

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 29.05.2025 12:26:00
Уникальный программный ключ:
sa953a0120d891083f939673078ef1a989dae18a

Федеральное государственное автономное образовательное учреждение высшего образования
«Российский университет дружбы народов имени Патриса Лумумбы»
Факультет искусственного интеллекта
(наименование основного учебного подразделения (ОУП)-разработчика ОП ВО)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

ГЛУБОКОЕ ОБУЧЕНИЕ ДЛЯ NLP

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

02.04.02 ФУНДАМЕНТАЛЬНАЯ ИНФОРМАТИКА И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

(код и наименование направления подготовки/специальности)

Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):

УПРАВЛЕНИЕ ДАННЫМИ И ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

(наименование (профиль/специализация) ОП ВО)

2025 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Глубокое обучение для NLP» входит в программу магистратуры «Управление данными и искусственный интеллект» по направлению 02.04.02 «Фундаментальная информатика и информационные технологии» и изучается во 2 семестре 1 курса. Дисциплину реализует Кафедра прикладного искусственного интеллекта. Дисциплина состоит из 3 разделов и 26 тем и направлена на изучение современных теорий и практик применения глубоких нейронных сетей в анализе, генерации, классификации, переводе и интерпретации текстовой информации. Курс выводит студентов за пределы основ NLP и базовых RNN/Word2Vec, знакомит с state-of-the-art архитектурами трансформеров, self-supervised обучением, большими языковыми моделями, а также их реальными промышленными, научными и этическими вызовами.

Целью освоения дисциплины является научить студентов строить, дообучать, анализировать и интегрировать глубокие NLP-модели, обеспечивать их интерпретируемость, этичность и работоспособность в реальных условиях, применять современные DL-инструменты для сложных многозадачных и мультязычных кейсов, а также критически разбирать сильные и слабые стороны таких систем для задач машинного интеллекта и коммуникации.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Глубокое обучение для NLP» направлено на формирование у обучающихся следующих компетенций (части компетенций):

Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
УК-1	Способен осуществлять поиск, критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Знает принципы сбора, отбора и обобщения информации; УК-1.2 Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности; УК-1.3 Имеет практический опыт работы с информационными источниками, опыт научного поиска, создания научных текстов.;
УК-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия	УК-5.1 Знает основные категории философии, законы исторического развития, основы межкультурной коммуникации; УК-5.2 Умеет вести коммуникацию с представителями иных национальностей и конфессий с соблюдением этических и межкультурных норм; УК-5.3 Имеет практический опыт анализа философских и исторических фактов, опыт эстетической оценки явлений культуры;
УК-7	Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования	УК-7.2 Умеет применять цифровые технологии для поиска, обработки, анализа, хранения и представления информации в области профессиональной деятельности; УК-7.3 Владеет навыками применения цифровых технологий и методов поиска, обработки, анализа, хранения и представления информации в области профессиональной деятельности;

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
	полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных	
ОПК-1	Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий	ОПК-1.2 Умеет осуществлять первичный сбор и анализ материала, интерпретировать различные математические объекты; ОПК-1.3 Имеет практический опыт работы с решением математических задач и применяет его в профессиональной деятельности;
ОПК-2	Способен применять компьютерные / суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности	ОПК-2.2 Умеет анализировать типовые языки программирования, составлять программы;
ОПК-3	Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования	ОПК-3.1 Знает методы теории алгоритмов, методы системного и прикладного программирования, основные положения и концепции в области математических, информационных и имитационных моделей; ОПК-3.2 Умеет соотносить знания в области программирования, интерпретацию прочитанного, определять и создавать информационные ресурсы глобальных сетей, образовательного контента, средств тестирования систем; ОПК-3.3 Имеет практический опыт применения и разработки программного обеспечения, тестирования программных продуктов;
ПК-1	Способен разрабатывать и применять алгоритмы интеллектуальной обработки данных для решения задач профессиональной деятельности	ПК-1.1 Знает существующие системы хранения и анализа данных, алгоритмы интеллектуальной обработки данных; ПК-1.2 Умеет модифицировать алгоритмы интеллектуальной обработки данных; ПК-1.3 Имеет навыки использования и применения существующих и модифицированных систем хранения и анализа данных, алгоритмов интеллектуальной обработки данных для решения задач профессиональной деятельности;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Глубокое обучение для NLP» относится к обязательной части блока 1 «Дисциплины (модули)» образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Глубокое обучение для NLP».

Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
УК-7	Способен: искать нужные	Правовые основы использования	Методы машинного

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных	искусственного интеллекта; <i>Современные устройства центров обработки больших данных**</i> ; Методы машинного обучения (продвинутого курса); SQL и NoSQL базы данных; Программирование на языке C++ (продвинутого курса);	обучения (продвинутого курса); Генеративный искусственный интеллект; Преддипломная практика;
УК-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия	Правовые основы использования искусственного интеллекта; Иностранный язык в профессиональной деятельности; Компьютерное зрение;	Технологическая (проектно-технологическая) практика (производственная); Иностранный язык в профессиональной деятельности;
УК-1	Способен осуществлять поиск, критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	Прикладная статистика и анализ данных; <i>Обработка мультимодальных данных**</i> ; Компьютерное зрение; Основы научных исследований;	Преддипломная практика; Научно-исследовательская работа (производственная);
ОПК-1	Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий	Прикладная статистика и анализ данных; Компьютерное зрение; SQL и NoSQL базы данных;	Обучение с подкреплением; Машинное обучение на больших данных;
ОПК-2	Способен применять компьютерные / суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности	Методы машинного обучения (продвинутого курса); Программирование на языке C++ (продвинутого курса); Компьютерное зрение;	Методы машинного обучения (продвинутого курса); Методы оптимизации; Технологическая (проектно-технологическая) практика (производственная);
ОПК-3	Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области	Прикладная статистика и анализ данных; SQL и NoSQL базы данных;	Методы оптимизации;

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	информатики и математического моделирования		
ПК-1	Способен разрабатывать и применять алгоритмы интеллектуальной обработки данных для решения задач профессиональной деятельности	<i>Обработка мультимодальных данных**;</i> <i>Современные устройства центров обработки больших данных**;</i>	Преддипломная практика; <i>Искусственный интеллект и интернет вещей**;</i> Генеративный искусственный интеллект; Большие языковые модели (на основании трансформеров); Методы оптимизации;

* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Глубокое обучение для NLP» составляет «4» зачетные единицы.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

Вид учебной работы	ВСЕГО, ак.ч.	Семестр(-ы)
		2
Контактная работа, ак.ч.	51	51
Лекции (ЛК)	17	17
Лабораторные работы (ЛР)	17	17
Практические/семинарские занятия (СЗ)	17	17
Самостоятельная работа обучающихся, ак.ч.	66	66
Контроль (экзамен/зачет с оценкой), ак.ч.	27	27
Общая трудоемкость дисциплины	ак.ч.	144
	зач.ед.	4

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы

Номер раздела	Наименование раздела дисциплины	Содержание раздела (темы)		Вид учебной работы*
Раздел 1	Современные архитектуры и принципы глубокого обучения для NLP	1.1	Обзор глубоких моделей для NLP: развитие от RNN к трансформерам	ЛК
		1.2	Архитектура Transformer и ее особенности: self-attention, position encoding, преимущества по сравнению с предыдущими моделями	ЛК
		1.3	Языковые модели (BERT, GPT, T5, BLOOM), decoder-only и encoder-only архитектуры, этапы их обучения. Pretraining и transfer learning	ЛК
		1.4	Реализация и тестирование seq2seq-модели для машинного перевода или суммаризации	ЛР
		1.5	Обучение модели с архитектурой Transformer на простой NLP-задаче	ЛР
		1.6	Применение и fine-tuning BERT/GPT для NER или классификации	ЛР
		1.7	Разбор современных публикаций по архитектурам NLP-моделей	СЗ
		1.8	Практика выбора деер-модели под задачу: классификация, генерация, QA	СЗ
		1.9	Анализ преимуществ и ограничений self-attention-transformer подходов	СЗ
Раздел 2	Решение задач NLP, LLM и методы их обучения, оценки и использования	2.1	Решение задач генерации, суммаризации, машинного перевода с помощью DL	ЛК
		2.2	LLM. Этапы обучения LLM – Pre-training, instruction tuning, preference optimization. Few-shot/zero-shot learning, PEFT, prompting, адаптация больших моделей	ЛК
		2.3	Методы оценки больших языковых моделей. Типовые бенчмарки, их особенности	ЛК
		2.4	Fine-tuning трансформера на задаче генерации (суммаризация, диалог, QA)	ЛР
		2.5	Применение zero-shot/few-shot learning, PEFT и prompting-a для задач NLP	ЛР
		2.6	Методы оценки LLM, использование типовых библиотек для оценки моделей на выбранной задаче	ЛР
		2.7	Разбор обучения моделей для нетипичных задач NLP	СЗ
		2.8	Дискуссия: применимость различных методов дообучения LLM для прикладных задач	СЗ
		2.9	Workshop по методам оценки моделей: примеры сбора данных для бенчмарка и оценка его качества	СЗ
Раздел 3	Интеграция, MLOps, интерпретируемость LLM и этика глубокого NLP	3.1	Интеграция DL-NLP в продакшен: пайплайны, API, оптимизация, мониторинг. Интерпретируемость, explainability и аудит решений глубоких NLP-систем	ЛК
		3.2	Мультимодальные и мультиязычные глубокие NLP-системы: архитектуры, вызовы, примеры	ЛК
		3.3	Bias и алгоритмическая справедливость в текстовых моделях. Этические, юридические и социальные аспекты применения LLM, NLP в бизнесе, государстве, образовании	ЛК
		3.4	Организация пайплайна развертывания LLM для применения в прикладной задаче	ЛР
		3.5	Применение мультимодальной или	ЛР

