

И. А. Любимов^{1}, П. Ю. Бугаков¹*

Разработка экспертной системы для автоматизации работы первой линии службы технической поддержки с использованием языковой модели GPT

¹ Сибирский государственный университет геосистем и технологий, г. Новосибирск,

Российская Федерация

* e-mail: lyubim.off@mail.ru

Аннотация. Данная статья посвящена разработке экспертной системы с использованием языковой модели GPT от компании OpenAI для автоматизации процессов первой линии технической поддержки. Основная цель исследования - создать эффективный инструмент, способный значительно улучшить качество и скорость обслуживания клиентов, снизив при этом нагрузку на операторов. Разработка системы проводилась на языке Python, ведь он обеспечивает гибкость и мощные возможности для работы с искусственным интеллектом и машинным обучением. Для хранения данных и управления их целостностью была использована база данных MongoDB с интерфейсом управления mongo-express. Применение контейнеризации через Docker и docker-compose позволило обеспечить удобное и надежное развертывание приложения, а также его масштабирование. В результате работы была создана система, которая не только адекватно реагирует на запросы пользователей, но и предоставляет возможность анализировать типичные проблемы и вопросы. Реализация такой системы на практике может значительно повысить уровень удовлетворенности клиентов и оптимизировать работу службы поддержки.

Ключевые слова: программное обеспечение, автоматизации, ChatGPT, Telegram, Python, Docker, MongoDB

I. A. Lyubimov^{1}, P. Yu. Bugakov¹*

Development of an Expert System to Automate the Work of the First line of Technical Support Using the GPT Language Model

¹ Siberian State University of Geosystems and Technologies, Novosibirsk, Russian Federation

* e-mail: lyubim.off@mail.ru

Abstract. This article is devoted to the development of an expert system using the GPT language model from OpenAI to automate the processes of the first line of technical support. The main purpose of the research is to create an effective tool that can significantly improve the quality and speed of customer service, while reducing the burden on operators. The system was developed in Python, because it provides flexibility and powerful capabilities for working with artificial intelligence and machine learning. A MongoDB database with a mongo-express management interface was used to store data and manage its integrity. The use of containerization through Docker and docker-compose made it possible to ensure convenient and reliable deployment of the application, as well as its scaling. As a result of the work, a system was created that not only responds adequately to user requests, but also provides an opportunity to analyze typical problems and questions. The implementation of such a system in practice can significantly increase the level of customer satisfaction and optimize the work of the support service.

Keywords: software, automation, ChatGPT, Telegram, Python, Docker, MongoDB

Введение

В современном мире монетизация продуктов часто сопрягается с необходимостью предоставления качественной клиентской поддержки. Однако на практике организации стремятся минимизировать расходы, в том числе за счет оптимизации затрат на обслуживание клиентов. Это приводит к сокращению числа сотрудников, которые обрабатывают запросы, что, в свою очередь, зачастую увеличивает время ожидания ответа, даже на наиболее типовые вопросы. Такая ситуация может негативно сказаться на удовлетворенности клиентов.

Из-за этих проблем возникает необходимость внедрения инновационных решений, способных улучшить эффективность работы служб поддержки, при этом не увеличивая общие затраты. Одним из таких решений является разработка и внедрение экспертной системы для автоматизации первой линии поддержки. Задачей данной системы является обработка стандартных обращений пользователей, что позволит сократить время реагирования на запросы и уменьшить нагрузку на живых операторов.

Для реализации предложенного решения было выбрано использование дообученной модели GPT от OpenAI, которая может аналогично оператору автоматически генерировать качественные ответы на типичные вопросы. Подобный подход не только оптимизирует работу службы поддержки, но и способствует повышению общей удовлетворенности пользователей за счет снижения времени ожидания и повышения качества обслуживания.

Методы и технологии

В рамках разработки экспертной я столкнулся с рядом задач, которые охватывали весь спектр от формулировки требований до реализации и тестирования системы. Для реализации некоторых функциональных требований необходимо было использовать одну из современных версий модели GPT.

GPT-3 – версия, отличающаяся улучшенной производительностью, способная генерировать тексты высокого качества на основе меньшего количества подсказок, в сравнении с другими версиями моделей. Но в данной версии ограничен доступ к API, а стоимость использования может быть высока.

