

ПРОБЛЕМЫ ОБРАЗОВАНИЯ

УДК 65.011.56

ПРИМЕНЕНИЕ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ОБУЧЕНИИ ИНОСТРАННЫМ ЯЗЫКАМ

© 2022 Е. В. Алференко

Воронежский институт высоких технологий (Воронеж, Россия)

В данной статье рассматриваются проблемы и перспективы применения искусственного интеллекта (ИИ) для обучения иностранному языку. Особое внимание уделяется их изучению и анализу с точки зрения использования в процессе преподавания немецкого языка как иностранного в неязыковом вузе. Предметом исследования служат приложения машинного перевода и машинной генерации текстов на основе искусственного интеллекта.

Ключевые слова: немецкий язык, преподавание, искусственный интеллект, естественный язык, программное обеспечение.

Технологии на основе искусственного интеллекта (ИИ) проникают во все сферы современной жизни, начиная с персонализированных поисковых запросов в Интернете и рекламных объявлений в социальных сетях, цифровых языковых ассистентов и устройств для «умного дома» (Smart-Home) вплоть до распознавания лица с целью разблокировки смартфона. Без них сегодня уже трудно представить себе обучение иностранным языкам. В данном исследовании дается представление об области обработки естественного языка и потенциале применения технологий на основе ИИ в преподавании иностранного языка на примере немецкого.

В фокусе исследования находятся две подобласти обработки естественного языка: машинный перевод и машинная генерация текстов.

Обработка естественного языка является дисциплиной, объединяющей лингвистику, информатику и исследования в области искусственного интеллекта. Их общей целью является «понимание» разговорного языка или текстов компьютерами, чтобы их можно было использовать, например, для оценки и генерации текстов, для машинного

перевода и распознавания речи. Чтобы обеспечить это понимание, нужны алгоритмы, способные находить и распознавать закономерности в огромном количестве неструктурированных данных – так называемые нейронные сети, которые иногда также называют «искусственным интеллектом». Они решают задачи, опираясь на предопределенные шаблоны. Когда нейронная сеть обучается переводу, она сначала «загружается» верными существующими переводами. Она строит связи между ними до тех пор, пока на их основе не может определить, как должны переводиться новые тексты. Этот метод также называется «машинное обучение».

Программное обеспечение для перевода в настоящее время прочно утвердилось в преподавании иностранных языков. Когда у обучающихся возникают проблемы с пониманием, они часто интуитивно тянутся к своим смартфонам, чтобы распознать значение незнакомых слов или фраз. По сравнению с обычными, тяжелыми словарями, легкий и компактный смартфон всегда под рукой, и словарный поиск в нем осуществить намного быстрее и проще. Программное обеспечение в виде специальных приложений также очень

Алференко Елена Вячеславовна – Воронежский институт высоких технологий, канд. филол. наук, доцент, e-mail: elena_alferenko@mail.ru.

популярно для создания текстов, изначально написанных на родном языке.

В то время как несколько лет назад тексты, переведенные с помощью машин, читатель мог сразу отличить из-за множества ошибок, рассогласованности и порой бессмысленности, сегодня почти невозможно сказать, человек это или компьютер сделал перевод благодаря тому, что исследования в области технологии перевода достигли огромного прогресса за последнее десятилетие. Конечно, преподаватели иностранного языка часто замечают это на занятиях, например, когда обучающийся может показать лишь посредственные результаты в области навыков продуктивного письма, но каждый текст, подготовленный им в качестве домашнего задания, выверен грамматически и орфографически корректен.

Примером быстрого прогрессивного развития программного обеспечения для перевода на основе ИИ, которым ежедневно пользуются более миллиона человек, является приложение DeepL, разработанное компанией из Кельна в 2017 году. На сегодняшний день переводы на 29 языков осуществляются с использованием статистических данных, описывающих вероятность употребления определенных слов в определенных контекстах. Учитывается одновременно несколько миллиардов параметров. Обучение нейронной сети требует довольно больших ресурсов. Например, DeepL использует суперкомпьютер с производительностью 5.1 PetaFLOPS, что примерно соответствует производительности 5000 настольных компьютеров. «Головной мозг» DeepL расположен не в Германии, а в центре обработки данных в Кеблавике, Исландия. Там программе перевода легче сохранять «холодную голову»: низкая температура окружающей среды на этом острове Атлантического океана облегчает кондиционирование центра обработки данных.

