

Авторы статьи:

Валиев Рустем Масхутович — генеральный директор, к.э.н., доцент кафедры менеджмента и маркетинга ВАВТ

Буглов Глеб Олегович — руководитель проектов

Сафонова Анастасия Александровна — ведущий аналитик

Янковская Екатерина Владимировна — ведущий эксперт

Введение

Что такое чат-боты

Чат-бот — это программа, которая имитирует человеческое общение. Алгоритм бота ориентирован как на ведение неструктурированного диалога, так и на решение поставленных клиентом задач. Большое распространение боты получили в бизнесе, как в высококонкурентной среде, где особенно важна клиентская поддержка. Аналитики KPMG¹ поставили чат-ботов на вторую строчку среди самых популярных инструментов цифровизации у российских компаний (на первом месте — большие данные и предиктивная аналитика).

По данным исследования Juniper, в среднем один чат-бот экономит 4 минуты времени клиента в сравнении со звонком в службу поддержки, благодаря чему бизнес, внедривший бота, к 2022 г. может сэкономить более \$8 млрд². Выгоду приносят несколько принципиальных преимуществ чат-бота:

- многозадачность: обрабатывает до нескольких тысяч запросов одновременно, что существенно снижает время ожидания клиента;
- доступность: под рукой у клиента ежедневно в любое время дня и ночи;
- надежность: отвечает по сценарию, не может нагрубить или сорваться на клиента.

Несмотря на то, что операторы до сих пор лучше программ справляются со сложными, многосоставными задачами, сегодня почти 3 из 4 клиентов ожидают увидеть чат-бота на сайте компании³.

Другим важным отличием ботов является простая интеграция с внутренними информационными системами заказчика, такими как CRM-системы и базы данных, причем не только для непосредственного взаимодействия с клиентами, но и для сбора данных. Голосовой бот запишет аудио разговора, автоматически переведет его в текст и при желании выделит и запомнит нужные сущности (имя клиента, возраст, адрес, номер телефона и др.). Собранные данные можно

¹ <https://ict.moscow/research/cifrovye-tehnologii-v-rossiyskih-kompaniyah/>

² <https://www.juniperresearch.com/new-trending/analystxpress/july-2017/chatbot-conversations-to-deliver-8bn-cost-saving>

³ <https://www.intercom.com/blog/the-state-of-chatbots/>

использовать для улучшения пользовательского опыта: согласно исследованию Deloitte, компании с клиентоориентированной бизнес-моделью на 60% более прибыльны, чем компании, для которых клиенты не находятся в центре внимания⁴. Собирая данные о посетителях, чат-бот становится более персонализированным, а компания лучше узнает свою аудиторию.

Чат-бот, благодаря интерактивности, чаще удерживает и возвращает клиента, что также приносит дополнительную выгоду, ведь вернувшиеся клиенты приносят выгоду на 67% больше, чем новые⁵. Этот эффект можно усилить, если научить бота говорить на отвлеченные темы и расширить его базу знаний. Даже если посетитель после общения с ботом не заинтересуется товаром, вы ничего не потеряете — он не занимал линию и не отвлекал сотрудника. Напротив, если коммуникация ему понравится, он вернется вновь и даже может привлечь новых потенциальных клиентов.

Цифровизация бизнес-процессов — очевидный тренд последних лет, а период карантина продемонстрировал преимущества автоматизированных и легко масштабируемых систем в кризисной ситуации. В начале марта 2020 года Gartner⁶ советовал руководителям больше инвестировать в каналы общения с клиентами и роботизировать рутинные операции, чтобы направить силы и время сотрудников на решение более сложных и комплексных задач. Чат-боты автоматизируют все нужные процессы, а в случае резкого роста нагрузки, их можно быстро адаптировать под любое количество операций и сохранить уровень сервиса. Наглядным примером здесь может служить сфера ритейл: после удачной маркетинговой кампании нагрузка на клиентскую поддержку стремительно возрастает и без роботизированного решения начинает страдать качество обслуживания.

Рынок чат-ботов

Рынок чат-ботов один из самых быстрорастущих — с 2.6 миллиардов долларов в 2019 году он вырастет до 9.4 миллиарда долларов к 2024 году⁷. Еще в 2017 году Gartner предсказал, что к 2021 году более половины предприятий будут тратить на создание чат-ботов больше, чем на разработку традиционных мобильных приложений⁸. Чат-боты уже применяются почти во всех сферах бизнеса, таких как транспорт, металлургия или IT. Однако лидирующие позиции занимают телеком, финансовые институты, ритейл и нефтегазовая отрасль, в которых

⁴ <https://www2.deloitte.com/content/dam/Deloitte/de/Documents/WM%20Digitalisierung.pdf>

⁵ <https://www.business.com/articles/returning-customers-spend-67-more-than-new-customers-keep-your-customers-coming-back-with-a-recurring-revenue-sales-model/>

⁶ <https://www.gartner.com/smarterwithgartner/coronavirus-cio-areas-of-focus-during-the-covid-19-outbreak>

⁷ <https://markets.businessinsider.com/news/stocks/global-chatbot-market-anticipated-to-reach-9-4-billion-by-2024-robust-opportunities-to-arise-in-retail-ecommerce-1028759508#>

⁸ <https://www.gartner.com/smarterwithgartner/gartner-top-strategic-predictions-for-2018-and-beyond/>

больше половины компаний используют чат-боты, и наиболее широко — в области продаж и поддержке. Эффективность использования ботов в торговле трудно поставить под сомнение: в 2019 году боты увеличили продажи в среднем на 67%, а 35% руководителей бизнеса заявили, что чат-боты помогли им заключать сделки. Использование ботов для клиентской поддержки также показывает хорошие результаты: время отклика увеличилось в среднем в 3 раза, а удовлетворенность клиентов выросла на 24%.⁹

Сфера чат-ботов имеет дело с естественным языком, из-за чего роль зарубежных IT гигантов не столь высока, как в других технологических отраслях. Из-за различий в языках в работе с каждым из них нужен индивидуальный и точный подход. Кроме того, каждая страна имеет свои популярные сервисы и нишевые отрасли. Всё это открывает возможности для развития небольшим стартапам, которые могут конкурировать с IBM, Google и Amazon, превосходя их с точки зрения более глубокой проработки языковых и культурных особенностей. Также, стартапы более гибкие и позволяют найти решение под каждую сферу.

