

УДК 311.42: 519.252

DOI 10.5281/zenodo.17849657

**КОЛОМЫЦЕВА Анна Олеговна<sup>1</sup>,  
ГЕЙЧЕНКО Евгений Игоревич<sup>2</sup>,  
КОЛОМЫЦЕВА Ирина Константиновна<sup>1</sup>,  
КОЖЕВНИКОВ Арсений Андреевич<sup>1</sup>**

<sup>1</sup> ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина», ул. Мира, 19, Екатеринбург, Россия, 620062

<sup>2</sup> СПбГЭТУ «ЛЭТИ» «Санкт-Петербургский государственный электротехнический университет имени В.И. Ульянова Ленина», Санкт-Петербург, Россия

## **ПРОЕКТИРОВАНИЕ МОДЕЛИ АДАПТИВНОЙ АРХИТЕКТУРЫ AI-АГЕНТА С ЕСТЕСТВЕННО-ЯЗЫКОВЫМ ИНТЕРФЕЙСОМ ДЛЯ ПЛАТФОРМ ТЕХНИЧЕСКОЙ ПОДДЕРЖКИ**

В статье представлен результат проектирования модели адаптивной архитектуры интеллектуального агента с естественным языковым интерфейсом, предназначенного для функционирования на платформе, обеспечивающей мультилинейную техническую поддержку пользователей. Проведен анализ научных публикаций исследователей, касающихся разработки интеллектуальных систем и AI-агентов, для создания высокопроизводительного инструмента автоматизированной поддержки пользователей, который снижает нагрузку на сотрудников команды технической поддержки и команды разработчиков, что обеспечивает повышение качества поддержки клиентов и сокращает время реагирования системы на их запросы. Архитектура AI-агента предназначена для предиктивного и контекстно-зависимого анализа обращений первого уровня техподдержки. Благодаря способности к самообучению и автоматической генерации решений, система способна обеспечить качественно новый уровень обслуживания клиентов и минимизировать задержку ответа. Статья описывает структуру основных модулей, включающих обоснованную выборку моделей до обучения и ключевые компоненты разработки. Важнейшие этапы разработки AI-агента включают сбор и предварительную обработку данных, обучение модели, создание основного программного модуля, проведение тестирования и интеграцию разработанной системы. Все перечисленные стадии играют ключевую роль в обеспечении эффективного взаимодействия AI-агента с пользователями, формировании точного и адекватного ответа, а также успешной интеграции с существующими каналами связи и эффективному решению поставленных задач. Исследование демонстрирует необходимость тщательного подхода ко всем этапам разработки для обеспечения высокого качества функционирования и надежности проектируемого AI-агента. Также в статье рассмотрен модуль оценки качества работы AI-агента, тестирование выполнено инструментом G-Eval и учитывающим семантику, что признано лучшим способом оценки точности ответов. В заключении представлен результат мониторинга качества генерируемых ответов и визуализация результатов.

**Ключевые слова:** AI-агент, адаптивная архитектура, естественно-языковой интерфейс, техническая поддержка, проектирование, моделирование, интеллектуальные системы, интерфейс.

**Введение.** Стремительное развитие искусственного интеллекта способствует трансформации современной рыночной среды и характеризуется ускоренными темпами развития и цифровизации всех процессов, а также минимизации рутинных задач,

выполнять которые способны AI-агенты, интеллектуальные помощники и чат-боты. Для обеспечения устойчивого и эффективного функционирования компаний в этих условиях, становится необходимым внедрение инновационных стратегий и инструментов, в том числе комплексная автоматизация ключевых бизнес-процессов на основе интеллектуализации процессов. Подобный подход позволяет не только оптимизировать операционную деятельность, но и рационально распределять ресурсы целевого действия, которые могут быть направлены на стратегическое планирование, определение приоритетных направлений развития и обеспечение конкурентоспособности компании в долгосрочной перспективе. Повышение эффективности и сокращение времени обработки больших массивов данных требует применения передовых и сложных методов анализа и управления информацией, что, в свою очередь, делает автоматизацию бизнес-операций стратегически важной задачей. Делегирование рутинных и стереотипных операций, а также стандартных бизнес-задач под контроль специализированных информационных систем или комплексных программно-аппаратных комплексов обеспечивает оптимизацию рабочих потоков, сокращение издержек, которые могут быть перенаправлены на повышение производительности, совершенствование процессов и стратегическое управление. AI-агенты с естественно-языковым интерфейсом для второй линии технической поддержки обладают способностью оперативно и эффективно и быстро отвечать на часто задаваемые вопросы пользователей, предоставлять исчерпывающую информацию о продуктах или услугах, и служить интеллектуальным помощником для сотрудников отдела технической поддержки. Такой подход позволяет значительно сократить время ожидания ответа на запрос, повысить скорость решения возникающих проблем и снизить общую нагрузку на специалистов технической поддержки [1].

Основой для проектирования модели адаптивной архитектуры AI-агента с естественно-языковым интерфейсом послужил анализ научных публикаций, подтвердивших актуальность разработки и внедрения таких систем. Практическая значимость данного исследования заключается в разработке и предложении стандартизированной адаптивной архитектуры AI-агента, способного к предиктивной и контекстно-зависимой обработке запросов первой линии технической поддержки. В статье Сианаки О.А. AI-агент определяется как компьютерная программа, разработанная с целью взаимодействия с пользователем посредством естественного языка таким образом, чтобы пользователь воспринимал это взаимодействие как диалог с живым человеком [2]. В работе Лэ Си Фонг Анг и Э. Раджа «Использование чат-ботов в обслуживании клиентов» исследуется удовлетворённость покупателей чат-ботом бронирования авиабилетов в Новой Зеландии. Исследование показывает, что сервис чат-бота подходит для решения несложных задач, в основном предоставляя заранее сформулированные ответы на повторяющиеся вопросы. В целом, респонденты показывают высокий уровень удовлетворённости сервисом, что говорит об актуальности данной технологии [3].