Компания регулярно принимает меры для усовершенствования своего ПО с целью обеспечения высокого качества выполненных переводов. Так, например, в 2017 году было проведено слепое исследование, в ходе которого группе профессиональных переводчиков было представлено 100 предложений, каждое из которых было переведено с помощью DeepL и трех конкурирующих программ для перевода от крупных технологических

компаний. Качество текстов, сгенерированных DeepL, оказалось примерно в три раза лучше, чем выходные тексты программного обеспечения для перевода от известной поисковой системы, поскольку люди-переводчики сочли их более естественными.

DeepL можно использовать бесплатно и без регистрации. Пользователю достаточно копировать текст для перевода в онлайн-форму, либо загрузить его как текстовый документ или презентацию. В форму можно вставить не более 5000 символов; размер загружаемых файлов не ограничен.

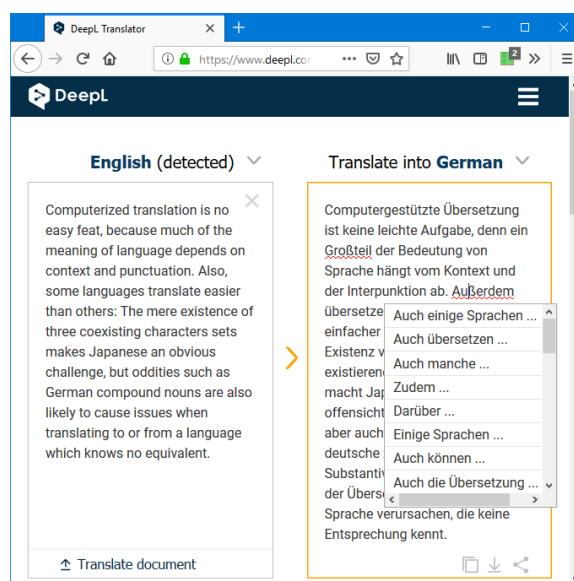


Рисунок 1. Пример перевода, сделанный переводчиком DeepL | © ДипЛ

Большим преимуществом перевода через форму является то, что результат можно улучшить в интерактивном режиме. Если получился «корявый» перевод, при нажатии на переведенный текст открывается всплывающее меню с альтернативными предложениями. Таким образом скорректировать перевод можно легко и быстро. DeepL исправляет не только отдельные слова: при необходимости программа предлагает совершенно новое предложение.

Помимо машинного перевода текстов, есть еще одна подобласть в сфере обработки естественного языка, которая в недавнем прошлом достигла большого технического прогресса – машинная генерация текстов. Одна из компаний, которая в последние годы неоднократно попадала в заголовки газет в этом контексте, – это OpenAI, организация,

основанная в 2015 году в Сан-Франциско, занимающаяся исследованиями искусственного интеллекта и разработкой, среди прочего, алгоритмов для написания и завершения текстов.

Последний из этих алгоритмов, получивший название *Generative Pretrained Transformer 3 (GPT-3)*, был впервые представлен в 2020 году в виде научной публикации. Языковая модель в настоящее время является крупнейшей в своем роде, она была обучена на 45 ТБ данных и основана на нейронной сети со 175 миллиардами параметров. Например, GPT-3 написала для издания «The Guardian» эссе о том, почему ИИ не угрожает людям. GPT-3 сочиняет стихи и прозу, выполняет переводы, ведёт диалоги, даёт ответы на вопросы, хотя никогда специально не училась выполнять эти задачи. Программное обеспечение отличается от обычных текстовых генераторов двумя ключевыми аспектами – с одной стороны, не используются готовые текстовые блоки, с другой стороны, алгоритм не специализирован на конкретной теме или области применения, а выдает подходящие предложения почти для каждой области. Как и DeepL, нейронная сеть GPT-3 основана на статистической модели, описывающей контекстно-зависимую вероятность слов и предложений.

