123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

Техническая спецификация

Содержание

Введение 3
Список функциональных возможностей 4
Особенности и преимущества 5
Техническая спецификация 6
Высокоуровневое проектирование в локальном варианте 7
Высокоуровневое проектирование в облачном варианте 8
Сеть заказчика 8
НЕЙРОСЕТЬ Contact Center AI 9
Центральный блок (Core Unit) 9
Блок пулов (Pool Unit) 10
NLU Engine (классификатор намерений) 11
NLU 12
Уникальные особенности NLU 14
Сервисы по обработке речи 15
Speech-to-text 15
Повышение точности 15
Text-to-Speech 16
Управление потоками звонка в Neuro Media Server 16
Требования к аппаратному обеспечению для локальной установки 17
Масштабируемость 19
Обработка персональных данных 20
Виды персональных данных, обрабатываемых Нейросеть Contact Center AI 20
Структура хранения и обработки данных 21
Доступ к персональным данным, обрабатываемым в рамках платформы Contact Center AI 22
Хранение персональных данных 23
Безопасность персональных данных в рамках платформы Contact Center AI 23
Меры контроля персональных данных 24
Сбор персональных данных 24
Шифрование 24
Управление доступом 24

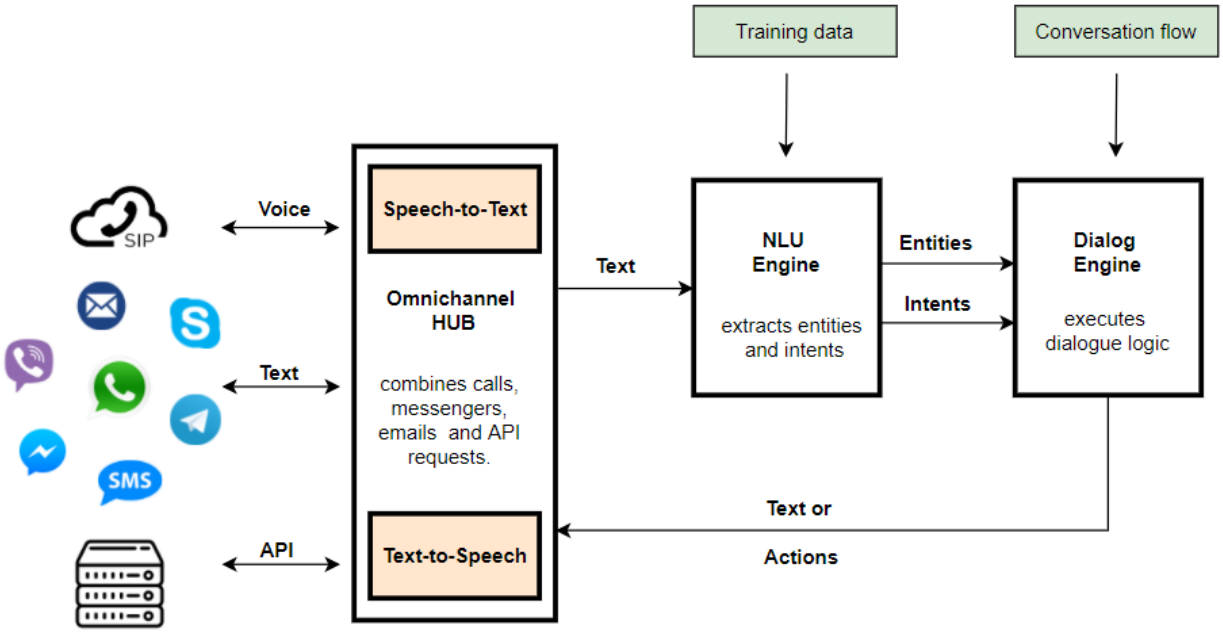
123557, город Москва, вн.тер. г. муниципальный Округ Пресненский, ул. Малая Грузинская, дом 28, помещение 6/14

Журналы аудита 25
Индивидуальные настройки клиента 25
Услуги по развертыванию платформы Contact Center AI **Error! Bookmark not defined.**
Установка 26
Интеграция 26
Ввод в эксплуатацию 26
Разработка виртуальных агентов 26
Проверка проектных решений 26
UAT 27
Условия технической поддержки 28

123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

Введение

Компания «Нейросеть» занимается разработкой, обучением и запуском виртуальных агентов в рамках своего решения Contact Center AI - полностью автономного контакт-центра на базе искусственного интеллекта. Контакт-центр работает в режиме омниканальности на основе собственной технологии и интегрирован с различными голосовыми решениями. Решение способно полностью заменить сотрудников контакт-центра виртуальными агентами. Оно может быть интегрировано с телефонной сетью, мессенджерами и системами поддержки бизнес-деятельности (англ. "Business Support System, BSS"). Система взаимодействует с клиентами и ведет диалог в соответствии с настройками потока разговора, не затрачивая человеческие ресурсы.



Решение Contact Center AI компании «Нейросеть» включает в себя следующие компоненты:

Omnichannel HUB объединяет звонки, мессенджеры и электронные письма в единую бесшовную коммуникационную среду, поддерживает запросы API для интеграции со сторонними приложениями и работает с минимальной задержкой, что имеет решающее значение в сфере голосовой связи.

NLU Engine обрабатывает естественный язык для извлечения сущностей и классификации намерений пользователя, работает с небольшими наборами данных и демонстрирует высокую точность (95-98%). Он разработан на базе архитектуры преобразователя DistilBERT с предварительно обученными моделями для более чем 100 языков.

123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

Dialog Engine выполняет логику диалога и поток разговора, поддерживает интерфейс drag-and-drop, язык программирования Python и нелинейные алгоритмы для выстраивания потока разговора.

Список функциональных возможностей

- Обработка входящих и исходящих звонков.
- Возможность переключения между языками (например, виртуальный агент начинает вести разговор на английском языке, а затем определяет, что клиент говорит на испанском языке, в данном случае, агент сразу же переключится на испанский).
- Возможность распознать автоответчик или голосовую почту.
- Возможность настроить обратные вызовы абоненту в течение указанного периода времени и с конфигурабельным количеством попыток вызова.
- Возможность отклика с малой задержкой (700 мс).
- Возможность записи аудио и транскрибации всех разговоров, а также возможность их отправки в соответствующее хранилище.
- Возможность хранения записи звонков.
- Возможность подготовить статистические отчеты по всем звонкам в формате *.xls или *.csv и отправить их согласованным способом в указанное время.
- Возможность определить просьбу абонента перезвонить в определенное время.
- Возможность сохранить историю абонента и при обратном звонке начать диалог с ним с точки, на которой прервался предыдущий звонок.
- Возможность перебивать собеседника (речь в целом или конкретные фразы).
- Возможность перевести вызов на оператора, если эта опция предусмотрена в настройках потока разговора.
- Возможность определить пол абонента по голосовой выборке.
- Встроенный визуальный конструктор разговоров на основе интерфейса drag-and-drop.
- Встроенный конструктор разговоров на основе языка программирования Python.
- Разделение логики диалога на логику в режиме реального времени (поток разговора во время звонка) и фоновую логику (управление попытками дозвониться, обмен сообщениями и данными).
- Поддержка нелинейной логики с возможностью переключения между единицами диалога без потери контекста.
- Использование естественного фонового шума во время звонка.
- Использование подхода активного слушания во время звонка.
- Двухуровневый NLU (понимание естественных языков) Engine, позволяющий обрабатывать текст на естественном языке с целью извлечения сущностей и классификации намерений пользователя с использованием алгоритмов

распознавания на первом уровне NLU и алгоритмов глубокого обучения на втором уровне NLU.

- Возможность отправки и приема SMS.
- Возможность отправки и приема электронной почты.
- Интеграция с WhatsApp.
- Интеграция с Viber.
- Интеграция с Telegram.
- Интеграция с FB Messenger.
- Интерфейс прикладного программирования (API) для интеграции по HTTP-протоколу.
- Выбор тона голоса и пола.
- Распознавание дат и дней недели.
- Автоматическая обработка адреса.
- Автоматическая обработка имени.
- Возможность логирования, позволяющая быстро и четко обнаруживать возникающие проблемы.
- Круглосуточные работа и техническая поддержка.
- Интерфейс управления и мониторинга.

