



Известия Балтийской государственной академии
рыбопромышленного флота. 2026. № 3(77). С. 212–217
Научная статья
УДК 378
Doi:10.46845/2071-5331-2026-3-77-212-217

Разработка тестовых оценочных заданий с использованием искусственного интеллекта

Светлана Александровна Любишина¹, Александр Юрьевич Михайлов²

^{1,2}Калининградский государственный технический университет, Калининград, Россия

¹lubishina@mail.ru; <https://orcid.org/0000-0003-4309-1531>

²mikhailov59@mail.ru; <https://orcid.org/0000-0001-8264-6291>

Аннотация. В условиях стремительного развития информационных технологий искусственный интеллект становится важной частью образовательной деятельности вузов. Рассмотрены основные понятия: оценочное средство, тестирование и тестовое задание, возможности некоторых нейросетей для оптимизации труда преподавателей в учебно-методической работе. Даны практические рекомендации при использовании нейросети для разработки тестовых заданий, приведены примеры написания «промтов». В качестве вывода приведены результаты, полученные при разработке тестовых заданий с использованием искусственного интеллекта.

Ключевые слова: искусственный интеллект, процесс обучения, образовательная организация, оценочные средства, тест, тестовое задание, критерий оценки

Для цитирования: Любишина, С. А., Михайлов, А. Ю. Разработка тестовых оценочных заданий с использованием искусственного интеллекта // Известия Балтийской государственной академии рыбопромышленного флота. – 2026. – № 3(77). – С. 212–217.

В образовательном процессе вуза инновационные технологии играют исключительно важную роль. К числу таких инновационных технологий в полной мере может быть отнесено внедрение искусственного интеллекта не только в образовательный процесс, но и в цифровую среду на уровне администрации вуза, что позволит существенно оптимизировать работу преподавателей и повысить качество обучения.

Под оптимизацией работы преподавателей мы прежде всего понимаем генерацию и автоматизированное составление рабочих программ, оценочных средств и других элементов учебно-методических комплексов по изучаемым дисциплинам.

Главным критерием такой оптимизации является сокращение трудозатрат на учебно-методическую работу. Высвободившееся время преподаватели могут использовать для более творческого подхода к учебному процессу, повышая его привлекательность и мотивацию учащихся, что, безусловно, является одними из важнейших критериев повышения качества обучения.

Стоит отметить, что под термином «искусственный интеллект» мы будем понимать комплекс инструментов, позволяющих решать задачи уровня человеческого интеллекта, такие как восприятие, обучение, рассуждение, решение проблем и принятие решений, реализованных машинами, в частности компьютерными системами [1].

Общепринятого термина «оценочное средство» в нормативных документах не значится, поэтому для определения его понятия мы воспользуемся анализом словарей современного русского языка. Так, в толковом словаре Ожегова С. И. под термином «средство» понимается прием, способ действия для достижения чего-нибудь [2].

В самом полном толковом словаре Ефремовой Т. Ф. в 3-х томах термин «средство» трактуется как прием, способ действия для чего-либо [3].

Аналогичная формулировка термина средство приводится и в Большом толковом словаре русского языка Кузнецова С. А. [4].

Следует также заметить, что прилагательное слово «оценочный» имеет значение, соотносящееся по значению с существительным «оценка» [3]. То есть выражает действие по глаголу «оценить/оценивать» – определить качество, уровень чего-либо, дать оценку чему-либо. Оценить работу, выступление. Оценить знания, способности. Оценить ответ студента [4].

Таким образом, «оценочные средства» – это приемы, способы действия для определения качества, уровня чего либо, а также предмет, приспособление (или их совокупность), необходимые для достижения соответствующего качества (уровня).

Применительно к системе высшего образования понятие «оценочное средство» может трактоваться как приемы, способы действия для определения качества овладения обучающимися учебным материалом, позволяющие посредством контрольных заданий определить уровень освоения учебной дисциплины, профессионального модуля и степень сформированности компетенции как в целом, так и отдельных ее компонентов [5].

В сборнике методических рекомендаций по разработке оценочных средств для проведения профессионального экзамена, разработанного Национальным Агентством развития квалификаций, под термином «оценочные средства» подразумевается комплекс заданий, критериев оценки, используемых центрами оценки квалификации при проведении профессионального экзамена [6].

