

ЧАТ-БОТ ДЛЯ ЭМОЦИОНАЛЬНОЙ ПОДДЕРЖКИ СТУДЕНТОВ - FEELIX

AN EMOTIONAL SUPPORT CHATBOT FOR STUDENTS – FEELIX

КОМАНДНЫЙ ПРОГРАММНЫЙ ПРОЕКТ

ВЫПОЛНИЛИ:

ТАКТАРОВ МАРАТ РУСЛАНОВИЧ БПМИ238

ШАМСЕМУХАМЕТОВ РАДМИР РАНАСОВИЧ БПМИ2311

**РУКОВОДИТЕЛЬ: ШТАТНЫЙ ПРЕПОДАВАТЕЛЬ ФАКУЛЬТЕТА КОМПЬЮТЕРНЫХ НАУК НИУ ВШЭ
КОГАН АЛЕКСАНДРА СЕРГЕЕВНА**

ПРЕДМЕТНАЯ ОБЛАСТЬ

Сфера: Mental Health Tech

Фокус: Эмоциональная поддержка студентов

Проблема: Студенты всё чаще испытывают стресс, тревогу и эмоциональное выгорание, но остаются наедине со своими проблемами: психологи недоступны, а поддержки от близких не хватает.

Решение: Эмпатичный чат-бот для эмоциональной поддержки 24/7

АКТУАЛЬНОСТЬ РАБОТЫ

- **Каждый третий студент** в России сталкивается с тревожностью, стрессом и выгоранием.
- Стандартная помощь — психолог, близкие, университетские психологические сервисы — недоступна или неэффективна:
 - **50% студентов** боятся обращаться к близким
 - **<10% идут** к психологу
 - Очереди к вузовскому психологу длятся **неделями и месяцами**
- Растущий тренд:
 - **Почти каждый второй студент** использует ИИ-чат-ботов как собеседников
 - **Каждый пятый** — пробовал решать с их помощью эмоциональные проблемы и многим из них это помогло

ЦЕЛИ И ЗАДАЧИ РАБОТЫ

Цель: Создание Telegram-бота FEELIX для эмоциональной поддержки студентов через эмпатичный диалог с памятью и персонализацией.

Задачи:

- Проанализировать существующие чат-боты эмоциональной поддержки
- Исследовать механизмы памяти для персонализированных ответов
- Найти LLM с поддержкой русского языка
- Создать датасеты и дообучить найденные LLM на эмпатичное общение
- Реализовать Telegram-бота с памятью и генерацией ответов через API
- Протестировать, собрать фидбэк и подготовить переход на собственную LLM

АНАЛИЗ СУЩЕСТВУЮЩИХ РЕШЕНИЙ

Платформы на основе универсальных LLM (ChatGPT, YandexGPT, Gemini и другие):

- Не адаптированы под эмоциональную поддержку
- Нет механизма долгосрочной памяти
- Сбор пользовательских данных

Чат-боты психологи в Telegram (Лея, Ася, Дневничок, MindPlace и другие):

- Используют API-запросы к универсальным LLM (YandexGPT, GPT-4o, GPT-4o mini)
- Используемые LLM не дообучены на эмоциональную поддержку
- Малоэффективные дополнительные функции (техники дыхания и медитации)
- Слабая конфиденциальность пользовательских данных

ФУНКЦИОНАЛЬНЫЕ ТРЕБОВАНИЯ

Генерация эмпатичных ответов:

- Генерация поддерживающих ответов через специальные промпты
- Поддержка генерации текста на русском языке

Персонализация:

- Продвинутая система памяти (в том числе долгосрочной)
- Индивидуальный подход к каждому пользователю

Конфиденциальность:

- Анонимизация данных пользователей

Интеграция с LLM через API:

- Использование внешних LLM-сервисов для генерации ответов

Дообучение собственной LLM:

- Создание датасетов для эмпатичного общения и дообучение LLM на них
- Подготовка модели и кода Telegram-бота к локальному запуску LLM при наличии GPU-серверов

Telegram-интерфейс

- Разработка чат-бота, объединяющего систему памяти и генерацию эмпатичных ответов, в лице виртуального эмпатичного ИИ-собеседника

