

Метод диалоговой диагностики знаний на основе двухконтурного взаимодействия с большой языковой моделью

П.В. Анни, М.Г. Жабицкий, В.С. Царев, Ю.В. Кальсина, И.А. Парфенов, М.О. Соколинская, М.Д. Забродин

Аннотация -- В статье предложен метод диалоговой диагностики знаний на основе двухконтурного взаимодействия с большой языковой моделью, обеспечивающий разделение процессов подготовки предметного содержания и проведения индивидуальной диагностики знаний. Метод основан на организации двух взаимосвязанных итерационных контуров: «эксперт-преподаватель – LLM», предназначенного для формирования и экспертной верификации экзаменационного контента, и «студент – LLM», реализующего адаптивный диагностический диалог с последовательным уточнением модели знаний обучающегося. Предложены принципы распределения функций между преподавателем, большой языковой моделью и информационной системой, при которых LLM выполняет интеллектуально трудоемкие операции подготовки контента и анализа ответов, тогда как окончательное предметное решение остается за преподавателем. На основе предложенного метода разработана архитектура интеллектуальной системы, реализован программный прототип и проведена его экспериментальная апробация в образовательном процессе. Полученные результаты подтверждают практическую реализуемость предложенного подхода, возможность существенного снижения трудоемкости подготовки фонда оценочных средств и проведения индивидуальной диалоговой диагностики при сохранении воспроизводимости, прослеживаемости и экспертной достоверности результатов.

Ключевые - слова: диалоговая диагностика знаний, большие языковые модели, LLM, интеллектуальные образовательные системы, адаптивный диалог, двухконтурное взаимодействие, экспертная валидация, фонд оценочных средств, контроль знаний, цифровая образовательная среда.

1. ВВЕДЕНИЕ

Современный образовательный процесс сталкивается с фундаментальным противоречием между качеством диагностики знаний и возможностью ее массового применения. Наиболее достоверной формой проверки остается индивидуальный устный экзамен, в ходе которого преподаватель не предъявляет обучающемуся заранее фиксированную последовательность вопросов, а последовательно исследует уровень его понимания, уточняя ответы и выявляя границы сформированных компетенций [1]. Однако именно такая форма диагностики требует значительных затрат времени

высококвалифицированного преподавателя и поэтому практически не масштабируется.

В результате массовый контроль знаний преимущественно строится на тестировании и стандартизованных экзаменационных заданиях. Такие технологии обеспечивают высокую производительность, воспроизводимость процедуры и объективность формальной оценки, однако ограниченно выявляют глубину понимания предметной области, способность к рассуждению, установлению взаимосвязей между понятиями и переносу знаний в новые ситуации. Тем самым традиционная организация контроля знаний вынуждает искать компромисс между качеством диагностики и эффективностью образовательного процесса.

Дополнительную сложность создает высокая динамика развития науки, техники и инженерной практики. Учебные материалы, профессиональные требования и состав формируемых компетенций непрерывно изменяются, тогда как подготовка и сопровождение экзаменационных материалов остаются одной из наиболее трудоемких задач преподавателя. В результате содержание контроля знаний нередко обновляется значительно медленнее, чем содержание самой дисциплины [2].

Появление больших языковых моделей (Large Language Models, LLM) впервые создает предпосылки для масштабирования содержательного индивидуального диалога между преподавателем и обучающимся. Вместе с тем непосредственное использование LLM для принятия экзаменационных решений не обеспечивает требуемых уровней достоверности, воспроизводимости и управляемости процедуры диагностики. Поведение языковой модели зависит от используемой архитектуры, настроечных промптов и контекста взаимодействия, что делает необходимым сохранение экспертного контроля над содержанием дисциплины и результатами проверки знаний.

В этой связи объектом автоматизации должны становиться не функции преподавателя по принятию итогового оценочного решения, а интеллектуально трудоемкие процессы подготовки экзаменационного контента и ведения диагностического диалога. Большие языковые модели способны существенно повысить производительность этих процессов, тогда как ответственность за формирование содержания дисциплины, критериев оценивания и итоговую

интерпретацию результатов должна сохраняться за преподавателем.

Таким образом, возникает научная задача разработки метода организации диалоговой диагностики знаний, сочетающего преимущества больших языковых моделей с экспертным управлением образовательным процессом. Представляется перспективным разделение подготовки экзаменационного контента и проведения проверки знаний на два взаимосвязанных итерационных контура: «эксперт – LLM», обеспечивающий формирование и верификацию вопросов, эталонных ответов и критериев оценивания, и «студент – LLM», реализующий последовательную диагностику знаний посредством адаптивного диалога.

Целью настоящей работы является разработка и экспериментальная апробация метода построения LLM-ориентированной системы диалоговой диагностики знаний, обеспечивающей формирование верифицированного экзаменационного контента, последовательное уточнение ответов обучающегося и экспертную валидацию результатов проверки.

Гипотеза исследования состоит в том, что разделение процессов подготовки экзаменационного контента и проведения проверки знаний посредством двух итерационных контуров взаимодействия «эксперт – LLM» и «студент – LLM» позволяет повысить содержательность, гибкость и эффективность диагностики знаний без передачи языковой модели окончательной ответственности за принимаемые решения.

В работе предложены метод построения LLM-ориентированной системы диалоговой диагностики знаний, принцип индивидуальной адаптивной диагностики, рассматривающий проверку знаний как процесс последовательного уточнения модели знаний обучающегося, а также архитектура программного прототипа, реализующего предложенный подход и экспериментально апробированного в реальном образовательном процессе.

II. АНАЛИЗ СУЩЕСТВУЮЩИХ ПОДХОДОВ И ПОСТАНОВКА ЗАДАЧИ

2.1. Контроль знаний как информационно-управляющая подсистема образовательного процесса

Контроль знаний традиционно рассматривается как процедура оценки результатов обучения. Однако в системе образования его роль значительно шире. Контроль представляет собой информационно-управляющую подсистему, формирующую обратную связь, необходимую для принятия решений на различных уровнях управления образовательным процессом [3].

На уровне отдельной дисциплины результаты контроля используются преподавателем для определения степени достижения образовательных целей, выявления пробелов в знаниях, выбора последующих педагогических воздействий и корректировки индивидуальной траектории освоения материала. Таким образом, основными функциями

контроля становятся не только оценочная, но прежде всего управляющая и развивающая [4].

При переходе к индивидуальным образовательным траекториям значение контроля возрастает. Для принятия обоснованных решений уже недостаточно итоговой оценки или результата отдельного тестирования. Необходима актуальная информация о состоянии знаний обучающегося, динамике их изменения, устойчивости усвоения материала и характере возникающих затруднений. Следовательно, главным результатом контроля становится не отметка, а формируемая модель текущего состояния знаний обучающегося.

Одновременно результаты контроля используются и на уровне образовательной программы. Руководители программ и деканат принимают решения о сопровождении индивидуальных образовательных траекторий, своевременном выявлении рисков академического отставания, корректировке последовательности освоения дисциплин и достижении целевых компетенций. При этом предметная интерпретация результатов остается исключительной компетенцией преподавателя, тогда как административный уровень работает с агрегированной информацией, полученной по результатам многих диагностических процедур.

Таким образом, система контроля знаний одновременно обслуживает несколько уровней управления: предметный, дисциплинарный и программный. Это предъявляет требования не только к точности диагностики, но и к своевременности получения информации, возможности ее накопления, сопоставления и последующего анализа.

2.2. Ограничения существующих подходов к контролю знаний

Современные формы контроля знаний обеспечивают различные сочетания глубины диагностики, производительности и трудоемкости, однако ни одна из них не обеспечивает одновременно всех требований, предъявляемых к современной образовательной системе [5].

Закрытые тесты отличаются высокой производительностью, стандартизованностью и простотой автоматической обработки, однако преимущественно фиксируют правильность отдельных ответов и дают ограниченную информацию о структуре знаний и логике рассуждений обучающегося [6].

Фиксированные экзаменационные билеты позволяют обеспечить необходимый охват содержания дисциплины, но практически не учитывают индивидуальный уровень подготовки студента и не позволяют адаптировать последовательность вопросов по результатам уже полученных ответов [7,8].

Индивидуальный устный экзамен, напротив, обладает максимальной диагностической информативностью. Опытный преподаватель не следует заранее заданному сценарию, а постепенно уточняет модель знаний обучающегося, прекращая проверку очевидно усвоенного материала и концентрируя внимание на областях неопределенности

и границах компетентности. Однако именно эта форма контроля требует значительных затрат времени высококвалифицированного преподавателя и поэтому плохо масштабируется.

Кроме того, традиционные формы контроля ориентированы преимущественно на календарную организацию образовательного процесса. Проверка проводится в заранее определенные моменты времени независимо от индивидуальной динамики освоения материала. Вместе с тем практика показывает, что скорость формирования знаний существенно различается между обучающимися. Один студент осваивает новый раздел постепенно, другому достаточно короткого периода интенсивной самостоятельной работы для достижения требуемого уровня понимания. Следовательно, наряду с календарной логикой контроля объективно существует событийная логика, при которой основанием для проверки становится изменение фактического состояния знаний.

Совместить регулярность календарного контроля с индивидуальной скоростью обучения традиционными средствами крайне сложно вследствие высокой трудоемкости индивидуальной диагностики.

2.3. Цифровизация и большие языковые модели как источник новых возможностей

Появление больших языковых моделей совпало с общим процессом цифровой трансформации образования и принципиально изменило технологические возможности организации контроля знаний.

В отличие от традиционных средств автоматизации, LLM способны интерпретировать свободно сформулированные ответы, поддерживать длительный содержательный диалог, учитывать контекст предыдущего взаимодействия, формировать уточняющие вопросы и сопоставлять ответы обучающегося с предметным содержанием на смысловом уровне [9,10]. Одновременно цифровая среда обеспечивает полную регистрацию процесса взаимодействия, хранение истории диалога и накопление структурированных данных о результатах диагностики.

