

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»**

Кафедра математической кибернетики и компьютерных наук

**РАЗРАБОТКА IOS-ПРИЛОЖЕНИЯ ДЛЯ ОБУЧЕНИЯ  
АНГЛИЙСКОМУ ЯЗЫКУ С СИСТЕМОЙ  
ПЕРСОНАЛИЗИРОВАННОГО ОБУЧЕНИЯ НА ОСНОВЕ  
ИСКУССТВЕННОГО ИНТЕЛЛЕКТА**

**АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ**

студента 4 курса 411 группы  
направления 02.03.02 — Фундаментальная информатика и информационные  
технологии  
факультета КНиИТ  
Куракина Эмиля Сергеевича

Научный руководитель

доцент, к. ф.-м. н.

\_\_\_\_\_

Ю. Н. Кондратова

Заведующий кафедрой

доцент, к. ф.-м. н.

\_\_\_\_\_

С. В. Миронов

Саратов 2026

## ВВЕДЕНИЕ

**Актуальность темы.** В условиях глобальной цифровой трансформации образования и стремительного роста мобильности общества потребность в эффективных, доступных и гибких инструментах для изучения иностранных языков становится как никогда острой. Традиционные педагогические подходы, ориентированные на «среднего» ученика в группе, зачастую не способны учитывать уникальный когнитивный профиль каждого обучающегося: индивидуальный темп усвоения лексики, специфику работы памяти, а также характерные паттерны совершаемых ошибок. Это приводит к снижению мотивации, быстрому забыванию изученного материала и, как следствие, низкой результативности обучения.

Современные достижения в области искусственного интеллекта, в частности технологии обработки естественного языка (NLP) и глубокого обучения, открывают новые возможности для создания адаптивных образовательных систем. Такие системы способны в режиме реального времени анализировать каждое действие пользователя, выявлять скрытые закономерности в его ошибках и динамически перестраивать образовательную траекторию. Интеграция подобных интеллектуальных алгоритмов непосредственно в мобильные приложения для платформы iOS позволяет сделать процесс изучения английского языка не только доступным «на ходу», но и превращает смартфон в персонального цифрового тьютора.

Особую значимость в процессе освоения иностранного языка имеет мгновенная и содержательная обратная связь. Традиционные тесты с выбором одного варианта ответа не позволяют практиковать свободную письменную речь. В то же время автоматическое обнаружение, классификация и интерпретация грамматических ошибок в произвольных текстах пользователя позволяют ему не просто видеть факт неточности, но и глубоко понимать её лингвистическую природу. Кроме того, применение алгоритмов интервальных повторений SRS на основе современных методов предиктивной аналитики (например, градиентного бустинга) позволяет оптимизировать процесс запоминания материала, фокусируя усилия пользователя именно на тех элементах, которые с наибольшей вероятностью будут забыты в ближайшее время.

Разработка подобных систем сопряжена с рядом технических вызовов: необходимостью обеспечения высокой скорости работы тяжёлых нейросетевых

моделей на мобильных устройствах с ограниченными ресурсами, обеспечением приватности данных пользователя через локальные вычисления (On-device ML) и созданием бесшовной интеграции между мобильным клиентом и облачными сервисами. Решение этих задач требует комплексного подхода, объединяющего знания в области лингвистики, педагогической психологии и современной программной инженерии.

**Цель бакалаврской работы** – проектирование и разработка iOS-приложения для обучения английскому языку с системой персонализированного обучения на основе искусственного интеллекта.

