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

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Наименование дисциплины (модуля):

Автоматизированная генерация текстов в профессиональной сфере

код и 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 и литература

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.05.

2. Цель и задачи освоения дисциплины (модуля)

Цель дисциплины: Формирование у обучающихся профессиональной компетенции в области применения технологий искусственного интеллекта для автоматизации создания, обработки и анализа текстовых данных в прикладных бизнес-процессах и корпоративной среде.

Задачи дисциплины:

1. Изучить фундаментальные технологии обработки естественного языка (NLP) и генеративного ИИ, определяющие возможности и ограничения современных текстовых моделей в корпоративном применении.
2. Сформировать навык проектирования и оптимизации промптов (промпт-инжиниринга) для решения прикладных задач: подготовки деловой переписки, создания маркетинговых материалов, составления технических описаний и аналитических отчетов.
3. Освоить методы интеграции генеративных моделей в существующие программные экосистемы (через API, плагины для офисных пакетов и корпоративные мессенджеры) для автоматизации рутинных текстовых операций.
4. Развить навыки экспертной верификации, фактчекинга и постредактирования сгенерированного контента для обеспечения его фактической точности, юридической безопасности и соответствия корпоративному стилю.
5. Провести комплексный лингвистический анализ сгенерированных текстов (на материале официально-делового, медийного или научного стилей) и выполнить их экспертное постредактирование, выявляя и исправляя типичные ошибки языковых моделей: нарушения лексической и грамматической сочетаемости, стилистические несоответствия жанру, смысловые искажения, синтаксическую омонимию и нарушения дискурсивной когезии.

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 задачами, условиями реальной и виртуальной коммуникации	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). • **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 нормативов различные типы текстов и документов в соответствии с отраслевыми, жанровыми, стилевыми требованиями и	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 редактирования (от light editing до deep editing); критерии оценки качества автоматического реферирования (сохранение фактологии, отсутствие искажений). • **Уметь:** составлять промпты для автоматического

практическими задачами, условиями реальной и виртуальной коммуникации		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 раздела и темы	Всего (часы)	Контактная работа (часы)	Форма промежуточной аттестации
Раздел 1. Лингвистические основы генеративного ИИ				
1	Тема 1. Генеративный ИИ как новый языковой уровень (Текст как объект моделирования, семантика, прагматика, проблема «галлюцинаций» и референции).	4	4	Тест
2	Тема 2. Промпт-инжиниринг: проектирование речевого акта для машины (Синтаксис и семантика запроса, управление стилем, регистром и иллокутивной силой).	4	4	Конспект
Раздел 2. Экспертиза и редактирование машинного текста				
3	Тема 3. Лингвистический аудит машинного текста (Нарушение лексической и грамматической нормы, стилистические диссонансы, разрушение когезии и когерентности).	4	4	Контрольная работа
4	Тема 4. Профессиональное постредактирование (Стратегии deep/light editing, адаптация под терминосистемы, приведение к эталонному стилю).	4	4	Тест
Раздел 3. Прикладная автоматизация и проектная работа				
5	Тема 5. Автоматизация лингвистических задач (Суммаризация, извлечение сущностей, машинный перевод и постредактура MTPE).	4	4	Конспект
6	Тема 6. Работа с неструктурированными данными и метаданными (NER,	4	4	Контрольная работа

№	Наименование раздела и темы	Всего (часы)	Контактная работа (часы)	Форма промежуточной аттестации
	классификация, тематическое моделирование, генерация аннотаций и хэштегов).			
7	Тема 7. Проектная работа: разработка лингвистического ассистента (Проектирование цепочек промптов, сборка библиотеки промптов для кейса).	4	4	
8	Тема 8. Лингвистика генеративного текста, этика и академические стандарты (Авторство, интертекстуальность, прагматика, этика данных, защита проектов).	4	4	Презентация
	Промежуточная аттестация	**Зачет -2**		
	ИТОГО	**34**	**32**	

6. Фонд оценочных материалов по дисциплине (модулю)

Типовые задания и иные материалы, необходимые для оценки результатов обучения

Примерные вопросы тестов:

Инструкция: Выберите один наиболее точный вариант ответа или дайте краткое определение.