Особенности и преимущества

- Микросервисная архитектура.
- Локальное или облачное (одно- или многопользовательское) развертывание.
- Разработанный компанией Media Server обеспечивает масштабируемость, надежность и широкий набор функций. Разработанная компанией платформа обработки голоса/вызовов (с использованием FreeSwitch только в качестве SBC).
- СУБД PostgreSQL v.11, нереляционная СУБД MongoDB v.4.2, файловое хранилище S3.
- Разработанный компанией NLU Engine поддерживает собственное распознавание паттернов в качестве решения для малых данных и глубокого обучения (архитектура преобразователя DistilBERT).
- Разработанный компанией Biometric Engine (альфа-версия).
- Разработанный компанией Logic Executor поддерживает логику в режиме реального времени (поток разговора во время звонка) и фоновую логику (управление попытками дозвониться, обмен сообщениями и данными) с возможностью хранения истории беседы всех попыток по всем каналам.
- Разработанный компанией Conversation Builder (конструктор разговоров) поддерживает визуальный интерфейс на основе интерфейса drag-and-drop и язык программирования Python.
- Разработанный компанией Speech-to-Text Engine.
- Поддержка сторонних решений для преобразования речи в текст и преобразования текста в речь, таких как Microsoft, Google, IBM, Yandex.

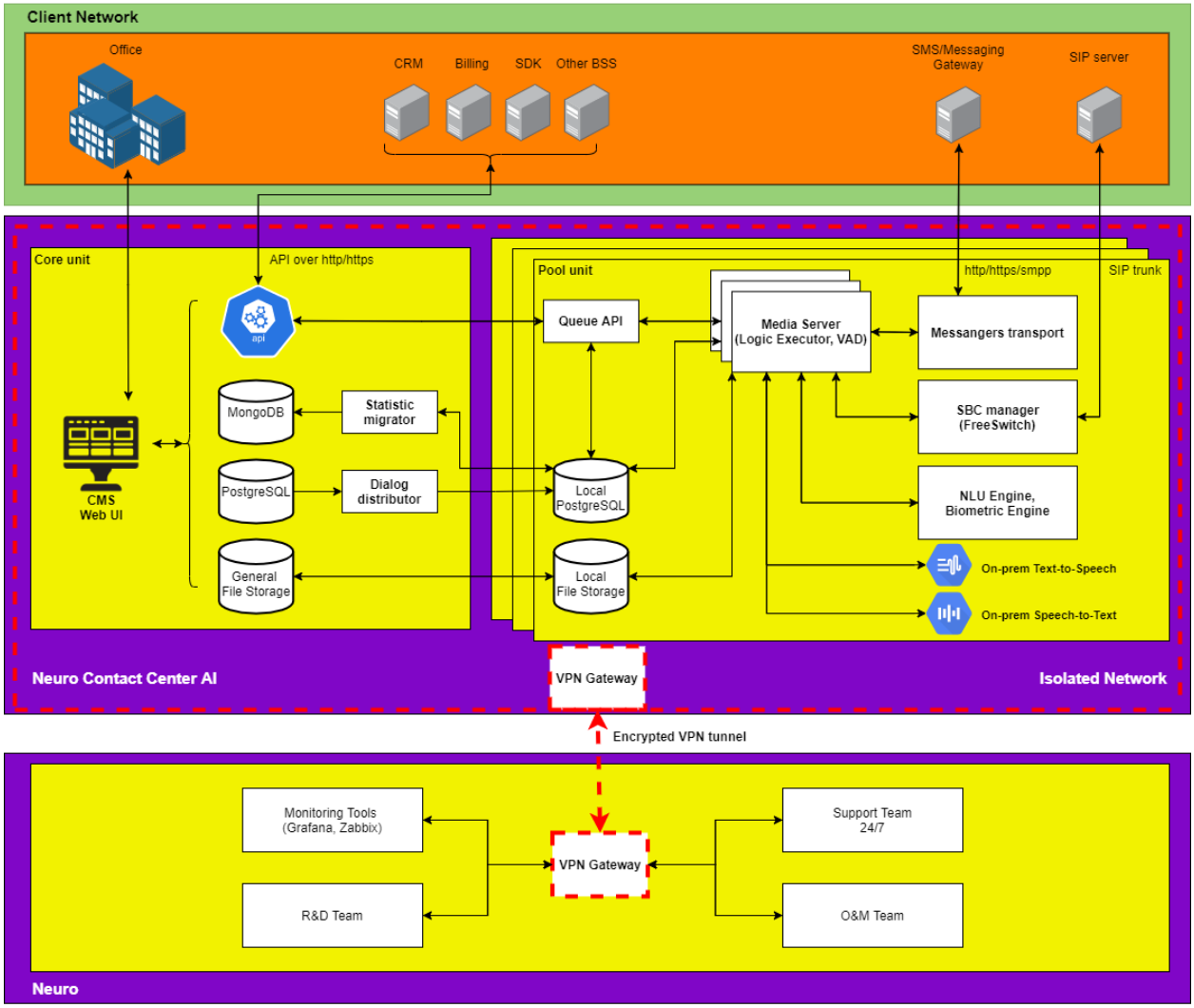
123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

- Разработанное компанией решение VAD в сочетании с остальными собственными высокопроизводительными модулями (например, NLU Engine и Media Server) демонстрирует низкое значение задержки (~ 700 мс). Можно настроить время ожидания завершения звонка в зависимости от сценария использования.
- Безопасное хранение и передача данных.
- Механизмы защиты данных во время обмена: HTTPS, IPsec и т. д.
- Возможность горизонтального масштабирования, использование технологии контейнерных приложений.
- Интеграция API с любой системой заказчика посредством связующего ПО.
- Интеграция с телефонной сетью посредством SIP.

Техническая спецификация

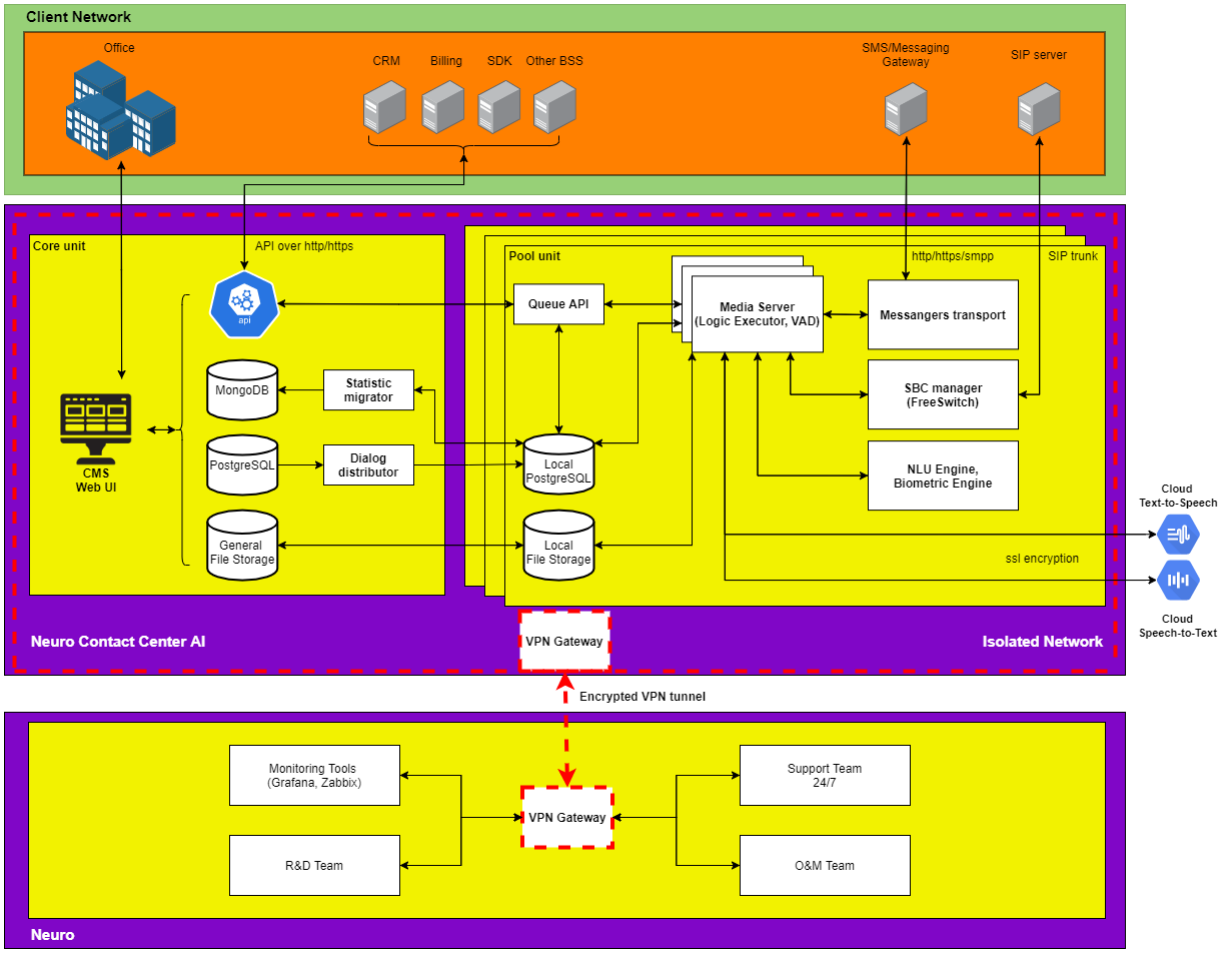
123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

