МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

«Рязанский государственный радиотехнический университет»

Кафедра «Безопасность жизнедеятельности и экология»

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

по дисциплине

«Большие языковые модели и генеративный ИИ»

Направление

09.03.04 Программная инженерия

Профиль

Программная инженерия систем искусственного интеллекта
Программное обеспечение систем искусственного интеллекта

Квалификация выпускника – бакалавр

Форма обучения – очная

Рязань

1. ОБЩИЕ ПОЛОЖЕНИЯ

Оценочные материалы – это совокупность учебно-методических материалов и процедур, предназначенных для оценки качества освоения обучающимися данной дисциплины как части основной образовательной программы.

Цель – оценить соответствие знаний, умений и уровня приобретённых компетенций обучающихся целям и требованиям основной образовательной программы в ходе текущего контроля и промежуточной аттестации.

Основная задача – обеспечить оценку уровня сформированности компетенций и индикаторов их достижения, приобретаемых обучающимися в соответствии с требованиями ФГОС ВО.

Контроль знаний обучающихся проводится в форме текущего контроля (лабораторные работы) и промежуточной аттестации (экзамен).

2. ПЕРЕЧЕНЬ КОМПЕТЕНЦИЙ С УКАЗАНИЕМ ЭТАПОВ ИХ ФОРМИРОВАНИЯ

При освоении дисциплины формируются следующие компетенции:

- **ПК-2** – способен проводить передовые исследования в области фундаментальных и генеративных моделей (FC-2);
  - **ПК-2.1** – исследует и разрабатывает большие языковые модели (LLM) и другие модели для символьных данных (FC-2.1);
  - **ПК-2.2** – исследует и разрабатывает диффузионные и другие модели для несимвольных данных (FC-2.2).
- **ПК-4** – способен определять требования к наборам данных для решения задач машинного обучения, проводить разметку и анализ наборов данных, оценивать качество данных, обеспечивать непрерывную интеграцию данных (BD-2);
  - **ПК-4.1** – определяет требования к наборам и качеству данных для решения задач машинного обучения (BD-2.1);
  - **ПК-4.2** – работает с данными, в том числе собирает данные из разрозненных источников, проверяет данные на корректность (BD-2.2).
- **ПК-6** – способен применять и (или) разрабатывать архитектуры глубоких нейронных сетей (DL-1);
  - **ПК-6.5** – разрабатывает, оптимизирует и применяет автокодировщики (AE) и вариационные автокодировщики (VAE) (DL-1.7);
  - **ПК-6.6** – разрабатывает, адаптирует и внедряет трансформерные архитектуры для решения задач обработки последовательностей (DL-1.9).
- **ПК-13** – способен применять и (или) разрабатывать алгоритмы, методы и технологии обработки естественного языка (DL-4);

- **ПК-13.1** – применяет известные алгоритмы и библиотеки обработки естественного языка, предобученные глубокие нейросетевые модели для прикладных задач анализа текстов (DL-4.1);
- **ПК-13.2** – определяет стек технологий, методов и алгоритмов для построения продуктов с обработкой естественного языка (DL-4.2).

Указанные компетенции формируются в соответствии со следующими этапами:

- формирование и развитие теоретических знаний (лекционные занятия, самостоятельная работа);
- приобретение и развитие практических умений (лабораторные работы, самостоятельная работа);
- закрепление знаний и навыков (выполнение и защита лабораторных работ, подготовка к экзамену).

3. ПОКАЗАТЕЛИ И КРИТЕРИИ ОЦЕНИВАНИЯ КОМПЕТЕНЦИЙ

Сформированность каждой компетенции оценивается по трёхуровневой шкале:

- **пороговый уровень** – обязательный для всех обучающихся; соответствует освоению более 40% приведённых знаний, умений и навыков;
- **продвинутый уровень** – соответствует освоению более 60% знаний, умений и навыков;
- **эталонный уровень** – соответствует освоению более 80% знаний, умений и навыков.

При освоении менее 40% компетенция считается несформированной.