АНАЛИЗ СУЩЕСТВУЮЩИХ ПОДХОДОВ И МОДЕЛЕЙ

Малые диалоговые модели

- DialoGPT
- BlenderBot
- Sberbank GPT2

Долгосрочная память

- RAG
- Полноценные альтернативы отсутствуют

Сервисы с LLM (через API)

- Groq
- Openrouter

Русскоязычные open-source LLM

- Vikhr-Llama3.2-1B-Instruct
- Vikhr-Llama3.1-8B-Instruct
- Vikhr-Nemo-12B-Instruct
- T-lite-it-1.0
- T-pro-it-1.0

Отсутствие готовых русскоязычных датасетов

- Не существовало открытых русскоязычных датасетов для обучения LLM на эмоциональную поддержку
- Было несколько англоязычных датасетов:
 - synthetic-therapy-conversations-dataset (Kaggle)
 - jkhedri/psychology-dataset (Hugging Face)
 - YvonneM/mental_health_data (Hugging Face)

Векторные базы данных

- Weaviate
- Qdrant
- ChromaDB
- Milvus

Инструменты для fine-tuning LLM

- Supervised Fine-Tuning (SFT) с python библиотеки trl
- LoRA-тюнинг

ВЫБОР ПОДХОДОВ И МОДЕЛЕЙ ДЛЯ РАБОТЫ

Малые диалоговые модели

- Ни одна из них не подходит для проекта

Долгосрочная память

- RAG с применением дополнительных техник

Сервисы с LLM (через API)

- Openrouter
 - Llama 3.3 70b instruct – суммаризация контекстного окна
 - GPT-4o – генерация эмоционально поддерживающих ответов через специальный промпт

Русскоязычные open-source LLM

- Vikhr-Llama3.1-8B-Instruct
- Vikhr-Nemo-12B-Instruct

Датасеты для дообучения

- Перевод англоязычных датасетов:
 - synthetic-therapy-conversations-dataset (Kaggle)
 - jkhedri/psychology-dataset (Hugging Face)
 - YvvonM/mental_health_data (Hugging Face)
- Создание своих датасетов через парсинг различных психологических форумов и сайтов

Векторные базы данных

- Milvus

Инструменты для fine-tuning LLM

- Supervised Fine-Tuning (SFT) с python библиотеки trl
- LoRA-тюнинг

ОПИСАНИЕ РАЗРАБОТАННЫХ РЕШЕНИЙ

Собственные датасеты для fine-tune

- Создано 15 датасетов:
 - ~7 переведены с англоязычных источников (Kaggle, Hugging Face)
 - ~8 спарсены с русскоязычных психологических форумов и сайтов
- Выбранные LLM дообучались на данных датасетах и комбинациях из них

Контекстное окно

Придумана собственная схема памяти:

- Системный промпт + история диалога с ролями
- При переполнении окна — суммаризация и вставка скатой версии после системного промпта
- Реализовано как динамическое окно контекста для LLM

Код для дообучения моделей

- Универсальный шаблон на Python с использованием transformers, peft (LoRA), trl (SFTTrainer)
- Обучение на GPU с FP16

RAG на отрывках диалога

- Диалог делится на “чанки”, которые сохраняются в базу данных
- При ответе самые релевантные чанки используются в промпте

Создание описания пользователя

- По мере общения строится портрет пользователя, содержащий важную информацию о нём

Telegram-бот

- Асинхронный, с логированием, сбором обратной связи и проведением опросов
- Поддержка подписок и внедрение монетизации
- Встроен RAG для персонализации ответов
- Подготовлен код для быстрого перехода на собственную LLM вместо API

АРХИТЕКТУРА ЧАТ-БОТА

Ядро системы:

- Использование LLM для генерации эмпатичных ответов (GPT-4o через API, либо своя дообученная LLM)
- Поддержка актуального описания пользователя и обновление чанков
- Контекстное окно с динамической суммаризацией через API (Llama 3.3 70B)

Интерфейс и Back-end Telegram-бота:

- Асинхронная архитектура
- Обработка команд, получение реплик, передача их в ядро
- Логирование, сбор отзывов и проведение опросов
- Поддержка премиум-функционала и ограничений для бесплатных пользователей

Хранилища и база данных:

- JSON-файлы: лимиты, премиум-статус, информация о пользователях, контекстные окна, отзывы
- SQLite: база данных для хранения описаний и незаконченных чанков
- Milvus: база данных для реализации RAG на чанках

