**Проблема политкорректности в нейросетевом переводе**

***Менчиков Андрей Леонидович***

*Студент*

*Российский Университет Транспорта (МИИТ), Институт экономики и финансов, кафедра «Лингвистика», Москва, Россия*

*E-mail: almenchikov@mail.ru*

В наши дни активно развивается ИИ. Бытует мнение, что он заменит переводчиков в силу достаточно быстрой и качественной работы. Однако, автоматизация перевода, с одной стороны, упростила работу лингвистам, а с другой, усложнила. В первую очередь, остро встал вопрос о репрезентативности найденной информации. В нашем исследовании мы попытались выявить и проанализировать существующие сложности автоматизированного перевода и риски, связанные с ним.

 Распространенными моделями нейросетевого машинного перевода являются: Sequence-to-Sequence (Seq2Seq); Transformer; BERT (Bidirectional Encoder Representations from Transformers); GPT (Generative Pre-trained Transformer).

Несомненно, машинный перевод имеет ряд всем очевидных преимуществ. Например, он гораздо быстрее и зачастую эффективнее переводчика. Однако, на данном этапе существуют весьма значительные недостатки, которые предстоит дорабатывать. Зачастую, перевод контекстуально зависим. Согласно Большому Энциклопедическому словарю, «контекст (от лат. contextus - соединение - связь), относительно законченный отрывок письменной или устной речи (текста), в пределах которого наиболее точно выявляется значение отдельных входящих в него слов, выражений и т. п.» [1]. Согласно «Контекст – относительно законченная в смысловом отношении часть текста, высказывания» [2]. Соответственно, от него зависит решение переводчика по применению того или иного эквивалента. Без контекста сложно понять точный смысл отдельных слов.

Отдельную трудность представляют специальные тексты: вопрос многозначности, грамматическая и лексическая полисемия и омонимия терминов. Рулевская Е.С. приводит следующий пример проблемы перевода многозначного термина «парк»: «русский термин «парк» может переводиться английской лексемой «park» лишь в отдельных случаях («car park», «automobile park»). Следовательно, в переводе русской лексемы на английский язык проявляют себя ограничения. Большинство русских коллокаций с данной лексемой будут иметь другие эквивалентные термины в английском языке в зависимости от контекста «fleet», «depot», «rolling stock»…» [3].

Машинный перевод с использованием нейросетей несовершенен и может повлечь ряд рисков, если его применяют без последующей постредактуры. ИИ может порождать неточности, неправильные интерпретации или неверные переводы. Это может привести к недовольству пользователя, особенно если перевод содержит ошибки или искажает смысл оригинального текста.

Отдельно мы отмечаем проблему нарушения прав и культурных норм. У каждого народа есть своя культурно-специфическая картина мира, с рядом особенностей. Компетентность в сфере межкультурной коммуникации позволяет переводчику «лавировать», выбирая наиболее адекватный и приемлемый перевод. Машинный же перевод, как показало исследование, может осуществлять некорректный перевод текста, вплоть до нарушения юридических норм.

В нашей работе мы рассмотрели и проанализировали существующие проблемы при переводе с помощью нейросетей, в том числе Chat GPT. Мы изучили ряд новостных статей в русскоязычных и англоязычных СМИ, сделали выборку тех, что при переводе подразумевают учет политкорректности. Далее мы перевели их при помощи разных систем ИИ. Был выявлен ряд нарушений при машинном переводе, таких как, например, эйджизм «pensioners» и сексизм «firemen». В Российских же системах ИИ присутствует ряд тем, которые машина обсуждать отказывается: присутствуют стоп-слова.

Приведем пример нарушения политкорректности, которая играет важную роль в англоязычных и западных культурах, в GigaChat:

1. A doctor opened *her* book. – Доктор *открыл* свою книгу.
2. *Disabled* people need help. – *Инвалидам* нужна помощь.
3. *Студент* принёс свою книгу. – The student brought *his* book.
4. *Студентка* принесла свою книгу. – The student brought *his* book.
5. Пожарные спасают *человечество*. – The fire*men* save *mankind*.

Отметим разницу с чатом GPT, который учитывает западные нормы политкорректности:

1. *Disabled* people need help. – *Людям с ограниченными возможностями* нужна помощь.
2. *Студент* принёс свою книгу. – The student brought *his* book.
3. *Студентка* принесла свою книгу. – The *female* student brought *her* book.
4. Пожарные спасают *человечество*. – The fire*fighters* save *humanity*.

В данных примерах мы наблюдаем влияние контекста культуры на автоматизированный перевод, что напрямую зависит от баз данных, на которых базируется нейросетевой перевод. Так или иначе, без анализа корректности перевода человеком на данном этапе не обойтись. В совокупности эти пункты могут повлечь колоссальные репутационные риски как для переводчика, так и для заказчика.Машинный перевод не может заменить переводчика во всех случаях. Однако, машинный перевод может быть полезен в некоторых ситуациях, например, при переводе большого количества текста или при работе с текстами на нескольких языках, при условии постредактирования профессиональным переводчиком.

**Литература**

1) Большой энциклопедический словарь / Гл. ред. А.М. Прохоров. – 2-е изд., перераб. и доп. – М.: Большая Российская энциклопедия; СПб.: Норинт, 2000. – 1434 с.

2) Контекст – Текст: электронный // Толковый словарь Ожегова онлайн: [сайт]. – 2008 – 2017. – URL: https://slovarozhegova.ru/word.php?wordid=11761 (дата обращения: 26.10.2023).

3) Рулевская Е.С. Роль корпусного анализа в обучении иностранным языкам (на материале многозначной русской лексемы "парк" и ее английских эквивалентов) / Е. С. Рулевская // Позиция. Философские проблемы науки и техники. – 2021. – № 16. – С. 236-243. – EDN BRKKEW.