

ТЕОРЕТИЧЕСКИЕ И ПРАКТИЧЕСКИЕ ВОПРОСЫ ОБУЧЕНИЯ ИНОСТРАННЫМ ЯЗЫКАМ

УДК 373.5:372.881.1:004.8
DOI 10.51955/2312-1327_2025_4_149

ПРИМЕНЕНИЕ ИНСТРУМЕНТОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ АВТОМАТИЧЕСКОЙ ПРОВЕРКИ ЯЗЫКОВЫХ НАВЫКОВ

Евгений Евгеньевич Никулин,
orcid.org/0009-0005-8956-6856,
Иркутский государственный университет,
Карла Маркса, 1
Иркутск, 664003, Россия
Eugnik303@gmail.com

Светлана Юрьевна Богданова,
orcid.org/0000-0002-4804-4788,
доктор филологических наук, профессор
Иркутский государственный университет,
Карла Маркса, 1
Иркутск, 664003, Россия
rusjap@mail.ru

Аннотация. В статье рассматривается применение инструментов искусственного интеллекта (ИИ) для автоматической проверки письменных работ учащихся 10–11 классов. Актуальность исследования обусловлена растущим запросом на интеграцию цифровых инструментов в образовательный процесс и необходимостью модернизации подходов к оценке языковых навыков. В ходе экспериментального исследования проведен сравнительный анализ эффективности, точности и объективности автоматизированной проверки эссе с помощью ИИ и традиционной экспертной оценки. Результаты демонстрируют высокую эффективность ИИ в выявлении грамматических и лексических ошибок, однако отмечаются значительные расхождения с экспертной оценкой в критериях структуры, логики и стиля текста. На основе анализа сформулированы рекомендации по гибридной модели оценивания, сочетающей технологические возможности ИИ и педагогическую экспертизу.

Ключевые слова: искусственный интеллект, автоматическая проверка, языковые навыки, оценка эссе, цифровизация образования, машинное обучение, английский язык.

APPLICATION OF ARTIFICIAL INTELLIGENCE TOOLS FOR AUTOMATIC ASSESSMENT OF LANGUAGE SKILLS

Evgeniy E. Nikulin,
orcid.org/0009-0005-8956-6856,
Irkutsk State University,
1, Karl Marx street
Irkutsk, 664003, Russia
Eugnik303@gmail.com

*Svetlana Yu. Bogdanova,
orcid.org/0000-0002-4804-4788,
doctor of philology, professor
Irkutsk State University,
1, Karl Marx street
Irkutsk, 664003, Russia
rusjap@mail.ru*

Abstract. The article discusses the use of artificial intelligence (AI) tools for the automatic assessment of written works of 10th–11th grade students. The relevance of the study is driven by the growing demand for the integration of digital tools into the educational process and the need to modernize approaches to assessing language skills. An experimental study conducted a comparative analysis of the effectiveness, accuracy, and objectivity of automated essay checking using AI and traditional expert assessment. The results demonstrate the high effectiveness of AI in identifying grammatical and lexical errors, but significant discrepancies with expert assessment are noted in the criteria of text structure, logic, and style. Based on the analysis, recommendations are formulated for a hybrid assessment model that combines the technological capabilities of AI and pedagogical expertise.

Keywords: artificial intelligence, automatic assessment, language skills, essay evaluation, education digitalization, machine learning, the English language.

Введение

Современное образование сталкивается с необходимостью модернизации подходов к оценке языковых навыков учащихся, особенно в контексте старшей школы. Традиционные методы проверки (эссе, тесты, дискуссии) обладают существенными ограничениями: субъективностью, временной затратностью и фрагментарностью обратной связи. Эти проблемы актуализируют поиск инновационных решений, среди которых ключевое место занимают технологии искусственного интеллекта (ИИ), способные автоматизировать рутинные задачи и повысить объективность оценивания.

Развитие алгоритмов обработки естественного языка (Natural Language Processing, NLP) и машинного обучения открывает возможности для автоматической проверки грамматики, лексики, структуры текста и даже стилистических особенностей. Интеграция технологий искусственного интеллекта в образование способна преодолеть ограничения традиционных методов, автоматизируя проверку рутинных элементов и дополняя её анализом связности текста, стилистического разнообразия. Однако эффективность таких систем требует верификации, особенно в контексте старших классов, где работы учащихся должны соответствовать строгим академическим стандартам.