Сочетание этих свойств создает ряд новых возможностей.

Во-первых, становится достижимым масштабирование индивидуального содержательного диалога без пропорционального увеличения нагрузки преподавателей.

Во-вторых, существенно возрастает производительность контроля знаний. Это позволяет перейти от эпизодических контрольных мероприятий к регулярному мониторингу образовательного процесса и принимать управленческие решения на основании значительно более оперативной информации.

В-третьих, цифровая фиксация всей процедуры делает возможной объективизацию контроля. Решение начинает опираться не только на итоговую оценку, но и на полный журнал взаимодействия: вопросы, ответы, уточнения, основания формирования предварительного

заключения и историю изменения знаний обучающегося.

В-четвертых, возникает возможность накопления больших массивов данных о знаниях. Если традиционная система сохраняет преимущественно итоговые оценки, то цифровая образовательная среда позволяет формировать модели знаний отдельных студентов, анализировать динамику освоения дисциплин, выявлять устойчивые ошибки, оценивать эффективность учебных материалов и получать агрегированную аналитику на уровне образовательных программ.

Наконец, существенно меняется сама философия повторной аттестации. Обучающийся получает возможность подтвердить повышение уровня подготовки не посредством субъективного пересмотра ранее принятого решения, а путем повторной диагностики, документально подтверждающей изменение состояния его знаний. Тем самым повторная проверка становится процедурой объективного подтверждения вновь приобретенных компетенций.

Вместе с тем генеративная природа LLM порождает и новые риски: возможность выхода за пределы предметной области, зависимость результатов от используемой модели и промптов, снижение воспроизводимости отдельных решений и необходимость сохранения экспертного контроля над содержанием и результатами проверки.

2.4. Распределение ролей в LLM-ориентированной системе контроля знаний

Использование больших языковых моделей не изменяет распределения ответственности между участниками образовательного процесса.

Эксперт-преподаватель остается конечным субъектом принятия решений по предметной области [11]. Именно он определяет содержание дисциплины, критерии достижения образовательных результатов, правила оценивания, допустимые границы интерпретации ответов и принимает окончательное решение по результатам диагностики.

Большая языковая модель выполняет интеллектуально трудоемкие операции, делегированные преподавателем: участвует в формировании экзаменационного контента, анализирует ответы обучающегося, ведет диагностический диалог, формирует уточняющие вопросы и подготавливает предварительное экспертное заключение.

Информационная система обеспечивает хранение предметного содержания, регистрацию истории взаимодействия, соблюдение установленных ограничений, накопление диагностических данных и формирование аналитической информации для различных уровней управления.

Руководители образовательных программ и деканат используют агрегированные результаты диагностики для сопровождения индивидуальных образовательных траекторий, анализа достижения компетенций и принятия организационных решений. При этом они не вмешиваются в предметную интерпретацию

результатов контроля, которая остается исключительной компетенцией преподавателя.

Таким образом, LLM усиливает интеллектуальные возможности эксперта, но не заменяет его как субъекта предметного решения [12].

2.5. Постановка задачи исследования

Проведенный анализ показывает, что современная образовательная система нуждается в средствах контроля знаний, обеспечивающих одновременно глубину индивидуальной диагностики, высокую производительность, регулярность обратной связи, прозрачность процедуры и накопление данных для управления образовательными траекториями.

Для достижения этой цели процессы подготовки предметного содержания и проведения диагностики должны быть разделены на два взаимосвязанных контура. В контуре «эксперт-преподаватель - LLM» формируются и верифицируются вопросы, эталонные ответы, критерии оценивания и правила проведения проверки. Организация взаимодействия специализированных интеллектуальных компонентов соответствует современному подходу к построению агентных LLM-систем [13]. В контуре «студент - LLM - преподаватель» реализуется индивидуальный диагностический диалог, формируется предварительное заключение и выполняется окончательная экспертная валидация результата.

Проверка знаний рассматривается не как предъявление фиксированной последовательности вопросов, а как последовательное получение и уточнение информации о состоянии знаний обучающегося. Каждый следующий вопрос должен учитывать всю предшествующую историю взаимодействия и быть направлен на проверку областей, в отношении которых сохраняется неопределенность.

Пусть C_E - верифицированная экспертом модель предметного содержания, включающая элементы знаний, связи между ними, вопросы, эталонные ответы и критерии оценивания;

- R_E - заданные экспертом правила проведения диагностики, формирования уточняющих вопросов, оценивания и завершения диалога;
- $\bar{K}_i$ - текущая оценка состояния знаний обучающегося после i -го шага взаимодействия;
- H_i - накопленная история диалога, включающая заданные вопросы, ответы обучающегося и сформированные уточнения.

На шаге i обучающемуся предъявляется вопрос Q_i , после чего формируется ответ A_i . На основании нового ответа, предыдущей оценки знаний, истории диалога и верифицированного предметного содержания выполняется уточнение модели знаний:

$$\bar{K}_i = \Phi(\bar{K}_{i-1}, Q_i, A_i, H_i, C_E).$$

Следующий вопрос определяется с учетом текущей оценки знаний, накопленной истории, предметного содержания и правил, установленных экспертом:

$$Q_{i+1} = \Psi(\bar{K}_i, H_i, C_E, R_E).$$

Общая последовательность диагностического взаимодействия имеет вид:

$$Q_1 \rightarrow A_1 \rightarrow \bar{K}_1 \rightarrow Q_2 \rightarrow A_2 \rightarrow \bar{K}_2 \rightarrow \dots \rightarrow Q_n \rightarrow A_n \rightarrow \bar{K}_n.$$

Принципиальным является то, что очередной вопрос формируется не только на основании последнего ответа, а с учетом всей накопленной оценки состояния знаний обучающегося.

После каждого шага система должна определить:

- какие элементы знания подтверждены;
- какие элементы содержат ошибки или противоречия;
- какие области остаются неопределенными;
- требуется ли дополнительное уточнение;
- какой вопрос целесообразно задать следующим;
- достигнуты ли условия завершения диалога.

После окончания диагностической сессии формируется предварительное заключение:

$$D_s = \Omega(\bar{K}_n, H_n, C_E, R_E),$$

где D_s включает предварительную оценку, подтвержденные элементы знания, выявленные пробелы, основания выставления балла и рекомендации по дальнейшему обучению.

Окончательный предметный результат определяется экспертом-преподавателем:

$$D_s^* = V_E(D_s, H_n, C_E),$$

где V_E - процедура экспертной проверки, корректировки и утверждения результата.

Тем самым языковая модель формирует диагностическое заключение, но не обладает полномочиями окончательного предметного решения.

Утвержденные результаты различных диагностических процедур и дисциплин формируют информационную основу индивидуальной образовательной траектории:

$$T_s = \Gamma(D_{s1}^*, D_{s2}^*, \dots, D_{sm}^*, P_s),$$

где T_s - текущее представление индивидуальной образовательной траектории студента s ;

D_{sj}^* - утвержденные результаты диагностики по отдельным дисциплинам и контрольным событиям;

P_s - сведения о последовательности, темпе и динамике освоения образовательной программы.

На предметном уровне решения принимает эксперт-преподаватель, тогда как на уровне образовательной программы деканат и руководители программы используют совокупность утвержденных данных для сопровождения индивидуальной траектории, выявления рисков и принятия организационных решений.

Таким образом, задача исследования состоит в разработке метода и программной архитектуры, обеспечивающих:

- формирование верифицированного предметного содержания;
- последовательную индивидуальную диагностику знаний;
- адаптивное формирование уточняющих вопросов;
- подготовку прозрачного предварительного заключения;

- окончательную предметную валидацию экспертом-преподавателем;
- накопление данных о состоянии и динамике знаний;
- использование утвержденных результатов для сопровождения индивидуальных образовательных траекторий;
- снижение трудоемкости и повышение регулярности контроля.

Целевую постановку можно выразить следующим образом: пребуется организовать последовательный диалог, обеспечивающий формирование достаточно достоверной и практически полезной модели знаний обучающегося при ограниченном числе вопросов, времени проверки и затратах преподавательского труда, с сохранением за экспертом-преподавателем окончательного предметного решения и возможностью последующего использования результатов на уровне управления образовательной программой.

В настоящей работе реализуется базовый механизм последовательного уточнения ответов, подготовки предварительного заключения и экспертной валидации результата. Формальный критерий информационно-оптимального выбора каждого следующего вопроса и полная междисциплинарная модель индивидуальной образовательной траектории рассматриваются как направления дальнейшего развития подхода.

III. ТРЕБОВАНИЯ К LLM-ОРИЕНТИРОВАННОЙ ТЕХНОЛОГИИ ДИАГНОСТИКИ ЗНАНИЙ

3.1. Требования к системе диагностики и управления образовательным процессом

Предлагаемая LLM-ориентированная технология рассматривает контроль знаний не как изолированную экзаменационную процедуру, а как элемент системы управления образовательным процессом. Следовательно, требования должны формулироваться не только по отношению к процедуре оценивания, но и по отношению к тем функциям, которые результаты

диагностики выполняют на различных уровнях образовательной системы. Онтологическая схема требований представлена на рис. 1.

Центральным объектом модели является индивидуальная модель знаний обучающегося. В отличие от традиционного подхода, где итогом контроля является преимущественно оценка, в предлагаемой системе основным результатом диагностической процедуры становится актуализированное представление о состоянии знаний студента. Итоговая оценка рассматривается как одно из производных представлений этой модели и служит, прежде всего, для формализации образовательного результата. Формирование модели знаний определяется совокупностью пяти взаимосвязанных требований.