Поставленная цель определила **следующие задачи:**

- провести аналитический обзор современных мобильных приложений для изучения английского языка и исследовать существующие подходы к применению машинного обучения в задачах NLP и адаптивного обучения;
- исследовать теоретические основы архитектуры Transformer и алгоритмов градиентного бустинга для их применения в образовательных целях;
- разработать и обучить модели машинного обучения для детектирования и классификации грамматических ошибок, а также генерации учебных упражнений на базе LLM;
- спроектировать семантическую графовую модель для реализации интеллектуальной системы рекомендаций учебного контента;
- реализовать подсистему персонализации обучения на основе алгоритма интервальных повторений, использующего модель CatBoost для оценки вероятности забывания материала;
- разработать гибридную архитектуру системы, сочетающую локальный инференс моделей (on-device ML) и облачный Backend;
- реализовать пользовательский интерфейс приложения с использованием Flutter, интегрировать все разработанные модули и провести комплексное тестирование производительности и удобства использования системы.

**Методологические основы** исследования в области персонализированного обучения и интеллектуальных обучающих систем представлены в работах В. Лейтнера, П. Возняка, А. Васвани, Д. Бахданау, Я. Бенджио, Д. Хинтона, Д. Сильвер, Д. Кшиштовски, а также в трудах исследователей Google Research и Hugging Face.

**Теоретическая и практическая значимость.** Теоретическая значимость

работы заключается в разработке и апробации гибридной архитектуры интеллектуальной обучающей системы, сочетающей локальный инференс нейросетевых моделей на мобильных устройствах с облачными вычислениями, а также в исследовании применимости модели CatBoost для задачи предсказания вероятности забывания лексических единиц в контексте интервальных повторений.

Практическая значимость подтверждается созданием полнофункционального мобильного приложения, способного функционировать в офлайн-режиме, обеспечивать персонализированную подачу учебного материала, автоматическую генерацию контекстуальных упражнений и интеллектуальную проверку грамматики в реальном времени. Разработанная система может быть использована в качестве образовательного инструмента для самостоятельного изучения английского языка.

**Структура и объём работы.** Бакалаврская работа состоит из введения, 2 разделов, заключения, списка использованных источников и 11 приложений. Общий объем работы – 82 страницы, из них 51 страница – основное содержание, включая 16 рисунков и 7 таблиц, цифровой носитель в качестве приложения, список использованных источников информации – 21 наименований.

## КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

**Первый раздел «Теоретические аспекты и анализ технологий разработки интеллектуальных обучающих систем»** посвящен обзору предметной области и обоснованию выбора технологий. Проведён сравнительный анализ популярных мобильных приложений для изучения английского языка (Duolingo, Rosetta Stone, Memrise, Babbel, Lingualeo), выявлены их сильные и слабые стороны в контексте персонализации обучения, качества обратной связи и используемых технологий машинного обучения.

Рассмотрены теоретико-методологические основы персонализированного обучения: формирование профиля когнитивного состояния студента, адаптивное управление интервалами повторения, диагностическая обработка учебных ошибок. Описаны ключевые методологические парадигмы: коммуникативный подход, метод интервальных повторений и микро-обучение.

Проанализированы технологические основы интеллектуальных обучающих систем: эволюция от классических подходов Лейтнера и SM-2 к вероятностному моделированию на основе Half-Life Regression, а также применение градиентного бустинга CatBoost для предсказания вероятности забывания. Рассмотрено развитие NLP-технологий в образовании: от N-грамм и LSTM до архитектуры Transformer и моделей BERT, T5. Описаны графовые модели знаний как инструмент структурирования лексики и интеллектуальные рекомендации на основе алгоритма Personalized PageRank.

Проведён сравнительный анализ существующих решений, обоснована необходимость разработки системы, объединяющей SOTA-модели NLP для глубокого анализа текстов и современные методы машинного обучения для предиктивного управления учебным процессом. Обоснован выбор технологического стека: Flutter и Dart для мобильного клиента, ONNX Runtime для локального инференса, FastAPI для серверной части, Hive для локальной базы данных.

**Второй раздел «Проектирование, программная реализация и экспериментальное исследование системы обучения»** посвящён практической реализации разработанной системы. Спроектирована гибридная архитектура, сочетающая локальные вычисления на устройстве и облачные сервисы. Система включает четыре уровня: Presentation Layer (Flutter UI), Service Layer (Dart-бизнес-логика с MLService и ApiService), Data Layer (Hive) и Intelligence Layer (Python/FastAPI).

