Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Государственный институт русского языка им. А.С. Пушкина»

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Наименование дисциплины (модуля):

Технологии автоматизированной генерации текстов

наименование дисциплины (модуля)

Уровень высшего образования:

бакалавриат

указывается: бакалавриат, магистратура

Направление подготовки:

45.03.01 Филология

(код и название направления подготовки)

Направленность (профиль) образования:

Русский язык как иностранный

Форма обучения:

очная

очная, очно-заочная, заочная

Рабочая программа дисциплины (модуля) разработана в соответствии с федеральным государственным образовательным стандартом высшего образования по направлению подготовки 45.03.01 «Филология», утвержденным приказом Минобрнауки России № 986 от 12.08.2020

1. Место дисциплины (модуля) в структуре ОПОП относится к части, формируемой участниками образовательных отношений ОПОП ВО Б1.В.ДВ.03.02

2. Цель и задачи освоения дисциплины (модуля)

Цель дисциплины: сформировать у студентов-лингвистов системное представление о технологиях автоматической генерации текстов как о практическом инструменте исследования и моделирования языка, позволяющем применять фундаментальные знания из области лексикологии, синтаксиса, семантики и дискурс-анализа для управления нейросетевыми языковыми моделями.

Задачи дисциплины:

1. **Проанализировать лингвистическую архитектуру языковых моделей.** Изучить, как фундаментальные уровни языка (от морфемы и синтаксической структуры до семантических ролей и прагматики дискурса) репрезентируются в векторном пространстве трансформеров. Научиться выявлять, как архитектура модели (например, механизм внимания) имитирует или упрощает реальные языковые процессы — управление анафорой, разрешение омонимии и сохранение синтаксической связности.
2. **Освоить методы лингвистического проектирования промптов.** Развить навык формализации лингвистических задач на языке инструкций для модели. Научиться управлять генерацией через контекстные примеры (few-shot), системные роли и лексико-грамматические ограничения, используя знания о валентностях глаголов, моделях управления и жанровых особенностях текста для получения предсказуемого и стилистически точного результата.
3. **Овладеть навыками лингвистической экспертизы и оценки качества машинного текста.** Научиться критически анализировать сгенерированные тексты с точки зрения языковой нормы и узуса. Выявлять специфические ошибки: морфологические (неверные флексии, агглютинации), синтаксические (нарушение порядка слов, неверное управление), семантические (нарушение лексической сочетаемости) и прагматические (потеря иллокутивной цели), а также фактологические галлюцинации, опираясь на компетенции лексикографа и редактора.

4. **Применить технологии генерации для решения прикладных лингвистических задач.** Использовать языковые модели как инструмент для автоматизации профессиональной деятельности: корпус-менеджмента (генерация и аугментация данных для обучения), лексикографии (составление словарных дефиниций и подбор контекстов), адаптации текстов (перевод на «простой язык», стилистическая трансформация) и создания диалоговых систем с заданными коммуникативными стратегиями.

3. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с требуемыми компетенциями выпускников

Компетенции выпускников (коды)	**Индикаторы достижения компетенций**	**Планируемые результаты обучения по дисциплине (модулю), соотнесенные с индикаторами достижения компетенций**
ПК-2 Способен дорабатывать, обрабатывать и создавать на основе стандартных методик и действующих нормативов различные типы текстов и документов в соответствии с отраслевыми, жанровыми, стилевыми требованиями и практическими	ПК – 2.1 Знает филологические основы текстообразования и нормативы создания текстов и документов в соответствии с отраслевыми, жанровыми, стилевыми требованиями и практическими задачами	• **Знать:** филологические основы текстообразования (категории связности, цельности, завершенности); нормативные требования к оформлению официально-деловых, научных и медийных текстов (ГОСТ, отраслевые стандарты); специфику влияния промпта на жанровые и стилистические характеристики генерируемого текста; лингвистические ограничения языковых моделей (длина контекста, разрушение когезии, стилистические сдвиги).

задачами, условиями реальной и виртуальной коммуникации		• **Уметь:** классифицировать профессиональные тексты по функциональным стилям и жанрам; определять нормативные требования к конкретному типу документа (например, к договору, пресс-релизу или аннотации); прогнозировать стилевые и структурные искажения, возникающие при машинной генерации в заданном жанре. • **Владеть:** терминологическим аппаратом лингвистики текста и функциональной стилистики для описания результатов машинной генерации; методикой лингвистического аудита сгенерированного контента на предмет соответствия жанровым нормам.
ПК-2 Способен дорабатывать, обрабатывать и создавать на основе стандартных методик и действующих	ПК-2.3 Осуществляет индексирование, аннотирование, реферирование текстов различной направленности, литературное редактирование и корректирование текстов различной степени глубины	• **Знать:** алгоритмы автоматического реферирования, суммаризации и извлечения ключевых слов; методы смыслового и формального индексирования текстов; уровни и стратегии литературного

нормативов различные типы текстов и документов в соответствии с отраслевыми, жанровыми, стилевыми требованиями и практическими задачами, условиями реальной и виртуальной коммуникации		редактирования (от light editing до deep editing); критерии оценки качества автоматического реферирования (сохранение фактологии, отсутствие искажений). • **Уметь:** составлять промпты для автоматического индексирования, аннотирования и реферирования текстов различной сложности; выполнять экспертное постредактирование сгенерированных аннотаций и рефератов, восстанавливая логические связи и устраняя фактические «галлюцинации»; проводить лексико-грамматическую и стилистическую корректуру текстов, прошедших через ИИ-обработку. • **Владеть:** навыками проектирования цепочек промптов (Prompt Chaining) для многоэтапной обработки текста (суммаризация → извлечение сущностей → стилистическая адаптация); методами глубокой редакторской правки

		машинного текста с восстановлением прагматики и авторского замысла; инструментами автоматизированной корректуры и проверки фактуры.

4. Объем дисциплины (модуля) _____2_____з.е., в том числе _____32_____академических часов на контактную работу обучающихся с преподавателем; _____38_____академических часов на самостоятельную работу обучающихся.

5. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов

Наименование и краткое содержание разделов и тем дисциплины (модуля). **Форма промежуточной аттестации по дисциплине (модулю)**	**Всего** **(часы)**	В том числе	
		Контактная работа **(работа во взаимодействии с преподавателем)** ***Виды контактной работы, часы***	**Самостоятельная работа обучающегося** ***Виды самостоятельной работы, часы***

		Занятия лекционного типа	Занятия семинарского типа (практические/ лабораторные)	Групповые консультации	Индивидуальная работа с обучающимися	**Всего**	**Тест**	**Конспект**	**КР**	**Всего**
Тема 1. Генеративный ИИ как новый языковой уровень (Текст как объект моделирования, семантика, прагматика, проблема «галлюцинаций» и референции).	8	2	2			4	2	2		4
Тема 2. Промпт-инжиниринг: проектирование речевого акта для машины (Синтаксис и семантика запроса, управление стилем, регистром и иллокутивной силой).	8	2	2			4	2	2		4
Тема 3. Лингвистический аудит машинного текста (Нарушение лексической и	8	2	2			4	2	2		4

грамматической нормы, стилистические диссонансы, разрушение когезии и когерентности).										
Тема 4. Профессиональное постредактирование (Стратегии deep/light editing, адаптация под терминосистемы, приведение к эталонному стилю).	10	2	2			4	2	2	2	6
Тема 5. Автоматизация лингвистических задач (Суммаризация, извлечение сущностей, машинный перевод и постредактура MTPE).	10	2	2			4	2	2	2	6
Тема 6. Работа с неструктурированными данными и метаданными (NER, классификация, тематическое моделирование,	10	2	2			4	2	2	2	6