Классификация диалоговых систем

Подходы и технологии, используемые при создании бота, зависят от типа и назначения системы. Классификация диалоговых систем опирается на 2 критерия:

- Ориентированность на задачу (goal-oriented) VS общее назначение
- Закрытая VS открытая предметная область (домен)

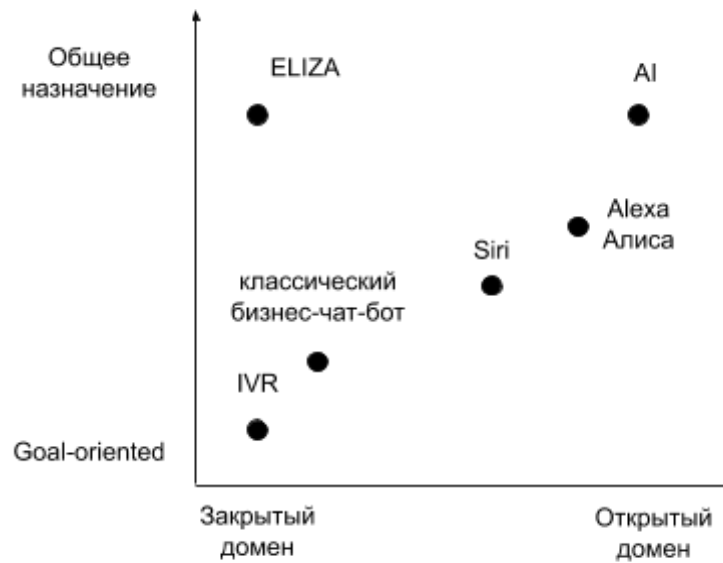
Goal-oriented боты разрабатываются для решения конкретных задач - выдачи информации по запросу, выполнения операций (проверка баланса, подключение тарифа). Диалоговые системы общего назначения (так называемые боты-"болталки") направлены исключительно на поддержание диалога и развлечение пользователя.

Системы с закрытым доменом выполняют задачи или ведут диалог, ограниченный узкой предметной областью, например банковской или медицинской сферой. Системы с открытым доменом предполагают универсальные возможности использования - к ним относятся ассистенты типа Алексы от Amazon или Алисы от Яндекс.

Сложность разработки чат-бота возрастает по мере расширения его предметной области и генерализации назначения (или увеличения количества задач, которые

⁹ <https://www.intercom.com/blog/the-state-of-chatbots/>

он способен выполнять). Вот пример классификации некоторых чат-ботов по данным критериям¹⁰:

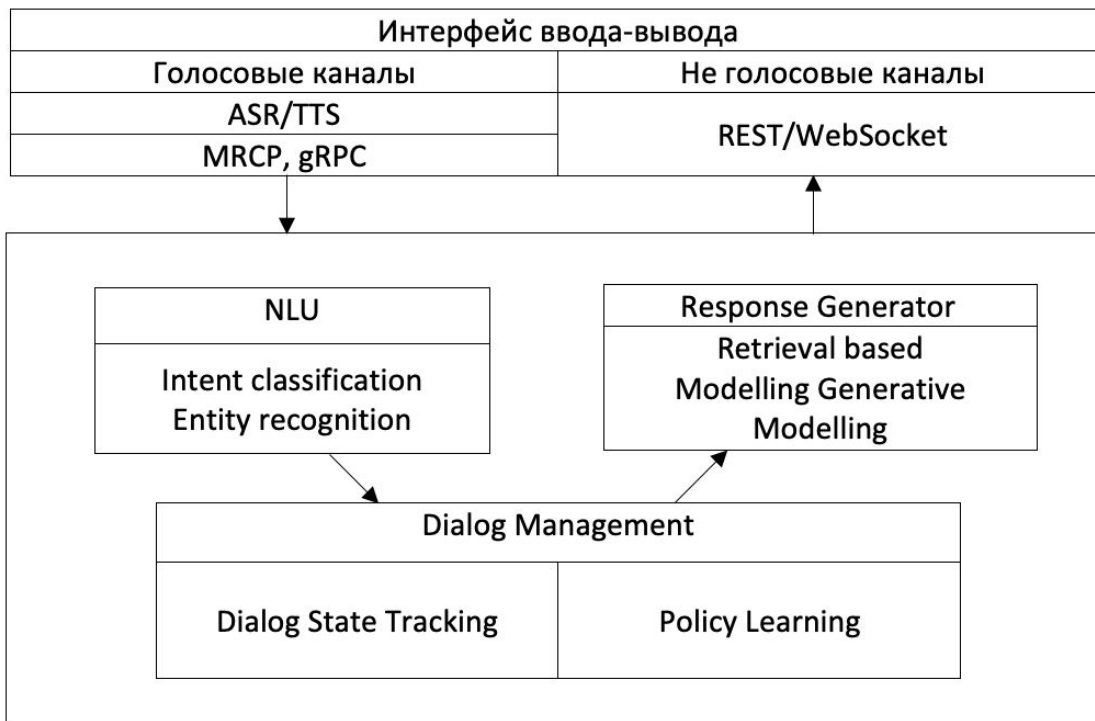


Для большинства бизнес-задач достаточно goal-oriented бота с ограниченной областью знаний; стремление охватить все предметные сферы и создать универсального собеседника в теории может привести к созданию “настоящего” искусственного интеллекта.

Технологии

Рассмотрим архитектуру классического бота:

¹⁰<https://medium.com/@madrugado/what-are-the-dialog-systems-or-something-about-eliza-9aefb551eaa>



Natural Language Understanding (NLU, NLP Engine) — один из ключевых модулей любой диалоговой системы. Он обеспечивает структуризацию и «понимание» (интерпретацию) сообщений пользователя. Существует несколько способов коммуникации с ботом: predetermined диалоговые кнопки (например, «Да» и «Нет»), ключевые слова («/start», «/bye» и т.д.) и общение в свободной форме с использованием произвольного текста на естественном языке. В общем случае свободу общения можно представить в виде следующего рисунка:



В современных системах наибольшую популярность получила свободная коммуникация в виде обмена сообщениями между клиентом и ботом. В связи с этим, основными задачами модуля NLU являются:

1) **Классификация сообщений** (Intent Classification). Она может производиться как с применением нейронных сетей (наиболее популярны

архитектуры на основе рекуррентных ячеек GRU/LSTM), так и классических алгоритмов машинного обучения (Naive Bayes, Logistic Regression), которые успешно применяются в случае небольшого количества обучающих данных. Любому сообщению пользователя должен соответствовать определенный класс (интент). Например, сообщения «привет», «хай», «добрый день» могут быть классифицированы в один класс intent_hello, а сообщения «до свидания», «пока», «удачи» — в класс intent_bye. Сообщения, класс которых с достаточной вероятностью определить не удалось, относятся к классу no_match. В общем и целом, все интенты разделяют на General intents (например, интенты приветствия и прощания) и Domain intents (например, интенты заказа конкретной пиццы или бронирования гостиницы на определенный день);

2) **Выделение именованных сущностей** (Named Entity Extraction, Named Entity Recognition, Slot Filling). Сущности представляют собой единый лексический элемент определенного типа (географическое название, имя, дата и т.п.). Так, в сообщении «Покажи расписание рейсов из Москвы в Париж 12 августа» именованными сущностями являются «Москва», «Париж» и «12 августа». Выделение сущностей может быть реализовано как на основе регулярных выражений (шаблонов, паттернов) и справочников, так и посредством NER Network — специально обученной нейронной сети с архитектурой many-to-many.

