

На правах рукописи



Дагаев Александр Евгеньевич

**РАЗРАБОТКА АЛГОРИТМОВ ИЗВЛЕЧЕНИЯ И ОБРАБОТКИ
ИНФОРМАЦИИ ДЛЯ ГЕНЕРАЦИИ МНОГОФОРМАТНЫХ ТЕСТОВЫХ
ЗАДАНИЙ**

Специальность 2.3.8. Информатика и информационные процессы

Автореферат
диссертации на соискание ученой степени
кандидата технических наук

Москва – 2026

Работа выполнена на кафедре «Информатика и информационные технологии» федерального государственного бюджетного образовательного учреждения высшего образования «Московский политехнический университет»

Научный руководитель: **Попов Дмитрий Иванович**
доктор технических наук, профессор, профессор кафедры «Информационные технологии и математика», ФГБОУ ВО «Сочинский государственный университет», г. Сочи

Официальные оппоненты: **Пылькин Александр Николаевич**
доктор технических наук, профессор, заслуженный работник высшей школы РФ, Почетный работник науки и техники РФ, профессор кафедры вычислительной и прикладной математики, ФГБОУ ВО «Рязанский государственный радиотехнический университет им. В.Ф. Уткина», г. Рязань

Остроух Андрей Владимирович
доктор технических наук, профессор, профессор кафедры «Автоматизированные системы управления», ФГБОУ ВО «Московский автомобильно-дорожный государственный технический университет (МАДИ)», г. Москва

Ведущая организация: Акционерное общество «Научно-исследовательский и конструкторский институт монтажной технологии – Атомстрой», г. Москва

Защита диссертации состоится 29.06.2026 года в 11 ч. 00 мин. на заседании объединенного диссертационного совета 99.2.113.02 на базе ФГБОУ ВО «Рязанский государственный радиотехнический университет имени В.Ф. Уткина», ФГБОУ ВО «Пензенский государственный технологический университет» по адресу: 390005, г. Рязань, ул. Гагарина, д. 59/1, РГРТУ, ауд. 235.

С диссертацией можно ознакомиться в библиотеке ФГБОУ ВО «Рязанский государственный радиотехнический университет имени В.Ф. Уткина», на сайте <http://rsreu.ru/>.

Автореферат разослан «___» _____ 2026 г.

Ученый секретарь
диссертационного совета
доктор технических наук, доцент



А.Н. Колесенков

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования

Актуальность исследования обусловлена активным развитием интеллектуальных образовательных систем и цифровых технологий обучения, ориентированных на автоматизацию процессов формирования и диагностики компетенций обучающихся. В условиях реализации компетентностной модели образования возрастает потребность в инструментах, обеспечивающих объективную, масштабируемую и воспроизводимую оценку степени сформированности отдельных компонентов компетенций, включая когнитивные, аналитические и прикладные аспекты.

Одним из ключевых средств такой оценки являются тестовые задания различных форматов, используемые для диагностики усвоения учебного материала и компонентов компетенций. В то же время традиционные подходы к их разработке, основанные преимущественно на ручной подготовке, характеризуются высокой трудоёмкостью, субъективностью и ограниченной масштабируемостью. Это существенно затрудняет их применение в современных образовательных системах, ориентированных на массовое и персонализированное обучение.

Следует подчеркнуть, что разрабатываемые алгоритмы рассматриваются не как замена педагогической деятельности, а как вспомогательный инструментарий цифровой образовательной среды, предназначенный для снижения трудоёмкости рутинных операций преподавателя.

В последние годы всё большее внимание уделяется применению методов искусственного интеллекта и обработки естественного языка для автоматизированной генерации тестовых заданий на основе учебных текстов. Несмотря на достигнутые результаты, существующие решения обычно ориентированы на генерацию отдельных типов заданий или на формальное соответствие исходному тексту без учёта дидактических требований, качества дистракторов и необходимости верификации корректности сформированных заданий. Также во многих работах отсутствует интеграция этапов извлечения информации, генерации заданий и последующей оценки их качества.

Наиболее сложной является задача автоматизированной генерации дистракторов, то есть неверных вариантов ответа, правдоподобных с точки зрения обучающегося, но логически или содержательно некорректных. Качество дистракторов во многом определяет диагностическую ценность тестового задания. Распространённые методы их формирования основываются на поверхностной семантической близости и не учитывают когнитивные механизмы типичных ошибок обучающихся.

Не менее значимой является проблема верификации сгенерированных тестовых заданий. Отсутствие автоматизированных процедур контроля приводит к появлению заданий с фактическими ошибками, неоднозначными формулировками, отсутствием корректного ответа либо несоответствием заявленному уровню сложности. Это снижает достоверность результатов диагностики и ограничивает практическое применение систем автоматизированной генерации заданий.

В связи с этим актуальной научно-технической задачей является разработка алгоритмов извлечения и обработки информации, обеспечивающих автоматизированную генерацию многоформатных тестовых заданий и дистракторов с последующей интеллектуальной верификацией их качества. Решение данной задачи связано с интеграцией методов обработки естественного языка, алгоритмов генерации контента и формализованных процедур оценки характеристик формируемых заданий, ориентированных на требования компетентностного подхода и практику применения в современных образовательных системах.

Степень разработанности темы.

Проблематика разработки алгоритмов извлечения и обработки информации для автоматизированной генерации тестовых заданий получила значительное развитие в современной научной литературе в области компьютерной лингвистики, искусственного интеллекта и образовательных технологий. Сформирована обширная исследовательская база, охватывающая задачи семантического анализа учебных текстов, автоматической генерации вопросов различных форматов, формирования дистракторов и последующей верификации результатов генерации.