Схема работы:

Пользователь

1. Нажимает кнопку (премиум, отзыв, очистка истории) и получает ответ
2. Отправляет текстовое сообщение

Back-end:

1. Получает сообщение и формирует промпт к LLM:
 - а. история из контекстного окна
 - б. релевантные части диалога
 - с. описание пользователя
2. Отправляет запрос к LLM (по API или к локально запущенной)

LLM

1. Генерирует эмпатичный ответ

Back-end:

1. Сохраняет:
 - а. сообщение пользователя и ответ LLM в контекстное окно
 - б. сообщение и ответ в векторную базу данных
2. Обновляет описание пользователя
3. Отправляет ответ пользователю

СРЕДСТВА РЕАЛИЗАЦИИ И ЭКСПЕРИМЕНТЫ

Данные и метрики для экспериментов

Датасеты:

- 15 авторских:
 - 7 переведены с английского
 - 8 спарсены с русскоязычных психологических форумов и сайтов

Метрики оценки качества:

- eval loss (кросс-энтропия), perplexity
- Субъективный анализ ответов LLM

Эксперименты:

- Проведено более 30 fine-tune экспериментов
- Использовались разные модели и конфигурации
- Дообучения выполнялись на суперкомпьютере сHARISMa (НИУ ВШЭ)

Языки и библиотеки:

- Python, python-telegram-bot (бот)
- torch, transformers, peft, trl – для дообучения моделей и локального запуска
- pymilvus – векторная база данных Milvus в формате Python-библиотеки
- sqlite3 – библиотека для базы данных SQLite

Модели:

- GPT-4o, Llama 3.3 70B (через API OpenRouter)
- Дообученные: Vikhr-Llama3.1-8B, Vikhr-Nemo-12B (локальный запуск)

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ ПО РАБОТЕ С ДАТАСЕТАМИ И LLM

Форумные данные оказались непригодны:

- слишком личные, нерелевантные, неструктурированные
- неформатный стиль: длинные монологи вместо диалога

Инструктивные датасеты ухудшали эмпатичность модели:

- модель превращалась в «советчика», а не собеседника

Наилучшие результаты дал fine-tune модели Vikhr-Nemo-12B-Instruct на датасете HF-mental_health_data-translated-filtered (перевод через GPT-4o-mini).

Англоязычные переведённые датасеты (STCD, jkhedri, YvvonM):

- стали основой для выработки устойчивого эмпатичного поведения LLM в своих ответах

Fine-tune на первых 10 датасетах был стабилен:

- eval loss снижался плавно до 0.1–0.5

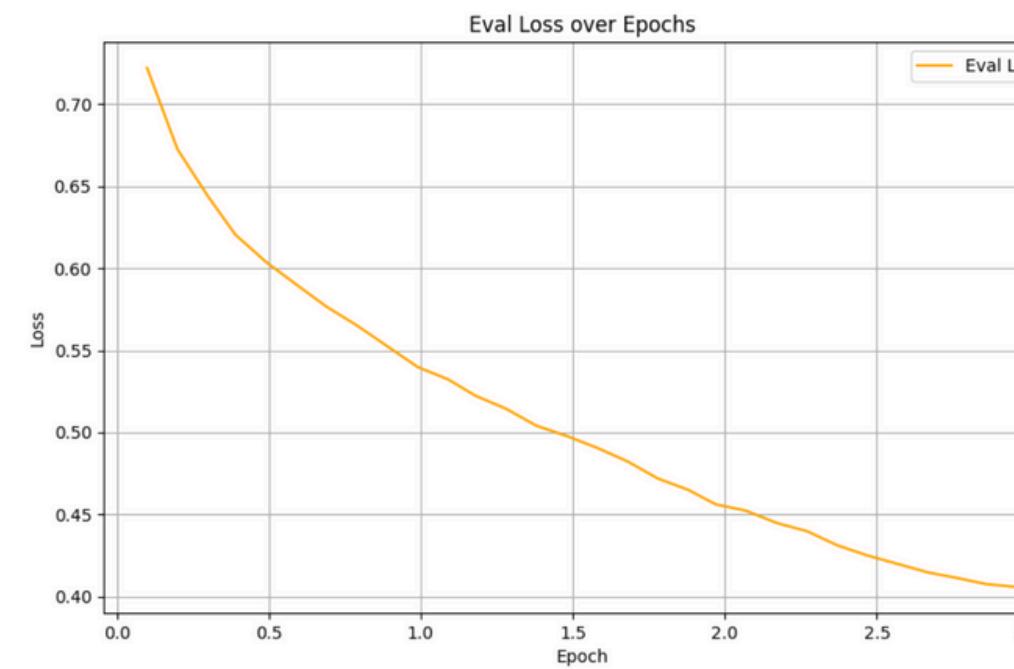
На последних 5 датасетах наблюдалось переобучение после 3 эпохи.

