

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

для проведения текущей и промежуточной аттестации
по учебной дисциплине

«Обработка естественного языка и большие языковые модели»

для направления подготовки

09.04.01 - Информатика и вычислительная техника

профиль подготовки:

Технологии разработки и сопровождения систем искусственного
интеллекта

1. Описание показателей(дескрипторов) и критериев оценивания компетенций на различных этапах их формирования

Контроль качества освоения дисциплины включает в себя текущий контроль успеваемости и промежуточную аттестацию. Текущий контроль успеваемости и промежуточная аттестация обучающихся проводятся в целях установления соответствия достижений обучающихся поэтапным требованиям образовательной программы к результатам обучения и формирования компетенций.

Компетенции	Показатели	Критерии в соответствии с уровнем освоения ОП			Оценочное средство (промежуточная аттестация)
		пороговый	стандартный	эталонный	
ПК-4	Знать	Основные концепции предварительно обученных моделей, их назначение и принцип работы.	Архитектуры популярных моделей (ResNet, BERT, GPT) и их области применения.	Подробные механизмы адаптации моделей, методы fine-tuning и техники увеличения данных	Создание компьютерных программ
	Уметь	Загружать и использовать предварительно обученные модели в простых задачах.	Настраивать модели под специфические задачи с использованием transfer learning.	Оптимизировать модели для производительности и точности, адаптировать для различных входных данных.	
	Владеть	Базовыми навыками работы с библиотеками для предварительно обученных моделей (PyTorch, TensorFlow).	Подходами к интеграции моделей в информационные системы.	Навыками развертывания fine-tuned моделей в продуктивной среде и их мониторинга.	
ОПК-7	Знать	Основы архитектуры трансформеров и механизма внимания.	Принципы работы моделей BERT, GPT и их различия.	Сложные механизмы оптимизации и применения трансформеров в специфических задачах.	

	Уметь	Использовать трансформеры для решения типовых задач анализа текста.	Выполнять fine-tuning трансформеров для нестандартных приложений.	Разрабатывать и внедрять трансформеры в программно-аппаратные комплексы с ограниченными ресурсами.	
	Владеть	Навыками работы с библиотеками для трансформеров (Hugging Face, Transformers).	Приемами адаптации моделей для распределенных систем и работы на GPU.	Методику развертывания трансформеров в реальном времени и их интеграции с API.	

2. Описание критериев и шкало оценивания результатов обучения по дисциплине

2.1 Критерии и шкалы оценивания результатов обучения при проведении текущего контроля успеваемости

Текущий контроль предназначен для проверки хода и качества формирования компетенций, стимулирования учебной работы обучающихся и совершенствования методики освоения новых знаний. Контролируемые разделы (темы) дисциплины, компетенции и оценочные средства представлены в таблице.

Модуль	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
все	все	ПК-4 ОПК-7.	Создание компьютерных программ

Критерии и шкала оценивания лабораторных работ

Оценка	Критерий оценки
«зачтено»	работа полностью выполнена. Даны ответы на теоретические вопросы по теме работы и вопросы о содержании лабораторной работы (исходного кода программы). Аргументирован выбор алгоритмов, структур данных. Проведена корректировка отдельных частей исходного кода, в соответствии с пожеланиями или замечаниями преподавателя.
«не зачтено»	работа не выполнена, выполнена не полностью или содержит грубые ошибки Не даны пояснения к отдельным частям исходного кода.

2.2 Критерии и шкалы оценивания результатов обучения при проведении промежуточной аттестации

Промежуточная аттестация предназначена для определения уровня освоения всего объема учебной дисциплины.

Экзамен

Экзаменационный билет включает один практический вопрос и два теоретических. Практический вопрос — задача, которую требуется решить на языке Python.

Шкала оценивания	Критерии оценивания	Уровень освоения компетенций
Отлично	Обучающийся правильно ответил на теоретические вопросы. Показал отличные знания в рамках учебного материала. Правильно выполнил практические задания. Ответил на все дополнительные вопросы	Эталонный
Хорошо	Обучающийся с небольшими неточностями ответил на теоретические вопросы. Показал хорошие знания в рамках учебного материала. С небольшими неточностями выполнил практические задания. Ответил на большинство дополнительных вопросов	Стандартный
Удовлетворительно	Обучающийся с существенными неточностями ответил на теоретические вопросы. Показал удовлетворительные знания в рамках учебного материала. С существенными неточностями выполнил практические задания. Допустил много неточностей при ответе на дополнительные вопросы	Пороговый
Неудовлетворительно	Обучающийся при ответе на теоретические вопросы и при выполнении практических заданий продемонстрировал недостаточный уровень знаний и умений. При ответах на дополнительные вопросы было допущено множество неправильных ответов	Компетенции не сформированы

3. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

3.1. Оценочные средства текущего контроля успеваемости

Примеры заданий для лабораторных работ

Задание 1. Разведочный анализ и предобработка текстов

Цель: Освоить базовые этапы подготовки текстовых данных к анализу.