Базовые подходы к автоматической генерации вопросов (Automatic Question Generation, AQG) были заложены в работах, посвящённых извлечению ключевой информации из текста, синтаксическому и семантическому анализу, а также преобразованию текстового контента в формализованные задания. К данному направлению относятся исследования М. Heilman, N. Smith, R. Mitkov, J. Mostow, S. Kurdi, Y. Zhang, L. Dong, M. Lapata и др., в которых сформированы теоретические основы автоматической генерации вопросов и определены основные классы алгоритмических решений.

Существенный прогресс в развитии AQG связан с внедрением контекстных языковых моделей и трансформерных архитектур. В работах J. Devlin, M. Liu, C. Raffel, A. Vaswani, Y. Tang, Z. Cao показано, что использование контекстно-зависимых представлений текста позволяет повысить семантическую согласованность автоматически формируемых заданий и расширить спектр поддерживаемых форматов.

Современный этап исследований характеризуется активным применением больших языковых моделей, способных выполнять генерацию тестовых заданий в рамках единого генеративного контура. Данные подходы представлены в работах T. Brown, J. Wei, T. Kojima, S. Yao, T. Schick, S. Min и др., где демонстрируется возможность совместной генерации текста вопроса, правильного ответа и набора дистракторов на основе исходного текстового контента. При этом подчёркивается необходимость разработки формализованных процедур контроля корректности и интерпретируемости результатов генерации.

Значительное число исследований посвящено автоматической генерации дистракторов (Automatic Distractor Generation, ADG) как ключевому компоненту заданий с выбором ответа. В работах S. Guo, Z. Jiang, C. Liang, J. Gao, K. Zhou, Y. Sakaguchi рассматриваются методы формирования правдоподобных неверных вариантов ответа с учётом контекста вопроса, семантической структуры текста и типичных ошибок обучающихся.

Отдельное направление исследований связано с автоматизированной верификацией качества сгенерированных тестовых заданий. В более ранних работах применялись преимущественно метрики текстового сходства, однако в современных исследованиях предлагаются многоаспектные подходы, включающие проверку фактической корректности, логической согласованности, релевантности содержания и соответствия уровня сложности. Данные подходы представлены в работах Y. Yuan, W. Wang, D. Kiela, E. Durmus, S. Min и др., включая многоагентные и LLM-ориентированные схемы верификации.

В отечественной научной школе значительный вклад в развитие методов обработки естественного языка и создание языковых моделей для русского языка внесли А. В. Куратов, М. С. Архипов, А. А. Сорокин, М. М. Бурцев, Т. И. Шаврина, А. А. Феногенова, А. А. Панченко, Е. А. Артёмова и др. Разработанные ими русскоязычные языковые модели и корпуса создают технологическую основу для реализации алгоритмов автоматизированной генерации тестовых заданий на русском языке.

Таким образом, при высокой степени разработанности отдельных компонентов автоматизированной генерации тестовых заданий в современной научной литературе отсутствуют комплексные исследования, предлагающие целостный алгоритмический подход, объединяющий указанные компоненты в рамках единого интегрированного контура, ориентированного на многоформатные задания, интеллектуальную верификацию и специфику русскоязычных образовательных текстов.

Цель исследования: разработка и исследование интеллектуальных алгоритмов для извлечения и обработки информации, генерации и верификации многоформатных тестовых материалов из цифрового контента.

Задачи исследования:

1. Провести анализ существующих алгоритмов автоматической генерации тестовых заданий и выявить их ограничения применительно к русскоязычному материалу.
2. Сформировать алгоритм семантического анализа образовательного контента и извлечения из него ключевой информации, релевантной для последующей генерации тестовых заданий.
3. Создать алгоритм генерации многоформатных тестовых заданий, включая задания с выбором одного или нескольких вариантов ответа, задания открытой формы, на установление соответствия и правильной последовательности.
4. Разработать алгоритм интеллектуального формирования дистракторов на основе когнитивного моделирования ошибок обучающихся.
5. Реализовать программный комплекс для генерации контрольно-измерительных материалов и провести его экспериментальную апробацию для подтверждения эффективности и оценки качества генерируемых контрольно-измерительных материалов.

Объект исследования: процессы автоматического извлечения и обработки информации из цифрового образовательного контента для создания тестовых материалов.

Предмет исследования: алгоритмы интеллектуального анализа текстовых данных и автоматической генерации многоформатных тестовых заданий.

Под интеллектуальными алгоритмами в работе понимается совокупность методов, основанных на технологиях искусственного интеллекта, таких как большие языковые модели, мультиагентные системы и когнитивное моделирование, в отличие от жестко детерминированных процедур.

Методологическая база исследования опирается на современные подходы к обработке естественного языка, анализу данных и проектированию алгоритмов. В работе используется сочетание методов компьютерной лингвистики, семантического извлечения информации и методов машинного обучения для разработки и валидации алгоритмов автоматической генерации тестовых заданий. Конкретные подходы включают глубокий семантический, синтаксический и морфологический анализ, а также применение больших языковых моделей для извлечения релевантных данных из образовательных текстов. Экспериментальная методология включает как качественные, так и количественные методы оценки, а также метрики качества сгенерированных тестовых заданий и сравнительный анализ с существующими алгоритмами.

Теоретическая база исследования опирается на фундаментальные труды в области обработки естественного языка, компьютерной лингвистики и педагогических измерений. Ключевыми теоретическими рамками являются семантическая разметка ролей, теория когнитивной нагрузки в образовательной оценке и принципы психометрического конструирования тестов. Разработанные алгоритмы генерации дистракторов опираются на принципы когнитивного моделирования ошибок, а система верификации – на теорию многоагентных систем. Исследование также учитывает современные достижения в области образовательных технологий, основанных на искусственном интеллекте, интегрируя результаты отечественной и зарубежной научной литературы. Важный теоретический фундамент для семантических и психометрических компонентов разработанных алгоритмов обеспечивают труды таких исследователей, как R. Mitkov, A. E. Goldberg, J. Sweller, F. M. Lord и L. M. Rudner.

