МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РЯЗАНСКИЙ ГОСУДАРСТВЕННЫЙ РАДИОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМЕНИ
В.Ф. УТКИНА»

Кафедра «Вычислительная и прикладная математика»

МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ

«Большие языковые модели и генеративный ИИ»

Направление подготовки – 09.03.04 Программная инженерия

ОПОП «Программная инженерия систем искусственного интеллекта»

Квалификация (степень) выпускника – бакалавр

Форма обучения – очная (4 года)

Рязань

1. СПИСОК ТЕОРЕТИЧЕСКИХ ВОПРОСОВ К ЭКЗАМЕНУ

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.

2. ПЕРЕЧЕНЬ ПРАКТИЧЕСКИХ И ЛАБОРАТОРНЫХ РАБОТ ПО ДИСЦИПЛИНЕ

№	Тема лабораторной работы	Кол-во часов
1	Разработка программ предварительной обработки слов и/или текстов (токенизатор, лемматизатор, стеммер) с использованием Python и NLTK/spaCy	2
2	Разработка собственного пайплайна предобработки текстов	2
3	Создание моделей Bag-of-words и TF-IDF. Анализ результатов	2
4	Работа с предобученными Word Embeddings (Word2Vec, GloVe). Визуализация эмбеддингов	2
5	Классификация текста с использованием Naive Bayes и Logistic Regression. Оценка качества модели (точность, полнота, F1-мера)	2
6	Обработка именованных сущностей (NER)	2
7	Знакомство с Hugging Face Transformers. Загрузка и использование предобученных токенизаторов и моделей	2
8	Генерация текста с использованием GPT-1/2/3 или аналогичных BERT-моделей. Эксперименты с различными параметрами генерации	2
Итого		**16**

3. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Самостоятельная работа как вид учебной работы может использоваться на лабораторных занятиях, а также иметь самостоятельное значение – внеаудиторная самостоятельная работа обучающихся – при подготовке к лабораторным работам и при подготовке к экзамену.

Самостоятельная работа включает в себя следующие этапы:

- изучение теоретического материала (работа над конспектом лекции);
- самостоятельное изучение дополнительных информационных ресурсов (доработка конспекта лекции);
- выполнение заданий текущего контроля успеваемости (подготовка к лабораторным работам);
- итоговая аттестация по дисциплине (подготовка к экзамену).

При подготовке к лабораторным работам необходимо:

1. Изучить конспект лекций и методические указания к данной работе.
2. Ознакомиться с рекомендуемой основной и дополнительной литературой.
3. Выполнить индивидуальное задание.
4. Оформить отчёт в соответствии с требованиями кафедры.
5. Защитить работу, ответив на вопросы преподавателя.

Важным этапом является защита лабораторной работы. В процессе защиты студент отвечает на вопросы преподавателя, касающиеся теоретического материала, относящегося к данной работе, и проекта, реализующего её задание, комментирует полученные результаты. При подготовке к защите рекомендуется ознакомиться со списком вопросов по изучаемой теме и попытаться самостоятельно на них ответить, используя конспект лекций и рекомендуемую литературу.

4. РЕКОМЕНДАЦИИ ПО ПЛАНИРОВАНИЮ И ОРГАНИЗАЦИИ ВРЕМЕНИ, НЕОБХОДИМОГО ДЛЯ ИЗУЧЕНИЯ ДИСЦИПЛИНЫ. ОПИСАНИЕ ПОСЛЕДОВАТЕЛЬНОСТИ ДЕЙСТВИЙ СТУДЕНТА («СЦЕНАРИЙ ИЗУЧЕНИЯ ДИСЦИПЛИНЫ»)

Рекомендуется следующим образом организовать время, необходимое для изучения дисциплины.

Для освоения лекционного материала следует: изучить конспект лекции в тот же день, после лекции: 10–15 минут, повторно прочитать конспект лекции за день перед следующей лекцией: 10–15 минут. Также следует изучить теоретический лекционный материал по рекомендуемому учебнику/учебному пособию: 1 час в неделю.