Высокоуровневое проектирование в локальном варианте



123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

Высокоуровневое проектирование в облачном варианте



Сеть заказчика

- **SIP-server** (SIP-прокси) - оборудование, полностью управляющее SIP-вызовами в сети и отвечающее за прием запросов от агентов пользователя для выполнения и завершения вызовов.
- **SMS/Messaging Gateway** (A2P SMSC, Messaging Gateway) - центр услуг передачи коротких сообщений в соответствии со стандартами обмена мгновенными сообщениями (например, FB Messenger, Telegram, WhatsApp, Viber и т. д.) или стандартами GSM.
- **CRM, Биллинг, прочие BSS** - интеграция с системой поддержки бизнес-деятельности в сети заказчика.
- **SDK** - комплект для разработки программного обеспечения, который помогает разработчикам создавать свои приложения на основе платформы Contact Center AI.
- **Офис**– менеджеры, техническая поддержка и т. д.

123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

НЕЙРОСЕТЬ Contact Center AI

VPN-Gateway - сетевой шлюз, который предоставляет членам команды «Нейронет» (сотрудникам группы поддержки, группы мониторинга, команды DevOps и разработчикам) доступ к IP-сети заказчика и программному обеспечению для мониторинга через защищенный зашифрованный канал.

Инструменты мониторинга (Monitoring Tools) - программное обеспечение, которое проверяет доступность модулей Contact Center AI.

Центральный блок (Core Unit)

Предоставляет услуги настройки агента и сбора статистики вызовов.

Backend включает в себя следующие модули:

- **CMS** (система управления контентом) - пользовательский веб-интерфейс.
- **API** - HTTP API для интеграции с внешними системами.
- **General File Storage** для хранения аудио-файлов фраз, сущностей, фоновых звуков и записей звонков.
- **PostgreSQL** – база данных для хранения метаданных: пользователей, агентов, компаний. PostgreSQL содержит загруженные звонки, потоки разговоров и прочие данные пользователей/агентов.
- **MongoDB** хранит статистику, контакты и пр.
- **Statistic migrator** – переносит статистику в MongoDB.
- **Dialog distributor** – переносит новые диалоги в нужную локальную БД PostgreSQL.

Интерфейс CMS предоставляет следующие функции:

- Конструктор разговоров.
- Управление агентами.
- Управление пользователями.
- Дашборд.
- Управление очередями.
- Загрузка диалогов.
- Статистика.
- Управление аудио-файлами и записями содержимого, управление диалогами и контактами и пр.

123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

Блок пулов (Pool Unit)

Набор модулей для выполнения логики диалогов и внедрения потока разговора:

- **NLU Engine** - извлекает намерения и сущности из высказываний, полученных от системы преобразования речи в текст:
  - Первый уровень - распознавание паттернов, работающее на небольших наборах данных.
  - Второй уровень - глубокое обучение с предварительно обученными моделями для более чем 100 языков. В основе решения лежит архитектура преобразователя DistilBERT.
- **Biometric Engine** - распознает пол и возраст, идентифицирует пользователя исходя из речевых особенностей (альфа-версия).
- **Media Server** - приложение, работающее напрямую с SIP и RTP. Проигрывает аудио-записи, слушает абонента.
- **Logic Executor** - выполняет логику потока разговора, логику в режиме реального времени и фоновую логику, поддерживает линейные и нелинейные алгоритмы.
- **VAD** (обнаружение речевых сигналов) - обнаруживает триггер завершения человеческой речи на основе алгоритмов машинного обучения.
- **Локальная БД PostgreSQL** - хранит очередь звонков, статистику звонков перед миграцией в MongoDB.
- **Local File Storage** – записи фраз, записи сущностей, записи звонков и пр.
- **SBC manager** - обеспечивает безопасный доступ к SIP-транкам.
- **Queue API** - обрабатывает очередь звонков, загружает детали звонков из базы данных в память и по запросу предоставляет их остальным модулям.
- **Messengers Transport** - интегрируется со сторонними системами, такими как: платформа A2P (SMS), WhatsApp, Viber, Telegram, FB Messenger для обмена сообщениями, а также поддерживает передачу сообщений по электронной почте.
- **Text-to-Speech, Speech-to-Text** - использует лингвистические алгоритмы для преобразования речи в текстовое высказывание и текстовое высказывание в речь с использованием символов Юникода. Блок пулов поддерживает интеграцию со сторонними решениями для преобразования речи в текст и преобразования текста в речь. Добавлена поддержка интеграции с Google, Yandex, Tinkoff, STC. Интеграция с новыми поставщиками решений для обработки речи занимает около 1 недели.

Платформа Contact Center AI поддерживает функции омниканальности (звонки, мессенджеры, электронная почта), тем самым предоставляя заказчикам единую бесшовную коммуникационную среду.

Можно начать диалог посредством звонка и продолжить его, например, в WhatsApp или текстовым сообщением. В то же время виртуальный агент может учитывать данные, которые были собраны во время предыдущих контактов.

Диалог разделен на логику в режиме реального времени (поток разговора во время звонка) и фоновую логику (управление попытками дозвониться, обмен сообщениями и данными).

Система создает среду звонка для каждого диалога, собирает данные во время каждого контакта и сохраняет их, пока диалог активен. Затем среда звонка используется для каждого последующего контакта. Среда звонка содержит исходные данные (данные, полученные до создания диалога) и выходные данные (данные, полученные во время диалога и обновляемые после каждого контакта). Под контактом подразумеваются входящие и исходящие звонки, сообщения и электронная почта.

Платформа Contact Center AI также хранит информацию обо всех контактах во время диалога: количество звонков, количество сообщений и тип последнего контакта. Информация используется во время выполнения потока разговора. Например, если последним контактом был исходящий звонок, клиент пропустил звонок и перезвонил, информация о последнем контакте хранится в среде звонка. С этой информацией можно свериться в ходе выполнения потока разговора - в данном примере виртуальный агент может поприветствовать клиента следующим образом: *«Здравствуйте! Вас беспокоит компания [...]. Благодарим за обратный звонок. Недавно мы пытались с вами связаться, чтобы ...»*