Вопрос 1 (Выбор определения) Как в лингвистике генеративного ИИ наиболее точно определяется феномен «галлюцинации»?

1. Техническая ошибка сервера при генерации ответа.
2. **Семантическая аномалия, при которой модель генерирует фактологически ложный или не соответствующий контексту текст, сохраняя при этом его грамматическую и стилистическую связность.**
3. Намеренное искажение информации моделью в целях безопасности.

Вопрос 2 (Выбор варианта) Какой параметр генерации в первую очередь отвечает за «креативность» и разнообразие текста, увеличивая вероятность выбора менее частотных слов?

1. Top-p (nucleus sampling).
2. **Температура (Temperature).**
3. Max tokens (максимальная длина).

Вопрос 3 (Выбор варианта) Какая стратегия постредактирования (light vs deep editing) применяется, если сгенерированный текст содержит грубые фактические ошибки, нарушает логику повествования и требует полной переработки структуры?

1. Light editing (легкая корректура опечаток и пунктуации).
2. **Deep editing (глубокая переработка с восстановлением логики и фактуры).**
3. Форматирование (изменение шрифтов и отступов).

Вопрос 4 (Дать определение) Дайте определение понятию «промпт-инжиниринг» в контексте лингвистики. *(Оценивается точность: проектирование и оптимизация текстовых запросов (промптов) для управления поведением языковой модели и получения предсказуемого лингвистического результата).*

Вопрос 5 (Выбор варианта) Нарушение лексической сочетаемости в машинном тексте (например, *«одержать неминуемое поражение»* вместо *«потерпеть»*) является ошибкой на уровне:

1. Прагматики.
2. **Лексической семантики.**
3. Синтаксиса.

Вопрос 6 (Выбор определения) Что такое «когерентность» текста, которая часто нарушается языковыми моделями на больших расстояниях?

1. Грамматическая связь между соседними предложениями.
2. **Смысловое единство, логическая связность и последовательность развития темы во всем тексте.**
3. Наличие вводных слов и союзов между абзацами.

Вопрос 7 (Выбор варианта) Какое лингвистическое явление наиболее ярко проявляется в «статистическом эхе» обучающей выборки, когда модель выдает шаблонные, стереотипные или устаревшие формулировки?

1. Нарушение валентностей глаголов.
2. **Лексико-стилистическая интертекстуальность (скрытое цитирование и воспроизведение частотных паттернов из данных).**
3. Ошибки в морфологическом разборе.

Вопрос 8 (Дать определение) Что такое «цепочка промптов» (Prompt Chaining)? *(Оценивается точность: метод обработки текста, при котором сложная задача разбивается на последовательные этапы, где результат выполнения одного промпта становится входными данными для следующего).*

Вопрос 9 (Выбор варианта) В чем заключается главная лингвистическая проблема использования ИИ для машинного перевода с точки зрения последующей редакуры (MTPE)?

1. Модель всегда переводит слишком дословно.
2. **Сохранение синтаксической структуры исходного языка (калькирование) при разрушении норм и узуса языка перевода.**
3. Невозможность перевода терминов.

Вопрос 10 (Выбор варианта) С точки зрения академической этики и ГОСТ, как следует квалифицировать использование ИИ для генерации черновика научной статьи, который затем полностью переписывается автором?

1. Как плагиат, требующий отказа от публикации.
2. **Как использование ИИ в качестве инструмента (подготовительный этап), требующее прозрачности, но не нарушающее авторство при условии полной переработки и верификации.**
3. Как техническое нарушение, не требующее упоминания в работе.

Ключи ответов:

1. – 2
2. – 2
3. – 2
4. – Проектирование и оптимизация текстовых запросов (промптов) для управления поведением языковой модели и получения предсказуемого лингвистического результата.
5. – 2
6. – 2
7. – 2
8. – Метод обработки текста, при котором сложная задача разбивается на последовательные этапы, где результат выполнения одного промпта становится входными данными для следующего.
9. – 2
10. – 2

Балл	**Критерии оценивания**
5 (Отлично)	Обучающийся демонстрирует системное владение метаязыком лексикографии. Все ответы верифицированы. В

Балл	Критерии оценивания
	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:

