

ОПИСАНИЕ ТЕЛЕГРАМ-БОТА, КОНСУЛЬТАНТА И ПОДБОРЩИКА ТОВАРОВ С ПОДКЛЮЧЕННЫМ АРІ GIGACHAT И CHATGPT

Обновлено 12.07.24



ОГЛАВЛЕНИЕ

| | |
|---|----|
| Вводная часть..... | 3 |
| Покупателям..... | 3 |
| О chatGPT и GigaChat | 3 |
| Протестировать бота | 4 |
| Как тестировать..... | 4 |
| Гарантии | 5 |
| Возможности бота | 6 |
| Общее описание | 6 |
| Раздел «Настройки бота»..... | 6 |
| Настройки бота в меню «Экраны» | 7 |
| Настройки бота в меню «Тексты» | 7 |
| Настройки бота в меню «Настройки» | 9 |
| «ИИ – Настройки» | 10 |
| Функционал подборщика | 14 |
| Функционал консультанта | 15 |
| Список параметров экрана «Нейросеть» | 16 |
| Текст экрана | 16 |
| Параметры для типа экрана "Нейросети" | 17 |
| Параметры консультации | 20 |
| Как пополнять базу данных | 21 |
| Технические требования..... | 22 |
| Минимальные обещания разработчика..... | 23 |



ПОКУПАТЕЛЯМ

В боте есть только то, что описано в документе и именно та логика работы, которая описана. Если что-то не указано текстовым описанием явно, значит этого функционала нет в боте. Изображения нельзя воспринимать, как описание, как обещание – они могут быть поняты субъективно, быть устаревшими – они просто для наглядности. Подходите к этому так: если чего-то нет в этом документе, но увидели в боте, считайте, что этого нет в боте.

Бот, в том виде, который есть не дорабатывается под индивидуальные требования. Если нужен дополнительный функционал, то необходимо сначала купить текущую версию. После этого обсуждаются индивидуальные задачи с отдельной оплатой. До покупки основной версии можно уточнить о возможности реализации индивидуальных задач поверх имеющегося функционала бота.

Бот будет дорабатываться. Поэтому новый функционал будет появляться, а в этом документе в блоке «Обновления» будет указана дата последнего обновления. Однако вы покупаете именно ту версию, которая сейчас доступна, вы не можете претендовать на будущие обновления и дополнения функционала (пока что).

Настройка из @BotFather производится вашими силами. Например, поменять описание боту, его имя, изображение, групповые настройки и прочие.

О CHATGPT И GIGACHAT

В данной системе, в этой версии, нейросети используются, как инструмент анализа диалога. Нейросети «живут» сами по себе, на них не может повлиять разработчик и нейросети обладают массой параметров, наиболее сложным из которых является инструкция для роли System, объясняющая нейронке, что от нее ждут.

В этой связи, если ответ нейронки не является надлежащим, то решение нужно искать в тонкой настройке, а не в коде. Код работает для всех ситуаций единообразно, код написан по типу конструктора, поэтому только то, что передается из админки и от пользователя по сути и влияет на ответ.

Ключ доступа к API chatGPT и GigaChat не передается. Для chatGPT существуют сервисы покупки, оплаты за ключи API, а также специалисты, которые занимаются предоставлением ключей доступа. Для GigaChat вы можете самостоятельно создать доступы или попросить отдельного специалиста, или обратиться к разработчику за консультацией. В настройках можно задать модель, температуру и прочее.

GigaChat имеет большую скорость, причем заметно. При этом используя модель GigaChat-Pro можно добиться результатов часто превосходящих GPT, особенно в задачах со строгими параметрами ответа. Для более качественной работы требуется «поиграть» с настройками, а не просто «Привет – Здравствуйте - Тест».

Демонстрация бота, его админки может быть проведена онлайн через Телеграм. Для тестирования самого бота в мессенджере нужны ваши ключи, причем ключи от версии ChatGPT 4 и GigaChat Pro дадут заметно лучший результат. Это также влияет на скорость, качество ответов.

В боте введено понятие модификатор преобразования символов в токены. Он вполне себе примерный. Это нужно, чтобы хоть как-то стараться учитывать расходы символов, однако, необходимо учитывать, что понятие токен – величина плавающая не только от нейронки к нейронке, но и со временем внутри одной нейросистемы. Модификатор настраивается вами самостоятельно в админке. На 5.12.23 он был установлен в 1.8 - для chatGPT и в 3 – для GigaChat



На каждого пользователя установлен через админку бота лимит в 30 000 токенов. Если понадобится больше, то запросите у разработчика в ЛС, сообщив свой TG ID.

Уточнение: лимиты пользовательские изменяются через админку – это не жесткое ограничение в коде.

КАК ТЕСТИРОВАТЬ

Этот раздел нужен потому что бот демонстрационный, не имеет множества реальных данных. Поэтому здесь конкретизируется, что следует делать, чтобы получить результат.

На 14.12.23 бот настроен, как под поиск квартир, имитируя виртуального консультанта-подборщика агентства недвижимости, у которой есть некий список квартир, так и под составителя текстов.

Во время онлайн демонстрации админки и возможностей бота,
можно провести настройки бота под другие товары, другие запросы.

ПОИСК-ПОДБОР НА ОСНОВЕ ПРЕДПОЧТЕНИЙ

В демо-боте размещены следующие квартиры:

1. Квартира 2хкомнатная в Коломне
2. Квартира 2хкомнатная в Питере
3. Квартира 4хкомнатная в Москве
4. Квартира 2хкомнатная в Москве

У каждой есть такие параметры:

1. Дверной звонок – да
2. Количество комнат – 2 (или 4)
3. Школа рядом – да
4. Наличие wifi – да.

То есть если боту вы отправите произвольного содержания текст, указывающий на то, что вы ищите 2-х комнатную квартиру с дверным звонком, чтобы школа была рядом, и был вайфай, то бот сможет предоставить вам для просмотра 3 квартиры. Если вы измените параметр «количество комнат» на 4, то бот выдаст одну квартиру.

Обратите внимание, что бот сделан по принципу конструктора. Вы
можете самостоятельно настроить, что искать: квартиры, авто,
мелкие товары, штучные товары, поиск по портфолио и т.д.

Не имеет смысла указывать город или область или другие параметры, потому что в настройках указано, что конкретизировать (выявлять, подбирать) нужно только по этим параметрам.

Версия GigaChat-Pro показала более интересные результаты в
сравнении с GPT.



СОЗДАНИЕ ТЕКСТОВ

Бот также умеет создавать тексты. Он (при помощи нейронки) выявит у пользователя заданные в админке потребности и на основе этого создаст текст. То, что вы задаете в админке, фактически является главным промптом (инструкцией). Если вы скажете, условно, «давай-ка, сам думай, что спросить», то нейронка так и поступит. Поэтому, лучше контролировать и продумывать этот аспект.

К этому разделу особых комментариев нет. Просто нужно тестировать. Если хотите проверить перед покупкой на какой-то конкретной задаче, то обратитесь в ЛС к разработчику и обговорим это.

ГАРАНТИИ

На исправление ошибок в коде кодеры должны давать пожизненную гарантию. Только покупатели путают ошибки в коде с ошибками в логике, с собственными ожиданиями, с орфографией, с несовместимостью с оборудованием, устареванием кода или другим софтом. Это примечание особо актуально в связи с использованием нейросетей, к которым нужно иметь особый подход.

Поэтому гарантию на этот код разработчик дает 2 недели и акцентирует внимание, что это гарантия именно на код: например, обещано, что кнопка будет заблокирована, а она не блокируется потому что в коде забыл предусмотреть – это ошибка разработчика, а если она не блокируется потому что браузер пользователя IE6.0, то это ошибка покупателя. По-человечески если: обращайтесь, потому что найденные ошибки в коде (если они в коде), помогут и разработчику бота улучшать бота.



