

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ОЦЕНИВАНИИ И ТЕСТИРОВАНИИ ПРИ ОБУЧЕНИИ ИНОСТРАННОМУ ЯЗЫКУ



Аннотация

В статье анализируются основные аспекты использования искусственного интеллекта в языковом оценивании и тестировании. На основе обращения к истории автоматизированного оценивания в обучении иностранным языкам проводятся параллели со специальным выпуском журнала *Language Testing* 2010 года и рассматриваются ключевые вопросы использования искусственного интеллекта с точки зрения определения тестового конструкта, разработки формата заданий, процессов оценивания и предоставления обратной связи. Приводятся конкретные примеры использования искусственного интеллекта с целью разработки контрольно-измерительных материалов. Сделан вывод о том, что искусственный интеллект может быть ценным ресурсом для составителя тестовых заданий и оценщика, но общедоступные модели на текущий момент не могут полноценно заменить работу эксперта.



Abstract

USING ARTIFICIAL INTELLIGENCE TECHNOLOGIES FOR ASSESSMENT AND TESTING IN LANGUAGE EDUCATION

The article analyzes the main aspects of using artificial intelligence in language assessment and testing. Referring to the history of automated assessment in foreign language education, parallels are drawn with a special issue of the *Language Testing* journal from 2010, exploring key issues of using AI in assessment from the perspective of defining the test construct, developing task formats, assessment processes, and providing feedback. The article provides specific examples of using artificial intelligence to develop assessment materials. The main conclusion of the article is that artificial intelligence can be a valuable resource for test developers and assessors, but current publicly available models cannot fully replace the work of an expert.



Ключевые слова

оценивание, тестирование, искусственный интеллект, тестовый конструкт, ИКТ-компетенция



Keywords

assessment, testing, artificial intelligence, ICT-competence

Использование технологий ИИ стало популярной и широко обсуждаемой в современной науке темой исследований в 2020-х годах. Подтверждение тому можно найти, ознакомившись с тематикой российских и международных конференций: например, международная конференция *EdCrunch* в 2023 году полностью сфокусировалась на данном вопросе; нейросети и искусственный интеллект в образовании стали одной из трех главных тем Восточного экономического форума,

прошедшего в 2023 году в ДВФУ. Большинство крупнейших российских вузов так или иначе касались данной темы за последние годы. Не стал исключением и Московский университет, где в 2023 году прошло две конференции, посвященные использованию ИКТ в обучении иностранным языкам. С. В. Титова отмечает, что современный этап развития педагогики характеризуется как переход от прочтения «с помощью компьютера или мобильного устройства целого курса или его фрагментов и контроль усвоенного ... к моделированию его существенных свойств, использованию элементов геймификации искусственного интеллекта, разработке так называемой обучающей метавселенной» [3, с. 28].

Профессор П. В. Сысоев на страницах журнала «Иностранные языки в школе» дает искусственному интеллекту в методике обучения иностранным языкам определение, которое будет использоваться и в данной статье: «ряд современных технологий, позволяющих компьютеру на основе сбора и анализа больших объемов данных и алгоритмов выстраивания паттернов речевого поведения разрабатывать и реализовывать методики обучения языку по индивидуальной траектории, имитировать иноязычную речемыслительную деятельность человека для решения коммуникативных задач, осуществлять автоматизированный контроль иноязычных умений и навыков обучающихся, предоставлять им обратную связь и осуществлять аналитическую работу» [2, с. 8–9]. Как следует из данного определения, автоматизированный контроль, предоставление обратной связи и аналитическая работа являются распространенными вариантами применения искусственного интеллекта в методике обучения иностранным языкам. Примечательно, что его использование с целью оценивания и тестирования широко обсуждалось в тестологии как части методики обучения иностранным языкам намного раньше, чем в обучении в целом. Во многом это связано со стремлением к повышению надежности и практичности оценивания. С другой стороны, использование компьютеров для оценивания стоит у истоков тестологии, и первые примеры подобной деятельности можно найти уже во второй половине прошлого века и в Советском Союзе, и за рубежом. В 2009 году появился первый полностью оцениваемый компьютером международный экзамен – *Pearson Test of English (Academic)*, а в 2010 году – специализированный выпуск ведущего мирового журнала, рассматривающего вопросы оценивания и тестирования, – *Language Testing* (27/3 за 2010 г.), в котором приняли участие ведущие тестологи того времени. Примечательно, что поднятые в этом выпуске его главным научным редактором Сиаоминг Си (*Xiaoming Xi*) вопросы не теряют своей актуальности и в настоящее время, несмотря на существенно упрощенную процедуру использования искусственного интеллекта. К ним, среди прочих, относятся следующие:

- 1) Насколько использование автоматизированного оценивания может ограничивать валидность оценивания с точки зрения представленности целевой сферы использования иностранного языка?
- 2) Насколько использование автоматизированного оценивания влияет на представленность целевого конструкта?
- 3) Насколько понимание алгоритмов, которые используются в процессе автоматизированного оценивания, тестируемыми может влиять на надежность оценивания и процесс прохождения тестирования?
- 4) Каким эффектом обратной связи (*washback effect*) может обладать использование автоматизированного оценивания? [5, с. 293]

Представляется полезным обратиться к данным вопросам в контексте современной ситуации, когда использование искусственного интеллекта стало более доступным, а основанные на его основе технологии широко используются в обучении иностранным языкам.

Искусственный интеллект и конструкт оценивания. Как известно, создание любой процедуры оценивания начинается с ответа на три главных вопроса: «зачем», «что» и «как». Едва ли искусственный интеллект на текущей ступени развития сможет принять решение о том, зачем проводится та или иная процедура оценивания. Таким образом, определение основной цели оценивания – мотивационной, обучающей, корректирующей, диагностической или другой – в настоящий момент целиком находится в зоне компетенций педагога. В то же время искусственный интеллект может быть использован для более подробного описания конструкта оценивания, под которым в данном случае понимается «детализированное описание объекта контроля, ... выделение совокупности измеряемых характеристик в контексте тестирования, которые составляют суть измеряемого объекта контроля ... дают возможность экстраполяции результатов теста на реальную жизненную ситуацию, в которой будет применяться измеряемый конструкт» [1, с. 4]. Предположим, что в рамках инструментальных компетенций ставится цель оценить письменную речь студентов: уровень «бакалавриат» в соответствии с ФГОС ВО по направлению 44.03.01 в соответствии с УК-4 способен осуществлять деловую коммуникацию в устной и письменной форме на государственном языке Российской Федерации и иностранном(ых) языке(ах). В ответ на запрос *«Make a list of all written documents that a professional teacher must be able to write, design or compile»* («Создай список всех документов, которые должен создавать учитель») чат *GPT Midjourney* на платформе Телеграм предлагает список из 11 позиций, включающий планы и материалы уроков, КИМ, отчеты об успеваемости, программы дисциплин, проекты, коммуникацию с родителями, исследовательские статьи, планы саморазвития и нормативные документы.

Безусловно, данный список полезен еще на этапе планирования, так как позволяет наполнить курс профессиональной коммуникации для будущих учителей и преподавателей иностранного языка целевым содержанием. Кроме того, тот вид, в котором представлен список, не позволяет четко разделить данные жанры речевой деятельности на те, которые следует формировать на уровне бакалавриата, и те, которые важны на этапе магистратуры. Кроме того, по какой-то причине искусственный интеллект не включил в состав целевых жанров письменной речи для будущих учителей и преподавателей иностранного языка резюме и сопроводительное письмо. Возникает вопрос, может ли искусственный интеллект быть использован также для уточнения тестового конструкта с точки зрения знаний, навыков и умений, которые подлежат оцениванию. Если по отношению к целевым знаниям можно признать ответ удовлетворительным, хоть и обобщенным, то в случае с тем, что искусственный интеллект обозначил как *skills*, очевидно смешаны умения, личностные качества и компетенции, что едва ли позволит использовать данное описание тестового конструкта. Таким образом, можно сделать вывод, что при определении тестового конструкта в настоящий момент искусственный интеллект может выполнять исключительно консультативную функцию, а для качественного описания требуется работа специалиста в области методики обучения иностранным языкам.

Искусственный интеллект и формат оценивания. Как было показано выше, искусственный интеллект едва ли может серьезно облегчить работу человека при описании тестового конструкта. Следующим шагом при создании любой процедуры оценивания является разработка формата оценивания, и представляется интересным рассмотреть, как могут быть использованы технологии ИИ при выборе формата и создании конкретных тестовых заданий. Очевидно, что искусственный интеллект по-прежнему может выполнять важную консультативную функцию. Так, в ответ на запрос *«Which task formats can be used to assess presentation skills of trainee teachers?»* («Какие форматы заданий могут быть использованы для оценивания умений презентации будущих преподавателей?») искусственный интеллект

предлагает восемь возможных форматов с их кратким описанием: проведение части или целого урока, предоставление взаимной обратной связи, видеопрезентации, панельное интервью, интерактивное интервью, письменная рефлексия и предоставление планов урока.

