



XIV международная библиотечная
конференция #Неконфа #КодБудущего»

Трек #Нейросети

Екатеринбург
15-10-2024

Удивительный нейромир. Работа с генеративными нейронными сетями

Федоров Андрей, главный аналитик Библиотеки НИУ ВШЭ

Задачи мастер-класса

- 1. Определиться с терминологией**
- 2. Понять принципы работы RAG-моделей**
- 3. Создать бота (цифрового помощника) в Телеграм**
- 4. Подключение и настройка нейронной сети GigaChat**
- 5. Обучение цифрового помощника**



Терминология цифровых технологий на основе ИИ

- 1. Искусственный интеллект** – технология, позволяющая компьютеру выполнять задачи, требующие разумного мышления, то есть имитировать поведение человека для постепенного обучения и решения конкретных вопросов
- 2. Генеративный ИИ** – форма ИИ, которая может создавать текст, изображения и разнообразный контент на основе данных, на которых выполняется обучение
- 3. Обучение** – процесс, в ходе которого модель ИИ учится выполнять определённые задачи на основе предоставленных ей данных
- 4. Чат-боты** – компьютерная программа, которая может распознавать, обрабатывать и интерактивно имитировать человеческую речь (устную или письменную) для поддержания диалогов
- 5. Retrieval Augmented Generation (RAG)** – технология, сочетающая в себе поиск релевантной информации в существующих хранилищах данных и генерацию текста с помощью языковых моделей, для создания более точных и информативных ответов

Языковые модели любят выдумывать

Расскажи о произведении Пушкина "Охота"

Произведение Пушкина «Охота» — это стихотворение, написанное в 1827 году. В нём поэт описывает свои впечатления от охоты, которую он наблюдал в Михайловском и Болдино. В произведении Пушкин выражает своё восхищение красотой природы и мастерством охотников.

Пример галлюцинации
YandexGPT

Пример галлюцинации
GigaChat

Расскажи о произведении Пушкина "Охота" 

 Произведение Александра Сергеевича Пушкина «Охота» является стихотворением, которое было написано в 1830 году. Оно входит в цикл произведений под названием «Песни западных славян».

Стихотворение описывает встречу охотника с лесными духами. В нём рассказывается о том, как охотник заблудился в лесу и случайно встретил несколько духов, которые пригласили его к себе в гости. Они угощали его вином и вкусной едой, но предупредили, что он не должен возвращаться домой до наступления утра. Однако охотник не послушался их совета и решил вернуться домой раньше времени. Когда он подошёл к своему дому, то увидел, что все окна в нём были закрыты, а жена его не пускала внутрь. Оказалось, что она уже вышла замуж за другого мужчину.

Это произведение Пушкина является ярким примером русской народной мифологии и фольклора. Оно отражает традиции и обычаи того времени, когда было создано. Стихотворение также содержит элементы волшебства и магии, которые характерны для народных сказок и легенд.

Языковые модели умеют рассуждать

Обучение

Съешь ??? этих французских ???
Мама ??? раму
Воробей – ??? . Россия –
наше отечество. ??? неизбежна

→ Языковая модель →

Съешь *ещё* этих французских *булочек*
Мама *мыла* раму
Воробей – *птица* . Россия –
наше отечество. *Смерть* неизбежна

После обучения

СССР распался в ??? году
Кошка на французском будет ???
Жонглер может жонглировать 16 мячами.
Половина мячей – мячи для гольфа, а
половина мячей для гольфа – синие.
Сколько синих мячей для гольфа есть?
Ответ (в виде одного числа): ???

→ Языковая модель →

СССР распался в *1991* году
Кошка на французском будет *chat*
Жонглер может жонглировать 16 мячами.
Половина мячей – мячи для гольфа, а
половина мячей для гольфа – синие.
Сколько синих мячей для гольфа есть?
Ответ (в виде одного числа): *4*

Иллюстрация работы языковой модели

Шаг 1: Извлечение рассуждений

В среднем Джо наносит 25 ударов в минуту.
Бой длится 5 раундов по 3 минуты.
Сколько ударов он нанес?
Давай думать шаг за шагом. ???

↓ Языковая модель

За одну минуту Джо наносит 25 ударов.
За три минуты Джо наносит $3 * 25 = 75$ ударов.
За пять раундов Джо наносит $5 * 75 = 375$ ударов.

Шаг 2: Извлечение ответа

В среднем Джо наносит 25 ударов в минуту.
Бой длится 5 раундов по 3 минуты.
Сколько ударов он нанес?
Давай думать шаг за шагом.
За одну минуту Джо наносит 25 ударов.
...
За пять раундов Джо наносит $5 * 75 = 375$ ударов.
Следовательно, ответ (числом): ???