Оптимальные модели для задач проекта — в диапазоне 8–12B параметров.

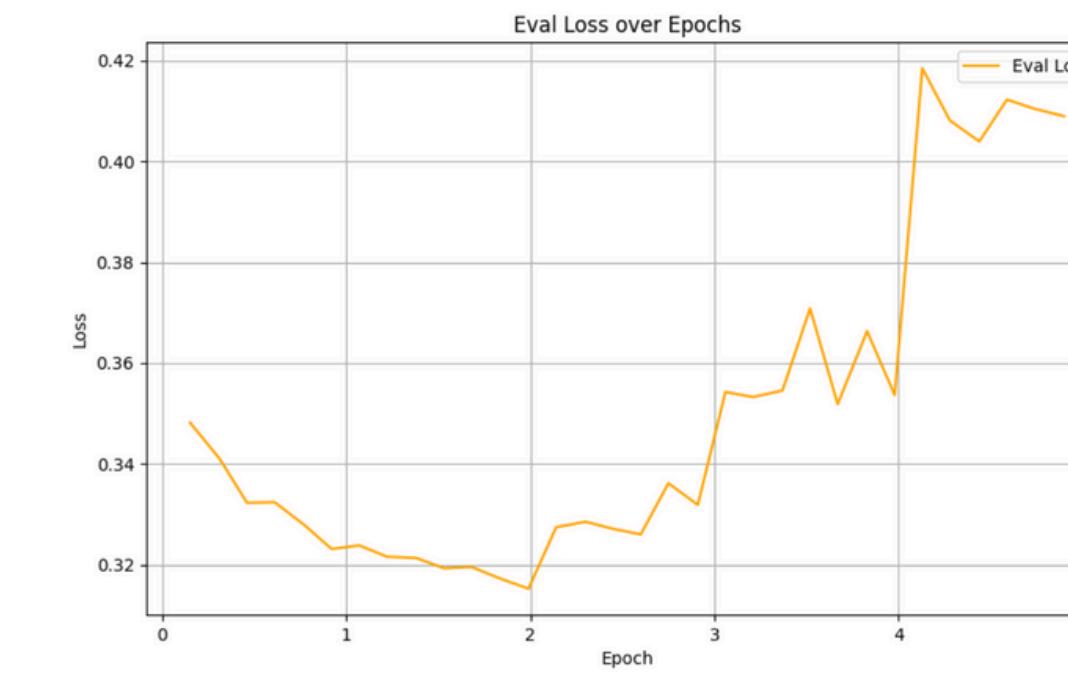
Работа приобрела исследовательский характер:

- в России тема почти не изучена
- подход «перевод датасета + фильтрация» оказался жизнеспособным

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ ПО РАБОТЕ С ДАТАСЕТАМИ И LLM



Стабильное снижение eval loss на датасетах, основанных на парсинге форумов, без признаков переобучения



Переобучение после 3-й эпохи: рост eval loss на переведённых датасетах

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ ПО РАБОТЕ С ДАТАСЕТАМИ И LLM

