

Обработка естественного языка — NLP

01	Кому подойдёт курс	Курс подойдёт тем, у кого есть опыт работы с данными, Machine Learning и Deep Learning Специалистам в области Data Science - Разберёте методы извлечения признаков из текста и сможете применять NLP для работы с большими данными и решения бизнес-задач DL- и ML-инженерам - Изучите способы обработки текстовых данных и современные архитектуры и будете использовать NLP в своих реальных проектах Разработчикам - Освоите актуальные библиотеки и инструменты, чтобы внедрить технологии NLP в разработку приложений, которые используют текстовые данные
02	Чему научитесь на курсе	- Понимать и применять современные NLP-модели: от BERT до LLM - Строить RAG и агентные системы - Обучать LLM с оптимизацией операций на GPU - Создавать модели генерации текста, перевода, NER и мультимодальные решения - Работать с поиском: от классического BM25 до векторного по эмбедингам
03	Как проходит курс	- Сопровождение кураторами - Обратная связь от опытных наставников - Воркшопы с экспертами - Теория на платформе Практикума - Практические задания с ревью на готовой инфраструктуре в облаке

Что вас ждёт на курсе

Ключевые NLP-архитектуры: от вариаций BERT до RAG и агентных систем	Фокус на актуальные задачи: NER, QA, VQA, машинный перевод и сценарии с RAG	Удостоверение о повышении квалификации
---	---	--

Обработка естественного
языка — NLP

2 месяца

продолжительность курса

4 проекта

в портфолио

2 ЧАСА

00

Бесплатная часть.
Нейросетевые
решения на практике

- Знакомство с курсом
- Введение в нейросети и их роль в решении сложных задач
- Работа с текстами
- Работа с изображениями
- Нейросети на практике

2 НЕДЕЛИ

01

Современные модели:
обучение и ключевые
механизмы

- Multi-Head Attention и BERT
- Эволюция моделей в NLP
- Эффективная тренировка моделей

2 НЕДЕЛИ

02

Большие языковые
модели

- Введение в большие языковые модели
- Архитектура и принципы работы LLM
- Оптимизации в LLM

2 НЕДЕЛИ

03

Путь генеративного
NLP: от Seq2Seq к RAG

- Архитектуры Seq2Seq в трансформерах
- Генерация текста для практических задач
- Retrieval-Augmented Generation — RAG

2 ЧАСА

04

Современный NLP:
поиск, агенты
и мультимодальность

- Полнотекстовый поиск
- Агенты
- Мультимодальные архитектуры

Бесплатная часть. Нейросетевые решения на практике

00

2 часа	Практическая работа
1 практическая работа	Научитесь использовать предобученные модели для анализа текста и изображений и интерпретировать результаты их работы
	Инструменты и технологии
	• PyTorch • CV • NLP

Содержание

01.	Знакомство с курсом	Узнаете, как организован курс: из чего состоит программа, как устроена платформа, какие активности вас ждут
02.	Введение в нейросети и их роль в решении сложных задач	Поймёте, в каких областях применяются нейросети и какие задачи они решают в реальной жизни
03.	Работа с текстами	Классифицируете тексты с использованием предобученной модели
04.	Работа с изображениями	Классифицируете изображения с помощью предобученной модели
05.	Нейросети на практике	Научитесь использовать нейросети в повседневной практике и для решения профессиональных задач

Современные модели: обучение и ключевые механизмы

01

2 недели 1 проект	Проект Решите задачу NER, выявив все сущности, и предскажете, сколько их в тексте, через регрессию по CLS Инструменты и технологии • RoBERTa • XLM-RoBERTa • DeBERTa • NER • PyTorch Lightning • DP • DDP • FSDP
----------------------	---

Содержание

01. Multi-Head Attention и BERT	Разберёте трансформер на уровне тензоров, attention-механизм, позиционные эмбединги и skip connections. Напишите полный encoder с нуля. Изучите архитектуру BERT, задачи MLM и NSP, CLS-токен, ограничения attention и такие решения, как flash/sparse attention.
02. Эволюция моделей в NLP	Сравните токенизаторы (BPE, WordPiece и другие) по стабильности и размеру словаря. Рассмотрите развитие моделей: RoBERTa, XLM-R, DeBERTa, e5, включая мультязычные и облегчённые версии для продакшна. Изучите NER: BIO-разметку, entity spans, лоссы.
03. Эффективная тренировка моделей	Освоите float16, bfloat16, mixed precision, включите AMP в PyTorch. Изучите квантизацию (PTQ, QAT) и распределённую тренировку (DataParallel, DDP, FSDP). Разберёте torch.compile и научитесь работать с PyTorch Lightning.