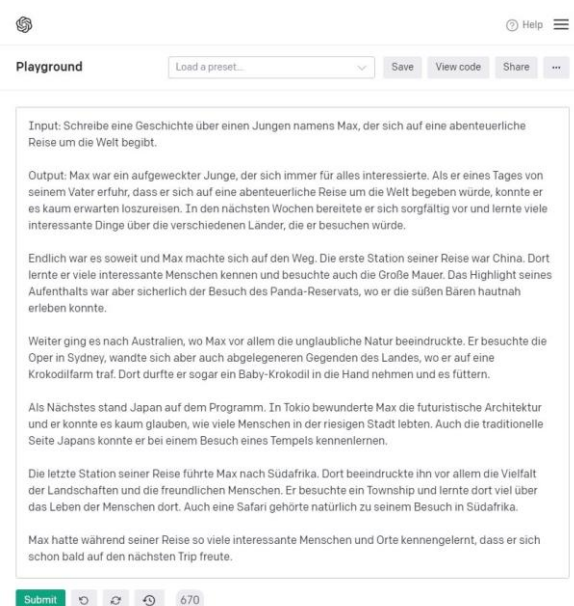


Рисунок 2. Пример истории, сделанной игровой площадкой OpenAI GPT-3 | © OpenAI

Если задать программе входные данные вроде «Напишите рассказ о юноше по имени Макс, который отправляется на увлекательную экскурсию по городу Москве», то алгоритм GPT-3 начнет делать именно это и напишет соответствующий рассказ. Однако алгоритм, который можно использовать на разных языках, не обладает собственным знанием мира, а лишь очень хорошо предсказывает, как может выглядеть следующее слово или предложение. Наконец, текстовые генераторы постепенно проникают на занятия по иностранному языку, поскольку обучающемуся, конечно, гораздо проще заставить программу сгенерировать эссе или диалог на определенной теме, чем написать его самому.

До недавних пор все возможности GPT-3 могли по достоинству оценить лишь англоязычные пользователи. «Сбер» решил исправить эту досадную оплошность и выложил в открытый доступ модель для генерации текстов GPT-3 Large с 760 млн параметров. GPT-3 можно обучить на основе русской литературы, русской и английской «Википедии», на новостных сайтах и сайтах с вопросами и ответами, на основе публичных разделов Pikabu, научно-популярного сайта 22century.ru и банковского «Банки.ру», а также с помощью проекта с открытым исходным кодом, содержащим 33 млрд русских слов, Omnia Russica. Это первый обучающий корпус на русском языке, отмечают в компании. Проектом занялись подразделения «Сбера» SberDevices, SberCloud и команда по разработке ИИ AGI NLP. Оригинальный корпус от OpenAI содержал в основном материалы на английском – соотношение последнего и других языков составляет 93:7. В случае «Сбера» разработчики сделали упор на русский – соотношение к другим языкам составляет примерно 9:1.

Помимо пользы применения ИИ для обучения иностранному языку, например, для написания текстов, существуют определенные риски использования текстовых генераторов. Например, с их помощью можно создавать без особых усилий уникальные спам-сообщения и фейковые новости. Кроме того, легко фальсифицировать научные статьи, что может способствовать распространению дезинформации. Программное обеспечение для машинного перевода также отлично подходит для составления спам-сообщений и в кон-

тексте различных форм онлайн-мошенничества, таких как мошенничество с лотереями или брачные аферы.

Еще одним риском использования программного обеспечения для перевода и генерации текста на основе ИИ является вывод проблемного контента или нецензурной лексики. Если обучающие данные программного обеспечения содержат нецензурные слова или расистские, сексистские или религиозные стереотипы, всегда существует риск того, что они также будут использованы или воспроизведены в генерируемых выходных данных.