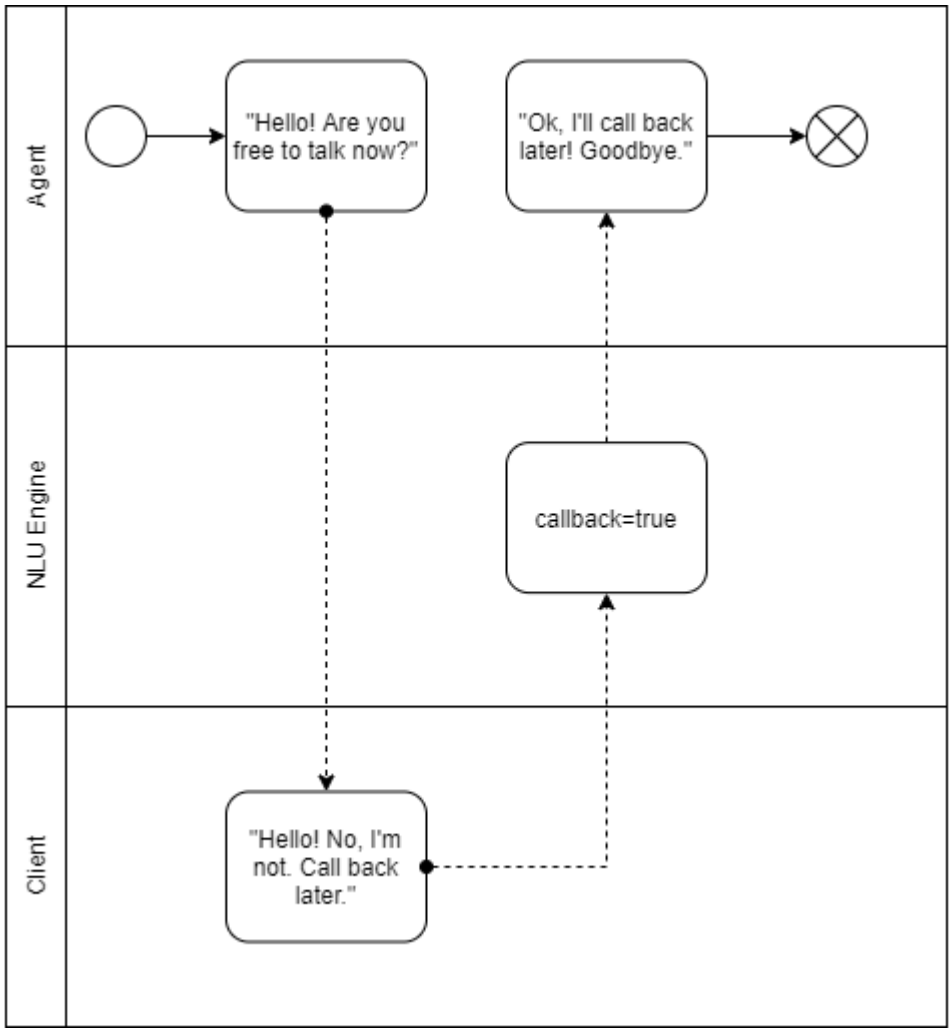
То же самое относится к смешанным текстовым и голосовым разговорам - например, клиент отправил запрос через чат-бота WA, и запрос был успешно обработан - намерения/сущности были извлечены посредством NLU Engine и записаны в среду звонка в виде выходных данных. Затем клиент делает звонок, чтобы уточнить статус запроса - система проверяет информацию о предыдущих контактах и уже собранных выходных данных в среде звонка. В этом конкретном примере система может ответить клиенту следующим образом: *«На основании информации, которую вы предоставили нам вчера в WhatsApp, мы создали запрос в нашей системе поддержки. Запрос все еще обрабатывается. Мы перезвоним вам, как только запрос будет обработан.»*

NLU Engine (классификатор намерений)

Под намерением (англ. “intent”) понимается предопределенный класс, выражающий намерение пользователя, выраженное им в речевом акте в процессе диалога.

Набор классов определяется требованиями проекта и особенностями скрипта/логики, и может быть специфичными для проектов или универсальными по всей предметной области (согласие/отказы/вопросы общего назначения).

Задача классификатора намерений состоит в том, чтобы по текстовому представлению фразы пользователя определить намерения (интенты), которые он выражает в рамках этой фразы.



Терминология:

- **Фраза/высказывание (phrase/utterance)** - текстовое представление высказывания пользователя системы.
- **Контекст (context aka bot stimula)** - категориальная переменная, характеризующая “на что ответил пользователь”.
- **Тренировочная выборка** - набор данных, каждый элемент которых представляет собой следующую комбинацию: фраза, контекст, список намерений.

NLU

Neuro NLU состоит из двух частей: классификатор намерений на основе паттернов (pattern-based intent classifier) и классификатор намерений на основе глубокого обучения (DL-based intent classifier). NLU может одновременно задействовать обе системы.

123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

Классификатор намерений на основе паттернов

Часть, основанная на паттернах, выполняет две задачи:

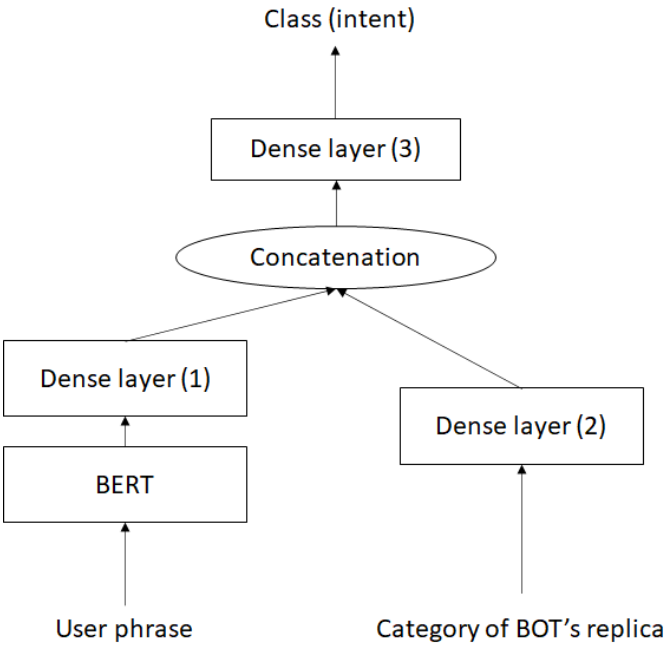
1. Первичный сбор данных. Любой проект начинается с классификатора намерений на основе паттернов, чтобы как можно скорее начать сбор данных из телефонных разговоров.
2. Коррекция возможных ошибок, допускаемых классификатором намерений на основе глубокого обучения. При работе с любой системой, основанной на машинном обучении, возможны ошибки. Однако, переобучение системы может занять некоторое время, что в ситуации продуктового трафика может быть неприемлемым. В этой ситуации можно прописать на ошибочную фразу паттерн. Паттерн будет нивелировать ошибку, возникающую в классификаторе намерений на основе глубокого обучения.

В основе данного классификатора лежит работа с регулярными выражениями (англ. "regular expressions") - используемый в компьютерных программах, работающих с текстом, формальный язык поиска и осуществления манипуляций с подстроками в тексте, основанный на использовании метасимволов. Для поиска используется строка-образец (англ. "pattern"), состоящая из символов и метасимволов и задающая правило поиска. Для манипуляций с текстом дополнительно задается строка замены, которая также может содержать в себе специальные символы.

Классификатор намерений на основе глубокого обучения

В основе классификатора намерений на основе глубокого обучения лежит BERT-based классификатор (используется DistillBERT: “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter” https://arxiv.org/abs/1910.01108), который обучается на достаточном числе данных из тренировочной выборки.

Архитектурное решение, реализованное в рамках NLU, принимает на вход предварительно обработанную фразу пользователя, а также категориальную переменную, которая кодирует категорию фразы бота - в ответ на какую фразу бота пользователь выдает нам ответ.



В классификаторе намерений используется предварительно обученная модель DistillBERT, которая в дальнейшем дообучается на тренировочных данных - фразах пользователя, контексте и целевом намерении.

Уникальные особенности NLU

- **Предварительная обработка.** Поскольку мы принимаем на вход результат распознавания ASR, который может содержать неточности и ошибки распознавания, разработан модуль предварительной обработки, который пытается скорректировать типовые ошибки, возникающие в ходе транскрибации. Производится дополнительная чистка от неинформативных слов.
- **Множественные намерения.** NLU Engine способен эффективно работать в условиях, когда в одной фразе пользователь выражает несколько намерений. То есть, рассматривается так называемая задача многозначной классификации.
- **Вероятность намерений.** Помимо предсказания непосредственно намерений, мы предсказываем вероятность того, что фраза относится к тому или иному намерению. За счет этого можно использовать пороги (thresholds), в соответствии с которыми можно фильтровать намерения (например, возможно снизить порог для критических намерений).
- **Использование контекста (bot stimula).** Изначально NLU разрабатывался для работы в качестве целеориентированной диалоговой системы - когда есть сценарий диалога, и пользователь проходит по данному сценарию. Контекст (bot stimula) выступает как некая категория, характеризующая в ответ на что пользователь

говорит ту или иную фразу. Это ответ на какой-то вопрос? Или ответ на приветствие? И тому подобное. Стоит отметить, что классификатор намерений может функционировать и без контекстной переменной, однако, ее использование приводит к повышению качества предсказания намерения.