ОБЩЕЕ ОПИСАНИЕ

Бот сделан на основе конструктора этого же разработчика. Улучшен код, логика.

Это означает, что бот обладает элементами конструктора, который позволяет создавать экраны, привязывать кнопки, что поможет организовать и информационный бот.

В совокупности с конструктором, данная система дает широкий функционал. С одной стороны, у вас есть консультант с подборщиком по вашей базе данных товаров или услуг, с другой стороны вы вольны создавать разные экраны, привязывать к ним кнопки.

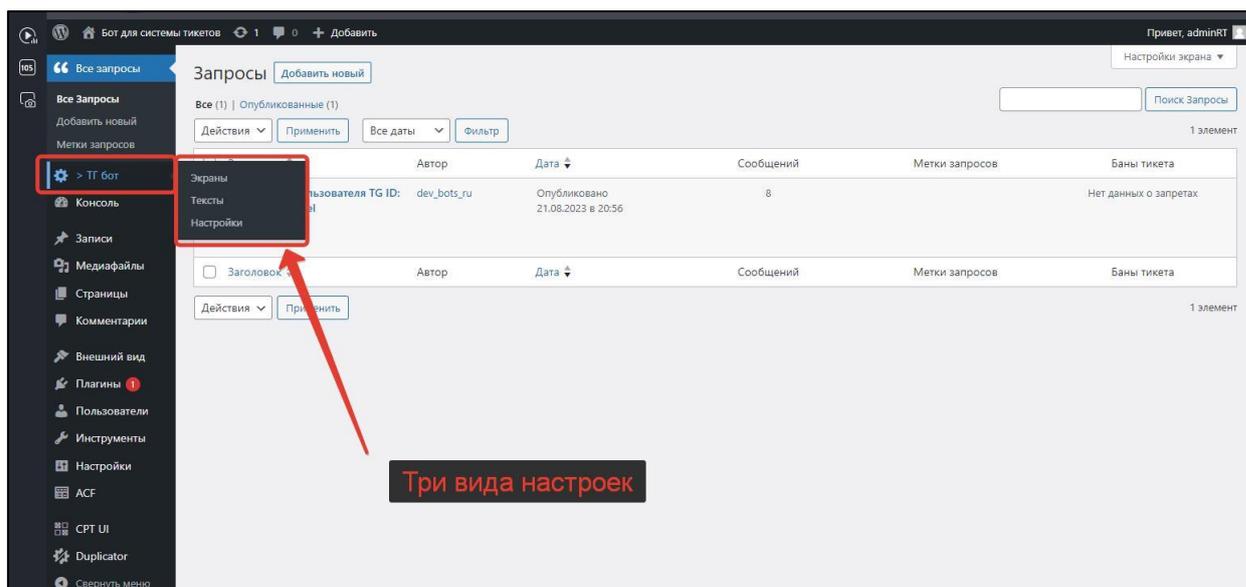
О базовом боте-конструкторе в данном документе будет мало информации. Основная часть будет посвящена функционалу консультации-подборке.

Помимо конструктора экранов, настройки бота тоже можно считать отдельным конструктором. Так, вы можете создать несколько экранов типа «нейросеть» и будет визуально разделение на несколько нейронов, каждая из которых будет работать по своей инструкции.

РАЗДЕЛ «НАСТРОЙКИ БОТА»

В боте есть три типа настроек:

1. Экраны – представляет собой настройку логики из конструктора.
2. Тексты – здесь прописываются тексты для пользователя, для операторов (иначе: модераторы, администраторы).
3. Настройки – здесь управляющие ботом элементы, например, количество загружаемых элементов, токен бота, данные администраторов и т.д.



НАСТРОЙКИ БОТА В МЕНЮ «ЭКРАНЫ»

Для получения представления о настройках «Экраны» необходимо изучить [примерное описание](#), скриншоты, видео к конструктору: <https://kwork.ru/script-programming/21707135/telegram-bot-konstruktor-s-adminkoy-na-wordpress> . По ссылке можно найти демонстрацию базового бота-конструктора.

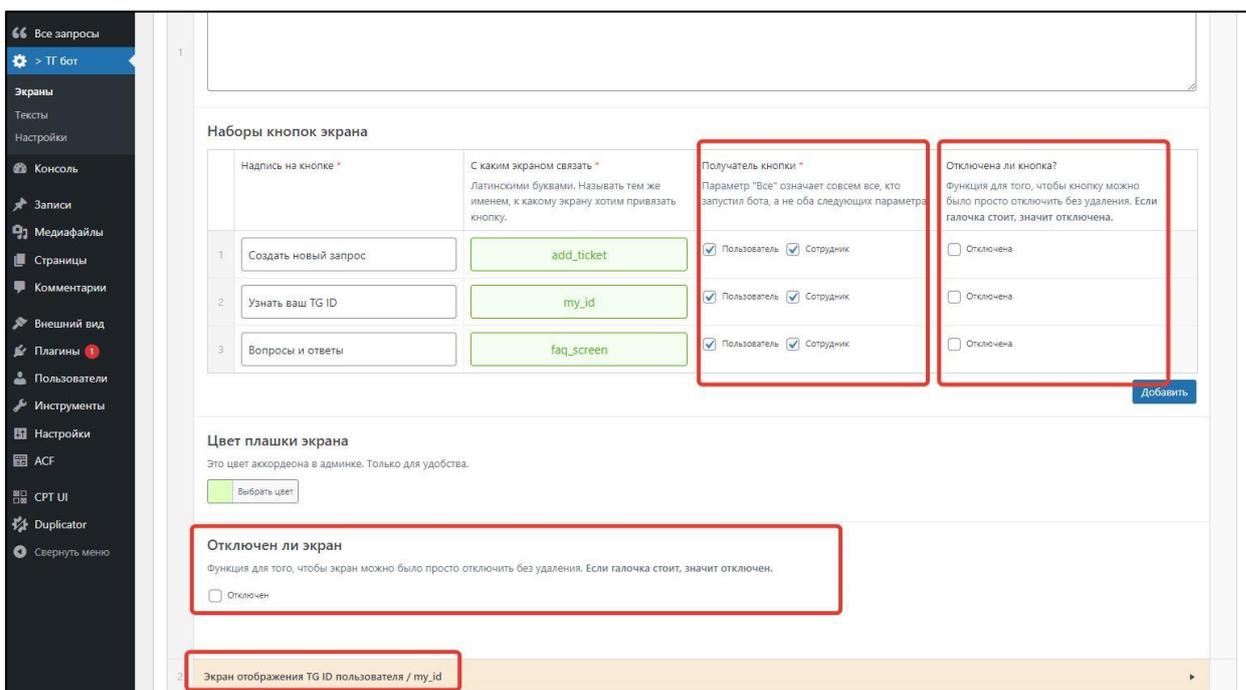
Основные отличия от изначального конструктора (см. скриншот).