Разработана семантическая графовая модель для интеллектуальных рекомендаций на основе предобученных эмбеддингов GloVe и алгоритма Personalized PageRank. Граф содержит более 50 тысяч узлов; веса рёбер определяются косинусным сходством между векторными представлениями слов.

Реализована система интервальных повторений на базе модели CatBoost. Обучение проводилось на датасете Settles (Duolingo), дополненном лингвистическими признаками, содержащем более 13 миллионов записей. Сформировано признаковое пространство, включающее темпоральные признаки (`delta_days`, `log_delta`, `days_since_first_seen`), статистику успеваемости (`history_seen`, `history_correct`, `history_accuracy`) и лингвистические признаки (`pos`, `word_complexity`, `user_proficiency`). На тестовой выборке достигнуты показатели: Accuracy — 0,896, F1-score — 0,923, ROC-AUC — 0,836.

Отдельным этапом реализации стала разработка механизма персонализации траектории обучения. Помимо базового прогнозирования вероятности забывания система формирует индивидуальный профиль пользователя, включающий динамику успеваемости, сложность изучаемых слов, частоту ошибок и скорость ответа. На основе этих параметров осуществляется адаптивное изменение интервалов повторения, а также подбор следующего учебного материала.

В отличие от классических реализаций алгоритмов интервальных повторений, использующих фиксированные математические формулы, предложенный подход позволяет учитывать многомерное пространство признаков и скрытые зависимости между ними. Это повышает точность прогнозирования забывания и обеспечивает более эффективное распределение учебной нагрузки.

Дополнительно реализован механизм накопления статистики обучения. Для каждого пользователя система хранит историю взаимодействий, включая правильность ответа, время выполнения задания, тип допущенной ошибки и общее количество повторений. Эти данные используются для последующего дообучения рекомендательной системы и повышения качества персонализации.

Разработан модуль детекции и классификации грамматических ошибок. Проведено сравнительное исследование моделей: прототип на основе k-NN с простыми лексико-статистическими признаками (Accuracy — 0,756) и модели на базе трансформеров. Для дообучения использован корпус BEA-2019. Модель DistilBERT (Optimized) для бинарной классификации достигла Accuracy — 0,92, F1-score — 0,88. Многоклассовая модель DistilBERT показала Accuracy

— 0,856, Weighted F1-score — 0,86. Модели экспортированы в формат ONNX с динамическим квантованием для локального инференса.

Особое внимание было уделено подготовке обучающих данных. Исходный корпус ВЕА-2019 представлен в формате M2, содержащем детализированные аннотации грамматических ошибок. Для автоматизации обработки был разработан специализированный парсер, преобразующий аннотированные предложения в обучающие примеры для бинарной и многоклассовой классификации.

Для повышения устойчивости моделей к реальным пользовательским ошибкам была применена аугментация данных. В обучающую выборку искусственно добавлялись синтетические ошибки: пропуски артиклей, нарушения временных форм, ошибки согласования, а также орфографические опечатки. Это позволило существенно повысить качество работы модели на пользовательских текстах, отличающихся от академических корпусов.

Для мобильной оптимизации использовалось динамическое квантование моделей DistilBERT. Такой подход позволил сократить объём модели, уменьшить потребление памяти и ускорить выполнение инференса без критической потери качества. Экспериментально было установлено, что среднее ускорение обработки составляет около 25 процентов.

Реализован сервис MLService на стороне мобильного приложения для загрузки ONNX-моделей и выполнения локального инференса. Процесс анализа текста включает три этапа: детекцию (DistilBERT, бинарная классификация), классификацию типа ошибки (DistilBERT, многоклассовая) и интерпретацию с выдачей методической рекомендации.

Разработан генератор контекстуальных упражнений на базе дообученной модели T5-small, преобразующей отдельную лексему в грамматически корректное предложение с последующим удалением целевого слова для создания задания на вставку.