№	Библиографическое описание	Где найти (открытый доступ)	Страницы для конспекта
1	**Большакова Е.И., Воронцов К.В. и др. Автоматическая обработка текстов на естественном языке и анализ данных. — М.: Изд. дом ВШЭ, 2017.**	Официально выложена в открытом доступе на сайте Высшей школы экономики (hse.ru) в разделе учебных пособий.	**Гл. 1 (стр. 10–35):** Лингвистические уровни и задачи автоматической обработки.**Гл. 4 (стр. 90–110):** Тематическое моделирование (как машина «понимает» смысл).
2	**Клышинский Э.С., Карпова И.С. Основы компьютерной лингвистики: учебное пособие. — М.: МИЭМ НИУ ВШЭ, 2015.**	Электронная версия доступна в репозитории НИУ ВШЭ и на портале «КиберЛенинка».	**Гл. 2 (стр. 25–45):** Формальные языки и грамматики (база для понимания архитектуры моделей).**Гл. 5 (стр. 80–105):** Машинный перевод и генерация текста.
3	**Сбоев А.В., Шагалова К.В. Нейросетевые модели для обработки естественного языка. — М.: ДМК Пресс, 2021.** *(Перевод и адаптация Y. Goldberg)*	В свободном доступе на крупных трекерах и в библиотеках (например, Libgen/Bookfi), так как издание научно-техническое.	**Гл. 2 (стр. 20–40):** Векторные представления слов (семантические пространства).**Гл. 5 (стр. 75–95):** Sequence-to-Sequence модели и механизмы внимания (основа генеративного ИИ).
4	**Падучева Е.В. Высказывание и его соотнесенность с действительностью.**	Оцифрованная версия доступна в Национальной электронной	**Введение (стр. 3–15):** Проблема референции и денотации.**Гл. 1, § 3 (стр. 45–55):** Анафорические связи и

№	Библиографическое описание	Где найти (открытый доступ)	Страницы для конспекта
	(Референциальные аспекты семантики местоимений). — М.: Наука, 1985.	библиотеке (НЭБ.рф) и в Президентской библиотеке.	местоимения (почему ИИ «теряет» нить повествования на длинных текстах). **Гл. 3 (стр. 90–110):** Прагматические аспекты высказывания (иллокутивная сила, пресуппозиции). **Семантика модальности (стр. 120–140):** Как язык выражает отношение говорящего к реальности (возможность, необходимость). Это поможет лингвистам конструировать промпты для управления «тональностью» (Tone of Voice) и «уверенностью» модели, используя лингвистические маркеры модальности вместо прямых технических инструкций.

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. **Лингвистическая переработка.** Студент не копирует текст, а переводит его на метаязык дисциплины. Теоретические положения (например, «референция» у Падучевой) сопровождаются пометками о применении к ИИ (напр.: *«анафора → разрушение контекста в LLM на >8k токенов»*). **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.

Балл	Критерии оценивания
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.

Дополнительные параметры для снижения балла (минус-баллы):

- **Небрежность оформления:** отсутствие полей для собственных пометок, сплошной нечитаемый текст (-1 балл).
- **Отсутствие библиографической привязки:** не указаны страницы или выходные данные источника (-1 балл).
- **Плагиат:** полное копирование текста из чужих рефератов или кратких содержаний (-автоматический «0»).

Контрольная работа:

Задание 1. Терминологический аппарат и метаязык (Теоретический блок) Дайте развернутые определения следующим понятиям в контексте взаимодействия лингвистики и генеративного ИИ. Объем ответа — 2–3 предложения на пункт.

1. **Галлюцинация языковой модели** (с позиции семантики и теории референции).
2. **Промпт-инжиниринг** (как вид искусственно сконструированного дискурса).
3. **Когезия и когерентность** (в аспекте устойчивости анафорических связей при генерации длинных документов).

(Оценивается точность использования метаязыка и разграничение смежных понятий).

Задание 2. Лингвистический аудит машинного текста (Аналитический блок) Проанализируйте представленный фрагмент, сгенерированный нейросетью для раздела «Цели и задачи» в рабочей программе дисциплины. *Фрагмент:* «Целью освоения данной дисциплины является овладение студентами компетенциями в сфере

цифровизации текстовых данных, а также для того, чтобы специалист умел работать с алгоритмами и повышать эффективность производства».