В своей научной работе «Типология чат-ботов» Трофименко А. и Задерейко А. предлагают семифакторную систематизацию чат-ботов [4], отдельные аспекты которой созвучны с принципами дифференциации, предложенными Ураевым Д. в его исследовании «Классификация и подходы к разработке чат-бот приложений». В зависимости от функционального назначения, виртуальные ассистенты делятся на генерализованные (предназначенные для общения на широкий спектр тем) и узкоспециализированные (сфокусированные на диалоге в рамках конкретной тематики или выполнении определенной задачи). По охвату аудитории выделяют индивидуальных и корпоративных виртуальных ассистентов. Разрабатываемый AI-агент является корпоративным узкоспециализированным виртуальным ассистентом. Основные характерные особенности интеллектуальных помощников с естественно-языковым интерфейсом и закрытой областью применения отражены в работе Хамза Эль-Алауи [5].

Также автором подробно описаны два основных подхода, применяемых при их построении. Для понимания пользовательского ввода, отслеживания состояния диалога и генерации ответов. Основное отличие интеллектуальных помощников с естественно-языковым интерфейсом и открытой областью (open-domain) от интеллектуальных помощников с естественно-языковым интерфейсом и узкой предметной областью (closed-domain) заключается в их способности обрабатывать запросы и вести диалог на разнообразные и широкие темы, не ограничиваясь небольшим набором predefined реакций или данных. Разрабатываемый AI-агент не предназначен для консультирования по широкому спектру тем и является узкоспециализированным сервисом, следовательно, обладает узкой предметной областью.

В ходе разработки применяются методы и подходы, зарекомендовавшие себя как наиболее эффективные в ходе проведения сравнительного анализа И. Саха и Дж. Ашер в статье «Исторический обзор подходов к разработке чат-ботов» [6]. В соответствии со словами автора, при разработке следует учитывать: цели разработки, аудиторию, каналы взаимодействия и потоки данных, так как это напрямую влияет на постановку задачи, а также на процесс тестирования.

Результаты многих исследований показывают, что правильное проектирование и внедрение интеллектуальных помощников может привести к существенному снижению операционных издержек, связанных с обслуживанием клиентов, повысить их удовлетворенность и улучшить общее впечатление от взаимодействия с компанией. Данные решения обладают значительным потенциалом для широкого внедрения во многих ИТ-компаниях и на предприятиях, поскольку позволяют существенно повысить эффективность работы служб поддержки, оптимизировать затраты и повысить уровень обслуживания клиентов.

Актуальность данного исследования продиктована потребностью в развитии стандартов архитектурного подхода для проектирования модели адаптивной архитектуры AI-агента, предназначенного для обработки входящих запросов первой линии технической поддержки. Предполагается, что данный агент будет выполнять функции дополнительного Hotline-разработчика, что приведет к повышению качества клиентской поддержки и сокращению времени ответа на обращения. Архитектура агента будет спроектирована для предиктивной и контекстно-зависимой обработки запросов от первой линии техподдержки. Выполняя роль «интеллектуального Hotline-разработчика», агент будет способен к самообучению и автономной генерации решений, что критически важно для достижения нового уровня качества клиентской поддержки и минимизации времени ответа. Оптимизация ресурсов Hotline-разработчика и сотрудников технической поддержки за счет автоматизации рутинных задач приведет к повышению оперативности и эффективности решения сложных клиентских запросов.

Основной целью данного исследования является развитие существующих подходов к разработке и внедрению модели адаптивной архитектуры AI-агента с естественно-языковым интерфейсом, в организацию процессов технической поддержки, при условии дообучения обновляемых данных и обеспечении высокой точности ответа по специфическим запросам, что значительно повышает производительность и качество работы отдела технической поддержки.

**Материалы и методы.** Анализ и сравнение базовых подходов к разработке интеллектуальных помощников и агентных систем определил архитектурный подход, как методологическую основу проводимого исследования. По нашему мнению, его применение к созданию виртуальных ассистентов обеспечит высокую производительность и повысит качество работы отдела технической поддержки международной компании, разрабатывающей широкий функционал ИТ-решений для бизнес-коммуникаций. В этой связи определена следующая последовательность

разработки: определение целей и домена, подготовка и адаптация данных, выбор и комбинирование методов генерации и извлечения исторических данных, оценка качества ответов, а также промышленная интеграция в корпоративную среду Microsoft Teams. Такой процесс как «Выбор закрытого домена» обусловлен требованиями точности и контекстной релевантности: AI-агент работает в пределах продуктовой линейки изучаемой IT-компании (рис. 1), использует термины и регламенты компании и опирается на закрытые корпоративные данные с контролируемой актуальностью. Это позволяет оптимизировать поиск и ранжирование знаний, минимизировать неоднозначности и сделать возможным цитирование первоисточников при получении ответов. С точки зрения подходов к ответу на вопросы предлагается применение гибридной схемы, которая синтезирует извлекающие и порождающие методы. Извлекающая компонента обеспечивает детерминированное нахождение релевантных фрагментов из корпоративного корпуса, тогда как порождающая компонента формирует связный ответ с учётом контекста диалога. Комбинация снижает вероятность фактических ошибок по сравнению с чисто порождающими моделями, благодаря опоре на явные источники знаний и их включению в контекст вывода.

