

Аннотация к рабочей программы дисциплины
Б1.В.07 «Обработка естественного языка»

Объем трудоемкости: 4 зачетных единиц

Цель дисциплины:

Формирование у студентов базовых знаний и практических навыков в области обработки естественного языка (NLP) с акцентом на применение глубокого обучения, включая архитектуры LLM и трансферное обучение.

Задачи дисциплины

Изучить математические основы NLP (векторизация, вероятностные модели).

Освоить архитектуры нейросетей для NLP (RNN, Transformer, BERT, GPT).

Научиться применять предобученные модели (Hugging Face) для классификации, генерации и анализа текста.

Разрабатывать и оптимизировать NLP-пайплайны (дообучение, RAG, развертывание).

Место дисциплины в структуре образовательной программы

Дисциплина «Обработка естественного языка» относится к «Часть, формируемая участниками образовательных отношений» Блока 1 «Дисциплины (модули)» учебного плана.

Дисциплина «Обработка естественного языка» в качестве прerreквизитов требует освоения «Программирования» (базовые навыки кодирования), «Алгоритмов и структур данных» (эффективные структуры для работы с текстами, очереди, деревья), «Объектно-ориентированного программирования и шаблонов проектирования» (архитектура NLP-пайплайнов), «Баз данных» (извлечение и хранение текстовых корпусов), «WEB-разработки» (работа с API и парсинг текстов из сети), а также «Численных методов» (понимание оптимизации и векторных операций). Постреквизитами выступают дисциплины, где NLP применяется как инструмент: «Генеративные нейронные сети» (языковые модели, трансформеры, генерация текста), «Технологии обработки больших данных» (масштабирование NLP-пайплайнов на большие корпуса) и «Мультиагентные системы» (лингвистический анализ коммуникации агентов). Смежными дисциплинами, изучаемыми параллельно или логически близко, являются «Нейросетевые технологии» (архитектуры для текстовых данных), «Современные методы компьютерного зрения» (мультимодальные задачи, текст + изображение), «Технологии управления данными NoSQL» (хранение неструктурированных текстов и графов знаний), а также «Численные методы и цифровая обработка сигналов» (векторизация текста и методы сжатия).

Требования к уровню освоения дисциплины

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

DL-1	<i>Способен применять и (или) разрабатывать архитектуры глубоких нейронных сетей</i>
DL-1.1	Объясняет и применяет математические основы нейронных сетей, включая расчет градиентов, методы оптимизации и алгоритм обратного распространения ошибки для эффективного обучения моделей.
	Знать: математические основы нейронных сетей, включая расчет градиентов, методы оптимизации и алгоритм обратного распространения ошибки Уметь: модифицировать популярные архитектуры графовых нейронных сетей (GCN, GraphSAGE, GIN); применяет графовые сети внимания (GAT).

	Адаптирует архитектуры под специфические типы графов (гетерогенные, динамические). Разрабатывает кастомные механизмы агрегации. Владеть: методами расчета градиентов и оптимизации нейронных сетей
DL-1.3	Применяет современные архитектуры глубоких сетей для решения различных задач, понимая их внутреннюю структуру и особенности обучения.
	Знать: современные архитектуры глубоких сетей и их внутреннюю структуру Уметь: применять глубокие сети для решения различных задач. Разрабатывает кастомные архитектуры для специфических комбинаций модальностей. Оптимизирует совместное обучение нескольких задач (multi-task learning). Владеть: методами обучения современных архитектур глубоких сетей. Применяет продвинутое техники (cross-modal distillation, contrastive learning). Создает эффективные пайплайны обработки мультимодальных данных
DL-1.5	Разрабатывает и оптимизирует специализированные архитектуры для работы с последовательностями, учитывая их уникальные свойства.
	Знать: архитектуры генеративных нейронных сетей. Понимает и применяет специфические архитектурные варианты для задач (например, Encoder-Decoder для перевода/суммаризации, Decoder-only для генерации). Уметь: разрабатывать и адаптировать генеративные сети для практических задач. Использует техники для работы с длинными последовательностями: эффективные схемы внимания (Linformer, Performer, Longformer, Sparse Transformers). Проводит детальный анализ ошибок модели, понимает типичные проблемы (bias в данных/модели, переобучение) и применяет методы борьбы с ними (регуляризация, аугментация данных) Владеть: промышленным развертыванием генеративных моделей
DL-2	<i>Способен применять и (или) разрабатывать современные архитектуры генеративных глубоких сетей</i>
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
	Знать: известные архитектуры генеративных глубоких нейронных сетей Уметь: применять генеративные сети для решения прикладных задач. Настраивает параметры генерации под конкретную задачу. Реализует техники контролируемой генерации (guidance scale, prompt engineering). Комбинирует несколько моделей в пайплайн. Проводит domain adaptation для специфических данных. Использует продвинутое методы эффективного обучения (QLoRA, DreamBooth). Оптимизирует процесс обучения (подбор lr, батчей). Владеть: методами дообучения генеративных моделей на наборах данных. 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.
DL-2.2	Имплементирует известные архитектуры генеративных сетей, реализует пайплайны их обучения на датасетах и вывод генеративных моделей в продуктивную среду
	Знать: архитектуры генеративных сетей и пайплайны их обучения Уметь: имплементировать архитектуры генеративных сетей. Разрабатывает кастомные реализации генеративных архитектур. Создает high-load