1. Выберите русскоязычный корпус текстов (например, отзывы, новости, комментарии).
2. Проведите разведочный анализ:
 - длина текстов, частоты слов, облако слов;
 - 10 самых частых слов и биграмм;
 - распределение длины предложений.
3. Выполните предобработку:
 - токенизация, лемматизация, удаление стоп-слов;
 - сравните результаты стемминга и лемматизации;
 - визуализируйте частоты слов до и после обработки.
4. *Бонус:* создайте интерактивную визуализацию с Plotly.

Задание 2. Кодирование и классификация текстов

Цель: Освоить методы векторизации текста и применение классических моделей МО.

1. Используйте корпус отзывов с метками «положительный / отрицательный».
2. Сравните методы кодирования:
 - CountVectorizer, TF-IDF, Word2Vec или SentenceTransformer;
 - визуализируйте эмбединги (PCA/UMAP).
3. Обучите модель классификации (логистическая регрессия, SVM, случайный лес).
4. Сравните метрики (Accuracy, F1, ROC AUC) для разных методов кодирования.
5. *Бонус*: используйте WandB или MLFlow для логирования экспериментов.

Задание 3. Работа с предобученными LLM

Цель: Освоить работу с трансформерами и основы промпт-инжиниринга.

1. Выберите одну из моделей Hugging Face (BERT, DistilBERT, Llama, Mistral).
2. Используйте модель для задачи классификации тональности или тематического анализа.
3. Сравните подходы:
 - zero-shot (через промпт),
 - few-shot (через примеры),
 - fine-tuning (через обучающую выборку).
4. Напишите 3 промпта и объясните, почему они эффективны.
5. *Бонус*: реализуйте RAG-систему, добавив поиск релевантных текстов перед генерацией.

Примеры вопросов для защиты лабораторных работы

1. В двух–трёх предложениях опишите цель и результат вашего задания.
2. Какие этапы предобработки текста вы использовали и почему выбрали именно их (токенизация, нормализация, лемматизация/стемминг)?
3. В задании 1: как вы определяли набор стоп-слов и нужно ли было исключать или сохранять некоторые из них по контексту?
4. В задании 2: какие методы кодирования текста вы сравнивали (CountVectorizer, TF-IDF, эмбединги) и какие получили результаты?
5. Почему один из методов кодирования показал лучше/хуже результаты — с чем это связано?
6. В задании 3: как вы выбрали предобученную модель (например, BERT, GPT-3, Llama и т.д.) и на каких критериях основывались?
7. Объясните, что такое концепция «внимания» (attention) в архитектуре трансформеров и как она влияет на качество обработки текста.
8. Что такое RAG (Retrieval-Augmented Generation) и в каком случае этот подход целесообразен?
9. При работе с локальными LLM (например через LangChain) — с какими практическими проблемами вы столкнулись (ресурсы, скорость, качество)?
10. Какие метрики вы использовали для оценки качества модели классификации или генерации и почему выбрали именно их?
11. Какие ограничения или слабые места вашей работы вы можете отметить? Что можно улучшить при дальнейшем

развитии проекта?

12. Какое влияние имеют данные (их качество, объем, сбалансированность) на итоговый результат вашей модели?
13. Если бы вам потребовалось применить вашу разработку в реальной системе (продукте), какие дополнительные шаги или требования вы бы включили?
14. Что такое промпт-инжиниринг (Prompt Engineering), и какие подходы к составлению промптов вы использовали/могли бы использовать?
15. Какие этические или социальные вопросы могут возникнуть при использовании больших языковых моделей или систем с RAG, и как вы их учитывали в своей работе?

