

Интеграция классификатора эмоций в архитектуру диалоговой системы для корректировки текстовых ответов

И. С. Косачев • Е. Ю. Сазонова

Уфимский университет науки и технологий

В статье рассматривается задача корректировки ответов чат-бота с учетом эмоционального состояния пользователя без дообучения языковой модели. Предложен подход, при котором отдельная мультимодальная модель определяет состояние человека по голосу и речи, после чего полученная метка объединяется с запросом и подается в GigaChat-2 от Сбера для управления тоном генерации. Проведенное ручное тестирование показало, что передача этой информации позволяет боту генерировать более персонализированные и эмпатичные ответы. Данный метод работает без дополнительного обучения, что значительно уменьшает вычислительные затраты на коррекцию ответов модели

Эмоциональный интеллект; коррекция поведения языковых моделей; человеко-компьютерное взаимодействие; управление контекстом; машинное обучение; мультимодальные данные.

ВВЕДЕНИЕ

С каждым годом взаимодействие человека и машины становится все более тесным. После появления в 2022 году ChatGPT произошли большие изменения в области организации человеко-машинного взаимодействия. Если раньше чат-боты были основаны на жестко установленных правилах, то сейчас они работают на основе больших языковых моделей (large language models, LLMs) и генерации с дополненной выборкой (retrieval augmented generation, RAG) на основе векторных баз данных, которые служат хранилищем знаний о предметной области. Такие боты умеют хорошо искать информацию, структурировать и выдавать в удобном и понятном для пользователя виде.

Этого может быть достаточно в случае, если бот используется как умный поисковик, но, когда дело доходит до общения с людьми, для эффективного взаимодействия этих функций может не хватить.

Эмоция – психофизиологическое состояние человека, отражающие его субъективное отношение к некоторой ситуации или объекту. Они могут выражаться человеком как с помощью вербальных средств коммуникации, так и с помощью невербальных. Вербальные средства коммуникации включают в себя устную и письменную речь человека, а невербальные мимику, язык телодвижения, тон голоса, жесты и др. Благодаря эмоциям мы можем понимать чувства других людей и выстраивать с ними более эффективное взаимодействие.

При взаимодействии с машиной можно испытывать те же самые чувства и эмоции, что и при общении с человеком. В таком случае, для поддержания естественного диалога чат-бот должен иметь информацию об эмоции пользователя и уметь реагировать на нее. Во время

общения с чат-ботом пользователи преимущественно используют текст, а значит, в большинстве случаев бот может анализировать только вербальные эмоциональные сигналы. Алан и Барбара Пиз [Пиз24] упоминали, что во время общения человек только 40% информации передает вербально, а остальные 60% невербально. Тогда получается, что для понимания запроса машина получает менее половины информации для ответа. В таком случае, для полноты понимания потребности пользователя машине важно учитывать его невербальные сигналы.

Изучением подходов к корректировке поведения чат-ботов для эмпатичного общения занимались Л. Сиян и др. [Siy21], Ц. Лонгран [Lon25], С. Абинаия и др. [Abi25].

Параллельно с этим в отечественной практике сформировался фундамент для проектирования программных комплексов и вычислительных алгоритмов фиксации аффекта; в частности, базовые методы глубокого обучения для оценки эмоционального фона и принципы построения соответствующих систем поддержки принятия решений детально исследованы Р. А. Шириновым [Шир25], И. С. Латкиным [Лат24], Е. Ю. Сазоновой [Саз23] и их коллегами. Кроме того, успешный опыт программной обработки узкоспециализированных изолированных каналов данных – от термографических снимков лица до графических характеристик подписи оператора – представлен в трудах С. С. Жумажановой [Жум23] и А. Е. Самотуги [Сам23]. Тем не менее в существующих разработках все еще наблюдается дефицит сквозных архитектурных решений, способных синхронизировать и объединять асинхронные мультимодальные потоки непосредственно в контуре реального времени.

В данной статье рассмотрены подходы корректировки поведения чат-бота путем добавления информации об эмоциональном состоянии человека на основе вербальных и невербальных сигналов, а также описана реализация одного из них с использованием мультимодальной нейронной сети и модели GigaChat-2 от компании Sber.

СОВРЕМЕННОЕ СОСТОЯНИЕ ПРОБЛЕМЫ

Л. Сиян и др. [Siy21] учили модель вести эмоционально поддерживающую беседу, основываясь на теории помогающих способностей Клары Хилл [Hil09]. Идея подхода заключается в том, чтобы обучить модель генерировать запрос с учетом стратегии общения, состоящей из этапов:

- исследование – определение проблемы пользователя;
- утешение – утешение пользователя посредством выражения эмпатии;
- действие – помощь в решении проблемы.

На этапе обучения генерация дополнительного токена в виде контекста позволяет модели самостоятельно определить стратегию. Результат тестирования модели показал, что такой подход обеспечивает возможность вести более комфортную беседу для пользователя, чем обычная генерация без учета стратегии. Стоит отметить, что в данном подходе нет прямого использования информации об эмоциональном состоянии пользователя.