Вопросы:

1. Выявите и назовите синтаксическую и стилистическую аномалии в конструкции «...компетенциями..., а также для того, чтобы...». Какое правило синтаксического управления нарушено?
2. Определите тип лексической ошибки в словосочетании «овладение компетенциями в сфере цифровизации». Какое семантическое ограничение (селекционное ограничение) существительного «овладение» здесь нарушено?
3. Предложите лингвистически выверенную (отредактированную) версию данного фрагмента, устраняющую нарушения норм официально-делового стиля.

Задание 3. Синтаксическое проектирование и управление пресуппозициями (Практический блок)

Вам необходимо сгенерировать текст, описывающий **разработку двуязычного тезауруса для предметной области**, с использованием одной и той же фактологической базы (Компания/Лаборатория «Лингво-Проект», дата запуска 15.10.2026, руководитель группы — к.ф.н. И.И. Сидоров). Составьте два промпта, которые с точки зрения лингвистики управляют глубинной структурой (синтаксисом) и пресуппозициями генерируемого текста.

1. **Промпт А (Активация пассивной диатезы и номинализации):** Сформулируйте промпт, который заставит языковую модель сгенерировать текст в официально-деловом научном стиле, характеризующийся максимальной номинализацией (превращением действий в отглагольные существительные) и использованием пассивного залога. В результате субъект действия (И.И. Иванов) должен быть либо полностью элиминирован, либо переведен в позицию агентивного дополнения. *Пример ожидаемого результата:* «Разработка двуязычного тезауруса в Лаборатории "Лингво-Проект" запланирована на 15.10.2026. Лексикографическая разметка корпуса будет осуществлена ответственным коллективом...»
2. **Промпт Б (Активация синтаксической проекции и фокусировки):** Сформулируйте промпт, который заставит модель использовать синтаксические конструкции с препозицией (вынесением в начало) обстоятельства времени и цели, а также активирует фокусировку на личности руководителя через использование выделительных

конструкций (именительного темы или псевдосложного предложения). *Пример ожидаемого результата:* «15 октября 2026 года стартует ключевой этап грантового исследования. Именно И.И. Иванов, как ведущий лексикограф проекта, возглавит работу по установлению семантических отношений между вокабулами...»

Вопросы к заданию:

1. Укажите, какие конкретные лингвистические операторы (синтаксические маркеры) вы включили в промпт А для принудительной активации номинализации (например, использование терминов «процесс», «осуществление», «фаза реализации», «проведение анализа»).
2. Объясните, как использование конструкций с именительным темы в промпт Б влияет на информационную структуру (актуальное членение) предложения, изменяя соотношение «данного» (темы) и «нового» (ремы) в генерируемом тексте.
3. Сравните две полученные генерации. Какое из двух предложений (из примеров ожидаемого результата) обладает большей информационной плотностью (*information density*) и почему? Обоснуйте ответ с позиции теории функциональной стилистики и синтаксической сложности.

Задание 4. Семантическая деривация и нарушение валентностей (Аналитический блок) В сгенерированном тексте нейросеть допустила ошибку: *«...департамент осуществил высокую текучесть кадров...»*.

1. Объясните природу данной ошибки с точки зрения теории валентности и семантических ролей. Почему глагол «осуществить» (агентивное действие) семантически несовместим с подлежащим «текучесть» (абстрактный процесс)?
2. Приведите два примера корректного управления для глагола «осуществить» и два примера корректного подлежащего для существительного «текучесть», доказывающие сохранение их исходных валентных свойств.

Задание 5. Теория референции и анафора (Теоретико-прикладной блок) Опираясь на концепцию Е.В. Падучевой о соотнесенности высказывания с действительностью, проанализируйте сгенерированный диалог:

- **Запрос:** «Опиши архитектуру собора Василия Блаженного. Он находится в Москве».

- **Ответ ИИ:** «Собор Василия Блаженного — это православный храм на Красной площади. Она находится в Москве».