Таблица 1. Показатели достижения индикаторов компетенций

Компетенция	Индикаторы (краткое содержание)	Этап	Оценочное средство
ПК-2.1	Знать принципы организации, предобучения и дообучения больших языковых моделей. Уметь производить анализ предметных областей, в том числе их фронтирные особенности. Владеть основными способами выбора больших языковых моделей для специфических и фронтирных задач.	7	Экзамен

Компетенция	Индикаторы (краткое содержание)	Этап	Оценочное средство
ПК-2.2	Знать основные методы исследования и разработки диффузионных, свёрточных, рекуррентных, поточных и графовых моделей. Уметь использовать в проектных решениях трансформеры, автокодировщики и графовые нейронные сети. Владеть приёмами работы с заказчиками и соисполнителями по настройке и эксплуатации больших генеративных языковых моделей.	7	Экзамен
ПК-4.1	Знать основные требования, предъявляемые к датасетам и естественно-языковым корпусам. Уметь численно оценивать качество исходных корпусов, определять достоверность, релевантность и пертинентность. Владеть практическими технологиями сбора, хранения и предварительной обработки данных для задач МО.	7	Экзамен
ПК-4.2	Знать приёмы и методы предварительной подготовки и комплексирования различных по типу и источникам данных. Уметь дообучать, адаптировать и оптимизировать генеративные модели под специфические задачи. Владеть способами добычи новых данных и поддержкой работоспособности языковых моделей в условиях дрейфа данных.	7	Экзамен
ПК-6.5	Знать основные методы оптимизации и настройки автокодировщиков, трансформеров и вариационных автокодировщиков. Уметь решать задачи	7	Экзамен

Компетенция	Индикаторы (краткое содержание)	Этап	Оценочное средство
	сокращения размерности признакового пространства LLM и обнаружения аномалий. Владеть технологией проектирования архитектур, настройки и обучения моделей для внедрения в продуктивную среду.		
ПК-6.6	Знать особенности внутреннего устройства трансформерных моделей типа BERT, DistilBERT, RoBERTa и др. Уметь создавать новые языковые, в том числе генеративные модели для оптимизации их обучения и внедрения в промышленную эксплуатацию. Владеть современной технологией разработки, настройки, оценки точности и внедрения LLM.	7	Экзамен
ПК-13.1	Знать особенности применения библиотек SpaCy, torch, transformers, keras и др. для NLP. Уметь применять предобученные языковые модели (BERT, DistilBERT) в прикладных задачах анализа текста с дообучением и валидацией. Владеть технологией анализа текстов на токсичность, гневность, достоверность (фейк) и другим семантическим классам.	7	Экзамен
ПК-13.2	Знать особенности построения программных продуктов с применением NLP (чат-боты, вопросно-ответные и справочные системы). Уметь выбирать программный инструментарий для проектирования, настройки и внедрения естественно-языковых систем. Владеть технологией программирования	7	Экзамен

Компетенция	Индикаторы (краткое содержание)	Этап	Оценочное средство
	пайплайнов для организации последовательной обработки естественного языка.		

4. ПАСПОРТ ОЦЕНОЧНЫХ МАТЕРИАЛОВ ПО ДИСЦИПЛИНЕ

Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или её части)	Наименование оценочного средства
Тема 1. Базовая технология проектирования ЕЯ систем (история, предметные области, архитектуры, введение в ИИ/МО/глубокое обучение, определение LLM)	ПК-2.1, ПК-6.5, ПК-6.6	Экзамен
Тема 2. Предварительная обработка текста (токенизация, стемминг, лемматизация, нормализация, стоп-слова, NER)	ПК-2.1, ПК-13.2	Экзамен
Тема 3. Векторные представления слов (Bag-of-words, TF-IDF, Word2Vec, GloVe, FastText, Sentence Embeddings)	ПК-2.1, ПК-13.2	Экзамен
Тема 4. Методы классификации текста (Naive Bayes, Logistic Regression, SVM, RNN, CNN, Transformer)	ПК-2.1, ПК-6.5, ПК-6.6	Экзамен
Тема 5. Архитектура трансформерных сетей (Self-Attention, Multi-Head Attention, Positional Encoding, Encoder-Decoder)	ПК-2.1, ПК-2.2, ПК-4.2, ПК-6.5, ПК-6.6	Экзамен
Тема 6. Предварительное обучение LLM (Masked Language Modeling, Next Sentence Prediction, Causal Language Modeling)	ПК-2.1, ПК-4.1, ПК-6.6, ПК-13.1	Экзамен

Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или её части)	Наименование оценочного средства
Тема 7. Большие языковые модели (BERT, LaMDA, GPT и др., сравнение архитектур, семантическое сходство, машинный перевод, анализ тональности, суммаризация)	ПК-2.1, ПК-2.2, ПК-4.1, ПК-4.2, ПК-13.1, ПК-13.2	Экзамен
Тема 8. Fine-tuning и адаптация LLM (адаптация к конкретным задачам, оценка качества LLM)	ПК-2.1, ПК-2.2, ПК-4.1, ПК-13.2	Экзамен
Тема 9. Генеративные модели (авторегрессивные модели, диффузионные модели, генерация текста, оценка качества)	ПК-6.6, ПК-13.1, ПК-13.2	Экзамен
Тема 10. Применение LLM (чат-боты, виртуальные помощники, диалоговые системы, кодогенерация)	ПК-13.2	Экзамен
Лабораторный практикум (8 лабораторных работ)	ПК-2.1, ПК-2.2, ПК-4.1, ПК-4.2, ПК-6.5, ПК-6.6, ПК-13.1, ПК-13.2	Защита лабораторных работ

5. ТИПОВЫЕ КОНТРОЛЬНЫЕ ЗАДАНИЯ

5.1. Промежуточная аттестация (экзамен)

Экзамен проводится в форме устного ответа по билетам, включающим теоретический вопрос и практическое задание (задачу). Также может использоваться тестирование.

5.1.1. Тестовые вопросы закрытого типа (20 вопросов)

1. Что такое большая языковая модель (LLM)?
а) Модель машинного обучения, обученная на большом объёме текстовых данных и способная генерировать текст;
б) База данных для хранения текстов;
в) Алгоритм сортировки;

г) Протокол передачи данных.
Правильный ответ: а)

2. Какая архитектура лежит в основе большинства современных LLM?
а) Трансформер;
б) Свёрточная нейронная сеть;
в) Рекуррентная нейронная сеть;
г) Перцептрон.
Правильный ответ: а)

3. Что такое токенизация?
а) Разбиение текста на токены (слова, подслова, символы);
б) Удаление стоп-слов;
в) Приведение слов к базовой форме;
г) Векторизация текста.
Правильный ответ: а)

4. Что такое лемматизация?
а) Приведение слова к его словарной форме (лемме);
б) Удаление окончаний;
в) Разбиение текста на предложения;
г) Векторизация слов.
Правильный ответ: а)

5. Что такое Bag-of-words?
а) Модель представления текста в виде мешка слов без учёта порядка;
б) Модель на основе трансформеров;
в) Метод кластеризации;
г) Алгоритм шифрования.
Правильный ответ: а)

6. Что такое TF-IDF?
а) Метрика, учитывающая частоту слова в документе и обратную частоту в корпусе;
б) Тип нейронной сети;
в) Метод оптимизации;
г) Алгоритм сжатия.
Правильный ответ: а)

7. Что такое Word2Vec?
а) Метод векторного представления слов на основе нейронных сетей;
б) Алгоритм кластеризации;
в) Тип базы данных;
г) Протокол передачи данных.
Правильный ответ: а)

8. Что такое механизм самовнимания (Self-Attention)?
а) Механизм, позволяющий модели учитывать взаимосвязи между всеми словами

последовательности;
б) Метод регуляризации;
в) Тип функции активации;
г) Алгоритм сжатия.
Правильный ответ: а)

9. Что такое BERT?
а) Предобученная языковая модель на основе трансформеров с двунаправленным обучением;
б) База данных;
в) Алгоритм сортировки;
г) Веб-фреймворк.
Правильный ответ: а)

10. Что такое GPT?
а) Генеративная предобученная трансформерная модель;
б) Тип базы данных;
в) Протокол передачи данных;
г) Алгоритм шифрования.
Правильный ответ: а)

11. Что такое fine-tuning?
а) Дообучение предобученной модели на конкретной задаче;
б) Обучение с нуля;
в) Метод кластеризации;
г) Алгоритм сжатия.
Правильный ответ: а)