Ц. Лонгран [Lon25] изучал способы корректировки поведения с помощью промптинга. Суть метода заключается в том, чтобы выбирать промпт для генерации в зависимости от ситуации, в которой находится пользователь. Например, если пользователь растерян и выражает тревогу, то боту будет дана инструкция для поддержки пользователя. Для этого были определены девять категорий, которые отражают различные ситуации и эмоциональные состояния пользователя. Каждая из этих категорий имеет свой набор промптов для генерации ответа. Принадлежность запроса к определенной категории определяется с помощью LLM. Автор провел сравнение данного подхода со следующими методами генерации:

- на основе нескольких примеров;
- с учетом определенной эмоции пользователя;
- с учетом ситуации, в которой находится пользователь;
- с учетом эмоции и ситуации, в которой находится пользователь.

Ситуация и эмоция пользователя в данных подходах определяются с помощью LLM. В результате рассмотренный подход обошел по метрикам качества, представленные выше подходы.

С. Абинаия и др. [Abi25] проводили корректировку ответа модели, включая информацию об эмоциональном состоянии пользователя в запрос. Для классификации эмоции они использовали две модели: BERT [Dev19] для определения эмоции по тексту и сверточную нейронную сеть для классификации по кадру из видео. В качестве чат-бота использовалась Gemini от Google. В результате тестирования пользователи, которые использовали чат-бота, информированного об эмоции, отмечают большую надежность и вовлеченность, чем те, которые использовали обычную версию.

В приведенных работах были рассмотрены подходы к коррекции поведения с помощью инструкции и с помощью включения информации о пользователе. В данной работе рассматривается комбинация этих двух подходов для корректировки ответа чат-бота на русском языке, где системный промпт выступает в качестве инструкции.

СПОСОБЫ КОРРЕКТИРОВКИ ПОВЕДЕНИЯ ЧАТ-БОТА

Пусть имеется обученная модель чат-бота на основе LLM, которая на вход принимает запрос пользователя и выдает ответ в виде текста. Требуется скорректировать её поведение так, чтобы она генерировала ответ с учетом эмоции пользователя.

Рассмотрены два варианта решения задачи. Первым вариантом является дообучение модели нейронной сети для генерации ответа с учетом невербальных сигналов пользователя. Плюсом такого подхода является то, что модель сможет самостоятельно находить связи между вербальными и невербальными данными, которые важны для генерации ответа. Однако для обучения требуется собрать большой набор данных в формате вопрос-ответ, включающий в себя и вербальную составляющую вопроса, и невербальную. При этом, дообучение такой модели будет требовать большого количества вычислительных мощностей.

С математической точки зрения этот подход сводится к минимизации функции потерь кросс-энтропии $L_{CE}(\theta)$ на мультимодальном датасете

$$D = \{(X_t^{(i)}, X_m^{(i)}, Y^{(i)})\}_{i=1}^N,$$

где X_t – текстовый запрос, X_m – сигнал невербальной модальности, а Y – ответ на запрос. Сама функция потерь $L_{CE}(\theta)$ имеет вид:

$$L_{CE}(\theta) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^S \log P(\hat{y}_j^{(i)} | y_{<j}^{(i)}, x_t^{(i)}, x_m^{(i)}; \theta),$$

где N – размер обучающей выборки, S – количество токенов ответа на запрос, $x_t \in X_t$ – токены текстового запроса, $x_m \in X_m$ – фрагменты невербального сигнала, $y \in Y$ – токены ответа на запрос, $\hat{y}$ – предсказанный токен.

Здесь оптимизируются все параметры θ базовой архитектуры, что требует квадратичных затрат памяти относительно длины контекста.

Второй вариант позволяет корректировать поведение модели без дополнительного обучения. Он основан на добавлении в запрос пользователя информации об его эмоциональном состоянии, полученном с помощью предобученной нейронной сети. Дополнительно для корректировки поведения модели можно задать системный промпт, в котором описывается желаемое поведение. Например, можно попросить модель действовать как психолог и давать ответ с учетом той эмоции, которая указана в запросе. Плюсом данного подхода является его простота, поскольку не требуется собирать большой набор данных для дообучения и иметь большое количество вычислительных мощностей. Однако в таком случае модель получает ограниченную информацию о состоянии человека (лишь дискретную метку класса эмоции e^*), что может снизить качество генерации ответа из-за потери взаимной информации

$I(Y; e^*) \leq I(Y; X_t, X_m)$. Такой подход был применен в работе С. Абинаия и др. [Abi25], но без использования системного промпта.

Формально процесс разделяется на два этапа. Сначала классификатор эмоций преобразует сырые данные X_t и X_m в вектор вероятности классов $p(e_k | X_t, X_m)$, где $k \in K$ – индекс класса эмоции, а затем выбирает класс с наибольшей вероятностью $e^* = \operatorname{argmax} p(e_k | x_m)$.

После этого происходит формирование расширенного контекста $C = [s; d; \text{"Эмоция: " } e^* + \text{"Вопрос: " } + X_t]$, где s – системный промпт, d – предыдущие сообщения диалога. Полученный расширенный контекст подается на вход языковой модели для генерации ответа.