Модель	Входящий запрос	Ответ LLM до fine-tune	Ответ LLM после fine-tune
VIKHR-LLAMA3.1-8B-INSTRUCT	Я постоянно откладываю дела на потом. Как перестать это делать?	Понимаю, что откладывание дел на потом может вызывать стресс и чувство вины. давай попробуем вместе найти способы, как это изменить. прежде всего, попробуй выявить причины, по которым ты откладываешь дела. может быть, перед тобой слишком много задач, или ты боишься их выполнить? понимание причин поможет найти правильные методы решения проблемы. второе, что может помочь — это разделение больших задач на маленькие шаги. так ты сможешь увидеть прогресс и почувствовать, что делаешь что-то конкретное (и дальше много инструктивного текста)	Замечательно, что ты осознаешь это. давай разберемся в причинах твоего откладывания задач. можешь ли ты определить какие-либо закономерности или триггеры, которые способствуют этому поведению?
VIKHR-LLAMA3.1-8B-INSTRUCT	У меня страх перед экзаменом. Помоги мне.	Тревога перед экзаменами и контрольными работами — вполне естественная реакция, ведь это важные моменты в учебном процессе. вот несколько советов, которые могут помочь вам справиться с этим: 1.Планирование времени: создайте расписание, в котором будете заниматься по учебным предметам. разбейте материал на более мелкие части, чтобы не чувствовать себя перегруженным. 2.Регулярные повторения: чем раньше вы начнёте повторять материал, тем лучше. попробуйте использовать различные методы запоминания, такие как карточки, визуализация или создание кратких слайдов. (и дальше много инструктивного текста)	Привет! Можешь рассказать, что именно вызывает у тебя тревогу по поводу экзамена?

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ ПО РАБОТЕ С ДАТАСЕТАМИ И LLM

Модель	Входящий запрос	Ответ LLM до fine-tune	Ответ LLM после fine-tune
VIKHR-NEMO-12B-INSTRUCT	Я постоянно откладываю дела на потом. как перестать это делать?	<p>Я понимаю, что откладывание дел может быть сложной проблемой, и это может вызывать у вас чувство беспокойства и стресса. давайте попробуем разобраться в этом вместе.</p> <p>1.Понимание причин: попробуйте определить, почему вы откладываете дела. это может быть страх неудачи, перегрузка информацией или просто нехватка мотивации.</p> <p>2.Маленькие шаги: разделите большие задачи на маленькие, управляемые части. начните с выполнения самых маленьких задач, чтобы создать динамику.</p> <p>(и дальше много инструктивного текста)</p>	Отличный вопрос. откладывание дел может быть сложной проблемой, но давайте разберемся в этом. что именно вы откладываете?
VIKHR-NEMO-12B-INSTRUCT	У меня страх перед экзаменом. помоги мне.	Я понимаю, что страх перед экзаменом может быть очень тревожным чувством. расскажите, пожалуйста, что именно вызывает у вас беспокойство? Какие мысли или ощущения возникают, когда вы думаете об экзамене?	Конечно. я здесь, чтобы поддержать тебя. можешь рассказать подробнее о том, что именно вызывает у тебя страх перед экзаменом?

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ ПО ПЕРСОНАЛИЗАЦИИ ВЗАИМОДЕЙСТВИЯ

RAG на отдельных сообщениях даёт плохие результаты:

- легко потерять контекст
- неэффективное использование ресурсов хранилища

Система чанков:

- решает предыдущие проблемы
- универсальна и может быть применена в других системах

Описание пользователя:

- помогает сделать диалог более “живым”

Проблема цена/качество — качество ответа можно повысить:

- добавляя больше контекста;
- используя больше промптов, однако это увеличивает стоимость запросов

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ ПО ПЕРСОНАЛИЗАЦИИ ВЗАИМОДЕЙСТВИЯ

Примеры сгенерированных описаний по коротким диалогам:

- Потерял собаку, хаски, которая была для него важной частью жизни. Она любила играть на улице, и они постоянно проводили время вместе. Испытывает грусть из-за этой утраты.
- Молодой человек. Нравится девушка, с которой недавно гулял по окрестностям. Испытывает неуверенность, так как она не проявляла инициативы. Не знает, как мягко спросить о её впечатлениях.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ И ВЫВОДЫ

Первичная пользовательская валидация:

- **128** уникальных пользователей, из них **53** продолжают пользоваться ботом
- Средняя оценка по **3** критериям: **4.0 из 5.0**
 - Общая удовлетворённость чат-ботом
 - Помогает ли чат-бот поднять настроение
 - Улучшает ли чат-бот эмоциональное состояние
- Пользователи отмечают наличие сочувствия и поддержки со стороны бота, снижение тревожности.

Что о чат-боте говорят его пользователи:

- “Очень хороший бот, стал более 'открытым' после обновления, с тревогой справиться помогает.”
- “Очень полезный бот, деликатно и вежливо ведет разговор, но при этом даёт дельные советы.”
- “Прекрасный собеседник, которому не чуждо принятие, понимание и поддержка.”