	генеративные сервисы. Разрабатывает системы кеширования генераций. Реализует сложные А/В тестирования. Написание кастом CUDA-ядер. Владеть: вывода генеративных моделей в продуктивную среду
DL-4	<i>Способен применять и (или) разрабатывать алгоритмы, методы и технологии обработки естественного языка</i>
DL-4.1	Применяет (проводя выбор и эксперименты) известные алгоритмы и библиотеки для обработки естественного языка, предобученные глубокие нейросетевые модели для прикладных задач анализа текстов, при необходимости дообучая и валидируя на собственных наборах данных
	Знать основные типы прикладных задач анализа текстов (классификация, извлечение сущностей, суммаризация, проверка тональности) и соответствующие им виды предобученных глубоких нейросетевых моделей (BERT, GPT, T5 и их легковесные версии). Уметь проводить простой выбор между несколькими библиотеками (Hugging Face Transformers, spaCy, NLTK) и предобученными моделями под фиксированную задачу на небольшом тестовом корпусе. Владеть навыком запуска дообучения (fine-tuning) предобученной модели на собственном размеченном наборе данных с последующей базовой валидацией по точности (accuracy) или F1-мере.
DL-4.2	Определяет стек технологий, методов и алгоритмов для построения продуктов с обработкой естественного языка (диалоговые системы, вопросно-ответные системы, рекомендательные системы и т.д.)
	Знать: основные архитектуры сетей, использующиеся для векторизации текстовых данных: Word2Vec, Doc2Vec, Glove, FastText, рекуррентные нейронные сети и сети-трансформеры (энкодеры). Уметь: применять простейшие классические модели для извлечения полезных признаков из текстов: мешок слов, TF/IDF, латентный семантический анализ. Владеть: самостоятельно находит подходящую модель для векторизации текстовых данных в открытых источниках и применить её для практической задачи.
DL-4.3	Имплементирует известные алгоритмы, архитектуры и модели обработки естественного языка на реальных данных, строит пайплайны обучения моделей и развертывания NLP-сервисов в продуктивной среде
	Знать: классические алгоритмы тематического моделирования и способен применить их в практической задаче: LSA, pLSA, BigARTM. Уметь: обучать модель классификации текстовых данных с использованием векторизованного представления текста - разреженного в случае применения классических статистических алгоритмов ML для классификации и плотного для применения в нейросетевых моделях) Владеть: архитектурой и применяет рекуррентные нейронные сети и сети типа "трансформер" для downstream задач, требующих языкового моделирования (контекстуализированные векторные представления, генерация текстов, аннотирование и реферирование текстов и т.п.)
DL-4.4	Разрабатывает новые алгоритмы и библиотеки обработки естественного языка, новые архитектуры глубоких нейронных сетей и методы их обучения для задач анализа текста
	Знать: архитектуру рекуррентных нейронных сетей и сетей типа "трансформер" для downstream задач Уметь: применяет рекуррентные нейронные сети и сети типа "трансформер" для downstream задач, требующих языкового моделирования

	(контекстуализированные векторные представления, генерация текстов, аннотирование и реферирование текстов и т.п.) Владеть: Использует предобученные модели без модификации архитектуры.
LLM-1	Способен применять и (или) разрабатывать генеративные модели и БЯМ
LLM-1.1	Знает архитектуры генеративных моделей.
	Знать: архитектуры генеративных моделей Уметь: различать архитектуры генеративных моделей. Сравнивает архитектуры и выбирает подходящую под задачу Владеть: знаниями об архитектурах генеративных моделей
FC-2	Способен проводить передовые исследования в области фундаментальных и генеративных моделей
FC-2.3	Исследует и создает мульти-модальные большие языковые модели (LLM)
	Знать: архитектуры мульти-модальных больших языковых моделей Уметь: исследовать и создавать мульти-модальные LLM Владеть: созданием мульти-модальных языковых моделей

Содержание дисциплины:

Распределение видов учебной работы и их трудоемкости по разделам дисциплины.

Разделы (темы) дисциплины, изучаемые в 6 семестре

№	Наименование разделов (тем)	Количество часов				
		Всего	Аудиторная работа			Внеаудиторная работа СРС
			Л	ПЗ	ЛР	
1	2	3	4	5	6	7
1.	Введение в NLP	5	2		2	1
2.	Лингвистические основы NLP	5	2		2	1
3.	Предобработка текста	6	2		2	2
4.	Векторизация текста: TF-IDF, Word2Vec	8	2		4	2
5.	Рекуррентные сети (RNN, LSTM)	8	2		4	2
6.	Механизм внимания и Transformer	8	2		4	2
7.	BERT и трансферное обучение	7	2		4	1
8.	Генеративные модели (GPT, T5)	8	2		4	2
9.	Оценка качества моделей (BLEU, ROUGE)	6	2		2	2
10	Дообучение и адаптация LLM	7	2		4	1
11	RAG (Retrieval-Augmented Generation)	7	2		4	1
12	Мультимодальные NLP-модели	8	2		4	2
13	Этика и bias в NLP	5	2		2	1
14	Промышленное внедрение NLP-моделей	5	2		2	1
15	Оптимизация NLP-моделей (квантование, дистилляция)	6	2		2	2
16	Итоговый обзор и перспективы NLP	5	2		2	1
ИТОГО по разделам дисциплины		104	32		48	24
Контроль самостоятельной работы (КСР)		4				
Промежуточная аттестация (ИКР)		0,3				
Подготовка к текущему контролю		35,7				
Общая трудоемкость по дисциплине		144				

Курсовые работы: не предусмотрена

Форма проведения аттестации по дисциплине: экзамен

Автор

Добровольская Наталья Юрьевна