Законом об образовании (ст. 28 часть 1) установлено, что образовательная организация обладает автономией, под которой понимается самостоятельность в осуществлении образовательной, научной, административной, финансово-экономической деятельности, разработке и принятии локальных нормативных актов в соответствии с настоящим Федеральным законом, иными нормативными правовыми актами Российской Федерации и уставом образовательной организации [7].

Законом об образовании не установлено никаких требований к форме и содержанию оценочных средств, это является исключительно компетенцией образовательной организации:

- разработка и принятие локальных нормативных актов;
- разработка и утверждение образовательных программ образовательной организации, если иное не установлено Законом об образовании;
- осуществление текущего контроля успеваемости и промежуточной аттестации обучающихся, установление их форм, периодичности и порядка проведения.

Способ представления оценочных материалов в составе образовательной программы определяется вузом в локальном нормативном акте с учетом требований Федерального государственного образовательного стандарта, который должен учитывать выбранную методику обучения. Среди большого разнообразия оценочных средств особое место занимает тестирование.

Под тестированием понимают метод измерения (оценки), который включает в себя разработку системы тестовых заданий с заданными качественными и количественными характеристиками для объективного и надежного оценивания предметов оценки, стандартизированную процедуру проведения измерения, методы статистической обработки, анализа и интерпретации полученных результатов.

К достоинствам тестирования относят: исключение субъективности оценивания; стандартизацию условий, для всех экзаменуемых; оперативность и экономичность; оптимизацию степени сложности; возможность использования в онлайн-режиме.

К недостаткам следует отнести: опасность «слепых» (автоматических) ошибок; присутствие элемента случайности; трудоемкость и затратность подготовки тестового инструментария; затруднительность диагностирования вероятностных, абстрактных, креативных и методологических профессиональных знаний.

Следует отметить, что именно наличие тестов в составе фондов оценочных средств по образовательным программам предписывается Приказом Минобрнауки России от 18.04.2023 № 409, в соответствии с которым осуществляется диагностическая работа по определению уровня достижения результатов обучения и (или) освоения образовательной программы высшего образования по соответствующему направлению подготовки или специальности [8].

Тест – система специально разработанных и структурированных заданий стандартной формы, выполнение которых проходит с ограничением времени в равных для всех испытуемых условиях, поддается количественному учету по заранее спроектированной технологии обработки и анализа результатов.

Особенностью современных образовательных нейросетей является их постоянное совершенствование. То, что было технологическим прорывом вчера, сегодня становится стандартом индустрии, а завтра будет дополнено ещё более впечатляющими возможностями.

Когда в конце 2022 года ChatGPT появился в открытом доступе, это стало своеобразным рубиконом для мировой системы образования. Преподаватели оказались перед лицом новой технологической реальности, важным аспектом которой является автоматизация рутинных задач



педагогов, к числу которых относится и составление тестовых заданий. В то же время выбор преподавателем нейросети для наилучшего решения этой задачи всё ещё представляет определенную сложность.

В ходе анализа научных исследований отечественных специалистов в области возможности эффективно использовать нейросети для оптимизации работы преподавателей и повышения качества учебного процесса был протестирован ряд известных нейросетей, итоги тестирования приведены в таблице 1 [9].

Таблица 1

Результаты тестирования нейросетей

№	Название нейросети	Точность	Полнота
1	GPT-3 и GPT-4 (OpenAI)	0,92	0,90
2	BERT и RoBERTa (Google)	0,91	0,89
3	DALL-E и Stable Diffusion	0,87	0,86
4	TesorFlow и PyTorch	0,93	0,91
5	DeepSpeech (Mozilla)	0,85	0,86
6	ChatGPT Gigachat	0,82	0,83

Из данных, представленных в таблице 1, видно, что все протестированные модели нейросетей продемонстрировали достаточно высокие показатели точности и полноты, что указывает на их высокую эффективность в решении задач, связанных с анализом данных и обработкой результатов. Основываясь на этих результатах, при необходимости можно проводить дальнейшие корректировки всех моделей нейросетей для их адаптации к образовательной среде вуза.

При разработке тестовых заданий стоял выбор между ChatGPT Gigachat и DeepSpeech. Обе нейросети способны создать тексты разной сложности, стиля, тематики, форматов, имеют четко прописанный «промт».