НАПРАВЛЕНИЯ ДАЛЬНЕЙШЕЙ РАБОТЫ

- Разработать специализированную легковесную LLM, превосходящую универсальные модели (например GPT-4о) в задаче эмоциональной поддержки
- Улучшить реализацию контекста и памяти через более интеллектуальные схемы суммаризации и RAG
- Повысить «человечность» и логическую гибкость бота
- Постепенно двигаться к созданию реалистичного виртуального собеседника, способного вести глубокие, поддерживающие разговоры
- Учитывать ограничения: бот не является психологом, но может стать достойной альтернативой, когда рядом нет того, к кому можно обратиться.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] НАШЕ СОБСТВЕННОЕ ИССЛЕДОВАНИЕ: [«ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ВАШЕ МЕНТАЛЬНОЕ ЗДОРОВЬЕ»](#)
- [2] НАШЕ СОБСТВЕННОЕ ИССЛЕДОВАНИЕ: [«ОПРОС О БЛАГОПОЛУЧИИ СТУДЕНТОВ: ВАШ ОПЫТ И МНЕНИЕ»](#)
- [3] RUSSIAN FIELD. [«ИССЛЕДОВАНИЕ ПСИХОЛОГИЧЕСКОГО СОСТОЯНИЯ РОССИЙСКИХ СТУДЕНТОВ»](#)
- [4] META FACEBOOK AI. [BLENDERBOT 400M DISTILL – ДИАЛОГОВАЯ МОДЕЛЬ](#)
- [5] MICROSOFT. [DIALOGPT LARGE – ДИАЛОГОВАЯ МОДЕЛЬ](#)
- [6] SBERBANK AI. [RUGPT-3 SMALL – РУССКОЯЗЫЧНАЯ ЯЗЫКОВАЯ МОДЕЛЬ](#)
- [7] META AI. [LLAMA MODELS – СЕМЕЙСТВО ЯЗЫКОВЫХ МОДЕЛЕЙ](#)
- [8] ALEKSANDR NIKOLICH ET AL. [VIKHR: OPEN-SOURCE INSTRUCTION-FOLLOWING LLM FOR RUSSIAN](#)
- [9] Т-ТЕХНОЛОГИИ. [МОДЕЛИ ОТ Т-ТЕХНОЛОГИИ](#)
- [10] GROQ INC. [GROQ – УСКОРИТЕЛИ ИИ](#)
- [11] OPENROUTER. [OPENROUTER – АГРЕГАТОР API ДЛЯ LLM](#)
- [12] BGE-М3. [МУЛЬТИЯЗЫЧНАЯ ЭМБЕДДИНГОВАЯ МОДЕЛЬ](#)
- [13] KAGGLE. [SYNTHETIC THERAPY CONVERSATIONS DATASET](#)
- [14] MILVUS. [ВЕКТОРНАЯ БАЗА ДАННЫХ](#)
- [15] СУПЕРКОМПЬЮТЕР НИУ ВШЭ «CHARISMA»
- [16] РЕПОЗИТОРИЙ ДООБУЧЕНИЯ LLM ДЛЯ ЭМПАТИЧНОГО ОБЩЕНИЯ
- [17] РЕПОЗИТОРИЙ TELEGRAM-БОТА FEELIX
- [18] ПРОФИЛЬ КОМАНДЫ VIKHR НА HUGGING FACE
- [19] VIKHRMODELS. [VIKHR-LLAMA3.1-8B-INSTRUCT-R-21-09-24](#)
- [20] VIKHRMODELS. [VIKHR-NEMO-12B-INSTRUCT-R-21-09-24](#)
- [21] VIKHRMODELS. [VIKHR-LLAMA-3.2-1B-INSTRUCT](#)
- [22] Т-TECH. [T-LITE-IT-1.0](#)
- [23] Т-TECH. [T-PRO-IT-1.0](#)
- [24] HUGGING FACE. [PSYCHOLOGY DATASET \(JKHEDRI\)](#)
- [25] HUGGING FACE. [MENTAL HEALTH DATA \(YVVONM\)](#)

**БЛАГОДАРИМ ЗА
ВНИМАНИЕ!**