Dialog Management Module (DM) — модуль управления диалогом. Ядро чат-бота, которое управляет текущим контекстом и реализует требуемый функционал. Контекст представляет собой «память» бота обо всех или последних действиях и сообщениях. Не все диалоговые модули поддерживают контекст и не всегда в нем есть необходимость. Например, в простых вопросно-ответных системах нет необходимости поддерживать контекст. В случае примитивных диалогов используют следующие типы DM: switch statement (т.е. полное отсутствие контекста, аналогично поисковым системам) или FSM (finite-state machine) — конечный автомат, представляющий собой однонаправленный диалог без сохранения полученных от клиента данных для других диалогов в рамках той же сессии. В случае goal-oriented чат-бота использование контекста обязательно, и тогда DM включает в себя:

1) **Dialog State Tracking (DST)** — компонент, отслеживающий текущее состояние (state) диалоговой машины и допустимые переходы между стеями. Сохраняет и отслеживает все сообщения клиента и ответные действия системы, чтобы помочь Policy Learning-компоненту оптимально выбрать следующий шаг.

2) **Policy Learning** (Dialog Policy, PL, DP) — компонент, определяющий дальнейшие действия бота на основе текущего состояния и контекста. Вариантами последующих действий могут быть: формирование ответного сообщения, обращение во внешнюю информационную систему за дополнительной информацией, условный переход в другой стейт (context

switching) и т.п. Policy Learning может быть реализован на основе правил (rule based), на основе предобученной классической нейронной сети (supervised learning), которая сама определяет, что лучше делать дальше, либо с использованием подхода reinforcement learning. Первый вариант получил наибольшее распространение из-за простоты реализации и отсутствия необходимости обучения на большом корпусе диалогов.

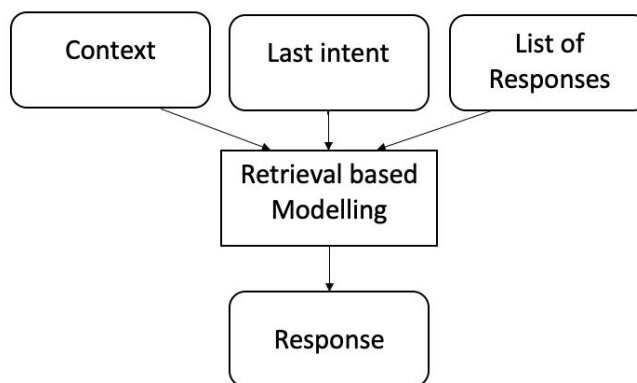
Response Generator Module (Natural Language Generator, NLG) — модуль генерации ответов. Модуль формирует ответ клиенту в соответствии с полученными из диалогового модуля и внешних информационных систем данными. В общем случае выделяют следующие архитектуры RGM:

1) **Retrieval Based Modelling** —

в простейшем случае ответом является статичный текст.

Для “оживления” сценария бот может случайным образом выбирать вариант из предопределенного набора статичных текстов (например, для прощания с пользователем всякий раз

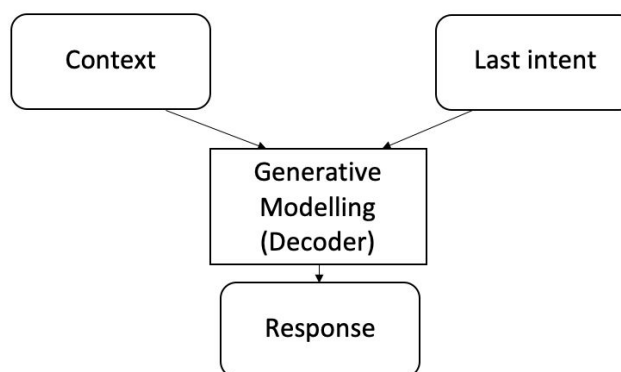
использовать случайный вариант из доступного списка: «пока», «приятно было пообщаться», «до свидания» и т.д.). Ответ может быть динамическим, т.е. включать в себя результаты, полученные из внешних систем: к примеру, значение баланса счета или время вылета самолета. Генерация ответа статичным или динамическим текстом являются самыми распространенными из-за простоты реализации, отсутствия необходимости обучения и предсказуемости результата.



2) **Generative Modelling** —

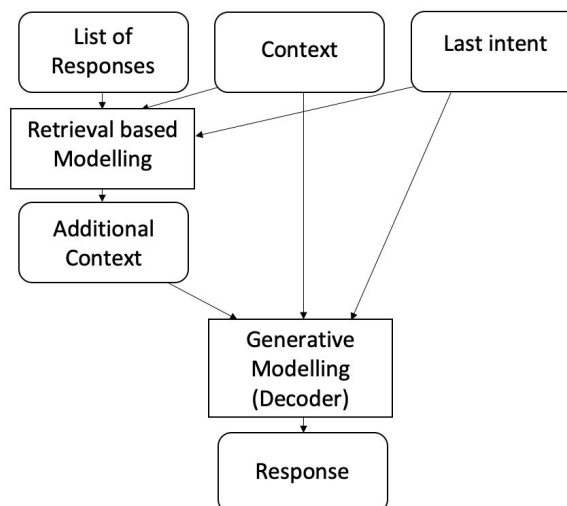
формирование ответа на основе работы нейронной сети (как правило, архитектура Encoder-Decoder). В таком случае содержание ответа полностью формируется предобученной нейронной сетью. Подход нашел свое отражение в

вопросно-ответных системах, но в классических чат-ботах применяется крайне редко из-за сложности обучения и риска неожиданных ответов (пример - известная ситуация с ботом Microsoft, который в результате



бесконтрольного самообучения стал генерировать расистские высказывания).

- 3) **Retrieve and Refine** - гибридный метод. Известно, что генеративные модели склонны давать общие, более “безопасные” ответы (“да”, “нет”, “не знаю”), в то время как Retrieval Based модели ограничены набором предзаписанных ответов, которые могут не полностью соответствовать контексту текущего диалога. Модель Retrieve and Refine комбинирует два описанных выше подхода. В гибридной модели Retrieve and Refine обнаруженные Retrieval Based моделью ответы подаются на вход генеративной модели в качестве дополнительного контекста. Генеративная модель формирует окончательный ответ бота.¹¹



Интерфейс ввода-вывода. Чат-боты могут взаимодействовать с клиентами в разных каналах связи. Это может быть традиционный виджет на сайте, мобильное приложение, мессенджер или телефонный канал. С точки зрения внутренней архитектуры, ядро бота не меняется, но для успешного взаимодействия ядра с внешним миром нужны соответствующие интерфейсы. Как правило, используют следующие дополнительные интерфейсные модули:

- для голосовых ботов — системы синтеза (TTS — Text to Speech) и распознавания (ASR — Automatic Speech Recognition) речи. Их качество достигло такого уровня, что при общении отличить живого человека от синтезированного голоса не всегда просто. Существует довольно много разработчиков этих решений в России (ЦРТ, Яндекс, АСМ-Решения) и в мире (IBM, Google, Microsoft, Nuance). Чтобы стандартизировать методы интеграции подобных систем, разработаны интерфейсы MRCP и gRPC.