- **Использование двухэтапной системы** - на основе паттернов и глубокого обучения, позволяет при возникновении каких-либо неточностей в модели, основанной на глубоком обучении, оперативно внести коррективы в работу системы.

Сервисы по обработке речи

Speech-to-text

Преобразование речи в текст (англ. "Speech-to-text, STT"), также известное как автоматическое распознавание речи (англ. "Automatic Speech Recognition, ASR"), позволяет транскрибировать аудио-потоки в текст в режиме реального времени.

В дополнение к стандартной модели сервиса по обработке речи могут быть добавлены пользовательские настройки. Кастомизация помогает преодолеть различные препятствия при распознавании речи, такие как стиль речи, словарный запас и фоновый шум.

Платформа Contact Center AI интегрируется с STT напрямую через протокол дуплексной связи WS. Это позволяет сократить время отклика и минимизировать задержки в обслуживании.

Платформа Contact Center AI интегрируется с поставщиками решений по обработке речи с использованием зашифрованных (SSL-шифрование) и анонимизированных методов (голос абонента передается небольшими частями и никоим образом не связан с персональной информацией человека), что обеспечивает полную анонимность данных абонента.

Повышение точности

Поставщики решений для преобразования речи в текст предлагают возможность проверки уровня точности путем измерения пословной вероятности ошибки (англ. "WER, Word Error Rate") - WER подсчитывает количество неверных слов, обнаруженных во время распознавания, а затем общее количество слов, представленных в размеченной человеком расшифровке (показано ниже как N). В итоге, это число умножается на 100% для расчета WER.

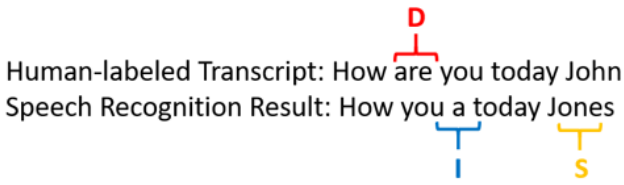
$$WER = \frac{I + D + S}{N} * 100\%$$

123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

Неверно обнаруженные слова делятся на три категории:

- Вставка (англ. "Insertion, I"): слова, которые неправильно добавлены в расшифровку гипотезы
- Удаление (англ. "Deletion, D"): слова, не обнаруженные в расшифровке гипотезы
- Замена (англ. "Substitution, S"): слова, которые отличаются в эталоне и гипотезе

Ниже приведен пример:



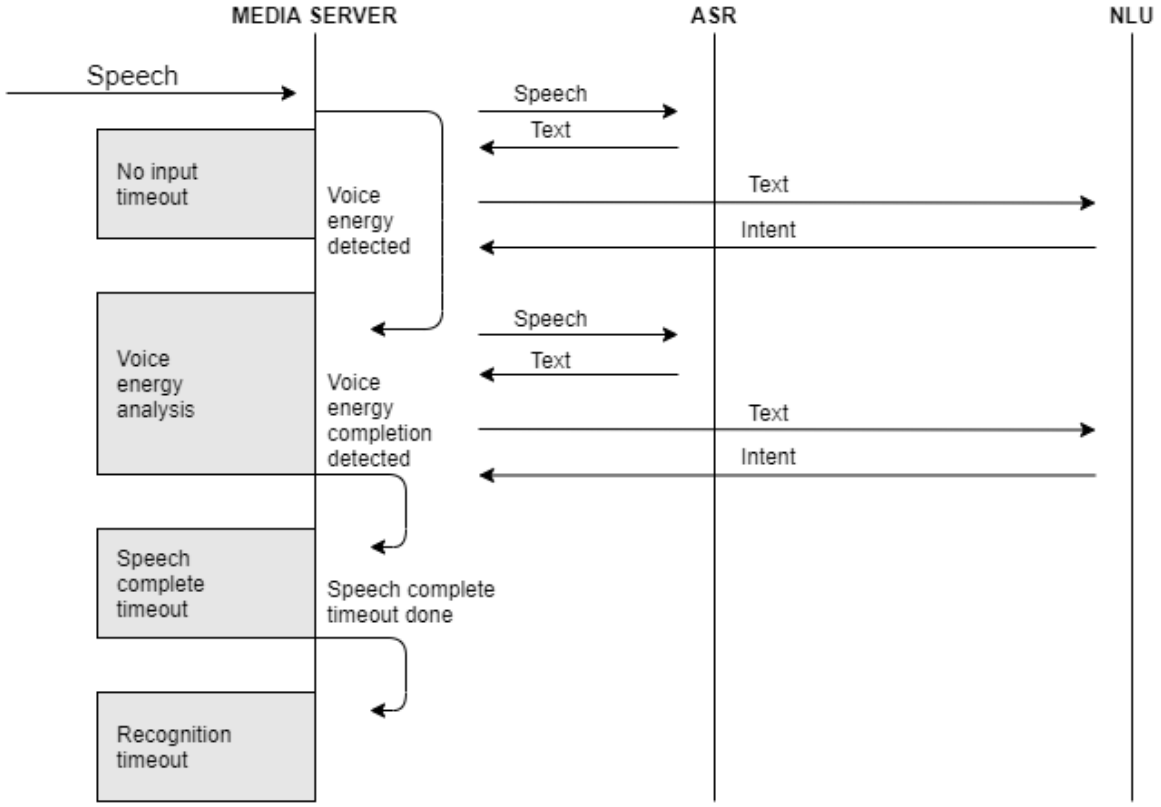
Text-to-Speech

Данный сервис позволяет преобразовывать текст в синтезированную человеческую речь. Сервис поддерживает нейронные голоса, похожие на человеческие, а также возможность создания собственного голоса, уникального для вашего продукта.

Нейронные голоса - глубокие нейросети используются для преодоления ограничений традиционного синтеза речи в плане ударения и интонации в разговорной речи.

Прогнозирование интонации и синтез голоса выполняются одновременно, что приводит к более плавным и естественно звучащим результатам. Нейронные голоса используются для того, чтобы сделать взаимодействие с чат-ботами и голосовыми помощниками более естественным и увлекательным. Благодаря естественной человеческой интонации и четкому произношению слов нейронные голоса значительно снижают слуховое утомление при взаимодействии с системами искусственного интеллекта.

Управление потоками звонка в Neuro Media Server



При получении аудио-потока Media Server отправляет его на ASR в потоковом режиме и получает текст в ответ в том же потоковом режиме.

Получив первый результат распознавания от Media Server, мы прекращаем отсчет таймаута no-input-timeout (заданное время, по истечении которого медиа-сервер будет считать, что абонент молчит) и запоминаем мощность голоса (англ. "VE, Voice Energy"). Запускается обработка текста и поиск намерений.

VAD включается в работу и анализирует поток во время распознавания - при уменьшении VE выполняется таймаут завершения речи (указанная длительность паузы после уменьшения VE), мы считаем, что абонент прекратил говорить и распознавание завершено.

Если абонент разговаривает в течение всего времени, отведенного на распознавание (recognition-timeout), то распознавание завершается принудительно с последним результатом распознавания.

Требования к аппаратному обеспечению для локальной установки

Отказоустойчивость платформы Contact Center AI составляет 99.9%, что достигается за счет дублирования (N + 1) сервисов платформы, а в некоторых случаях и утроения. HAProxy

выступает в качестве балансировщика нагрузки, который в случае инцидента на одном экземпляре как можно скорее передаст нагрузку другому.

Требования к установке платформы Contact Center AI с учетом 1500 одновременных звонков и отказоустойчивости 99,9%.