Следует максимально использовать лекционное время для изучения дисциплины, понимания лекционного материала и написания конспекта лекций. В процессе лекционного занятия студент должен уметь выделять важные моменты и основные положения. При написании конспекта лекций следует придерживаться следующих правил и рекомендаций:

1. При ведении конспекта рекомендуется структурировать материал по разделам, главам, темам. Вести нумерацию формул. Выделять по каждой теме постановку задачи, основные положения, выводы. Кратко записывать те пояснения лектора,

которые показались особенно важными. Это позволит при подготовке к сдаче экзамена не запутаться в структуре лекционного материала.
2. Лекционный материал следует записывать в конспект лишь после того, как излагаемый лектором тезис будет вами дослушан до конца и понят.
3. При конспектировании следует отмечать непонятные на данном этапе положения, доказательства и пр.
4. Рекомендуется по каждой теме выразить своё мнение, комментарий, вывод.

Доработка конспекта лекции с применением учебника, методической литературы, дополнительной литературы, интернет-ресурсов: этот вид самостоятельной работы студентов особенно важен в том случае, когда одну и ту же задачу можно решать различными способами, а на лекции изложен только один из них. Кроме того, рабочая программа предполагает рассмотрение некоторых относительно несложных тем только во время самостоятельных занятий, без чтения лектором.

Подготовка к лабораторным работам состоит в теоретической подготовке (изучение конспекта лекций, методических указаний к данной работе и дополнительной литературы) и выполнении индивидуального задания. Выполнение каждой из запланированных работ заканчивается предоставлением отчёта. Требования к форме и содержанию отчёта приведены в методических указаниях к лабораторным работам или определяются преподавателем на первом занятии. Допускаясь к лабораторной работе, каждый студент должен представить преподавателю «заготовку» отчёта, содержащую: оформленный титульный лист, цель работы, задание, проект решения, полученные результаты, выводы.

Подготовка к сдаче экзамена.

Экзамен – форма промежуточной проверки знаний, умений, навыков, степени освоения дисциплины. Главная задача экзамена состоит в том, чтобы у студента по окончании изучения данной дисциплины сформировались определённое представление об общем содержании дисциплины, определённые теоретические знания и практические навыки, определённый кругозор. Готовясь к экзамену, студент приводит в систему знания, полученные на лекциях, на лабораторных занятиях, разбирается в том, что осталось непонятным, и тогда изучаемая им дисциплина может быть воспринята в полном объёме с присущей ей строгостью и логичностью, её практической направленностью.

Экзамен даёт возможность преподавателю определить теоретические знания студента и его практические навыки при решении определённых прикладных задач. Оцениваются: понимание и степень усвоения теоретического материала; степень знакомства с основной и дополнительной литературой, а также с современными публикациями; умение применить теорию к практике, решать определённые практические задачи данной предметной области, правильно проводить расчёты и т. д.; знакомство с историей данной науки; логика, структура и стиль ответа, умение защищать выдвигаемые положения.

Планируйте подготовку к экзамену, учитывая сразу несколько факторов: неоднородность в сложности учебного материала и степени его проработки в ходе обучения, свои индивидуальные способности. Рекомендуется делать перерывы в занятиях через каждые 50–60 минут на 10 минут. После 3–4 часов занятий следует сделать часовой перерыв. Чрезмерное утомление приведёт к снижению тонуса интеллектуальной деятельности. Целесообразно разделять весь рабочий день на три рабочих периода – с утра до обеда, с обеда до ужина и с ужина до сна. Каждый рабочий период дня должен заканчиваться отдыхом не менее 1 часа. Работая в сессионном режиме, студент имеет возможность увеличить время занятий с 10 (как требовалось в семестре) до 12 часов в сутки.

Подготовку к экзамену следует начинать с общего планирования своей деятельности. С определения объёма материала, подлежащего проработке, необходимо внимательно сверить конспекты с программой дисциплины, чтобы убедиться, все ли разделы отражены в лекциях, отсутствующие темы изучить по учебнику. Второй этап предусматривает системное изучение материала по данному предмету с обязательной записью всех выкладок, выводов, формул. На третьем этапе – этапе закрепления – полезно чередовать углублённое повторение особенно сложных вопросов с беглым повторением всего материала.