генерация аннотаций и хэштегов).										
Тема 7. Проектная работа: разработка лингвистического ассистента (Проектирование цепочек промптов, сборка библиотеки промптов для кейса).	8	2	2			4	2	2		4
Тема 8. Лингвистика генеративного текста, этика и академические стандарты (Авторство, интертекстуальность, прагматика, этика данных)	10	2	2			4	2	2	2	6
Промежуточная аттестация *зачет*	*2*									
Итого	72	32					38			

Рабочая программа дисциплины «Технологии автоматизированной генерации текстов»

6. Фонд оценочных материалов по дисциплине (модулю)

Типовые задания и иные материалы, необходимые для оценки результатов обучения

Раздел 1. Примерные вопросы для тестового контроля (10 вопросов)

1. **Выберите наиболее точное определение феномена «галлюцинация» в контексте больших языковых моделей:** а) Случайное искажение токенов из-за ошибки квантования весов модели. б) Генерация правдоподобного по форме, но фактологически ложного или не подтвержденного в контексте утверждения. в) Нарушение синтаксической связности предложения при высокой температуре декодирования. г) Неверная референция местоимения к антецеденту в предыдущем предложении.
2. **Установите соответствие между лингвистическим уровнем и его репрезентацией в архитектуре Transformer:**
   1. Морфология и синтаксис
   2. Семантика и референция
   3. Прагматика и регистр а) Матрица внимания (Attention weights), фиксирующая синтаксические зависимости. б) Векторные смещения (steering vectors) и системные инструкции (system prompt). в) Семантическое пространство эмбеддингов, где близкие векторы соответствуют близким смыслам.
3. **Дайте определение понятию «Top-p (nucleus) sampling» в генеративных моделях.** а) Метод, при котором модель всегда выбирает слово с максимальной вероятностью. б) Метод декодирования, при котором отсекаются все варианты продолжения, суммарная вероятность которых не достигает порога p, что позволяет динамически менять размер словаря на каждом шаге. в) Фиксированное ограничение выбора только из k наиболее вероятных токенов. г) Увеличение вероятности редких слов для борьбы с повторениями.
4. **В чем заключается ключевое отличие экстрактивной суммаризации от абстрактивной с точки зрения лингвистической обработки?** а) Экстрактивная меняет порядок предложений, а абстрактивная — нет. б) Экстрактивная использует только исходные токены (фрагменты текста), тогда как абстрактивная генерирует новые лексические единицы

и синтаксические конструкции для передачи смысла. в) Экстрактивная всегда короче по объему, чем абстрактивная. г) Экстрактивная требует обучения с подкреплением, а абстрактивная — нет.

5. **Выберите верное утверждение о влиянии параметра «температура» (temperature) на генерацию текста:** а) Значение 0.0 делает текст максимально креативным и разнообразным. б) При повышении температуры распределение вероятностей следующего токена выравнивается, что увеличивает лексическое разнообразие и риск семантических ошибок. в) Температура управляет длиной генерируемого ответа. г) Температура влияет только на скорость инференса, но не на качество текста.

6. **Дайте определение «контекстному окну» (context window) языковой модели и объясните его лингвистическое ограничение.** а) Объем оперативной памяти видеокарты, выделенный под модель. б) Максимальное количество токенов (подслов), которое модель может удерживать в оперативной памяти при обработке одного запроса, что ограничивает анализ длинных дискурсивных связей и интертекстуальных отсылок. в) Количество примеров в few-shot промпте. г) Максимальный размер файла, который можно загрузить в модель.

7. **Какое лингвистическое явление чаще всего нарушается при «разрушении когезии» в машинном тексте?** а) Логическая связь между тезисом и аргументом (когерентность). б) Грамматическая и лексическая связность на поверхностном уровне: неверное согласование, разрыв анафорических цепочек (местоимений), нарушение эллипсиса. в) Стилистическое единство регистра. г) Фактологическая точность.

8. **Выберите стратегию «light editing» (поверхностного постредактирования) в отличие от «deep editing» (глубокого):** а) Полная переработка структуры текста и замена терминов. б) Исправление только критических ошибок, мешающих пониманию (опечатки, грубые фактические ошибки, нарушение базовой грамматики), без изменения авторского синтаксиса и стиля модели. в) Перевод текста на другой язык. г) Изменение тональности текста с нейтральной на рекламную.

9. **Дайте определение «инференсу» (inference) в контексте работы с языковыми моделями.** а) Процесс первоначального обучения модели на массиве данных (тренинг). б) Процесс применения обученной модели для получения ответа на новый запрос (генерация), не влекущий за собой изменения ее весов. в) Векторное представление слова в многомерном пространстве. г) Очистка данных перед подачей в модель.

10. **В чем заключается прагматическая проблема «потери иллокутивной силы» при генерации текста?** а) Модель генерирует текст, который грамматически верен, но не выполняет заложенную в запросе коммуникативную задачу (например, вместо убеждающей речи выдает сухую справку). б) Модель отказывается отвечать на вопросы из-за фильтров безопасности. в) Модель использует слишком сложную терминологию. г) Модель нарушает авторские права.

Ключи к тестовым вопросам

1. **б)** Генерация правдоподобного по форме, но фактологически ложного или не подтвержденного в контексте утверждения. *(Пояснение: вариант «а» описывает техническую ошибку, «в» — проблему декодирования, «г» — проблему разрешения анафоры, которая может быть следствием, но не определением галлюцинации).*
2. **1 — а**, **2 — в**, **3 — б**. *(Пояснение: синтаксические зависимости считываются через паттерны внимания; семантика зашита в плотные векторные пространства эмбеддингов; прагматика и стиль управляются через внешние инструкции и смещения векторов).*
3. **б)** Метод декодирования, при котором отсекаются все варианты продолжения, суммарная вероятность которых не достигает порога *p*, что позволяет динамически менять размер словаря на каждом шаге. *(Пояснение: вариант «а» — greedy search, «в» — top-k sampling).*
4. **б)** Экстрактивная использует только исходные токены (фрагменты текста), тогда как абстрактивная генерирует новые лексические единицы и синтаксические конструкции для передачи смысла. *(Пояснение: это фундаментальное различие в лингвистической обработке: копирование vs синтез).*
5. **б)** При повышении температуры распределение вероятностей следующего токена выравнивается, что увеличивает лексическое разнообразие и риск семантических ошибок. *(Пояснение: высокая температура делает выбор модели менее детерминированным).*
6. **б)** Максимальное количество токенов (подслов), которое модель может удерживать в оперативной памяти при обработке одного запроса, что ограничивает анализ длинных дискурсивных связей и интертекстуальных отсылок. *(Пояснение: это жесткое техническое ограничение, влияющее на глубину дискурс-анализа).*
7. **б)** Грамматическая и лексическая связность на поверхностном уровне: неверное согласование, разрыв анафорических цепочек (местоимений), нарушение эллипсиса. *(Пояснение: когезия — это «клей» на уровне формы текста; когерентность — на уровне смысла, упомянутая в варианте «а»).*
8. **б)** Исправление только критических ошибок, мешающих пониманию (опечатки, грубые фактические ошибки, нарушение базовой грамматики), без изменения авторского синтаксиса и стиля модели. *(Пояснение: deep editing подразумевает полную стилистическую и структурную переработку).*
9. **б)** Процесс применения обученной модели для получения ответа на новый запрос (генерация), не влекущий за собой изменения ее весов. *(Пояснение: ключевое отличие от обучения/дообучения — инференс не меняет параметры модели).*