По мере того, как программное обеспечение совершенствуется и расширяет сферу своего применения, также возрастают возможности для злоупотреблений им. Конечные пользователи должны осознавать свои обязанности в отношении того, как они используют программное обеспечение и что происходит с выходными данными.

Не вызывает сомнений, что преподаватели могут и должны конструктивно использовать программное обеспечение на основе ИИ и повышать осведомленность об ответственном использовании этих технологий в учебном процессе. Один из способов – работать с переведенными и сгенерированными машинами текстами, чтобы повысить металингвистическую осведомленность обучающихся и показать им как преимущества, так и ограничения использования программного обеспечения.

Когда дело доходит до машинного перевода, например, с английского или испанского, программное обеспечение вынуждено бороться с полом или даже с пропущенными личными местоимениями. При этом неизбежно могут возникать ошибки. Программное обеспечение для перевода также часто конфликтует с переводом метафор и аллегорий, где формулировка сильно отличается от фактического смысла высказывания. Программное обеспечение также не всегда попадает в нужную ноту с точки зрения контекста, адресатов и языкового регистра. Осознанию этих проблем может способствовать разработка заданий, направленных на исправление учащимися семантических, грамматических и стилистических ошибок и несоответствий в тексте, переведенном с помощью компьютера.

Автоматические генераторы текста также можно легко интегрировать в препода-

вание иностранного языка. Поскольку текстовые генераторы не имеют собственных знаний о мире, а генерируют последовательности слов на основе статистических данных, выходные данные не всегда соответствуют тому, что на самом деле ожидалось от программного обеспечения. Для текстовых генераторов не проблема написать фэнтезийный рассказ с конкретным сюжетом, но им гораздо сложнее написать фактические тексты на специализированные темы. На генераторы текста также влияют стилистические проблемы, упомянутые в контексте машинного перевода. Таким образом, в рамках занятий можно работать с машинно-генерируемыми текстами, предлагая обучающимся пересматривать их различными способами – будь то адаптация текстов к определенному регистру или адресату, или выполнение определенных требований к содержанию при написании художественного текста.

Преподаватели также могут использовать генераторы текстов при подготовке к уроку, например, для введения новой лексики. Например, можно задать алгоритму определенную тему и назвать некоторые слова, которые должны быть включены в генерируемый текст, и алгоритм реализует это.



Рисунок 3. Образец текста, подготовленный OpenAI GPT-3 Playground для введения заданной лексики по теме «Отпуск»| ©

СПИСОК ИСТОЧНИКОВ

1. Faes, Florian (2017): Почему DeepL занялась машинным переводом и как планирует зарабатывать деньги. Slator – Language Industry Intelligence. Доступно по: <https://slator.com/deepl-got-machine-translation-plans-make-money/>.

2. Sciforce (2021): Что такое GPT-3, как он работает и что он на самом деле делает. Доступно по: <https://medium.com/sciforce/what-is-gpt-3->

[how-does-it-work-and-what-does-it-actually-do-9f721d69e5c1](https://habr.com/ru/news/t/590193/).

3. <https://habr.com/ru/news/t/590193/>.

4. <https://www.deepl.com/ru/translator>.

5. Свиридов В. И. Лингвистическое обеспечение автоматизированных систем

управления и взаимодействие пользователя с компьютером / В. И. Свиридов, Е. И. Чопорова, Е. В. Свиридова // Моделирование, оптимизация и информационные технологии. – 2019. – Т. 7. – № 1 (24). – С. 430-438.

THE USE OF I-BASED SOFTWARE IN FOREIGN LANGUAGE TEACHING

© 2022 E. V. Alferenko

Voronezh Institute of the High Technologies (Voronezh, Russia)

This article deals with the problems and perspectives of artificial intelligence (AI) application in foreign language teaching. Particular attention is paid to their study and analysis in terms of their use in the process of teaching German as a foreign language in a non-language university. The subject of the research is machine translation and machine text generation applications based on artificial intelligence.

Keywords: german, teaching, artificial intelligence, natural language, software.