- для не голосовых систем широко применяются уже зарекомендовавшие себя протоколы HTTP REST/WebSocket. Они гораздо проще в реализации, т.к. не требуют распознавания голоса в режиме он-лайн. Разработчики вправе как использовать общепринятые интерфейсы, так и создавать свои собственные, в случае, если новая разработка по каким-либо критериям лучше существующих.

¹¹ <https://www.aclweb.org/anthology/W18-5713.pdf>

Сквозное обучение

Ранее была рассмотрена классическая архитектура goal-oriented чат-бота и его ключевые компоненты. Как уже было отмечено, каждый из них может быть реализован как с применением нейронной сети, так и без нее. Классическая архитектура допускает также частичное использование нейронных сетей: например, компонент NLU может быть реализован нейросетью, а модули управления диалогом и генерации ответов - на основе правил (rule-based). Таким образом, можно отметить гибкость и адаптируемость архитектуры классического бота, в которой каждый компонент может оптимизироваться и дорабатываться отдельно от других.

Развитие нейронных сетей и глубокого обучения привело к появлению в последние 3-4 года новых подходов к реализации ботов:

- **End-to-end Learning** - в данном подходе сохраняется классическая архитектура бота, но все ее компоненты представлены нейронными сетями. Обучение производится сквозным методом, т.е. на каждом экземпляре обучающего диалога все компоненты обучаются как единое целое, а не по отдельности.
- **Fully data driven Conversation Models** - в данном подходе уже не применяется классическая архитектура. Весь бот представляет собой одну глубокую нейронную сеть. Такая реализация архитектуры целесообразна для ботов-"болталок" и не подходит для goal-oriented чат-ботов.

Стоит отметить, что указанные новые подходы еще не получили широкого практического применения и пока носят научно-исследовательский характер. В первую очередь это связано с тем, что для обучения подобных архитектур требуются огромные объемы тренировочных данных, количество которых исчисляется десятками миллионов и сотнями гигабайт диалогов (методы расширения обучающих данных (data augmentation) для текстов недостаточно эффективны и не решают проблему). Ни одна компания не располагает подобным корпусом, а исследователи обычно используют данные из Public Domain Social Media (таких как Reddit, Twitter и т.п.).

Во вторых, для обучения моделей на таких объемах данных требуются значительные вычислительные ресурсы (GPU, TPU). Например, для обучения open-domain чат-бота Meena, разработанного в Google Brain в 2020 году, потребовалось 30 дней и 2048 TPU.¹² Привлечь подобные ресурсы для обучения модели могут только компании масштаба Google.

¹² <https://arxiv.org/pdf/2001.09977.pdf>

В третьих, на сегодняшний день не существует устоявшихся архитектурных решений нейронных сетей, которые наилучшим образом подходили бы для решения таких задач.

Поиск подходов к реализации архитектур диалоговых систем продолжается, количество публикаций по данной тематике в последние несколько лет показывает, что интенсивные исследования ведутся как научными институтами (MIT, Stanford и др.), так и ведущими ИТ компаниями мира, и в будущем можно ожидать существенных прорывов. Однако на данный момент можно достаточно уверенно утверждать, что в ближайшие несколько лет открытий, которые приведут к коммерческим реализациям подобных решений, не предвидится.

Сферы применения

Наиболее распространенная сфера применения чат-ботов — это клиентский сервис (55% всех российских компаний, внедривших ботов, согласно отчету KPMG¹³). Сегодня боты способны самостоятельно обрабатывать до 80% обращений, что позволяет компаниям более чем на треть снизить затраты на соответствующие службы¹⁴. Однако оптимизация работы контактных центров — далеко не единственная их функция. Роботы внедряются во все сферы бизнеса, от HR до здравоохранения. Вот несколько менее очевидных вариантов использования чат-ботов:

Сфера HR:

- подбор персонала (опрос кандидата и назначение собеседования);
- адаптация новых сотрудников (помощь с документами, создание инструкций, проверка усвоенного материала);
- ответы на вопросы (сотрудники компании могут узнать про праздничные дни, текущие задачи и т.п.);
- подача заявлений (автоматическое заполнение документов в диалоговой форме);
- экзит-интервью (с ботом увольняющийся сотрудник может быть откровеннее);
- обучение сотрудников (разработка индивидуального плана обучения и развития сотрудника, своевременные советы курсов и литературы).

Банки, брокеры, коллекторские агентства:

- манипуляции со счётом (проверка баланса, перевод денег, контроль транзакций);
- персонализированный советчик в инвестициях;
- получение справок, документов без заполнения форм;

¹³ <https://home.kpmg/ru/ru/home/insights/2019/01/digital-technologies-in-russian-companies-survey.html>

¹⁴ <https://www.content-review.com/articles/46933/>

- управление кредитом, сбор информации по погашению задолженности.

Телеком:

- подбор и подключение тарифов и услуг
- консультация по условиям использования
- проверка баланса и пополнение счета
- помощь с настройкой оборудования (роутер, приставка, сат-модуль)

Продажи и маркетинг:

- консультация по ассортименту;
- персонализация предложений (подбор индивидуального маркетингового контента на основе данных о клиенте, его поведении на сайте и истории разговора);
- лидогенерация;

Туризм:

- подбор туров, отелей, авиабилетов;
- туристический гид (информация о стране, достопримечательностях, помощь в новом городе);
- экскурсии/квесты (бот проведет участников по маршруту, загадает загадки, проверит ответы и даст подсказку).

Здравоохранение:

- запись на прием (бот предложит клиенту свободные окна, вовремя напомнит о записи и перенесет ее в случае необходимости);
- индивидуальные рекомендации (на основе истории назначений бот напомнит о приеме лекарств, соблюдении диеты и необходимости повторного осмотра);
- консультант по здоровью (в любое время поможет рассчитать дозировку лекарств, допустимую физическую нагрузку, расскажет о противопоказаниях);
- ассистент врача (мгновенно предоставит всю информацию о препаратах, дозировках и современных медицинских стандартах).

Крупный бизнес

В начале 2018 года МТС запустила умного чат-бота в личных кабинетах абонентов на сайте, а позднее и в приложении. Спустя год эффективность бота достигла 75%, бот отвечает на 8 из 10 вопросов, поступающих в колл-центр МТС, без передачи звонка оператору. Робот МТС отвечает на вопросы, с которыми абоненты чаще всего обращаются в компанию, подключает и отключает услуги, при необходимости меняет тариф, предоставляет информацию о балансе клиента и детализацию счета. Чат-бот демонстрирует

более высокие показатели эффективности по ряду параметров по сравнению с традиционными каналами: средняя длительность ожидания начала разговора с клиентом не превышает 2 секунд вместо минуты в контактном центре, средняя продолжительность разговора до успешного решения проблемы – 2 минуты, что в четыре раза быстрее средней длительности разговора с оператором. Руководитель Центра искусственного интеллекта МТС Аркадий Сандлер заявляет: «Роботы уже способны освободить специалистов от рутинных операций, чтобы люди могли сосредоточиться на более сложной и творческой работе. Это позволит компаниям сократить в среднем на 40% затраты на соответствующие службы».