Убрано:

- функционал рассылки,
- статистики,

Добавлено:

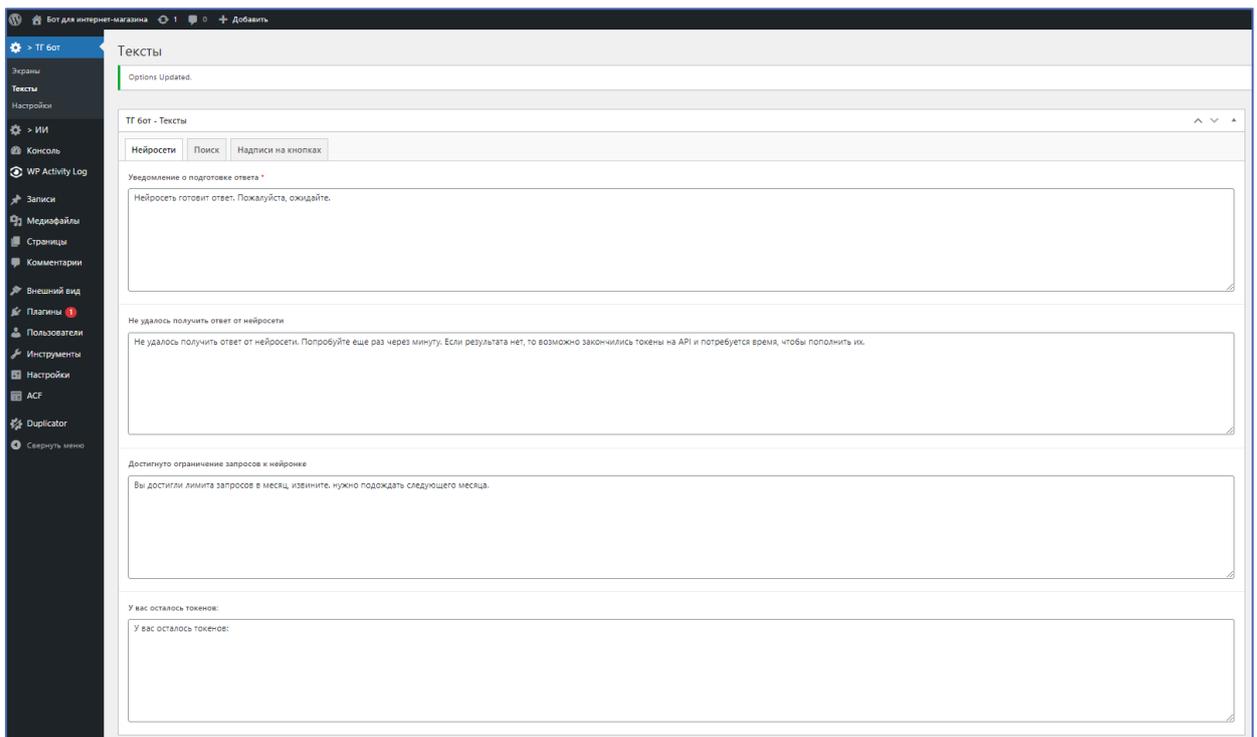
- оформление экранов цветами,
- вставка названия экрана и его команды в плашку имени,
- назначение кнопок по типу пользователя (Сотрудник или Пользователь),
- возможность отключить кнопку или экран без их физического удаления.
- Тип экрана «Нейросеть»



НАСТРОЙКИ БОТА В МЕНЮ «ТЕКСТЫ»

Список настроек «Тексты» разделен визуально на вкладки. Все указанные ниже тексты владелец бота может менять самостоятельно. Далее подробно о каждом блоке текстов.





| Название настройки | Описание настройки |
|--|--|
| Уведомление о подготовке ответа * | Какой текст показывать пользователю, когда нейросеть приступила к анализу и подготовке ответа. Это момент сразу после того, как бот отправил данные в нейросеть. |
| Не удалось получить ответ от нейросети | По любой причине может нейросеть не вернуть человеческий ответ, с уведомлением об ошибке. В таком случае пользователю показывается сообщение из этого поля, а администраторам направляется сообщение о причине. Причины могут быть: <ul style="list-style-type: none"> • Превышение длины диалога, включая текст для роли System • Исчерпан лимит токенов • Ошибка сервера (на стороне нейронки) |
| Достигнуто ограничение запросов к нейронке | Сообщение из этого поля отобразится пользователю, если у него закончились лимиты. |
| У вас осталось токенов: | Если у отдельного пользователя кончились лимиты, то будет показано это сообщение |
| Недостаточно данных для поиска | Текст из этого поля будет показан пользователю, если он попал на экран поиска, но невозможно установить в каких местах производить поиск. На 5.12.23 к такой ошибке может привести отсутствие указания на таксономия, идентификатор категории, тип записи. |
| Нет совпадений по вашим критериям поиска | Если параметры поиска указаны, но по ним невозможно найти никакую запись. |

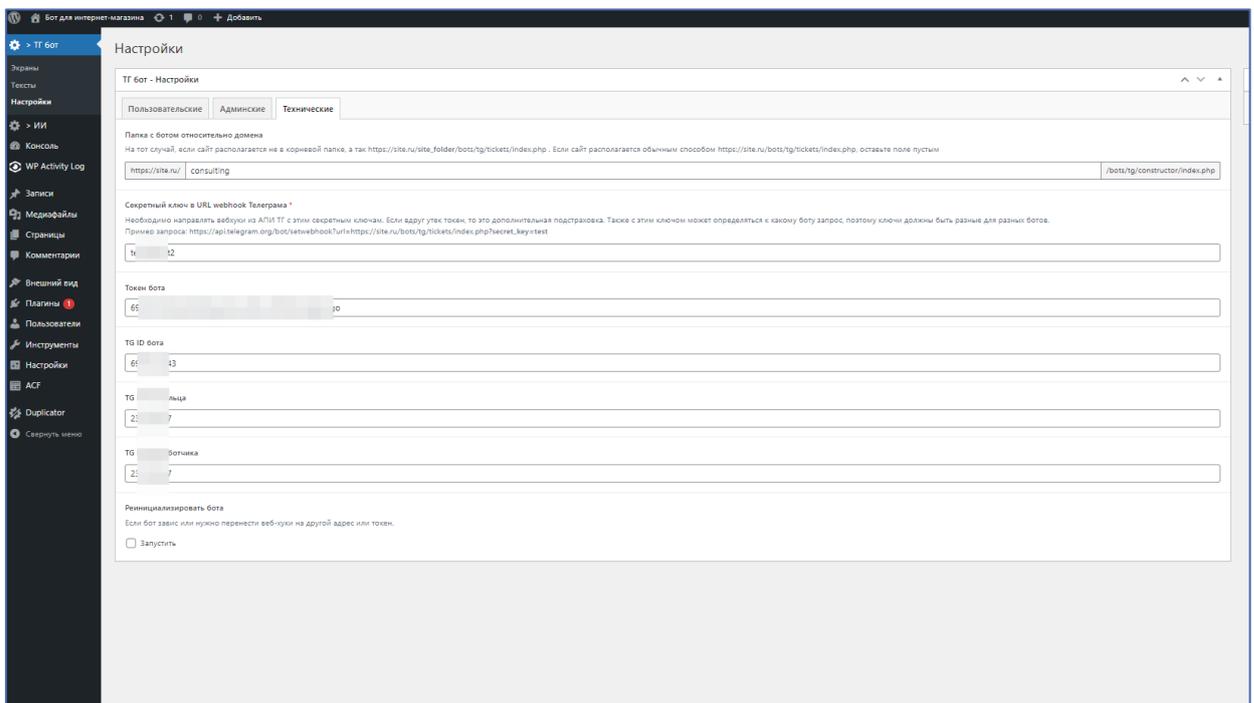
| | |
|---|--|
| Все поля заполнены, кнопка искать активна | Такое поле появляется после того, как на основе данных от нейросети были установлены все необходимые параметры для осуществления поиска. Данное сообщение может появиться и в ситуации, когда пользователь перешел к работе с нейронкой, но имел к этому моменту ранее заполненные параметры поиска в данной категории. |
| Искать | Надпись на кнопке искать, которая появляется после ... см. пункт выше. |

НАСТРОЙКИ БОТА В МЕНЮ «НАСТРОЙКИ»

Список настроек «Настройки» разделен визуально на:

- Пользовательские
- Админские
- Технические

Все указанные ниже настройки владелец бота может менять самостоятельно.