Эмпирическая база исследования включает обширные наборы данных цифрового образовательного контента на русском языке: учебники, научные публикации и предметно-ориентированные материалы по различным дисциплинам. Среди эмпирических методов исследования – сравнительный анализ работы предложенных алгоритмов с их базовыми решениями, оценка качества результатов как с помощью автоматизированных метрик, так и на основе экспертной оценки, а также экспериментальная проверка адаптивности алгоритмов к различным образовательным контекстам. Интеграция указанных эмпирических данных с методологическими и теоретическими рамками обеспечивает надёжность и практическую применимость разработанных алгоритмов в условиях реальных образовательных практик.

Научная новизна исследования заключается в следующем:

1. Предложен алгоритм абстрактивного извлечения информации, отличающийся от существующих алгоритмов, основанных на статистических или общих семантических оценках качества тем, что вводит

параметрическую интегральную метрику с доменно-специфичными весовыми коэффициентами, что позволяет повысить релевантность извлекаемой информации в среднем на 13,2 % относительно лучшей из рассмотренных базовых моделей (O1).

2. Разработан алгоритм автоматической генерации многоформатных тестовых заданий, отличающийся от существующих подходов на основе больших языковых моделей тем, что интегрирует модуль мультиагентной верификации, позволяющий снизить долю дефектных заданий на 84,5 %.
3. Предложен алгоритм генерации дистракторов, отличающийся от алгоритмов, основанных на семантической близости к правильному ответу тем, что моделирует процесс построения корректной цепочки рассуждений с последующим внесением когнитивных ошибок для генерации неверных ответов, что позволяет повысить качество дистракторов на 22,7 % относительно базовых алгоритмов.

Теоретическая значимость исследования состоит в развитии теоретических основ компьютерной лингвистики применительно к образовательным технологиям, расширении научных представлений о методах автоматического извлечения и структурирования знаний, формулировании новых принципов создания интеллектуальных образовательных систем.

Результаты исследования вносят вклад в теорию педагогических измерений через разработку новых подходов к автоматизации процессов создания тестовых материалов, в теорию искусственного интеллекта путём создания специализированных алгоритмов обработки образовательного контента.

Практическая значимость работы определяется возможностью использования разработанных алгоритмов в системах электронного обучения, платформах массовых открытых онлайн-курсов и корпоративных системах обучения.

Практические результаты исследования могут быть применены для:

- автоматизации создания тестовых материалов в образовательных учреждениях,
- повышения эффективности процессов разработки электронных образовательных ресурсов,
- создания адаптивных систем контроля знаний,
- разработки инструментов поддержки преподавательской деятельности.

Положения, выносимые на защиту.

1. Алгоритм абстрактного извлечения информации из образовательного контента. Применение в алгоритме интеллектуального отбора аннотаций на основе откалиброванной доменно-настроенной интегральной метрики обеспечивает формирование релевантной базы для генерации тестов с количественным преимуществом в интегральной оценке качества извлекаемой информации на 13,2 % по сравнению с лучшей из рассмотренных современных больших языковых моделей (O1).
2. Алгоритм генерации многоформатных тестовых заданий на основе LLM, интегрированный с мультиагентной системой верификации. Использование ансамбля узкоспециализированных агентов позволяет проводить

фильтрацию логически и фактически некорректных формулировок, обеспечивая снижение количества дефектных контрольно-измерительных материалов на 84,5 %.

3. Алгоритм интеллектуального формирования дистракторов, основанный на когнитивном моделировании ошибок. Явное построение цепочки рассуждений с имитацией типичных заблуждений обучающихся обеспечивает повышение правдоподобности и семантического разнообразия неверных вариантов ответа, что дает интегральный прирост качества дистракторов на 22,7 % по сравнению с базовыми алгоритмическими подходами.

Соответствие научно-квалификационной работы Паспорту научной специальности.

Диссертационное исследование соответствует паспорту научной специальности 2.3.8 «Информатика и информационные процессы» по пп. 4 и 5.

Пункт 4. Разработка методов и технологий цифровой обработки аудиовизуальной информации с целью обнаружения закономерностей в данных, включая обработку текстовых и иных изображений, видео контента. Разработка методов и моделей распознавания, понимания и синтеза речи, принципов и методов извлечения требуемой информации из текстов.

Пункт 5. Лингвистическое обеспечение информационных систем и процессов. Методы и средства проектирования словарей данных, словарей индексирования и поиска информации, тезаурусов и иных лексических комплексов. Методы семантического, синтаксического и прагматического анализа текстовой информации для представления в базах данных и организации интерфейсов информационных систем с пользователями.

Разработанные в диссертации алгоритмы по своей функциональной роли являются прямой реализацией «методов и технологий» цифровой обработки текстовой информации и извлечения из нее требуемых сведений, что соответствует пункту 4 паспорта специальности.

В то же время, будучи реализованными в виде программного комплекса, данные алгоритмы выступают в качестве «средств» лингвистического обеспечения информационных систем, направленных на семантический и прагматический анализ текста. Это устанавливает соответствие работы пункту 5 паспорта специальности.

Апробация и реализация результатов исследования.

Результаты исследования апробированы на 4 международных научных конференциях:

- III Международная научно-практическая конференция «Компьютерные приложения для управления и устойчивого развития производства и промышленности»,
- XVI Международная научно-практическая конференция «Управление устойчивым развитием России в условиях цифровой трансформации»,
- XII Международная научная конференция «Математическое и компьютерное моделирование»,

- IX Международная научно-практическая конференция студентов, аспирантов и молодых учёных «Фундаментальные и прикладные исследования молодых учёных».

Основные результаты исследования представлены в 9 работах, включая 4 статьи в журналах, входящих в перечень ВАК (из них 3 (К2) по направлению 2.3.8), 5 публикаций в изданиях РИНЦ, а также 11 зарегистрированных в Роспатенте программ для ЭВМ.