Номер раздела	Наименование раздела дисциплины	Содержание раздела (темы)		Вид учебной работы*
			мультязычной модели к прикладной задаче. Анализ и визуализация токенов важности/предвзятости при ответах (LIME, AttentionViz, SHAP)	
		3.6	Семинар-кейс: особенности внедрения моделей NLP в реальном проекте (support, banking, образование)	СЗ
		3.7	Анализ работы мультимодальных систем (text+vision, text+speech и др.)	СЗ
		3.8	Workshop по методам оценки моделей: кейсы провалов объяснимости, вредных советов, анализа bias. Дебаты: этика, приватность, ответственность разработчика NLP, предотвращение misuse	СЗ

* - заполняется только по **ОЧНОЙ** форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
Лекционная	Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций.	
Компьютерный класс	Компьютерный класс для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная персональными компьютерами (в количестве 25 шт.), доской (экраном) и техническими средствами мультимедиа презентаций.	
Семинарская	Аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная комплектом специализированной мебели и техническими средствами мультимедиа презентаций.	
Для самостоятельной работы	Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом	

	специализированной мебели и компьютерами с доступом в ЭИОС.	
--	---	--

* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Макмахан Брайан, Рао Делип. Знакомство с PyTorch: глубокое обучение при обработке естественного языка. – СПб.: Питер, 2020 с.: ил. – (Серия «Бестселлеры O'Reilly»). ISBN: 978-5-4461-1241-8
2. Глубокое обучение - Гудфеллоу Я., Бенджио И., Курвилль А., 2017 г.
3. Jurafsky D., Martin J. H. Speech and language processing. – 2-е и 3-е издания. <http://web.stanford.edu/~jurafsky/slp3/>

Дополнительная литература:

1. Математические и программные методы построения моделей глубокого обучения: учебное пособие / А. В. Протождяконов, А. В. Дягилева, П. А. Пылов, Р. В. Майтак. - Москва; Вологда: Инфра-Инженерия, 2023. - 176 с. - ISBN 978-5-9729-1484-5. - Текст: электронный. - URL: <https://znanium.com/catalog/product/2094440>

Ресурсы информационно-телекоммуникационной сети «Интернет»:

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров
 - Электронно-библиотечная система РУДН – ЭБС РУДН <https://mega.rudn.ru/MegaPro/Web>
 - ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>
 - ЭБС «Юрайт» <http://www.biblio-online.ru>
 - ЭБС «Консультант студента» www.studentlibrary.ru
 - ЭБС «Знаниум» <https://znanium.ru/>
2. Базы данных и поисковые системы
 - Sage <https://journals.sagepub.com/>
 - Springer Nature Link <https://link.springer.com/>
 - Wiley Journal Database <https://onlinelibrary.wiley.com/>
 - Научометрическая база данных Lens.org <https://www.lens.org>

Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля:*

1. Курс лекций по дисциплине «Глубокое обучение для NLP».

* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**

РАЗРАБОТЧИК:

Заведующий кафедрой
прикладного искусственного
интеллекта

Должность, БУП

Подпись

Подолько Павел
Михайлович

Фамилия И.О.

РУКОВОДИТЕЛЬ БУП:

Заведующий кафедрой
прикладного искусственного
интеллекта

Должность БУП

Подпись

Подолько Павел
Михайлович

Фамилия И.О.

РУКОВОДИТЕЛЬ ОП ВО:

Заведующий кафедрой
прикладного искусственного
интеллекта

Должность, БУП

Подпись

Подолько Павел
Михайлович

Фамилия И.О.