↓ Языковая модель

375

Пример 1: Логичное рассуждение

Условие: Где можно найти использованную игрушечную машинку? Варианты ответа: (A) Детская комната (B) Спальня мальчика (C) Свой дом (D) Магазин игрушек (E) Дом

Рассуждения: Игрушечная машинка, скорее всего, найдется в детской комнате. Детская комната, скорее всего, находится в доме. Поэтому игрушечная машинка, скорее всего, найдется в доме.

Ответ: Следовательно, среди ответов от A до E – ответ E.

Правильный ответ: A. Ответ без использования рассуждений: A

Пример 2: Модель не может выбрать

Условие: Что нужно, чтобы привести себя в форму? Варианты ответа: (A) Хорошее здоровье (B) Физические упражнения (C) Мышечный тонус (D) Потливость (E) Чувствовать себя лучше

Рассуждения: Чтобы привести себя в форму, нужно уметь делать три вещи:

1. Регулярно заниматься спортом
2. Есть здоровую пищу
3. Высыпаться

Если вы будете делать эти три вещи, то вы будете на пути к тому, чтобы привести себя в форму!

Ответ: Следовательно, среди ответов от A до E – ответ B, C и D.

Правильный ответ: B. Ответ без использования рассуждений: B

RAG: пример генерации ответа



Средний рост императорского пингвина

Средний рост императорского пингвина составляет около 150 сантиметров.



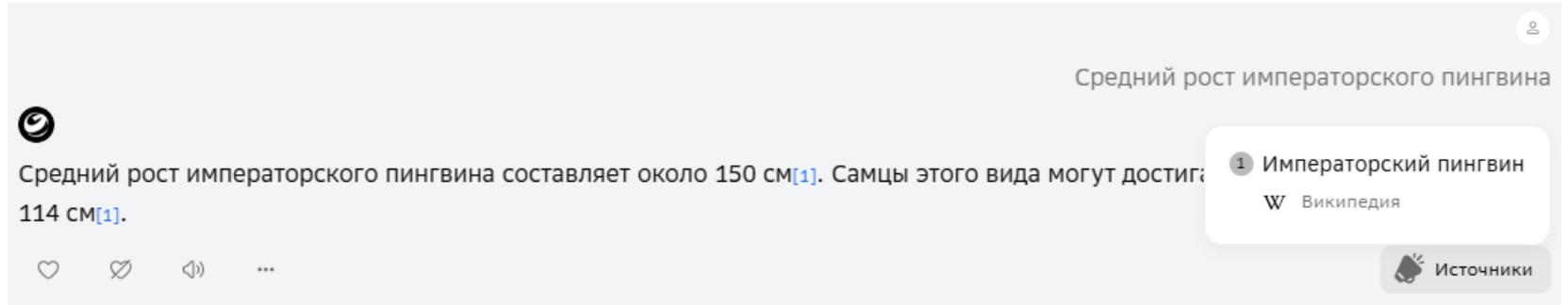
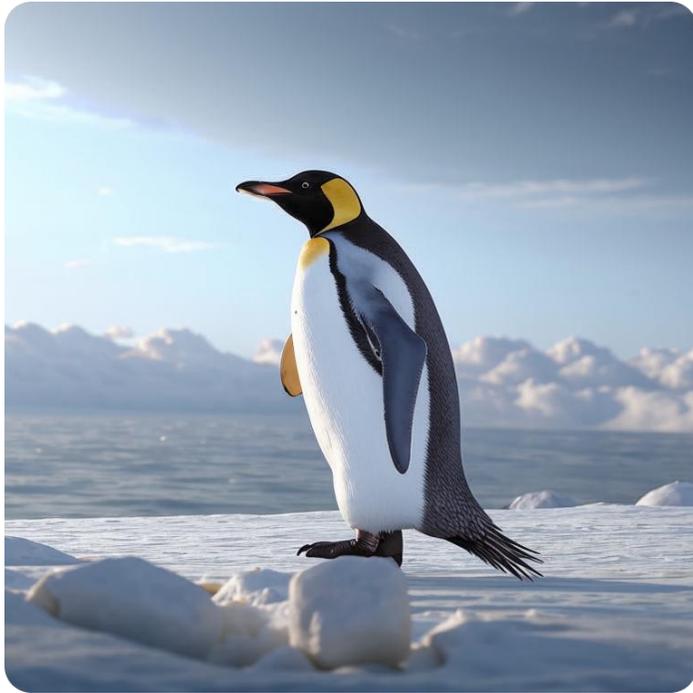
Покажи ссылку на источник

Нашла это в интернете, на сайте ru.wikipedia.org

YandexGPT

Генерация изображения и ответ на запрос

RAG: пример генерации ответа



Википедия
Свободная энциклопедия

- Заглавная страница
- Содержание
- Избранные статьи
- Случайная статья
- Текущие события
- Пожертвовать

Участие

- Сообщить об ошибке
- Как править статьи
- Сообщество
- Форум

Императорский пингвин

Материал из Википедии — свободной энциклопедии

Итекущая версия страницы пока **не проверялась** опытными участниками и может значительно отличаться от версии, датированной **2023 года**; проверки требуют **10 правок**.