| Название настройки | Описание настройки |
|---|---|
| Количество последних действий пользователя для хранения | Сколько системе запоминать последних действий пользователя. Это может пригодиться для, например, кнопки «назад», которую может реализовать ваш разработчик. Сейчас этой кнопки нет, код написан без ее учета. |

| | |
|--|--|
| Данные сотрудников компании | В этой настройке следует указывать TG ID всех, кто должен получать некоторые системные сообщения, в частности, об ошибке со стороны нейронки. |
| Папка с ботом относительно домена | На тот случай, если сайт располагается не в корневой папке, а так https://site.ru/site_folder/bots/tg/constructor/index.php . Если сайт располагается обычным способом https://site.ru/bots/tg/constructor/index.php , оставьте поле пустым |
| Секретный ключ в URL webhook Телеграма | Необходимо направлять вебхуки из АПИ ТГ с этим секретным ключам. Если вдруг утек токен, то это дополнительная подстраховка. Также с этим ключом может определяться к какому боту запрос, поэтому ключи должны быть разные для разных ботов. |
| Токен бота | - |
| TG ID бота | - |
| TG ID владельца | - |
| TG ID разработчика | - |
| Реинициализировать бота | Опция если бот завис или нужно перенести веб-хуки на другой адрес или токен. |

«ИИ – НАСТРОЙКИ»

ИИ – НАСТРОЙКИ – CHATGPT

| Название настройки | Описание настройки |
|---|---|
| Токен chatGPT | - |
| Модель chatGPT | Если установить в этом параметре модель на 16К знаков, то нейронка будет принимать очень длинные тексты. По умолчанию стоит gpt-3.5-turbo на 4К знаков. |
| Всего токенов разрешено при запросе к GPT * | Суммарное количество токенов (зависит от модели) при запросе к нейронке, учитывая исходящую длину диалога, длину ответа, длину текста System. Если установить другую модель, то это количество можно поменять. |
| Модификатор для конвертации символов в токены для GPT * | Модификатор позволяет определить, на какое число делить общую длину символов в диалоге (включая текст роли System), чтобы не превысить максимальную длину запроса к нейросети. |
| Максимальное количество токенов в ответе chatGPT | То есть вы разрешаете нейронке дать ответ, в котором будет определенное количество токенов, и она не будет превышать это число, хотя может и обрезать при этом ответ. |
| Температура ответа chatGPT | temperature - используется для контроля случайности и креативности сгенерированного текста. Более высокое значение температуры приводит к более разнообразным и креативным результатам, тогда как более низкое значение производит более сфокусированные и детерминированные ответы. (принимает значение от 0 до 1) |
| Повторение слов и фраз в ответе ии | presence_penalty - влияет на вероятность того, что модель повторит слова или фразы в своем ответе. Более высокое значение этого параметра уменьшает вероятность повторений, тогда как более низкое значение позволяет использовать больше повторений (от -2 до 2) |
| Использование распространенных слов и фраз | frequency_penalty - контролирует предпочтение модели использовать более распространенные слова или фразы. Более высокое значение этого параметра поощряет использование менее распространенных слов, тогда как более низкое значение отдает предпочтение более распространенным словам (от -2 до 2) |
| URL запроса к chatGPT | - |



ИИ – НАСТРОЙКИ – GIGACHAT

| Название настройки | Описание настройки |
|--|--|
| Токен client ID GigaChat | |
| Токен client secret GigaChat | |
| Токен access data GigaChat | От этого ключа происходит генерация основного токена авторизации. Подробности: https://developers.sber.ru/docs/ru/gigachat/api/integration-individuals |
| Тип аккаунта (scope) GigaChat | Тип аккаунта: бизнес или частный |
| Модель GigaChat | - |
| Всего токенов разрешено при запросе к GigaChat * | Суммарное количество токенов (зависит от модели) при запросе к нейронке, учитывая исходящую длину диалога, длину ответа, длину текста System |
| Модификатор конвертации символов в токены для GigaChat * | Модификатор позволяет определить, на какое число делить общую длину символов в диалоге (включая текст роли System), чтобы не превысить максимальную длину запроса к нейросети. |
| Максимальное количество токенов в ответе GigaChat * | То есть вы разрешаете нейронке дать ответ, в котором будет определенное количество токенов, и она не будет превышать это число, хотя может и обрезать при этом ответ. |
| Температура ответа GigaChat | temperature - Температура выборки в диапазоне от нуля до двух. Чем выше значение, тем более случайным будет ответ модели. |
| Повторение слов и фраз в ответе GigaChat | repetition_penalty, количество повторений слов: При значениях 0 до 1 модель будет повторять уже использованные слова. Значение 1.0 — нейтральное значение. |

| | |
|---------------------------------------|--|
| | При значении больше 1 модель будет стараться не повторять слова. |
| URL запроса к GigaChat | - |
| URL регенерации access token GigaChat | - |

ПРОЧИЕ НАСТРОЙКИ ИИ

| | |
|---|--|
| Сколько сообщений хранить * | Не используется. Для обрезки длинных диалогов скрипт прибегает к настройке общего числа токенов в запросе и числа токенов в ответе нейронки. |
| Лимиты запросов от пользователя (в токенах) * | Обновляется раз в месяц. |



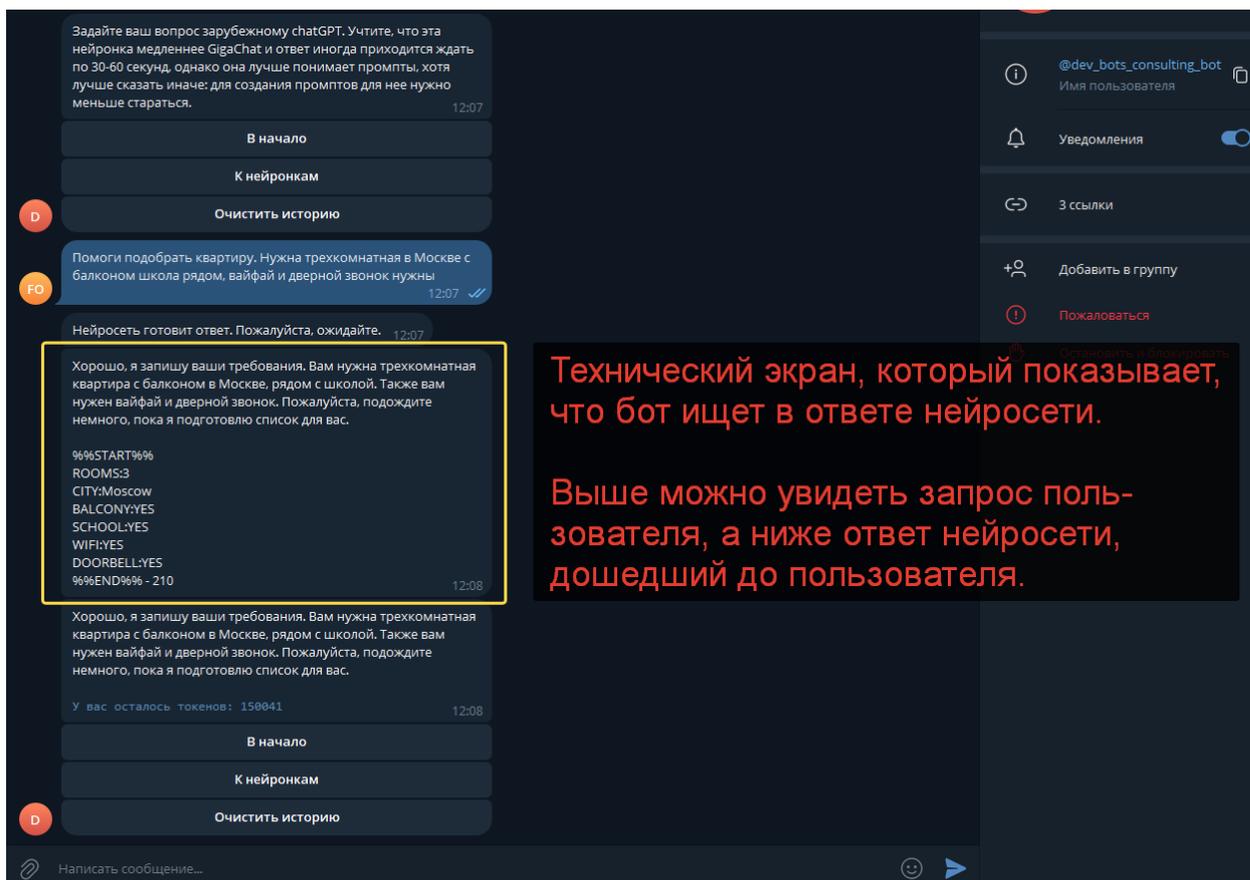
Итак, этот бот имеет роль консультанта и подборщика. Как это работает на схематично:



Таким образом, следует сделать вывод, что нейросети служат в данном боте переводчиком (фильтром) словестного хаоса в технический язык. Сами нейросети не работают с базами данных, да и это невозможно из-за объема информации в БД.