«Промт» (от англ. prompt – подсказка) – это запрос, команда или набор инструкций, которые пользователь сообщает нейросети для выполнения определенной задачи. Написание «промта» имеет целью помочь нейросети лучше понять намерение человека и выдать наиболее релевантный результат.

ChatGPT Gigachat и DeepSpeech ориентированы на русскоязычных пользователей, но при недостатке исходных данных обе модели склонны к «галлюцинациям» – генерации правдоподобной, но фактически неверной информации. Нейросеть может с уверенностью предложить несуществующие источники, исказить исторические факты или выдать ошибочные научные данные.

В конечном итоге, взвесив все «за» и «против», в нашем исследовании предпочтение было отдано нейросети DeepSpeech. Стоит также отметить, что разработка тестовых заданий с использованием искусственного интеллекта должна быть подчинена определенному алгоритму.

Шаг 1. Определение предмета оценивания. На этом этапе отбираются знания и умения по соответствующим компетенциям. К примеру, для дисциплин профессионального цикла это будут общепрофессиональные (ОПК) либо профессиональные компетенции (ПК), которые выбираются вузом самостоятельно по одной-двум общепрофессиональным функциям соответствующего профессионального стандарта. Знания и умения определяют содержание теста, являются фундаментом выполнения трудовых функций (ТФ) и трудовых действий (ТД), входящих в описание квалификации и играют важную роль в обеспечении успешной профессиональной деятельности.

Тогда на первом шаге, написание «промта», к примеру, по дисциплине «Организация строительного производства», соотносимой с профессиональным стандартом «Специалист по организации строительства» (утвержден Приказом Минтруда России от 21.10.2021 №747), может иметь следующий вид.

Промт 1. Проанализируй планируемые результаты обучения по дисциплине «Организация строительного производства», соотнесенные с установленными общепрофессиональными компетенциями (ОПК) или профессиональными компетенциями (ПК).

Выпиши необходимые знания (З) и умения (У) строго из текста Фонда оценочных средств (ФОС). Включи только те знания и умения, которые непосредственно относятся к рассматриваемой компетенции. Запрещается добавлять или изменять перечень.

Прикрепи исходный документ к «промту» (например, ФОС).

Шаг 2. Выбрать тип и количество заданий для каждого предмета оценки. Как правило, тесты разрабатываются четырех основных типов:

– Задание закрытой формы – испытуемый выбирает ответ из нескольких предложенных вариантов, при этом только один из вариантов ответов правильный, остальные – отвлекающие, так называемые дистракторы, которых должно быть не менее 4-х.

– Задание на установление соответствия – испытуемый устанавливает соответствие между элементами двух множеств (двух списков), количество единиц в каждом из множеств не должно совпадать (одно из множеств должно быть хотя бы на один элемент больше, или вариант ответа может быть выбран более одного раза).

– Задание на установление правильной последовательности (перегруппировка, тест на систематизацию) – например, испытуемый восстанавливает правильную последовательность технологических операций, действий, событий.

– Задание открытой формы (тест со свободно конструируемым ответом, как правило, содержит одно-два слова) – дополнение, самостоятельное формулирование ответа или подстановка пропущенного слова на месте прочерка, не допускается развернутых ответов.

К наиболее общим требованиям к тестам следует отнести следующие: длина или количество тестовых заданий должно быть рассчитано на среднее время тестирования, которое должно составлять не более 60–90 минут, поэтому длина теста, как правило, составляет от 30 до 50 заданий для технических, от 60 до 100 – для экономических и гуманитарных предметов. Количество заданий в базе данных тестовых заданий должно быть примерно в 10–20 раз больше усредненной длины теста.

Формулировка заданий должно быть краткой, длина основной части не должна превышать 10–15 слов (кроме кейсов), а длина ответов не превышать 3–5 слов, при этом в основе задания допускается одно придаточное предложение.

При этом шаге «промт» не составляется.

Шаг 3. Проранжируй знания и умения по степени их важности (соответствующие трудовым функциям) – для технических дисциплин это особо значимо.