⚠Данные требования носят рекомендательный характер и подлежат уточнению после получения подробной информации о периоде наибольшей нагрузки.

сервер	**количество**	**ВЦП**	**ОЗУ**	**Диск**
mediaservers	50	4	8	130 HDD
maindb	3	6	6	130 SSD
pooldb	2	14	14	600 SSD
apicms	2	8	16	130 HDD
glusterfs	3	6	6	200 SSD
nluapi	2	4	8	130 HDD
elasticmasters	3	4	8	60 SSD
elastidata	2	4	8	100 SSD
MongoDB	2	8	8	400 SSD
mongodbarbitr	1	2	2	60 SSD
redis	3	6	6	80 SSD
rabbitmq	3	8	8	80 SSD
dnsoptional	2	4	4	60 HDD
haproxy	2	4	4	60 SSD
sbc	2	8	8	60 SSD
etcd	3	6	6	60 SSD
nluengine	2	25	4	130 HDD
Связующее ПО (middleware)	2	4	8	130 SSD

123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

⚠ Убедитесь, что на диске достаточно свободного места для хранения статистики и записей звонков, если возникает такая необходимость.

Количество минут в месяц	MongoDB	Хранение записей
1.000.000	20 GB SSD	2.4 TB HDD
2.000.000	40 GB SSD	4.8 TB HDD
5.000.000	100 GB SSD	12 TB HDD
10.000.000	200 GB SSD	24 TB HDD

Масштабируемость

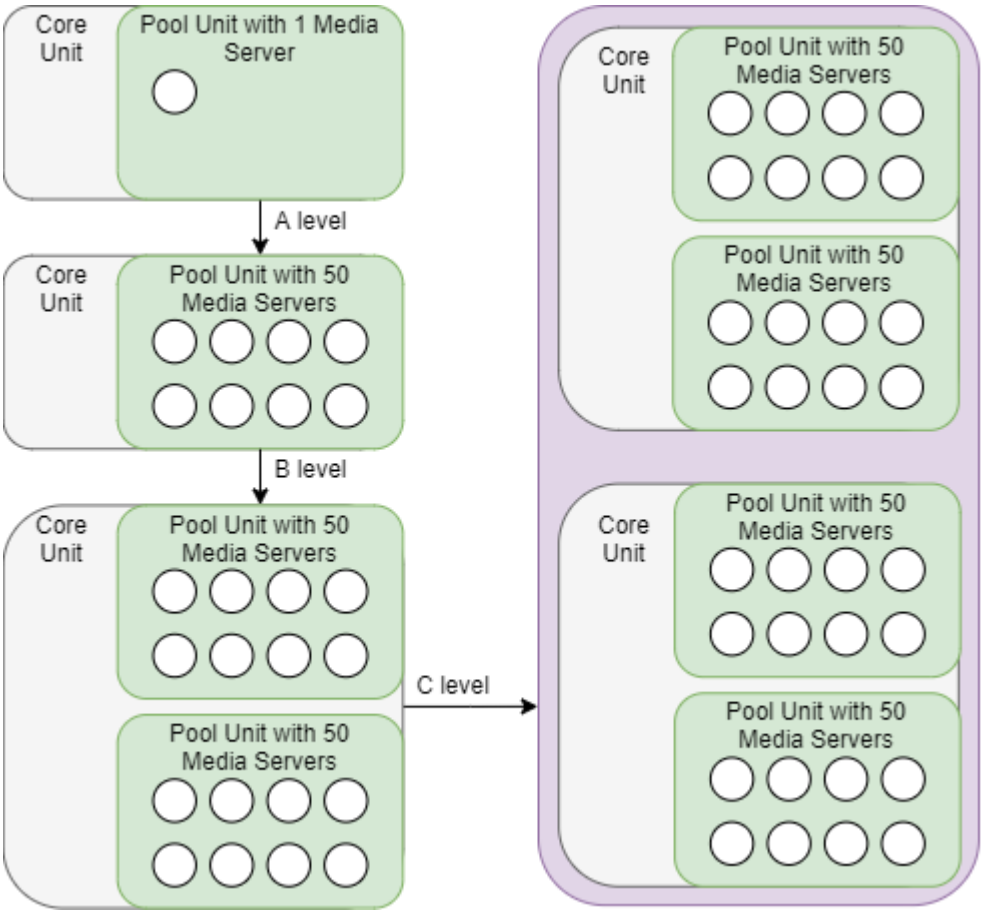
Платформа Contact Center AI — гибкое и чрезвычайно надежное решение с высоким уровнем масштабируемости, основанное на целом ряде технологий в сфере голосовых роботов. Представители больших и малых предприятий находят в нем решение конкретно своих коммуникационных задач, что позволяет им максимально автоматизировать работу колл-центра или других отделов.

Платформа Contact Center AI эффективно масштабируется от 30 до 15000 одновременных звонков. Подобные цифры легко достигаются путем автоматического горизонтального масштабирования, которое заключается в возможности быстрого развертывания дополнительных Media Server, Pool Unit, платформ.

- **Уровень масштабирования A.** Развертывание до 50 Media Server на каждый Pool Unit (1500 одновременных звонков).
- **Уровень масштабирования B.** Развертывание до 10 Pool Unit на каждую платформу (15000 одновременных звонков).
- **Уровень масштабирования C.** Развертывание неограниченного количества платформ Contact Center AI (неограниченное количество одновременных звонков).

Вы платите только за используемые экземпляры и экономите деньги, запуская экземпляры только при необходимости и завершая их работу, когда они больше не требуются.

123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14



Обработка персональных данных

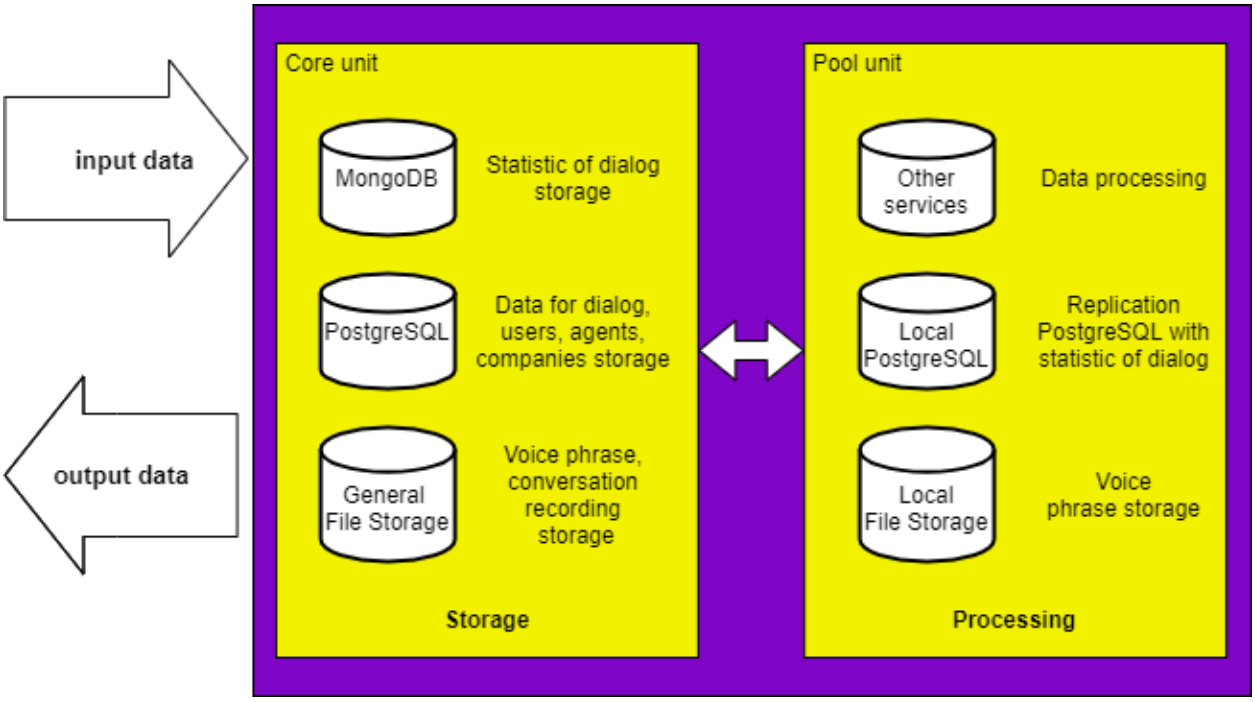
Сбор и обработка персональных данных зачастую необходимы и целесообразны для юридических и физических лиц. Цель компании «Нейросеть» - обеспечить баланс между преимуществами нашего бизнеса и бизнеса наших корпоративных клиентов и правами физических лиц в отношении их персональных данных. Исходя из вышеизложенного, в соответствующий продукт компании «Нейросеть» встроены определенные технологические функции, которые позволяют нашим заказчикам выполнять соответствующие требования, предписанные законами о конфиденциальности.