Однако, это лишь сухое изложение. На самом деле, то, что делают нейросети сильно упрощает жизнь и разработчику, и владельцу и пользователю. Единственный, кому жизнь усложняется – создателю промптов.

Технический пример того, что бот ожидает (это не видно пользователю – скриншот для ясности) от нейросети на скриншоте ниже.



В примере-скриншоте выше имитируем подбор квартиры для покупки. Допустим, пользователь захотел изменить требования к школе, вайфай, дверному звонку и балкону с «требуется» на «не требуется». Чтобы не устраивать пошаговую сессию вопрос-ответ пользователь может написать «Я передумал, сделай эти требования не обязательными» и нейронка старается понимать, что каждый YES нужно изменить на NO. Опять же, если промпт доходчиво написан.

Это лишь малая часть возможностей интерпретации человеческого произвольного языка в рамках технических возможностей данного бота.

ФУНКЦИОНАЛ КОНСУЛЬТАНТА

Консультант и подборщик в коде не сильно отличаются. Трактовка зависит от настроек, где среди главных – текст для роли System.

Возьмем модель GPT 3.5 на 16K. Если сделать приблизительное преобразование на русский язык, то это около 8000 слов. То есть вы можете создать экран типа «Нейросеть» для отдельной группы товаров, для отдельного товара, дать обширную инструкцию на 2-5K токенов и пусть нейронка отвечает на все вопросы, пользуясь подсказками, которые вы описали в инструкции.

Тем не менее, есть одно маленькое отличие между консультантом и подборщиком: **переключатель режима работы**. Вы можете задать работу в режиме чата или в режиме консультанта. В первом случае будет происходить простое общение (на заданную тему) с тем же текстом для роли System, с сохранением диалогов. Во втором случае бот будет дополнительно направлять сценарий работы кода по пути вычленения технических параметров, сохранения их настройках пользователя. Но об этом ниже.

СПИСОК ПАРАМЕТРОВ ЭКРАНА «НЕЙРОСЕТЬ»

Весь функционал консультанта базируется на дополнительном типе экрана под названием «Нейросеть». Также используется некоторый функционал от базовых типов экранов, вроде «Текст экрана», «Кнопки экрана».

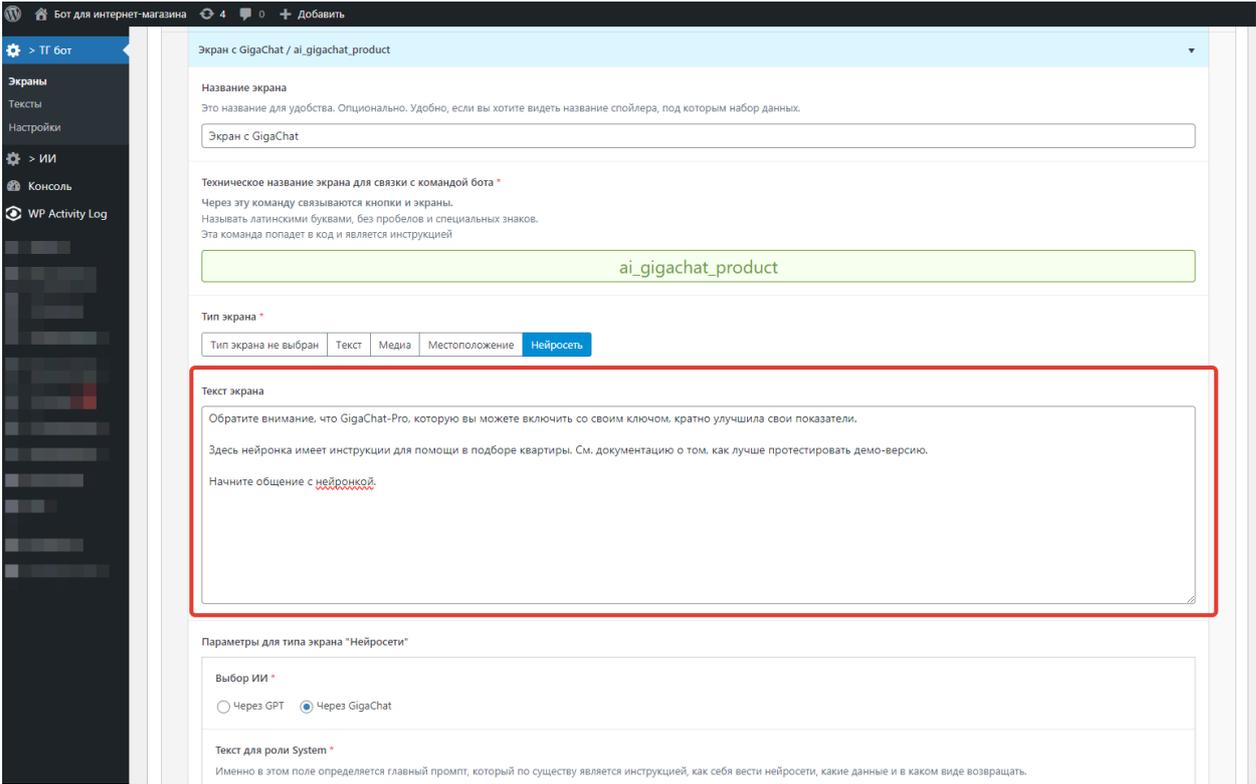
Что такое экран: это то, что вы видите в боте: текст сообщения от бота, кнопки под сообщением.

Тип экрана «Нейросеть» открывает некоторые специализированные настройки, позволяющие тонко настраивать работу нейросети.

Обратите внимание, что вы можете создать несколько экранов с типом «Нейросеть», настроить каждый экран для работы под определенные условия, например, для определенной категории, для определенного типа записи, для определенного режима, с определенной инструкцией и так далее.

Ваши возможности действительно ограничены только фантазией и умением работать с генеративными текстовыми моделями. Вы можете настроить вплоть до того, чтобы одна нейросеть рассказывала байки в режиме простого чата, а вторая была строго по тригонометрии, третья помогала выявлять потребности относительно искомых товаров, а третья консультировала по конкретному товару, описание которого вы передадите в роль System.