1. Какое именно референциальное правило (правило анафорического связывания) нарушила модель при выборе местоимения «Она»?
2. Объясните, почему статистическая вероятность выбора местоимения женского рода (опираясь на слово «храм») оказалась для модели выше, чем сохранение референциальной связи с именем собственным «Собор Василия Блаженного» (мужской род).
3. Сформулируйте правило для промпта, которое заставит модель строго соблюдать грамматическое и референциальное согласование местоимений в подобных задачах.

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 (Задания 2, 4, 5) 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 (Задания 2, 4, 5) факты идентифицированы верно, но объяснение лингвистического механизма представлено без достаточной глубины или с незначительными неточностями в терминологии. В задании на промпты (Задание 3) 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.

Балл	Критерии оценивания
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 заданиях на анализ (Задания 2, 4, 5) 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

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)

Автоматизированная генерация текстов в профессиональной сфере

(название дисциплины (модуля))

Код и наименование направления подготовки: 45.03.01 Филология

Уровень высшего образования бакалавриат

Направленность (профиль) образования: Русский язык и литература

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 «**Автоматизированная генерация текстов в профессиональной сфере**».

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. Теоретические основы генеративной лингвистики		
Тема 1. Генеративный ИИ как новый языковой уровень		
1.	**Вопросы для устного опроса:** – Разграничьте предметные области традиционной лингвистики и компьютерной лингвистики в контексте генеративных моделей. – Дайте определение понятию «контекстное окно» с точки зрения обработки дискурса. – Какова роль пресуппозиции в формировании запроса (промпта)?	ПК-2.1
2.	**Тест:** 1. Что является объектом моделирования для LLM? а) Звуковая система языка. б) Статистические закономерности распределения лексем в корпусе. в) Правила построения предложения. 2. Феномен «галлюцинации» с точки зрения семантики – это: а) Техническая ошибка сервера. б) Нарушение референции и максимы качества. в) Намеренное искажение информации.	ПК-2.1
Тема 2. Промпт-инжиниринг: проектирование речевого акта		
3.	**Вопросы для устного опроса:** а) Перечислите обязательные компоненты лингвистической структуры промпта (Роль-Контекст-Задача-Формат). б) Какова функция лингвистических маркеров (модальных слов, вводных конструкций) в управлении регистром генерируемого текста?	ПК-2.1
4.	**Тест:** 1. Параметр «Температура» (Temperature) в генерации текста отвечает за: а) Скорость ответа. б) Креативность и разнообразие (выбор менее частотных слов). в) Длину текста. 2. Какую функцию выполняет техника Chain of Thought (CoT) в промпте? а) Увеличивает объем текста. б) Принуждает модель к пошаговому логическому выводу. в) Изменяет язык ответа.	ПК-2.1
Тема 3. Лингвистический аудит и типология ошибок		
5.	**Вопросы для устного опроса:** – В чем заключается принципиальное отличие лексической несочетаемости от нарушения валентности глагола? – Охарактеризуйте разницу между разрушением когезии (формальной связи) и когерентности (смыслового единства).	ПК-2.3
6.	**Тест (на основе анализа фрагментов):** 1. Фрагмент: *«Департамент осуществил высокую текучесть кадров»*. Ошибка допущена на уровне: а) Синтаксиса. б) Лексической семантики (нарушение валентности/селективных ограничений).	ПК-2.3

№ п/п	Содержание задания	Шифр индикатора достижения компетенции (ИДК)
	в) Орфографии. 2. Фрагмент: *«Собор Василия Блаженного — это храм. Она находится в Москве»*. Нарушено: а) Правило грамматического согласования. б) Референциальная связь (анафорический провал). в) Стилистическая норма.	
Тема 4. Профессиональное постредактирование		
7.	**Вопросы для устного опроса:** – Какие требования предъявляются к этапу фактчекинга при стратегии Deep Editing? – В чем разница между адаптацией под терминосистему и общей стилистической правкой?	ПК-2.3
8.	**Тест (на основе анализа предложений):** 1. Стратегия редактирования, при которой текст полностью переписывается с восстановлением логики, называется: а) Light editing. б) Deep editing. в) Форматирование. 2. В предложении *«Целью является овладение компетенциями, а также для того, чтобы...»* нарушена: а) Синтаксическая норма (нанизывание падежей и конструкций). б) Лексическая норма. в) Пунктуационная норма.	ПК-2.3
Тема 5. Автоматизация прикладных лингвистических задач		
9.	**Вопросы для устного опроса:** – Какова роль лингвистических аннотаций (парсинга, NER) в обучении моделей для извлечения информации? – В чем заключается главная лингвистическая проблема машинного перевода (MTPE) при работе с прецизионной терминологией?	ПК-2.1
10.	**Тест:** 1. Извлечение именованных сущностей (NER) — это задача уровня: а) Морфологии. б) Семантики и информационного поиска. в) Фонетики. 2. Проблема «статистического эха» обучающей выборки в генерируемом тексте – это: а) Быстрое устаревание фактов. б) Воспроизведение моделью скрытых стереотипов и частотных паттернов из данных. в) Ошибки в написании слов.	

