



Академия АйТи  
a Softline Company



## NLP / Natural Language Processing

Код курса: NLP-1

# NLP / Natural Language Processing

Код курса: NLP-1

<b>Длительность</b>	72 ак. часа
<b>Формат</b>	
<b>Разработчик курса</b>	Академия АйТи
<b>Тип</b>	Учебный курс
<b>Способ обучения</b>	Под руководством тренера

## О курсе

Фундаментальный курс по обработке естественного языка, охватывающий весь путь от классических методов до современных трансформерных моделей. Программа построена по принципу "от простого к сложному": от Python-инструментов для работы с текстом через основы глубокого обучения и рекуррентные сети к архитектуре Transformer, моделям BERT и GPT, промптингу LLM и созданию конечных продуктов на их основе.

## Подробная информация

Профиль аудитории:

- начинающие специалисты в области Data Science и машинного обучения
- Python-разработчики, желающие освоить NLP
- аналитики данных, расширяющие компетенции в области обработки текстов
- студенты технических специальностей, изучающие искусственный интеллект

Предварительные требования:

- базовые знания Python (переменные, циклы, функции, классы)
- знание основ линейной алгебры и теории вероятностей
- опыт работы с библиотеками pandas и numpy будет преимуществом
- базовое понимание концепций машинного обучения

По окончании курса слушатели смогут:

- выполнять предобработку текстовых данных: токенизация, лемматизация, работа с регулярными выражениями
- создавать датасеты с помощью парсинга и API
- строить и обучать нейронные сети на PyTorch, включая рекуррентные архитектуры (RNN, LSTM, GRU)
- применять классические методы NLP: Word2Vec, FastText, seq2seq с attention
- работать с трансформерными моделями: BERT, GPT, T5 для задач классификации, NER, генерации текста

- применять техники промптинга LLM: Zero-shot, Few-shot, Chain-of-Thought, ReAct
- оценивать качество языковых моделей и создавать телеграм-ботов на их основе

## Программа курса

### Модуль 1. Python для работы с текстами

- Цель модуля: Заложить фундамент для работы с текстовыми данными с помощью Python и освоить базовые инструменты для сбора и предобработки информации.
- Тема 1: Введение в NLP
- Область и задачи NLP; современные достижения.
- Тема 2: Recap Python / Data Analysis / визуализации
- Основные библиотеки Python для анализа данных и визуализации: pandas, numpy, matplotlib, seaborn.
- Тема 3: Работа со строками + регулярные выражения
- Методы работы со строками в Python; регулярные выражения; практика.
- Тема 4: Парсинг данных
- Открытые источники данных; использование API и парсинга; создание датасетов. Домашнее задание: "Откуда берутся датасеты".

### Модуль 2. Введение в DL (Deep Learning)

- Цель модуля: Познакомить с основами глубокого обучения, от базовых нейронов до современных фреймворков и рекуррентных архитектур.
- Тема 1: Введение в нейросети
- Искусственный нейрон; полносвязная нейронная сеть.
- Тема 2: Градиентный спуск и backpropagation
- Градиентный спуск; chain rule; граф вычислений; backpropagation.
- Тема 3: PyTorch. Часть 1
- Фреймворк PyTorch; тензоры; dataset; dataloader; обучение на MNIST. Домашнее задание: "Практика по PyTorch".
- Тема 4: PyTorch. Часть 2
- PyTorch 2.x; torch.compile(); Transfer Learning; CNN; CIFAR/MNIST.
- Тема 5: Рекуррентные сети. Часть 1
- Рекуррентные нейронные сети: RNN, LSTM, GRU.
- Тема 6: Рекуррентные сети. Часть 2
- Классификация и генерация текста рекуррентной сетью; shallow-parsing модели.

### Модуль 3. Классические методы NLP

- Цель модуля: Изучить традиционные подходы в NLP: предобработку текста, векторные представления и применение RNN для языкового моделирования.
- Тема 1: Предобработка данных и векторные представления слов
- Pipeline предобработки (удаление стоп-слов, токенизация, лемматизация); Word2Vec; FastText.
- Тема 2: Понятие языковой модели, RNN для работы с текстом
- Языковые модели на основе LSTM; практика.
- Тема 3: Машинный перевод и seq2seq
- Задача seq2seq; encoder-decoder; концепция attention mechanism.

#### Модуль 4. Трансформерные модели

- Цель модуля: Глубоко погрузиться в архитектуру трансформеров и научиться применять их для решения практических задач.
- Тема 1: Архитектура Transformer
- Идея self-attention mechanism; архитектура Transformer.
- Тема 2: Transfer learning; BERT model
- Transfer learning; pretraining+finetuning; BERT; стратегии токенизации; практика дообучения.
- Тема 3: Работа с предобученными языковыми моделями. Задача NER
- Named Entity Recognition; подходы для решения; практика. Домашнее задание: "Почувствуй мощь трансформеров в бою".
- Тема 4: Генеративные языковые модели и энкодер-декодер архитектуры
- GPT-3; few-/zero-shot learning; архитектура T5.
- Тема 5: Towards ChatGPT
- RLHF; тюнинг инструкций; модель InstructGPT.
- Тема 6: Теория промптинга LLM
- Методы промптинга: Zero-shot, Few-shot, CoT, Self-consistency, Knowledge-generated prompting, Agency prompting, ReAct.
- Тема 7: Q&A
- Разбор типичных ошибок; ответы наставников и преподавателей.

#### Модуль 5. Дополнительные главы NLP

- Цель модуля: Рассмотреть прикладные аспекты NLP: оценку моделей и создание конечных продуктов.
- Тема 1: Оценка LLM
- Англоязычные, русскоязычные и мультязычные бенчмарки и датасеты; практика оценки трансформеров.
- Тема 2: Создание телеграм-бота
- Практика по созданию телеграм-бота. Домашнее задание: "И пусть твой бот заговорит!".

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

**Обращайтесь по любым вопросам**  
к менеджерам Академии АйТи

**+7 (495) 150 96 00** | [academy@academyit.ru](mailto:academy@academyit.ru)