ТЕКСТ ЭКРАНА



The screenshot shows the configuration interface for a bot screen. The left sidebar contains navigation options: 'ТГ бот', 'Экраны', 'Тексты', 'Настройки', '> ИИ', 'Консоль', and 'WP Activity Log'. The main content area is titled 'Экран с GigaChat / ai_gigachat_product'. It includes the following fields:

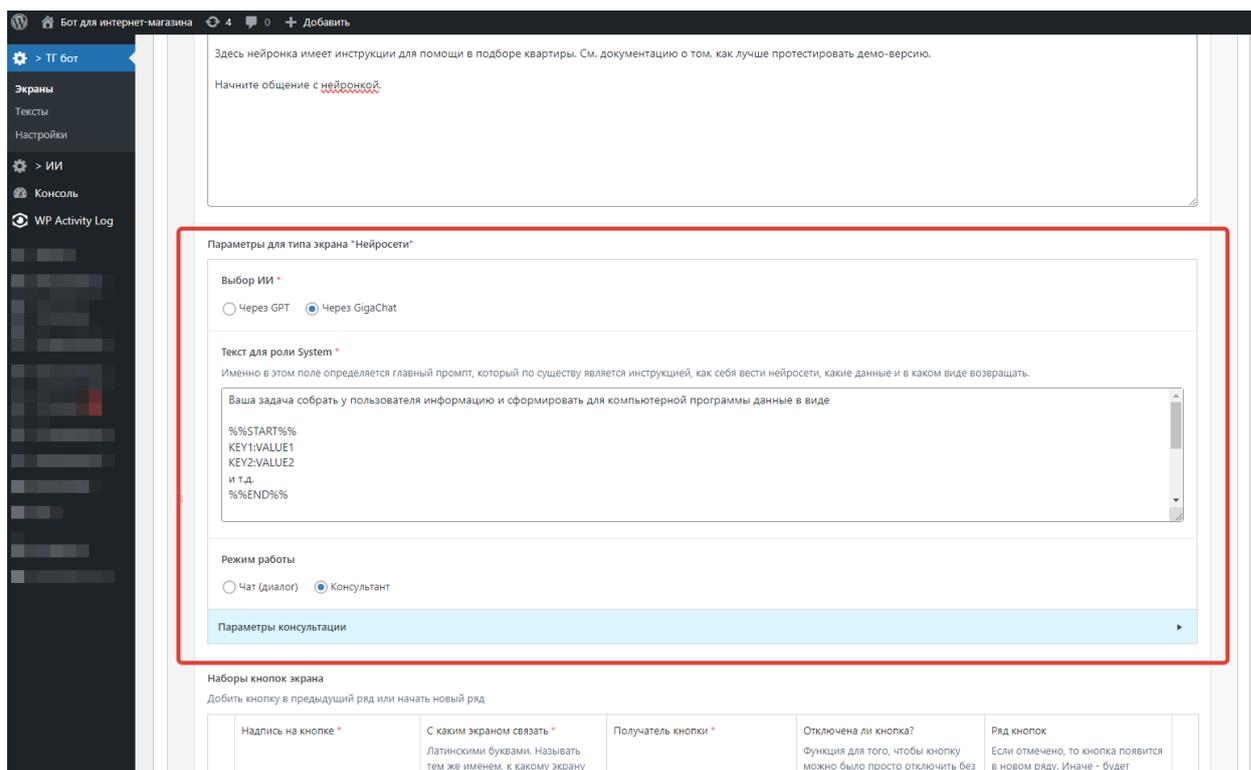
- Название экрана:** 'Экран с GigaChat'
- Техническое название экрана для связи с командой бота:** 'ai_gigachat_product'
- Тип экрана:** 'Нейросеть' (selected)
- Текст экрана:** A text area containing the following text:

Обратите внимание, что GigaChat-Pro, которую вы можете включить со своим ключом,кратно улучшила свои показатели.
Здесь нейронка имеет инструкции для помощи в подборе квартиры. См. документацию о том, как лучше протестировать демо-версию.
Начните общение с нейронкой.
- Параметры для типа экрана "Нейросети":**
 - Выбор ИИ:** 'Через GigaChat' (selected)
 - Текст для роли System:** 'Именно в этом поле определяется главный промпт, который по существу является инструкцией, как себя вести нейросети, какие данные и в каком виде возвращать.'



Здесь вы устанавливаете тот текст, который появится перед пользователем при его переходе на экран взаимодействия с нейросетью. Это может быть приглашение ввести какой-то текст, объяснение того, как работает нейросеть, инструкция для пользователя, что нужно сделать.

ПАРАМЕТРЫ ДЛЯ ТИПА ЭКРАНА "НЕЙРОСЕТИ"



Все параметры для настройки нейронки выведены в отдельную область «Параметры для типа экрана "Нейросети"»

ВЫБОР ИИ

См. скриншот выше. Этим переключателем вы решаете, через какую нейросеть пропускать запросы от пользователя. Удобно, если вам нужно моментально переключить на запасную или провести быстрые тесты.



ТЕКСТ ДЛЯ РОЛИ SYSTEM

См. скриншот выше. Успех работы бота зависит на 80-90% от этой настройки. Именно в этом поле определяется главный промпт, который по существу является инструкцией, как себя вести нейросети, какие данные и в каком виде возвращать.

Пример инструкции, размещенной в этом поле в демо-боте на 5.12.23 следующий (с 14.12.23 - для GigaChat другой – проще):

Ты - фильтр. Твоя задача предоставить ТОЧНУЮ информацию для скрипта, которому разрешено общаться ТОЛЬКО ПО ТЕМЕ АРЕНДЫ ИЛИ ПРОДАЖИ КВАРТИРЫ, а остальные темы ты игнорируешь. Твоя задача вы выявить конкретные параметры, чтобы вернуть их моему боту. Сделай это, задавая следующие вопросы:

Список требований:

%%CONS_DATA%%

Необходимые данные верни в виде КЛЮЧ:ЗНАЧЕНИЕ. На основе предоставленных тобой КЛЮЧ:ЗНАЧЕНИЕ мой скрипт подготовит нужный ответ. НЕ ПИШИ В ЗНАЧЕНИИ что-то в скобках или дополнительную информацию, потому что значение нужно для обработки скриптом.

ВНИМАНИЕ! Раздели свой ответ на две части:

Первая часть: ты по-человечески ответишь пользователю или задашь дополнительные вопросы. Вторая часть: вторая часть является технической и ее НУЖНО ОФОРМИТЬ в теге %%START%% и %%END%%, между тегами расположив ответы пользователя, преобразованные в КЛЮЧ:ЗНАЧЕНИЕ. БЕЗ ВТОРОЙ ЧАСТИ НЕ ВОЗВРАЩАЙ ОТВЕТ!

ПРИМЕР ПРАВИЛЬНОГО ОТВЕТА:

Спасибо, ваши данные приняты. Ваши данные следующие: количество комнат - 5, школа рядом - да, наличие wifi в квартире - нет, наличие дверного звонка - да. Эта информация корректна? Хотели бы вы добавить что-то еще?

%%START%%

ROOMS:5

SCHOOL:YES

WIFI:NO

DOORBELL:YES

%%END%%

ПРИМЕР НЕПРАВИЛЬНОГО ОТВЕТА:

Спасибо за предоставленные данные. Вот данные, которые я получил от вас:

Количество комнат: 3

Наличие школы рядом: Да

Наличие WIFI: Нет

Спасибо за предоставленную информацию! Я начну поиск квартир, удовлетворяющих вашим требованиям. Пожалуйста, подождите немного, пока я подготовлю список для вас.

Точно можно сказать, что данная инструкция для нейронки не совершенная, но разработчик специализируется на программировании, а не составлении текстов для нейронок.



Feelosoov: <https://kwork.ru/user/feelosoov>

РЕЖИМ РАБОТЫ

Здесь два параметра:

1. Чат (диалог)
2. Консультант.

В первом случае будет происходить простое общение с тем же текстом для роли System, с сохранением диалогов. Во втором случае бот будет дополнительно направлять сценарий работы кода по пути вычленения технических параметров, сохранения их в настройках пользователя.