Целью настоящего исследования являлось проведение комплексного анализа функциональных возможностей инструментов искусственного интеллекта в контексте автоматизированной оценки письменных работ учащихся старших классов (10-11 классы) с фокусом на систематическое сравнение ключевых операционных характеристик алгоритмов с традиционными методами экспертной проверки, осуществляющей педагогами.

Материалы и методы

В качестве основного инструмента искусственного интеллекта для автоматической проверки языковых навыков в данном исследовании использовалась нейросетевая языковая модель DeepSeek, представляющая собой генеративную AI-систему на основе архитектуры трансформер, обученную на обширных корпусах текстовых данных на множестве языков. Данная модель способна к пониманию, анализу и генерации естественного языка, что делает её эффективным инструментом для обработки и оценки письменных работ. Для использования DeepSeek в исследовании применялась веб-версия платформы, доступная по ссылке: [<https://chat.deepseek.com>]. Методика применения включала несколько этапов: сначала разрабатывались специализированные промты (text prompts) – структурированные запросы на русском языке, содержащие четкие инструкции для модели по анализу эссе по пяти ключевым критериям (грамматика, лексика, структура, логика и стиль); затем каждый текстовый материал последовательно вводился в интерфейс чата DeepSeek вместе с разработанным промтом, после чего модель выполняла комплексный лингвистический анализ текста, идентификацию и классификацию ошибок, оценку по установленным критериям и генерацию развернутой обратной связи; на заключительном этапе результаты работы DeepSeek фиксировались в формате балльных оценок по каждому критерию, детализированного перечня выявленных ошибок и рекомендаций по улучшению текста, что в последующем подвергалось сравнительному анализу с экспертными оценками для верификации эффективности системы [Галагузова и др., 2024].

В рамках исследования предполагалось осуществить многоаспектную верификацию ИИ-систем по четырём критически значимым параметрам:

Эффективность – это способность алгоритмов к идентификации и категоризации лингвистических ошибок различного типа.

Точность – это способность минимизировать как ложноположительные (ошибочная маркировка корректных языковых конструкций), так и ложноотрицательные (пропуск фактических ошибок) результаты.

Скорость – временные затраты на выполнение полного цикла обработки текста.

Объективность – это независимость от субъективных факторов, характерных для человеческого оценивания.

Эмпирическая база и методы исследования

Эмпирической базой исследования выступили письменные работы (эссе) пяти учащихся 10-11 классов, отобранные из 40 работ с учётом вариативности уровня языковой компетенции (базовый A2 и средний B1 по адаптированной шкале CEFR [North, 2014]).

Методология включала поэтапное сопоставление результатов автоматической проверки, выполненной с использованием ИИ-инструментов, с детализированными экспертизами, проведёнными квалифицированным преподавателем.

Для количественной оценки эффективности применялась матрица ошибок, фиксирующая совпадения и расхождения в идентификации недочётов, в то время как качественный анализ направлен на выявление паттернов, связанных со стилистическими и логическими особенностями текстов.

Критерии и инструментарий оценивания



Рисунок 1 – Критерии оценки качества текста

Оценка проводилась по пяти заранее определённым критериям, максимальный балл за эссе составлял 45:

Грамматика (10 баллов): оценка правильности формообразования и синтаксических конструкций.

Лексика (10 баллов): оценка разнообразия словарного запаса, использования синонимов и терминов.

Структура (10 баллов): наличие и качество введения, основной части с аргументами и заключения.

Логика (10 баллов): связность аргументов, наличие обоснований и примеров, плавность переходов.

Стиль (5 баллов): соответствие академическому регистру, синтаксическое разнообразие.

Для автоматической оценки использовался промт (prompt), адаптированный под возможности нейросетевой модели. Пример промта, приведённый в работе, гласил:

«Проанализируй представленное эссе по критериям: Лексика: оцени разнообразие словарного запаса... Структура: проверь наличие введения с формулировкой проблемы... Логика: определи, связаны ли абзацы между собой... Стиль: выяви разговорные выражения... Оцени по данным тебе критериям, поставь соответствующую оценку».