GPT-3,5 – версия с улучшенной способностью к обобщению, совершающая меньшее количество ошибок, отличающаяся более высокой степенью универсальностью в ответах и имеющая доступ к do-it-yourself-дообучению через сервисы OpenAI. Но для развертывания в коммерческих целях требуется разрешение от OpenAI.

GPT-4 – самая современная версия с улучшенными функциями и возможностями, но имеющая также высокую стоимость, сложность в интеграции и потенциальную недоступность для широкой аудитории.

Для выбора версии модели были определены такие критерии, как: стоимость, доступность, точность ответов, возможность своевременного дообучения,

а также наличие технической поддержки и документации. Проанализировав перечисленные версии, для реализации проекта была выбрана GPT-3.5.

Для реализации web-ресурса был выбран стек таких технологий, как PHP+bootstrap. Данная связка дает такие преимущества, как: адаптивность, кросс-браузерность, сокращение времени на разработку.

Результаты

Машинное обучение (ML) – ветвь искусственного интеллекта, с основной идеей самообучения компьютера поставленной задаче вместо использование заранее написанного алгоритма. В целом все обучение можно разделить на два модуля обучение с учителем и обучение без учителя.

При обучении с учителем имеется набор данных, на основании которых нужно что-то предсказать. При обучении без учителя отсутствует возможность вмешаться человеку, а в наборе данных есть то, какие свойства нужны.

В машинном обучении присутствует большое число алгоритмов, причем некоторые являются довольно универсальными (метод опорных векторов, бустинг над решающими деревьями, нейронные сети). Нейронная сеть – это сеть нейронов, где каждый нейрон является математической моделью реального нейрона. Это одна из передовых технологий, используемых в машинном обучении и применяемая в огромном количестве существующих решений.

Для обучения модели был собран датасет различных обращений и отзывов от пользователей компании X с ее странички в PlayMarket и AppStore.

Следующим шагом было подготовить датасет для работы с нейронной сетью. После обучения нейронная сеть стала выдавать первые корректные ответы пользователям.

Для улучшения реагирования нейронной сети на обращения пользователя, а также для обеспечения более быстрой генерации ответов, было принято решение добавить категории на существующие знания. Это необходимо для разбивки базы знаний на смысловые сегменты.

После этого нейронная сеть стала разбивать обращения по таким категориям, как баг, фичреквест, благодарность, другое.

Таким образом, после обучения и подключения нейронной сети к telegram-боту, мы имеем следующее рабочее поведение экспертной системы (рис. 1).

Пользовательская часть экспертной системы представлена web-ресурсом, написанным с использованием PHP и bootstrap. Адрес: tipo-knowledgebase.ru

При входе в систему пользователя встречает окно авторизации. Для экстренных случаев реализована функция восстановления пароля через почту или номер телефона, а также создание нового аккаунта. После входа пользователь попадает на главную страницу ресурса (рис. 2).