Первое требование связано с образовательными функциями системы. Диагностика должна обеспечивать не только измерение достигнутого уровня подготовки, но и выполнение управляющей и развивающей функций. Полученные результаты должны использоваться преподавателем для выбора дальнейших педагогических воздействий, корректировки последовательности изучения материала и определения готовности обучающегося к переходу на следующий этап освоения дисциплины. Таким образом, контроль знаний становится источником информации для управления образовательным процессом.

Второе требование состоит в обеспечении индивидуальной диагностики. Процедура проверки должна учитывать содержание предшествующего диалога и формировать индивидуальную последовательность вопросов, направленную на исследование областей неопределенности. Повторная проверка уже подтвержденных знаний не должна становиться самоцелью. Основная задача диагностического взаимодействия заключается в максимально полном восстановлении модели знаний обучающегося при минимально необходимом объеме диалога.



Рис. 1. Онтологическая схема требований к системе диагностики и управления образовательным процессом.

Третье требование определяет распределение ролей субъектов образовательного процесса. Эксперт-преподаватель формирует предметную модель дисциплины, определяет критерии достижения

образовательных результатов и принимает окончательное решение по итогам диагностики.

Студент является активным участником диагностического диалога и получает возможность подтверждать изменение уровня своей подготовки. Руководители образовательной программы используют только предметно утвержденные результаты для сопровождения индивидуальных образовательных траекторий и принятия организационных решений. Тем самым предметные и административные функции остаются разделенными, несмотря на использование общей информационной среды.

Четвертое требование связано с организацией календарного и событийного контроля знаний. Обязательные контрольные мероприятия обеспечивают регулярность образовательного процесса и своевременное получение обратной связи. Одновременно система должна поддерживать диагностику, инициируемую изменением фактического состояния знаний обучающегося. Такое сочетание позволяет учитывать индивидуальную скорость освоения материала, не нарушая общей логики образовательного процесса.

Пятое требование заключается в поддержке повторного подтверждения знаний. Поскольку объектом управления является текущее состояние знаний, а не результат отдельной экзаменационной процедуры, система должна обеспечивать возможность повторной диагностики после дополнительного обучения. При этом новый результат не отменяет предыдущий, а дополняет историю наблюдений, позволяя документально подтвердить образовательный прогресс. Такой подход переводит передачу из

административной процедуры в процедуру объективного подтверждения изменения уровня подготовки.

Совместная реализация перечисленных требований обеспечивает переход от эпизодического контроля к непрерывному сопровождению образовательного процесса. Индивидуальная модель знаний становится связующим звеном между предметным решением преподавателя, образовательной деятельностью студента и управлением образовательной программой. Именно она обеспечивает возможность накопления информации о динамике формирования компетенций и создает основу для сопровождения индивидуальных образовательных траекторий.

Таким образом, требования к системе диагностики определяют не только принципы организации экзаменационной процедуры, но и место контроля знаний в общей системе управления образованием. Их реализация требует создания интеллектуального программного инструмента, способного поддерживать подготовку предметного содержания, проведение индивидуального диагностического диалога, экспертную валидацию результатов и их интеграцию в цифровую образовательную среду. Эти требования рассматриваются в следующем подразделе.

3.2. Требования к интеллектуальному программному инструменту

Из требований к системе диагностики и управления образовательным процессом непосредственно следуют требования к интеллектуальному программному инструменту, реализующему предложенный подход.



Рис. 2. Онтологическая схема требований к интеллектуальному программному инструменту.

В отличие от традиционных систем тестирования, такой инструмент должен поддерживать не только

проведение экзаменационной процедуры, но и полный жизненный цикл диагностического процесса - от подготовки предметного содержания до передачи утвержденных результатов в образовательную информационную среду. Обобщенная онтологическая

схема требований представлена на рис. 2. В центре схемы располагаются пять взаимосвязанных функциональных требований, определяющих основные этапы обработки информации. Первым этапом является подготовка предметного содержания, включающая формирование и актуализацию вопросов, критериев оценивания и других диагностических материалов. Далее реализуется проведение индивидуального диагностического диалога, в ходе которого анализируются ответы обучающегося и формируется последовательность уточняющих вопросов. Следующий блок обеспечивает использование больших языковых моделей как интеллектуального механизма анализа естественного языка и генерации новых элементов диалога. Полученные результаты передаются в контур экспертной валидации, где преподаватель принимает окончательное предметное решение. Завершающим этапом является интеграция утвержденных результатов с внешними информационными системами образовательной организации.

Предложенная структура принципиально разделяет интеллектуальные функции языковой модели и полномочия преподавателя [14]. LLM участвует в подготовке диагностического содержания, анализе ответов и формировании предварительных выводов, однако итоговое решение остается за экспертом-преподавателем. Такое разделение обеспечивает одновременно высокую производительность обработки информации и сохранение предметной достоверности результатов.

Функционирование инструмента основано на взаимодействии нескольких категорий участников образовательного процесса. Студент выступает источником диагностической информации, преподаватель определяет предметное содержание и утверждает результаты контроля, администратор обеспечивает функционирование программной среды, а внешние образовательные системы используют только утвержденные данные для сопровождения образовательного процесса. Такое разделение ролей исключает смешение предметных и административных функций и позволяет реализовать разграничение прав доступа к информации.

Существенным элементом архитектуры является выделение информационных объектов, сопровождающих диагностическую процедуру. К ним относятся предметный контент, конфигурация диагностической процедуры, диалоговая сессия, предварительное заключение LLM, экспертное решение и служебные метаданные. Подобное разделение обеспечивает прослеживаемость происхождения результатов, возможность повторного анализа диагностической процедуры и воспроизводимость проведенных экспериментов.

Использование интеллектуального программного инструмента предполагает интеграцию с существующей цифровой образовательной инфраструктурой. Наиболее естественными направлениями интеграции являются системы управления обучением (LMS), электронные

образовательные среды, системы учета контингента обучающихся, электронные ведомости, сервисы аутентификации и аналитические подсистемы сопровождения индивидуальных образовательных траекторий. При этом во внешние системы передаются только результаты, прошедшие экспертную валидацию, тогда как первичные диалоги и промежуточные выводы LLM используются исключительно внутри диагностического контура.

Особое место в архитектуре занимают сквозные требования, распространяющиеся на все функциональные компоненты инструмента. К ним относятся приоритет экспертного решения, прозрачность и прослеживаемость диагностической процедуры, независимость архитектуры от конкретной языковой модели, сохранность данных и возможность гибкой реконфигурации предметного содержания без изменения программной реализации. Именно эти свойства обеспечивают возможность постепенного развития системы по мере повышения уровня технологической зрелости и подключения новых языковых моделей.

Таким образом, интеллектуальный программный инструмент рассматривается не как автономная система оценивания знаний, а как управляемая программная среда, обеспечивающая взаимодействие преподавателя, обучающегося и больших языковых моделей в рамках единой воспроизводимой процедуры диалоговой диагностики [15].

3.3. Требования к данным, их жизненному циклу и разграничению доступа

Одной из принципиальных особенностей предлагаемого подхода является рассмотрение экзамена не только как процедуры оценки знаний, но и как процесса формирования информации, используемой на последующих этапах управления образовательным процессом. При этом объектом проектирования становится не только программная система, но и информационная модель диагностического процесса. Именно структура данных, правила их преобразования и условия использования в значительной степени определяют воспроизводимость результатов, возможность их последующего анализа и интеграции в цифровую образовательную среду. Обобщенная схема информационной модели представлена на рис. 3.

Верхний уровень схемы отражает последовательность формирования информационных объектов. Исходной информацией является предметный контент дисциплины, подготовленный и утвержденный преподавателем. На его основе формируется диагностическая сессия, в ходе которой накапливаются первичные данные диалога - последовательность вопросов, ответов и параметров взаимодействия. Эти данные служат основанием для формирования предварительного заключения языковой модели. Однако данный результат не является итоговым и рассматривается исключительно как промежуточная интеллектуальная интерпретация результатов диагностики. Окончательное экспертное

решение принимается преподавателем, после чего актуализируется индивидуальная модель знаний

обучающегося, используемая для сопровождения его образовательной траектории.



Рис. 3. Информационная модель диагностического процесса, жизненного цикла данных и разграничения доступа.

Таким образом, в информационной модели принципиально разделяются первичные данные, результаты интеллектуальной обработки и официально подтвержденные сведения, обладающие различным статусом и различной областью применения. Такое разделение обеспечивает прозрачность процедуры оценивания и исключает автоматическое превращение выводов LLM в официальный результат образовательной деятельности.

Второй уровень схемы определяет сквозные свойства всех информационных объектов, сохраняющиеся на протяжении их жизненного цикла. Каждый объект существует в определенной версии, проходит последовательность состояний, сопровождается историей изменений и служебными метаданными. Фиксация версии предметного контента, параметров используемой языковой модели, промптов, времени проведения диагностики, участников процесса и других атрибутов обеспечивает полную прослеживаемость происхождения результатов. Благодаря этому любое экспертное решение может быть восстановлено до исходной диагностической сессии и использованного предметного содержания. Такая прослеживаемость представляет собой необходимое условие воспроизводимости результатов исследования, объективного разрешения спорных ситуаций и последующего совершенствования интеллектуального инструмента.

Нижний уровень схемы отражает принцип разграничения доступа к данным. В предлагаемой архитектуре права доступа определяются не только

ролью пользователя, но и типом информационного объекта и его состоянием в жизненном цикле. Студент получает доступ к собственной диагностической информации и утвержденным результатам обучения. Эксперт-преподаватель работает с полным набором данных, необходимым для принятия предметного решения, включая первичные материалы диалога и предварительные выводы языковой модели. Руководители образовательной программы используют только официально утвержденные результаты, необходимые для сопровождения индивидуальных образовательных траекторий и управления образовательным процессом. Исследователи получают доступ к обезличенным данным, предназначенным для анализа эффективности диагностических процедур и совершенствования используемых моделей. Администратор обеспечивает функционирование информационной инфраструктуры, не вмешиваясь в предметное содержание образовательной деятельности.