Расширением второго подхода может быть дообучение модели с использованием низкоранговой адаптации (Low-Rank Adaptation, LoRA [Hu21]) с целью научить модель лучше проявлять эмпатию в зависимости от определенной эмоции. Метод предполагает, что изменение весов ΔW обученной матрицы W_0 имеет низкий ранг r . Вместо полной оптимизации замораживаются исходные веса, а к ним добавляется произведение двух матриц малого ранга $W = W_0 + BA$, где $B \in \mathbb{R}^{d \times r}$ и $A \in \mathbb{R}^{r \times k}$ (d – количество признаков на входе, k – количество признаков на выходе слоя). Тогда, для учета конкретной эмоции e^* могут быть обучены только эти матрицы весов, в результате чего можно произвести корректировку параметров трансформера при обновлении лишь нескольких миллионов дополнительных параметров вместо миллиардов основных.

В следующем разделе рассматривается реализация второго подхода с использованием модели GigaChat-2 от Сбера без дообучения с помощью LoRA и мультимодальной модели для классификации эмоции [Koc25].

ФОРМАЛИЗАЦИЯ ПРОБЛЕМЫ ИССЛЕДОВАНИЯ И РАЗРАБОТКА ПРИЛОЖЕНИЯ

В формальном виде проблема заключается в поиске оптимального баланса между сложностью реализации и качеством работы композиции $G \circ I \circ C$, где G – модель для генерации текста, I – оператор введения контекста в виде эмоции, C – модель для классификации эмоции.

Полное дообучение модели G подразумевает изменение её параметров ϕ , что требует квадратичных затрат памяти относительно длины контекста и наличие датасета в виде кортежей (e, X_t, Y) , где e – класс эмоции, X_t – текстовый запрос, Y – ответ на запрос.

Использование только алгоритма I без изменения параметров модели G ограничивает выразительную мощность системы, так как классификатор C выдает лишь индекс класса $e^* = C(X_t, X_m)$, теряя большое количество информации об исходных входных признаках X_t и X_m , где X_m – невербальный сигнал.

Задача разработки заключается в том, чтобы реализовать процедуру взаимодействия компонентов $I(C(X_t, X_m), X_t)$ таким образом, чтобы сделать ответ более персонализированным и эмпатичным, не прибегая к модификации параметров ϕ .

Алгоритм генерации ответа с учетом эмоции пользователя выглядит следующим образом:

1. Получение запроса от пользователя.
2. Предобработка входных данных.
3. Классификация эмоции по входным данным.
4. Генерация ответа с учетом определенной эмоции.
5. Вывод ответа пользователю.

За основу было взято приложение для классификации эмоций из предыдущей работы автора [Koc24], которое было дополнено модулем для взаимодействия с чат-ботом. Архитектура модуля представлена на рис. 1.

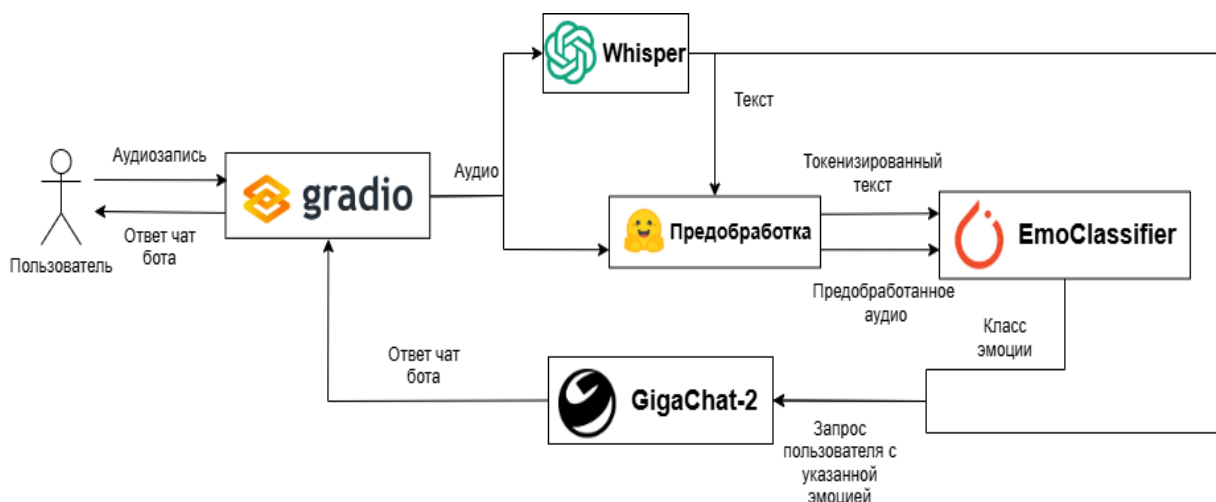


Рис. 1. Архитектура модуля для взаимодействия с чат-ботом

В качестве модели для классификации эмоций была использована мультимодальная нейронная сеть, предложенная автором ранее [Koc25], которая производит классификацию по голосу и речи. Процесс обучения модели заключался в минимизации функции потерь категориальной кросс-энтропии L_{CE} на множестве обучающих примеров:

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K y_k^{(i)} \log(p_k^{(i)}),$$

где N – размер обучающей выборки, $K = 4$ – количество классов (радость, грусть, злость, нейтральная эмоция), $y_k^{(i)}$ – бинарный индикатор принадлежности примера к классу, а $p_k^{(i)}$ – предсказанная моделью вероятность класса с индексом k , вычисленная путем применения оператора softmax к выходному вектору модели.