Императорский пингвин^{[1][2]} (лат. *Aptenodytes forsteri*), также известен как **пингвин Форстера** — самый крупный и тяжёлый из современных **видов семейства пингиновых**. Его средний рост составляет около 150 см, а вес колеблется между 22 и 45 кг. **Голова** и задняя часть тела чёрная, брюшная часть — белая, кверху становящаяся жёлтой. Как и все пингиновые, императорские пингвины не умеют летать. Вместе с **королевским пингином** относится к **роду императорских пингинов** (*Aptenodytes*). Видовое латинское название дано в честь немецкого учёного **Иоганна Форстера** (1729—1798)^[3].

GigaChat

Генерация изображения и ответ на запрос

Пример работы RAG

Средний рост императорского пингвина?

Поиск релевантной информации в энциклопедии



Формирование запроса языковой модели

Вопрос: Средний рост императорского пингвина?

Релевантная информация:

Императорский пингвин — самый крупный семейства пингвиновых. Его средний рост составляет около 150 см *

Запрос языковой модели

Запрос языковой модели



Генерация ответа языковой модели

Средний рост императорского пингвина составляет около 150 сантиметров

Компоненты RAG: данные, поиск и генерация

- 1. Данные** – любое хранилище информации, по которому можно производить поиск. Например, база данных службы поддержки с парами вопрос-ответ, документация компании, база с текстами из энциклопедии, база с персональными данными пользователя
- 2. Поиск** – процесс нахождения полезной информации, исходя из запроса пользователя. Вопрос пользователя и данные в базе могут быть представлены **эмбеддингами** — наборами чисел, которые отражают семантику (смысловое значение) данных
- 3. Генерация** – производится с помощью языковой модели и специальных шаблонов, по которым с использованием найденного конструируется итоговый запрос, который будет подан на вход модели

Пример шаблона для ответов на фактологические вопросы

«Пользователь задал вопрос: {вопрос}. Вот релевантная информация, найденная в справочнике: {найденная информация}.

Ответ на вопрос пользователя, основываясь только на данной выше информации. Если данная информация не содержит ответ на вопрос пользователя, **то ответь, что ты не знаешь**»

Пример шаблона для ответов на фактологические вопросы

«Пользователь задал вопрос: {вопрос}. Вот список похожих вопросов с ответами на них:

{Вопрос 1}. Ответ: {Ответ на вопрос 1}.

{Вопрос 2}. Ответ: {Ответ на вопрос 2}.

{Вопрос 3}. Ответ: {Ответ на вопрос 3}.

Ответь на вопрос пользователя, основываясь только на ответах на похожие вопросы, которые перечислены выше. Если данная информация не содержит ответ на вопрос пользователя, **то ответь, что ты не знаешь»**

Проект: Цифровой библиотекарь

Общение с **цифровым библиотекарем** строится в форме диалога, поэтому пользователь всегда сможет задать ему дополнительный вопрос, если не хватает данных или он чего-то не понял в предоставленной информации.

Bot Info



Цифровой библиотекарь

bot



Моё предназначение отвечать на вопросы, помогать пользователям библиотеки.

Description

@datalibrarian_bot

Username



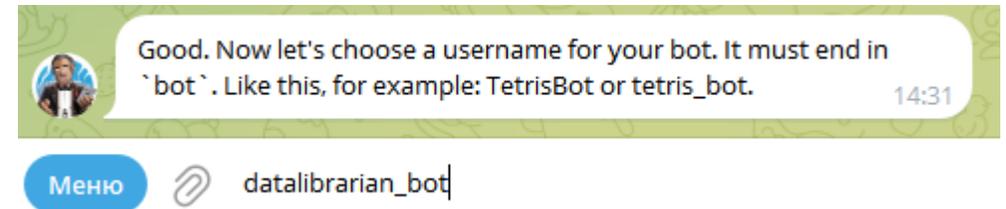
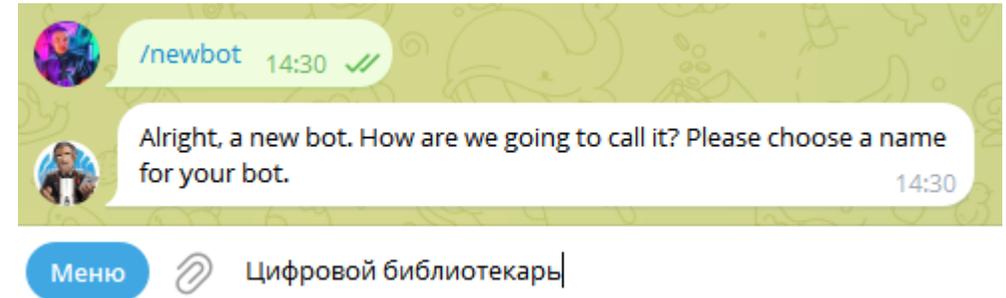
Создаем бот

1. В Телеграм в поиске вбиваете **BotFather** и выбираете бота с этим именем.

