УДК 004.8:004.946:331.363
DOI: 10.17277/vestnik.2026.03.pp.373-389

ПРИМЕНЕНИЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ ДЛЯ ОРГАНИЗАЦИИ ПРОФЕССИОНАЛЬНОЙ ПОДГОТОВКИ НА БАЗЕ ИММЕРСИВНЫХ СИСТЕМ

© А. Ю. Свешников[1] ✉, Ю. В. Никитников[1], М. А. Шильцын[1], Д. Л. Дедов[2]

[1] *Кафедра «Системы автоматизированной поддержки принятия решений», hammer68@mail.ru; ФГБОУ ВО «ТГТУ», Тамбов, Российская Федерация;*
[2] *Управление фундаментальных и прикладных исследований, ФГБОУ ВО «ТГТУ», Тамбов, Российская Федерация*

Ключевые слова: большие языковые модели; виртуальные тренажеры; иммерсивные системы; профессиональная подготовка; ранжирование альтернатив; рекомендательные системы; система поддержки принятия решений; цифровой профиль сотрудника.

Аннотация: Рассмотрена возможность применения локальных больших языковых моделей (LLM) для ранжирования виртуальных тренажеров и иммерсивных средств профессиональной подготовки по цифровому профилю сотрудника. Поставлена задача семантического ранжирования с использованием LLM на основе анализа слабоструктурированных сведений о пользователе и их сопоставления с каталогом доступных тренажерных решений. Экспериментальная оценка выполнена на наборе из 20 цифровых профилей сотрудников различных отраслей и каталоге виртуальных тренажеров, охватывающих различные сферы (охрану труда, электробезопасность, пожарную безопасность, эвакуацию, первую помощь и т.д.). В сравнении участвовали десять локальных LLM. Качество рекомендаций оценивалось по различным метрикам точности, дополнительно анализировались структурированная валидность ответа, корректность возвращаемых кодов тренажеров и среднее время выполнения. Сделан вывод о применимости локальных LLM как интеллектуального компонента системы поддержки выбора иммерсивных средств профессиональной подготовки.

Введение

Профессиональная подготовка персонала в технических, промышленных, медицинских, энергетических организациях для обеспечения высокой эффективности активно применяет виртуальные и иммерсивные технологии [1, 2].

Виртуальные тренажеры позволяют воспроизводить опасные, редкие или дорогостоящие производственные ситуации без прямого риска для работника и оборудования. За счет интерактивной среды обучаемый может отрабатывать действия в условиях пожара, эвакуации, работы на высоте, взаимодействия с оборудованием, управления транспортно-складской техникой, проведения первой помощи или выполнения регламентных операций в аварийной ситуации [3, 4]. Развитие этой отрасли приводит к внедрению в обучающие средства программно-аппаратных комплексов, включая имитационные компоненты, системы захвата движений, обратной связи, экзоскелетные модули, что позволяет повысить реалистичность и качество профессиональной подготовки [5, 6].

Однако расширение набора обучающих средств, широкое варьирование их сложности и стоимости приводит к необходимости решения методологической задачи по выбору и компоновке тренажеров в индивидуальную образовательную траекторию, минимально достаточную для выработки требуемых компетенций у обучаемого. Для автоматизации поддержка принятия решений в этой области необходимо учитывать не только отрасль и должность обучаемого, но и уровень его базовой подготовки, стаж, должностные обязанности и индивидуальные особенности [7, 8].

Традиционные подходы к назначению учебных модулей обычно строятся на экспертных правилах, матрицах компетенций, должностных инструкциях и типовых программах обучения [9, 10]. Эти механизмы хорошо работают для формализованных требований, например при назначении обязательного вводного инструктажа или периодической проверки знаний. Но когда профиль сотрудника является набором слабоструктурированной информации (описание истории тренировок, характеристика, оценки), то автоматизированная обработка таких данных затруднена. В рамках данной предметной области мы рассматриваем задачу сопоставления такого цифрового профиля с набором доступных средств подготовки и получение на основе анализа профиля упорядоченного (ранжированного) списка альтернатив. Тогда первая позиция такого списка соответствует основному рекомендуемому тренажеру, а последующие позиции формируют расширенную траекторию подготовки.

Большие языковые модели (LLM) являются перспективным инструментом для решения данной задачи, поскольку способны интерпретировать естественно-языковые описания, извлекать смысловые признаки из слабоструктурированных профилей и сопоставлять их с описаниями учебно-тренировочных средств. В отличие от классической модели машинного обучения, требующей заранее подготовленного большого размеченного набора признаков, LLM может работать с комбинированным представлением: результатами входного тестирования, описанием профессиональной роли, перечнем компетентностных дефицитов, историей ошибок и текстовыми характеристиками тренажеров. Это позволяет использовать LLM в качестве основы системы поддержки принятия решений при организации профессиональной подготовки.

Вместе с тем применение LLM включает и ряд возможных ограничений: возвращенный моделью ответ не соответствовать заданной структуре, включать повторы и артефакты; прогноз модели может быть некорректным и нерелевантным для пользователя. Таким образом, без строгой, комплексной оценки модели ее применение для решения поставленной задачи невозможно.

Цель работы – исследование возможности применения локальных больших языковых моделей для выбора виртуальных тренажеров и иммерсивных средств профессиональной подготовки на основе цифрового профиля сотрудника. Для достижения поставленной цели формализуется цифровой профиль сотрудника и каталог виртуальных тренажеров, разрабатывается единый протокол формирования запроса к LLM и выполняется сравнительная оценка локальных LLM по метрикам ранжирования, структурированной валидности и времени ответа. Научная новизна исследования состоит в том, что выбор иммерсивного средства профессиональной подготовки рассматривается как задача семантического ранжирования альтернатив по цифровому профилю сотрудника.