Малый бизнес

Вопреки распространенному мнению, чат-боты не являются привилегией крупного бизнеса. Разработка собственного высокоуровневого бота, как, скажем, Олег от Tinkoff, с нуля — масштабный и дорогостоящий проект. Но такой чат-бот нужен не всем. Для создания роботов, умеющих выполнять ограниченный набор задач, таких как подбор товара, проверка статуса заказа и обзвон клиентов, давно существуют готовые решения: SaaS или open source фреймворки с графическим интерфейсом, позволяющие построить простого бота, не имея специальных знаний в области программирования, или «коробочные» продукты, подстраиваемые под специфику заказчика. Их очевидные плюсы — доступная цена и скорость внедрения.

Внедрение чат-бота. С чего начать?

Прежде чем принимать решение о создании чат-бота, важно провести детальный анализ бизнес-процессов и получить четкое представление о ваших потребностях и ожидаемых результатах. На их основе в дальнейшем может быть сформировано ТЗ для внутреннего или аутсорсингового департамента разработки (в случае обращения к сторонней компании-разработчику, такой бизнес-анализ может входить в спектр оказываемых услуг). Вот основные пункты, определяющие требуемые характеристики чат-бота:

- Форма бизнеса: малый/средний/крупный
- Бизнес-цели: клиентская поддержка/удержание клиентов/рост продаж и пр.
- Бизнес-потребности, которые должен удовлетворять чат-бот: улучшение клиентского опыта/сбор данных/автоматизация процессов и пр.
- Характеристики целевой аудитории: возраст/род деятельности/интересы/посещаемые платформы/сайты и пр.
- Платформы и каналы взаимодействия с целевой аудиторией: текстовый/голосовой, мессенджеры/приложение/виджет на сайте/телефония

- Ожидаемая нагрузка: количество диалогов в день, количество параллельных сессий
- Функциональность бота: консультация/проверка баланса/перевод на нужную группу операторов и пр.
- Количество и глубина (сложность) сценариев диалогов чат-бота
- Необходимость интеграции с ИС и базами данных
- Информационная безопасность и возможность размещения в облаке/on-premise
- Наличие ИТ-компетенций внутри компании (собственные программисты/лингвисты/интеграторы)
- Допустимая стоимость внедрения чат-бота
- Допустимое время разработки и внедрения чат-бота

Понимание ответов на эти вопросы позволяет сформулировать ТЗ и выбрать один из следующих вариантов:

1) Купить готовое решение

Это вариант для малого и среднего бизнеса, имеющего ограниченные временные и финансовые ресурсы и не обладающего собственным отделом разработки из 3-х и более человек. Готовые боты реализует типовые сценарии взаимодействия, например оформление заказа, бронирование, HR-интервью.

Плюсы: скорость внедрения (может составлять 2-3 дня), доступная стоимость, качество продукта (над созданием бота работали профессионалы, часто это крупные фирмы “с именем”, занимающиеся NLP-технологиями), от заказчика требуется минимум компьютерных навыков, предоставляются готовые отлаженные механизмы интеграции с мессенджерами и CRM/ИС, доступ к техподдержке по подписке.

Минусы: мало возможностей индивидуализации и изменения сценария, на рынке представлен ограниченный ассортимент типовых сценариев, ограниченное количество предоставляемых интеграций (только с популярными CRM/ИС), часто нет возможности установки on-premise.

2) Создать собственное решение на базе готовой платформы для разработки чат-ботов

а) Сделать бота самостоятельно

Также подходит для малого и среднего бизнеса, но требует больших навыков программирования (иногда - существенных, платформы разнятся в этом отношении; несколько примеров таких платформ будут приведены ниже). Скорость внедрения полностью зависит от человеческих ресурсов и уровня компетенций в компании.

Плюсы: полная свобода в написании сценария в постоянном тесном контакте с бизнес-заказчиком, возможность реализации не типовых

сценариев и индивидуализации чат-бота, доступная стоимость, доступ к техподдержке по подписке.

Минусы: самостоятельно продумать структуру сценария, предусмотрев все возможные ответвления, реакции пользователя и подводные камни, без соответствующего опыта сложно. На отладку в процессе эксплуатации потребуется время и опыт, приобретенный ценой недовольства и раздражения клиентов. Также часто платформы предоставляют готовые интеграции только с несколькими каналами связи, для настройки других потребуется привлечение дополнительных ИТ-ресурсов.

б) Заказать разработку профильной компании

Исходные требования могут варьироваться от небольшого бота с узким функционалом (к примеру, оформление заказов для интернет-магазина или обзвон клиентов с маркетинговым предложением) до масштабного чат-бота с продуманной личностью, разветвленным сценарием на сотни интенгов, выполняющего большой спектр задач и выдерживающего ежедневную нагрузку в тысячи обращений.

Плюсы: высокое качество разработки, полная индивидуализация и персонификация бота, возможность редактировать и дополнять сценарий, интеграции с любыми необходимыми системами, размещение on-premise с возможностью выполнить все требования по безопасности данных.

Минусы: серьезные временные (от нескольких недель до нескольких месяцев) и финансовые вложения, необходимость наличия внутренних специалистов для составления технического задания и для грамотного выбора компании-интегратора (разработчика).

3) Создать собственную чат-бот-платформу и разработать бота “с нуля”

Вариант имеет смысл только в том случае, если компания планирует в дальнейшем распространять свое решение в качестве готовой платформы и/или заниматься разработкой чат-ботов на заказ (см. пункт 2). В зависимости от наличия внутри компании специалистов по технологиям NLU, ASR/TTS и др. такой проект займет от полугода до нескольких лет.

Плюсы: возможность продемонстрировать качество технологии на собственном чат-боте будет значимым конкурентным преимуществом.

Минусы: использование для разработки только своего бота нерентабельно, требуется команда разработчиков от 5 и более человек с глубоким знанием процессов обработки текста средствами нейронных сетей, необходимо наладить бизнес-процессы разработки программного обеспечения.

Вне зависимости от исполнителя, платформы и технологий, которые были выбраны для создания чат-бота, на введении его в эксплуатацию работа не

заканчивается. Для получения результата необходимо постоянно совершенствовать бота, анализируя диалоги с пользователями, расширять функциональность и внедрять новые технологии.

Как оценить качество чат-бота?