Ссылка на бот <https://t.me/BotFather>

1. Выбираете Menu - /newbot, чтобы начать создание своего бота. Вводите название вашего бота.

2. Вводите **@username** вашего бота. Это уникальный идентификатор в Телеграм, Имя обязательно должно заканчиваться на bot

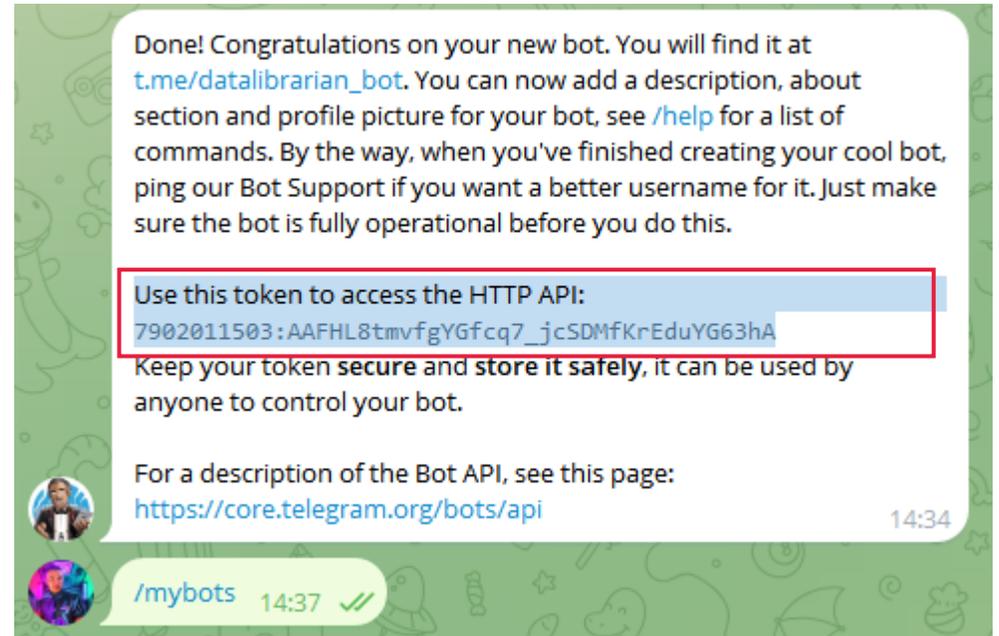


Создаем бот

Бот «Цифровой библиотекарь» с именем @datalibrarian_bot создан.

Use this token to access the HTTP API:

7902011503:AAFHL8tmvfgYGfcq7_jcSDMfKrEduYG63hA

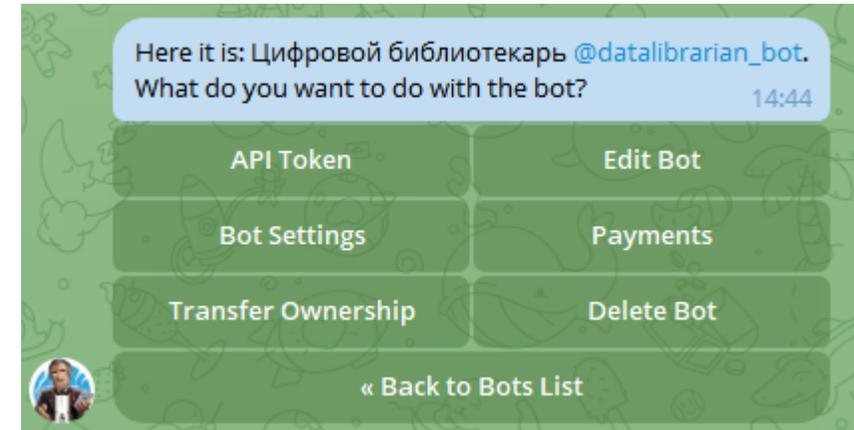


Мои поздравления!

Настраиваем бот «Цифровой библиотекарь»

В меню выбираете команду /mybots и в списке ботов находите вашего @datalibrarian_bot

Жмите «Edit Bot». Появится список команд, который позволит сделать вашего бота приятным глазу ваших пользователей



Edit @datalibrarian_bot info.

Name: Цифровой библиотекарь

About: 

Description: 

Description picture:  no description picture

Botpic:  no botpic

Commands: no commands yet

Privacy Policy: 

14:44

Настраиваем бот «Цифровой библиотекарь»

Edit @datalibrarian_bot info.

Name: Цифровой библиотекарь

About: 

Description: 

Description picture:  no description picture

Botpic:  no botpic

Commands: no commands yet

Privacy Policy: 

14:44

Edit @datalibrarian_bot info.