12. Что такое промпт-инжиниринг?
а) Практика составления эффективных запросов (промптов) к LLM для получения желаемого результата;
б) Программирование на Python;
в) Метод оптимизации;
г) Алгоритм шифрования.
Правильный ответ: а)

13. Что такое RAG (Retrieval Augmented Generation)?
а) Подход, сочетающий поиск информации и генерацию текста LLM;
б) Тип базы данных;
в) Метод сжатия;
г) Протокол передачи данных.
Правильный ответ: а)

14. Что такое Hugging Face Transformers?
а) Библиотека для работы с предобученными языковыми моделями;
б) Операционная система;
в) Веб-браузер;

г) Текстовый редактор.
Правильный ответ: а)

15. Что такое NER (Named Entity Recognition)?
а) Задача распознавания именованных сущностей в тексте;
б) Метод кластеризации;
в) Тип нейронной сети;
г) Алгоритм шифрования.
Правильный ответ: а)

16. Что такое диффузионная модель?
а) Генеративная модель, основанная на постепенном добавлении и удалении шума;
б) Тип базы данных;
в) Метод оптимизации;
г) Алгоритм сжатия.
Правильный ответ: а)

17. Что такое метрика BLEU?
а) Метрика оценки качества машинного перевода;
б) Тип базы данных;
в) Метод кластеризации;
г) Алгоритм шифрования.
Правильный ответ: а)

18. Что такое эмбеддинг (embedding)?
а) Векторное представление слов, предложений или документов;
б) Тип базы данных;
в) Метод оптимизации;
г) Протокол передачи данных.
Правильный ответ: а)

19. Что такое LangChain?
а) Фреймворк для создания приложений на основе LLM;
б) Операционная система;
в) База данных;
г) Веб-сервер.
Правильный ответ: а)

20. Что такое дрейф модели (model drift)?
а) Снижение качества модели со временем из-за изменения данных;
б) Ошибка в коде;
в) Тип нейронной сети;
г) Метод оптимизации.
Правильный ответ: а)

5.1.2. Тестовые вопросы открытого типа (10 вопросов)

1. Опишите основные этапы предварительной обработки текста для задач NLP.

2. Сравните подходы Bag-of-words, TF-IDF и Word Embeddings для представления текста.

3. Объясните архитектуру трансформера и механизм самовнимания (Self-Attention).

4. Что такое предобучение LLM? Опишите методы Masked Language Modeling и Causal Language Modeling.

5. Что такое fine-tuning? Опишите процесс адаптации LLM к конкретной задаче.

6. Опишите методы оценки качества больших языковых моделей.

7. Что такое промпт-инжиниринг? Приведите примеры техник (Few-shot, Chain-of-Thought, Tree-of-Thought).

8. Опишите подход RAG (Retrieval Augmented Generation) и его применение.

9. Сравните авторегрессивные и диффузионные генеративные модели.

10. Какие этические и безопасностные аспекты применения LLM вы знаете?

5.1.3. Типовые практические задания (задачи) к экзамену

1. Разработать программу предварительной обработки текста (токенизатор, лемматизатор, стеммер) на Python с использованием NLTK/spaCy.
2. Построить пайплайн предобработки текста и применить его к заданному корпусу.
3. Создать модели Bag-of-words и TF-IDF, проанализировать результаты.
4. Работа с предобученными Word Embeddings (Word2Vec, GloVe), визуализировать эмбеддинги.
5. Выполнить классификацию текста с использованием Naive Bayes и Logistic Regression, оценить качество (точность, полнота, F1-мера).
6. Реализовать распознавание именованных сущностей (NER).
7. Загрузить и использовать предобученные токенизаторы и модели из Hugging Face Transformers.
8. Выполнить генерацию текста с использованием GPT-модели, экспериментировать с параметрами генерации.
9. Провести fine-tuning предобученной модели BERT для задачи классификации текста.
10. Реализовать чат-бота на основе LLM с использованием LangChain.

5.1.4. Вопросы к экзамену (30 вопросов)

1. История развития естественно-языковых систем. Предметные области.
2. Основные понятия и архитектуры систем обработки естественного языка.
3. Введение в концепцию ИИ, машинного обучения и глубокого обучения.
4. Определение больших языковых моделей (LLM).

