

РЕРАЙТИНГ ТЕКСТА С ПОМОЩЬЮ ИИ НА УРОВНЕ, ПОЗВОЛЯЮЩЕМ ПРОХОДИТЬ ФИЛЬТРАЦИИ РАСПОЗНАЮЩИХ ИИ СИСТЕМ

AnomalyDetection

Участники:

Миронов Владимир (DS, Analytics)

Гордеев Евгений (DS, Analytics)

Шумский Игорь (DS, Analytics)

Солонченко Роман (Support, DataBase)

Ушаков Иван (Project, Visualization)

ИНСТРУМЕНТЫ ДЛЯ РЕШЕНИЯ ЗАДАЧИ

- Весь код для этого сервиса написан на Python.
- Были использованы следующие основные библиотеки: torch, transformers, nltk, steamlit.
- Использовались модели и токенизаторы: BERT, T5, Pegasus, BERT, NMT (из библиотеки HuggingFace Transformers).



ЦЕЛИ

- Разработать сервис для ререйтинга текста с помощью ИИ, который на вход получает текст, а на выходе выдаёт оценку присутствия следов ИИ по 100-бальной шкале и может очеловечить данный текст.
- Сделать простейшую реализацию сервиса, используя Streamlit.



ПРЕДСТОЯЩИЕ ЗАДАЧИ

- Проверка текста на ИИ-сгенерированность по 100-балльной шкале.
- Применить различные модели для ререйтинга, чтобы понять какая модель лучшая, используя различные оценки.
- Добавить «камуфлирование» текста.
- Сделать простейшую реализацию сервиса, используя streamlit.



РЕАЛИЗОВАНО

- Перефразирование текста с использованием семейства моделей T5, а также с помощью двойного перевода.
- Различные камуфлирования текста, которые используются как постобработка.
- Метрики сходства исходного и перефразированного текста на основе модели `rubert-base-cased-dp-paraphrase-detection`.
- Поддержка двух языков: часть моделей работает с русским, часть — с английским языком.
- Оценка качества генерации: адаптирована модель(`nlptown/bert-base-multilingual-uncased-sentiment`), определяющая, был ли текст сгенерирован, и выдающая оценку качества перефразы по 100 шкале.



РЕРАЙТИНГ С ПОМОЩЬЮ МОДЕЛЕЙ T5

- Для рерайтинга текстов на английском языке мы использовали предобученную модель `google/flan-t5-xl`, а для русского языка - `cointegrated/rut5-base-paraphraser`. Обе модели принадлежат к семейству моделей T5.

Исходный текст	Текст полученный с помощью модели
Wings of Fire: An Autobiography of Abdul Kalam” is written by Dr. APJ Abdul Kalam and Arun Tiwari. This book is an autobiography of Dr. APJ Abdul Kalam, former president of India.	This book is an autobiography of Dr. APJ Abdul Kalam, former President of India.
Многие считают, что искусственный интеллект способен полностью заменить человека, но на самом деле его возможности ограничены.	Многие полагают, что интеллектуальный интеллект способен заменить человека полностью, но в действительности возможности его ограничены.

РЕРАЙТИНГ С ПОМОЩЬЮ МЕТОДА ДВОЙНОГО ПЕРЕВОДА

- Метод основан на двустороннем машинном переводе: исходный английский текст переводится на русский, после чего выполняется обратный перевод на английский. Для этого использовались модели Helsinki-NLP/opus-mt-en-ru и Helsinki-NLP/opus-mt-ru-en. Для текстов на русском последовательность была обратной.

Исходный текст	Текст полученный с помощью модели
"Wings of Fire: An Autobiography of Abdul Kalam" is written by Dr. APJ Abdul Kalam and Arun Tiwari. This book is an autobiography of Dr. APJ Abdul Kalam, former president of India.	Fire Wings: Abdul Kalam's autobiography is written by Dr. A J Abdul Kalam and Arun Tiwari, the autobiography of the former President of India.
Многие считают, что искусственный интеллект способен полностью заменить человека, но на самом деле его возможности ограничены.	Многие считают, что искусственный интеллект может полностью заменить людей, но на самом деле его возможности ограничены.

РЕРАЙТИНГ С ПОМОЩЬЮ МОДЕЛИ PEGASUS

- Модель PEGASUS перефразирует исходный текст, создавая несколько вариантов с разной структурой и формулировками при сохранении СМЫСЛА.