10. **а)** Модель генерирует текст, который грамматически верен, но не выполняет заложенную в запросе коммуникативную задачу (например, вместо убеждающей речи выдает сухую справку). *(Пояснение: модель справилась с синтаксисом и семантикой, но провалилась на уровне прагматики).*

Балл	**Критерии оценивания**
5 (Отлично)	Обучающийся демонстрирует системное владение метаязыком лексикографии. Все ответы верифицированы. В задании на множественный выбор неверный вариант определен безошибочно, приведено краткое лингвистическое обоснование ошибочности тезиса (ссылка на теорию рода или хронологические границы словника).
4 (Хорошо)	Ответы в целом верифицированы. Допущена 1 ошибка в задании на установление соответствий (путаница смежных понятий) или дан верный ответ в тесте на исключение без развернутой лингвистической аргументации.
3 (Удовлетворительно)	Обучающийся демонстрирует знание базовых определений, однако допускает 2–3 ошибки в системных связях (например, путает задачи аспектных словарей). Отсутствует понимание дифференциальных признаков лексикографических типов.
2 (Неудовлетворительно)	Более 50% заданий выполнены неверно. Обучающийся не владеет базовой терминологией (не дифференцирует понятия «словник», «вокабула», «дефиниция»).

Конспектирование

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. — М.: МИЭМ, 2011 — 272 с. ISBN 978–5–94506–294–8

2. Тошович Б. Лингвистика, стилистика и поэтика виртуальности: монография. — Москва: ФЛИНТА, 2024. — 520 с. — ISBN 978-5-9765-5503-7.

Страницы для конспектирования: Главы «Генераторы» (с. 21–44), «Модели порождения текстов» (с. 61–112) и «Генераторская лингвистика» (с. 113–130). **Доступность:** Электронная версия доступна в ЭБС Znanium (znanium.ru) .

3. Грудева Е. В. Корпусная лингвистика: учебное пособие. — 4-е изд., стер. — Москва: ФЛИНТА, 2025. — 165 с. — ISBN 978-5-9765-1497-3.

Страницы для конспектирования: Темы, связанные с использованием корпусов как инструмента для обучения генеративных моделей.

Балл	Критерии оценивания
5 (Отлично)	**Полнота и системность.** Конспект охватывает все заявленные в плане страницы. Присутствует иерархическая структура: выделены основные главы, параграфы, ключевые определения и примеры. **Критичность.** В конспекте зафиксированы не только тезисы автора, но и границы их применимости. Присутствуют собственные микро-выводы или вопросы к тексту.
4 (Хорошо)	**Полнота с незначительными лакунами.** Охват материала составляет 85–90%. Пропущены второстепенные примеры или сноски, но основные определения и структура сохранены. **Точность терминологии.** Ключевые понятия (валентность, пресуппозиция, когезия) выписаны верно. Связь с ИИ указана, но носит характер простого сопоставления, а не глубокого анализа (например, просто пометка «важно для ИИ» без пояснения механизма). **Структура.** Текст структурирован (списки, подзаголовки), но иерархия выдержана не везде.
3 (Удовлетворительно)	**Репродуктивность.** Конспект представляет собой сплошной или слабо структурированный текст, близкий к пересказу или фрагментарному копированию (copy-paste) исходника. **Базовые знания.** Выписаны только прямые определения. Отсутствует понимание системных связей (например, студент выписал, что такое «анафора», но не связал это с ошибками языковых моделей). **Фактология.** Допущены 1–2 фактические ошибки при выписывании определений или искажен смысл авторского тезиса.
2 (Неудовлетворительно)	**Фрагментарность.** Охват материала менее 50%. Записаны только вводные или заключительные абзацы, отсутствует суть центральных глав. **Отсутствие метаязыка.** Студент не дифференцирует авторскую терминологию и общепринятый лингвистический метаязык. Понятия подменены бытовой лексикой. **Бессвязность.** Текст выглядит как набор разрозненных цитат без логической связи между ними. Отсутствуют любые попытки структурирования или осмысления прочитанного.

Контрольная работа

Задания выполняются письменно. Для заданий на создание промптов необходимо привести точную формулировку запроса и кратко обосновать выбор лингвистических маркеров.

Блок 1. Лингвистическая экспертиза (анализ готового текста)

Задание 1. Аудит когезии (поверхностной связности). *Текст для анализа:* «Нейросеть проанализировала рукопись. Она была написана в XIX веке. Исследователи долго спорили над ними, но они оказались слишком сложными для перевода». *Задача:* Найдите в тексте 3 случая нарушения когезии (разрыва анафорических цепочек, неверного

согласования или эллипсиса). Укажите номер предложения и тип ошибки. Предложите исправленный вариант. (Оценивается точность лингвистического диагноза).

Задание 2. Семантический аудит и фактчекинг (детекция галлюцинаций). *Текст для анализа:* «Александр Пушкин в 1829 году совершил путешествие в Японию, где изучал поэзию хокку. Под влиянием восточной эстетики он написал стихотворение "Я помню чудное мгновенье", заменив традиционную рифму белым стихом». *Задача:* Выпишите 2 фактологические галлюцинации и 1 лингвистическую (стилистическую) ошибку. Объясните, почему модель допустила стилистическую ошибку, опираясь на понятие «вероятностного распределения токенов». (Оценивается знание истории языка и понимание природы ошибок ИИ).

Задание 3. Синтаксический аудит (управление порядком слов). *Текст для анализа:* «Очень интересный и полезный для студентов, изучающих компьютерную лингвистику, был разработан новый алгоритм». *Задача:* Опишите тип синтаксической ошибки (смещение синтаксического дерева, нарушение валентностей). Предложите два варианта исправления: один с сохранением исходной лексики, второй — с изменением структуры предложения для улучшения читаемости. (Оценивается владение нормами русского синтаксиса).

Задание 4. Прагматический аудит (иллокутивная сила). *Текст для анализа:* «Ваше заявление получено. Обработка займет 30 дней. Ждите». *Задача:* Проанализируйте текст с точки зрения теории речевых актов. Какую коммуникативную неудачу (коммуникативный провал) совершает этот текст, если его цель — удержать клиента? Предложите 3 конкретных лингвистических изменения (замена лексики, модальность, структура), которые превратят этот текст в клиентоориентированный. (Оценивается понимание прагматики).

Блок 2. Проектирование промптов (Prompt Engineering)

Задание 5. Управление регистром и стилем (стилизация). *Задача:* Составьте промпт для генерации отрывка из протокола судебного заседания (официально-деловой стиль). В промпте должны быть использованы минимум 3 лингвистических ограничения: лексические (канцеляризмы), синтаксические (пассивный залог, цепочки существительных в родительном падеже) и прагматические (объективность, отсутствие эмотивов).

Задание 6. Контекстное обучение (Few-shot) для лексикографической задачи. *Задача:* Составьте промпт для создания словарной статьи неологизма «нейродеградация». Используйте формат Few-shot, приведя в промпте два примера существующих словарных статей (например, для слов «гаджет» и «дедлайн»). Ваша инструкция должна заставить модель сгенерировать: толкование, грамматические пометы, пример употребления и стилистическую помету.