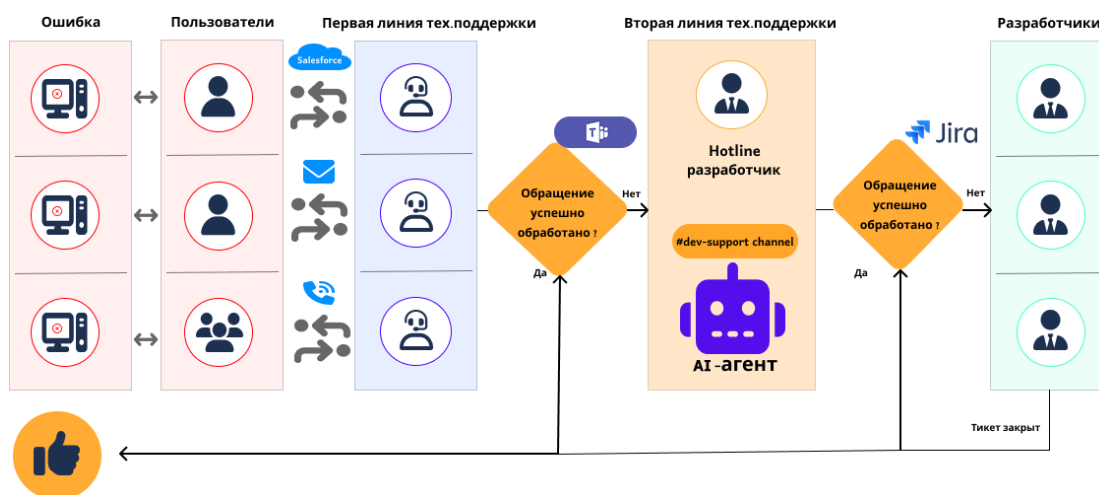


Рис. 1. Принципы взаимодействия пользователей, команды технической поддержки и команды разработчиков в организационной структуре IT-компании

Технологический стек подбирался, исходя из критериев производительности, стоимости владения, управляемости данными и независимости от вендоров. В таблице 1 представлено сравнение облачных технологий.

Таблица 1. Облачные технологии создания интеллектуальных помощников\*

Платформа	Функционал	Применение	Ценообразование
1	2	3	4
IBM Watson	NLP, алгоритмы ML, многоязычная поддержка, аналитика и отчетность	Обслуживание клиентов, электронная коммерция, управление социальными сетями	Lite (1000 запросов в месяц), Plus (0,0025 USD за API call), Enterprise (цены по запросу)
Amazon Lex	NLP, автоматическое распознавание речи, несколько языков, служба поддержки, полная интеграция с другими AWS	Широкая область применения	Пробная версия (с лимитами). Платные тарифы: 0,00075 USD за вызов API для текста, \$0.004 за выступление запроса.

Окончание табл. 1

1	2	3	4
Google DialogFlow	NLP, алгоритмы ML, предопределенные сущности, анализ тональности, многоязычная поддержка, интеграция с такими сервисами, как Google Assistant, Slack и Twitter	Широкий спектр применения	Стандарт (бесплатный, с ограничениями), Доступны платные тарифные планы (0,007 USD за вызов API для текста, 0.06 за минуту для аудио)
Rasa	Открытый исходный код, настраиваемые функции, алгоритмы машинного обучения, поддержка локального и облачного хостинга	Может создавать чат-ботов и виртуальных помощников	Open-source
Azure Bot Service	NLP, алгоритмы ML, многоязычная поддержка, аналитика и отчетность, интеграция с различными языками программирования и SDK	Виртуальные помощники, корпоративные интеграции, интернет вещей	\$0.50 за 1,000 сообщений

\* Составлено авторами по [7].

На основании проведенного сравнительного анализа облачных технологий для разработки интеллектуальных помощников обоснован выбор в пользу платформы Azure Bot Service. Платформа демонстрирует высокую адаптивность, позволяя эффективно настраивать ресурсы в соответствии с потребностями интеллектуального помощника. Она обеспечивает простую интеграцию с такими услугами, как Azure Cognitive Services и Azure Active Directory, что расширяет функциональность решения. Кроме того, Azure Portal предлагает обширные возможности для мониторинга, анализа производительности и обслуживания, а удобная загрузка кода через Deployment Center, Git, Azure DevOps и другие инструменты обеспечивают быстрое развертывание приложений. Необлачные технологии подразумевают наличие у разработчика навыков программирования, а разработка AI-агента осуществляется локально. В таблице 2 представлены ключевые характеристики и ограничения инструментов, предназначенных для разработки интеллектуальных помощников с интерфейсом естественного языка.

Сопоставление облачных платформ и локальных инструментов показало, что open-source стек позволяет повысить уровень контроля над данными и гибкость настройки AI-агента, а отсутствие вендор-зависимости, в свою очередь, предопределяет выбор функциональности в пользу программной оркестрации компонентов и интеграции AI-агента в существующую корпоративную инфраструктуру. Программная оркестрация определяется как процесс кодового управления цепочками обработки запроса, включающей загрузку и нормализацию данных, семантическое извлечение, формирование промпта, генерацию, постобработку и логирование. Оркестрация обеспечивает детерминированность, воспроизводимость и лёгкое подключение источников знаний при промышленной нагрузке.