Ход эксперимента

1. Подготовительный этап: Учащиеся написали эссе на заданную тему, соответствующую требованиям ЕГЭ. Тексты были оцифрованы.
2. Этап автоматической проверки: Оцифрованные эссе загружались в ИИ-систему, где проходили комплексный анализ по заданным критериям. Система генерировала оценку и развернутую обратную связь.
3. Этап экспертной оценки: Тексты тех же эссе независимо оценивались учителем-экспертом по идентичной критериальной шкале.
4. Сравнительно-сопоставительный анализ: Производилось детальное сравнение оценок ИИ и эксперта по каждому критерию для каждого эссе. Фиксировались и анализировались случаи радикального расхождения.
5. Дополнительные методы: Эксперимент также включал анкетирование участников для оценки их восприятия ИИ-обратной связи и сравнение времени проверки алгоритмом и учителями.

Ключевые термины:

- NLP (Natural Language Processing / Обработка естественного языка): это область исследований и приложений, которая изучает, как компьютеры могут использоваться для понимания текста или речи на естественном языке и манипулирования ими для выполнения полезных задач [Chowdhary, 2020, р. 603].
- Промт (Prompt): это запрос, с помощью которого ставится задача языковой модели [Алексеева и др., 2024, с. 51].
- Верификация: процедура проверки и подтверждения точности и объективности работы ИИ-системы. Проверка на соответствие этическим нормам. Основной формой этой верификации является использование комплексных тестов Тьюринга [Карпов и др., 2018, с. 86].
- Матрица ошибок (Confusion Matrix): является таблицей с четырьмя различными комбинациями прогнозируемых и фактических значений [Пупков, 2021, с. 35]. Другими словами, инструмент для количественной оценки точности классификации, показывающий соотношение верно и ошибочно идентифицированных случаев.
- Критериальное оценивание: позволяет не только учитывать возрастные особенности детей, их академический опыт, но и точно воздействовать на наиболее проблемные зоны с помощью «удорожания» соответствующего критерия. Или система оценки, основанная на сравнении работы учащегося с заранее установленными и описанными критериями [Ступницкая, 2015, с. 53].

- Ложноположительные/ложноотрицательные результаты – типы ошибок алгоритма: ложное срабатывание на правильной конструкции и пропуск существующей ошибки [Свинцов, 2016, с. 63].
- Гиперкоррекция: избыточное и часто необоснованное исправление алгоритмом конструкций, которые могут быть допустимы в определенном контексте, проявляет себя на разных уровнях и в разных аспектах языка: фонетическом, грамматическом, лексико-семантическом, стилистическом, орфографическом [Петров, 2022, с. 105].
- Гибридная (этапная) модель оценки: предложенная в исследовании модель, сочетающая последовательную автоматизированную и экспертную проверку.

Таким образом, разработанная и апробированная методика представляет собой интегративный подход, сочетающий количественные метрики (время проверки, процент точности) с качественными критериями (адаптивность к контексту, объективность), что позволяет всесторонне оценить потенциал и ограничения ИИ в образовательной практике.

Дискуссия

Применение систем искусственного интеллекта в образовании неоднократно становилось предметом обсуждения [AI Technologies in Training..., 2024; The future of teaching and learning..., 2025; Tan et al., 2025], при этом обсуждаются как практические вопросы использования различных систем и те проблемы, которые можно успешно решать с помощью ИИ, так и теоретические вопросы, связанные с местом учителя в образовании, влиянием ИИ на его профессиональное развитие и с вероятным изменением отношения «учитель-ученик» [Tan et al., 2025]. Основными сферами образовательной практики, в которых появление ИИ наиболее приветствуется, являются планирование, поиск учебных материалов и система оценивания [Проверка домашних заданий..., 2021]. Многие авторы анализируют возможности применения ИИ в преподавании иностранных языков [Ekizer, 2025; Rahimi, 2025], отмечая улучшение результатов обучения английскому языку как иностранному (EFL) благодаря основанной на ИИ педагогике, особенно в таких аспектах как точность написания, беглость речи, мотивация обучающихся. В России с 2022 года разрабатываются нейросети, способные качественно проверять эссе на русском и английском языках, обращая внимание на речевые, грамматические, логические, фактические и этические погрешности [Создана программа..., 2022].