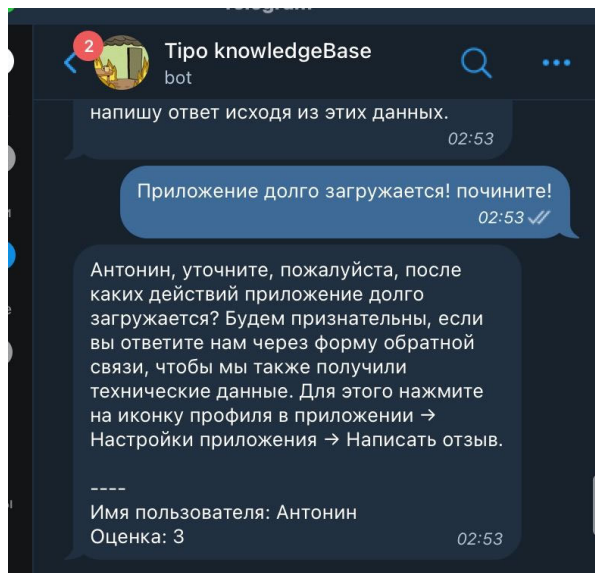


Рис. 1. Работа telegram бота как точки входа в экспертную систему

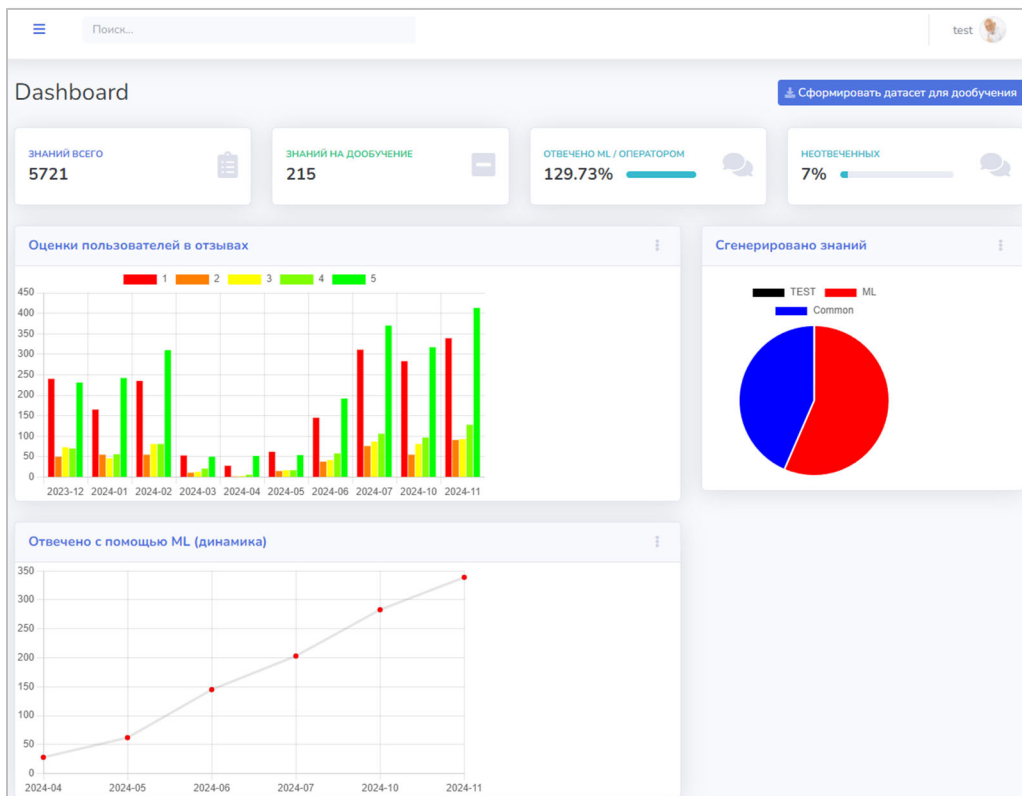


Рис. 2. Главная страница личного кабинета

На главной странице отображена следующая информация:

- общее количество знаний о продукте;
- общее количество знаний для дообучения;
- статистика о количестве ответов ML по отношению к ответам оператора;

- процентное количество неотвеченных обращений;
- сводка об оценках пользователей в отзывах продукта;
- сводка о соотношении количества сгенерированных знаний;
- динамика ответов с помощью ML.

Для удобства поиска информации в знаниях, в главной шапке ресурса расположен инструмент поиска, при помощи которого можно найти необходимое знание. В каждой из перечисленных веток дерева проекта содержатся знания, разбитые по категориям. Их общий вид представлен на рис. 3.

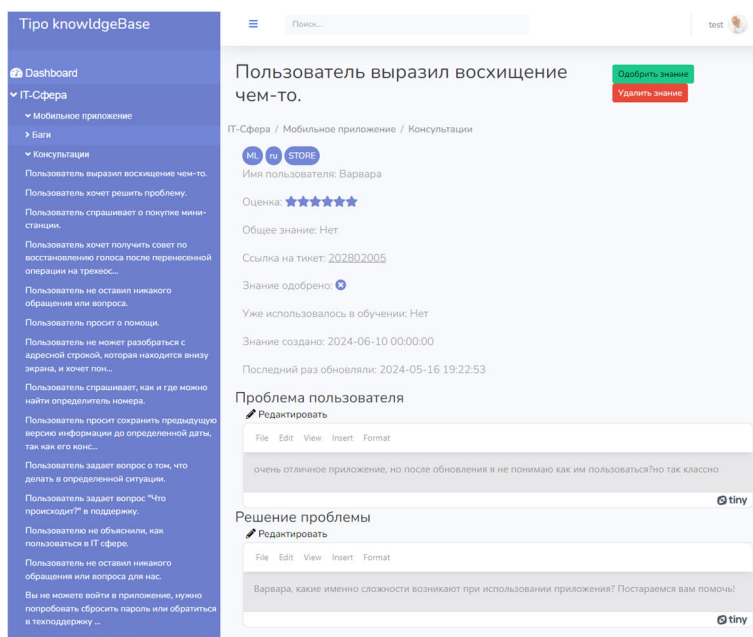


Рис. 3. Шаблонный вид знания

Пользователь может внести какие-либо правки в обращение или знание, если в целом оно удовлетворяет требованиям. Если в тексте ML содержится много неточностей или ошибочных данных, в верхней части экрана администратор может удалить его (кнопка «удалить знание»), иначе – принять (кнопка «одобрить знание»).