2 недели	Проект
1 проект	Дообучите языковую модель, чтобы управлять генерацией текстов на разных этапах — pretrain, SFT, alignment
	Инструменты и технологии
	- LoRA - QLoRa - SFT - TRL - vLLM - FlashAttention - Triton - ALiBi - RoPE

Содержание

01. Введение в большие языковые модели	Изучите фундаментальные концепции LLM. Рассмотрите архитектуры для генерации текста: decoder-only, encoder-decoder, диффузионные модели. Разберёте устройство attention-масок, их влияние на генерацию. Реализуете вызов API через OpenRouter. Исследуете ограничения LLM.
02. Архитектура и принципы работы LLM	Узнаете принципы обучения декодера. Реализуете цикл генерации текста с чат-шаблонами Jinja. Рассмотрите методы улучшения генерации: In-Context Learning, reasoning, CoT. Изучите стратегии генерации текста: beam search, sampling, параметры temperature, top-k, top-p, repetition penalty. Поймёте причины их выбора.
03. Оптимизации в LLM	Попробуете методы обучения и оптимизации LLM: LoRA, QLoRa, адаптеры, PEFT и Unsloth. Поймёте проблему длинного контекста. Освоите методы расширения контекста: KV cache, paged attention, speculative decoding, continuous batching. Проведёте оптимизацию на GPU, с использованием Triton, fused layers и FlashAttention. Примените gradient checkpointing и выберете стратегии обучения под ресурсы и скорость.

Путь генеративного NLP: от Seq2Seq к RAG

03

2 недели
1 проект

Проект
Разработаете retrieval-систему по статьям из arXiv, с поиском по документам и генерацией ответов на естественном языке

- Инструменты и технологии
- T5
 - LoRA
 - Seq2Seq
 - BLEU
 - ROUGE
 - chrF
 - COMET
 - RAG
 - Векторные базы данных
 - LangChain

Содержание

01. Архитектуры Seq2Seq в трансформерах	На примере T5 и русскоязычных аналогов разберёте архитектуру encoder-decoder. Поймёте принципы cross-attention и teacher forcing. Изучите метрики BLEU и ROUGE. Примените T5 к разным NLP-задачам. Освоите предобучение T5 с помощью span corruption. Рассмотрите ключевые бенчмарки.
02. Генерация текста для практических задач	Изучите модели машинного перевода mT5 и NLLB. Освоите их дообучение для доменов. Разберёте методы улучшения параллельных корпусов: выравнивание, фильтрацию, back-translation и paraphrasing. Решите NER-задачу в формате генерации с промптингом и constrained decoding. Дообучите Seq2Seq-модель и оцените качество перевода.
03. Retrieval-Augmented Generation — RAG	Узнаете причины галлюцинаций и как RAG снижает риск ошибок. Изучите bi-cross-encoder, гибридный поиск и векторные базы (FAISS, Chroma, Qdrant). Рассмотрите методы снижения размерности и индексации. Построите RAG-пайплайн: от подготовки документов до генерации ответа. Освоите LangChain и оценку качества поиска.

Современный NLP: поиск, агенты и мультимодальность

04

2 недели 1 проект	Проект Разработаете модель, которая будет искать изображения по описанию, проверять их на релевантность и отвечать на вопросы по содержанию Инструменты и технологии • rapidfuzz • OpenSearch • datasketch • LangChain • AutoGen • smolagents • MCP • CLIP • SigLIP • BLIP • LLaVA • Florence
----------------------	--

Содержание

01. Полнотекстовый поиск	Рассмотрите нечёткий поиск и алгоритм Левенштейна. Освойте rapidfuzz и BM25. Поработаете с индексами в OpenSearch. Разберёте расширенные техники поиска. Изучите LSH с datasketch. Сравните подходы через хэши и эмбединги. Реализуете пайплайны поиска с опечатками, а также индексацию и дедупликацию коллекций.
02. Агенты	Изучите function calling, structured output, а также интеграцию функций через LangChain. Разберёте агентные системы: ReAct, LLM-as-judge и современные фреймворки (AutoGen, smolagents), а также протокол MCP. Создадите агентов для работы с интернетом и веб-страницами. Рассмотрите архитектуры Perplexity и Deep Research.
03. Мультимодальные архитектуры	Узнаете, как создаются мультимодальные датасеты и бенчмарки. Разберёте CLIP и его варианты (SigLIP, ViT-L/14), а также модели BLIP, Florence и LLaVA для VQA. Научитесь выявлять ошибки аннотаций с помощью CLIP и применять его для поиска, очистки датасетов и VQA-задач.