Исходный текст	Текст полученный с помощью модели
The ultimate test of your knowledge is your capacity to convey it to another.	<p>The test of your knowledge is your ability to convey it.</p> <p>'The ability to convey your knowledge is the ultimate test of your knowledge.</p> <p>The ability to convey your knowledge is the most important test of your knowledge.</p> <p>Your capacity to convey your knowledge is the ultimate test of it.</p> <p>The test of your knowledge is your ability to communicate it.</p>

КАМУФЛИРОВАНИЕ ТЕКСТА

- Для того, чтобы улучшить рерайтинг с помощью моделей мы также ввели постобработку, которая должна также скрывать следы ИИ в тексте. Так мы попытались с некоторой вероятностью добавлять знаки препинания, пробелы.

Исходный текст	Текст полученный с помощью камуфлирования
Wings of Fire: An Autobiography of Abdul Kalam” is written by Dr. APJ Abdul Kalam and Arun Tiwari. This book is an autobiography of Dr. APJ Abdul Kalam, former president of India. This story sheds light on the journey of a young boy from Rameswaram to become a renowned scientist. It reflects how simple living, dedication, strong will and hard work led to success. It also shows how cultural unity impacts the lives of individuals.	Wings of Fire : An Autobiography of Abdul Kalam ” is written by Dr . APJ Abdul Kalam and Arun Tiwari . This book is an autobiography of Dr . APJ Abdul Kalam , former president of India . This story sheds light on the journey of a young boy from Rameswaram to become a renowned scientist . It reflects how simple living , dedication , strong will and hard work led to success . It also shows how cultural unity impacts the lives of individuals

КАМУФЛИРОВАНИЕ ТЕКСТА

- Также для сокрытия следов ии в текстах мы решили изменять длину предложений и переставлять некоторые слова в предложении.

Исходный текст	Текст полученный с помощью камуфлирования
Wings of Fire: An Autobiography of Abdul Kalam is written by Dr. APJ Abdul Kalam and Arun Tiwari. This book is an autobiography of Dr. APJ Abdul Kalam, former president of India. This story sheds light on the journey of a young boy from Rameswaram to become a renowned scientist. It reflects how simple living, dedication, strong will and hard work led to success. It also shows how cultural unity impacts the lives of individuals.	Wings by of is Autobiography Kalam written of Abdul An Fire: Dr. APJ of This autobiography Kalam is Abdul and an Tiwari Arun book Dr. APJ boy on This renowned India the Kalam, young to sheds of of former a a Abdul president journey light story become from Rameswaram scientist. It to will living, simple hard strong and led dedication, work how reflects success. It also how shows of unity lives the cultural impacts individuals. . .

КАМУФЛИРОВАНИЕ ТЕКСТА

- Мы решили поменять стиль текста, который получаем на выходе модели, путем замены пассивных конструкций на активные, добавления эмоциональных элементов, замены некоторых слов на их разговорные аналоги.

Исходный текст	Текст полученный с помощью камуфлирования
Предложения в активном залоге, как правило, легче воспринимаются. Кроме того, включение в текст слов, передающих эмоции или личное отношение автора, может сделать его более живым и выразительным.	Предложения в активном залоге, как правило, легче понимаются. Еще, включение в текст слов, передающих эмоции или личное отношение автора, может сделать его более живым и выразительным.

КАМУФЛИРОВАНИЕ ТЕКСТА

- Также мы добавляли в слова грамматические ошибки и заменяли слова их синонимами.

Исходный текст	Текст полученный с помощью камуфлирования
Wings of Fire: An Autobiography of Abdul Kalam is written by Dr. APJ Abdul Kalam and Arun Tiwari. This book is an autobiography of Dr. APJ Abdul Kalam	fender of Fire: sodium autobiography of Abdul Kalam is drop_a_line past medico APJ Abdul Kalam and Arun Tiwari. This book is Associate_in_Nursing aooghtryiubap of medico APJ Abdul Kalam

РЕЗУЛЬТАТ АНАЛИЗА

- Рассмотрев различные модели и оптимизировав их, в финальной версии мы решили внедрить все эти модели в сервис и дать пользователю выбирать модель самому, однако мы также ввели опцию подбора лучшей модели по метрикам, основанных на входном тексте.



СРАВНЕНИЕ С ТЕХНИЧЕСКИМ ЗАДАНИЕМ

- Мы смогли разработать сервис, используя Streamlit, который оценивает текст на следы ИИ и очеловечивает данный текст. Данная реализация полностью соответствует техническому заданию, а также мы смогли реализовать ререйтинг текстов на другой язык, внедрить несколько моделей, а также сделали функцию автоподбора лучшей модели по входному тексту и постобработку текста.



**СПАСИБО ЗА
ВНИМАНИЕ**

