**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ**

**РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ «АСТРАХАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМ. В.Н.ТАТИЩЕВА»**

Факультет цифровых технологий и кибербезопасности

Кафедра цифровых технологий

**ЛАБОРАТОРНАЯ РАБОТА №1**

**ОПРЕДЕЛЕНИЕ ГРАНИЦ ПРОЕКТА**

**«Машинный перевод с английского языка на русский язык»**

выполнена в рамках изучения дисциплины

«Архитектура систем ИИ»

Направление подготовки: 09.04.02 Информационные системы и технологии Направленность (профиль): «Проектирование и разработка

систем искусственного интеллекта»

Исполнитель:

студент группы ИИ-15

\_\_\_\_\_\_\_\_\_ Исназаров Р.К.

Научный руководитель:

к.т.н., доцент кафедры ПМИ,

\_\_\_\_\_\_\_\_\_\_\_\_ Гордеев И.И.

Астрахань – 2023

**Оглавление**

[Определение границ проекта 3](#_Toc133775249)

[Обзор существующих решений 3](#_Toc133775250)

[Наборы данных 3](#_Toc133775251)

[Список литературы 4](#_Toc133775252)

# Определение границ проекта

Одним из направлений машинного обучения является машинный перевод. Машинный перевод – одна из наиболее важных областей компьютерной лингвистики, которая включает в себя все проблемы обработки речи [1]. Машинный перевод позволяет обрабатывать большие объемы данных с высокой скоростью, что способствует уменьшению затратов на переводы.

Планируется использование языка программирования Python, библиотека для глубокого обучения Keras.

# Обзор существующих решений

Существует несколько разновидностей технологий машинного перевода. Некоторые их них:

1. Статистический машинный перевод - разновидность машинного перевода, где перевод генерируется на основе статистических моделей, параметры которых являются производными от анализа двуязычных корпусов текста
2. Машинный перевод на основе правил - перевод на основе лингвистической информации об исходном и переводящем языках.
3. Нейронный машинный перевод - перевод, который использует искусственную нейронную сеть для прогнозирования вероятности последовательности слов, обычно моделируя целые предложения в единой интегрированной модели.

Наиболее лучшее качество дает нейронный машинный перевод. Нейронные сети в переводе используются Google Translate[2], Яндекс.Переводчик[3] и Promt Neural[4]. Однако у нейронного перевода есть свои недостатки, которые зачастую решаются в совместном использовании с другими технологиями перевода.

# Наборы данных

Планируется использовать англо-русский словарь для машинного перевода с Kaggle «English-Russian dictionary for machine translate» (url: https://www.kaggle.com/datasets/hijest/englishrussian-dictionary-for-machine-translate).

Набор состоит из двух полей: слово, фраза или предложение на английском языке и их перевод на русский язык. В него входит около 36 тысяч пар переводов.

Примеры пар из словаря:

* Fire! - Пожар!
* I fell. - Я упал.
* I'm 19. - Мне девятнадцать лет.
* I went out for dinner. - Я вышел поужинать.
* I work as much as you. - Я работаю столько же, сколько и ты.
* After talking to Tom for a few minutes, it became obvious that his French wasn't very good, so Mary switched to English. - После нескольких минут разговора с Томом стало очевидно, что его французский был не очень хорошим, поэтому Мэри перешла на английский.
* Eight years ago, we were in the early stages of what would become the worst economic crisis of our lifetimes. - Восемь лет назад мы находились у истоков того, что впоследствии стало тяжелейшим экономическим кризисом нашего времени.

# Цели и задачи

**Цель:** разработать переводческую программу для англоязычных текстов на русский язык.

**Задачи:**

1. анализ существующих решений;

2. сбор данных и их обновление;

3. обучение и оценка моделей на исходном датасете;

4. выбор двух моделей с наибольшим значением точности;

5. оценка качества и скорости работы модели на новых значениях биржевых котировок;

6. выбор и развертывание наилучшей модели;

7. оптимизация выбранной модели путем квантизации и дистилляции;

8. развертывание оптимизированной модели, количественная оценка эффекта оптимизации.

# Список литературы

1. Арзамасцева И.В., Подгорный И.В. Подход к корректному машинному переводу на основе автономных адаптивных интеллектуальных систем // Современные технологии обучения иностранным языкам. Международная научно-практическая конференция (Ульяновск, 25 января 2012 года): сб. науч. тр. / отв. ред. Н.С. Шарафутдинова. Ульяновск: УлГТУ, 2012. С. 181-186
2. Wu Y. et al. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation // arXiv.org e-Print archive [сайт]: 1609.08144v2 [cs.CL]. 8 Oct. 2016. https://arxiv.org/pdf/1609.08144.pdf
3. David Talbot, Две модели лучше одной. Опыт Яндекс.Переводчика, 2018. URL: https://habr.com/ru/company/yandex/blog/350002/ (Дата обращения: 30.04.2023)
4. Нейросетевой перевод: рутина - машине, творчество – человеку. 2019. URL: https://www.promt.ru/press/blog/neyrosetevoy-perevod-rutina-mashine-tvorchestvo-cheloveku/ (Дата обращения: 30.04.2023)