Подобная организация доступа позволяет одновременно решить несколько задач. Во-первых, обеспечивается соблюдение требований к защите персональных данных обучающихся. Во-вторых, сохраняется независимость экспертного предметного решения от административных процедур. В-третьих, становится возможным накопление обезличенных массивов диагностических данных для последующего исследования эффективности различных языковых моделей, алгоритмов формирования вопросов и стратегий диалоговой диагностики без нарушения конфиденциальности образовательной информации.

Следовательно, информационная модель диагностического процесса представляет собой не просто структуру хранения данных, а основу функционирования всей интеллектуальной системы. Она обеспечивает непрерывную прослеживаемость

формирования результатов, разделение ответственности между участниками образовательного процесса и возможность накопления объективных данных о динамике формирования знаний обучающихся. Именно такая организация информационного пространства создает предпосылки для масштабируемого применения LLM в образовательной практике при сохранении определяющей роли эксперта-преподавателя и доверия к результатам диагностики.

3.4. Требования к разграничению доступа и защите информации

Использование LLM в образовательном процессе требует разделения не только информационных объектов, но и полномочий участников диагностического процесса. Предлагаемая архитектура основана на принципе, согласно которому искусственный интеллект является интеллектуальным инструментом обработки информации, но не субъектом принятия образовательных решений [11]. Он может формировать вопросы, интерпретировать ответы и подготавливать предварительные заключения, однако

не обладает полномочиями по изменению предметного содержания дисциплины, утверждению результатов диагностики или принятию административных решений. Обобщенная схема распределения ответственности представлена на рис. 4.

Эксперт-преподаватель формирует и утверждает предметное содержание дисциплины, определяет критерии оценивания и принимает окончательное решение по результатам диагностического диалога. Только после экспертной валидации результаты приобретают статус официальной образовательной информации и могут использоваться в информационных системах университета.

Административные подразделения работают исключительно с утвержденными результатами диагностики, не вмешиваясь в предметную интерпретацию ответов обучающихся. В свою очередь, студент получает доступ к результатам собственной диагностики и основаниям принятого решения, что обеспечивает прозрачность процедуры оценивания и возможность повторного подтверждения уровня знаний после дополнительной подготовки.



Рис. 4. Онтология доверия и ответственности в LLM-ориентированной системе диагностики знаний.

Особое внимание должно уделяться использованию внешних языковых моделей. При их применении в

облачной инфраструктуре передаваемые данные должны быть обезличены и не содержать сведений, позволяющих идентифицировать обучающегося. Для всех критически важных операций - утверждения предметного содержания, изменения критериев

оценивания, экспертного подтверждения результатов и корректировки официальных оценок - должна обеспечиваться регистрация событий в журнале аудита с сохранением информации о времени выполнения операции, исполнителе и версии используемой модели. Таким образом, разграничение полномочий, прослеживаемость критических операций и сохранение приоритета экспертного решения являются необходимыми условиями доверия к LLM-ориентированной системе диагностики знаний и ее применения в образовательном процессе

3.5. Функционально-ролевая архитектура системы

Предлагаемая функционально-ролевая архитектура объединяет участников образовательного процесса, интеллектуальные программные модули и внешние языковые модели в единую систему диагностики знаний. В отличие от существующих решений, архитектура ориентирована не на прямое использование LLM, а на построение закрытого интеллектуального контура, внутри которого обеспечиваются предметная управляемость, воспроизводимость результатов диагностики и независимость системы от конкретной языковой модели [16]. Функционально-ролевая архитектура системы представлена на рис. 5.

Центральным элементом архитектуры является закрытое ядро системы, внутри которого выполняются все критически важные операции по подготовке диагностических сессий, управлению диалогом, экспертной валидации результатов, хранению данных и регистрации событий. Пользователи не взаимодействуют непосредственно с языковой

моделью и не имеют доступа к внутренним механизмам формирования запросов. Вся работа осуществляется через стандартизованные интерфейсы и специализированные сервисы ядра системы.

Предметное сопровождение системы осуществляет эксперт-преподаватель. В его функции входят загрузка и актуализация базового учебного контента (лекции, учебные пособия, нормативные материалы и другие обязательные источники), а также расширенного контента (монографии, научные статьи, справочные материалы, методические рекомендации и иные источники, определяющие профессиональный контекст дисциплины). На основе этих материалов формируются стандартизованный банк диагностических вопросов и параметры проведения экзаменационных сессий.

Важнейшей функцией архитектуры является предварительное предметное обогащение контекста языковой модели. Перед формированием каждого запроса система автоматически извлекает релевантные фрагменты загруженного преподавателем контента и использует их при построении контекста обращения к LLM [17]. Таким образом, оценивание выполняется не на основе общей модели знаний языковой модели, а относительно предметной модели дисциплины, сформированной преподавателем. Именно преподаватель определяет содержание курса, используемую терминологию, требуемую глубину проверки и критерии оценивания, тогда как LLM выполняет функции интеллектуальной обработки информации в рамках заданного предметного контекста [18].

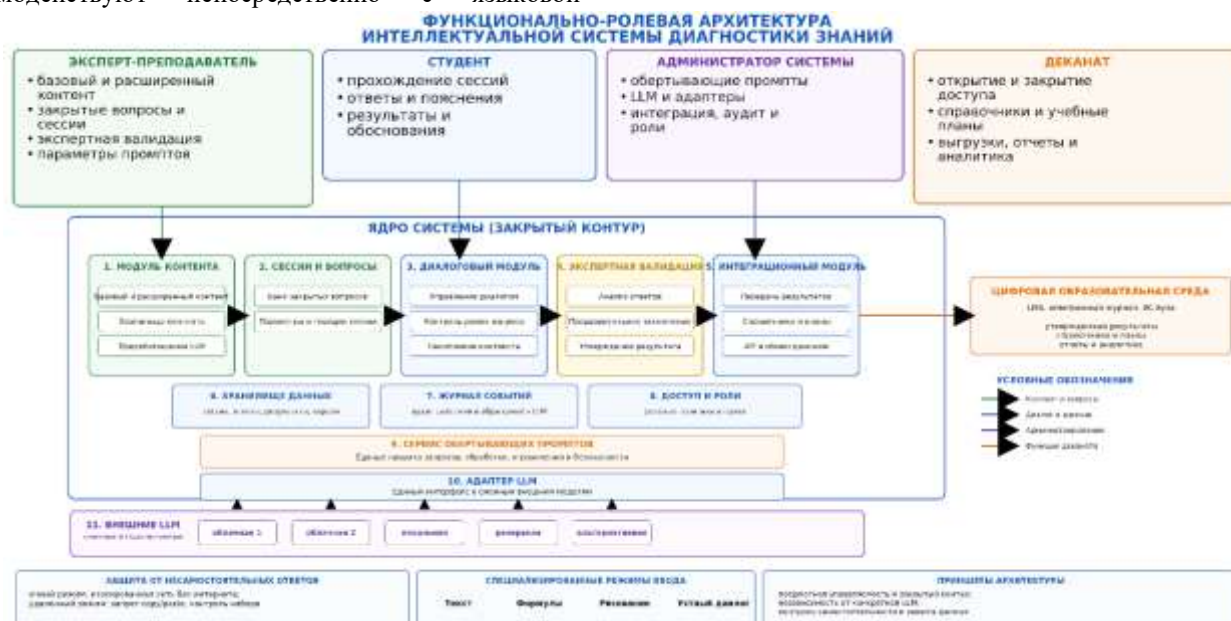


Рис. 5. Функционально-ролевая архитектура интеллектуальной системы диагностики знаний.

Студент взаимодействует исключительно с диалоговым контуром системы. Ему предъявляются стандартизованные формулировки вопросов,

сформированные на основе утвержденного банка заданий. Ответы обучающегося анализируются внутри закрытого контура системы и только после этого преобразуются в унифицированные запросы к языковой модели.

Такая организация исключает непосредственное взаимодействие студента с внешней LLM и обеспечивает единообразие процедуры диагностики независимо от используемой модели. Для обеспечения воспроизводимости и управляемости взаимодействия с внешними языковыми моделями в архитектуре выделен специализированный сервис системных (обертывающих) промптов. Он формирует единый формат запросов к LLM, включает предметный контекст, определяет ограничения генерации, требования к структуре ответа и правила интерпретации результатов [19]. Преподавателю предоставляется возможность настройки ограниченного набора параметров диагностического процесса без доступа к внутренней структуре системных промптов. Полное управление версиями и содержанием обертывающих промптов относится к функциям администратора системы.

Взаимодействие с внешними языковыми моделями осуществляется через специализированный адаптер, обеспечивающий изоляцию внутренней архитектуры от особенностей конкретных API и позволяющий использовать различные локальные и облачные LLM без изменения программной логики системы. Такая организация обеспечивает возможность сравнительной оценки различных языковых моделей, их поэтапной замены и модернизации без изменения архитектуры интеллектуальной системы.

Архитектура предусматривает средства повышения достоверности результатов диагностики и минимизации использования внешних интеллектуальных помощников обучающимися. При проведении экзамена в локальной вычислительной сети образовательной организации реализуется полностью изолированный режим функционирования. Для удаленного режима архитектура допускает подключение специализированных механизмов контроля происхождения ответов, включая ограничение операций копирования и вставки, анализ характеристик клавиатурного ввода и другие средства контроля самостоятельности выполнения работы. При этом соответствующие механизмы рассматриваются как подключаемые сервисы архитектуры и не изменяют логики функционирования интеллектуального ядра.