В настоящее время существуют платформы для автоматизированной оценки письменной речи (automated writing assessment (AWE) platforms), такие как Grammarly, ProWritingAid, Ginger Software, PaperRater, которые могут находить грамматические, структурные и стилистические ошибки, экономя время учителя, обеспечивая справедливую оценку и помогая индивидуальному обучению [Automated writing evaluation..., 2024]. Представляется, что при самостоятельной подготовке обучающихся к

экзаменам по иностранному языку подобные инструменты ИИ представляют наибольший интерес, поскольку, в отличие от тестовой части экзамена, где можно проверить правильность выполнения заданий по «ключам», задания с открытым ответом по-прежнему требуют обращения к учителю или репетитору. Модели, обеспечивающие проверку письменных работ, включая эссе, открывают новые возможности для обучающихся.

Современные системы ИИ функционируют на основе машинного обучения, где алгоритмы обучаются на обширных массивах данных. Модель, предназначенная для проверки эссе, анализирует тысячи текстов, размеченных педагогами, чтобы научиться распознавать грамматические ошибки и стилистические недочеты. Продвинутые NLP-модели, такие как GPT-4, способны оценивать логическую связность эссе и предлагать рекомендации по улучшению структуры текста.

Методики автоматической проверки можно классифицировать по типам решаемых задач: грамматический анализ, оценка стиля и связности текста, антиплагиат. Однако их внедрение поднимает этические и практические вопросы, включая риск алгоритмической предвзятости и чрезмерной стандартизации, когда творческие работы «подгоняются» под шаблоны, понятные алгоритмам.

Результаты

В рамках экспериментального исследования для проведения детального сравнительного анализа была сформирована репрезентативная выборка из пяти письменных работ учащихся 10-11 классов. Исходная совокупность состояла из 15 эссе, написанных на заданную тему в соответствии с требованиями к итоговой аттестации. Отбор конкретных работ для последующего сопоставления оценок, выставленных искусственным интеллектом и экспертом-педагогом, осуществлялся на основе строгих критерии, главным из которых стало техническое качество оцифровки текста.

Первостепенным и определяющим фактором выбора являлась высокая точность и читаемость электронных версий работ. Эссе с нечеткими сканами, распознанные с ошибками или содержащие технические артефакты, были исключены из углубленного анализа. Это было обусловлено необходимостью минимизировать сторонние погрешности, которые могли бы исказить результаты проверки как со стороны ИИ, работающего строго с предоставленным цифровым текстом, так и со стороны эксперта, анализирующего его содержание. Таким образом, пять отобранных работ гарантировали чистоту экспериментальной базы, позволяя сфокусироваться именно на сопоставлении методик оценки, а не на исправлении технических недочетов исходных данных.

Кроме того, в рамках обеспечения репрезентативности выборки учитывался диапазон исходного качества работ от слабых до достаточно уверенных, что позволило проанализировать эффективность ИИ-инструментов применительно к разным уровням языковой подготовки учащихся.

В Таблице 1 представлены результаты проверки 5 эссе школьников искусственным интеллектом и экспертом, а также разница в выставленных ими баллах.

Таблица 1 – Сравнение итоговых оценок

Эссе	Итоговая оценка ИИ	Итоговая оценка эксперта	Разница
№1	25/45	29/45	-4
№2	22/45	31/45	-9
№3	15/45	21/45	-6
№4	23/45	29/45	-6
№5	30/45	35/45	-5

ИИ-системы продемонстрировали высокую результативность в идентификации грамматических, орфографических и лексических ошибок. Алгоритмы успешно выявляли нарушения согласования подлежащего и сказуемого («Baikal have» → «has»), некорректное употребление артиклей («in world» → «in the world»), а также лексические повторы («pollution» – 5-7 раз в одном тексте). Средняя точность алгоритмов в грамматическом анализе составила 6,8/10, что на 9,7% превысило экспертные оценки (6,2/10), благодаря ориентации на формальные шаблоны и корпусные данные.