Промт 2. Оцени степень влияния рассматриваемых знаний и умений из промта 1 на общую безопасность или тяжесть ошибки, допущенной при осуществлении трудовой функции по трехбалльной шкале. Объясни свой выбор для каждого действия:

– какие знания и умения, соответствующие общепрофессиональным или профессиональным компетенциям, наиболее критичны для безопасности?

– какие потенциальные последствия ошибки возможны при осуществлении трудовых функций (несчастный случай, аварии, штраф)?

– насколько вероятны такие ошибки?

Ответ представь в табличной форме, ранжируя ошибки от высшей степени к низшей.

Шаг 4. Составь тестовое задание. На этом этапе в «промте» следует конкретизировать требования к тестовым заданиям, их вид, количество и другие требования. Чем подробнее будут описаны требования, тем качественнее в конечном итоге будет тест. Предпочтение следует отдавать тестовым заданиям, относящимся с высшей степени критичности при ранжировании на основании промта 2, к тестовым заданиям с высокой критичностью следует относить до 80 % от общего количества тестовых заданий.

Промт 3. Возможное написание «промта» в данном случае не приводится из-за его чрезмерной громоздкости и целого ряда особенностей.

Шаг 5. Установить правила обработки результатов тестирования. Тестовое задание должно иметь критерий оценки, которую лучше производить баллами с пересчетом в проценты. Самый простой и надежный вариант – оценка каждого задания по дихотомической системе: «правильный ответ – 1 балл, неправильный – 0 баллов». При этом следует отметить, что нет единого мнения по поводу тестовых заданий, содержащих множество ответов. Вводить ли весовые значения для каждого правильно выбранного ответа, либо оценивать их на «ноль» при наличии хотя бы одного неправильно выбранного ответа. Авторы этих строк придерживаются мнения, что наиболее правильным была бы оценка результатов тестирования с учетом весовых значений за каждый правильный ответ.



Распределение баллов может зависеть от процента правильных ответов и может лежать в различных границах, например, результат «сдан» выставляется для границы «более 65 % правильных ответов». Однако при наличии «неравновесных» заданий, заданий «особой важности», получение 65 % без выполнения таких заданий не может обеспечить положительный результат тестирования. Остается только резюмировать, что выбор критериев оценки тестирования остается прерогативой разработчика.

Промт 4 (в простейшем виде, без учета весов). На основании промта 2 и 3 установи правила обработки результатов тестирования. Базовый принцип оценивания: Задание – Правильный ответ 1 балл – Неправильный ответ 0 баллов. Система: Дихотомическая (бинарная) оценка. Запрещено (или разрешено): Частичные баллы (веса), субъективные интерпретации.

Расчет итогового результата без весов: $(\sum \text{баллов} / \text{кол-во заданий}) \times 100 \%$

Порог сдачи теста – 65 % (рекомендуемое) – сдал (зачтено) или устанавливаются другие значения при использовании четырехбалльной системы в качестве оценки.

Шаг 6. Выполнить методическую и содержательную экспертизу. Экспертиза тестовых заданий также может быть выполнена с помощью нейросети. При положительном заключении методиста или эксперта тестовое задание может быть рекомендовано для проведения промежуточной или итоговой аттестации.

Шаг 7. Оформить спецификацию теста и составить ключ. На данном этапе производится оформление оценочного средства в соответствии с принятыми в вузе правилами или шаблонами локальных нормативных документов.

Таким образом, нами рассмотрена одна из вспомогательных функций современных нейросетей, направленная на сокращение трудозатрат при разработке тестовых заданий. В этой связи вполне уместным будет привести данные об информированности преподавателей о возможностях нейросетей по результатам анкетирования.

Результаты анкетирования приведены в таблице 2.

Таблица 2

Результаты анкетирования

Вопросы	Варианты ответов	Количество ответов	Доля ответов в %
1. Осведомлены ли вы о возможностях использования нейросетей в сфере образования?	Да	414	53,91
	Нет	354	46,09
2. Используете ли вы нейросети в своей преподавательской практике?	Да	300	39,07
	Нет	468	60,93
3. Видите ли вы полезность нейросетей для оптимизации работы преподавателя?	Да	300	39,07
	Нет	468	60,93
4. Видите ли вы полезность нейросетей для повышения качества образования?	Да	369	48,05
	Нет	399	51,95
5. Исключаете ли вы для себя технические сложности при использовании нейросетей в образовании?	Да	270	35,16
	Нет	498	64,84
6. Считаете ли вы, что необходимо разработать нормативную документацию для использования нейросетей в вузе?	Да	402	52,34
	Нет	366	47,66
7. Считаете ли вы, что интеграция нейросетей может изменить качество образования в вашем вузе?	Да	390	50,78
	Нет	378	49,22
8. Поддерживаете ли вы интеграцию нейросетей в учебный процесс вашего вуза?	Да	380	49,48
	Нет	388	50,52

