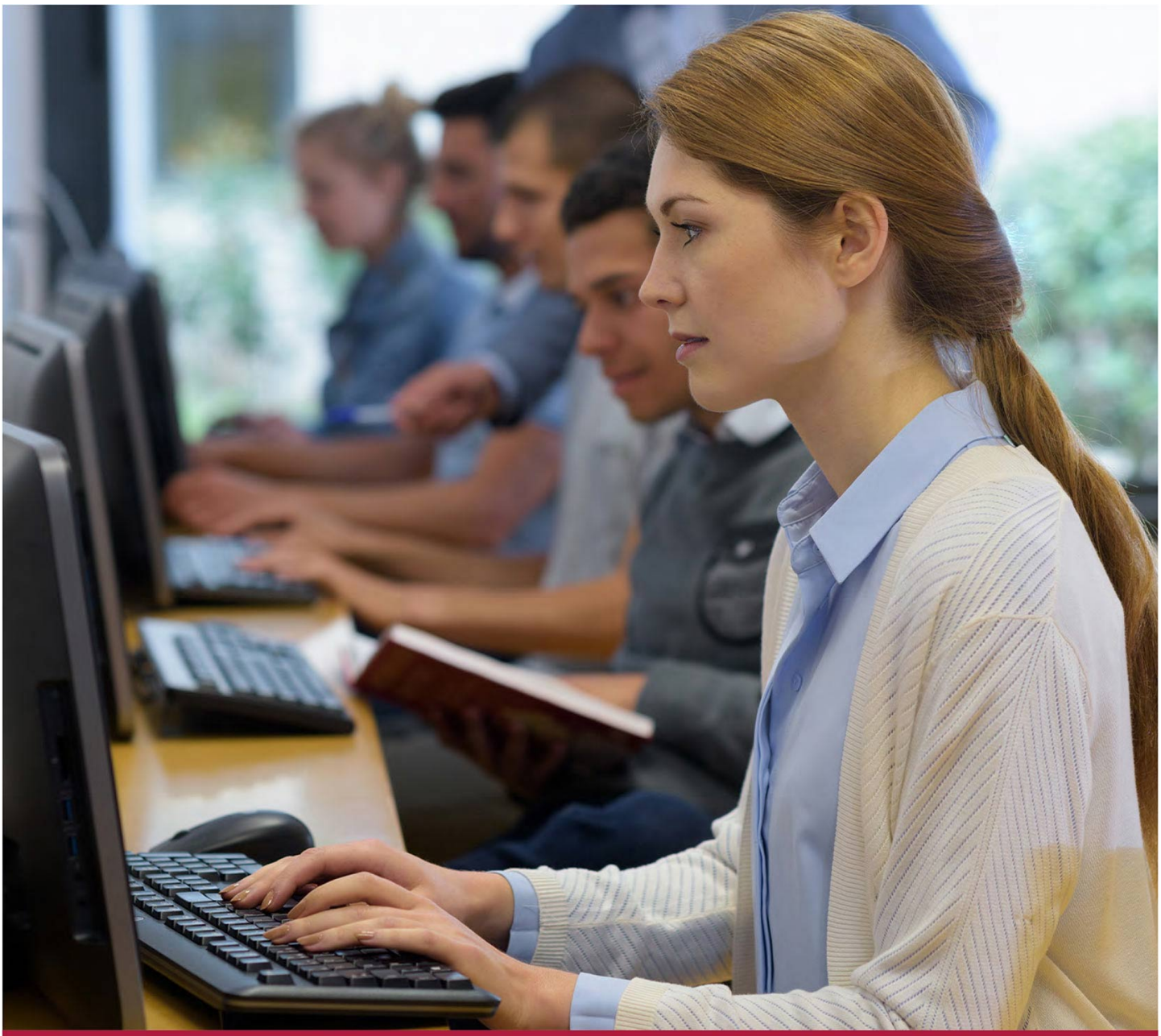
Академия АйТи
a Softline Company



Reinforcement Learning. Алгоритмы обучения с подкреплением для задач NLP

Код курса: RL-ML

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

Reinforcement Learning. Алгоритмы обучения с подкреплением для задач NLP

Код курса: RL-ML

Длительность	16 ак. часов
Формат	
Разработчик курса	Академия АйТи
Тип	Учебный курс
Способ обучения	Под руководством тренера