Задание 7. Управление референцией и пресуппозициями. *Задача:* Вам нужно, чтобы модель написала текст о пользе редкого лекарственного растения, но строго избегала упоминания его названия (например, «зверобой»), чтобы не

нарушать закон о рекламе. Составьте промпт, который использует лингвистические механизмы управления референцией (замена на дескрипции: «это растение семейства зверобойных», пресуппозиции), чтобы модель не сгенерировала запрещенное слово.

Задание 8. Сложная суммаризация с изменением модальности. *Задача:* Дан фрагмент агрессивного поста из соцсетей (предоставляется преподавателем). Составьте промпт, который превратит этот текст в нейтральную новостную заметку. В промпте должны быть использованы инструкции по замене эмотивной лексики на фактологическую, трансформации прямой речи в косвенную и изменению синтаксиса с простого на сложноподчиненный.

Балл	Критерии оценивания
5 (Отлично)	Ответы характеризуются высокой степенью научной абстракции и верифицированы релевантным иллюстративным материалом. Обучающийся последовательно использует профессиональный лингвистический метаязык (диатеза, агенс, номинализация, пресуппозиция, актуальное членение). В заданиях на анализ текста продемонстрировано глубокое понимание механизмов: четко разграничены уровни языка (синтаксис, семантика, прагматика), выявлены глубинные причины ошибок генеративной модели. В задании на проектирование промптов (Задание 3) продемонстрировано владение синтаксическим и лексическим инструментарием для управления стилем.
4 (Хорошо)	Лингвистический анализ выполнен корректно, однако наблюдается смешение строгого научного и общеупотребительного метаязыка (например, использование описательных конструкций вместо терминов «диатеза» или «компонентный анализ»). В заданиях на анализ текста факты идентифицированы верно, но объяснение лингвистического механизма представлено без достаточной глубины или с незначительными неточностями в терминологии. В задании на промпты синтаксические маркеры подобраны верно, но их прагматический эффект описан неполно.
3 (Удовлетворительно)	Обучающийся идентифицировал языковые факты (нашел ошибки в тексте, привел примеры), однако не сумел эксплицировать лингвистический механизм их функционирования, подменяя анализ общими рассуждениями («звучит неестественно», «неправильный порядок слов»). Допущены существенные ошибки в терминологическом аппарате (смешение грамматических категорий, уровней языка или семантических ролей). В задании на промпты (Задание 3) предложены верные тексты, но лингвистическое обоснование их конструкции отсутствует или ошибочно.
2 (Неудовлетворительно)	Ответ носит репродуктивный (пересказный) или интуитивный характер. Лингвистический анализ подменен субъективной оценкой или рассуждениями, не имеющими отношения к метаязыку дисциплины. Отсутствует релевантный иллюстративный материал, либо он подобран некорректно (примеры не подтверждают тезис). В

Балл	Критерии оценивания
	заданиях на анализ допущены грубые фактические ошибки, демонстрирующие непонимание базовых синтаксических и семантических категорий.

Шкала и критерии оценивания

Текущий контроль успеваемости учитывает следующие виды текущей аттестации: выполнение обучающимся всех видов контрольных мероприятий, предусмотренных рабочей программой дисциплины, и самостоятельную работу обучающегося. Все формы текущего контроля и активность участия студентов в практических и (или) лабораторных занятиях, критерии их оценивания представлены в рабочей программе дисциплины, технологической карте и фонде оценочных материалов

В Институте используется балльно-рейтинговая система оценивания уровня знаний обучающихся (далее – БРС). При БРС оценивание уровня знаний обучающихся осуществляется по 5-балльной шкале.

0 - 2 баллов = «неудовлетворительно», «не зачтено»

3 балла = «удовлетворительно», «зачтено»

4 балла = «хорошо», «зачтено»

5 баллов = «отлично», «зачтено».

К промежуточной аттестации допускаются обучающиеся, выполнившие необходимые тесты, контрольные и самостоятельные контрольные работы в ходе проведения текущего контроля успеваемости. В случае наличия учебной задолженности обучающийся отрабатывает пропущенные занятия в форме, предложенной преподавателем и представленной в настоящей рабочей программе дисциплины в разделе самостоятельная работа.

Промежуточная аттестация осуществляется в форме экзамена.

Результаты обучения оцениваются: *«отлично» / «хорошо» / «удовлетворительно» / «неудовлетворительно»*.

7. Методические материалы

7.1. Перечень основной и дополнительной учебной литературы

Основная литература

1. **Лукашевич Н.В., Сорокин А.А. Компьютерная лингвистика и автоматическая обработка текстов.** — Москва: МАКС Пресс, 2025. — 608 с. — ISBN 978-5-317-07437-1.

Аннотация: Базовый учебник для студентов, обучающихся по специальностям «Фундаментальная и прикладная лингвистика» и «Прикладная математика и информатика». Подробно описаны методы обработки текста, включая нейросетевые подходы, архитектуру трансформер и большие языковые модели. Учебник доступен для чтения и скачивания на платформе «Истина» МГУ .

2. **Зильберман Н.Н. Нейросети в педагогическом дизайне: практика эффективного промтинга для создания текстовых учебных материалов.** — Томск: Издательство Томского государственного университета, 2026. — 64 с. — ISBN 978-5-908040-65-5. *Аннотация*: Практическое пособие, ориентированное на создание текстового образовательного контента с помощью нейросетей. Рассматриваются типы контента, базовые и продвинутые техники промтинга, а также ограничения моделей. Пособие доступно в электронной библиотеке ТГУ .

3. **Бессмертный И.А., Нугуманова А.Б., Платонов А.В. Интеллектуальные системы: учебник и практикум для вузов.** — Москва: Издательство Юрайт, 2024. — 243 с. — ISBN 978-5-534-01042-8. *Аннотация*: Учебник охватывает разделы, связанные с естественными языками и машинным анализом текстов, а также моделированием текстов на естественном языке. Доступен на образовательной платформе Юрайт .

4. **Воронов М.В., Пименов В.И., Небаев И.А. Системы искусственного интеллекта: учебник и практикум для вузов.** — Москва: Издательство Юрайт, 2024. — 268 с. — ISBN 978-5-534-17032-0. *Аннотация*: Издание включает материалы по моделированию текстов на естественном языке и генеративным технологиям. Доступно на платформе Юрайт .

🕮 Дополнительная литература

1. **Станкевич Л.А. Интеллектуальные системы и технологии: учебник и практикум для вузов.** — 2-е изд., перераб. и доп. — Москва: Издательство Юрайт, 2024. — 495 с. — ISBN 978-5-534-16238-7. *Аннотация*: Содержит разделы, посвящённые моделированию текстов на естественном языке. Доступен на платформе Юрайт .

2. **Кудрявцев В.Б., Гасанов Э.Э., Подколзин А.С. Интеллектуальные системы: учебник и практикум для вузов.** — 2-е изд., испр. и доп. — Москва: Издательство Юрайт, 2024. — 165 с. — ISBN 978-5-534-07779-7. *Аннотация*: Учебник затрагивает теоретические аспекты интеллектуальных систем, релевантные для понимания принципов автоматической генерации. Доступен на платформе Юрайт .

3. **Пензенский государственный университет. Технологии искусственного интеллекта: учебное пособие.** — Пенза: Изд-во ПГУ, 2025. — ISBN 978-5-908089-22-7.
Аннотация: Пособие включает разделы по большим языковым моделям, эволюции GPT, промпт-инжинирингу и интеллектуальным агентам. Доступно в электронной библиотеке ПГУ .