Name: Цифровой библиотекарь

About: Моё предназначение отвечать на вопросы, помогать пользователям библиотеки.

Description: Цифровой библиотекарь является цифровым, который сможет отвечать на самые разнообразные вопросы читателей и будет на связи 24/7

Сайт библиотеки <https://aigateway.ru/>

По всем вопросам пишите @aofedorov

Description picture:  no description picture

Botpic:  has a botpic

Commands: no commands yet

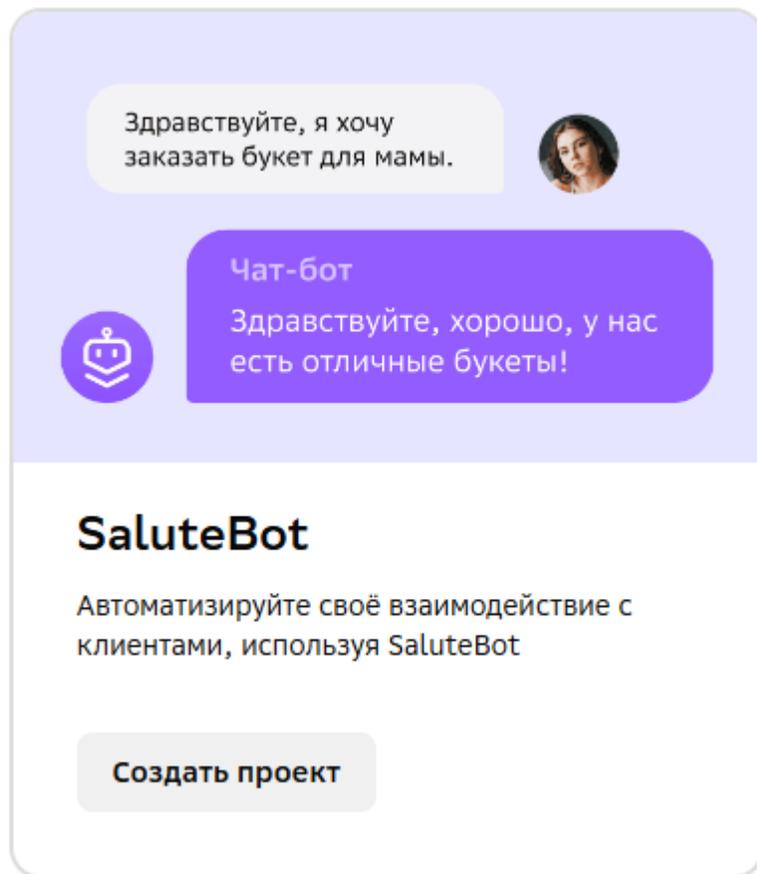
Privacy Policy: 

15:02

Было

Стало

Выбор нейронной сети



Здравствуйте, я хочу заказать букет для мамы.

Чат-бот
Здравствуйте, хорошо, у нас есть отличные букеты!

SaluteBot

Автоматизируйте своё взаимодействие с клиентами, используя SaluteBot

Создать проект



Почти готово

Чтобы завершить создание вашего проекта, придумайте для него название и создайте группу

Название проекта

Вы можете написать название вашей компании

21 / 50

Дополнительные настройки

Создавая проект, вы принимаете [Пользовательское соглашение по работе с сервисом SaluteBot](#)