Ключи к оценочным материалам

Раздел 1. Теоретические основы генеративной лингвистики

Тема 1. Генеративный ИИ как новый языковой уровень

1. **Устный опрос:**

- Лексикология изучает словарный состав как систему, лексикография — принципы и методы его практического описания и словарного представления. В контексте ИИ: лингвистика изучает природу языка, а генеративные модели — его статистическую репрезентацию в данных.
- Контекстное окно — максимальный объем текста (в токенах), который модель способна удерживать в оперативной памяти при генерации следующего слова, определяющий границы анафорических связей.

- Пресуппозиция — исходный фоновый компонент смысла, который должен быть истинным для того, чтобы вопрос или утверждение в промпте были уместными; используется для управления «знанием» модели.

2. **Тест:**

- 1 — б; 2 — б.

Тема 2. Промпт-инжиниринг: проектирование речевого акта 3. **Устный опрос:** * Роль (Role/Persona), Контекст (Context), Задача (Task), Формат (Format), Ограничения (Constraints). * Функция маркеров — управление иллокутивной силой и регистром (Tone of Voice). Они задают прагматические параметры (убедительность, нейтральность, эмпатия), не меняя техническую суть запроса. 4. **Тест:** * 1 — б; 2 — б.

Тема 3. Лингвистический аудит и типология ошибок 5. **Устный опрос:** * Лексическая несочетаемость — нарушение семантической совместимости слов (красный синий). Нарушение валентности — неспособность глагола сочетаться с данным типом актантов (субъекта/объекта) по смыслу (*течь осуществила департамент*). * Когезия — формальные средства связи (местоимения, союзы). Когерентность — логическая и смысловая связность текста. Модель может сохранить когезию (повторить слово), но разрушить когерентность (потерять логику). 6. **Тест:** * 1 — б (глагол «осуществить» требует одушевленного агенса и целенаправленного действия, «текучесть» — абстрактный процесс). * 2 — б (модель выбрала местоимение «она» по грамматическому роду ближайшего существительного «храм», нарушив референциальную связь с исходным объектом «Собор»).

Тема 4. Профессиональное постредактирование 7. **Устный опрос:** * Сверка всех фактологических утверждений (дат, имен, терминов) с независимыми верифицированными источниками. Модель не знает правды, она знает вероятности. * Адаптация под терминосистему — замена лексики на строго зафиксированные в отрасли эквиваленты. Стилистическая правка — приведение текста к нормам жанра (официально-деловой, научный) без изменения терминологии. 8. **Тест:** * 1 — б; 2 — а.

Тема 5. Автоматизация прикладных лингвистических задач 9. **Устный опрос:** * Парсинг и NER (извлечение сущностей) создают структурированные данные (деревья зависимостей, метки), которые служат «золотым стандартом» (ground truth) для обучения моделей синтаксису и семантике. * Проблема «ложных друзей переводчика» и отсутствия эквивалентов: модель может перевести термин по частям или выбрать неверный аналог из многозначности, что требует жесткой постредактуры специалистом-лингвистом. 10. **Тест:** * 1 — б; 2 — б.