Пример: В эссе №1 ИИ идентифицировал 4 грамматические ошибки, включая неправильное использование существительного вместо прилагательного («environment pollution» → «environmental pollution») и ошибки в артиклах («instead of the car» → «a car»).

Наиболее значимые расхождения между автоматизированной и экспертной оценкой выявлены в критериях структуры (33,3%), логики (24,3%) и стиля (38,1%). Алгоритмы фиксируют количественные показатели (наличие подзаголовков, частотность переходных слов), но игнорируют смысловую связность, глубину аргументации и синтаксическое разнообразие. ИИ не способен оценить прагматическую значимость ошибок, например, стилистическую неуместность разговорной лексики в академическом контексте.

В категориях грамматики и лексики различия оказались минимальными. Средняя оценка ИИ по грамматике на 9,7% превысила экспертную (6,8/10 и 6,2/10), что обусловлено ориентацией алгоритмов на формальные ошибки.

Наибольшие расхождения между автоматизированной и экспертной оценкой были выявлены в критериях структуры (33,3%), логики (24,3%) и стиля (38,1%). Алгоритмы фиксировали количественные показатели (наличие подзаголовков, частотность переходных слов), но игнорировали смысловую связность, глубину аргументации и синтаксическое разнообразие.

Пример: В эссе №4 ИИ снизил балл за структуру (6/10) из-за отсутствия четкого введения и заключения. Эксперт, напротив, повысил оценку (7/10) после реструктуризации текста, отметив добавление тезиса («Lake Baikal faces a serious environmental crisis») и систематизацию аргументов. Аналогично, в эссе №5 ИИ не распознал улучшение логики после добавления статистики

(«500 tons of tourist waste annually»), тогда как эксперт повысил балл за усиление убедительности аргументации.

Автоматизированные системы проявили избыточную строгость в оценке стиля, трактуя эмоциональные выражения («I feel sad», «It's very sad») как грубое нарушение академических норм, оценивая их в 0-2 балла. В то же время эксперты допускали подобные элементы в учебном контексте, рассматривая их как попытку авторского самовыражения, и оценивали стиль выше (2-4 балла). ИИ также не способен оценить синтаксическое разнообразие, такое как использование номинализаций («deterioration») или пассивного залога.

Например: В эссе №3 фраза «I feels sadly» была исправлена экспертом на «I feel sad» с сохранением эмоциональной окраски как части авторского замысла, тогда как ИИ рекомендовал полностью устраниТЬ подобные конструкции.

Сводный анализ пяти эссе показал, что экспертная оценка в среднем на 15-20% превышает автоматизированную (Таблица 1).

Результаты подтвердили гипотезу о комплементарности методов: ИИ эффективен для первичной коррекции формальных ошибок, сокращая временные затраты на 40%, тогда как экспертная оценка обеспечивает содержательный анализ структуры, логики и терминологической точности.

Практическое исследование позволило сформулировать ключевой вывод: инструменты искусственного интеллекта являются эффективным вспомогательным ресурсом для оптимизации проверки работ, демонстрируя высокую точность в выявлении формальных ошибок. Однако их применение для оценки структурно-логической связности и стилистического разнообразия ограничено. Оптимальной стратегией является внедрение гибридной этапной модели проверки «ИИ → эксперт → ИИ», которая минимизирует риски гиперкоррекции, обеспечивает соответствие работ академическим стандартам и позволяет педагогам сосредоточиться на развитии у учащихся навыков критического мышления и академического письма.

Заключение

Таким образом, проведённое исследование вносит вклад в трансформацию образовательных практик в условиях цифровизации. Результаты демонстрируют, что симбиоз технологий ИИ и педагогической экспертизы не только оптимизирует процесс оценки, но и способствует формированию у старшеклассников функциональной грамотности. Важно подчеркнуть, что внедрение ИИ должно быть направлено на усиление роли педагога как наставника, способного развивать творческий потенциал и критическое мышление учащихся. Это соответствует стратегическим целям современного образования, где технологии выступают инструментом, а человек остаётся центральным звеном в достижении образовательных результатов.