5. Предварительная обработка текста: токенизация, стемминг, лемматизация.
6. Нормализация текста. Удаление стоп-слов. Работа с пунктуацией.
7. Обработка именованных сущностей (NER).
8. Векторные представления слов (Word Embeddings).
9. Bag-of-words. TF-IDF. Недостатки bag-of-words моделей.
10. Word2Vec, GloVe, FastText. Сравнение методов.
11. Представление предложений (Sentence Embeddings).
12. Основные методы классификации текста (Naive Bayes, Logistic Regression, SVM).
13. Нейронные сети для классификации текста (RNN, CNN, Transformer).
14. Архитектура трансформерных сетей. Механизм самовнимания (Self-Attention).
15. Multi-Head Attention, Positional Encoding, Encoder-Decoder.
16. Предварительное обучение LLM: Masked Language Modeling.
17. Предварительное обучение LLM: Next Sentence Prediction, Causal Language Modeling.
18. Основные типы больших языковых моделей: BERT, LaMDA, GPT и др.
19. Сравнение архитектур и возможностей LLM.
20. Применение LLM в задачах определения семантического сходства предложений.
21. Применение LLM в машинном переводе, анализе тональности, суммаризации текстов.
22. Fine-tuning и адаптация больших языковых моделей к конкретным задачам.
23. Оценка качества больших языковых моделей.
24. Принципы работы генеративных моделей. Авторегрессивные и диффузионные модели.
25. Генерация текста: подходы и техники. Оценка качества генерируемого текста.
26. Применение LLM в чат-ботах и виртуальных помощниках.
27. Диалоговые системы и управление диалогом.
28. Применение LLM в кодогенерации. Инструменты и фреймворки для автоматизации программирования.
29. Промпт-инжиниринг: Few-shot, Chain-of-Thought, Tree-of-Thought.
30. Этические и безопасностные аспекты применения LLM. Оценка дрейфа модели.

6. КРИТЕРИИ ОЦЕНИВАНИЯ

6.1. Шкала оценивания тестирования

Шкала	Критерий
5 (отлично)	Процент верных ответов от 90 до 100%
4 (хорошо)	Процент верных ответов от 75 до 89%
3 (удовлетворительно)	Процент верных ответов от 60 до 74%
2 (неудовлетворительно)	Процент верных ответов менее 60%

6.2. Шкала оценивания теоретического вопроса на экзамене

Шкала	Критерий
5 (отлично)	Дан полный ответ, показаны глубокие систематизированные знания, приведены примеры, даны ответы на дополнительные вопросы.
4 (хорошо)	Дан полный ответ, но на некоторые дополнительные вопросы ответы даны с помощью наводящих вопросов.
3 (удовлетворительно)	Дан неполный ответ, на дополнительные вопросы ответы даны только с помощью преподавателя.
2 (неудовлетворительно)	Студент не смог ответить на вопрос.

6.3. Шкала оценивания практического задания на экзамене

Шкала	Критерий
5 (отлично)	Задание решено верно, получены корректные результаты, сделаны выводы.
4 (хорошо)	Задание решено верно, но имеются незначительные технические неточности.
3 (удовлетворительно)	Задание решено верно с дополнительными наводящими вопросами преподавателя.
2 (неудовлетворительно)	Задание не решено.

6.4. Итоговая оценка на экзамене

На экзамене студент отвечает на теоретический вопрос и выполняет практическое задание. Максимальная сумма баллов – 10 (5 за теорию + 5 за практику).
Итоговая оценка выставляется по пятибалльной шкале:

- «отлично» – 9–10 баллов;

- «хорошо» – 7–8 баллов;
- «удовлетворительно» – 5–6 баллов;
- «неудовлетворительно» – менее 5 баллов.

Оператор ЭДО ООО "Компания "Тензор"

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ

СОГЛАСОВАНО **ФГБОУ ВО "РГРТУ", РГРТУ,** Овечкин Геннадий Владимирович, Заведующий кафедрой ВПМ **05.11.26** 14:46 (MSK) Простая подпись