Предлагаемая архитектура ориентирована на развитие мультимодального взаимодействия. Помимо текстового ввода она предусматривает подключение специализированных компонентов ввода математических формул, графических схем и рукописных изображений, создаваемых непосредственно в процессе экзамена. В качестве перспективного направления развития рассматривается переход к полноценному устному диалоговому взаимодействию студента с системой при сохранении описанных принципов предметной управляемости и экспертной валидации.

Функции администратора системы ограничиваются сопровождением технологической инфраструктуры, управлением системными промптами, адаптерами

взаимодействия с LLM, ролями пользователей и параметрами функционирования системы. Деканат не участвует в диагностическом процессе и не взаимодействует с языковой моделью, а использует исключительно утвержденные результаты диагностики для управления доступом пользователей, интеграции с цифровой образовательной средой и формирования отчетности.

Таким образом, предложенная функционально-ролевая архитектура реализует разделение предметных, диагностических, административных и технологических функций, обеспечивает независимость интеллектуальной системы от конкретной реализации LLM и создает условия для последовательного развития программного комплекса при сохранении неизменных принципов проведения диалоговой диагностики знаний.

4. БАЗОВАЯ РЕАЛИЗАЦИЯ ПРОГРАММНОГО ПРОТОТИПА

4.1. Назначение и границы реализации

Для экспериментальной проверки предложенной архитектуры был реализован программный прототип интеллектуальной системы диалоговой диагностики знаний [20]. Целью разработки являлось не создание законченной промышленной информационной системы образовательной организации, а практическая проверка реализуемости предложенных архитектурных решений, механизмов взаимодействия с большой языковой моделью и организации диагностического диалога между системой и обучающимся.

Реализованный прототип обеспечивает полный цикл проведения экзаменационной сессии, включающий аутентификацию пользователей, формирование индивидуальной сессии, выдачу вопросов, ведение диалога с использованием LLM, предварительное оценивание ответов по заранее заданным критериям, экспертную корректировку результатов преподавателем и сохранение итоговой информации в базе данных. При этом языковая модель используется исключительно как интеллектуальный исполнитель отдельных функций анализа и формирования ответов, тогда как управление экзаменационным процессом полностью остается за программным приложением.

При разработке сознательно были ограничены границы реализации. В состав первой версии не включались функции интеграции с корпоративной информационной системой университета, формирования официальной отчетности, долговременного архивирования результатов и развитых средств администрирования. Эти возможности рассматриваются как направления дальнейшего развития после завершения экспериментальной верификации предложенного подхода.

4.2. Компонентная архитектура

Компонентная архитектура реализованного программного прототипа представлена на рис. 6. В отличие от целевой функционально-ролевой архитектуры, сформированной в разделе 3, данный

рисунок отражает фактический состав первой версии системы и реализованные на текущем этапе связи между ее компонентами.

Прототип построен как веб-ориентированная система с централизованным серверным приложением Examiner, в котором сосредоточена основная бизнес-логика проведения экзамена [21]. Проведенный анализ реализации показал, что система не является совокупностью независимых микросервисов. Она представляет собой модульный монолит, взаимодействующий с отдельными инфраструктурными компонентами - reverse проху, базой данных и внешним сервисом большой языковой модели. Такое решение соответствует экспериментальному назначению прототипа: оно обеспечивает целостность диагностического процесса и одновременно позволяет независимо изменять инфраструктуру развертывания и подключаемую LLM.

Пользовательское взаимодействие осуществляется через веб-браузер [22]. В текущей версии реализованы три роли: студент, преподаватель и администратор. Студент проходит экзаменационные сессии, вводит ответы и пояснения, а после утверждения результата получает доступ к странице просмотра. Преподаватель работает с предметными вопросами и экзаменационными сессиями, просматривает сформированные системой оценки и при необходимости корректирует их. Администратор управляет учетными записями, ролями, загрузкой наборов вопросов и конфигурационными параметрами системы. Поддержка этих ролей реализована

средствами серверного приложения, а анонимный доступ к функциям системы исключен.

Прием и маршрутизация пользовательских запросов выполняются компонентом Caddy, который действует как reverse проху между браузерами пользователей и приложением Examiner. Он обеспечивает терминацию защищенных HTTPS-соединений и перенаправление запросов во внутренний контур приложения. Такое разделение позволило вынести транспортную защиту и сетевую маршрутизацию за пределы бизнес-логики экзамена. При развертывании на испытательном стенде использовался встроенный механизм выпуска локальных TLS-сертификатов Caddy, обеспечивающий шифрование данных внутри закрытой сети образовательной организации.

Центральным компонентом прототипа является приложение Examiner. Его внутренняя структура на рисунке представлена пятью укрупненными функциональными модулями: аутентификации и управления ролями, управления экзаменами, ведения диагностического диалога, формирования запросов к LLM и экспертной валидации. Такое выделение не означает их физического развертывания в отдельных сервисах: все перечисленные функции реализованы в составе единого серверного приложения и используют общую модель данных.

Модуль аутентификации и ролей обеспечивает вход пользователей, проверку их полномочий и управление пользовательскими сессиями.



Рис. 6. Компонентная архитектура реализованного программного прототипа.

Модуль управления экзаменами формирует состав вопросов, сохраняет параметры диагностической сессии и контролирует изменение ее состояния.

Диалоговый модуль принимает ответы студента, хранит историю взаимодействия по каждому вопросу и управляет выдачей уточняющих вопросов. Модуль формирования запросов собирает предметный и диалоговый контекст, необходимый для обращения к внешней языковой модели.

Модуль экспертной валидации предоставляет преподавателю доступ к результатам автоматизированного анализа, позволяет корректировать баллы и фиксировать итоговое предметное решение.

Предметный контент поступает в систему в виде структурированного набора вопросов. В реализованной версии используется формат JSON, содержащий формулировку вопроса, тему, уровень сложности, рубрику оценивания, эталонный ответ и максимальный балл. Загруженный набор импортируется приложением Examiner и используется при формировании экзаменационных сессий. Таким образом, внешний файл является источником верифицированного содержания, но не участвует непосредственно во взаимодействии с LLM. Подготовка и экспертная валидация этого контента рассматриваются отдельно в разделе 5.

Для хранения эксплуатационных данных используется база данных SQLite [23]. В ней сохраняются учетные записи и роли пользователей, загруженные вопросы, параметры экзаменационных сессий, сообщения диагностического диалога, предварительные и утвержденные оценки, а также текущие статусы прохождения экзамена. Размещение этих данных в серверной части обеспечивает восстановление незавершенных сессий и позволяет отделить состояние экзамена от состояния внешней языковой модели.

Связь приложения Examiner с LLM реализована через OpenAI-совместимый программный интерфейс. В материалах реализации зафиксировано подключение локальных моделей Llama 3.1 [24] и DeepSeek-R1:8B [25], а также настройка внешней облачной модели для проведения испытаний. Благодаря этому в качестве интеллектуального сервиса могут применяться локально развернутые модели через Ollama или vLLM, а также внешние облачные провайдеры с совместимым API. В материалах реализации зафиксировано подключение локальных моделей llama3.1 и deepseek-r1:8b, а также настройка внешней облачной модели для проведения испытаний. Замена модели осуществляется через конфигурационные параметры и не требует изменения логики экзаменационного приложения.

Принципиальной особенностью реализации является то, что внешняя LLM не хранит состояние экзаменационной сессии. Для каждого обращения приложение заново формирует запрос, включающий текущий вопрос, рубрику, эталонный ответ, ответ студента и релевантную историю диалога. Полученный результат возвращается в Examiner, проверяется, сохраняется в базе данных и используется для продолжения сценария. Поэтому целостность экзамена, состояние отдельных вопросов и последовательность диагностического взаимодействия определяются серверным приложением, а не используемой моделью.



Рис. 7. Сценарий диагностической сессии в программном прототипе.

Таким образом, реализованный прототип представляет собой модульный монолит с внешними инфраструктурными компонентами. Такая архитектура

обеспечивает минимально необходимое разделение ответственности: Caddy отвечает за защищенный сетевой доступ, Examiner - за бизнес-логику диагностического процесса, SQLite - за сохранение состояния, структурированный набор вопросов - за предметное содержание, а внешняя LLM - за выполнение делегированных операций семантического анализа и генерации уточняющих вопросов. Это позволило реализовать работоспособный экспериментальный контур без преждевременного усложнения системы и сохранить возможность последующего перехода к более развитой целевой архитектуре [16].

4.3. Сценарий диагностической сессии

Сценарий диагностической сессии реализован как управляемый приложением Examiner последовательный процесс, включающий формирование индивидуальной сессии, предъявление вопросов, диалоговое уточнение ответа, предварительное оценивание и экспертное утверждение результата. Общая организация процесса представлена на рис. 7.

Для каждого студента создается отдельная экзаменационная сессия, а взаимодействие по каждому вопросу ведется как самостоятельный диалоговый поток. Состояние сессии, история сообщений и результаты промежуточной обработки сохраняются в базе данных; внешняя LLM не управляет

последовательностью экзамена и не хранит его состояние.

При обработке ответа Examiner формирует запрос, включающий вопрос, рубрику, эталонный ответ и накопленный контекст диалога. В зависимости от результата анализа LLM формирует уточняющий вопрос либо предварительную оценку и комментарий. Число последовательных уточнений ограничивается параметрами экзаменационной сессии.

Полученная оценка имеет предварительный характер. После завершения сессии преподаватель просматривает диалог, при необходимости корректирует баллы и утверждает итоговый результат. Таким образом, LLM обеспечивает семантический анализ и поддержку диалога, тогда как управление процессом и окончательное предметное решение остаются в контуре программной системы и преподавателя-эксперта.