4. **Haralambous, Yannis. A Course in Natural Language Processing.** — Cham: Springer, 2024. — 534 p. — ISBN 978-3-031-27226-4.
Аннотация: Фундаментальный курс, рассматривающий NLP с трёх точек зрения: лингвистика, символический ИИ и статистические методы. Подходит для углублённого изучения теоретических основ генерации текста .

🔍 Ресурсы для поиска источников

Для самостоятельного поиска актуальных публикаций рекомендуется использовать **Научную электронную библиотеку** eLIBRARY.RU и **ЭБС «Лань»**, в которой доступен генеративный сервис **ЛаньGPT**, работающий исключительно с верифицированным контентом библиотеки .

7.2. Перечень лицензионного программного обеспечения

1. Microsoft Windows 10.
2. Microsoft Office 2016 стандартная.
3. Big Blue Button

8.1. Перечень профессиональных баз данных и информационных справочных систем

1. СЛОВАРИ.РУ — портал, объединяющий словари русского языка различных типов. Режим доступа: https://slovari.ru

2. Грамота.РУ — справочно-информационный портал по русскому языку, включающий словарную базу и консультации. Режим доступа: https://gramota.ru

3. Академик — словарно-энциклопедический портал, предоставляющий доступ к толковым, орфографическим, этимологическим и другим словарям. Режим доступа: https://academic.ru

4. Национальный корпус русского языка — информационно-справочная система, обеспечивающая доступ к текстам на русском языке с лингвистической разметкой. Режим доступа: https://ruscorpora.ru

8.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Институт лингвистических исследований РАН — официальный сайт, содержащий информацию о словарных проектах и лексикографических исследованиях. Режим доступа: https://iling.spb.ru
2. Институт русского языка им. В. В. Виноградова РАН — официальный сайт с ресурсами по русской лексикографии и словарными материалами. Режим доступа: https://ruslang.ru
3. Электронно-библиотечная система Znanium — база данных, включающая учебные и научные издания по лексикографии. Режим доступа: https://znanium.ru

9. Язык преподавания русский

10. Разработчики программы: Федорова Е.Л., кандидат филологических наук, доцент

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Государственный институт русского языка им. А.С. Пушкина»

ФОНД ОЦЕНОЧНЫХ МАТЕРИАЛОВ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

Технологии автоматизированной генерации текстов
(название дисциплины (модуля))

Код и наименование направления подготовки: 45.03.01 Филология

Уровень высшего образования бакалавриат

Направленность (профиль) образования: Русский язык как иностранный

Форма (формы) обучения: очная

Квалификация: бакалавр

г. Москва

1. ОЦЕНИВАНИЕ РЕЗУЛЬТАТОВ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель фонда оценочных материалов – контроль и оценка образовательных достижений обучающихся в ходе освоения рабочей программы дисциплины «Технологии автоматизированной генерации текстов».

Цель текущей аттестации – обеспечение своевременной обратной связи, коррекция обучения, активизация самостоятельной работы обучающихся.

Цель промежуточной аттестации – оценивание промежуточных и окончательных результатов обучения по дисциплине, обеспечивающих достижение планируемых результатов освоения ОПОП.

2. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕРКИ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ

№ п/п	Содержание задания	Шифр индикатора достижения компетенции (ИДК)
Раздел 1. Лингвистические основы генеративного ИИ		
Тема 1. Генеративный ИИ как новый языковой уровень		
1.	**Вопросы для устного опроса:** – Как архитектура Transformer решает проблему контекстной зависимости (длинной дистанции) в отличие от рекуррентных сетей? – В чем заключается лингвистическая природа «галлюцинаций» при генерации референциальных цепочек (местоимений и анафоры)? – Почему LLM не «понимает» смысл слова, а оперирует его векторным представлением?	ПК-2.1
2.	**Тест:** 1. Архитектура Transformer базируется на:а) Последовательном чтении текста слева направо.б) Механизме внимания (Attention), позволяющем взвешивать значимость всех слов контекста одновременно.в) Только на рекуррентных связях.2. С точки зрения лингвистики, «галлюцинация» — это:а) Сбой в работе сервера.б) Генерация правдоподобной, но фактологически ложной информации, нарушающей максиму качества (Грайс).в) Намеренная дезинформация.	ПК-2.1
Тема 2. Промпт-инжиниринг: проектирование речевого акта		
3.	**Вопросы для устного опроса:** – Как синтаксическая структура промпта (сложноподчиненные предложения vs простые перечисления) влияет на иерархию приоритетов в ответе модели? – Каким образом лингвистические маркеры вежливости и модальные глаголы в промпте меняют иллокутивную силу генерируемого текста? – В чем разница между Zero-shot и Few-shot промптом с точки зрения объема передаваемой пресуппозиции?	ПК-2.1

№ п/п	Содержание задания	Шифр индикатора достижения компетенции (ИДК)
4.	**Тест:** 1. Параметр «Top-p (nucleus sampling)» в декодировании:а) Ограничивает выбор слова только из *k* самых вероятных вариантов.б) Динамически формирует словарь из наиболее вероятных токенов, сумма вероятностей которых равна *p*.в) Увеличивает длину ответа.2. Техника «Chain of Thought» (CoT) в промптинге направлена на:а) Сокращение времени генерации.б) Вербализацию промежуточных логических шагов для повышения точности решения сложных задач.в) Изменение языка вывода.	ПК-2.1
Раздел 2. Экспертиза и редактирование машинного текста		
Тема 3. Лингвистический аудит машинного текста		
5.	**Вопросы для устного опроса:** – Как нарушение валентности (селективных ограничений) глагола помогает отличить машинный текст от человеческого? – Приведите пример разрушения когезии (формальной связи) при сохранении когерентности (смыслового единства) в сгенерированном тексте. – Какова роль просодии (интонации) в восприятии сгенерированного устного текста (TTS) и как она соотносится с пунктуацией в промпте?	ПК-2.3
6.	**Тест (анализ фрагментов):** 1. Фрагмент: «Мы провели высокую температуру в колбе». Ошибка:а) Синтаксическая.б) Лексическая (нарушение селективных ограничений: температура нельзя «провести»). в) Орфографическая. 2. Фрагмент: «М. Ю. Лермонтов написал "Героя нашего времени". Он умер в 1856 году». Ошибка: а) Фактологическая (нарушение хронологической референции). б) Грамматическая.в) Стилистическая.	ПК-2.3
Тема 4. Профессиональное постредактирование		
7.	**Вопросы для устного опроса:** – Какова лингвистическая методика проверки фактологической точности (фактчекинга) при стратегии Deep Editing? – В чем заключается проблема «галлюцинации ссылок» и как ее выявлять на этапе редактирования? – Как адаптировать машинный текст под узкую терминосистему (например, юриспруденцию), не потеряв при этом общую связность?	ПК-2.3
8.	**Тест (анализ стратегий):** 1. Стратегия, при которой исправляются только критические ошибки, мешающие пониманию, без изменения авторского синтаксиса модели: а) Deep editing.б) Light editing.в) Рефакторинг. 2. Ошибка в предложении «Подлежащее выражено существительным, а сказуемое — является глаголом» (избыточность «а также» или тавтология конструкций): а) Лексическая. б) Синтаксическая (нанизывание однотипных конструкций). в) Пунктуационная.	ПК-2.3
Раздел 3. Прикладная автоматизация и проектная работа		
Тема 5. Автоматизация прикладных лингвистических задач		
9.	**Вопросы для устного опроса:** – Как лингвистическая разметка (dependency parsing) помогает модели лучше справляться с	ПК-2.1