В качестве подхода для сборки AI-агента был принят DIY-подход из открытых библиотек с настройкой под домен компании. Данный подход обеспечивает прозрачность, воспроизводимость и управляемость, снижает вендор-локин также усиливает

безопасность за счёт локального контроля и ускоряет адаптацию под сценарии второй линии поддержки.

**Таблица 2. Технические решения для создания интеллектуальных помощников, не требующие облачной инфраструктуры\***

Наименование технического решения	Язык программирования	Свободное программное обеспечение	Преимущество интеллектуальной разработки
Botkit	JavaScript	+	Создание пользовательских интеллектуальных помощников с использованием NLP, ML, интеграцией с внешними API и аналитикой.
Chatterbot	Python	+	Обработка естественного языка, интеграция со сторонними API, перевод языков и аналитические возможности.
Chatscript	C++	+	Обработка естественного языка, формирование ответов на основе декларативной логики, интеграция с внешними базами данных и API.
BotMan	PHP	+	Обработка естественного языка, интеграция с API сторонних разработчиков, языковой перевод и аналитика.

\* Составлено авторами по [8-9].

В качестве ядра оркестрации выбран фреймворк LangChain, модульная архитектура которого упрощает подключение ретриверов, векторных хранилищ и языковых моделей, а также стандартизирует промпт-инжиниринг и поддерживает воспроизводимые конвейеры.

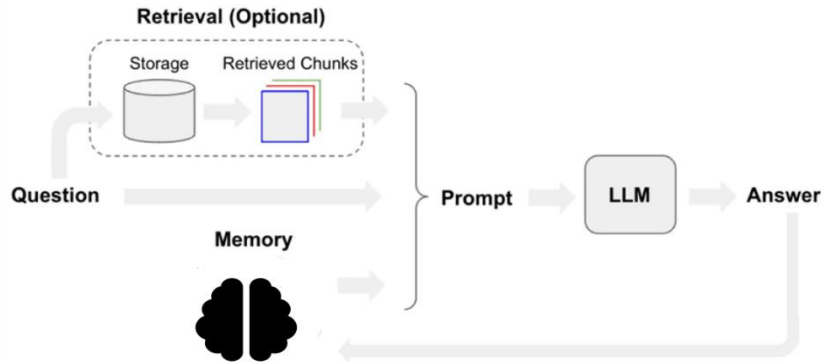
Использование фреймворка LangChain для создания AI-агента с естественно-языковым интерфейсом для платформы, включающей несколько линий технической поддержки, снижает интеграционную сложность, позволяет управлять контекстной памятью диалога, что существенно повышает качество работы AI-агента.

Ключевым механизмом снижения вероятности искажения эмпирической информации и обеспечения актуальности знаний является Retrieval Augmented Generation [5]. В рамках данного подхода корпоративный корпус данных импортируется, затем разбивается на фрагменты, кодируется в эмбединги (индексы для семантического поиска) и в завершении индексируется [10]. При запросе выполняется семантический поиск, а релевантные фрагменты подмешиваются в промпт [11]. Это повышает точность и специфичность ответов и позволяет цитировать первоисточники. Визуальные схемы конвейера отражены на рисунке 2.

Как показано на рисунке 2, по входному вопросу извлекаются релевантные фрагменты из корпоративного хранилища с использованием RAG и вместе с историей диалога собираются в промпт, после чего LLM формирует ответ. Схема поддерживает как режим с извлечением, так и упрощённый без извлечения для тривиальных запросов. Из представленной схемы следует управляемый цикл с обратной связью памяти. Память фиксирует ключевые факты и принятые решения, повышая связность диалога между ходами. Retrieval снижает вероятность искажения эмпирической информации и позволяет цитировать первоисточники. Программируемая оркестрация конвейера обеспечивает воспроизводимость, контроль качества и интеграцию в корпоративную среду.

Поскольку конвейер опирается на извлечение знаний и память диалога, для итогового качества критичны сохранение контекста, корректная работа с цитируемыми источниками и латентность. В связи с этим выполнен прикладной сравнительный анализ

LLM в составе RAG-контура с перебором схем извлечения, моделей встраивания и повторного ранжирования на исторических данных ИТ-компаний. Результаты сравнительного анализа метрик оценки качества работы двух больших языковых моделей представлены в таблице 3.



**Рис. 2. Retrieval-augmented generation для использования поиска при генерации естественно-языковых ответов на запросы пользователей**

**Таблица 3. Сравнительный анализ метрик оценки качества LLM: GPT-3.5-turbo и LLM: GPT-4o**

LLM: GPT-3.5-turbo					
Модель повторного ранжирования (Reranking model) Модель встраивания (Embedding model)	нет повторного ранжирования (no reranking)	<i>ms-marco-TinyBERT-L-2-v2</i>	<i>ms-marco-MultiBERT-L-12</i>	<i>ms-marco-MiniLM-L-12-v2</i>	<i>rank-T5-flan</i>
all-MiniLM-L6-v2	0.65	0.56	0.61	0.56	0.67
<i>text-embedding-3-small</i>	0.87	0.89	0.85	0.87	0.74
text-embedding-ada-002	0.80	0.77	0.82	0.74	0.70
LLM: GPT-4o					
Модель повторного ранжирования (Reranking model) Модель встраивания (Embedding model)	нет повторного ранжирования (no reranking)	<i>ms-marco-TinyBERT-L-2-v2</i>	<i>ms-marco-MultiBERT-L-12</i>	<i>ms-marco-MiniLM-L-12-v2</i>	<i>rank-T5-flan</i>
all-MiniLM-L6-v2	0.54	0.55	0.54	0.53	0.55
<i>text-embedding-3-small</i>	0.78	0.82	0.75	0.80	0.58
text-embedding-ada-002	0.73	0.70	0.76	0.74	0.62

В результате проведенного эксперимента можно обосновать выбор и доказать целесообразность использования LLM: GPT-3.5-turbo, так как замеры LLM: GPT-4o существенно уступают в метриках оценки качества.