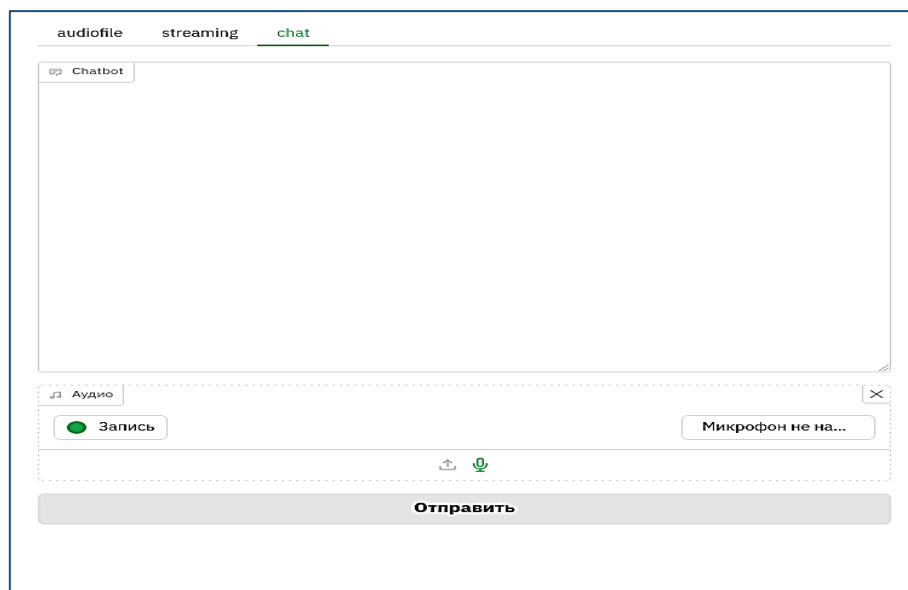
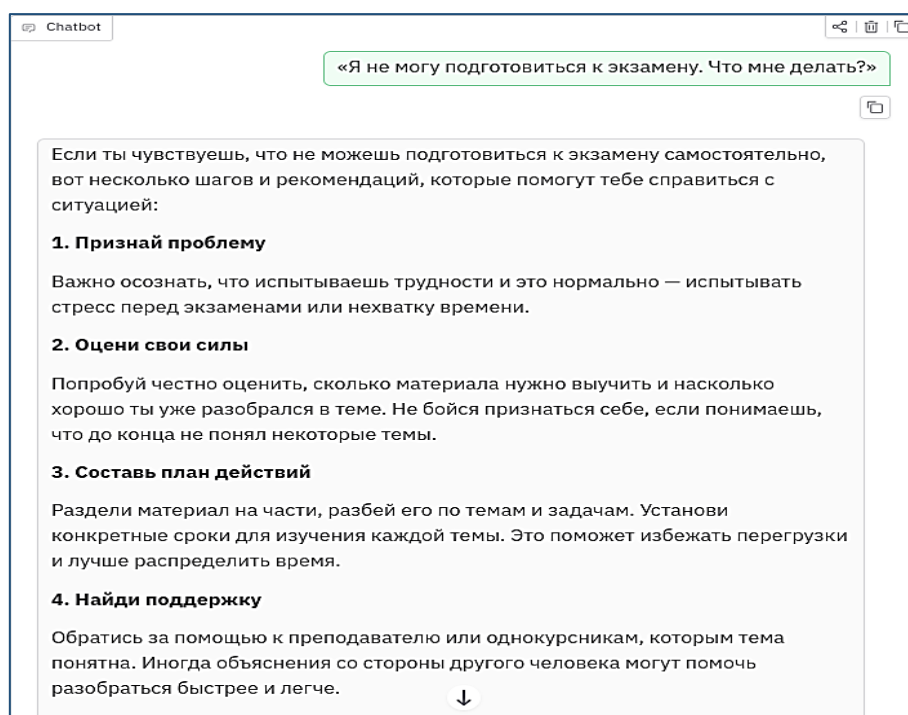
Для извлечения признаков из аудио используется модель Wav2Vec2 [Bae20], а для извлечения признаков из текста использует DistilBert-tiny [Kol22]. Модель обучалась на наборе данных Dusha [Kon22]. Набор состоит из аудиозаписей, размеченных по четырем классам эмоций: радость, грусть, злость и нейтральная эмоция. Обучение и запуск модели производился с использованием фреймворка PyTorch.

Предобработка данных выполнялась с помощью фреймворка transformers от Hugging Face. Она включает в себя приведение аудиозаписи к 16 кГц, извлечение текста из аудио и его токенизацию. Для извлечения текста использовалась модель Whisper [Rad23].