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

№ п/п	Вопросы к зачету	Шифр ИДК
2	Архитектура генеративных моделей (Трансформеры, механизм внимания). Лингвистическая интерпретация «понимания» контекста.	ПК-2.1, ПК-2.3
3	Феномен «галлюцинации» языковой модели: лингвистическая природа (нарушение референции, семантические аномалии).	ПК-2.1, ПК-2.3
4	Понятие промпта как искусственно сконструированного дискурса. Синтаксис и прагматика запроса.	ПК-2.1, ПК-2.3
5	Управление стилем и регистром через лингвистические маркеры в промпте (модальность, вводные конструкции).	ПК-2.1, ПК-2.3
6	Параметры генерации (температура, Top-p) и их влияние на лексико-семантическое разнообразие текста.	ПК-2.1, ПК-2.3
7	Контекстное окно и проблема разрушения когезии и когерентности на больших расстояниях.	ПК-2.1, ПК-2.3
8	Типология ошибок генеративного ИИ: лексические (валентность), грамматические и стилистические.	ПК-2.1, ПК-2.3
9	Разрушение анафорических связей (местоимений) как референциальный сбой модели.	ПК-2.1, ПК-2.3
10	Нарушение лексической сочетаемости и семантических ограничений (селективных ограничений) слов.	ПК-2.1, ПК-2.3
11	Стилистические диссонансы: смешение регистров и разрушение тональности (Tone of Voice).	ПК-2.1, ПК-2.3
12	Стратегии постредактирования: Light editing vs Deep editing. Лингвистические критерии выбора.	ПК-2.1, ПК-2.3
13	Адаптация машинного текста под узкопрофессиональные терминосистемы (юридическая, медицинская).	ПК-2.1, ПК-2.3
14	Восстановление прагматики и авторского замысла в процессе редакторской правки.	ПК-2.1, ПК-2.3
15	Академическая этика использования ИИ. Правила цитирования сгенерированного текста (ГОСТ, ВАК).	ПК-2.1, ПК-2.3
Раздел 2. Прикладная генеративная лингвистика		
16	Автоматическое реферирование (суммаризация). Лингвистические критерии сохранения фактологии.	ПК-2.1,

№ п/п	Вопросы к зачету	Шифр ИДК
		ПК-2.3
17	Извлечение именованных сущностей (NER) и атрибутов. Лингвистическая классификация объектов.	ПК-2.1, ПК-2.3
18	Машинный перевод и постредактирование (MTPE). Проблема «ложных друзей переводчика» в ИИ.	ПК-2.1, ПК-2.3
19	Генерация параллельных текстов для преподавания иностранных языков. Методический потенциал.	ПК-2.1, ПК-2.3
20	Классификация текстов через естественный язык (Zero-shot classification). Лингвистические маркеры категорий.	ПК-2.1, ПК-2.3
21	Тематическое моделирование через генеративные модели (вместо LDA). Семантические поля.	ПК-2.1, ПК-2.3
22	Цепочки промптов (Prompt Chaining) для многоэтапной обработки текста.	ПК-2.1, ПК-2.3
23	Работа с API языковых моделей. Формат запроса и парсинг лингвистически значимых данных.	ПК-2.1, ПК-2.3
24	Интеграция ИИ в офисные пакеты (Word, Excel) для автоматизации рутинных текстовых операций.	ПК-2.1, ПК-2.3
25	Проблема авторства в цифровую эпоху: промптер, модель, редактор.	ПК-2.1, ПК-2.3
26	Интертекстуальность машинного творчества. Понятие «статистического эха» обучающей выборки.	ПК-2.1, ПК-2.3
27	Прагматика и иллокутивные акты. Может ли машина совершить истинный речевой акт?	ПК-2.1, ПК-2.3
28	Дискриминация и этика данных (Data Ethics). Лингвистические маркеры скрытой предвзятости (bias).	ПК-2.1, ПК-2.3
29	Трансформация профессиональных компетенций лингвиста: от автора к AI-тренеру и лингво-аудитору.	ПК-2.1, ПК-2.3
30	Промпт как гибридный речевой жанр. Структура и конвенции.	ПК-2.1, ПК-2.3
31	Диалоговый след (chat log) как объект дискурс-анализа.	ПК-2.1,

№ п/п	Вопросы к зачету	Шифр ИДК
		ПК-2.3
32	Сравнительный анализ эффективности ручного и автоматизированного создания текстов (метрики качества).	ПК-2.1, ПК-2.3
33	Верификация фактологической точности сгенерированного текста. Методология фактчекинга.	ПК-2.1, ПК-2.3
34	Проектирование библиотеки промптов для решения конкретной профессиональной задачи.	ПК-2.1, ПК-2.3
35	Футурология лингвистики: новые объекты исследования в эпоху генеративного ИИ.	ПК-2.1, ПК-2.3