Методы оценки качества работы чат-бота (его KPI) можно разделить на две категории:

1. Метрики бизнес-показателей
2. Метрики технических показателей

К **бизнес-метрикам** относятся следующие измерения (в зависимости от функционала бота):

- **Уровень снижения загрузки колл-центра.** На сегодняшний день достижим уровень в 70 - 80%
- **Конверсия в покупки.** Актуально для чат-ботов сегмента e-commerce.
- **Процент вернувшихся клиентов.** После добавления в мессенджер пользователя чат-бот может присылать ему маркетинговый контент и напоминать о неоплаченном товаре в корзине. Успешность ремаркетинга можно оценить, добавляя в сообщения UTM-метки.
- **Количество пользователей**
 - **Вовлеченные пользователи** - все пользователи, хотя бы раз вступившие в контакт с ботом (написавшие сообщение). Вовлеченность бота измеряется в процентах от общего числа людей, добавивших бота в друзья, и оценивается в сравнении с другими каналами (email, соцсети).
 - **Активные пользователи** - написавшие боту самостоятельно, а не в ответ на его сообщение. Рост процента активных пользователей означает полезность сервиса и успешность его рекламы (коррелирует с увеличением ROMI).
 - **Повторные пользователи** - ежемесячный рост числа клиентов, использующих бота, особенно вернувшихся пользователей, означает удовлетворенность клиентов новым каналом коммуникации.
- **Читаемость сообщений и CTR (Click-Through Rate).** Эта метрика особенно актуальна при использовании бота для веерной рассылки. Читаемость сообщений бота составляет 70-80%, а CTR (количество

переходов по ссылкам) - 15-60%¹⁵ против читаемости email-рассылок 11-25% и CTR 4-8%¹⁶.

- **NPS (Net Promoter Score)** - индекс потребительской лояльности по отношению к продукту и компании. В конце диалога пользователю предлагается оценить разговор по шкале и/или выбрать один из вариантов ответа (“Вопрос решен”, “Ответ не решил проблему”, “Я спрашивал о другом” и пр). Значение NPS рассчитывается на основе ответов по формуле:

$$NPS = (\text{кол-во оценок от 9 до 10} - \text{кол-во оценок от 0 до 6}) / \text{общее кол-во оценок} * 100$$

Нормы значений NPS различаются по отраслям. В целом нормальным считается NPS > 0, NPS > 30 - хороший показатель, NPS > 50 - отличный.

Технические метрики:

- **Средняя продолжительность сессии.** В зависимости от функционала бота о его эффективности свидетельствует большая или меньшая длительность диалога. В клиентской поддержке время решения запроса ботом должно быть меньше аналогичного для оператора-человека. Бот-представитель бренда или консультант онлайн-магазина, напротив, призван дольше удерживать внимание пользователя.
- **GCR (Goal Completion Rate).** Рассчитывается по формуле:
$$GCR = \text{кол-во диалогов, приведших к достижению цели} / \text{общее кол-во диалогов} * 100,$$
где цель - удовлетворение запроса пользователя (ответ на вопрос, выполнение операции) либо решение бизнес-задачи (подписка, продажа, получение контактов).
- **Процент нераспознанных реплик.** Вопросы, на которые бот еще “не знает” ответ, классифицируются как no match. Сокращение доли нераспознанных реплик свидетельствует об успешном обучении бота.
- **Качество классификации сообщений.** Для оценки работы ядра NLU применяются традиционные метрики, самые распространенные - точность (precision), полнота (recall) и F-мера. Для подсчета значений проводится разметка диалогов и каждая реплика пользователя получает одну из меток:

TP — true positive, бот верно отнёс реплику к рассматриваемому интену.

TN — true negative, бот верно полагает, что реплика не относится к рассматриваемому интену.

¹⁵<https://chatbotslife.com/chatbots-vs-email-when-to-use-them-as-marketing-support-channels-3a9e5567dc20>

¹⁶<https://www.smartinsights.com/email-marketing/email-communications-strategy/statistics-sources-for-email-marketing/>

FP — false positive, бот неверно отнёс реплику к рассматриваемому интену.

FN — false negative, бот неверно полагает, что реплика не относится к рассматриваемому интену.

Значения точности и полноты классификации вычисляются по формулам:

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN}$$

Поскольку повышение точности классификации автоматически ведет к снижению полноты и наоборот, используется третья метрика - F-мера - гармоническое среднее между полнотой и точностью:

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

F-мера достигает максимума при полноте и точности, равными единице, и близка к нулю, если один из аргументов близок к нулю. Хорошим показателем F-меры для систем NLU считается 0.95 и выше.

- **Качество выделения сущностей.** Аналогичные метрики - precision, recall, F-measure - применяются для оценки качества извлечения данных из сообщений пользователя (имен, географических наименований, адресов, телефонов и пр.)

Помимо оценки качества работы самого чат-бота, анализ пользовательских диалогов позволяет получить полезную информацию, которую можно использовать в развитии маркетинговых стратегий и самого продукта:

- **Часто открываемые ссылки.** Наиболее часто посещаемые страницы из предложенных чат-ботом отражают интересы и предпочтения клиентов. Эта информация позволяет изучать и формировать спрос пользователей на конкретные услуги или товары и планировать маркетинговые кампании.
- **Темы диалогов.** Анализ количественного распределения сообщений по тематикам позволяет отслеживать наиболее востребованные темы и совершенствовать их. Например, размещать на сайте более развернутые описания сервисов, вызывающих много вопросов, или изменить интерфейс приложения, чтобы было легче найти нужный раздел.

Платформы для разработки чат-ботов

С ростом популярности ботов появляется всё больше платформ, которые позволяют создать голосового и/или текстового бота. Решения различаются степенью свободы, которую имеет разработчик. Самые популярные и крупные из платформ на международном и российском рынках:

- Dialogflow — облачный сервис распознавания естественного языка от Google, который поддерживает различные языки, в том числе и русский. Имеет визуальный редактор, что дает возможность создавать простых ботов без навыков программирования (для более сложных ботов потребуется база написания и редактирования кодов на любом распространенном языке программирования). Интегрируется со всеми популярными мессенджерами. Имеется возможность синтеза и распознавания речи. Есть бесплатный тариф с лимитом 180 текстовых запросов в минуту. Такое решение отлично подойдет для малого бизнеса.
- Azure Bot Service — облачное решение от Microsoft как для текстовых, так и для голосовых ботов, поддерживает множество языков. Для разработки обязательно уверенное владение языками программирования C# или JavaScript. Платная недорогая подписка, тарифы зависят от количества сообщений. Интегрируется со всеми мессенджерами.
- IBM Watson — открытая облачная платформа для создания ботов с графическим интерфейсом. Позволяет создать простого бота без знаний программирования. Поддерживает 13 самых распространенных языков, включая русский. Можно пользоваться бесплатно с ограниченным набором функций и с лимитом в количестве обращений.
- ChatNavigator — омниканальная платформа для создания голосовых и текстовых ботов одного из лидеров рынка, российской компании ЦРТ. Специализируется на русском, английском и казахском языках. В зависимости от задачи, используется система на правилах, нейронная или гибридная сеть. Платформа позволяет создать сложных и многоуровневых ботов, имеет собственную технологию синтеза, распознавания и биометрии речи. Такое решение отлично подойдет крупному и среднему бизнесу.
- JAICP — платформа для разработки голосовых ассистентов и чат-ботов с искусственным интеллектом российской компании Just AI. Комплексные сложные боты создаются на основе правил и машинного обучения. Имеет собственные ASR и TTS технологии, интегрируется со всеми популярными мессенджерами и Алисой. Отличное решение для крупного бизнеса.
- Aimylogic — еще одна платформа компании Just AI с простым графическим интерфейсом и открытым доступом. Позволяет создать простого бота без навыков программирования. Бюджетные тарифы, отлично подходит для малого бизнеса.
- NLab — платформа российской компании Наносемантика, включающая среду разработки, словари, диалоговый процессор и язык лингвистического программирования Dialog Language. Работает с

собственной технологией распознавания речи, а также поддерживает технологии ASR/TTS от Yandex, ЦРТ, Google. Поставляется по модели SaaS или On-premise.