Виды персональных данных, обрабатываемых Нейросеть Contact Center AI

Платформа Contact Center AI может обрабатывать различные персональные данные для конкретных нужд - например, сбор среды звонка и номера телефона в телефонный справочник для последующего соединения с абонентом или сбор наименования назначенного SIP-транка и IP-адрес сервера для маршрутизации вызовов. Мы делаем все возможное, чтобы информировать наших клиентов о возможных действиях по обработке

данных в рамках нашего продукта и предоставлять клиентам контроль над такими данными. Персональные данные в частности могут включать в себя (но не ограничиваться) контактную информацию (например, компанию, адрес электронной почты, номер телефона, физический адрес), данные подключения (например, IP-адрес, операционную систему, интернет-провайдер, браузер, GPS/данные о местоположении) данные связи (например, идентификатор получателя вызова/вызывающего абонента, номер телефона получателя вызова/вызывающего абонента, продолжительность/время/дата вызовов, записанные голосовые сообщения, транскрипцию вызова, сохраненные контакты), сетевую информацию (например, загрузку процессора Media Server), данные об устранении неполадок (например, файлы журналов) и полученные из них метаданные.

Структура хранения и обработки данных



Input data – информация, которую пользователь может вносить на платформе Contact Center AI. Например, информация для диалога, информация о новом пользователе, настройки агента и т. д.

Output data – информация, которую платформа Contact Center AI выдает в зависимости от запроса пользователя. Например, статистика по диалогу, запись разговора, настройки агента и т. д.

Центральный блок (хранение):

- **MongoDB** – хранит текстовую статистику диалогов (например, номер телефона, дата создания, дата звонка, дата ответа, дата завершения звонка, транскрипция звонка, длительность звонка, значение уникальной сущности); данные, содержащие персональную информацию, которые пользователю необходимо собрать в ходе разговора (например, ФИО, адреса, номера документов, любая другая персональная информация, по которой

можно однозначно определить человека), биллинговая информация (например, расходы компании по определенному агенту, SIP-транку).

- **PostgreSQL** – хранит информацию, которую пользователь вносит на платформе Contact Center AI для дальнейшего создания голосового робота. Информация о компании (например, наименование, описание, часовой пояс), информация об агентах (например, наименование, описание, часовой пояс, язык, поток разговора, текстовое представление голосовой фразы), информация о пользователях (например, имя пользователя, номер телефона, email-адрес, дата регистрации), данные, содержащие персональную информацию, которые пользователь может добавить самостоятельно (например, сущности, которые будут необходимы в рамках диалога – это могут быть адреса, имена, должности, названия компаний, номера документов и другая персональная информация, по которой можно однозначно определить человека).
- **General file storage** – хранит записи фраз робота в формате *.wav, записи разговоров (пользователь сам решает, что будет говорить робот, поэтому фразы могут содержать адреса, имена, должности, номера документов и другую персональную информацию, по которой можно однозначно определить человека).

Блок пулов (обработка):

- **Local PostgreSQL** – реплицирует данные таблиц базы PostgreSQL для дальнейшего диалога с абонентом (например, информацию об агенте, информацию о компании; данные, содержащие персональную информацию, которые пользователь может добавить самостоятельно). Хранит текстовую статистику диалога до ее миграции в MongoDB (например, номер телефона, дата создания, дата звонка, дата ответа, дата завершения звонка, транскрипция звонка, длительность звонка, значение уникальной сущности); данные, содержащие персональную информацию, которые пользователю необходимо собрать в ходе разговора (например, ФИО, адреса, номера документов, любая другая персональная информация, по которой можно однозначно определить человека).
- **Local file storage** – реплицирует записи фраз робота в формате *.wav из General file storage для дальнейшего разговора с абонентом.
- **Other services** – остальные сервисы системы, которые обрабатывают информацию из Local PostgreSQL и Local file storage для построения верного диалога с абонентом. Могут хранить персональную информацию в лог-файлах, такую как: результаты распознавания (например, адреса, имена, должности, номера документов и другую персональную информацию, по которой можно однозначно определить человека), текстовое представление голосовой фразы, результаты работы потока разговора (например, значения переменных, в которых может содержаться персональная информация о человеке). По запросу клиента платформа Contact Center AI может быть установлена локально или в облаке.

Доступ к персональным данным, обрабатываемым в рамках платформы Contact Center AI

Компания Нейросеть может получить доступ к определенным персональным данным только в ходе обычных коммерческих операций (например, путем выполнения соглашения/указаний корпоративного клиента для целей, которые были доведены до сведения корпоративного клиента или субъектов персональных данных, с соблюдением

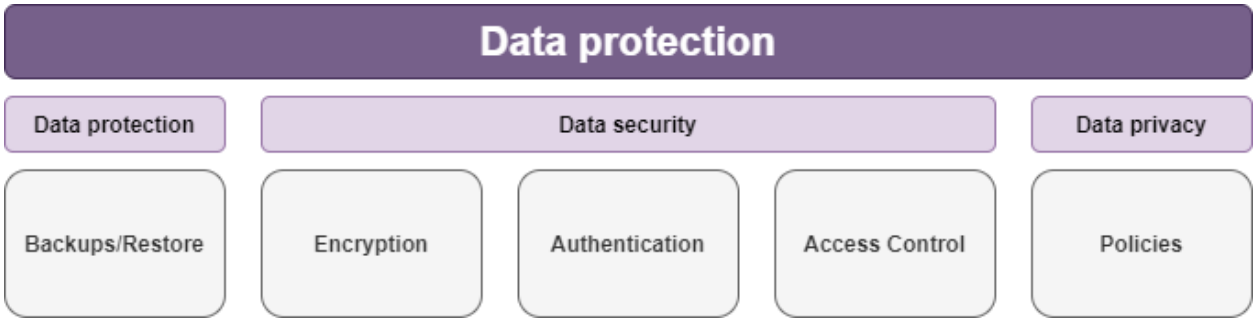
всех требований применимого законодательства и т.д.) при предоставлении запрошенного продукта или услуг.

Хранение персональных данных

Компания Нейросеть будет хранить и использовать персональные данные по мере необходимости для достижения целей, для которых они были собраны, или по мере необходимости для разрешения споров, обеспечения соблюдения условий контрактов и/или соблюдения наших юридических обязательств. Платформа Contact Center AI предоставляет клиентам (то есть, контроллерам данных) возможность принять определенные меры технического характера, позволяющие решить, как долго персональные данные должны храниться в рамках продукта.

Безопасность персональных данных в рамках платформы Contact Center AI

Одним из приоритетов компании «Нейросеть» является обеспечение безопасности данных. У нас работают высококвалифицированные специалисты, которые помогают обеспечить безопасную обработку информации и персональных данных, находящихся под контролем и в зоне ответственности компании, независимо от того, связаны ли они с услугами удаленного обслуживания платформы Contact Center AI, нашими облачными предложениями или любыми другими решениями, в рамках которых компания «Нейросеть» обрабатывает данные. Компания «Нейросеть» внедрила и будет поддерживать меры безопасности технического и организационного характера, соответствующие характеру персональных данных, которые собираются и обрабатываются платформой Contact Center AI. Все передаваемые и хранящиеся персональные данные будут защищены с помощью, например, шифрования и/или мер контроля доступа.