4.4. Ролевая модель и экспертная валидация

Распределение функций между участниками диагностического процесса представлено на рис. 8. В реализованном прототипе используются три основные роли: студент, преподаватель и администратор. При этом ответственность за проведение экзаменационной сессии, управление ее состоянием и взаимодействие с большой языковой моделью сосредоточена в серверном приложении Examiner.



Рис. 8. Ролевая модель и экспертная валидация реализованного программного прототипа.

Студент участвует только в прохождении экзаменационной сессии, формируя ответы и пояснения

по поставленным вопросам. Преподаватель выполняет экспертную проверку результатов, сформированных системой, при необходимости корректирует предварительную оценку и утверждает итоговое решение. Администратор обеспечивает сопровождение программного комплекса, управление пользователями и настройку его функционирования.

Принципиальной особенностью реализованного подхода является разделение ответственности между интеллектуальной моделью и человеком-экспертом. LLM используется для анализа ответа и формирования предварительного заключения, однако не принимает окончательного решения о результатах диагностики. Управление экзаменационной сессией полностью осуществляется программной системой, а итоговая оценка утверждается преподавателем после экспертной валидации результатов [14]. Такой подход обеспечивает сохранение экспертного контроля и соответствует требованиям прослеживаемости и воспроизводимости процесса оценки знаний

4.5. Механизмы надежности и безопасности

При разработке программного прототипа особое внимание было уделено обеспечению надежности функционирования и защите диагностического процесса.

Аутентификация пользователей реализована на основе ролевой модели доступа с хранением паролей в виде криптографических хешей. Исключена возможность самостоятельной регистрации пользователей, а обращения к ресурсам контролируются механизмами разграничения доступа. Для защиты веб-приложения используются HTTPS-соединение, защищенные сессионные cookie, контроль принадлежности пользовательских сессий и защита от атак класса CSRF.

Дополнительное внимание уделено обеспечению устойчивости взаимодействия с большой языковой моделью. Все обращения к LLM формируются серверным приложением по стандартизированному шаблону, включающему критерии оценивания, эталонный ответ и контекст текущего диалога [19]. Пользовательский ввод отделяется от служебной части промпта специальными разделителями, что снижает

вероятность успешной реализации атак класса prompt injection. Кроме того, сервер осуществляет контроль допустимости возвращаемых моделью оценок и ограничивает объем обрабатываемых текстовых ответов, предотвращая некорректное поведение системы.

В совокупности реализованные решения обеспечили функционирование программного прототипа как управляемой интеллектуальной системы, в которой использование большой языковой модели встроено в контролируемый программный контур и не нарушает требований к воспроизводимости, прослеживаемости и экспертной управляемости диагностического процесса.

5. ПОДГОТОВКА И ЗАГРУЗКА ПРЕДМЕТНОГО КОНТЕНТА

Одной из принципиальных особенностей предложенного подхода является отделение процесса формирования предметного содержания от процесса проведения экзаменационной сессии. В реализованном прототипе большая языковая модель не определяет самостоятельно содержание дисциплины и не формирует экзаменационные вопросы непосредственно во время диагностики. Источником предметной модели выступает преподаватель-эксперт, который определяет состав используемых материалов, структуру дисциплины, глубину изложения и критерии оценивания. LLM используется как интеллектуальный инструмент автоматизированной обработки подготовленного корпуса знаний, обеспечивающий существенное сокращение трудоемкости формирования экзаменационного контента. Общая схема подготовки предметного содержания представлена на рис. 9

Предлагаемая технология принципиально отличается от использования языковой модели в режиме свободного вопросно-ответного взаимодействия. На вход интеллектуального конвейера поступает не отдельный запрос пользователя, а целостный предметный корпус, сформированный преподавателем. На выходе формируется структурированный и экспертно верифицированный набор диагностических материалов, который после загрузки используется системой без изменения его содержания в ходе проведения экзамена.



Рис. 9. Итерационный конвейер подготовки предметного контента для интеллектуальной системы диагностики знаний.

5.1. Исходные учебные материалы

Исходным этапом является формирование предметного корпуса. В экспериментальной реализации использовались разработанные преподавателем планы дисциплины, лекционные материалы и презентации, дополненные рекомендованной учебной литературой и другими проверенными источниками, определенными преподавателем в соответствии с содержанием курса. Для апробации системы был подготовлен корпус по дисциплине «Теория автоматического управления», включающий материалы трех лекций, охватывающих основные разделы курса.

Таким образом, источником предметного содержания является не языковая модель, а сформированный преподавателем корпус материалов, отражающий авторскую концепцию дисциплины.

5.2. Первичное формирование вопросов и ответов

Первичная интеллектуальная обработка корпуса выполняется с использованием специально подготовленных промптов. На данном этапе LLM выделяет смысловые блоки учебного материала, формирует черновые варианты экзаменационных вопросов, краткие эталонные ответы и предварительные критерии оценивания. При этом промпт ограничивает обработку исключительно представленным корпусом материалов и требует полного тематического покрытия без привлечения внешних знаний.

Полученный результат рассматривается как промежуточная версия предметного содержания, предназначенная для последующей экспертной обработки.

5.3. Методическая и предметная валидация

Ключевым элементом технологии является итерационный контур взаимодействия преподавателя и языковой модели. После первичной генерации преподаватель анализирует полученные материалы, оценивает полноту охвата дисциплины, корректность терминологии, глубину раскрытия вопросов, инженерный контекст и соответствие собственному методическому замыслу. Замечания используются как исходные данные для последующих итераций автоматизированной обработки.

В отличие от традиционного ручного формирования экзаменационных материалов преподаватель не создает каждую формулировку заново, а управляет процессом интеллектуальной обработки, последовательно уточняя требования к содержанию. Такой режим обеспечивает одновременное повышение качества итогового материала и существенное сокращение трудоемкости его подготовки. Отчетные материалы показывают, что в ходе доработки основная часть изменений была связана не с исправлением фактических ошибок, а с совершенствованием структуры ответов, терминологии, логики изложения и уровня детализации.

5.4. Формирование рубрик и JSON-набора

После завершения экспертной валидации каждому вопросу назначаются тема, уровень сложности, максимальный балл, эталонный ответ и рубрика оценивания. Эти данные формируют формализованное описание вопроса, используемое системой при проведении экзаменационной сессии.

В реализованном прототипе каждый элемент экзаменационного набора содержит формулировку вопроса, тематическую принадлежность, уровень сложности, эталонный ответ, рубрику оценивания и максимальный балл. На основании этих данных автоматически формируется JSON-набор, являющийся

стандартной формой представления предметного контента для последующей загрузки в систему.

5.5. Импорт в систему

Заключительным этапом является импорт подготовленного набора в программный комплекс Examiner. После загрузки вопросы становятся частью внутреннего хранилища системы и используются при формировании экзаменационных сессий. В процессе диагностики система обращается только к предварительно подготовленным и утвержденным преподавателем материалам; генерация новых вопросов или изменение критериев оценивания непосредственно во время экзамена не выполняются.

Таким образом, предложенная технология разделяет функции преподавателя и языковой модели. Преподаватель определяет содержание дисциплины и управляет последовательностью экспертных итераций, тогда как LLM выполняет наиболее трудоемкие операции анализа, структурирования и первичного формирования материалов. Это обеспечивает одновременное снижение трудоемкости подготовки экзаменационного контента, повышение его качества и полноты, а также возможность формирования специализированных диагностических наборов, отражающих особенности конкретной образовательной программы.

6. РАЗВЕРТЫВАНИЕ НА ИСПЫТАТЕЛЬНОМ СТЕНДЕ И СИСТЕМНЫЕ ИСПЫТАНИЯ

Для проверки работоспособности разработанного программного прототипа был развернут испытательный стенд, включающий сервер приложений, сервер баз данных, систему хранения документов, средства контейнеризации, а также программные компоненты интеллектуального конвейера подготовки фонда оценочных средств. В составе стенда были развернуты веб-интерфейс преподавателя, сервер приложений, база данных проекта, файловое хранилище исходных материалов, сервисы обработки документов и подготовки контекстов, модуль взаимодействия с большими языковыми моделями, модуль формирования фонда оценочных средств и модуль экспорта результатов. Проведена проверка совместной работы всех компонентов в составе единой программной системы.

Все программные компоненты были развернуты в виде отдельных Docker-контейнеров с использованием внутренней виртуальной сети. В ходе развертывания выполнены настройка взаимодействия сервисов, разграничение внутренних и внешних интерфейсов, организация защищенного обмена данными между программными модулями, настройка постоянного хранения данных и конфигурирование параметров подключения к внешним сервисам. Проверена устойчивость взаимодействия компонентов при последовательном запуске, перезапуске и обновлении отдельных сервисов.

В составе испытательного стенда реализована возможность использования как локально развернутых, так и облачных больших языковых моделей [26].

Проверены подключение локальных моделей через Ollama [27], взаимодействие с удаленными моделями по API, переключение между различными LLM без изменения прикладной логики системы, сохранение унифицированного формата обмена и повторное выполнение этапов генерации после смены модели. Использование унифицированного программного интерфейса подтвердило независимость прикладных модулей от конкретной реализации языковой модели.

Программа функциональных и интеграционных испытаний предусматривала последовательную проверку полного технологического цикла подготовки фонда оценочных средств. Выполнена проверка загрузки учебных материалов, формирования внутреннего корпуса документов, подготовки контекстов для LLM, генерации экзаменационных вопросов, эталонных ответов и рубрик оценивания, сохранения промежуточных результатов, повторной генерации после внесения замечаний преподавателя, формирования итогового фонда оценочных средств и его экспорта в формат, пригодный для использования системой интеллектуальной диагностики знаний. Дополнительно проверена корректность совместной работы всех программных модулей в составе единого технологического конвейера.