5. РЕКОМЕНДАЦИИ ПО РАБОТЕ С ЛИТЕРАТУРОЙ

Основная литература, литература для дополнительного чтения и методические указания к лабораторным работам указаны в рабочей программе данной дисциплины и доступны посредством ЭБС.

Теоретический материал курса становится более понятным, когда дополнительно к прослушиванию лекции и изучению конспекта изучаются и книги по данному предмету. Литературу по дисциплине рекомендуется читать как в бумажном, так и в электронном виде (если отсутствует бумажный аналог). Полезно использовать несколько учебников и пособий по дисциплине. Рекомендуется после изучения очередного параграфа ответить на несколько вопросов по данной теме. Кроме того, полезно мысленно задать себе следующие вопросы (и попробовать ответить на них): «о чём этот параграф?», «какие новые понятия введены, каков их смысл?», «зачем мне это нужно по специальности?».

Рекомендуется самостоятельно изучать материал, который ещё не прочитан на лекции и не применялся на лабораторном занятии, тогда занятия будут гораздо понятнее. В течение недели рекомендуется выбрать время (1 час) для работы с литературой.

6. БИБЛИОГРАФИЧЕСКИЙ СПИСОК

6.1. Основная литература

1. Каширин И. Ю. Проектирование систем искусственного интеллекта: учеб. пособие. – Рязань, 2000. – 48 с. – ISBN 5-772-0142-5.
2. Каширин И. Ю., Маркова В. В. Математическая логика и теория алгоритмов: метод. указ. – Рязань, 2003. – 28 с.
3. Каширин И. Ю. Основы искусственного интеллекта: метод. указ. к лаб. работам: Методические указания. – Рязань, 2022. – URL: https://elib.rsreu.ru/ebs/download/3425
4. Цуканова Н. И. Применение нейронных сетей к решению практических задач обработки изображений и структурированных данных: учеб. пособие. – Москва: КУРС, 2022. – 256 с. – ISBN 978-5-907535-60-2.
5. Каширин И. Ю., Маликова Л. В., Маркова В. В. Основы формальных систем: учеб. пособие. – М.: НИЦПриС, 1999. – 84 с. – ISBN 5-900881-10-5.

6.2. Дополнительная литература

1. Каширин И. Ю. Оценочные и методические материалы по дисциплине «Основы систем искусственного интеллекта»: Методические указания. – Рязань, 2023. – URL: https://elib.rsreu.ru/ebs/download/3874
2. Каширин И. Ю. Проектирование систем искусственного интеллекта: учеб. пособие. – Рязань, 2000. – 48 с. – ISBN 5-772-0142-5.
3. Цуканова Н. И. Программирование глубоких нейронных сетей на языке Python: учеб. пособие. – Рязань: КУРС, 2022. – URL: https://elib.rsreu.ru/ebs/download/3187
4. Судаков В. А., Титов Ю. П. Методы искусственного интеллекта в информационных системах: учебник. – Москва: РТУ МИРЭА, 2024. – 150 с. – ISBN 978-5-7339-2354-3. – URL: https://e.lanbook.com/book/457079
5. Каширин И. Ю., Коричнев Л. П. Формальное исследование интеллектуальных программных систем. – М.: Радио и связь, 1997. – 160 с. – ISBN 5-256-01368-8.
6. Верхова Г. В., Прокофьев П. А. Методы и модели искусственного интеллекта в управлении техническими системами: учебно-методическое пособие по выполнению лабораторных работ. – Санкт-Петербург: СПбГУТ им. М. А. Бонч-Бруевича, 2024. – 53 с. – URL: https://e.lanbook.com/book/426122
7. Степанов Ю. А., Вылегжанина А. В., Бурмин Л. Н. Системы искусственного интеллекта: учебное пособие. – Кемерово: КемГУ, 2024. – 102 с. – ISBN 978-5-8353-3166-6. – URL: https://e.lanbook.com/book/427532