Личный вклад автора. Все основные результаты, выносимые на защиту, получены автором лично. Автором выполнены постановка научной задачи, анализ состояния исследований по теме диссертации, разработка алгоритмов абстрактного извлечения информации, генерации многоформатных тестовых заданий, интеллектуального формирования дистракторов, программная реализация предложенных решений, организация и проведение экспериментальных исследований, обработка, анализ и интерпретация полученных результатов.

Объем и структура диссертации. Диссертационная работа состоит из введения, четырёх глав, заключения, списка литературы и приложений. Общий объём работы составляет 146 страниц машинописного текста, содержит 30 таблиц; список литературы включает 136 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность исследования, сформулированы его цель и задачи, определены объект и предмет, а также описаны методологическая и теоретическая базы. Показано, что автоматизированная генерация тестовых заданий представляет собой значимое направление развития интеллектуальных образовательных технологий, позволяющее сократить трудоёмкость формирования контрольно-измерительных материалов и повысить их адаптивность и обновляемость. Особое внимание уделено проблемам генерации многоформатных тестовых заданий и качественных дистракторов для русскоязычного образовательного контента, а также необходимости автоматизированной верификации корректности и дидактической состоятельности формируемых заданий. На основе анализа современных исследований выявлены нерешённые научно-технические задачи и сформулирована научная проблема, заключающаяся в отсутствии интегрированных алгоритмов и программных средств, обеспечивающих воспроизводимый полный цикл автоматизированной генерации тестовых заданий – от извлечения релевантной информации из текстовых источников до формирования заданий различных форматов и проверки их корректности.

В первой главе представлен развернутый обзор современных подходов к автоматизированной генерации тестовых заданий и дистракторов, отражающий эволюцию данной области в контексте развития методов обработки естественного языка и искусственного интеллекта. Показано, как от первоначальных решений, основанных на синтаксических шаблонах и грамматических правилах, исследователи перешли к статистическим моделям и технологиям машинного перевода, трактующим генерацию вопросов как задачу преобразования исходного текста в последовательность, соответствующую «языку вопросов». Такие методы

позволили учитывать вероятностные зависимости между словами и фразами, однако при ограниченности параллельных корпусов часто сводились к поверхностному перефразированию исходных утверждений.

Дальнейшее развитие направления связано с использованием нейронных архитектур типа *encoder–decoder*, первоначально реализованных на основе рекуррентных нейронных сетей, а затем трансформеров. Их применение продемонстрировало возможность обучения моделей на парах «контекст – вопрос» и формирования более вариативных и семантически релевантных формулировок. Последующие исследования дополнили данные архитектуры механизмами внимания и фокусировки на ключевых элементах текста, что позволило повысить точность и содержательную насыщенность генерируемых вопросов.

Особое внимание в обзоре уделено анализу предварительно обученных языковых моделей, включая BERT, GPT и T5. Использование априорных знаний о языке и фактах обеспечивает данным моделям высокую связность и гибкость генерации, однако одновременно выявляет проблемы интерпретируемости результатов и риска появления фактически некорректных утверждений. Переход к большим языковым моделям последнего поколения расширил возможности генерации заданий различных типов на основе исходного текста и инструкций на естественном языке, но потребовал разработки специализированных процедур инженерии подсказок и внедрения механизмов контроля достоверности получаемых результатов.

В главе также рассмотрены гибридные стратегии, объединяющие нейросетевые алгоритмы с внешними источниками знаний, включая онтологии, графы сущностей, структурированные и мультимодальные данные. Подобные подходы применяются в задачах, требующих интеграции информации из разнородных источников или построения заданий на основе нескольких типов представлений знаний.

Значительное внимание уделено проблеме формирования дистракторов для заданий с выбором ответа. Показано, что качество неверных, но педагогически значимых вариантов ответа во многом определяет диагностическую ценность тестового задания. Эволюция соответствующих методов прослеживается от лексических подстановок и эвристических правил к вероятностным моделям ошибок обучающихся и алгоритмам, основанным на больших языковых моделях. Несмотря на достигнутые результаты, сохраняются сложности, связанные с поверхностной модификацией правильного ответа либо генерацией семантически удалённых вариантов, не выполняющих функцию эффективного отвлечения. В литературе предлагаются различные подходы к их преодолению, включая многоэтапную фильтрацию кандидатов, ранжирование по семантическим и дидактическим критериям и моделирование типичных когнитивных заблуждений.

Проведённый анализ показывает, что при наличии значительного прогресса в отдельных компонентах – генерации вопросов, формировании дистракторов, использовании внешних знаний и мультимодальных данных – по-прежнему отсутствует воспроизводимая целостная концепция, объединяющая извлечение релевантной информации, формирование заданий различных форматов и

автоматизированную верификацию их корректности, особенно в контексте русскоязычных образовательных материалов.

Во второй главе излагаются результаты разработки и исследования алгоритма абстрактного извлечения информации, предназначенного для формирования аннотаций, пригодных для последующей генерации контрольно-измерительных материалов. Показано, что существующие алгоритмы суммаризации, включая абстрактные, ориентируются преимущественно на универсальные метрики качества и, как правило, не учитывают специфику конкретных предметных областей. Такой подход приводит к снижению как точности, так и дидактической ценности автоматически формируемых аннотаций, особенно при их использовании в качестве основы для построения тестовых заданий.

Для преодоления данного ограничения предложен алгоритм, сочетающий автоматическую классификацию области знаний, генерацию нескольких аннотаций и их отбор на основе интегральной метрики качества с доменно-настроенными весовыми коэффициентами.