Модель адаптивной архитектуры AI-агента с естественно-языковым интерфейсом для платформ технической поддержки, помимо обоснованного выбора модели

дообучения, включает также обязательные программные модули. Для проведения качественного и детального анализа на рисунке 3 представлены основные модули, использованные в процессе проектирования интеллектуального помощника, которые составляют неотъемлемую часть разработки.



Рис. 3. Схема программных модулей AI-агента

Все этапы разработки AI-агента, включая сбор и предобработку данных, обучение модели, разработку основного модуля AI-агента, валидацию и интеграцию, являются критически важными. Они определяют способность AI-агента понимать запросы пользователей, предоставлять точные и релевантные ответы, взаимодействовать через различные каналы и эффективно решать поставленные задачи. Пропуск или некачественное выполнение любого из этих этапов непосредственно влияет на качество и функциональность конечного продукта.

Сбор данных для обучения AI-агента осуществляется из действующих корпоративных каналов коммуникации и документации: Microsoft Teams как основной канал взаимодействия первой и второй линии, Confluence и Jira как источники требований, регламентов и историй инцидентов, а также архив Slack для ретроспективы обсуждений.

Этап предобработки данных включает: удаление паролей, спецсимволов, символов Unicode и упоминаний пользователей, разбиение текста на chunk-и (фрагменты) и создание и сохранение эмбедингов (векторных представлений текста) в векторной базе данных.

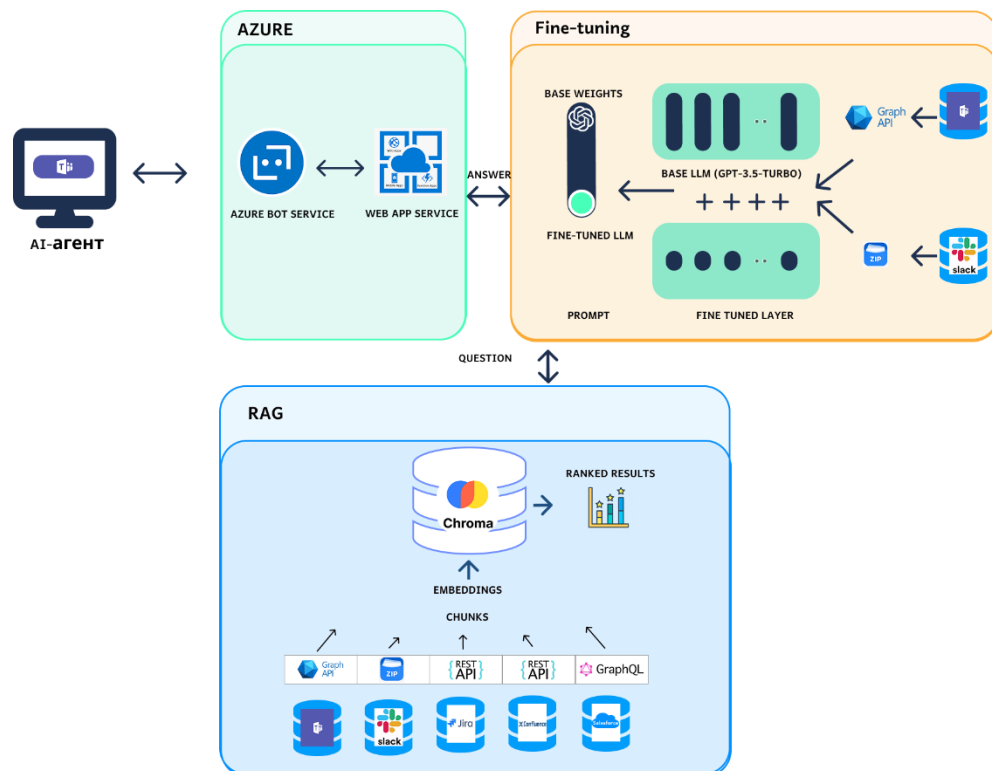
Модуль обучение модели – Fine-tuning включает загрузку данных из предварительно созданной векторной базы данных, создание обучающего файла и отправку его на сервер OpenAI для запуска процесса обучения.

Модуль валидации – evaluator включает классы для оценки качества ответов LexiTech. Алгоритм основан на наборе данных с вопросами и эталонными ответами. Сгенерированные моделью ответы сравниваются с эталонными с помощью специального промпта, который задает критерии оценки [12].

Основной модуль AI-агент – инициализация языковой модели, буфера памяти, настройки доступа к векторным БД и промптов для генерации ответов. Основной модуль инициализирует AI-агент, его языковую модель и буфер памяти, а также настраивает доступ к векторным базам данных и специальным промптам для генерации ответов.

Модуль интеграции с Teams: REST API методы для взаимодействия и файлы для развертывания в Azure. Во внутреннем корпоративном сервисе IT-компании Microsoft Teams будет интегрирован AI-агент, обладающий контекстной памятью и предоставляет ссылки на внутренние ресурсы, использованные для генерации ответа.