- DeepPavlov — open source фреймворк от лаборатории нейронных сетей и глубокого обучения МФТИ, построенный на базе библиотек TensorFlow и Keras с использованием классификатора Bert. Позволяет строить диалоговые системы с использованием предобученных NLP-компонентов (named entity recognition, intent classification и др.), а также обучать и тестировать собственные модели. Инструмент для разработчиков, не имеет графического интерфейса.

На всех платформах есть инструменты мониторинга таких метрик как динамика количества пользователей и обращений, частые запросы и ключевые слова, инструменты просмотра неудачных коммуникаций. Кроме того, присутствует система накопления диалогов, которая помогает улучшить сценарий.

Интеграционные платформы

Большинство используемых сегодня чат-ботов призваны решить конкретную задачу клиента или компании. В большинстве случаев для этого требуется интеграция бота с информационными системами (ИС) заказчика. Чаще всего на момент принятия решения о внедрении чат-бота у заказчика уже существуют внутренние ИС со своими программными интерфейсами (API) и требованиями информационной безопасности к передаче данных и интеграции. В этом случае интеграция осуществляется с использованием текущих интерфейсов либо посредством доработки/разработки с нуля недостающих интерфейсов самим заказчиком или подрядной организацией, сопровождающей конкретную систему.

В тоже время, с развитием облачных технологий многие компании переходят к использованию облачных ресурсов для размещения своих систем и/или меняют политики безопасности так, чтобы они позволяли облачным сервисам получать доступ к внутренним ИС. В этом случае процесс интеграции бота с ИС значительно упрощается. Существует значительное количество облачных сервисов, предоставляющих широкие средства интеграции со сторонними популярными решениями (такими как Google Sheets, облачные CRM системы, базы данных PostgreSQL/Oracle/MySQL и т.п., сервисы OpenWeatherMap и многие другие), а также предлагающих свои собственные сервисы (отправка SMS или push-уведомлений, инициация звонков и пр.).

Среди наиболее популярных облачных сервисов интеграции стоит отметить следующие:

- IntegroMat
- Zapier
- IFTTT (If This Then That)

Использование облачных интеграционных платформ позволяет не только решить вопрос быстрой разработки и интеграции бота в случае, если заказчик широко использует облачные технологии, но и быстро реализовать пилот или тестовый стенд будущего on-premise решения.

Как правило, разработка бота осуществляется внешней компанией-исполнителем, которая не всегда имеет доступ к внутренним ИС заказчика и не может оперативно вносить в них изменения для тестирования поведения бота. В этом случае быстрая разработка эмулятора интерфейса ИС заказчика средствами облачных интеграционных решений является гибким инструментом разработки и тестирования бота. Таким образом, использование облачных интеграционных платформ положительно сказывается одновременно на сроках, стоимости и качестве разработки чат-бота.

Подводные камни

Несмотря на все преимущества использования чат-ботов, они пока не в состоянии полностью заменить людей. Согласно исследованию Userlike, для 77% пользователей перевод запроса на человека-оператора является наиболее важной опцией при общении с ботом¹⁷. Одна из причин — ограниченная область знаний, бот не умеет импровизировать или сам создавать новые решения. В этом случае для повышения эффективности следует тщательно продумать возможные запросы пользователей на этапе пилотного проекта, а также постоянно обновлять базу знаний действующего бота.

Кроме того, проблемы могут возникнуть при обработке сложных реплик, которые содержат в себе сразу несколько запросов. Чтобы избежать непонимания, следует предусмотреть возможность использования нескольких запросов в одной реплике, а затем предложить клиенту их последовательное решение.

Другой проблемой во взаимодействии бота с человеком может стать «плохая память», другими словами, отсутствие контекста. Чат-боты не всегда могут запомнить уже состоявшийся разговор или отреагировать с учетом предыдущих реплик, что заставляет пользователя писать одно и то же несколько раз. Решением проблемы может стать сбор аналитики, позволяющей увидеть общие закономерности в диалоге и подготовить бота к ситуациям, разрешаемым в зависимости от контекста.

Боты не могут улавливать настроение пользователя. Люди чувствительны к различным типам эмоций в разговоре, к тому, как они передаются через тон и изменения формулировок. При возникновении проблемы клиенты хотят, чтобы

¹⁷ <https://www.userlike.com/en/blog/consumer-chatbot-perceptions>

их разочарования были встречены с сочувствием, а его отсутствие воспринимают негативно. Решение проблемы лежит в области сентимент-анализа (или анализа тональности), который занимается определением эмоциональной оценки авторов в тексте. Эта технология до сих пор не идеальна, однако быстро развивается и все чаще используется в современных ботах.

Разработка и интеграция бота — длительный процесс: от задумки до полного внедрения может пройти от месяца до года в зависимости от требований и сложности исполнения. Разработка включает в себя не только формирование и написание сценария, но и его тестирование с последующей коррекцией и повторными обучениями моделей. Кроме того, создание сложного бота требует высокого уровня компетенций у исполнителя, поэтому выбор правильного партнера также займет время.

Будущее

Боты вытеснят приложения?

С ростом рынка чат-ботов виртуальные помощники постоянно совершенствуются, становятся всё более надежными и человечными. Исследование Twilio показало, что около 66% клиентов предпочитают получать информацию через текстовое общение с представителями бренда, а не через специальные приложения и сайты¹⁸. Вполне вероятно, что скоро боты заменят существенную долю приложений и вот несколько тому причин:

Приложение	Чат-бот
Нужно специально скачивать, оно занимает место и при этом не используются часто	Уже установлен внутри всех популярных мессенджеров, надо только знать название бота и начать с ним диалог
Нужно искать информацию самостоятельно	Быстро и четко даст ответ на поставленный вопрос
Долгая и дорогая разработка	В зависимости от задачи, разработка бота может быть дешевле быстрее

¹⁸

<https://www.twilio.com/press/releases/twilio-study-shows-consumers-arent-paying-attention-to-brands-social-media-or-mobile-apps-they-prefer-email-or-text-instead>

Универсально для всех пользователей	Может быть персонализировано за счет полученных в ходе общения знаний о пользователе
Давно на рынке, приелось пользователям	Пользователь проявляет больше интереса, как к новой, еще не приевшейся технологии

Будет неправильно утверждать, что в будущем чат-боты полностью вытеснят приложения, однако для многих бизнесов бот может стать хорошей и выгодной альтернативой.