1. Teaching Demonstrations:

- *Ask the trainee teacher to prepare and deliver a short teaching demonstration on a specific topic. This can include preparing lesson plans, engaging students, and managing classroom interactions.*

2. Microteaching Sessions:

- *Conduct microteaching sessions where the trainee teacher delivers a short lesson to a small group of peers or students. This allows for focused feedback and evaluation.*

3. Peer Feedback Sessions:

- *Have trainee teachers observe and provide feedback on each other's presentations. This fosters collaboration and gives them a chance to analyze presentations from an audience perspective.*

4. Video Presentations:

- *Ask trainee teachers to record themselves delivering a presentation. This allows for self-reflection and evaluation of their own delivery style, body language, and content organization.*

Анализируя представленный ИИ список возможных форматов, можно отметить, что среди них есть как широко распространенные и используемые в нашей стране (составление планов уроков, педагогическая практика, письменная рефлексия, проведение части урока), так и те, которые используются в значительно меньшем объеме (видеопрезентации и панельные интервью). Выбирая менее распространенный формат задания, преподаватель может задействовать искусственный интеллект для разработки инструкции к нему. Так, при подготовке описания к панельным интервью стало очевидно, что это схоже с подготовкой интервью для приема на работу. Для письменной же рефлексии искусственный интеллект в течение всего нескольких секунд предложил достаточно исчерпывающее описание задания для студентов, включающее его описание, инструкцию для студентов, содержательный план (методы, оценка эффективности, реакции обучающихся, сложности, сильные и слабые стороны, план по адаптации обратной связи в дальнейшую практику и цели на будущее), критерии оценивания и инструкции по сдаче задания.

Примечательно, что данная рубрика задания (*task rubric*) содержит как инструктирующую, так и мотивирующую части и в целом может быть полноценно использована преподавателем. Также представляется интересным, что искусственный интеллект предложил критерии оценивания данного задания по содержанию. В случае его реального использования было бы необходимо добавить еще и языковые критерии оценивания. Так, дав ему задание «*Develop assessment criteria for the linguistic aspects of written reflection on a scale from 0 to 5*» («Разработай критерии оценивания для языковых аспектов письменной рефлексии по шкале от 0 до 5») можно получить в течение нескольких секунд описание 6 уровней достижений по 4 параметрам: владение языком, словарный запас и терминология, структура и когерентность, ясность изложения. Безусловно, данные критерии могут быть доработаны и уточнены, однако стоит отметить, что в течение нескольких секунд искусственный интеллект сделал работу, которая могла бы занять от получаса до часа времени опытного преподавателя. Кроме того, в качестве сильных сторон разработанной шкалы оценивания можно отметить четкую группировку внутри одного параметра по одному или двум отличительным объектам оценивания. В то же время ряд моментов может нуждаться в уточнении.

Искусственный интеллект может быть применен и для составления более традиционных тестовых заданий. Ряд отечественных и зарубежных платформ уже используют искусственный интеллект для создания тестовых заданий множественного выбора для оценивания лексических и грамматических навыков, заданий на чтение и аудирование (восстановление соответствий, «верно/неверно/не сказано», заданий с кратким ответом). Технический потенциал искусственного интеллекта для создания подобных заданий не вызывает сомнений, однако, как и в любой ситуации разработки тестовых заданий, в любом случае необходима проверка специалистом на предмет их методической адекватности. В рамках данной статьи нам особенно интересно рассмотреть использование искусственного интеллекта для разработки форматов альтернативного оценивания, описав его потенциал и ограничения.

Процесс оценивания и обратная связь. На текущий момент технологии ИИ в большинстве случаев могут эффективно осуществлять проверку письменных работ обучающихся на предмет языковой грамотности. Кроме того, особенно полезной представляется способность искусственного интеллекта в течение нескольких секунд генерировать метalingвистические комментарии, то есть объяснения допущенных обучающимся ошибок и рекомендации по их исправлению (и именно этот вид обратной связи представляется обучающимся наиболее полезным) [4]. Также при помощи технической проверки можно определять уровень используемых в работе лексических и грамматических структур. Важным ограничением в данном случае является то, что при отсутствии ошибок в рассматриваемой работе искусственный интеллект может иногда «изобретать» их в своем комментарии, подменяя корректные формулировки в работе обучающегося некорректными, в связи с чем повторная проверка специалистом остается необходимым условием.