Процесс тестирования разработанной экспертной системы состоял из нескольких этапов, направленных на оценку функциональности, производительности и удобства использования системы. Тестирование проводилось как разработчиками, так и группой пользователей, представляющих первую линию службы технической поддержки.

Функциональное тестирование включало проверку корректности работы базы знаний, в том числе добавление, редактирование и удаление записей, поиск информации, навигацию по разделам. Также проводилось тестирование телеграм-бота, включая отправку сообщений, получение ответов, обработку запросов пользователей. Кроме того, была проверена интеграция между базой знаний и телеграм-ботом, обеспечивающая бесперебойную передачу данных.

Тестирование производительности включало оценку времени отклика системы при нагрузках, проверку стабильности работы системы при одновременном доступе нескольких пользователей, а также мониторинг потребления ресурсов сервера и контейнеров Docker.

Для оценки удобства использования было проведено юзабилити-тестирование с привлечением группы пользователей из службы технической поддержки. Собиралась обратная связь и предложения по улучшению, оценивалась интуитивность и простота навигации и взаимодействия с ботом.

Тестирование безопасности включало проверку системы на устойчивость к несанкционированному доступу и атакам, тестирование механизмов аутентификации и авторизации пользователей.

Результаты тестирования показали, что разработанная экспертная система успешно выполняет свои основные функции. База знаний обеспечивает эффективное хранение и доступ к информации, а telegram-бот корректно обрабатывает запросы пользователей и предоставляет релевантные ответы. Система демонстрирует хорошую производительность и стабильность работы даже при высоких нагрузках.

Заключение

Проект достиг своих основных целей. Основное достижение заключается в значительном улучшении качества и скорости обработки запросов пользователей за счет внедрения интеллектуальной системы, основанной на языковой модели GPT-3.5.

База знаний доступна по адресу <https://tipo-knowledgebase.ru/> (логин test, пароль 123). Точка входа для пользователей реализована в виде Телеграм бота, доступного по ссылке: https://t.me/tipo_knb_bot.

Эффективность системы обуславливается рядом факторов: во-первых, точной настройкой модели под специфику вопросов, свойственных сфере; во-вторых, созданием базы знаний, которая облегчила процесс интеграции полученных ответов в рабочие процессы; в-третьих, разработкой интерфейсов, которые позволяют пользователям легко взаимодействовать с системой.

Кроме того, благодаря анализу и визуализации статистики использования системы удалось выявить ключевые точки для дальнейшего улучшения функционала. Программная реализация на базе технологий Docker и MongoDB обеспечила устойчивую и масштабируемую структуру, что позволяет смотреть в будущее на возможности дальнейшего расширения и улучшения системы.

Таким образом, разработанная система не только устранила многие изначальные слабости процесса обслуживания клиентов, но и подготовила почву для будущих инноваций в этом направлении, что гарантирует принципиально новый уровень взаимодействия с пользователем.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Инструкция: как создавать ботов в Telegram – 2015. – [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/articles/262247/>.

2. Как сделать бота в Телеграм: пошаговая инструкция – 2022. – [Электронный ресурс]. – Режим доступа: <https://vc.ru/dev/530248-kak-sdelat-bota-v-telegram-poshagovaya-instrukciya>.
3. How to create your own Telegram bot – 2021. – [Электронный ресурс] – Режим доступа: <https://medium.com/spidernitt/how-to-create-your-own-telegram-bot-63d1097999b6>.
4. A guide to Telegram bot library – 2021. – [Электронный ресурс] – Режим доступа: <https://telegambots.github.io/book/1/quickstart.html>.
5. Makey P.I. Generative Pre-Trained Transformer 3 (GPT-3) / Makey P.I. // Электронные системы и технологии. – Минск: Белорусский государственный университет информатики и радиоэлектроники, 2022. – с. 901-902.

© И. А. Любимов, П. Ю. Бугаков, 2024