6.3. Методические разработки

1. Каширин И. Ю. Проектирование систем искусственного интеллекта: метод. указ. к лаб. работам: Методические указания. – Рязань, 2021. – URL: https://elib.rsreu.ru/ebs/download/2852
2. Пылькин А. Н., Соколова Ю. С. Использование Python в научных вычислениях (Часть 2): метод. указ. к лаб. работам: Методические указания. – Рязань: РИЦ РГРТУ, 2024. – URL: https://elib.rsreu.ru/ebs/download/3908
3. Панина И. С. Машинное обучение: метод. указ. к практ. занятиям: Методические указания. – Рязань: РИЦ РГРТУ, 2024. – URL: https://elib.rsreu.ru/ebs/download/3909

4. Каширин И. Ю. Методы машинного обучения: метод. указ. к лаб. работам: Методические указания. – Рязань, 2022. – URL: https://elib.rsreu.ru/ebs/download/3427
5. Каширин И. Ю. Сетевые технологии и администрирование в информационных системах: метод. указ. к лаб. работам: Методические указания. – Рязань, 2022. – URL: https://elib.rsreu.ru/ebs/download/3428
6. Цуканова Н. И. Оценочные и методические материалы по дисциплине «Основы искусственных нейронных сетей»: Методические указания. – Рязань, 2023. – URL: https://elib.rsreu.ru/ebs/download/3873
7. Каширин И. Ю., Новичков В. С. Объектно-ориентированное программирование на Си++: метод. указ. к лаб. работам № 9–12. – Рязань, 1993. – 56 с.
8. Каширин И. Ю., Маликова Л. В., Маркова В. В. Дискретная математика: метод. указ. к практ. занятиям. – Рязань, 2002. – 47 с.
9. Каширин И. Ю. Оценочные и методические материалы по дисциплине «Машинная обработка текстов на естественном языке»: Методические указания. – Рязань, 2023. – URL: https://elib.rsreu.ru/ebs/download/3869
10. Цуканова Н. И., Дмитриева Т. А. Системы искусственного интеллекта. Нейронные сети и генетические алгоритмы: метод. указ. к лаб. работам 1, 2: Методические указания. – Рязань, 2020. – URL: https://elib.rsreu.ru/ebs/download/2842
11. Каширин И. Ю. Методы интеллектуального анализа данных: метод. указ. к лаб. работам: Методические указания. – Рязань, 2021. – URL: https://elib.rsreu.ru/ebs/download/2858

6.4. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Электронная библиотека РГРТУ: https://elib.rsreu.ru/ebs/
2. Электронное учебное пособие профессора И. Ю. Каширина.

6.5. Перечень программного обеспечения и информационных справочных систем

Лицензионное и свободно распространяемое программное обеспечение:

Наименование	Описание
Операционная система Windows	Коммерческая лицензия
Kaspersky Endpoint Security	Коммерческая лицензия
OpenOffice	Свободное ПО

Наименование	Описание
Python	Свободное ПО
Notepad++	Свободное ПО
Firefox	Свободное ПО
Adobe Acrobat Reader	Свободное ПО
LibreOffice	Свободное ПО
Visual Studio Community	Свободное ПО
Интерпретатор Python	Свободное ПО
Microsoft Visual Studio 12.0	Microsoft Imagine, номер подписки 700102019
Deductor Academic	Свободное ПО
Apache OpenOffice	Свободный пакет офисных приложений. Лицензия Apache License 2.0
Операционная система Windows XP/Vista/7/8/10	Microsoft Imagine: Номер подписки 700102019, бессрочно
Система программирования Microsoft Visual Studio 2010	Коммерческая лицензия
Java Runtime Environment	Свободное ПО
Protege	Свободное ПО

Наименование	Описание
СУБД MySQL	Свободное ПО
PHP	Свободное ПО
PyCharm	Свободное ПО
Visual Studio Code (язык C#)	Свободное ПО
Windows 10	Коммерческая лицензия

Информационные справочные системы:

1. Информационно-правовой портал ГАРАНТ.РУ: http://www.garant.ru
2. Система КонсультантПлюс: http://www.consultant.ru
3. Справочная правовая система «КонсультантПлюс» (договор об информационной поддержке №1342/455-100 от 28.10.2011 г.)