



Академия АйТи
a Softline Company



Продвинутый NLP: LLM, RAG, мульти-modalность и агенты

Код курса: NLP-ADV

Продвинутый NLP: LLM, RAG, мультимодальность и агенты

Код курса: NLP-ADV

Длительность	120 ак. часов
Формат	
Разработчик курса	Академия АйТи
Тип	Учебный курс
Способ обучения	Под руководством тренера

О курсе

Продвинутый курс по обработке естественного языка, построенный на глубоком изучении современных трансформерных моделей и методов работы с большими языковыми моделями. Программа охватывает архитектурные оптимизации LLM, техники промптинга и адаптации (LoRA), семантический поиск, RAG-системы, мультимодальные модели, AI-агентов и работу с временными рядами. Курс предназначен для специалистов, желающих выйти на уровень промышленного применения NLP.

Подробная информация

Профиль аудитории:

- ML-инженеры и Data Scientists с опытом работы в NLP
- разработчики, прошедшие базовый курс NLP и желающие углубить знания
- специалисты по обработке текстов, работающие с трансформерными моделями
- исследователи в области искусственного интеллекта и NLP

Предварительные требования:

- уверенное владение Python и библиотеками для ML (PyTorch, Hugging Face)
- знание основ архитектуры Transformer, BERT, GPT
- опыт дообучения (fine-tuning) нейросетевых моделей
- базовое понимание RNN/LSTM, Word2Vec, механизма attention

По окончании курса слушатели смогут:

- работать с современными LLM (LLaMA, Mistral, DeepSeek) и понимать их архитектурные оптимизации (RoPE, Flash Attention, GQA)
- применять продвинутые техники промптинга и метод LoRA для доменной адаптации моделей
- создавать системы семантического поиска с использованием sentence-transformers и bi-encoder/cross-encoder архитектур
- оптимизировать модели: дистилляция, квантизация, прунинг; работать с vllm и ollama
- строить RAG-системы на LangChain для генерации на основе базы знаний

- применять мультимодальные и Vision-трансформеры (CLIP, LLaVA, QWEN), работать с временными рядами (TimeGPT)
- проектировать AI-агентов и интегрировать LLM в конечные продукты (телеграм-боты)

Программа курса

Модуль 1. Базовые понятия трансформерных моделей

- Цель модуля: Укрепить и систематизировать знания об архитектуре трансформеров, от классических энкодеров и декодеров до современных подходов, лежащих в основе ChatGPT.
- Тема 1: Задача языкового моделирования и понятие языковой модели
- Постановка задачи, обучение моделей, языковые модели на основе RNN/LSTM. Методы генерации текста (beam search, sampling, top-k sampling).
- Тема 2: Архитектура трансформер и задача машинного перевода
- Задача машинного перевода, seq2seq, механизм attention, архитектура Transformer. Домашнее задание: "Собираем трансформер своими руками".
- Тема 3: BERT и другие трансформерные энкодеры
- Transfer learning, pretraining + fine-tuning, стратегии токенизации, BERT, RoBERTa. Домашнее задание: "Дообучение энкодерных моделей".
- Тема 4: GPT и другие декодерные модели для генерации текста
- Стратегии генерации, архитектура GPT-3/4, подходы few-shot/zero-shot.
- Тема 5: Seq2seq модели. Дообучение трансформеров на практическом примере
- Сравнение архитектур, модель T5, практический пример дообучения. Домашнее задание: "Дообучение энкодер-декодерных моделей".
- Тема 6: Методы за ChatGPT + Q&A сессия
- Инструктивное обучение и RLHF, архитектура ChatGPT, вопросы alignment и безопасности.

Модуль 2. Новая эра LLM: базовые методы

- Цель модуля: Изучить современные LLM, их архитектурные оптимизации, продвинутые техники промптинга и адаптации.
- Тема 1: Современные LLM и оптимизация архитектуры
- Обзор LLM (LLaMA, Mistral), оптимизации: RoPE, Flash Attention, Grouped-Query Attention.
- Тема 2: Практическое занятие: теория промптинга LLM
- System Prompt, ChatTemplate, multi-turn. Домашнее задание: "Промптинг языковых моделей".
- Тема 3: Методы за моделью DeepSeek — на пути к современному ризонингу
- Модели рассуждений, обучение ризонингу, архитектура DeepSeek, GRPO.
- Тема 4: LoRA и доменная адаптация
- Метод LoRA, доменная адаптация, кейсы. Домашнее задание: "LoRA и доменная адаптация".
- Тема 5: Мультязычные трансформеры
- Проблемы мультязычности, популярные архитектуры, адаптация к новым языкам.
- Тема 6: Работа с длинным контекстом
- Модификации Attention для длинного контекста, бенчмарки.
- Тема 7: Q&A сессия и дискуссионный клуб

Модуль 3. Продвинутые методы работы с LLM

- Цель модуля: Освоить инструментарий для создания сложных NLP-решений: семантический

- поиск, оптимизация моделей, RAG.
- Тема 1: Sentence-transformers
- Эмбеддеры, bi-encoder/cross-encoder, Labse, примеры для русского языка. Домашнее задание: "Работа с sentence-transformers".
- Тема 2: Эффективные трансформеры
- Дистилляция, квантизация, прунинг; практика.
- Тема 3: Фреймворки для эффективной работы с LLM
- vllm, ollama; практика работы. Домашнее задание: "Работа с vllm".
- Тема 4: LangChain
- Библиотека LangChain, сегментация текста.
- Тема 5: RAG — генерация на основе базы знаний
- Понятие RAG, основные фреймворки и подходы, практика.

Модуль 4. Дополнительные главы работы с LLM

- Цель модуля: Прикладные аспекты: оценка, распределенное обучение и интеграция в продукты.
- Тема 1: Оценка LLM
- Бенчмарки, LLM-as-judge, MERA для русского языка, читерство в бенчмарках.
- Тема 2: Распределенное обучение
- Методы распределенного обучения языковых моделей.
- Тема 3: Интеграция LLM в телеграм-бота
- Создание и запуск LLM в тг-боте. Домашнее задание: "Создание тг-бота с LLM".
- Тема 4: Q&A сессия и дискуссионный клуб

Модуль 5. Трансформеры для других модальностей

- Цель модуля: Расширить знания за пределы текста: изображения, видео, временные ряды, табличные данные, AI-агенты.
- Тема 1: Мультимодальные и Vision трансформеры
- Трансформеры для изображений: CLIP, DALL-E, Kandinsky; практика с CLIP.
- Тема 2: Работа с мультимодальными моделями
- LLaVA, QWEN для видео, ASR через Whisper. Домашнее задание: "Работа с Vision моделями".
- Тема 3: Введение в AI-агентов
- Основные концепции AI-агентов, ролей и мультиагентных систем; модель Operator от OpenAI.
- Тема 4: Трансформеры для временных рядов
- Временные ряды, модель TimeGPT; практика.
- Тема 5: Трансформеры для табличных данных
- Адаптация архитектуры, сравнение с бустингом; практика.
- Тема 6: Q&A сессия и дискуссионный клуб
- Тема 7: Собеседование. Как подготовиться и что нужно знать
- Что спрашивают на собеседованиях по NLP.

[Посмотреть расписание курса и записаться на обучение](#)

Обращайтесь по любым вопросам

к менеджерам Академии АйТи

+7 (495) 150 96 00 | academy@academyit.ru