Взаимодействие с пользователем реализовано в виде веб-интерфейса, написанного с помощью библиотеки Gradio. Интерфейс представляет собой чат, поле для записи аудио с помощью микрофона и кнопку для отправки запроса. Интерфейс приложения представлен на рис. 2, а.

Интеграция интерфейсного компонента системы реализована на базе GigaChat-2. Применение данной платформы аргументировано ее совместимостью с русскоязычными текстовыми данными, простотой интеграции программных интерфейсов в общую архитектуру разрабатываемого ПО, а также доступностью вычислительных ресурсов (токенов) для проведения этапа тестирования и экспериментальной валидации системы.

Новизна работы состоит в создании метода программного управления формированием выходных текстовых последовательностей в интерфейсах. Эффект достигается за счет алгоритмического обогащения входного вектора запроса метаданными о текущем состоянии оператора. Предложенная программная архитектура не требует переконфигурирования базового вычислительного модуля или формирования специализированных баз данных для предварительной настройки системы, что существенно снижает временную и вычислительную сложность этапа развертывания и интеграции программного обеспечения.

*а – интерфейс приложения**б – запрос без корректировки поведения***Рис. 2.** Приложение для общения с чат ботом

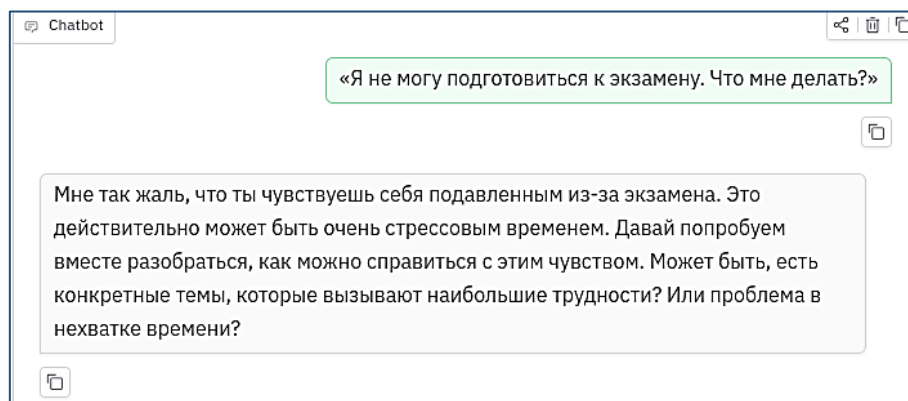
ЭКСПЕРИМЕНТЫ И ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Был проведен эксперимент диалога с чат-ботом без корректировки поведения и с учетом корректировки поведения. Пример ответа чат-бота без корректировки поведения изображен на рис. 2, б. Бот дает подходящий ответ по смыслу, но никак не учитывает эмоциональное состояние пользователя.

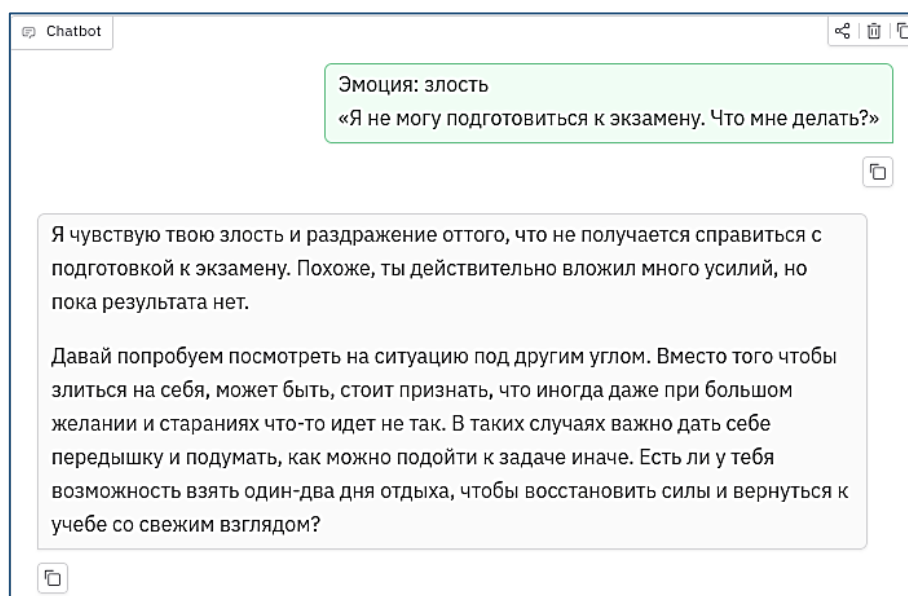
На рис. 3 представлены примеры ответов чат-бота с использованием только системного промпта и с добавлением информации об эмоциональном состоянии пользователя.

В первом случае (см рис. 3, а) чат-бот дал ответ с учетом того, что пользователь находится в подавленном состоянии. Это хороший результат, однако пользователь при подготовке к экзамену может испытывать не только подавленность, но и страх и даже злость.

Во втором случае (см рис. 3, б) в первом абзаце выражено понимание к чувствам пользователя, а во втором чат-бот дает совет по изменению ситуации с учетом эмоции пользователя.