Человечность

Голосовые боты будущего будут лучше понимать интонации, настроение человека и выстраивать тактику ведения диалога с учётом состояния собеседника. Этот навык не только существенно повысит качество беседы, но и уменьшит затраты на звонок, т.к. технология распознавания эмоций сможет автоматизировать сбор фидбека от общения. Бот будет сам понимать, когда клиент остался доволен обращением, а когда его задача не была решена.

Встроенные технологии

Уже сегодня в чат-боты встраивают всё больше функций и технологий. Например, компания Nike одна из первых стала использовать технологию дополненной реальности (AR) в своем боте на платформе Facebook: в 2018 году перед выпуском лимитированных пар кроссовок компания создала квест через чат-бота, в финале которого пользователю открывалась 3D копия новых кроссовок с функцией AR, которые он мог внимательно рассмотреть и после этого купить. В будущем всё больше технологий будут встроены в чат-ботов, сильнее всего пользователи ждут внедрение автоматических платежей, что существенно ускорит и упростит процесс покупки.

Заключение

Благодаря развитию технологий NLU общение с ботами становится естественным и комфортным, и все больше людей предпочитают переписку с ботом другим способам взаимодействия: по данным Forbes¹⁹, их доля в 2019 г. увеличилась в 2 раза по сравнению с 2018 г. Во всем мире компании автоматизируют коммуникационные процессы, и люди привыкают мгновенно

¹⁹<https://www.forbes.com/sites/gilpress/2019/09/25/ai-stats-news-chatbots-lead-to-80-sales-decline-satisfied-customers-and-fewer-employees/#3011b6d848e0>

получать ответы на возникающие вопросы онлайн. В таких условиях чат-бот — это не веяние моды, а необходимость для тех, кто хочет соответствовать ожиданиям своих клиентов.

СПИСОК ИСТОЧНИКОВ

Malykh V. (2017) *What are the Dialog Systems, or Something about Eliza* // medium.com
(<https://medium.com/@madrugado/what-are-the-dialog-systems-or-something-about-eliza-9aefb551eaaa>)

Лемешковская Н. (2018) *Возможно ли измерить KPI чат-бота?* // Zen.Yandex
(<https://zen.yandex.ru/media/e-xecutive.ru/vozmojno-li-izmerit-kpi-chatbota-5c18c8070c4c2700ab6e0d09>)

Mansuri Sh. (2018) *How much time goes into developing a Messenger chatbot* // chatbotsmagazine.com
(<https://chatbotsmagazine.com/how-much-time-goes-into-developing-a-messenger-chatbot-ce86afdfb969>)

Зубкова А. (2018) *KPI для робота. Как измерить эффективность чат-ботов* // sostav.ru
(<https://www.sostav.ru/publication/kpi-dlya-robota-kak-izmerit-effektivnost-chat-botov-31678.html>)

КПМГ (2019) *Цифровые технологии в российских компаниях* // KPMG
(<https://assets.kpmg/content/dam/kpmg/ru/pdf/2019/01/ru-ru-digital-technologies-in-russian-companies.pdf>)

YIN S. (2019) *Where chatbots are headed in 2020* // intercom.com
(<https://www.intercom.com/blog/the-state-of-chatbots/>)

Schmidt D. Majonek A. Moulliet D. Grimm F. (2017) *Wealth Management Digitalization changes client advisory more than ever before* // Deloitte
(<https://www2.deloitte.com/content/dam/Deloitte/de/Documents/WM%20Digitalisierung.pdf>)

Clark M. (2013) *Returning Customers Spend 67% More Than New Customers - Keep Your Customers Coming Back with a Recurring Revenue Sales Model* // business.com
(<https://www.business.com/articles/returning-customers-spend-67-more-than-new-customers-keep-your-customers-coming-back-with-a-recurring-revenue-sales-model/>)

Starita L. (2020) *Coronavirus: CIO Areas of Focus During the COVID-19 Outbreak* // Gartner
(<https://www.gartner.com/smarterwithgartner/coronavirus-cio-areas-of-focus-during-the-covid-19-outbreak>)

Wood L. (2019) *Global Chatbot Market Anticipated to Reach \$9.4 Billion by 2024 - Robust Opportunities to Arise in Retail & eCommerce* // [markets.businessinsider.com](https://markets.businessinsider.com/news/stocks/global-chatbot-market-anticipated-to-reach-9-4-billion-by-2024-robust-opportunities-to-arise-in-retail-ecommerce-1028759508#)
(<https://markets.businessinsider.com/news/stocks/global-chatbot-market-anticipated-to-reach-9-4-billion-by-2024-robust-opportunities-to-arise-in-retail-ecommerce-1028759508#>)

Panetta K. (2017) *Gartner Top Strategic Predictions for 2018 and Beyond* // Gartner
(<https://www.gartner.com/smarterwithgartner/gartner-top-strategic-predictions-for-2018-and-beyond/>)

Juniper Research (2017) *Chatbot Conversations to deliver \$8 billion in Cost savings by 2022* // Juniper Research
(<https://www.juniperresearch.com/new-trending/analystxpress/july-2017/chatbot-conversations-to-deliver-8bn-cost-saving>)

Weston J., Dinan E., Miller A. H. (2018) *Retrieve and Refine: Improved Sequence Generation Models For Dialogue* // Association for Computational Linguistics
(<https://www.aclweb.org/anthology/W18-5713.pdf>)

Adiwardana D., Luong M.-T., David R. So, Hall J., Fiedel N., Thoppilan R., Yang Z., Kulshreshtha A., Nemade G., Lu Y., Quoc V. Le (2020) *Towards a Human-like Open-Domain Chatbot* // Semantic Scholar
(<https://arxiv.org/pdf/2001.09977.pdf>)

Micro Startups (2018) *Chatbots Vs. Email: When To Use Them As Marketing & Support Channels* // chatbotslife.com
(<https://chatbotslife.com/chatbots-vs-email-when-to-use-them-as-marketing-support-channels-3a9e5567dc20>)

Chaffey D. (2020) *How do you compare? 2020 email marketing statistics compilation* // Smart Insights
(<https://www.smartinsights.com/email-marketing/email-communications-strategy/statistics-sources-for-email-marketing/>)

Leah (2020) *What Do Your Customers Actually Think About Chatbots?* // Userlike
(<https://www.userlike.com/en/blog/consumer-chatbot-perceptions>)

Twilio (2019) *Twilio Study Shows Consumers Aren't Paying Attention to Brands' Social Media or Mobile Apps — They Prefer Email or Text Instead* // Twilio

(<https://www.twilio.com/press/releases/twilio-study-shows-consumers-arent-paying-attention-to-brands-social-media-or-mobile-apps-they-prefer-email-or-text-instead>)

Press G. (2019) *AI Stats News: Chatbots Lead To 80% Sales Decline, Satisfied Customers And Fewer Employees* // Forbs

(<https://www.forbes.com/sites/gilpress/2019/09/25/ai-stats-news-chatbots-lead-to-80-sales-decline-satisfied-customers-and-fewer-employees/#77c1dc6748e0>)