№ п/п	Содержание задания	Шифр индикатора достижения компетенции (ИДК)
	задачами суммаризации длинных текстов? – В чем заключается главная проблема машинного перевода промптов (Prompt Translation) при переносе инструкций с русского на английский? – Какую роль играют тезаурусы и онтологии в снижении предвзятости (bias) генеративных моделей?	
10.	**Тест:** 1. Задача извлечения именованных сущностей (NER) относится к уровню:а) Морфологического анализа.б) Информационного поиска и семантической обработки. в) Фонологического анализа. 2. Проблема «статистического эха» (bias) в генеративном тексте проявляется как: а) Опечатки в редких словах. б) Воспроизведение моделью скрытых стереотипов и частотных паттернов из обучающих данных (например, «программист — всегда мужчина»). в) Слишком медленный ответ сервера.	ПК-2.1

Раздел 4. Практические задания (для письменной части зачета/экзамена)

№	Содержание задания	Шифр ИДК
11.	**Проектирование промпта (ПК-2.1).** Составьте промпт для генерации текста в жанре «научная рецензия». В промпте должны быть жестко заданы: 1) синтаксическая структура (использование пассивного залога и безличных конструкций); 2) лексические ограничения (запрет на оценочную лексику, использование только академических глаголов: *аргументирует*, *констатирует*, *подвергает сомнению*); 3) формат вывода (таблица с критериями оценки).	ПК-2.1
12.	**Лингвистический аудит (ПК-2.3).** Проанализируйте предложенный преподавателем сгенерированный текст (объем 150–200 слов). Составьте таблицу ошибок: 1) Нарушение лексической сочетаемости (3 примера); 2) Анафорический провал (2 примера); 3) Нарушение логической когерентности (1 пример). Для каждой ошибки укажите номер предложения и предложите вариант исправления.	

Ключи к оценочным материалам

Раздел 1. Лингвистические основы генеративного ИИ

Тема 1. Генеративный ИИ как новый языковой уровень

1. **Устный опрос:**

- *Архитектура:* Transformer использует механизм внимания (Self-Attention) для параллельного взвешивания связей между всеми токенами, что позволяет моделировать зависимости на неограниченной дистанции, в отличие от RNN, обрабатывающих текст последовательно и теряющих контекст.
- *Галлюцинации:* С точки зрения лингвистики — это нарушение референции. Модель генерирует объект (сущность), который не был введен в дискурс, или приписывает

ему свойства, нарушающие максиму качества (Грайс), так как оперирует вероятностными связями токенов, а не реальным положением дел.

- *Понимание:* Модель не оперирует семантическими примитивами или образами, а вычисляет наиболее вероятную последовательность векторных представлений (эмбеддингов) на основе статистических закономерностей корпуса.

2. **Тест:**
   - 1 — б (Статистические закономерности распределения лексем в корпусе).
   - 2 — б (Нарушение референции и максимы качества).

Тема 2. Промпт-инжиниринг: проектирование речевого акта 3. **Устный опрос:**

- *Компоненты:* Role (Роль эксперта), Context (Фоновые знания и вводные данные), Task (Четкая инструкция/речевой акт), Format (Структура вывода: JSON, таблица, список).
- *Функция маркеров:* Модальные слова и вводные конструкции («Вероятно», «С точки зрения», «Важно отметить») задают прагматическую рамку и модальность текста, управляя иллокутивной силой (убеждение, информирование, сомнение) и регистром (формальный/неформальный).

4. **Тест:**
   - 1 — б (Креативность и разнообразие).
   - 2 — б (Принуждает модель к пошаговому логическому выводу).

Раздел 2. Экспертиза и редактирование машинного текста

Тема 3. Лингвистический аудит и типология ошибок 5. **Устный опрос:**

- *Отличие:* Лексическая несочетаемость (оксюморон, семантический конфликт) — это конфликт значений целых слов («фиолетовая идея»). Нарушение валентности — это невыполнение глаголом своих требований к актантам (например, переходный глагол употреблен без прямого объекта: *«Он читает»* вместо *«читает книгу»*, или нарушение семантических ролей).
- *Когезия vs Когерентность:* Разрушение когезии — это формальный разрыв (неверное местоимение, нарушение согласования: *«Дом старый. Она скрипит»*). Разрушение когерентности — текст грамматически гладкий, но логически абсурдный (смыслы предложений не складываются в единую картину).

6. **Тест:**
   - 1 — б (Лексическая семантика: нарушение валентности/селективных ограничений; «провести» требует одушевленного деятеля или мероприятия).
   - 2 — б (Референциальная связь: анафорический провал; местоимение «Она» не может относиться к слову «Собор» в строгой норме, хотя допустимо в разговорной, но здесь также конфликт с полом слова «храм» — «Он»).

Тема 4. Профессиональное постредактирование 7. **Устный опрос:**

- *Фактчекинг:* Сверка каждого факта (даты, имена, цифры, цитаты) с независимыми авторитетными источниками (первоисточники, энциклопедии), так как модель склонна к «галлюцинации» правдоподобных, но ложных данных.
- *Разница:* Адаптация под терминосистему — это замена лексики на узкоспециализированные эквиваленты (динамик — громкоговоритель, в зависимости от

ГОСТа). Стилистическая правка — это изменение синтаксиса, ритма и регистра (превращение канцелярита в публицистику).

8. **Тест:**
   - 1 — б (Deep editing).
   - 2 — а (Синтаксическая норма: нанизывание падежей и конструкций; смешение управления: «овладение (чем?) компетенциями» и «для того, чтобы (что сделать?)»).

Раздел 3. Прикладная автоматизация и проектная работа

Тема 5. Автоматизация прикладных лингвистических задач 9. **Устный опрос:**

- *Роль аннотаций:* Парсинг (синтаксические деревья) и NER (разметка сущностей) служат «сверхточным сигналом» (supervision), обучая модель структуре языка и связям между словами, что критически важно для задач извлечения информации (QA).
- *Проблема MTPE:* Диссоциация эквивалентов. Модель может перевести прецизионный термин (например, юридический или медицинский) общеупотребительным словом, потеряв специфику, так как в обучающих данных частотность общеупотребительного слова выше.

10. **Тест:**
    - 1 — б (Семантики и информационного поиска).
    - 2 — б (Воспроизведение моделью скрытых стереотипов и частотных паттернов из данных).

Раздел 4. Практические задания

11. **Проектирование промпта (Пример эталона):**
    - *Текст промпта:* «Ты — строгий научный рецензент. Проанализируй статью. Твоя задача — оценить аргументацию. Правила вывода:
      1. Используй только пассивный залог и безличные конструкции (например: "Утверждается, что...", "Было выявлено...").
      2. Запрещены оценочные слова ("хорошо", "плохо", "интересно"). Используй только академические глаголы: "аргументирует", "констатирует", "игнорирует", "подвергает сомнению".
      3. Выведи результат строго в виде таблицы Markdown с колонками: "Критерий", "Оценка", "Цитата-иллюстрация"».
12. **Лингвистический аудит (Пример эталона):**
    - *Ожидаемая структура ответа:* Таблица из 3 колонок (№, Тип ошибки, Пример из текста -> Исправленный вариант).
    - *Анафорический провал:* «Нейросеть обучалась на корпусе. Они [вместо "Она"] показала высокую точность». -> «Она показала...».
    - *Нарушение валентности:* «Мы провели высокую температуру». -> «Мы нагрели образец до высокой температуры» / «Мы зафиксировали высокую температуру».
    - *Нарушение когерентности:* Текст, где предложения грамматически верны, но вывод не следует из посылки (ошибка логики, которую студент должен выявить и переписать связный фрагмент).

3. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ (зачет)

Промежуточная аттестация осуществляется в форме зачета. Способ проведения: устный (по билетам).

3.1. Перечень вопросов к зачету

№ п/п	Вопросы к зачету	Шифр ИДК
Раздел 1. Теоретические и технологические основы		
1	Объект, предмет и задачи дисциплины. Разграничение понятий «автоматизация» и «генерация» текста.	ПК-2.1, ПК-2.3
2	История развития технологий генерации текста: от шаблонных систем к нейросетевым моделям.	ПК-2.1, ПК-2.3
3	Архитектура генеративных моделей (Трансформеры, механизм внимания). Лингвистическая интерпретация «понимания» контекста.	ПК-2.1, ПК-2.3
4	Понятие токенизации и векторного представления слов (эмбеддинги). Лингвистические проблемы морфологической сегментации.	ПК-2.1, ПК-2.3
5	Феномен «галлюцинации» языковой модели: лингвистическая природа (нарушение референции, семантические аномалии).	ПК-2.1, ПК-2.3
6	Понятие промпта как искусственно сконструированного дискурса. Синтаксис и прагматика запроса.	ПК-2.1, ПК-2.3
7	Управление стилем и регистром через лингвистические маркеры в промпте (модальность, вводные конструкции).	ПК-2.1, ПК-2.3
8	Параметры генерации (температура, Top-p) и их влияние на лексико-семантическое разнообразие текста.	ПК-2.1, ПК-2.3
9	Контекстное окно и проблема разрушения когезии и когерентности на больших расстояниях.	ПК-2.1, ПК-2.3
10	Типология ошибок генеративного ИИ: лексические (валентность), грамматические и стилистические.	ПК-2.1, ПК-2.3

№ п/п	Вопросы к зачету	Шифр ИДК
11	Разрушение анафорических связей (местоимений) как референциальный сбой модели.	ПК-2.1, ПК-2.3
12	Нарушение лексической сочетаемости и семантических ограничений (селективных ограничений) слов.	ПК-2.1, ПК-2.3
13	Стилистические диссонансы: смешение регистров и разрушение тональности (Tone of Voice).	ПК-2.1, ПК-2.3
14	Стратегии постредактирования: Light editing vs Deep editing. Лингвистические критерии выбора.	ПК-2.1, ПК-2.3
15	Адаптация машинного текста под узкопрофессиональные терминосистемы (юридическая, медицинская).	ПК-2.1, ПК-2.3
16	Восстановление прагматики и авторского замысла в процессе редакторской правки.	ПК-2.1, ПК-2.3
17	Академическая этика использования ИИ. Правила цитирования сгенерированного текста (ГОСТ, ВАК).	ПК-2.1, ПК-2.3
Раздел 2. Прикладная генеративная лингвистика		
18	Автоматическое реферирование (суммаризация). Лингвистические критерии сохранения фактологии.	ПК-2.1, ПК-2.3
19	Извлечение именованных сущностей (NER) и атрибутов. Лингвистическая классификация объектов.	ПК-2.1, ПК-2.3
20	Машинный перевод и постредактирование (MTPE). Проблема «ложных друзей переводчика» в ИИ.	ПК-2.1, ПК-2.3
21	Генерация параллельных текстов для преподавания иностранных языков. Методический потенциал.	ПК-2.1, ПК-2.3
22	Классификация текстов через естественный язык (Zero-shot classification). Лингвистические маркеры категорий.	ПК-2.1, ПК-2.3

№ п/п	Вопросы к зачету	Шифр ИДК
23	Тематическое моделирование через генеративные модели (вместо LDA). Семантические поля.	ПК-2.1, ПК-2.3
24	Цепочки промптов (Prompt Chaining) для многоэтапной обработки текста.	ПК-2.1, ПК-2.3
25	Работа с API языковых моделей. Формат запроса и парсинг лингвистически значимых данных.	ПК-2.1, ПК-2.3
26	Интеграция ИИ в офисные пакеты (Word, Excel) для автоматизации рутинных текстовых операций.	ПК-2.1, ПК-2.3
27	Проблема авторства в цифровую эпоху: промптер, модель, редактор.	ПК-2.1, ПК-2.3
28	Интертекстуальность машинного творчества. Понятие «статистического эха» обучающей выборки.	ПК-2.1, ПК-2.3
29	Прагматика и иллокутивные акты. Может ли машина совершить истинный речевой акт?	ПК-2.1, ПК-2.3
30	Дискриминация и этика данных (Data Ethics). Лингвистические маркеры скрытой предвзятости (bias).	ПК-2.1, ПК-2.3
31	Трансформация профессиональных компетенций лингвиста: от автора к AI-тренеру и лингво-аудитору.	ПК-2.1, ПК-2.3
32	Промпт как гибридный речевой жанр. Структура и конвенции.	ПК-2.1, ПК-2.3
33	Диалоговый след (chat log) как объект дискурс-анализа.	ПК-2.1, ПК-2.3
34	Сравнительный анализ эффективности ручного и автоматизированного создания текстов (метрики качества).	ПК-2.1, ПК-2.3

№ п/п	Вопросы к зачету	Шифр ИДК
35	Верификация фактологической точности сгенерированного текста. Методология фактчекинга.	ПК-2.1, ПК-2.3
36	Проектирование библиотеки промптов для решения конкретной профессиональной задачи.	ПК-2.1, ПК-2.3
37	Оценка качества генерации: метрики BLEU, ROUGE, BERTScore и их лингвистическая интерпретация.	ПК-2.1, ПК-2.3
38	Футурология лингвистики: новые объекты исследования в эпоху генеративного ИИ.	ПК-2.1, ПК-2.3

Краткие ответы на вопросы к зачету

Раздел 1. Теоретические и технологические основы

1. Объект, предмет и задачи дисциплины. Разграничение понятий «автоматизация» и «генерация» текста. Объект — процессы создания текстов с помощью компьютерных технологий. Предмет — лингвистические и технологические аспекты автоматизированной генерации. «Автоматизация» — выполнение рутинных операций по заданному алгоритму (шаблоны, подстановки). «Генерация» — создание нового текста на основе вероятностных моделей без жёстких шаблонов.

2. История развития технологий генерации текста: от шаблонных систем к нейросетевым моделям. Этапы: 1) шаблонные системы (ELIZA, 1966); 2) статистические модели (n-граммы, HMM); 3) нейросетевые модели (RNN, LSTM); 4) трансформеры (2017) и большие языковые модели (GPT, BERT, LLaMA).

3. Архитектура генеративных моделей (Трансформеры, механизм внимания). Лингвистическая интерпретация «понимания» контекста. Трансформер использует механизм self-attention, позволяющий модели учитывать связи между всеми словами последовательности. «Понимание» контекста — это статистическое взвешивание вероятностей совместной встречаемости токенов, а не семантическое осмысление в человеческом смысле.

4. Понятие токенизации и векторного представления слов (эмбеддинги). Лингвистические проблемы морфологической сегментации. Токенизация — разбиение текста на минимальные единицы (токены). Эмбеддинги — числовые векторы, отражающие

семантическую близость слов. Проблема: для русского языка морфологическая сегментация сложна из-за флективности и омонимии.

5. Феномен «галлюцинации» языковой модели: лингвистическая природа. Галлюцинация — порождение фактически неверной информации при сохранении формальной правильности текста. Лингвистические причины: нарушение референции, ложная кореферентность, семантические аномалии, подмена фактов статистически вероятными.