Алгоритм включает следующие основные этапы:

- автоматическую классификацию исходного текста T по предметной области с использованием предварительно обученной языковой модели, дообученной на размеченных корпусах;

- генерацию набора абстрактных аннотаций

$$S = \{S_1, S_2, \dots, S_n\}$$

- вычисление набора метрик релевантности и полноты

$$M = \{m_1, m_2, \dots, m_k\}, m_j(S_i, T) \rightarrow [0,1]$$

- расчёт интегральной оценки с доменными весами и выбор наилучшего кандидата

Для каждой предметной области $d \in \mathcal{D}$ (набор доменов) задаётся вектор весов:

$$w^{(d)} = \{w_1^{(d)}, w_2^{(d)}, \dots, w_k^{(d)}\}, \sum_{i=1}^k w_i^{(d)} = 1$$

Интегральная оценка аннотации определяется выражением:

$$Q(S_i, T; d) = \sum_{j=1}^k w_j^{(d)} \cdot m_j(S_i, T)$$

Оптимальная аннотация выбирается по правилу:

$$S^* = \arg \max_{S_i \in S} Q(S_i, T; d)$$

Для различных предметных областей сформированы отдельные профили весов интегральной метрики, что позволяет учитывать особенности структуры знаний и характер изложения учебного материала. Весовые коэффициенты интегральной оценки определялись автоматически на основе процедуры эмпирической калибровки, направленной на максимизацию согласованности с эталонными аннотациями, и приведены в таблице 1.

Таблица 1 – Весовые коэффициенты интегральной метрики

Метрика	Формальные науки	Естественные науки	Гуманитарные науки	Социально-экономические науки	Юридические науки
ROUGE-1	0,08	0,07	0,14	0,15	0,15
ROUGE-2	0,38	0,32	0,21	0,09	0,22
ROUGE-L	0,11	0,14	0,23	0,36	0,26
BLEU	0,04	0,07	0,08	0,05	0,07
METEOR	0,12	0,13	0,20	0,11	0,11
BERTScore	0,14	0,17	0,09	0,13	0,12
BLEURT	0,13	0,10	0,05	0,11	0,07

Экспериментальная проверка алгоритма проведена на сбалансированном корпусе русскоязычных образовательных текстов. В качестве базовых решений использовались реализации современных алгоритмов абстрактивной суммаризации на основе моделей O1, O1-mini, GPT-4o и GPT-4o-mini. Качество итоговых аннотаций оценивалось по совокупности стандартных метрик и сопоставлялось с эталонными резюме, подготовленными экспертами. Результаты эксперимента представлены в таблице 2.

Таблица 2 – Сравнительная оценка качества аннотаций

Метрика Модель	ROUGE-1	ROUGE-2	ROUGE-L	BLEU	METEOR	BERTScore	BLEURT
O1	0,35	0,15	0,33	0,17	0,23	0,78	0,65
O1-mini	0,29	0,12	0,27	0,14	0,20	0,75	0,60
GPT-4o	0,33	0,13	0,31	0,16	0,22	0,78	0,63
GPT-4o-mini	0,27	0,11	0,25	0,13	0,19	0,74	0,59
Предложенный алгоритм	0,39	0,18	0,37	0,20	0,27	0,82	0,69
Прирост относительно лучшей SOTA-модели	11,4%	20%	12,1%	17,6%	17,4%	5,1%	6,2%

Относительный прирост по сравнению с лучшей базовой моделью составил от 11,4 % до 20 % по метрикам семейства ROUGE, 17,4–17,6 % по METEOR и BLEU, а также 5,1–6,2 % по семантическим метрикам BERTScore и BLEURT. Прирост рассчитывался как среднее процентное улучшение значений метрик по корпусу текстов.

Усреднённый по предметным областям относительный прирост интегральной оценки качества аннотаций, рассчитанный на основе доменно-настроенной метрики составляет 13,2 %.

Достигнутый эффект обусловлен не только использованием языковой модели для генерации аннотаций, но и введением механизма интеллектуального отбора, основанного на доменно-настроенной интегральной метрике.

В заключение главы отмечается, что использование доменно-настроенной интегральной оценки позволяет улучшить характеристики извлекаемой информации и обеспечивает возможность последующей автоматической генерации тестовых заданий.

В третьей главе изложены результаты разработки алгоритма, который автоматически формирует тестовые задания различных типов и обеспечивает их последующую проверку на корректность и педагогическую пригодность. Показано, что для повышения качества контрольно-измерительных материалов целесообразно интегрировать процедуру генерации заданий с модулем многоаспектной верификации, предназначенным для выявления и устранения широкого спектра потенциальных дефектов.

Алгоритм принимает на вход текстовый контент \mathcal{C} , параметр формата задания

$$\tau \in \mathcal{T} = \{\text{single, multiple, open, match, sequence}\},$$

а также уровень сложности

$$\sigma \in \{1, 2, 3\}$$

где 1 соответствует низкому, 2 – среднему, 3 – высокому уровню. На основе указанных параметров большая языковая модель формирует текст задания и структуру вариантов ответа:

$$(q, V(q)) = G(\mathcal{C}, \tau, \sigma),$$

где q – текст задания, $V(q)$ – набор ответов или элементов последовательности, G – генератор, реализованный на базе LLM.

Поскольку результаты генерации не всегда гарантируют фактическую корректность и соответствие педагогическим требованиям, каждое сформированное задание проходит процедуру мультиагентной бинарной верификации. Для анализа различных аспектов качества используется множество агентов

$$\mathcal{A} = \{A_1, A_2, \dots, A_7\},$$

Каждый агент A_j формирует бинарное решение $s_j \in \{0, 1\}$ на основе специализированного подмножества входных данных (формулировка задания, структура ответа, исходный контент, параметр сложности) и инструментальных модулей.

Выбор состава агентов обусловлен необходимостью охвата основных характеристик тестового задания. Назначение агентов представлено в таблице 3.