**Результаты.** Основным результатом исследования является предложенная модель адаптивной архитектуры интеллектуального агента, оснащённого системой естественного языкового взаимодействия, функционирующего в рамках мультиканальной платформы техподдержки. Проектирование и реализация всех программных модулей AI-агента позволили собрать необходимые компоненты информационной архитектуры и за счет правильной и оптимальной их интеграции, построить общую схему работы модели адаптивной архитектуры AI-агента с естественно-языковым интерфейсом для платформы, включающей несколько линий технической поддержки.



**Рис. 4. Модель адаптивной архитектуры AI -агента с естественно-языковым интерфейсом для платформы, включающей несколько линий технической поддержки**

Исходя из модели адаптивной архитектуры, представленной на рисунке 4, следует, что разрабатываемый в рамках данного исследования AI-агент с естественно-языковым интерфейсом обучается с применением подходов RAG на исторических данных IT-компании из разных корпоративных сервисов. Для стилизации ответов максимально приближенных к человеческой речи использовался метод для тонкой настройки глубокого обучения – Fine-tuning. А интеграция готового разработанного AI-агента в корпоративную

среду Microsoft Teams для запуска в эксплуатацию реализована через Azure Bot Service с программным кодом в GitHub и развертыванием через Azure Web Service.

Принцип работы AI-агента, как Hotline-разработчика с естественно языковым интерфейсом, в рамках деятельности команды технической поддержки AI-компании, происходит следующим образом: Запрос пользователя векторизуется, далее происходит семантический поиск, и результаты используются для создания промпта для fine-tuned модели, которая и формирует ответ.

**Заключение.** Представленное исследование посвящено проектированию модели адаптивной архитектуры AI-агента для автоматизации рутинных и стандартизированных операций, что позволит сократить временные затраты сотрудников технической поддержки ИТ-компании. Это способствует оптимизации рабочих процессов, повышению производительности команды технической поддержки. Полученные результаты работы AI-агента демонстрируют стилистику ответов, максимально приближенную к естественной человеческой коммуникации, а также высокий уровень точности ответов, с вероятностью 95% AI-агент даст верный ответ, оставшиеся 5% зависят от точности формулировки и контекста задаваемого вопроса.

В качестве тестирования работы AI-агента, используется инструмент G-Eval, который базируется на LLM, а также учитывает семантику, представляется наиболее подходящим методом для оценки точности ответов AI-агента [14]. Для наглядного представления результатов использован фреймворк LangSmith, предоставляющий удобный интерфейс для мониторинга качества ответов LLM и отслеживания регрессии при дообучении или изменении алгоритма.

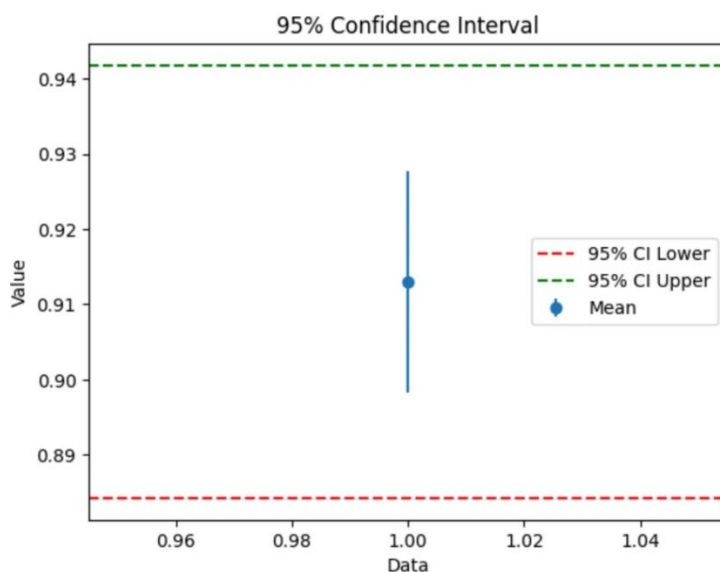


Рис. 5. 95% доверительный интервал для оценки качества генерируемых ответов

Для оценки качества ответов AI-агента использовался предварительно созданный датасет из 100 вопросов и эталонных ответов, взятых из базы знаний, что позволило определить доверительный интервал, который составляет 95% и находится в диапазоне от 0.8842 до 0.9417, как показано на рисунке 5. Следовательно, с вероятностью 95% средняя точность ответов GPT-3.5-turbo находится в этом интервале. Средняя корректность ответов (0.913) свидетельствует о высоком качестве модели. Следовательно, ответы AI-агента в целом точны и соответствуют заданным критериям.

Научная новизна работы состоит в том, что получил дальнейшее развитие архитектурный подход для проектирования адаптивного контура на основе до обучения

AI-агента с естественно-языковым интерфейсом для платформ технической поддержки. Таким образом, представляемое исследование характеризуется высоким уровнем практической значимости и включает обоснованные рекомендации, направленные на обеспечение устойчивой динамики развития ИТ-компаний.