6. Понятие промпта как искусственно сконструированного дискурса. Синтаксис и прагматика запроса. Промпт — текст-инструкция, адресованная модели. Обладает синтаксической структурой (императивы, вопросительные конструкции) и прагматикой (цель, адресат, контекст). Является гибридным речевым жанром.

7. Управление стилем и регистром через лингвистические маркеры в промпте. Стиль задаётся через: модальность («напиши строго», «вырази сомнение»), вводные конструкции («с точки зрения науки»), лексические маркеры («официально», «разговорно»), указание тональности.

8. Параметры генерации (температура, Top-p) и их влияние на лексико-семантическое разнообразие. Температура регулирует случайность выбора: низкая — предсказуемые, шаблонные тексты; высокая — креативные, но рискованные. Top-p ограничивает выборку наиболее вероятными токенами, балансируя разнообразие и связность.

9. Контекстное окно и проблема разрушения когезии и когерентности на больших расстояниях.

Контекстное окно — максимальный объём текста, который модель удерживает. При превышении модель «забывает» начало, разрушаются анафорические связи, теряется логическая связность.

10. Типология ошибок генеративного ИИ. Лексические (нарушение валентности, ложная сочетаемость), грамматические (согласование, управление), стилистические (смешение регистров), семантические (аномалии), прагматические (несоответствие цели).

11. Разрушение анафорических связей как референциальный сбой модели. Модель может использовать местоимение без антецедента, менять референт или дублировать подлежащее. Причина — потеря информации за пределами контекстного окна.

12. Нарушение лексической сочетаемости и семантических ограничений. Модель порождает сочетания, невозможные в языке (*«сильный дождь» вместо «сильный ливень»*). Причина — статистическая генерация без учёта селективных ограничений.

13. Стилистические диссонансы: смешение регистров и разрушение тональности. В одном тексте могут сочетаться официально-деловые и разговорные элементы, научные и публицистические. Tone of Voice нарушается при смене эмоциональной окраски.

14. Стратегии постредактирования: Light editing vs Deep editing. Light editing — минимальная правка (орфография, пунктуация, отдельные лексемы). Deep editing — переработка структуры, стиля, фактологии. Выбор зависит от цели текста и степени его публичности.

15. Адаптация машинного текста под узкопрофессиональные терминосистемы. Требуется верификация терминов, замена ложных синонимов, приведение к стандартам ГОСТ, отраслевым регламентам. Для юридических и медицинских текстов — обязательная экспертная проверка.

16. Восстановление прагматики и авторского замысла в процессе редакторской правки.

Редактор восстанавливает цель высказывания, интенцию автора, адресность. Устраняет «статистическое эхо», придаёт тексту индивидуальность.

17. Академическая этика использования ИИ. Правила цитирования сгенерированного текста. Согласно ГОСТ и требованиям ВАК, сгенерированный текст не является объектом авторского права. При использовании ИИ необходимо указывать модель, промпт и дату генерации. Плагиатом считается выдача машинного текста за собственный.

Раздел 2. Прикладная генеративная лингвистика

18. Автоматическое реферирование (суммаризация). Extractive — выбор ключевых предложений из исходного текста. Abstractive — порождение нового сжатого текста. Критерии: сохранение фактологии, устранение повторов, сохранение логики.

19. Извлечение именованных сущностей (NER) и атрибутов. NER — распознавание имён, организаций, локаций, дат. Лингвистическая классификация: одушевлённые/неодушевлённые, конкретные/абстрактные, собственные/нарицательные.

20. Машинный перевод и постредактирование (MTPE). Проблема «ложных друзей переводчика».

MTPE — редактирование машинного перевода. «Ложные друзья» — слова, сходные по форме, но разные по значению (*«академик» — academic*). Требуется верификация по словарям.

21. Генерация параллельных текстов для преподавания иностранных языков. Создание адаптированных текстов одного содержания на разных языках. Методический потенциал: обучение переводу, сопоставительному анализу, лексике.

22. Классификация текстов через естественный язык (Zero-shot classification). Модель классифицирует текст без предварительного обучения на размеченных данных. Лингвистические маркеры: ключевые слова, тематическая лексика, стилевые особенности.

23. Тематическое моделирование через генеративные модели. Вместо LDA используется кластеризация эмбеддингов или генерация описаний тем. Семантические поля выявляются через векторную близость.

24. Цепочки промптов (Prompt Chaining). Последовательность запросов, где результат одного становится входом для другого. Позволяет выполнять многоэтапную обработку: анализ → план → черновик → редактура.

25. Работа с API языковых моделей. Запрос формируется в формате JSON: модель, сообщения, параметры. Парсинг ответа извлекает текст, метаданные, вероятности. Лингвистически значимые данные — токены, эмбеддинги.

26. Интеграция ИИ в офисные пакеты. Word: генерация текста, реферирование, перевод. Excel: генерация формул, анализ данных, создание отчётов. Автоматизация рутинных операций.

27. Проблема авторства в цифровую эпоху. Автор — промптер (формулирует задачу), модель (генерирует), редактор (верифицирует). Юридически автор — человек, но вклад модели требует указания.

28. Интертекстуальность машинного творчества. Понятие «статистического эха». Модель воспроизводит фрагменты обучающей выборки. «Статистическое эхо» — непроизвольное цитирование без ссылки. Требуется проверка на плагиат.

29. Прагматика и иллокутивные акты. Может ли машина совершить истинный речевой акт? Иллокутивный акт предполагает намерение и ответственность. Машина имитирует речевой акт, но не несёт ответственности. Истинный акт совершает человек, использующий машину.

30. Дискриминация и этика данных. Лингвистические маркеры скрытой предвзятости.

Bias проявляется в стереотипных ассоциациях (профессия + гендер), оценочной лексике, умолчаниях. Требуется аудит данных и промптов.

31. Трансформация профессиональных компетенций лингвиста. Новые роли: AI-тренер (обучение моделей), промпт-инженер, лингво-аудитор (верификация), редактор машинных текстов, эксперт по этике.

32. Промпт как гибридный речевой жанр. Сочетает черты инструкции, вопроса, команды. Структура: контекст + задача + ограничения + формат. Конвенции: чёткость, однозначность, вежливость.

33. Диалоговый след (chat log) как объект дискурс-анализа. Изучается структура диалога, стратегии уточнения, речевые акты, ошибки. Позволяет выявить типичные проблемы взаимодействия человека и модели.

34. Сравнительный анализ эффективности ручного и автоматизированного создания текстов.

Метрики: скорость, стоимость, качество (BLEU, ROUGE, экспертные оценки). Автоматизация быстрее и дешевле, но требует верификации.

35. Верификация фактологической точности. Методология фактчекинга. Проверка фактов по авторитетным источникам, базам данных, первоисточникам. Автоматические методы: сопоставление с Knowledge Graph.

36. Проектирование библиотеки промптов. Систематизация промптов по задачам: генерация, реферирование, перевод, классификация. Включает шаблоны, переменные, примеры, ограничения.

37. Оценка качества генерации: метрики BLEU, ROUGE, BERTScore. BLEU — совпадение n-грамм с эталоном. ROUGE — полнота и точность для реферирования. BERTScore — семантическая близость через эмбеддинги.

38. Футурология лингвистики: новые объекты исследования. Промпт-дискурс, машинные тексты, диалоговые следы, гибридные жанры, этика ИИ, цифровая коммуникация, языковая политика в эпоху генерации.