По результатам системных испытаний подтверждена работоспособность всех основных компонентов программного прототипа. В ходе испытаний подтверждены корректное взаимодействие программных сервисов, устойчивое выполнение полного технологического конвейера, возможность многократного итерационного уточнения результатов, совместимость с локальными и облачными LLM, сохранение промежуточных результатов обработки, корректное формирование итогового фонда оценочных средств и экспорт подготовленных материалов для последующего использования в интеллектуальной экзаменационной системе. Полученные результаты подтвердили техническую готовность разработанного программного прототипа к проведению пользовательской апробации.

7. ПОЛЬЗОВАТЕЛЬСКАЯ АПРОБАЦИЯ В ФОРМЕ ТЕСТОВОГО ЭКЗАМЕНА

После завершения системных испытаний проведена пользовательская апробация разработанного программного прототипа в форме тестового экзамена. Целью апробации являлась проверка возможности практического использования автоматически сформированного системой фонда оценочных средств при проведении интеллектуальной диагностики знаний, а также оценка корректности взаимодействия преподавателя, обучающегося и программной системы в условиях, максимально приближенных к реальному образовательному процессу.

В апробации приняли участие преподаватели и обучающиеся, ранее не работавшие с разработанным программным прототипом. В ходе испытаний последовательно выполнены загрузка сформированного фонда оценочных средств, проведение экзаменационной сессии, анализ

автоматически сформированных вопросов, эталонных ответов и критериев оценивания, экспертная оценка качества полученных материалов, фиксация замечаний и повторная генерация отдельных элементов после внесения экспертных корректировок. Дополнительно проверена возможность многократного повторения полного цикла подготовки фонда оценочных средств без изменения общей технологии работы системы.

Проведение тестового экзамена подтвердило возможность практического использования автоматически сформированного фонда оценочных средств при организации интеллектуальной диагностики знаний. Все этапы подготовки и проведения экзамена были выполнены штатными средствами разработанного программного прототипа, включая загрузку предметного контента, генерацию материалов, экспертную корректировку, формирование итоговой версии фонда оценочных средств и проведение диагностической сессии. Подтверждена возможность последовательного применения интеллектуального конвейера на всех этапах подготовки экзамена.

Экспертная оценка результатов показала, что автоматически сформированные вопросы, эталонные ответы и рубрики оценивания в целом соответствуют структуре учебной дисциплины, методическому замыслу преподавателя и требованиям к уровню подготовки обучающихся. Отмечено существенное сокращение объема рутинной работы по формированию фонда оценочных средств при сохранении ведущей роли преподавателя в определении содержания дисциплины, экспертной корректировке результатов и утверждении итоговой версии материалов.

В ходе апробации выявлено, что качество автоматически формируемых материалов в значительной степени определяется полнотой исходного предметного контента и эффективностью его контекстного обогащения перед передачей в LLM [18]. Перспективными направлениями развития являются расширение механизмов автоматического контекстного обогащения, совершенствование средств экспертной корректировки, поддержка специализированных предметных областей, развитие коллективной подготовки фондов оценочных средств и интеграция программного комплекса с цифровой образовательной средой университета. В целом проведенная пользовательская апробация подтвердила содержательную работоспособность предложенного подхода и возможность его практического использования при подготовке и проведении интеллектуальных экзаменов.

8. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Целью работы являлась разработка программного прототипа интеллектуальной системы, обеспечивающей автоматизированную подготовку фонда оценочных средств и проведение интеллектуальной диагностики знаний на основе больших языковых моделей при сохранении определяющей роли преподавателя в постановке

задачи, формировании предметного контента и принятии итоговых решений. В отличие от распространенного использования LLM как универсального источника ответов, предложенный подход ориентирован на использование языковой модели в качестве элемента интеллектуального технологического конвейера, функционирующего в рамках заранее определенной архитектуры, фиксированных процедур и экспертных ограничений [28].

Для достижения поставленной цели разработана функционально-ролевая архитектура системы, определены требования к данным, их жизненному циклу, разграничению доступа и взаимодействию пользователей с внешней LLM. Реализован программный прототип, включающий средства подготовки предметного контента, итерационной генерации фонда оценочных средств, экспертной корректировки результатов, интеграции с локальными и облачными языковыми моделями, а также проведения интеллектуального экзамена. Проведены системные испытания и пользовательская апробация, подтвердившие работоспособность реализованного технологического конвейера и возможность его использования в условиях реального образовательного процесса.

Полученные результаты позволяют сделать вывод о практической реализуемости предложенного подхода. Использование интеллектуального конвейера обеспечивает существенное сокращение трудоемкости подготовки фонда оценочных средств, повышение однородности и воспроизводимости результатов, а также возможность многократной итерационной доработки материалов без нарушения общей технологии подготовки экзамена. При этом определяющая роль преподавателя сохраняется на всех этапах жизненного цикла фонда оценочных средств - от формирования предметного контента до экспертного утверждения итоговой версии экзаменационных материалов.

Проведенная работа рассматривается как первый этап создания интеллектуальной системы диагностики знаний. Дальнейшее развитие связано с расширением средств контекстного обогащения предметных материалов, поддержкой специализированных режимов ввода ответов (формулы, графические схемы, устный диалог), развитием механизмов противодействия использованию внешних интеллектуальных помощников, интеграцией с цифровой образовательной средой университета и проведением полномасштабной экспериментальной проверки на различных учебных дисциплинах. Полученные результаты подтверждают перспективность предложенной архитектуры как основы для построения интеллектуальных образовательных систем нового поколения.

ЗАКЛЮЧЕНИЕ

Представленная работа демонстрирует возможность перехода от использования больших языковых моделей как универсальных генераторов текстов к построению специализированных интеллектуальных систем, в которых LLM выступает лишь одним из компонентов управляемого технологического конвейера [28]. Принципиальной особенностью предложенного подхода является сохранение определяющей роли преподавателя в формировании предметного содержания, методологии оценки и принятии итоговых решений при одновременной передаче интеллектуальной системе значительной части рутинной когнитивной работы [11]. Тем самым закладываются предпосылки для создания нового поколения образовательных систем, сочетающих возможности современных языковых моделей с требованиями воспроизводимости, управляемости, прозрачности и экспертной ответственности.

Разработанный и апробированный программный прототип рассматривается как первый этап создания полнофункциональной интеллектуальной системы диагностики знаний. Следующим этапом работы является разработка первой промышленно ориентированной версии программного продукта, ее реализация и развертывание в образовательной инфраструктуре университета в течение осеннего семестра 2026 года с последующим переходом к опытной эксплуатации на реальных учебных дисциплинах. Полученные результаты создают необходимую научно-техническую основу для этого этапа и подтверждают практическую реализуемость выбранного направления развития.

БИБЛИОГРАФИЯ

- [1] Bloom B. S., Engelhart M. D., Furst E. J., Hill W. H., Krathwohl D. R. Taxonomy of Educational Objectives: Handbook I. Cognitive Domain. New York: David McKay Company, 1956.
- [2] Biggs J., Tang C. Teaching for Quality Learning at University. 4th ed. Maidenhead, UK: Open University Press, 2011.
- [3] Black P., William D. Assessment and Classroom Learning // Assessment in Education: Principles, Policy & Practice. 1998. Vol. 5, no. 1. P. 7–74. DOI: 10.1080/0969595980050102.
- [4] Nicol D. J., Macfarlane-Dick D. Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice // Studies in Higher Education. 2006. Vol. 31, no. 2. P. 199–218. DOI: 10.1080/03075070600572090.
- [5] Pellegrino J. W., Chudowsky N., Glaser R. (Eds.). Knowing What Students Know: The Science and Design of Educational Assessment. Washington, DC: National Academies Press, 2001. DOI: 10.17226/10019.
- [6] Bennett R. E. Inexorable and Inevitable: The Continuing Story of Technology and Assessment // Journal of Technology, Learning, and Assessment. 2002. Vol. 1, no. 1.
- [7] Vygotsky L. S. Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press, 1978.
- [8] Bruner J. S. The Process of Education. Cambridge, MA: Harvard University Press, 1960.
- [9] Graesser A. C., Chipman P., Haynes B. C., Olney A. AutoTutor: An Intelligent Tutoring System with Mixed-Initiative Dialogue // IEEE Transactions on Education. 2005. Vol. 48, no. 4. P. 612–618. DOI: 10.1109/TE.2005.856149.
- [10] Zhao W. X., Zhou K., Li J., et al. A Survey of Large Language Models. arXiv:2303.18223, 2023. DOI: 10.48550/arXiv.2303.18223
- [11] Shneiderman B. Human-Centered AI. New York: Oxford University Press, 2022. DOI: 10.1093/oso/9780192845290.001.0001.
- [12] Russell S. Human Compatible: Artificial Intelligence and the Problem of Control. New York: Viking, 2019. ISBN 978-0-525-55861-3.
- [13] Wu Q., Bansal G., Zhang J., et al. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. arXiv:2308.08155, 2023. DOI: 10.48550/arXiv.2308.08155
- [14] Amershi S., Weld D. S., Vorvoreanu M., Fournay A., Nushi B., Collisson P., Suh J., Iqbal S. T., Bennett P. N., Inkpen K., Teevan J., Kikin-Gil R., Horvitz E. Guidelines for Human-AI Interaction // Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). New York: ACM, 2019. Article 3. DOI: 10.1145/3290605.3300233.
- [15] Park J. S., O'Brien J. C., Cai C. J., Morris M. R., Liang P., Bernstein M. S. Generative Agents: Interactive Simulacra of Human Behavior // Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23). New York: Association for Computing Machinery, 2023. Article No. 2. P. 1–22. DOI: 10.1145/3586183.3606763.
- [16] Anthropic. Building Effective AI Agents. Dec. 19, 2024. [Online]. Available: <https://www.anthropic.com/engineering/building-effective-agents>.
- [17] Lewis P., Perez E., Piktus A., et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks // Advances in Neural Information Processing Systems (NeurIPS 2020). Vol. 33. 2020. P. 9459–9474. [Online]. Available: <https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html>
- [18] Gao Y., Xiong Y., Gao X., Jia K., Pan J., Bi Y., Dai Y., Sun J., Guo Q., Wang M., Wang H. Retrieval-Augmented Generation for Large Language Models: A Survey. CoRR, abs/2312.10997, 2023. DOI: 10.48550/arXiv.2312.10997.
- [19] White J., Fu Q., Hays S., Sandborn M., Olea C., Gilbert H., Elnashar A., Spencer-Smith J., Schmidt D. C. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. arXiv:2302.11382, 2023. DOI: 10.48550/arXiv.2302.11382.
- [20] Python Software Foundation. Python 3 Documentation. [Online]. Available: <https://docs.python.org/3/>
- [21] Pallets Project. Flask Documentation. [Online]. Available: <https://flask.palletsprojects.com/>
- [22] Mozilla. MDN Web Docs: HTML, CSS, and JavaScript. [Online]. Available: <https://developer.mozilla.org/>
- [23] SQLite Consortium. SQLite Documentation. [Online]. Available: <https://sqlite.org/docs.html>
- [24] Meta AI. The Llama 3 Herd of Models // arXiv:2407.21783, 2024. [Online]. Available: <https://arxiv.org/abs/2407.21783>.
- [25] DeepSeek-AI. DeepSeek-R1: Incentivizing Reasoning Capability in Large Language Models // arXiv:2501.12948, 2025. [Online]. Available: <https://arxiv.org/abs/2501.12948>
- [26] Radford A., Kim J. W., Xu T., Brockman G., McLeavey C., Sutskever I. Robust Speech Recognition via Large-Scale Weak Supervision. arXiv:2212.04356, 2022. DOI: 10.48550/arXiv.2212.04356
- [27] Ollama. Ollama Documentation. [Online]. Available: <https://ollama.com/docs>
- [28] OpenAI. A Practical Guide to Building AI Agents. 2025. [Online]. Available: <https://openai.com/business/guides-and-resources/a-practical-guide-to-building-ai-agents>