123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

Меры контроля персональных данных

Сбор персональных данных

Платформа Contact Center AI может обрабатывать различные персональные данные для конкретных нужд - например, сбор среды звонка и номера телефона в телефонный справочник для последующего соединения с абонентом или сбор наименования назначенного SIP-транка и IP-адрес сервера для маршрутизации вызовов. Некоторая информация может быть сохранена в системных журналах для последующей диагностики или аудита. Когда информация больше не нужна, эти журналы могут быть удалены. Системные резервные копии могут также собирать некоторые персональные данные в той степени, в которой они присутствуют в данных, для которых выполняется резервное копирование. По этой причине при оценке персональных данных необходимо учитывать как активные системные данные, так и резервные копии.

Шифрование

Шифрование в состоянии покоя обеспечивает безопасность содержимого файла или базы данных таким образом, что делает его непригодным для использования кем-либо, у кого нет надлежащего разрешения. Некоторые сервисы предлагают опции для поддержки шифрования, а другие нет. Для тех сервисов, которые их не предлагают, могут быть введены компенсирующие меры (см. «Управление доступом» ниже). Шифрование при передаче применяется ко всем передачам данных вовне. Платформа Contact Center AI поддерживает протоколы SSL, HTTPS, IPsec.

Управление доступом

Платформа Contact Center AI обеспечивает средства управления доступом, которые могут быть использованы для ограничения возможности доступа отдельных лиц или систем к собранным данным. Могут быть предоставлены следующие средства контроля доступа:

- **Пароли** – пароли применяются для получения доступа к продукту. Нижеследующие правила применяются в отношении паролей: правила сложности, блокировки при неудачных попытках, требуемые интервалы смены паролей.
- **Двухфакторная аутентификация** – пользователь получает доступ к платформе Contact Center AI только после успешного получения SMS.
- **Управление доступом на основе ролей** – средства управления доступом на основе ролей позволяют системе предоставлять настраиваемые возможности для каждого

входа в систему, который был назначен «роли», для управления тем, к чему они могут получить доступ и что изменить в системе.

- **Управление доступом к файловой системе** – средства управления доступом к файлам ограничивают владение и тип доступа к информации, предоставляемый отдельным учетным записям в операционной системе продукта. Данные средства управления должны быть настроены в соответствии с принципом безопасности наименьших привилегий.
- **Списки контроля доступа к сети** – списки контроля доступа ограничивают сетевые соединения в соответствии с заранее определенными списками «разрешить» и «запретить», которые хранятся локально в системе.

Журналы аудита

Журналы аудита, в особенности журналы аудита безопасности, также являются ключевым элементом в управлении соблюдением нормативных требований. Журналы аудита фиксируют активность системы и могут использоваться для выявления потенциальных проблем.

Индивидуальные настройки клиента

Платформа Contact Center AI является многоцелевой и может быть настроена и интегрирована в общую архитектуру обработки бизнес-информации клиента. Ожидается, что платформа Contact Center AI и оборудование, не относящееся к компании «Нейросеть», будут функционировать вместе для выполнения общей обработки информации для бизнеса.

123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

Услуги по развертыванию платформы Contact Center AI

Установка

Установка и обновление оборудования, настройка программного обеспечения и сервисов. Также на данном этапе должна быть настроена сетевая политика - правила брандмауэра, VPN доступ и т. д.

Интеграция

Необходима интеграция со следующими сервисами заказчика:

1. Интеграция с SIP - установка SIP-соединения между телефонными системами заказчика и платформой AI Нейросеть Contact Center для входящих вызовов
2. Обмен сообщениями - необходимо обеспечить интеграцию с Messaging API, чтобы иметь возможность отправлять сообщения клиентам.
3. Биллинг, CRM, SDK и т. д. - для обмена данными между платформой Нейросеть Contact Center AI и системами поддержки бизнес-деятельности заказчика.

Ввод в эксплуатацию

Процесс обеспечения установки, настройки и тестирования всех сервисов и компонентов платформы Contact Center AI в соответствии с эксплуатационными требованиями поставщиков и операционной команды заказчика.

Разработка виртуальных агентов

Стандартный цикл разработки виртуального агента включает в себя следующие этапы:

- Разработка потока разговора - выстраивание потока разговора для каждого сценария использования.
- Обучение NLU - обучение, тестирование и проверка NLU Engine на основе наборов данных
- Разработка логики - подготовка логической части, написанной на Python или построенной в Visual Builder

Проверка проектных решений

123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

Процесс тестирования разработанного виртуального агента.

Включает в себя следующие пункты для проверки:

- Поток разговора
- Логическая часть и переходы между логическими единицами
- Взаимодействия между интегрированными сервисами (получение/передача данных через API)
- Проверка выходных данных и отчетов/аналитики

UAT

Пользовательское приемочное тестирование (англ. "User Acceptance Testing, UAT") является заключительной ступенью жизненного цикла разработки виртуального агента. Данный процесс подразумевает тестирование виртуального агента сотрудниками заказчика для того, чтобы убедиться, что он способен выполнять требуемые задачи, для решения которых он был разработан, в реальных ситуациях. Тесты UAT соответствуют требованиям заказчиков. Тестировщики UAT стремятся проверить внесенные изменения в соответствии с исходными требованиями.

Сеансы передачи команде заказчика

Команда компании «Нейросеть» предоставляет операционной команде заказчика поддерживаемую и работающую платформу Contact Center AI.

123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

Условия технической поддержки

Компания «Нейросеть» предоставляет условия технической поддержки, которые гарантируют высочайший уровень обслуживания, достаточный для оптимального функционирования системы.

Доступность услуг

Таблица 1

Услуга	Доступность/время отклика
Горячая линия	24x7 (7 дней в неделю, 24 часа в сутки)
Высокий приоритет	Время отклика: в течение 15 минут круглосуточно (24x7) Время разрешения: в течение 4 часов
Средний приоритет	Время отклика: в течение 60 минут круглосуточно (24x7) Время разрешения: в течение 5 дней
Низкий приоритет	Время отклика: в течение 8 часов круглосуточно (24x7) Время разрешения: в течение 10 дней
Выезд специалиста на место	По соглашению
Эскалация проблем	Да
Поддержка удаленного доступа	Да
Удаленное подключение	через VPN
Регулярные встречи	ежемесячно
Статистические отчеты	ежемесячно
Обслуживание	

Услуга	Комментарий
Устранение ошибок	Да
Услуги по обучению	Да
Обновление документации	Да
Регулярная информация о разработке продукта	Да
Исправления	Да
Новые релизы и обновления	Включая установку обновлений, новые функции и опции
Информация о профилактических мерах и обновлениях	Да

123557, город Москва, вн.тер.
г. муниципальный Округ
Пресненский, ул. Малая
Грузинская, дом 28,
помещение 6/14

Порядок взаимодействия для обеспечения качества услуг

Заказчик регистрирует запрос в системе отслеживания ошибок компании «Нейросеть», назначая приоритет данной неисправности на свое усмотрение.

В течение времени, указанного для выбранного приоритета в Таблице 2, компания «Нейросеть» связывается с заказчиком для получения дополнительной информации о неисправности и согласования приоритета запроса. Компания «Нейросеть» классифицирует неисправность, описанную в заявке, определяет степень сложности работ по устранению неисправности, запрашивает у заказчика необходимую дополнительную информацию.

Предварительный технический ответ в соответствии с проведенной классификацией предоставляется заказчику в течение времени отклика, указанного в Таблице 2.

Для приоритета высокого и среднего уровня приемлемым вариантом обходного решения проблемы считается выполнение компанией «Нейросеть» таких действий, в результате которых неисправность в программном обеспечении может быть переведена на более низкий уровень приоритета. Компания «Нейросеть» прилагает все усилия, чтобы предоставить обходное решение проблемы в сроки, указанные в Таблице 2. При этом компания «Нейросеть» обязуется продолжить работу по устранению неисправности в программном обеспечении в соответствии с новым уровнем приоритета неисправности.

В зависимости от классификации проблемы по приоритету, компания «Нейросеть» предоставляет решение проблемы в течение времени разрешения, указанного в Таблице 2.

Таблица 2

Уровень приоритета	Время отклика	Время обходного решения	Время разрешения
Высокий	15 минут	2 часа	4 часа
Средний	60 минут	24 часа	5 дней
Низкий	8 часов	Н/Д	10 дней