а – с использованием системного промпта



*б – с использованием системного промпта
и информации о эмоции пользователя*

Рис. 3. Ответы чат-бота с корректировкой поведения

ЗАКЛЮЧЕНИЕ

Анализ существующих решений показал, что корректировка поведения модели возможна двумя путями: через внутреннее дообучение параметров или внешнее управление контекстом.

Первый подход требует обучения модели на мультимодальных данных и значительных вычислительных ресурсов для оптимизации её весов. Второй предполагает использовать системный промпт в качестве инструкции для чат-бота и включить в запрос пользователя информацию об его эмоциональном состоянии в виде метки, определенной с помощью модели для классификации.

В данном исследовании реализован второй подход: состояние человека определяется мультимодальной сетью по голосу и речи, а итоговый ответ формируется языковой моделью на основе расширенного текстового контекста, содержащего информацию об эмоциях собеседника. Такой выбор обусловлен скоростью разработки и отсутствием необходимости собирать масштабные размеченные датасеты для дообучения нейросети.

Интерфейс приложения реализован на языке Python с использованием библиотеки Gradio, который представляет собой чат с полем для записи аудио и кнопкой отправки запроса. Для извлечения текста из аудиодорожки применяется модель Whisper [Rad23], а предобработка данных осуществляется средствами фреймворка transformers от Hugging Face. За распознавание эмоционального состояния отвечает обученная автором на предыдущем этапе разработки мультимодальная модель. Обучение классификатора эмоций проводилось во фреймворке PyTorch на наборе данных Dusha. В качестве языковой модели для генерации ответа был выбран GigaChat-2 от Сбера, обеспечивающий поддержку русского языка.

Ручное тестирование показало, что добавление информации об эмоции позволяет скорректировать ответ модели. Если без корректировки бот дает подходящий по смыслу, но нейтральный ответ, то системный промпт помогает учесть подавленное состояние пользователя. Наилучший результат достигнут при передаче конкретной метки состояния – чат-бот выразил понимание чувств человека и сформулировал совет с учетом его ситуации. Это подтверждает эффективность использования классифицированной метки для генерации более адресной поддержки.

Одним из возможных вариантов развития данного подхода является улучшение качества ответов с использованием RAG на основе векторной базы данных. В этой базе данных могут храниться правила поведения модели, учитывающие различные ситуации и эмоциональное состояние пользователя. В случае использования открытых локальных моделей, качество ответа также может быть улучшено с помощью дообучения модели с использованием LoRA.

В дальнейшем планируется расширить количество распознаваемых классов эмоций и улучшить качество распознавания с помощью анализа выражения лица пользователя. Помимо улучшения качества распознавания, также планируется более детальное тестирование модели с использованием отложенной выборки.

СПИСОК ЛИТЕРАТУРЫ | REFERENCES

- | | |
|---|--|
| <p>[Abi25] Abinaya S. Ashwin K. S., Alphonse A. S. Enhanced Emotion-Aware Conversational Agent: Analyzing User Behavioral Status for Tailored Responses in Chatbot Interactions // IEEE Access. 2025. Vol. 13, pp. 19770-19787. DOI: 10.1109/ACCESS.2025.3534197.</p> <p>[Bae20] Baevski A., et al. Wav2Vec 2.0: A Framework for Self-Supervised Learning of Speech Representations // Proc. 34th Int. Conf. on Neural Information Processing Systems. 2020. pp. 12449-12460. DOI: 10.48550/arXiv.2006.11477.</p> <p>[Dev19] Devlin J., et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // North American Chapter of the Association for Computational Linguistics. 2019. DOI: 10.48550/arXiv.1810.04805.</p> <p>[Жум23] Жумажанова С. С., Сулавко А. Е., Ложников П. С. Распознавание психофизиологического состояния субъектов-операторов на основе анализа термографических изображений лица с применением сверточных нейронных сетей // СИИТ. 2023. Т. 5, № 2(11). С. 41-55. EDN: NNZWLV.</p> <p>[Kol22] Kolesnikova A., Kuratov Yu., et al. (2022). Knowledge Distillation of Russian Language Models with Reduction of Vocabulary. DOI: 10.48550/arXiv.2205.02340.</p> <p>[Kon22] Kondratenko V., Sokolov A, et al. (2022). Large Raw Emotional Dataset with Aggregation Mechanism. DOI: 10.48550/arXiv.2212.12266.</p> <p>[Koc24] Косачев И. С. Программное обеспечение для классификации эмоций человека ПО вербальным и невербальным признакам // Мавлютовские чтения: Матер. XVIII Всеросс. молодёжн. науч. конф. В 9-ти томах, Уфа, 25–</p> | <p>Abinaya S. Ashwin K.S., Alphonse A.S. Enhanced Emotion-Aware Conversational Agent: Analyzing User Behavioral Status for Tailored Responses in Chatbot Interactions // IEEE Access. 2025. Vol. 13, pp. 19770-19787. DOI: 10.1109/ACCESS.2025.3534197.</p> <p>Baevski A., et al. Wav2Vec 2.0: A Framework for Self-Supervised Learning of Speech Representations // Proc. 34th Int. Conf. on Neural Information Processing Systems. 2020. pp. 12449-12460. DOI: 10.48550/arXiv.2006.11477.</p> <p>Devlin J., et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // North American Chapter of the Association for Computational Linguistics. 2019. DOI: 10.48550/arXiv.1810.04805.</p> <p>Zhumazhanova S. S., Sulavko A. E., Lozhnikov P. S. Recognition of the psychophysiological state of human operators based on the analysis of facial thermographic images using convolutional neural networks // SIIT. 2023. Vol. 5, No 2(11), pp. 41-55. EDN: NNZWLV. (In Russian).</p> <p>Kolesnikova A., Kuratov Yu., et al. (2022). Knowledge Distillation of Russian Language Models with Reduction of Vocabulary. DOI: 10.48550/arXiv.2205.02340.</p> <p>Kondratenko V., Sokolov A, et al. (2022). Large Raw Emotional Dataset with Aggregation Mechanism. DOI: 10.48550/arXiv.2212.12266.</p> <p>Kosachev I. S. Software for classifying human emotions based on verbal and non-verbal cues // Mavlyutov Readings: Proceedings of the XVIII All-Russian Youth Scientific Conference. In 9 volumes, Ufa, November 25–29, 2024.</p> |
|---|--|