Статья получена 20 июля 2026

Анни П.В. – специалист компании Red Hat (e-mail: pavel.anni@gmail.com)

Жабицкий М.Г. - заместитель директора ВИШ НИЯУ МИФИ (e-mail: jabitsky@mail.ru);

Царев В.С. – руководитель департамента АО ВНИИАЭС (e-mail: tsarevvs@bk.ru);

Кальсина Ю.В.- студент ВИШ НИЯУ МИФИ; (e-mail: yuliya-kalsina@mail.ru)

Парфенов И.А.- студент ВИШ НИЯУ МИФИ; (e-mail: parfenov314ivan@gmail.com)

Соколинская М.О.- студент ВИШ НИЯУ МИФИ; (e-mail: maryodyvan@gmail.com)

Забродин М.Д. - студент ВИШ НИЯУ МИФИ; (e-mail: maks.zabrodin.04@gmail.com)

A Method for Dialog-Based Knowledge Assessment Using a Two-Loop Human–LLM Interaction

P.V. Anni, M.G. Zhabitskii, V.S. Tsarev, Yu.V. Kalsina, I. A. Parfenov, M.O. Sokolinskaya, M.D. Zabrodin

Abstract — This paper proposes a method for dialog-based knowledge assessment based on a two-loop interaction with a large language model (LLM), enabling the separation of subject content preparation from individual knowledge assessment. The method is organized around two interconnected iterative loops: the expert–LLM loop, intended for the generation and expert validation of examination content, and the student–LLM loop, which implements an adaptive diagnostic dialogue through the progressive refinement of the learner's knowledge model. The paper defines principles for distributing functions among the instructor, the large language model, and the information system, whereby the LLM performs cognitively intensive tasks related to content generation and response analysis, while the final subject-specific judgment remains the responsibility of the instructor. Based on the proposed method, the architecture of an intelligent assessment system was developed, a software prototype was implemented, and its experimental validation was carried out in an educational environment. The experimental results confirm the practical feasibility of the proposed approach and demonstrate its potential to substantially reduce the effort required for developing assessment item banks and conducting individualized dialog-based knowledge assessment while preserving reproducibility, traceability, and expert reliability of the assessment results.

Keywords - dialog-based knowledge assessment, large language models, LLM, intelligent educational systems, adaptive dialogue, two-loop interaction, expert validation, assessment item bank, knowledge assessment, digital learning environment.

REFERENCES

- [1] Bloom B. S., Engelhart M. D., Furst E. J., Hill W. H., Krathwohl D. R. Taxonomy of Educational Objectives: Handbook I. Cognitive Domain. New York: David McKay Company, 1956.
- [2] Biggs J., Tang C. Teaching for Quality Learning at University. 4th ed. Maidenhead, UK: Open University Press, 2011.
- [3] Black P., Wiliam D. Assessment and Classroom Learning // Assessment in Education: Principles, Policy & Practice. 1998. Vol. 5, no. 1. P. 7–74. DOI: 10.1080/0969595980050102.
- [4] Nicol D. J., Macfarlane-Dick D. Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice // Studies in Higher Education. 2006. Vol. 31, no. 2. P. 199–218. DOI: 10.1080/03075070600572090.
- [5] Pellegrino J. W., Chudowsky N., Glaser R. (Eds.). Knowing What Students Know: The Science and Design of Educational Assessment. Washington, DC: National Academies Press, 2001. DOI: 10.17226/10019.
- [6] Bennett R. E. Inexorable and Inevitable: The Continuing Story of Technology and Assessment // Journal of Technology, Learning, and Assessment. 2002. Vol. 1, no. 1.
- [7] Vygotsky L. S. Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press, 1978.
- [8] Bruner J. S. The Process of Education. Cambridge, MA: Harvard University Press, 1960.
- [9] Graesser A. C., Chipman P., Haynes B. C., Olney A. AutoTutor: An Intelligent Tutoring System with Mixed-Initiative Dialogue // IEEE Transactions on Education. 2005. Vol. 48, no. 4. P. 612–618. DOI: 10.1109/TE.2005.856149.
- [10] Zhao W. X., Zhou K., Li J., et al. A Survey of Large Language Models. arXiv:2303.18223, 2023. DOI: 10.48550/arXiv.2303.18223
- [11] Shneiderman B. Human-Centered AI. New York: Oxford University Press, 2022. DOI: 10.1093/oso/9780192845290.001.0001.
- [12] Russell S. Human Compatible: Artificial Intelligence and the Problem of Control. New York: Viking, 2019. ISBN 978-0-525-55861-3.
- [13] Wu Q., Bansal G., Zhang J., et al. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. arXiv:2308.08155, 2023. DOI: 10.48550/arXiv.2308.08155
- [14] Amershi S., Weld D. S., Vorvoreanu M., Fournay A., Nushi B., Collisson P., Suh J., Iqbal S. T., Bennett P. N., Inkpen K., Teevan J., Kikin-Gil R., Horvitz E. Guidelines for Human-AI Interaction // Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). New York: ACM, 2019. Article 3. DOI: 10.1145/3290605.3300233.
- [15] Park J. S., O'Brien J. C., Cai C. J., Morris M. R., Liang P., Bernstein M. S. Generative Agents: Interactive Simulacra of Human Behavior // Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23). New York: Association for Computing Machinery, 2023. Article No. 2. P. 1–22. DOI: 10.1145/3586183.3606763.
- [16] Anthropic. Building Effective AI Agents. Dec. 19, 2024. [Online]. Available: <https://www.anthropic.com/engineering/building-effective-agents>.
- [17] Lewis P., Perez E., Piktus A., et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks // Advances in Neural Information Processing Systems (NeurIPS 2020). Vol. 33. 2020. P. 9459–9474. [Online]. Available: <https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html>
- [18] Gao Y., Xiong Y., Gao X., Jia K., Pan J., Bi Y., Dai Y., Sun J., Guo Q., Wang M., Wang H. Retrieval-Augmented Generation for Large Language Models: A Survey. CoRR, abs/2312.10997, 2023. DOI: 10.48550/arXiv.2312.10997.
- [19] White J., Fu Q., Hays S., Sandborn M., Olea C., Gilbert H., Elnashar A., Spencer-Smith J., Schmidt D. C. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. arXiv:2302.11382, 2023. DOI: 10.48550/arXiv.2302.11382.
- [20] Python Software Foundation. Python 3 Documentation. [Online]. Available: <https://docs.python.org/3/>
- [21] Pallets Project. Flask Documentation. [Online]. Available: <https://flask.palletsprojects.com/>
- [22] Mozilla. MDN Web Docs: HTML, CSS, and JavaScript. [Online]. Available: <https://developer.mozilla.org/>
- [23] SQLite Consortium. SQLite Documentation. [Online]. Available: <https://sqlite.org/docs.html>
- [24] Meta AI. The Llama 3 Herd of Models // arXiv:2407.21783, 2024. [Online]. Available: <https://arxiv.org/abs/2407.21783>.
- [25] DeepSeek-AI. DeepSeek-R1: Incentivizing Reasoning Capability in Large Language Models // arXiv:2501.12948, 2025. [Online]. Available: <https://arxiv.org/abs/2501.12948>
- [26] Radford A., Kim J. W., Xu T., Brockman G., McLeavey C., Sutskever I. Robust Speech Recognition via Large-Scale Weak Supervision. arXiv:2212.04356, 2022. DOI: 10.48550/arXiv.2212.04356
- [27] Ollama. Ollama Documentation. [Online]. Available: <https://ollama.com/docs>
- [28] OpenAI. A Practical Guide to Building AI Agents. 2025. [Online]. Available: <https://openai.com/business/guides-and-resources/a-practical-guide-to-building-ai-agents>