Библиографический список

- Алексеева Л. Г. Язык промптов, или особенности формулирования запросов к генеративным нейросетям для создания изображений / Л. Г. Алексеева, П. С. Алексеев // Verba. Северо-Западный лингвистический журнал. 2024. № 3 (13). С. 50-61. DOI 10.34680/VERBA-2024-3(13)-50-61. EDN ZVOBVA.
- Галагузова М. А. Искусственный интеллект в педагогике: от понятия к функции / М. А. Галагузова, Ю. Н. Галагузова, Г. Н. Штинова // Педагогическое образование в России. 2024. №. 2. С. 48-55. EDN JBMYDO.
- Карпов В. Э. К вопросу об этике и системах искусственного интеллекта / В. Э. Карпов, П. М. Готовцев, Г. В. Ройзензон // Философия и общество. 2018. №. 2 (87). С. 84-105. DOI 10.30884/jfio/2018.02.07. EDN YAEVYT.
- Лапина А. Проверка домашних заданий школьников станет автоматической // [Электронный ресурс]. – 2021. URL: <https://skillbox.ru/media/education/proverka-domashnikh-zadaniy-shkolnikov-stanet-avtomaticheskoy/> (дата обращения: 16.10.2025).
- Лапина А. Создана программа, способная проверять эссе на английском языке эффективнее учителя // [Электронный ресурс] – 2022. URL: <https://skillbox.ru/media/education/sozdana-programma-sposobnaya-proveryat-esse-na-angliyskom-yazyke-effektivnee-uchitelya/> (дата обращения: 16.10.2025).
- Петров С. В. Гиперкорректность как причина ошибок в речи учащихся // Международный научно-исследовательский журнал. 2022. № 2-3 (116). С. 160-162. DOI 10.23670/IRJ.2022.116.2.100. EDN QOMYRF.
- Пупков Т. С. Применение матрицы ошибок (confusion matrix) для оценки качества обучения моделей классификации // Актуальные вопросы эксплуатации систем охраны и защищенных телекоммуникационных систем : Сборник материалов Всероссийской научно-практической конференции, Воронеж, 10 июня 2021 года. Воронеж: Воронежский институт Министерства внутренних дел Российской Федерации, 2021. С. 35-36. EDN EKDEYO.
- Свинцов Р. Ложноположительные и ложноотрицательные результаты анализов // Ветеринария сельскохозяйственных животных. 2016. № 5. С. 62-65. EDN YXFPPN.
- Ступницкая М. Критериальное оценивание // Педагогические измерения. 2015. №. 1. С. 52-74. EDN TUFWDF.
- AI Technologies in Training PhD Students / S. Bogdanova, M. Stepanova, M. Matytcina [et al.] // 4th International Conference on Technology Enhanced Learning in Higher Education (TELE), Lipetsk, Russian Federation, 2024. Pp. 153–156. DOI 10.1109/TELE62556.2024.10605642. EDN VPRQNC.
- Automated Writing Evaluation Tools: Guide 2025 // [Электронный ресурс]. – 2024. URL: <https://www.yomu.ai/blog/automated-writing-evaluation-tools-guide-2025> (дата обращения: 17.10.2025).
- Chowdhary K. R. Natural language processing // Fundamentals of artificial intelligence. New Delhi: Springer, 2020. Pp. 603–649.
- Ekizer F. N. Exploring the impact of artificial intelligence on English language teaching: A meta-analysis // Acta Psychologica, 2025. № 260. Pp. 105649. DOI 10.1016/j.actpsy.2025.105649.
- North B. The CEFR in practice. Cambridge: Cambridge University Press, 2014. Vol. 4. 270 p.
- Rahimi A. R. Developing and validating the scale of language teachers' design thinking competency in artificial intelligence language teaching (LTDTAILT) // Computers and Education: Artificial Intelligence. 2025. Vol. 8. Pp. 100420. DOI <https://doi.org/10.1016/j.caeari.2025.100420>.
- Tan X. Artificial intelligence in teaching and teacher professional development: A systematic review / X. Tan, G. Cheng, M. H. Ling // Computers and Education: Artificial Intelligence. 2025. № 8. Pp. 100355. DOI <https://doi.org/10.1016/j.caeari.2024.100355>.