Что происходит в режиме консультанта

The screenshot shows the configuration interface for a Telegram bot in 'Консультант' mode. The left sidebar contains navigation options like 'Экраны', 'Тексты', 'Настройки', and 'Пользователи'. The main area is titled 'Параметры консультации' and includes a 'Параметры категорий' section with input fields for 'Таксономия *' (category), 'ID таксономии *' (1), and 'Тип записи *' (post). Below this is a 'Режим работы с ключами *' section with a dropdown menu for 'Запрос у нейросети'. The bottom section, 'Ключи полей для сравнения', contains a table with columns for 'Обязательное', 'Ключ записи *', 'Ключ сопоставления *', and 'Запрос ключа сопоставления'. The table lists three keys: 'rielt_rooms' (ROOMS), 'rielt_school' (SCHOOL), and 'rielt_wifi' (WIFI), each with a corresponding prompt for user clarification.

| Обязательное | Ключ записи * | Ключ сопоставления * | Запрос ключа сопоставления |
|-------------------------------------|--|---|--|
| <input checked="" type="checkbox"/> | rielt_rooms Название ключа записи с дополнительной информацией, на латинице. Это типовая линейная сущность WordPress типа string или integer. | ROOMS Этот ключ нужен, чтобы сопоставить ключ записи с ответом от нейросети. К этому ключу нейронка крепит значение, этот ключ вы указываете в промпте роли System, значение этого ключа бот присваивает ключу записи. | Используйте псевдокод %%CONS_DATA%% в поле "Текст для роли System", чтобы в указанном месте вставить описанные в соответствующих полях требования. Также вы можете сформулировать вопросы в этих полях для пользователя. Зависит от выбора "Режим работы с ключами". Уточни у пользователя требуемое количество комнат для ключа ROOMS. Ответ пользователя интерпретируй, как точную цифру, или диапазон через дефис, например, ROOMS:1 или ROOMS:3-5 и не пиши |
| <input checked="" type="checkbox"/> | rielt_school | SCHOOL | Уточни у пользователя необходимость школы для ключа SCHOOL. Ответ пользователя интерпретируй, как YES или NO. Например, SCHOOL:YES или SCHOOL:NO |
| <input checked="" type="checkbox"/> | rielt_wifi | WIFI | Уточни у пользователя необходимость WIFI для ключа WIFI. Ответ пользователя интерпретируй, как YES или NO. Например, WIFI:YES или WIFI:NO |

Когда нейросеть возвращает ответ, то он проходит через дополнительную обработку. Ее суть заключается в поиске определенных ключей в ответе нейронки. Соответственно, нейронка должна/обязана вернуть эти ключи, причем в определенном псевдокоде.

Режим консультанта позволяет сравнить, все ли из обязательных значений нейронка сформулировала в виде ключей. Конечно, нейронка может упрямо не отправлять нужный формат ответа и такое наблюдалось во время тестов в 5-10% случаев. Здесь нужно решение лексического характера – не через код, а через промпт (поле Текст для роли System).

Когда бот получил все необходимые ключи, то записывает (закрепляет) их за пользователем, а бот предоставляет нужную кнопку пользователю. Пользователь переходит к экрану поиска и бот осуществляет поиск (подбор) необходимых товаров, услуг, показывая на экране доступные данные.

Обратите внимание на маленькую деталь в длинной инструкции выше: **%%CONS_DATA%%**. Вместо этого псевдокода бот подставляет текст из полей «Запрос ключа сопоставления» по каждому искомому ключу.

ПАРАМЕТРЫ КОНСУЛЬТАЦИИ

ПАРАМЕТРЫ КАТЕГОРИЙ

Сюда отнесены настройки полутехнические. Это может показаться сложным совершенному новичку, но все очень просто.

1. Таксономия.
2. ID таксономии.
3. Тип записи.

Сложно объяснить простыми словами, поэтому разработчик попросил объяснить GPT:

Таксономия:

Пример: Разделение книг по жанрам (фантастика, детективы, романы).

Аналогия: Если ваш сайт - это библиотека, то таксономия - это система классификации жанров книг. Таким образом, книги легко группируются и находятся в определенных разделах.

ID таксономии (термина):

Пример: Уникальный номер или идентификатор жанра (например, ID для фантастики может быть 1, для детективов - 2).

Аналогия: Каждый жанр книги получает свой уникальный номер, чтобы можно было точно определить, о каком именно жанре идет речь. Например, если вам нужна фантастика, вы просто уточняете по ее уникальному номеру (ID).

Тип записи:

Пример: Определение, является ли книга романом, статьей или справочным материалом.

Аналогия: Тип записи - это указание на то, какую именно информацию содержит каждая книга. Если это роман, вы знаете, что это художественное произведение, а если это справочный материал, то это, вероятно, информация для обучения.

Как бы там ни было, если вы погрузитесь и поэкспериментируете, то увидите, что возможность самостоятельно определять эти параметры дает много больший простор для управления. Особенно, если ваш сайт на WordPress уже имел какие-то настройки, он был готов и нужно данного бота внедрить аккуратно. Для облегчения работы установлены параметры по умолчанию.

РЕЖИМ РАБОТЫ С КЛЮЧАМИ

Здесь на выбор дается два параметра, но рабочий на 5.12.23 только один – запрос у нейросети. Второй параметр предполагал вынесение части работы по выявлению потребности пользователя на самого пользователя. Что-то вроде вопрос-ответ, при котором от нейросети требуется почти только приветствие и итог. Возможно, второй параметр будет реализован позже (пока нет спроса). Пока же все запросы режим консультанта старается направлять на нейросеть.



КЛЮЧИ ПОЛЕЙ ДЛЯ СРАВНЕНИЯ

Очень интересный и важный набор параметров. Нужно задаться вопросом. Бот является просто (и простым) кодом, он не интерпретирует человеческую речь, но он должен выдавать пользователю результаты-записи, подходящие по критериям. Как это сделать?

Любая запись в базе данных должна иметь характеристики. Будь-то автомобиль, одежда, квартира, примеры выполненных работ (портфолио) или что-то иное. Возьмем квартиру: у нее есть параметр «количество комнат», «рядом ли школа» и прочие. Каждый параметр имеет свой уникальный ключ, по которому его можно найти в базе данных. У каждой записи таких ключей может быть неограниченно много.

Следовательно, мы каким-то образом должны через эти ключи синхронизировать бота, нейросеть и пользователя. Как? Один ответ – через промежуточные данные, которые понимают все три стороны. Итак, нейросеть понимает пользователя, бот понимает нейросеть, пользователь понимает ответы бота и нейросети.

Возьмем связку параметров Ключ записи – Ключ сопоставления – Запрос ключа сопоставления:

Здесь мы указываем, ключ под которым хранится параметр у записи. Например, под ключом «rielt_rooms» в записи хранится сколько комнат. Через ключ сопоставления ROOMS бот связывает rielt_rooms и ROOMS. Через ключ сопоставления ROOMS нейронка передает ответ пользователя.

Условный пример:

Пользователь сообщает: «Мне нужна квартира четырехкомнатная». Причем может написать с ошибками, в виде цифры и так далее. Нейросеть, тем не менее, это понимает и сохраняет в ключ ROOMS значение 4. Далее бот ищет ключ ROOMS, получает из него значение 4 и сохраняет в rielt_rooms у пользователя в параметрах поиска. Когда пользователь ищет квартиру, то это значение вставляется в условие поиска и бот находит запись у которой в rielt_rooms тоже есть значение 4.

Вся прелесть заключается в том, что таких связок вы можете указывать сколько угодно и называть как угодно, а также описывать детали «Запроса ключа сопоставления» более конкретно, чтобы нейросеть лучше понимала задачу. Например, в демо-версии на 5.12.23 указан запрос ключа сопоставления так:

Уточни у пользователя требуемое количество комнат для ключа ROOMS. Ответ пользователя интерпретируй, как точную цифру, или диапазон через дефис, например, ROOMS:1 или ROOMS:3-5 и не пиши здесь длинные слова.

Главное, чтобы аналогичные ключи реально у записи были.

КАК ПОПОЛНЯТЬ БАЗУ ДАННЫХ

Вопрос логичный. Ответ простой: бот сделан на основе WordPress, а потому есть решения.