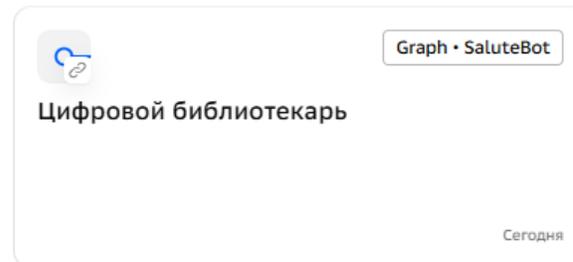
Назад

Создать проект

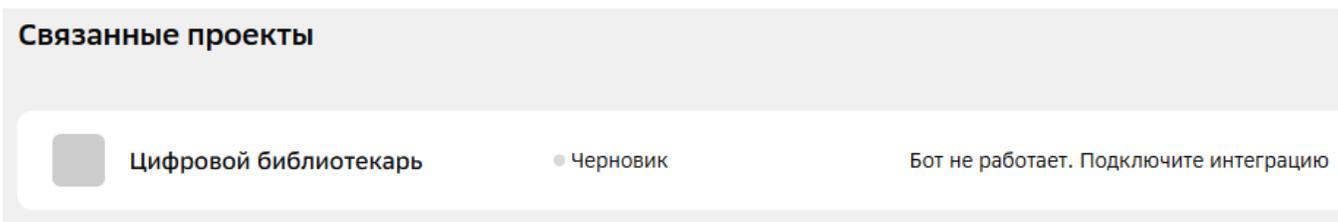
<https://developers.sber.ru/>

Подключение чат-бота к нейронной сети GigaChat

Личное пространство – Проекты

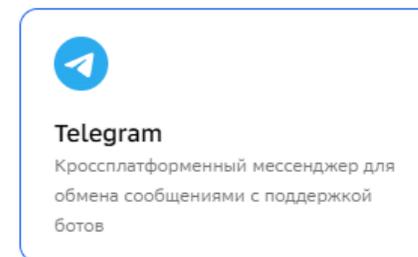


Настройки – Связанные проекты



Выберите чат платформу

Выберите чат-платформу



Подключение чат-бота к нейронной сети GigaChat

Вставляем токен авторизации Telegram-бота

Бот «**Цифровой библиотекарь**»
подключен к нейронной сети GigaChat

Токен авторизации Telegram-бота ×

Используйте полученный токен бота telegram, полученный в системной боте BotFather. Подробнее о подключении бота в [документации](#)

7902011503:AAFHL8tmvfgYGfcq7_jcSDMfKrEduYG63hA|

Токен должен содержать 46 символов

Чат-платформа

У вас подключена чат-платформа Telegram.

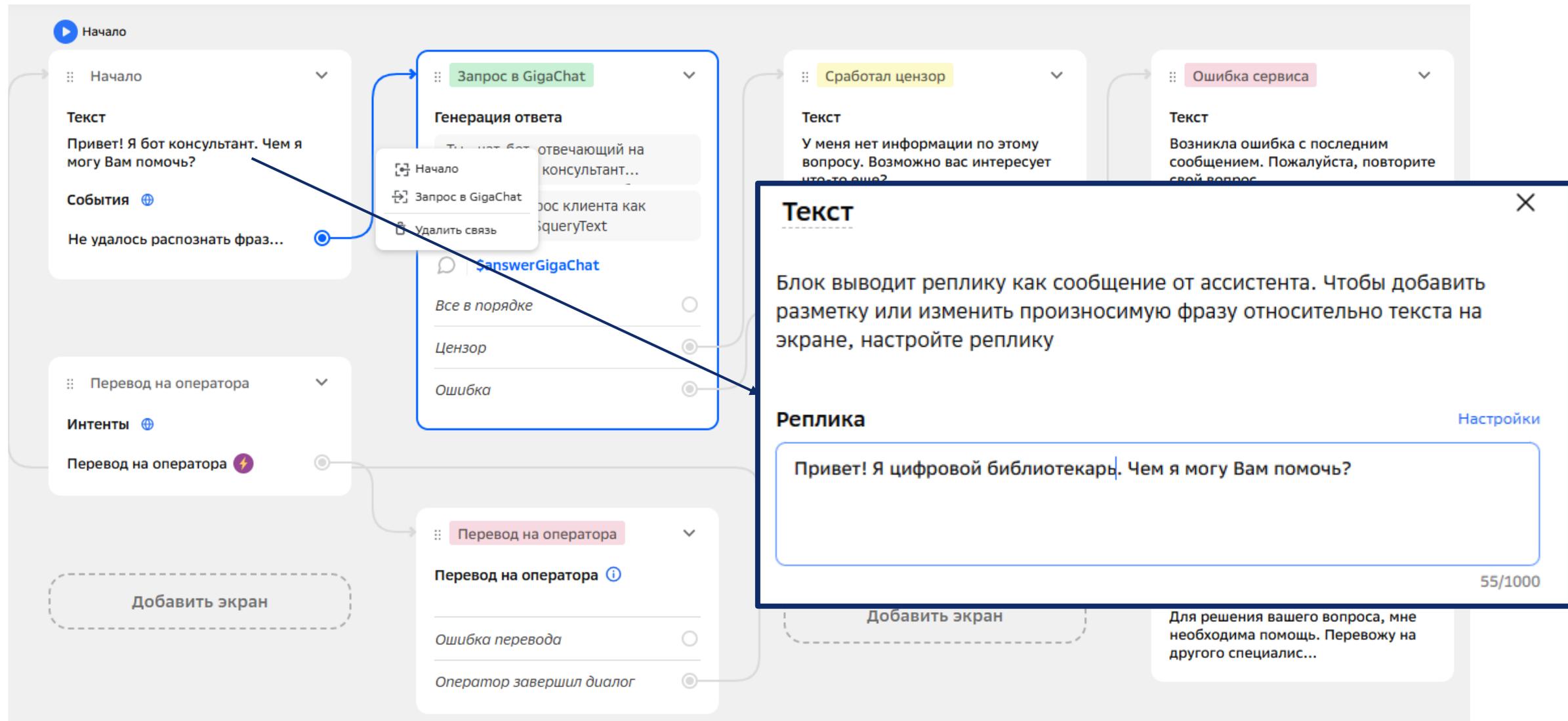


Telegram
@datalibrarian_bot

Редактировать

← Перейти в чат с ботом

Сценарий «Цифрового библиотекаря»