Анкетирование проводилось в ноябре-декабре 2024 года, в нём приняли участие 768 преподавателей из Кузбасского государственного технического университета им. Т. Ф. Горбачева и Кемеровского государственного медицинского университета [9].

Проанализировав данные, приведенные в таблице 2, нетрудно заметить, что более половины преподавателей имеет низкую осведомленность о возможностях применения нейросетей в своей работе.

И, в заключении, приведем собственные данные, полученные по результатам разработки тестовых заданий для дисциплин профессионального цикла по направлению «Строительство» с использованием нейросети DeepSpeech. Если в обычных условиях на разработку теста в трех вариантах с общим количеством 150 тестовых заданий уходило в среднем 3–4 рабочих дня, то после прохождения авторами этих строк повышения квалификации при Национальном Агентстве развития квалификаций времени на разработку такого же количество тестовых заданий потребовалось всего лишь около 3 часов.

Список источников

1. Рассел, С., Норвиг, П. Искусственный интеллект: современный подход / пер. с англ. и ред. А. В. Слепцова. – 4-е изд. – Санкт-Петербург : Диалектика, 2021. – Т. 1–3.
2. Ожегов, С. И. Толковый словарь русского языка / под ред. проф. Л. И. Скворцова. – 28-е изд. – перераб. – Москва : Мир и образование, 2014. – 1376 с.
3. Ефремова, Т. Ф. Новый словарь русского языка. Толково-словообразовательный. – Москва : Русский язык, 2000. – URL: <http://www.efremova.info> (дата обращения: 05.01.2026).
4. Большой толковый словарь русского языка: А–Я / РАН ; Ин-т лингв. исслед. ; сост., гл. ред. канд. филол. наук С. А. Кузнецов. – Санкт-Петербург : Норинт, 1998. – 1534 с.
5. Коновалова, Ю. О. Оценочные средства : определение, формирование, использование : методические рекомендации / Ю. О. Коновалова, Т. А. Борзова, М. Ю. Криницкая. – Владивосток : Издательство ВВГУ, 2022. – 24 с.
6. Разработка и применение оценочных средств для проведения профессиональных экзаменов : сборник методических рекомендаций – Москва : АНО «Национальное агентство развития квалификаций», 2023. – 294 с.
7. Федеральный закон «Об образовании в Российской Федерации» от 29.12.2012 № 273-ФЗ (в редакции 2025 года). – URL: https://minobr.gov-murman.ru/activities/federalnyy-zakon-ot-29.12.2012-n-273_fz-_red.-ot-31.07.2025.pdf?ysclid=mu211opa3r109421789 (дата обращения: 11.02.2026).
8. Приказ Минобрнауки Российской Федерации от 18.04.2023 № 409 «Об утверждении аккредитационных показателей по образовательным программам высшего образования, методики расчета и применения аккредитационных показателей по образовательным программам высшего образования». – URL: https://obrnadzor.gov.ru/wp-content/uploads/2023/08/prikaz-minobrnauki-rossii-ot-18_04_2023-№-409.pdf (дата обращения: 11.02.2026).
9. Ширококолобова, А. Г., Ларионова, Ю. С. Алгоритм интеграции нейросетей в цифровую среду вуза для повышения качества образовательного процесса // Педагогика. Вопросы теории и практики. – 2025. – Т. 10. – Вып. 6. – URL: <http://www.pedagogy-journal.ru> (дата обращения: 06.01.2026).

Информация об авторах

С. А. Любишина – кандидат педагогических наук, доцент;
А. Ю. Михайлов – кандидат педагогических наук, доцент.

Статья поступила в редакцию 01.07.2026; одобрена после рецензирования 03.08.2026; принята к публикации 10.08.2026.