Искусственный интеллект может эффективно выделять основные идеи, выраженные автором написанной работы, и в большинстве случаев верно оценивать использование в работе средств логической связи. В то же время стоит отметить, что на текущий момент имеющиеся в открытом доступе технологии ИИ едва ли могут осуществлять качественную проверку работы по содержанию, а также производить оценивание по заданным критериям. Так, при использовании шкалы оценивания экзамена *IELTS* искусственный интеллект на базе *GPT* будет тяготеть к оценкам 6, 6.5, 7.0. Это можно объяснить тем, что именно эти оценки наиболее часто получают обучающиеся, следовательно, в доступной для обучения искусственного интеллекта базе данных именно эти результаты и комментарии по ним являются наиболее частотными. Отдельным вопросом становится использование обучающимися искусственного интеллекта при создании письменных работ, а также возможности распознавания подобного использования самим искусственным интеллектом. К сожалению, на текущий момент искусственный интеллект не в 100 % случаев может достоверно определить, создана ли письменная работа с его использованием. Кроме того, он также будет определять как созданные искусственным интеллектом работы, которые были переведены с использованием электронных переводчиков.

Заключение. Таким образом, искусственный интеллект может быть использован на различных этапах языкового оценивания и тестирования. Примечательно, что основные вопросы, касающиеся применения искусственного интеллекта, едва ли претерпели значительные изменения за последние 15 лет с момента выхода первого полностью автоматизированного международного экзамена, оценивающего все виды речевой деятельности. Они по-прежнему касаются определения конструктора оценивания, самого процесса оценивания и предоставления обратной

связи. По-прежнему актуальным остается вопрос о том, насколько понимание механизмов и алгоритмов оценивания тестируемыми может повлиять на их поведение.

Ряд предпринятых попыток использования искусственного интеллекта для уточнения тестового конструкта показал, что ИИ может выполнять вспомогательную функцию при описании целевых ситуаций использования иностранного языка. В то же время при описании конструкта оценивания следует отметить, что ИИ едва ли справляется с данной задачей, а представленные им модели могут обладать рядом значительных терминологических неточностей, нарушений классификации и логики построения.

Представляется оправданным использование искусственного интеллекта для формулирования, проверки и уточнения рубрик тестовых заданий. Это может в значительной мере сократить затрачиваемые временные ресурсы и помочь избежать языковых неточностей. Искусственный интеллект может быть эффективно использован для создания заданий открытого и закрытого типов, однако необходима их последующая проверка экспертом. Искусственный интеллект в целом может эффективно оценивать грамотность письменной речи, вариативность и уровень используемых лексических и грамматических единиц, а также предоставлять обратную связь в форме металингвистических комментариев и пояснений. В то же время общедоступные модели не вполне позволяют полностью проводить критериальное оценивание исключительно с использованием искусственного интеллекта. Остается дискуссионным вопрос о целенаправленном обучении студентов использованию искусственного интеллекта. С одной стороны, это может позволить повысить качество представляемых работ, с другой – потенциально снизить надежность оценивания, так как способность общедоступных моделей искусственного интеллекта определять текст, созданный на его основе, пока не является абсолютно точной. Таким образом, можно прийти к выводу о том, что в большинстве ситуаций оценивания и тестирования представляется целесообразным использовать искусственный интеллект как ресурс, не заменяя им полностью работу разработчика и оценщика.



Литература

1. **Кожевникова Л. А.** Определение конструкта при составлении тестов по чтению для академических целей. Диссертация на соискание степени кандидата педагогических наук. – М.: МГУ им. М. В. Ломоносова, 2001. – 262 с.
2. **Сысоев П. В.** Технологии искусственного интеллекта в обучении иностранному языку // Иностр. языки в школе. – 2023. – № 3. – С. 6–16.
3. **Титова С. В.** Карта компетенций педагога иностранных языков в условиях цифровизации образования. – М.: Эдитус, 2023. – 270 с.
4. **Ellis R.** A typology of written corrective feedback types // ELT Journal. – 2008. – № 2. – Pp. 97–107.
5. **Xi X.** Automated scoring and feedback systems: Where are we and where are we heading? // Language Testing. – 2010. – № 27 (3). – Pp. 291–300.



Сведения об авторе

Коренев Алексей Александрович, канд. пед. наук, доцент факультета иностранных языков и регионоведения МГУ имени М. В. Ломоносова.
E-mail: studywithkorenev@gmail.com