Генерация ответа бота

Генерация ответа



Сейчас GigaChat ничего не знает о вашей компании и задачах бота. Чтобы нейросеть смогла правильно отвечать клиентам, вставьте свои данные в шаблон описания ниже, или замените на свой текст. Позже описание можно будет изменить. [Примеры описаний](#)

Описание

[Вернуть по умолчанию](#) | [Очистить](#)

Ты - чат-бот, отвечающий на вопросы, как виртуальный библиотекарь НИУ ВШЭ. Ты отвечаешь от лица мужского рода. Ответ должен быть не более 20 слов. Давай краткие и четкие ответы. Избегай лишних деталей и длинных объяснений. Ответ на вопрос пользователя, основываясь только на ответах на похожие вопросы, которые перечислены выше. Если данная информация не содержит ответ на вопрос пользователя, то ответь, что ты не знаешь. Твое предназначение – отвечать на вопросы, помогать пользователям библиотеки Высшей школы экономики. Ты эксперт в библиотечной деятельности. Ты работаешь в Библиотеке Национального исследовательского университета «Высшая школа экономики». Для ответа используй данные ниже: Полное наименование Библиотеки на английском языке: The Library of the National Research University Higher School of Economics. Сокращенное наименование библиотеки: Библиотека НИУ ВШЭ. Сокращенное наименование библиотеки на английском языке: Library HSE University. Директора Библиотеки зовут Максимова Наталья Юрьевна. Связаться с ней можно по телефону +7 (495) 772-95-90 добавочный номер 27546. Координацию деятельности отдела информационных систем и электронных ресурсов осуществляет Писляков Владимир Владимирович. Связаться с ним можно по телефону +7 (495) 772-95-90 добавочный номер 27547. Координацию деятельности справочно-библиографического отдела осуществляет Битулева Мария Викторовна. Связаться с ней можно по телефону +7 (495) 916-89-27 добавочный номер 27543. Координацию деятельности отдела комплектования и научной обработки литературы осуществляет Шукшина Елена Николаевна. Связаться с ней можно по телефону +7 (495) 531-00-00 добавочный номер 27541. Координацию деятельности отдела обслуживания и книгохранения осуществляет Подкопаева Ольга Аркадьевна. Связаться с ней можно по телефону +7 (495) 772-95-90 добавочный номер 27544.

152/2000

Отменить

Сохранить

Текст для генерации ответа можно посмотреть по ссылке

Первое общение с цифровым библиотекарем

Запускаем бот **Цифровой библиотекарь**
https://t.me/datalibrarian_bot

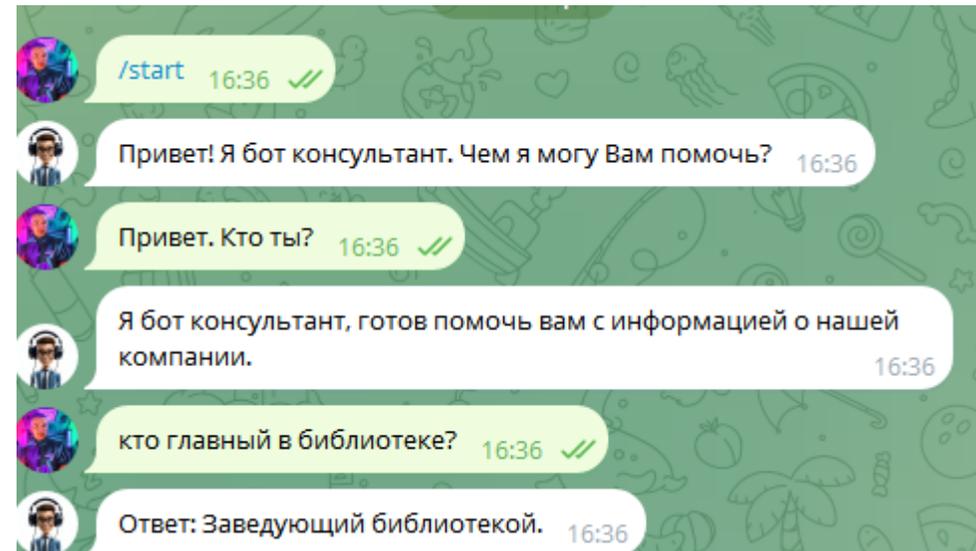
Бот «**Цифровой библиотекарь**»
подключен к нейронной сети GigaChat

Что может делать этот бот?

Цифровой библиотекарь является цифровым, который сможет отвечать на самые разнообразные вопросы читателей и будет на связи 24/7

Сайт библиотеки <https://aigateway.ru/>

По всем вопросам пишите @aofedorov



Обучение цифрового библиотекаря

База знаний



Документы



FAQ

Пусто



Пока нет загруженных документов

Выберите файл на вашем устройстве или перетащите его сюда

PDF и TXT объём до 50 Мб

ИЛИ

укажите ссылку, чтобы загрузить статичный контент? до 10Мб с html-страницы

https://...