- 29 ноября 2024 года. Уфа: УУНИТ, 2024. С. 265-270. EDN: [NGNAMP](#).
- [Koc25] Косачев И. С., Сметанина О. Н., Сазонова Е. Ю. Построение нейросетевого классификатора эмоций для мультимодальных данных // Вестник УГАТУ. 2025. Т. 29, № 4(110). С. 39-47. EDN: [SCXTEM](#).
- [Лат24] Латкин И. С., Богданова Д. Р., Ахметвалеев Р. Р. Разработка программного обеспечения для построения карт эмоций: методологии моделирования и сценарии использования // СИИТ. 2024. Т. 6, № 2(17). С. 39-49. EDN: [JHOJSU](#).
- [Lon25] Longrun Z. GuidedEmpathy: Guiding Large Language Models for Empathetic Response Generation with Situational Awareness. // CSAI '24: Proc. 2024 8th Int. Conf. Computer Science and Artificial Intelligence. 2025, pp. 354-360. DOI: [10.1145/3709026.3709029](#).
- [Пиз24] Пиз А., Пиз Б. Новый язык телодвижений. М.: Изд-во Бомбора, 2024.
- [Rad23] Radford A., Kim J., et al. (2022). Robust Speech Recognition via Large-Scale Weak Supervision. DOI: [10.48550/arXiv.2212.04356](#).
- [Саз23] Сазонова Е. Ю. и др. Интеллектуальная СППР при управлении психофизическим состоянием человека // СИИТ. 2023. Т. 5, № 6(15). С. 38-49. EDN: [JFLRCX](#).
- [Сам23] Смотуга А. Е. Распознавание субъектов и их психофизиологических состояний на основе параметров подписи для защиты документооборота // СИИТ. 2023. Т. 5, № 2(11). С. 56-65. EDN: [JCFRBU](#).
- [Siy21] Siyang L., et al. Towards Emotional Support Dialog Systems // Proc. 59th Annual Meeting of the Association for Computational Linguistics and 11th Int. Joint Conf. Natural Language Processing. 2021. Vol. 1, pp. 3469-3483. DOI: [10.18653/v1/2021.acl-long.269](#).
- [Hil09] Hill C. E. Helping skills: Facilitating, exploration, insight, and action. American Psychological Association, 2009. DOI: [10.1037/0000147-000](#).
- [Hu21] Hu E., et al. LoRA: Low-Rank Adaptation of Large Language Models. 2021. DOI: [10.48550/arXiv.2106.09685](#).
- [Шир25] Ширинов Р. А., Гардашова Л. А., Богданова Д. Р. Краткий анализ методов Deep Learning для распознавания эмоционального состояния человека для принятия решений // СИИТ. 2025. Т. 7, № 2(21). С. 68-77. EDN: [DAWGEL](#).
- Ufa: Ufa University of Science and Technology, 2024. Pp. 265-270. EDN: [NGNAMP](#). (In Russian).
- Kosachev I. S., Smetanina O. N., Sazonova Y. Y. Construction of a neural network emotion classifier for multimodal data // Vestnik UGATU. 2025. Vol. 29, No. 4(110), pp. 39-47. EDN: [SCXTEM](#). (In Russian).
- Latkin I. S., Bogdanova D. R., Akhmetvaleev R. R. Software Development for Emotion Mapping: Modeling Methodologies and Use Cases // SIIT. 2024. Vol. 6, No. 2(17), pp. 39-49. EDN: [JHOJSU](#). (In Russian).
- Longrun Z. GuidedEmpathy: Guiding Large Language Models for Empathetic Response Generation with Situational Awareness. // CSAI '24: Proc. 2024 8th Int. Conf. Computer Science and Artificial Intelligence. 2025, pp. 354-360. DOI: [10.1145/3709026.3709029](#).
- Pease, A., Pease, B. The New Body Language. Moscow: Bombora Publ., 2024. (In Russian).
- Radford A., Kim J., et al. (2022). Robust Speech Recognition via Large-Scale Weak Supervision. DOI: [10.48550/arXiv.2212.04356](#).
- Sazonova Y. Y., et al. Intelligent Decision Support System for Managing a Person's Psychophysical State // SIIT. 2023. Vol. 5, No. 6(15), pp. 38-49. EDN: [JFLRCX](#). (In Russian).
- Samotuga A. E. Recognition of subjects and their psychophysiological states based on signature parameters for document workflow protection // SIIT. 2023. Vol. 5, No. 2(11), pp. 56-65. [JCFRBU](#). (In Russian).
- Siyang L., et al. Towards Emotional Support Dialog Systems // Proc. 59th Annual Meeting of the Association for Computational Linguistics and 11th Int. Joint Conf. Natural Language Processing. 2021. Vol. 1, pp. 3469-3483. DOI: [10.18653/v1/2021.acl-long.269](#).
- Hill C. E. Helping skills: Facilitating, exploration, insight, and action. American Psychological Association, 2009. DOI: [10.1037/0000147-000](#).
- Hu E., et al. LoRA: Low-Rank Adaptation of Large Language Models. 2021. DOI: [10.48550/arXiv.2106.09685](#).
- Shirinov R. A., Gardashova L. A., Bogdanova D. R. A Brief analysis of deep learning methods for recognizing human emotional states for decision-making // SIT. 2025. Vol. 7, No. 2(21), pp. 68-77. EDN: [DAWGEL](#). (In Russian).