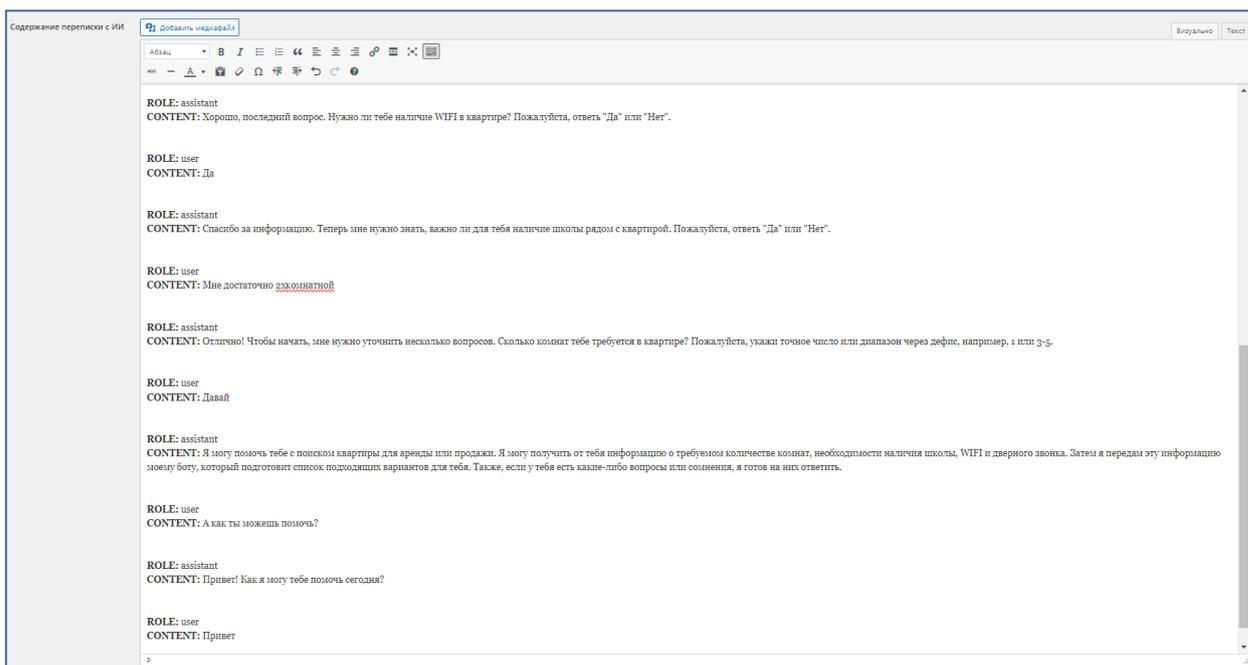
- Встроенные инструменты импорта в формате XML.
- Установка тонконастраиваемых плагинов для импорта из CSV, XLS, XML, TXT.
- Создание API под конкретные потребности.
- Создавать вручную, если штучные товары, но к ним применима работа с нейросетью (продажа какой-нибудь подложки, создание ТЗ для сайта, заказ мебели).

При достаточном спросе разработчик бота может реализовать общий API, подходящий под большинство условий.



СОХРАНЕНИЕ ДИАЛОГОВ

Все диалоги сохраняются в профилях каждого пользователя и их можно изучить для улучшения понимания работы нейросети, для ведения аналитики, для уточнения деталей. Сохраняется та длина диалога, которая может быть принята нейросетью по критерию «длина», которая в свою очередь определяется максимальным количеством токенов в настройках нейросетей.



ТЕХНИЧЕСКИЕ ТРЕБОВАНИЯ

Основные:

1. Бот был протестирован на версии PHP 8.2, на хостинге Бегет, на нем же работает демонстрационная версия. Поэтому работа проверена с этой версией и на этом хостинге. С прочими хостерами вы решаете самостоятельно.
2. Библиотека для работы с функциями *mb_* должна быть подключена, например, *mb_strlen()*:
3. База данных должна иметь поддержку *utf8mb4_unicode_ci*.
4. SSL-сертификат (для *https* в адресной строке).
5. Домен второго уровня или поддомен. Бота также можно разместить в подпапку, вида сайт внутри сайта.
6. Бот на вебхуках.

Ограничения разные:

1. На стороне кода обрезаются теги HTML и PHP.

Ограничения ТГ примерные:

С такими ограничениями ничего не поделаться на стороне разработчика, в рамках данного бота.

- Размер фото не должен превышать 5Мб.
- Размер видео не должен превышать 50Мб (на деле - 10Мб).
- За один раз (в одном экране) нельзя отправить более 10 медиа
- За один раз (в одном экране) нельзя отправить разные типы медиа. Должны быть или все фото, или все видео, или все документы и т.д.
- Ссылки на медиа можно вставлять, как со своего сайта с ботом, так и с других сайтов.
- Телеграм уверенно работает с форматами MP4, PDF, JPG, PNG, ZIP. За остальные форматы он не ручается.
- Обратите внимание, что у Telegram свое понимание, что такое видео. Иногда вместо реального видео формата mp4 Telegram считает, что ему прислали анимацию. Причины разные: видео без звука, файл видео слишком маленького размера или видео слишком короткое. Учитывая, что в одной партии медиа (если их больше 1-го) должны быть одинаковые типы медиа, то весьма трудным оказывается найти два файла видео, которые Telegram принял бы за один тип – видео. Говоря кратко: если отправляете видео, то лучше отправляйте по одному.

Это не все ограничения, но наиболее значимые для определения ответственности разработчика. Telegram может их добавлять/менять без отдельного уведомления. Подробнее и точнее на странице Телеграма: <https://core.telegram.org/bots/api#recent-changes>

Ограничения на скрипт:

- Запрещается бота продавать в целом виде или частично.
- Запрещается передавать безвозмездно, включая аффилированным лицам.
- Скрипт бота будет частично зашифрован.
- Лицензия на ACF Pro не передается, но плагин работать будет. При необходимости, покупатель самостоятельно, под свой домен приобретает лицензию на этот плагин.

Что вам будет передано:

Вам передается весь код бота, как он есть у меня, используемый для демонстрации, а также WordPress со всеми предустановленными темами и плагинами, включая работоспособный и официальный(!) ACF Pro (без кода лицензии) версии на момент разработки, включая настройки WordPress, созданные через плагины и код.

МИНИМАЛЬНЫЕ ОБЕЩАНИЯ РАЗРАБОТЧИКА

Бот будет точно хорошо работать, при соблюдении технических требований и условий, включая набор плагинов на сайте разработчика, хостинг, оборудование.

Одновременно может (но не обязан) обрабатывать до 30 запросов из Телеграма на отдельных потоках. При отправке запроса в нейросеть запрос из Телеграм обрывается намеренно, чтобы не висел веб-хук в ожидании.

На каждый запрос к боту (не к нейросети) уходит 0.08-0.3 секунды при текстовом сообщении, при одном исходящем фото - 0.3-0.8, при видео сложно сказать, но ориентируйтесь на значение от 0.7 сек и выше (не считая загрузки от пользователя в Телеграм), потому что зависит от скорости между сайтом и Телеграмом, потому что зависит от скорости обработки на стороне сайта, от размера видео.



Рекомендации по оптимальному использованию:

1. Запросы 2-3 в секунду максимум.
2. В конструкторе рекомендованное количество экранов – до 30.

Если отправляется в бот больше 5 запросов одновременно, то примите поздравления: «у вас много клиентов» и это причина, чтобы ваш разработчик переписал небольшим усилием код так, чтобы загружать минимальную версию WordPress, что даст ускорение до 0.04-0.08 на текстовых сообщениях. Или даже написал свою отдельную систему с отдельным ботом.

Разработчик бота настоятельно рекомендует ознакомиться с документацией chatGPT и GigaChat, а также понять, как составлять промпты, чтобы точно понимать, что и как должно работать. Например, для физлиц в ГигаЧат выдается 1 000 000 токенов сроком на 12 месяцев, а не на каждый месяц и запросы обрабатываются в одном потоке на аккаунте физлица. Например, ЧатГПТ никогда и ни за что не несет ответственности.