Важным компонентом системы стала реализация отказоустойчивого режима работы. В случае отсутствия интернет-соединения или недоступности серверной части приложение автоматически переключается на локальные шаблоны генерации упражнений и встроенные словари. Это позволяет сохранить непрерывность обучения и обеспечивает возможность полноценной работы в офлайн-режиме.

При проектировании пользовательского интерфейса особое внимание уде-

лялось удобству взаимодействия и минимизации когнитивной нагрузки. Навигационная структура приложения построена на основе BottomNavigationBar и включает основные функциональные разделы: словарь, упражнения, уроки, рекомендации и статистику. Такое разделение обеспечивает быстрый доступ ко всем ключевым функциям системы.

Для повышения вовлечённости пользователя была реализована система визуальной и тактильной обратной связи. При правильном выполнении упражнений используются анимации, изменение цветовой индикации и вибрационный отклик устройства. Такой подход формирует положительное подкрепление и способствует более устойчивому закреплению учебного материала.

Проведено исследование производительности на мобильных устройствах Apple. Время обработки одного предложения на iPhone 8 (A11 Bionic) — 280 мс (BERT) и 65 мс (SRS), на iPhone 15 Pro (A17 Pro) — 75 мс и 20 мс соответственно. Потребление оперативной памяти: idle state — 95 МБ, ML Session Active — 240 МБ, peak load — 310 МБ.

Реализован пользовательский интерфейс приложения на Flutter с интеграцией всех модулей. Серверная часть реализована на Python с использованием FastAPI, включает эндпоинты для генерации упражнений, предсказания SRS и семантических рекомендаций.

## ЗАКЛЮЧЕНИЕ

В ходе выполнения бакалаврской работы была спроектирована и разработана система интеллектуального обучения английскому языку в виде iOS-приложения с гибридной архитектурой, сочетающей локальный инференс нейросетевых моделей и облачные вычисления.

Проведённый анализ существующих решений показал, что большинство популярных образовательных приложений используют либо жёсткие шаблоны, либо упрощённые алгоритмы адаптивности. Разработанная система устраняет выявленные пробелы за счёт интеграции современных NLP-моделей для глубокого анализа текстов и методов машинного обучения для предиктивного управления учебным процессом.

Теоретические результаты работы включают: апробацию гибридной архитектуры on-device ML и облачных вычислений для образовательных приложений; исследование применимости модели CatBoost для задачи предсказания вероятности забывания лексики с использованием комплексного признакового пространства, включающего темпоральные, статистические и лингвистические характеристики; сравнительный анализ подходов к детекции грамматических ошибок на основе статистических методов и трансформерных архитектур.

Практические результаты подтверждены созданием полнофункционального мобильного приложения, демонстрирующего следующие возможности:

- автоматическая генерация контекстуальных упражнений на основе модели T5-small;
- интеллектуальная проверка грамматики в реальном времени с локальным инференсом моделей DistilBERT;
- персонализированная система интервальных повторений на базе CatBoost;
- семантические рекомендации новой лексики через графовую модель с алгоритмом Personalized PageRank;
- работа в офлайн-режиме с гарантией приватности пользовательских данных.

Экспериментальные исследования подтвердили высокое качество разработанных моделей (F1-score до 0,923 для системы повторений, Accuracy до 0,92 для детекции ошибок) и приемлемую производительность на мобильных устройствах (время отклика менее 300 мс даже на устройствах пятилетней давности, потребление памяти до 310 МБ).

Таким образом, поставленная цель достигнута, все задачи решены. Разработанная система представляет собой комплексное решение для персонализированного обучения английскому языку, отличающееся от существующих аналогов наличием интеллектуальной обратной связи на основе современных методов NLP, адаптивной системы запоминания с предиктивной аналитикой и гибридной архитектурой, обеспечивающей баланс между качеством обучения, производительностью и защитой данных пользователя.