О курсе

Специализированный курс по методам обучения с подкреплением, адаптированным для задач обработки естественного языка. Программа охватывает путь от фундаментальных концепций RL (агент, среда, награда) через ключевые алгоритмы (Q-Learning, Policy Gradient, Actor-Critic) к практическому применению RLHF для выравнивания больших языковых моделей. Курс завершается разработкой адаптивной системы генерации ответов на базе LLaMA.

Подробная информация

Профиль аудитории:

- ML-инженеры и Data Scientists, работающие с языковыми моделями
- Python-разработчики, специализирующиеся на NLP
- специалисты по глубокому обучению, желающие освоить методы RL
- исследователи в области alignment и безопасности LLM

Предварительные требования:

- уверенное владение Python и библиотеками глубокого обучения
- базовые знания нейронных сетей и принципов обучения
- опыт работы с Hugging Face Transformers
- знакомство с языковыми моделями (GPT, LLaMA) будет преимуществом

По окончании курса слушатели смогут:

- понимать фундаментальные отличия обучения с подкреплением от классического ML
- владеть базовыми компонентами RL: Агент, Среда, Действие, Награда, Политика, Функция ценности
- реализовывать ключевые алгоритмы: Q-Learning, Policy Gradient, Actor-Critic
- применять RL для задач генерации текста, машинного перевода и диалоговых систем
- освоить подход RLHF для выравнивания (alignment) моделей под предпочтения пользователей
- разрабатывать адаптивные системы генерации ответов с интеграцией LLaMA и Q-Learning

Программа курса

Блок 1. Введение в обучение с подкреплением (Основы)

- Тема 1. От классического ML к RL: адаптация в реальном времени
- Проблемы статичных данных и фиксированных распределений. Зависимость последовательных решений. Концепция обучения в процессе взаимодействия (чат-боты, рекомендательные системы, умный дом).
- Тема 2. Анатомия RL: Агент, Среда и Награда
- Ключевые термины: Agent, Environment, Reward, Action. Цикл обучения RL. Марковский процесс принятия решений. Примеры: AlphaGo, OpenAI Five, AlphaZero, Boston Dynamics, Tesla.

Блок 2. Ключевые алгоритмы глубокого обучения с подкреплением

- Тема 3. Формализация задачи и алгоритм Q-Learning
- Проектирование Environment (reset, step). Функция ценности и Политика. Уравнение Беллмана, Q-функции, жадная стратегия, дисконтирование. Практикум: реализация QLearningAgent.
- Тема 4. Продвинутые алгоритмы: Policy Gradient и Actor-Critic
- Policy Gradient: прямая оптимизация политики. Actor-Critic: объединение подходов, TD-ошибка. Критерии выбора алгоритма. Практикум: реализация PolicyGradientAgent и ActorCriticAgent.

Блок 3. Специфика применения RL в NLP

- Тема 5. RL для работы с текстами
- RL в машинном переводе: фокус на связности текста. Адаптация повествования в играх. Оптимизация диалоговых систем.
- Тема 6. RLHF: Обучение с подкреплением на основе отзывов людей
- Пайплайн RLHF: предобучение, сбор оценок, Reward Model, дообучение. PPO (Proximal Policy Optimization). Проблемы: предвзятость, галлюцинации, субъективность.

Блок 4. Практика и проектный кейс

- Тема 7. Практикум: Разработка среды и политик
- Восстановление пропущенных элементов Environment. Усложнение логики среды. Модификация класса Policy: стратегия выбора действий.
- Тема 8. Разбор кейса: Создание адаптивной системы генерации ответов
- Интеграция LLaMA с Q-Learning и принципами RLHF. Генерация ответов разного уровня сложности. Использование Q-таблицы для выбора оптимального стиля. Адаптация под предпочтения пользователя в реальном времени.

Посмотреть расписание курса и записаться на обучение

Обращайтесь по любым вопросам
к менеджерам Академии АйТи

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