Ключи и опорные тезисы для оценки ответов

№	
1	Различие между правилами (лингвистика) и вероятностями (ИИ). Автоматизация — применение правил, генерация — предсказание токена.
2	Трансформер: токенизация, эмбеддинги, self-attention. Понимание как статистическая вероятность следования слова в данном контексте.
3	Галлюцинация = нарушение максимы качества Грайса. Семантическая аномалия при сохранении грамматической связности.
4	Структура RPCF (Role-Persona-Context-Task-Format). Промпт как речевой акт с заданной иллокуцией.
5	Использование модальных слов (должен, может), перформативов (напиши, составь), маркеров жанра (в стиле...).
6	Температура = креативность (энтропия). Top-p = ограничение словаря наиболее вероятными токенами (nucleus sampling).
7	Когезия (формальные связи: союзы, местоимения) vs Когерентность (смысл). Потеря контекста за пределами окна.
8	Нарушение валентности (ожидание одушевленного субъекта), лексическая несочетаемость (оксюморон), разрушение стилистического регистра.
9	Анафорический провал: модель выбирает ближайшее существительное по роду, а не референт по смыслу (Собор -> храм -> она).
10	Нарушение селективных ограничений (департамент осуществил текучесть). Семантические роли (агенс vs патиенс).
11	Смешение официально-делового и разговорного стилей. Несоответствие Tone of Voice бренду или задаче.
12	Light — исправление опечаток. Deep — фактчекинг, восстановление логики, смена структуры. Критерий: сохранение фактуры.
13	Замена общеупотребительной лексики на терминосистемы (например, юриспруденция:

№	
	«сделал» -> «совершил», «документ» -> «акт»).
14	Восстановление иллокуции (угроза, обещание), которую модель лишь симулирует, но не совершает.
15	ИИ как инструмент (черновик) vs ИИ как автор. Обязательная прозрачность использования. Ссылка на ГОСТ Р 7.0.5-2008.
16	Экстрактивная vs Абстрактивная суммаризация. Сохранение ключевых сем (смысловых компонентов).
17	Именованные сущности (имена, даты, локации). Атрибуты (должность, цена, возраст).
18	Калькирование синтаксиса. Проблема многозначности: выбор неверного эквивалента без учета контекста.
19	Генерация упражнений, диалогов, параллельных корпусов для контрастивного анализа.
20	Использование лингвистических дескрипторов в промпте для классификации без дообучения.
21	Выделение тем через семантическую близость слов в векторном пространстве.
22	Разбиение сложной задачи на этапы (суммаризация -> извлечение -> рерайт). Выход одного промпта — вход другого.
23	JSON-структура запроса. API-ответ как объект для парсинга.
24	Макросы в Word, плагины, скрипты для Google Docs. Автоматизация форматирования и вставки шаблонов.
25	Разделение ответственности. Промптер — автор идеи, модель — инструмент, редактор — финальный гарант качества.
26	«Статистическое эхо» — воспроизведение моделью частотных паттернов и стереотипов из обучающих данных.
27	Иллокутивный акт требует интенции. Модель симулирует интенцию через статистические паттерны, но не обладает ею.
28	Bias в данных: гендерные роли (врач - он, няня - она), социальные стереотипы в генерации.
29	Смена ролей: создатель текста -> разметчик данных (датасетов) -> аудитор ИИ -> промпт-инженер.
30	Жанр: наличие цели, образа автора (модели), образа адресата (пользователя), формата.
31	Анализ логов как корпуса диалогов: стратегии уточнения запроса, исправления ошибок модели пользователем.
32	Сравнение трудозатрат (API vs человек) и качества (оценка по критериям ЗУВ, АП5).
33	Перекрестная проверка фактов в независимых источниках. Верификация дат, имен, цитат.
34	Создание набора промптов под одну задачу (библиотека) для стандартизации качества.
35	Промпт как объект лингвистики. Chat log как материал для дискурс-анализа. Смерть автора (Барт) vs новая риторика.