3.2. Оценочные средства промежуточной аттестации

Вопросы для экзамена

1. Дайте определение понятию **токенизация (tokenization)**. Почему этот этап важен в NLP?
1. Чем лемматизация (**lemmatization**) отличается от стемминга (**stemming**)? Приведите примеры
2. Что такое метод **TF-IDF** и в каких случаях его применение предпочтительно по сравнению с CountVectorizer?
3. Объясните, что такое **word2vec** и как он работает (в двух вариантах: CBOW и Skip-gram).
4. Что такое модель видов **GloVe** и как она концептуально отличается от word2vec?
Опишите архитектуру трансформера (Transformer): какие ключевые компоненты входят и какова роль механизма внимания (attention)?
5. Чем отличается **Encoder-Decoder** архитектура от чисто энкодера или чисто декодера в трансформерах?
Что такое большая языковая модель (LLM – large language model)? Перечислите типы и основные различия между BERT, GPT-3 и Llama.
6. Как использовать предобученную модель (например BERT) для задачи классификации текста? Опишите основные шаги.
7. Что такое **промт-инжиниринг (prompt engineering)**? Приведите минимум два примера эффективных приёмов.
8. Какие проблемы возникают при работе с локальными LLM (например, модель, развернутая локально с помощью LangChain)?
9. Опишите концепцию **Retrieval-Augmented Generation (RAG)**: поиск (retrieval) + генерация, где и зачем она применяется.
10. Какова роль векторного поиска (vector retrieval) при реализации RAG-системы? Какие форматы данных и индексы при этом могут использоваться?
11. В контексте NLP и LLM: что такое **zero-shot**, **few-shot** и **fine-tuning**? В каких ситуациях применимо каждое?
12. Перечислите и объясните три метрики качества для задач классификации текста (например, accuracy, F1-score, ROC AUC). Почему важно смотреть не только accuracy?
13. Какие сложности возникают при работе с несбалансированными классами в задаче классификации текста? Как их решать?
14. Как влияет качество данных (объём, разнообразие, язык, разметка) на итоговое качество LLM-системы или классификатора?
15. Какие этические и социальные вопросы могут возникнуть при использовании LLM и систем RAG (например, предвзятость моделей, безопасность, приватность)?
16. Представьте, что вы ведёте проект: хотите развернуть локальную LLM-систему для аналитики текстов на русском языке. Какие практические шаги необходимо предпринять: выбор модели, подготовка данных, развертывание, эксплуатация?

Опишите, как можно визуализировать эмбединги текстов (например, полученные с помощью word2vec или BERT) и зачем это может быть полезно (PCA, UMAP, TSNE).

17. Практические вопросы для экзамена включают выполнение нескольких пунктов из вышеприведенных заданий.

4. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

4.1. Описание процедур проведения текущего контроля успеваемости студентов

В таблице представлено описание процедур проведения контрольно-оценочных мероприятий текущего контроля успеваемости студентов, в соответствии с рабочей программой дисциплины, и процедур оценивания результатов обучения с помощью запланированных оценочных средств.

Наименование оценочного средства	Описания процедуры проведения контрольно-оценочного мероприятия и процедуры оценивания результатов обучения
лабораторная работа (компьютерная программа)	Список лабораторных работ выдаётся каждому студенту в начале семестра. Лабораторная работа представляет собой задачу, которую требуется решить с использованием языка программирования общего назначения. Работа выполняется студентом самостоятельно. Защита лабораторных работ проходит в виде собеседования. Преподаватель задаёт вопросы по теме лабораторной работы

4.2. Описание процедур проведения промежуточной аттестации

Экзамен

При определении уровня сформированности компетенций обучающихся на экзамене учитывается:

- знание программного материала дисциплины (блок 1 «знать»);
- знания, необходимые для выполнения практических заданий (блок 2 «уметь»);
- владение методологией дисциплины, умение применять теоретические и практические знания в нестандартных ситуациях при решении практических заданий, обосновывать свои действия (блок 3 «владеть»).

До экзамена допускаются студенты выполнившие все лабораторные работы и защитившие курсовую работу (если она предусмотрена в семестре).

Экзамен проводится в письменной форме по билетам. Билет состоит из трех вопросов (два теоретических, один практический). Время подготовки заранее оговаривается преподавателем.

Обязательным условием сдачи экзамена является решение практического задания.

В процессе ответа обучающегося на вопросы и задания билета, преподаватель может задавать дополнительные вопросы.

При определении уровня достижений обучающихся на экзамене обращается особое внимание на следующее:

1. дан полный, развернутый ответ на поставленный вопрос;
2. показана совокупность осознанных знаний об объекте, проявляющаяся в свободном оперировании понятиями, умении выделить существенные и несущественные признаки, причинно-следственные связи;
3. знание об объекте демонстрируется на фоне понимания его в системе данной дисциплины и междисциплинарных связей;
4. ответ формулируется в терминах дисциплины, изложен литературным языком, логичен, доказателен, демонстрирует авторскую позицию студента;
5. теоретические постулаты подтверждаются примерами из практики.