The future of teaching and learning in the context of emerging artificial intelligence technologies / E. Ukwandu, O. Omisade, K. Jones [et al.] // *Futures*. 2025. № 171. Pp. 103616. DOI <https://doi.org/10.1016/j.futures.2025.103616>.

References

- Alekseeva L. G., Alekseev P. S. (2024). The language of prompts, or the peculiarities of the formation of queries to generative neuro systems for creating images *Verba. North-West linguistic journal*. 3 (13): 50-61. (In Russian)
- Automated Writing Evaluation Tools: Guide 2025 (2024). Available at: <https://www.yomu.ai/blog/automated-writing-evaluation-tools-guide-2025> (accessed 17 October 2025).
- Bogdanova S., Stepanova M., Matytcina M., Plastinina N., Plekhanova Y., Kudritskaya M. (2024). AI Technologies in Training PhD Students. *4th International Conference on Technology Enhanced Learning in Higher Education (TELE)*. 153-156. DOI 10.1109/TELE62556.2024.10605642.
- Chowdhary K. R. (2020). Natural language processing. *Fundamentals of artificial intelligence*. New Delhi: Springer. 603-649.
- Ekizer F. N. (2025). Exploring the impact of artificial intelligence on English language teaching: A meta-analysis. *Acta Psychologica*. 260: 105649. DOI 10.1016/j.actpsy.2025.105649.
- Galaguzova M. A., Galaguzova Yu. N., Shtinova G. N. (2024). Artificial intelligence in pedagogy: from concept to function. *Pedagogical survey in Russia*. 2: 48-55. (In Russian)
- Karpov V. E., Gotovcev P. M., Rojzenzon G. V. (2018). On the issue of ethics and the systems of artificial intelligence. *Philosophy and society*. 2 (87): 84-105. (In Russian)
- Lapina A. (2021). Checking pupils' homework will become automatic. Available at: <https://skillbox.ru/media/education/proverka-domashnikh-zadaniy-shkolnikov-stanet-avtomaticheskoy/>. (accessed 16 October 2025) (In Russian)
- Lapina A. (2022). A program has been created that can check English essays more effectively than a teacher. Available at: <https://skillbox.ru/media/education/sozdana-programma-sposobnaya-proveryat-esse-na-angliyskom-yazyke-effektivnee-uchitelya/> (accessed 16 October 2025) (In Russian)
- North B. (2014). The CEFR in practice. Vol. 4. Cambridge: *Cambridge University Press*, 2014. 270 p.
- Petrov S. V. (2022). Hypercorrectness as a reason for mistakes in the learners' speech. *International research journal*. 2-3 (116): 160-162. (In Russian)
- Pupkov T. S. (2021). Application of the confusion matrix for the assessment of the quality of training of the classification models. *Topical issues of operation of security systems and protected telecommunication systems*: Collection of materials from the All-Russian scientific and practical conference, Voronezh, 10/06/2021. Voronezh: Voronezhskij institut Ministerstva vnutrennih del Rossiijskoj Federacii. 35-36. (In Russian)
- Rahimi A. R. (2025). Developing and validating the scale of language teachers' design thinking competency in artificial intelligence language teaching (LTDTAILT). *Computers and Education: Artificial Intelligence*. 8: 100420. DOI <https://doi.org/10.1016/j.caear.2025.100420>.
- Stupnickaya M. (2015). Criteria-based assessment. *Pedagogical measurements*. 1: 52-74. (In Russian)
- Svincov R. (2016). False positive and false negative test results. *Veterinary science of farm animals*. 5: 62-65. (In Russian)
- Tan X., Cheng G., Ling M. H. (2025). Artificial intelligence in teaching and teacher professional development: A systematic review. *Computers and Education: Artificial Intelligence*. 8. 100355. DOI <https://doi.org/10.1016/j.caear.2024.100355>.
- Ukwandu E., Omisade O., Jones K., Thorne S., Castle M. (2025). The future of teaching and learning in the context of emerging artificial intelligence technologies. *Futures*. 171: 103616. DOI <https://doi.org/10.1016/j.futures.2025.103616>.