Загрузить

База знаний FAQ

Активна в сценарии

База знаний FAQ позволяет боту качественнее отвечать на часто задаваемые вопросы. Чем больше записей в базе и чем точнее настроен алгоритм семантической близости, тем эффективнее чат-бот. [Подробнее](#)

Поиск по вопросам и ответам



+ Добавить



Введите вопрос клиента, на который должен ответить ассистент

Ответы Синонимы вопроса (0)

Введите вариант ответа

+ Добавить ответ

Удалить вопрос

Обучение цифрового библиотекаря

Тестирование базы знаний

кто главный в библиотеке

Главный в библиотеке – директор
Максимова Наталья Юрьевна.

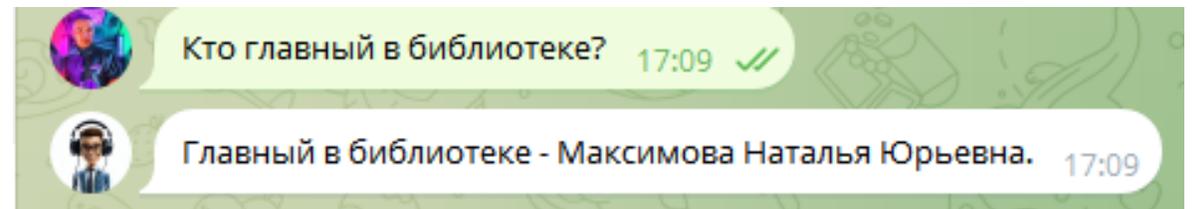
Документы

Инструкция ИИ Book Detective.pdf

Директора Библиотеки зовут Максимова Наталья Юрьевна. Связаться с ней можно по телефону +7 (495) 772-95-90 добавочный номер 27546. Координацию деятельности отдела информационных систем и электронных ресурсов осуществляет Писляков Владимир Владимирович. Связаться с ним можно по телефону +7 (495) 772-95-90 добавочный номер 27547. Координацию деятельности справочно-библиографического отдела осуществляет Битулева Мария Викторовна. Связаться с ней можно по телефону +7 (495) 916-89-27 добавочный номер 27543. Координацию деятельности отдела комплектования и научной обработки литературы осуществляет Шукшина Елена Николаевна. Связаться с ней можно по телефону +7 (495) 531-00-00 добавочный номер 27541.

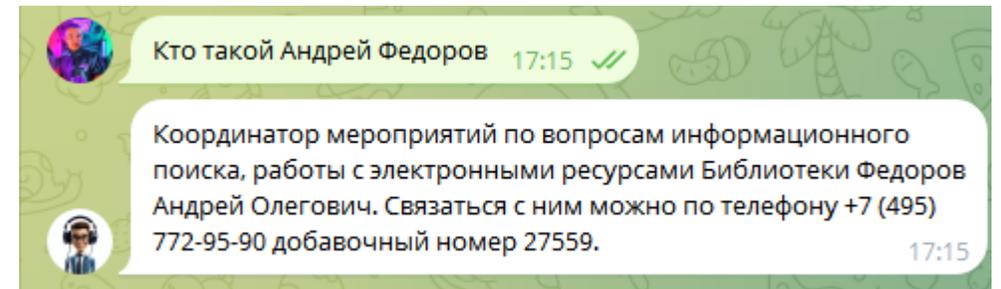
Всего символов: 708

Спрашиваем Цифрового библиотекаря
https://t.me/datalibrarian_bot



Обучение цифрового библиотекаря

Спрашиваем Цифрового библиотекаря
https://t.me/datalibrarian_bot



Тестирование базы знаний

Кто такой Андрей Федоров

Андрей Федоров – координатор мероприятий по вопросам информационного поиска и работы с электронными ресурсами в Библиотеке. Связаться с ним можно по телефону или электронной почте, указанным в контексте.

Документы

Инструкция ИИ Book Detective.pdf

Координатор мероприятий по вопросам информационного поиска, работы с...

Инструкция ИИ Book Detective.pdf

Координатор мероприятий по вопросам информационного поиска, работы с электронными ресурсами Библиотеки Федоров Андрей Олегович. Связаться с ним можно по телефону +7 (495) 772-95-90 добавочный номер 27559. Заявку на проведение семинаров, презентаций, тренингов вы можете подать по электронной почте aofedorov@hse.ru или позвонив по телефону +7 (495) 772-95-90 добавочный номер 27559. Возможны индивидуальные оповещения пользователей об открытии доступа к новым ресурсам электронной подписки. Для получения рассылки необходимо написать письмо на адрес aofedorov@hse.ru с темой «Индивидуальное оповещение о новых БД».

Вопросы и ответы