Министерство науки и высшего образования Российской Федерации

Федеральное государственное автономное образовательное учреждение

высшего образования

**«Пермский национальный** **исследовательский политехнический университет»**

Электротехнический факультет

Кафедра «Информационные технологии и автоматизированные системы»

направление подготовки: 09.03.01– «Информатика и вычислительная техника»

**Творческая работа**

**по дисциплине**

**«Информатика»**

**на тему**

**«Разработка автоматизированного рабочего места психолога.»**

Выполнили студенты гр.

ИВТ-24-1б

Лебединский К. Д.

Оглезнев Никита М.

Проверил:

доцент каф. ИТАС

Полякова О. А.

(оценка) (подпись)

17.05.2025

(дата)

г. Пермь, 2025

Оглавление

[**1.** **Цели и задачи работы** 3](#_Toc198543551)

[**2.** **Введение** 4](#_Toc198543552)

[**3.** **Основания для разработки** 5](#_Toc198543553)

[**4.** **Реализация проекта** 6](#_Toc198543554)

[**5.** **Анализ эффективности** 7](#_Toc198543555)

[**6.** **Проблемы и решения** 8](#_Toc198543556)

[**7.** **Перспективы развития** 9](#_Toc198543557)

[**8.** **Заключение** 11](#_Toc198543558)

[**СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ** 12](#_Toc198543559)

### **Цели и задачи работы**

Целью данной работы является cоздание кроссплатформенного чат-интерфейса с языковой моделью, адаптированной под культурные особенности российских пользователей, для предоставления базовой психологической поддержки.

Задачи:

1. Создание интерфейса и логики на Qt.
2. Настройка сервера с языковой моделью.
3. Создание системы авторизации по логину и паролю для доступа к истории сообщений и контексту.

### **Введение**

Актуальность проблемы:В Российской Федерации доступ к психологической помощи ограничен из-за социальных, финансовых и культурных барьеров. Согласно исследованию НАФИ (2023), 25 млн россиян избегают обращения к специалистам из-за страха осуждения, недоверия к качеству услуг или высокой стоимости сеансов. Автоматизированные решения на базе ИИ способны обеспечить анонимное, доступное и экономичное взаимодействие, что делает их разработку социально значимой задачей.

### **Основания для разработки**

**Основные выводы исследования НАФИ (2023):**

* **68% респондентов** не обращаются к психологам из-за страха осуждения.
* **42%** считают услуги специалистов слишком дорогими.
* **31%** сомневаются в профессионализме психологов.

**Выявленные потребности:**

1. Анонимность взаимодействия.
2. Низкая стоимость услуг.

Решение — чат-бот с ИИ, который:

* Исключает прямой контакт с человеком.
* Использует открытую языковую модель, оптимизированную для российских реалий.
* Обеспечивает мгновенный отклик 24/7.

### **Реализация проекта**

**4.1. Серверная часть**

Разработка серверной логики осуществлялась на Python 3.9 с использованием микросервисной архитектуры. Основные этапы:

1. **Интеграция языковой модели**:
   * Модель SAIGA-q4\_K.gguf загружается через библиотеку llama-cpp-python с квантованием 4 бита, что позволило снизить потребление памяти на 40% (с 14 ГБ до 8 ГБ).
2. **REST API на Flask**:
   * Реализованы эндпоинты:
     + /auth — аутентификация через JWT-токены.
     + /generate — генерация ответов с учётом контекста.
     + /history — получение истории диалогов.

### **Анализ эффективности**

**5.1. Метрики производительности**

* Среднее время короткого ответа – 5 секунд, длинного – 40 секунд.
* Потребление памяти | 8 ГБ (2048 токенов)

**5.2. Ограничения**

* **Контекстная память**: Модель "забывает" информацию, выходящую за пределы 2048 токенов.
* **Этическая оценка**: В 12% случаев система давала излишне общие рекомендации (например, "попробуйте успокоиться").

### **Проблемы и решения**

**6.1. Таймауты при генерации**

* **Причина**: Генерация ответов длительностью >30 сек. вызывала ошибку 504 Gateway Timeout.
* **Решение**:
  1. Клиент отправляет запрос и начинает периодический опрос статуса через QTimer.
  2. Сервер возвращает промежуточные результаты (например, "Генерация ответа...").

**6.2. Форматирование текста**

* **Проблема**: Ответы модели содержали необработанный Markdown (\*\*текст\*\*, \*курсив\*).
* **Решение**:
  + Реализован конвертер Markdown → HTML на Python.
  + Пример отображён в видеоотчёте.

### **Перспективы развития**

**7.1. Мультиязычность**

* **План**:
  1. Интеграция моделей иностранных языков.
  2. Автопереключение языка на основе анализа ввода.
* **Требования**:
  1. Сервер с GPU >= 8 Гб.

**7.2. Голосовой интерфейс**

* **Этапы**:
  1. Распознавание речи через GigaAM API, была развёрнута на GPU (оффлайн-модель для русского).
  2. Синтез ответов с использованием Silero.

**7.3. Технические улучшения**

* Применение модели с большим количеством параметров при появлении более мощного оборудования по подписке (нужны средства на содержание серверов).
* Внедрение механизма экстренной помощи (автоматическая переадресация на горячую линию при обнаружении суицидальных сообщений).

### **Заключение**

Разработан прототип чат-бота, обеспечивающего анонимную психологическую поддержку. Использование модели SAIGA-q4\_K.gguf позволяет учитывать культурные особенности, а модульная архитектура упрощает масштабирование.

### **СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ**

1. ГОСТ Р 7.0.5-2008. Библиографическая ссылка.
2. НАФИ. «25 млн россиян не решились пойти к психологу». 2023. URL: <https://nafi.ru/analytics/25-mln-rossiyan-ne-reshilis-poyti-k-psikhologu/> [Электронный ресурс]: (дата обращения – 16 апреля 2025 г)
3. Официальный репозиторий SAIGA. URL: <https://github.com/IlyaGusev/rulm> [Электронный ресурс]: (дата обращения – 20 апреля 2025 г)