Таблица 3 – Назначение агентов системы верификации

№	Агент	Проверяемый аспект
1	FactChecker	Фактическая корректность
2	ContextGrounding	Соответствие теме исходного текста
3	AnswerValidity	Корректность структуры ответа
4	Clarity	Ясность и однозначность формулировки
5	DifficultyMatch	Соответствие уровня сложности σ
6	BiasDetection	Отсутствие предвзятости
7	Originality	Отсутствие прямого копирования исходного фрагмента

По результатам работы агентов формируется вектор решений:

$$\mathbf{s}(q) = (s_1, s_2, \dots, s_7) \in \{0, 1\}^7$$

Задание принимается системой только при условии выполнения всех критериев:

$$\text{Accept}(q) = \prod_{j=1}^7 s_j = 1$$

Использование данного правила обеспечивает исключение заданий, содержащих хотя бы один критический дефект, и ориентировано на формирование надёжных контрольно-измерительных материалов. Эффективность системы верификации оценивалась на корпусе из 1000 автоматически сгенерированных заданий, что обеспечивает статистически устойчивую оценку качества и сопоставимо с объёмами выборок, используемыми в аналогичных исследованиях. Результаты эксперимента приведены в таблице 4.

Таблица 4 – Сравнение качества заданий до и после верификации

Показатель	До верификации	После верификации	Снижение (%)
Всего заданий	1000	1000	—
Задания с любыми дефектами	419	65	84,5%
– фактические ошибки	106	7	93,4%
– тематическая нерелевантность	76	17	77,6%
– некорректная структура ответа	54	9	83,3%
– неясные формулировки	64	13	79,7%
– несоответствие уровня сложности	47	10	78,7%
– наличие предвзятости	41	5	87,8%
– копирование исходного текста	31	4	87,1%

Анализ экспериментальных данных показал, что внедрение многоагентной верификации позволяет существенно снизить долю заданий с дефектами: общее количество ошибок сокращается на 84,5%, при этом наибольший эффект достигается для фактических неточностей (93,4%) и проявлений предвзятости (87,8%). Существенно уменьшается также количество заданий с неясными формулировками, несоответствием заявленного уровня сложности и прямым копированием фрагментов исходного текста.

В завершение главы сформулированы основные выводы. Разработан комплексный алгоритм, разделяющий процессы генерации и контроля качества тестовых заданий. Первый этап обеспечивает гибкое формирование заданий различных типов и уровней сложности на основе больших языковых моделей. Второй этап реализует мультиагентную систему верификации, позволяющую автоматически выявлять и исключать задания с фактическими ошибками, логическими несоответствиями и иными дефектами. Проведённые эксперименты подтверждают эффективность предложенного подхода и его способность повысить надёжность и педагогическую ценность автоматически формируемых тестовых материалов.

В четвёртой главе изложены результаты исследования, направленного на разработку алгоритма автоматической генерации дистракторов для тестовых

заданий на русском языке. Основная цель исследования заключалась в создании метода, способного формировать неверные варианты ответов, которые являются правдоподобными, тематически релевантными и отражают типичные когнитивные ошибки обучающихся, что позволяет повысить диагностическую ценность контрольно-измерительных материалов.

В рамках предлагаемого подхода реализуется генерация дистракторов на основе моделирования процесса рассуждений, приводящего к правильному ответу, с последующим целенаправленным внесением типичных ошибок в полученную цепочку. Такой механизм позволяет формировать варианты ответов, которые не только семантически согласуются с формулировкой вопроса, но и воспроизводят характерные заблуждения, возникающие при усвоении учебного материала.

Алгоритм реализует трёхэтапную процедуру генерации. На первом этапе формируется корректная цепочка рассуждений R_c , описывающая последовательность логических шагов, ведущих к правильному ответу A на вопрос Q :

$$R_c = \text{LLM}(P_{\text{correct}}), \quad P_{\text{correct}} = f(Q, A, C),$$

где функция f формирует запрос к языковой модели, объединяя текст вопроса, правильный ответ, исходный контент и инструктивные элементы.

На втором этапе на основе цепочки R_c строятся ошибочные рассуждения, каждое из которых завершается неверным выводом:

$$R_e = \text{LLM}(P_{\text{mistakes}}), \quad P_{\text{mistakes}} = g(R_c, N),$$

где g определяет формат запроса для получения N вариантов с внесёнными когнитивными искажениями.

На заключительном этапе из ответа модели извлекается набор дистракторов:

$$D = \{d_1, d_2, \dots, d_N\} = \text{parse}(R_e),$$

после чего выполняется их автоматическая валидация на уникальность и достаточное семантическое отличие от правильного ответа. В случае недостаточного количества валидных элементов процедура генерации повторяется.

Экспериментальная проверка алгоритма проводилась на двух русскоязычных наборах данных. Первый набор – RuOpenBookQA, который содержит вопросы по естественно-научным дисциплинам с вручную составленными неверными вариантами ответов. Второй набор – RuWorldTree, включающий задания с графами объяснений и ориентированный на проверку способности алгоритма работать с вопросами, требующими понимания причинно-следственных связей. Из каждого корпуса случайным образом отбиралось по 200 вопросов; для каждого вопроса формировались три дистрактора, что соответствует стандартной структуре тестовых заданий с четырьмя вариантами ответа. Таким образом, для каждого алгоритмического подхода было сформировано по 1200 дистракторов, что обеспечивает достаточный объём данных для устойчивой сравнительной оценки.

Характеристики сгенерированных дистракторов оценивались по следующим показателям:

– правдоподобие, определяемое на основе трёхбалльной экспертной шкалы (0 — «неправдоподобно», 1 — «частично правдоподобно», 2 — «полностью правдоподобно»);

- семантическое сходство дистрактора с вопросом;
- семантическое различие дистрактора с правильным ответом;
- сходство с эталонными дистракторами, подготовленными вручную;
- внутригрупповое разнообразие, характеризующее различие между дистракторами одного задания.