### Список литературы

1. Чижик А.В. Создание чат-бота: обзор архитектур и векторных представлений текста / А.В. Чижик, Ю.А. Жеребцова // *Международный журнал открытых информационных технологий*. 2020. – 8, №7.
2. Кан Дж. Чатбот: Архитектура, Дизайн и Разработка / Дж. Кан 2017.
3. Раджа Е. Использование чат-ботов в обслуживании клиентов: тематическое исследование компании «Эир Нью Зиланд» / Е. Раджа, Ши Фон Ан Л. е-пресс, 2021. – С. 161-176.
4. Трофименко О. Типология чатботов / О. Трофименко, И. Прокоп, Н. Логинова, А. Задерейко // *Международная научная и практическая конференция «Интеллектуальные системы и информационные технологии»*, 13-19 сентября 2021 г., Одесса, Украина. – С. 181-195.
5. Шейх С., Тивари В., Сингал С. Порождающая модель чат-бота управления персоналом с использованием глубокого обучения / С. Шейх, В. Тивари, С. Сингал // *V Международная конференция «Науки о данных и Инженерии»*, 2019, Патна, Индия. – С. 126-132.
6. Вейценбаум Дж. Вычислительная лингвистика / Дж. Вейценбаум // *Коммуникации ассоциации вычислительной техники*. – 1966. – 9, № 1. – С. 36-45.
7. Набоков, Т.Н. Облачные технологии как технология рынка информационно коммуникационных технологий: динамика развития и особенности / Т.Н. Набоков, М.И. Попова // *Цифровизация экономики: направления, методы, инструменты : сборник материалов VI Всероссийской научно-практической конференции (Краснодар, 15–19 января 2024): в 3-х т.* – Краснодар: КубГАУ – 2024. – С. 147-148.
8. Ураев Д.А. Классификация и методы создания приложений чат-ботов / Д.А. Ураев // *Международный научный обзор*. – 2019. –Т. LXIV. – С. 30-33.
9. Крылов, В.С. Компьютерный анализ эмоциональной компоненты научных публикаций на примерах в физике и экономике / В.С. Крылов, А.А. Кудрявцев, Л.Н. Абдурайимов // *Онтология проектирования*. – 2020. – Т. 10, № 4(38). – С. 449-462. – DOI 10.18287/2223-9537-2020-10-4-449-462. – EDN VJWTKW.
10. Ханов, Р.Г. Компоненты пользовательского интерфейса в нативных мобильных приложениях с дополненной реальностью / Р.Г. Ханов, Д.А. Евдокименко // *Электронные библиотеки*. – 2020. –Т. 23. – №5. – С. 1104-1118.
11. Zachman J.A. The Framework for Enterprise Architecture – Cell Definitions // ZIFA, 2003. 17. Roberts, E.B. The Dynamics of Research and Development. New York: Harper & Row. 1964.
12. Sojasingarayar, A. Seq2Seq AI Chatbot with Attention Mechanism. IA School // University-GEMA Group Conference Materials. – 2020. – P. 1-15.
13. Roberts, E.B. The Dynamics of Research and Development. New York: Harper & Row. 1964.
14. Zhang L. LLM-AIDSim: LLM-Enhanced Agent-Based Influence Diffusion Simulation in Social Networks / L. Zhang, Yu. Hu, W. Li [et al.]. – *Systems*, 2025. – Vol. 13, No. 1. – P. 29.

**Коломыцева Анна Олеговна**, канд. экон. наук, доцент, доцент кафедры УНЦ «Искусственный интеллект», ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина», Екатеринбург, Россия

E-mail: [a.o.kolomytseva@urfu.ru](mailto:a.o.kolomytseva@urfu.ru)

ORCID: 0000-0002-2797-5487

AuthorID: 839547

**Гейченко Евгений Игоревич**, магистрант кафедры информационных систем, СПбГЭТУ «ЛЭТИ» «Санкт-Петербургский государственный электротехнический университет имени В.И. Ульянова Ленина», Санкт-Петербург, Россия

Email: [geychenko.1995@mail.ru](mailto:geychenko.1995@mail.ru)

ORCID: 0009-0009-7076-0367

AuthorID: 130874

**Коломыцева Ирина Константиновна**, магистрант кафедры информационных технологий и систем управления, ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина», Екатеринбург, Россия

E-mail: [irinakolomyceva128@gmail.com](mailto:irinakolomyceva128@gmail.com)

ORCID: 0009-0001-7764-5064

AuthorID: 1277986

**Кожевников Арсений Андреевич**, магистрант кафедры информационных технологий и систем управления, ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина», Екатеринбург, Россия

E-mail: [Kozhevnikovarseniy@urfu.me](mailto:Kozhevnikovarseniy@urfu.me)

ORCID: 0009-0006-7892-8409

AuthorID: 53247

*Поступила в редакцию 10.09.2025 г.*

UDC 311.42: 519.252

DOI 10.5281/zenodo.17849657

**KOLOMYTSEVA Anna**<sup>1</sup>,  
**GEICHENKO Evgeny**<sup>2</sup>,  
**KOLOMYTSEVA Irina**<sup>1</sup>,  
**KOZHEVNIKOV Arseniy**<sup>1</sup>

<sup>1</sup> Ural Federal University named after the First President of Russia B.N. Yeltsin, Mira St., 19, Yekaterinburg, Russia, 620062

<sup>2</sup> St. Petersburg State Electrotechnical University named after V.I. Ulyanov Lenin, St. Petersburg, Russia

## **DESIGNING AN ADAPTIVE AI AGENT ARCHITECTURE MODEL WITH A NATURAL LANGUAGE INTERFACE FOR TECHNICAL SUPPORT PLATFORMS**

The article presents the result of designing an adaptive architecture model for an intelligent agent with a natural language interface designed to operate on a platform that provides multi-line technical support for users. The analysis of scientific publications of researchers related to the development of intelligent systems and AI agents has been carried out to create a high-performance automated user support tool that reduces the burden on the technical support team and the development team, which improves the quality of customer support and reduces the system response time to their requests. The AI agent architecture is designed for predictive and context-sensitive analysis of first-level technical support requests. Due to the ability to self-learn and automatically generate solutions, the system is able to provide a qualitatively new level of customer service and minimize response delay. The article describes the structure of the main modules, which include an informed selection of models before training and key development components. The most important stages of AI agent development include data collection and preprocessing, model training, creation of the main software module, testing and integration of the developed system. All these stages play a key role in ensuring effective interaction of the AI agent with users, forming an accurate and adequate response, as well as successful integration with existing communication channels and effective solution of tasks. The study demonstrates the need for a thorough approach to all stages of development to ensure high-quality functioning and reliability of the designed AI agent. The article also discusses the module for evaluating the quality of an AI agent's work. Testing was performed using the G-Eval tool and taking into account semantics, which is recognized as the best way to evaluate the accuracy of responses. In conclusion, the result of monitoring the quality of the generated responses and visualization of the results is presented.

