Владеет продвинутыми техниками (adapter layers, LoRA, prefix-tuning). Комбинирует различные стратегии адаптации. Работает с малыми датасетами (few-shot learning)

Модифицирует архитектуры под специфические требования. Разрабатывает гибридные подходы (например, диффузионные модели + GAN). Оптимизирует архитектуры для целевых аппаратных платформ. Разрабатывает новые методы дообучения для генеративных моделей. Применяет few-shot/zero-shot learning техники. Реализует reinforcement learning для генерации.

Сравнивает архитектуры и выбирает подходящую под задачу

Использует мультимодальные модели для captioning и tagging

Применяет fine-tune к предобученным моделям на новых датасетах

Знает передовые архитектуры в основных триадах: архитектура-данные-задача, принципы их построения, сильные и слабые стороны. Знает особенности наиболее часто встречающихся вычислителей, умеет подбирать архитектуры, адекватные особенностям вычислительных устройств.

Дообучает готовые мультимодальные модели

модальные большие языковые модели (LLM)

(Flamingo, LLaVA). Строит пайплайны согласования данных разных модальностей. Владеет техниками базового выравнивания модальностей через CLIP-подобные энкодеры. Оценивает качество через стандартные метрики (cross-modal retrieval accuracy)

Содержание дисциплины:

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа СРС
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Введение и основы	27,8	8		8	11,8
2.	Современные архитектуры и мультимодальность	42	14		14	14
3.	Оптимизация, адаптация и инженерия	36	12		12	12
ИТОГО по разделам дисциплины		103,8	34		34	35,8
Контроль самостоятельной работы (КСР)		4				
Промежуточная аттестация (ИКР)		0,2				
Подготовка к текущему контролю						
Общая трудоемкость по дисциплине		108				

Курсовые работы: учебным планом не предусмотрены.

Форма проведения аттестации по дисциплине: экзамен.

Автор

С. Г. Сеница, доцент КИТ, к.т.н.