ОБ АВТОРАХ | ABOUT THE AUTHORS

КОСАЧЕВ Илья Сергеевич

Уфимский университет науки и технологий, Россия.
ilyastalk@bk.ru ORCID: [0009-0005-2812-6777](#).
 Аспирант спец. «Математическое и программное обеспечение выч. систем, комплексов и компьютерных сетей».

САЗОНОВА Екатерина Юрьевна

Уфимский университет науки и технологий, Россия.
sazonova.eyu@ugatu.ru ORCID [0000-0001-8834-2635](#).
 Доц. каф. вычислительной математики и кибернетики. Дипл. экон.-математик (Уфимск. гос. авиац. техн. ун-т). Канд. техн. наук по системному анализу, управлению и обработке информации (там же). Иссл. в обл. управления, сист. анализа, интел. обр. данных.

KOSACHEV Ilya Sergeevich

Ufa University of Science and Technology, Russia.
ilyastalk@bk.ru ORCID: [0009-0005-2812-6777](#).
 PhD student specializing in "Mathematical and Software Support for Computing Systems, Complexes, and Computer Networks".

SAZONOVA Ekaterina Yurevna

Ufa University of Science and Technology, Russia.
sazonova.eyu@ugatu.ru ORCID [0000-0001-8834-2635](#).
 Assoc. Prof. of Computational Mathematics and Cybernetics dept. Dipl. Economist-Mathematician (Ufa State Aviation Tech. Univ.). Cand. Tech. Sci. on System Analysis, Management and Information Processing (ibid.). Research in the field of management, systems analysis, data mining.

МЕТАДАННЫЕ | METADATA

Заглавие: Интеграция классификатора эмоций в архитектуру диалоговой системы для корректировки текстовых ответов.

Авторы: Косачев И. С., Сазонова Е. Ю.

Аннотация: В статье рассматривается задача корректировки ответов чат-бота с учетом эмоционального состояния пользователя без дообучения языковой модели. Предложен подход, при котором отдельная мультимодальная модель определяет состояние человека по голосу и речи, после чего полученная метка объединяется с запросом и подается в GigaChat-2 от Сбера для управления тоном генерации. Проведенное ручное тестирование показало, что передача этой информации позволяет боту генерировать более персонализированные и эмпатичные ответы. Данный метод работает без дополнительного обучения, что значительно уменьшает вычислительные затраты на коррекцию ответов модели.

Ключевые слова: Эмоциональный интеллект; коррекция поведения языковых моделей; человеко-компьютерное взаимодействие; управление контекстом; машинное обучение; мультимодальные данные.

Язык: Русский.

Статья поступила в редакцию 1 сентября 2026 г.

Title: Integration of an emotion classifier into the architecture of dialogue systems for the adjustment of text responses.

Authors: Kosachev I. S., Sazonova E. Yu.

This article examines the task of adjusting chatbot responses based on the user's emotional state without fine-tuning the language model. An approach is proposed in which a separate multimodal model assesses the user's state via voice and speech; the resulting label is then combined with the query and fed into Sber's GigaChat-2 to control the tone of the generated output. Manual testing demonstrated that incorporating this information enables the bot to generate more personalized and empathetic responses. This method operates without additional training, significantly reducing the computational costs associated with adjusting the model's responses.

Key words: Emotional intelligence; language model behavior correction; human-computer interaction; context management; machine learning; multimodal data..

Language: Russian.

The editors received the article on 1 September 2026.