**Key words:** *AI agent, adaptive architecture, natural language interface, technical support, design, modeling, intelligent systems, interface.*

### **References**

1. Chizhik, A., Zhrebtsova, Y. (2020). [Challenges of Building an Intelligent Chatbot]. *International Conference "Internet and Modern Society"*. (In Russian).
2. Cahn, J. (2017). [Chatbot: Architecture, Design, & Development]. (In Russian).
3. Rajah, E., Thi Phuong Anh, L. (2021). [Using Chatbots in Customer Service: A Case Study of Air New Zealand]. ePress, 161–176. (In Russian).
4. Trofymenko, O., Prokop, Y., Loginova, N. & Zadereyko, A. (2021) [Classification of chatbots]. In: *International Science and Practical Conference «Intellectual Systems & Information Technology»*, 13–19 September 2021, Odessa, Ukraine. pp. 181–195. (in Ukraine).

5. Sheikh S., Tiwari V., Singal S. *Generative model chatbot for Human Resource using Deep Learning*. / S. Sheikh, V. Tiwari, S. Singal. // V International conference of data science and Engineering, Patna, India. – 2019. – pp. 126-132. (In Russian).
  6. Weizenbaum, J. (1966) [Computational Linguistics]. *Communications of the ACM*. 9(1), 36–45. (In Russian).
  7. Nabokov, T.N. & Popova, M.I. (2024) [Cloud technologies as a technology of the information and communication technology market: dynamics of development and features]. In: *Digitalization of the Economy: Directions, Methods, Tools: Collection of Materials of the VI All-Russian Scientific and Practical Conference*, 15–19 January 2024, Krasnodar, Russia. Krasnodar: Kuban State Agrarian University named after I.T. Trubilin. pp. 147–148. (In Russian).
  8. Urayev, D.A. (2019) [Classification and methods for creating chatbot applications]. *International Scientific Review*. LXIV, 30–33. (In Russian).
  9. Krylov, V.S., Kudryavtsev, A.A. & Abdurayimov, L.N. (2020) [Computer analysis of the emotional component of scientific publications using examples in physics and economics]. *Ontology of Design*. 10(4), 449–462. DOI: 10.18287/2223-9537-2020-10-4-449-462. EDN: VJWTKW. (In Russian).
  10. Khanov, R.G. & Evdokimenko, D.A. (2020) [Components of the user interface in native mobile applications with augmented reality]. *E`lektronny`e biblioteki = Electronic libraries*. Vol. 23. No.5, pp. 1104-1118. (In Russian).
  11. Zachman J.A. *The Framework for Enterprise Architecture – Cell Definitions* // ZIFA, 2003.17. Roberts, E.B. *The Dynamics of Research and Development*. New York: Harper & Row. 1964.
  12. Sojasingarayar, A. *Seq2Seq AI Chatbot with Attention Mechanism*. IA School // University-GEMA Group Conference Materials. – 2020. – P. 1-15.
  13. Roberts, E.B. *The Dynamics of Research and Development*. New York: Harper & Row. 1964.
  14. Zhang L. *LLM-AIDSim: LLM-Enhanced Agent-Based Influence Diffusion Simulation in Social Networks* / L. Zhang, Yu. Hu, W. Li [et al.]. – Systems, 2025. – Vol. 13, No. 1. – P. 29.
- 

**Kolomytseva Anna**, Candidate of Economic Sciences, Associate Professor of the Department of UNC «Artificial Intelligence», Ural Federal University named after the First President of Russia B.N. Yeltsin, Yekaterinburg, Russia

E-mail: [a.o.kolomytseva@urfu.ru](mailto:a.o.kolomytseva@urfu.ru)

ORCID: 0000-0002-2797-5487

AuthorID: 839547

**Geichenko Evgeny**, Master's Student of the Department of Information Systems, St. Petersburg State Electrotechnical University named after V.I. Ulyanov Lenin, St. Petersburg, Russia

Email: [geychenko.1995@mail.ru](mailto:geychenko.1995@mail.ru)

ORCID: 0009-0009-7076-0367

AuthorID: 130874

**Kolomytseva Irina**, Master's Student of the Department of Information Technology and Management Systems, Ural Federal University named after the First President of Russia B.N. Yeltsin, Yekaterinburg, Russia

E-mail: [irinakolomytseva128@gmail.com](mailto:irinakolomytseva128@gmail.com)

ORCID: 0009-0001-7764-5064

AuthorID: 1277986

**Kozhevnikov Arseniy**, Master's Student of the Department of Information Technology and Management Systems, Ural Federal University named after the First President of Russia B.N. Yeltsin, Yekaterinburg, Russia

E-mail: [Kozhevnikovarseniy@urfu.me](mailto:Kozhevnikovarseniy@urfu.me)

ORCID: 0009-0006-7892-8409

AuthorID: 53247

*Received 10.09.2025*