Для сравнительного анализа использовались два базовых алгоритмических подхода: прямая генерация неверных ответов без моделирования рассуждений и алгоритм семантической трансформации правильного ответа. Результаты экспериментальной оценки представлены в таблицах 5 и 6.

Таблица 5 – Оценка качества дистракторов на наборе RuOpenBookQA

Способ	Экспертная оценка	Сходство дистрактора с вопросом	Различие дистрактора с ответом	Сходство с эталонными дистракторами	Внутригрупповое разнообразие
Семантическое изменение	1,07	0,60	0,55	0,35	0,52
Базовая генерация	1,20	0,68	0,62	0,42	0,61
Предлагаемый алгоритм	1,66	0,72	0,68	0,58	0,74

Как показывают полученные данные, предложенный алгоритм превосходит оба базовых решения по всем анализируемым показателям. Наиболее выраженное преимущество наблюдается по экспертной оценке правдоподобия, сходству с эталонными дистракторами и внутригрупповому разнообразию. Для проверки устойчивости полученных результатов была проведена дополнительная серия экспериментов, в которой рассматривались задания, характеризующиеся более сложной структурой заданий.

Таблица 6 – Оценка качества дистракторов на наборе RuWorldTree

Способ	Экспертная оценка	Сходство дистрактора с вопросом	Различие дистрактора с ответом	Сходство с эталонными дистракторами	Внутригрупповое разнообразие
Семантическое изменение	0,98	0,47	0,52	0,31	0,51
Базовая генерация	1,13	0,53	0,60	0,40	0,57
Предлагаемый алгоритм	1,60	0,69	0,65	0,56	0,64

Анализ экспериментальных данных показал, что включение этапа явного построения рассуждений и целенаправленного внесения ошибок обеспечивает улучшение диагностически значимых характеристик дистракторов по сравнению с альтернативными алгоритмическими подходами. Экспертные оценки составили 1,66 и 1,60 балла против 1,0–1,2 у базовых решений. Значения показателя сходства с эталонными дистракторами 0,58 и 0,56 свидетельствуют о близости автоматически сформированных вариантов к дистракторам, подготовленным вручную.

Сводный показатель 22,7 % определялся как среднее значение относительных приростов по пяти показателям качества дистракторов относительно алгоритма прямой генерации на наборе RuOpenBookQA.

В заключение главы сделан вывод о том, что разработанный алгоритм обеспечивает автоматическую генерацию дистракторов с улучшенными диагностически значимыми характеристиками, сопоставимыми с результатами их ручного проектирования и превосходящими исследованные алгоритмические решения автоматической генерации дистракторов.

В заключении обобщены основные результаты выполненного исследования, направленного на разработку алгоритмов автоматизированной генерации тестовых заданий для русскоязычных образовательных текстов. В работе последовательно решены задачи извлечения значимой информации, формирования тестовых заданий различных форматов и автоматической верификации их корректности и педагогической пригодности. Показано, что использование доменно-настроенных интегральных оценок, многоагентной верификации и генерации дистракторов на основе моделирования рассуждений позволяет повысить диагностическую ценность автоматически формируемых контрольно-измерительных материалов.

Полученные результаты подтверждают возможность построения воспроизводимых алгоритмических решений и вносят вклад в развитие методов извлечения и обработки информации для интеллектуальных образовательных систем. Перспективы дальнейших исследований связаны с расширением используемых источников знаний, включая внешние базы и онтологии, а также с адаптацией предложенных алгоритмов для применения в специализированных интеллектуальных образовательных системах.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

Основные результаты работы заключаются в следующем:

1. Проведен комплексный анализ существующих алгоритмов автоматической генерации тестовых заданий, в результате которого выявлены их ключевые ограничения применительно к русскоязычному материалу (в частности, поверхностное перефразирование, отсутствие механизмов верификации и слабый учет дидактических требований). Доказана необходимость перехода к интегрированным контурам генерации.
2. Сформирован алгоритм семантического анализа образовательного контента и извлечения из него ключевой информации, релевантной для последующей генерации заданий. Доказано экспериментально, что применение параметрической метрики с доменно-специфичными весовыми коэффициентами позволило превзойти все рассмотренные базовые решения, при этом по сравнению с лучшей базовой моделью (O1) усреднённый по предметным областям прирост интегральной оценки качества аннотаций составил 13,2 %.
3. Создан алгоритм генерации многоформатных тестовых заданий, успешно формирующий задания с выбором одного или нескольких вариантов ответа, задания открытой формы, а также на установление соответствия и правильной последовательности. Внедрение в данный алгоритм разработанной мультиагентной системы верификации позволило автоматически исключать некорректные формулировки и снизить общую долю дефектных заданий на 84,5 %.

4. Разработан алгоритм интеллектуального формирования дистракторов на основе когнитивного моделирования ошибок обучающихся. Эксперименты на наборах данных RuOpenBookQA и RuWorldTree подтвердили, что имитация искажений в цепочке рассуждений превосходит подходы на основе семантической близости, повышая интегральное качество (правдоподобность и дидактическую ценность) неверных вариантов ответа на 22,7 %.
5. Созданные алгоритмы реализованы в виде программного комплекса для автоматизированной генерации контрольно-измерительных материалов (получено 11 свидетельств о государственной регистрации программ для ЭВМ). Проведенная экспериментальная апробация подтвердила высокую эффективность предложенных решений и дидактическое качество генерируемых материалов, готовых к интеграции в современные образовательные системы.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в научных журналах, входящих в перечень ВАК:

1. **Дагаев А.Е., Попов Д.И.** Сравнение автоматического обобщения текстов на русском языке // Программные системы и вычислительные методы. 2024. № 4. С. 13-22. DOI: 10.7256/2454-0714.2024.4.69474
2. **Дагаев А.Е., Попов Д.И.** Оценка качества интеллектуального перефразирования текстов на русском языке. // Моделирование, оптимизация и информационные технологии. 2024. Т. 12. № 4 (47). DOI: 10.26102/2310-6018/2024.47.4.038
3. **Дагаев А.Е.** Метод генерации вопросов закрытого типа с использованием LLM // Моделирование, оптимизация и информационные технологии. 2025. Т. 13. № 1 (48). DOI: 10.26102/2310-6018/2025.48.1.021
4. **Дагаев А.Е.** Искусственный интеллект в задаче генерации дистракторов для тестовых заданий. // Моделирование, оптимизация и информационные технологии. 2025. Т. 13. № 2 (49). DOI: 10.26102/2310-6018/2025.49.2.028

Публикации в материалах (трудах) международных научных конференций:

5. **Дагаев, А. Е.** Методы интеллектуальной генерации тестовых материалов / А. Е. Дагаев, Д. И. Попов // Математическое и компьютерное моделирование : сборник материалов XII международной научной конференции, Омск, 14 марта 2025 года. – Омск: Омский государственный университет им. Ф.М. Достоевского, 2025. – С. 303-305.
6. **Дагаев, А. Е.** Использование искусственного интеллекта в системах управления обучением сотрудников / А. Е. Дагаев // Управление устойчивым инновационным развитием России в условиях цифровой трансформации : Материалы XVI Международной научно-практической конференции памяти заслуженного деятеля науки РФ В.И. Кравцовой, Москва, 06 декабря 2024 года. – Москва: Московский политехнический университет, 2025. – С. 529-532.

7. **Дагаев, А. Е.** Применение моделей искусственного интеллекта в задаче извлечения информации из текста / А. Е. Дагаев // *Фундаментальные и прикладные исследования молодых учёных : сборник материалов IX Международной научно-практической конференции студентов, аспирантов и молодых учёных, приуроченной к 95-летию основания ФГБОУ ВО «СибАДИ»*, Омск, 24 апреля – 25 2025 года. – Омск: Сибирский государственный автомобильно-дорожный университет (СибАДИ), 2025. – С. 659-662.
8. **Дагаев, А. Е.** Алгоритм интеллектуальной генерации вопросов с множественным выбором / А. Е. Дагаев // *Фундаментальные и прикладные исследования молодых учёных : сборник материалов IX Международной научно-практической конференции студентов, аспирантов и молодых учёных, приуроченной к 95-летию основания ФГБОУ ВО «СибАДИ»*, Омск, 24 апреля – 25 2025 года. – Омск: Сибирский государственный автомобильно-дорожный университет (СибАДИ), 2025. – С. 651-654.
9. **Дагаев, А. Е.** Методы верификации автоматически сгенерированных тестовых заданий / А. Е. Дагаев // *Фундаментальные и прикладные исследования молодых учёных : сборник материалов IX Международной научно-практической конференции студентов, аспирантов и молодых учёных, приуроченной к 95-летию основания ФГБОУ ВО «СибАДИ»*, Омск, 24 апреля – 25 2025 года. – Омск: Сибирский государственный автомобильно-дорожный университет (СибАДИ), 2025. – С. 655-658.

Свидетельства о государственной регистрации программ для ЭВМ:

10. Свидетельство о государственной регистрации программы для ЭВМ №2023686997. «Автоматизированная система генерации контрольно-измерительных материалов» / Дагаев А.Е. – Роспатент, 2023.
11. Свидетельство о государственной регистрации программы для ЭВМ №2024612459. «Подсистема верификации контрольно-измерительных материалов» / Дагаев А.Е. – Роспатент, 2024.
12. Свидетельство о государственной регистрации программы для ЭВМ №2024691569. «Адаптивный модуль формирования дистракторов» / Дагаев А.Е. – Роспатент, 2024.
13. Свидетельство о государственной регистрации программы для ЭВМ №2024691716. «Интеллектуальная система анализа качества дистракторов» / Дагаев А.Е. – Роспатент, 2024.
14. Свидетельство о государственной регистрации программы для ЭВМ №2025610015. «Инструмент оценки контентной валидности теста». / Дагаев А.Е. – Роспатент, 2025.
15. Свидетельство о государственной регистрации программы для ЭВМ №2025614018. «Подсистема автоматической оценки правильности ответов» / Дагаев А.Е. – Роспатент, 2025.
16. Свидетельство о государственной регистрации программы для ЭВМ №2025661041. «Модуль анализа соответствия тестовых заданий образовательным стандартам» / Дагаев А.Е. – Роспатент, 2025.

17. Свидетельство о государственной регистрации программы для ЭВМ №2025661449. «Инструмент автоматической категоризации тестовых вопросов» / Дагаев А.Е. – Роспатент, 2025.
18. Свидетельство о государственной регистрации программы для ЭВМ №2025661372. «Модуль автоматического извлечения структурированной информации из графических данных» / Дагаев А.Е. – Роспатент, 2025.
19. Свидетельство о государственной регистрации программы для ЭВМ №2025666271. «Инструмент генерации вопросов причинно-следственных связей» / Дагаев А.Е. – Роспатент, 2025.
20. Свидетельство о государственной регистрации программы для ЭВМ №2025666095. «Модуль формирования тестовых заданий на основе табличных данных» / Дагаев А.Е. – Роспатент, 2025.

Дагаев Александр Евгеньевич

**РАЗРАБОТКА АЛГОРИТМОВ ИЗВЛЕЧЕНИЯ И ОБРАБОТКИ
ИНФОРМАЦИИ ДЛЯ ГЕНЕРАЦИИ МНОГОФОРМАТНЫХ ТЕСТОВЫХ
ЗАДАНИЙ**

Автореферат

диссертации на соискание ученой степени
кандидата технических наук