Анализ сопутствующих исследований

Существующие обзорные работы связывают эффективность иммерсивного обучения с проектированием сценария, педагогической моделью и измеряемым образовательным результатом [1, 5]. Выбор тренажера должен учитывать не только

соответствие профессии, но и ограничения пользователя, уровень реалистичности, характер обратной связи и формируемых навыков [11, 12].

Эмпирические исследования показывают, что более высокая степень присутствия в immersive VR не всегда приводит к лучшему усвоению материала [13], что дополнительно обосновывает необходимость индивидуального выбора средства подготовки. VR/AR широко применяются в промышленной безопасности и охране труда [14], однако большинство решений остаются предметно специализированными, а задача выбора тренажера с учетом уровня подготовки, ошибок и текущей учебной задачи сотрудника проработана недостаточно [15].

Иммерсивные сценарии пожарной безопасности и эвакуации позволяют моделировать редкие и опасные события [16, 17], однако типовое назначение одного и того же тренажера не учитывает различия в опыте и ранее допущенных ошибках сотрудников.

Образовательные рекомендательные системы применяются для персонализации учебных ресурсов и траекторий [7]. Однако виртуальный тренажер как объект рекомендации сложнее традиционного учебного материала, поскольку дополнительно характеризуется аппаратной конфигурацией, уровнем иммерсивности, безопасностью, длительностью, сложностью, обратной связью и ограничениями пользователя.

Исследования искусственного интеллекта в образовании подчеркивают значение профилирования и адаптивной поддержки [18]. Для рассматриваемой задачи это означает необходимость связать цифровой профиль сотрудника с релевантным средством обучения. LLM расширяют возможности обработки слабоструктурированных образовательных данных и персонализации обучения [19]. Вместе с тем требуется контролировать корректность ответа, поскольку модель может нарушать структуру выдачи, возвращать недопустимые коды или игнорировать ограничения безопасности.

Проведенный анализ позволяет сделать следующие выводы. С одной стороны, существует развитое направление исследований по VR/AR-тренажерам и иммерсивному обучению, где изучаются проектирование сценариев, эффект присутствия, когнитивная нагрузка, мотивация, безопасность и учебные результаты; с другой – существует направление образовательных рекомендательных систем, где анализируются методы подбора ресурсов и курсов на основе профиля пользователя. Наконец, активно развивается применение LLM в образовательных и профессиональных сценариях. Однако на пересечении этих направлений остается недостаточно проработанной задача ранжирования иммерсивных тренажеров, программных и аппаратных средств профессиональной подготовки по цифровому профилю сотрудника.

Материалы и методы

Общая схема исследования

Методологическая схема исследования включает пять последовательных этапов. На первом этапе формируется набор цифровых профилей сотрудников, отражающий разные отраслевые и профессиональные ситуации. На втором этапе формируется каталог доступных средств подготовки. В него включаются виртуальные тренажеры и имитационные сценарии, отличающиеся тематикой, режимом реализации, длительностью, сложностью, целевыми компетенциями и ограничениями применения. На третьем этапе для каждого цифрового профиля формируется унифицированный запрос к большой языковой модели; на четвертом – выполняется вызов локальных LLM; на пятом – полученные рекомендации сопоставляются с экспертно заданным эталонным ранжированием.

Формальная постановка задачи ранжирования тренажеров

Пусть задано множество цифровых профилей сотрудников

$$U = \{u_1, u_2, ..., u_N\}, \quad (1)$$

где N – число анализируемых профилей. Каждый профиль сотрудника задается кортежем

$$u_i = (r_i, s_i, e_i, c_i, d_i, h_i, l_i), \quad (2)$$

где r_i – профессиональная роль; s_i – отрасль или производственный контекст; e_i – стаж и предшествующий опыт; c_i – результаты входной оценки компетенций; d_i – выявленные дефициты и потребности подготовки; h_i – история ранее пройденных тренажеров; l_i – ограничения использования иммерсивной среды и условия безопасности.

Множество доступных иммерсивных средств обозначим как

$$T = \{t_1, t_2, ..., t_M\}, \quad (3)$$

где M – число тренажеров в каталоге. Каждый тренажер описывается кортежем

$$t_j = (q_j, n_j, a_j, m_j, k_j, z_j, b_j), \quad (4)$$

где q_j – уникальный код; n_j – название; a_j – отрасль или область применения; m_j – режим реализации; k_j – формируемые компетенции; z_j – параметры сложности и длительности; b_j – ограничения и условия безопасного применения.

Для каждого профиля u_i требуется получить упорядоченный список

$$\pi_i = (q_{i,1}, q_{i,2}, ..., q_{i,K}), \ \ q_{i,K} \in T, \quad (5)$$

где $K = 5$, элементы списка не повторяются и расположены по убыванию ожидаемой релевантности. В отличие от задачи классификации, здесь требуется не только определить один целевой тренажер, но и сформировать верхний фрагмент индивидуальной траектории подготовки.

Экспертное эталонное ранжирование обозначим как

$$\pi_i^* = \left(q_{(i,1)}^*, q_{(i,2)}^*, ..., q_{(i,K)}^*\right), \quad (6)$$

где $q_{(i,K)}^*$ соответствует основному тренажеру, а последующие позиции задают допустимые дополнительные альтернативы. Эталон используется только для расчета метрик и не передается в контур формирования рекомендаций.

В общем виде соответствие между цифровым профилем сотрудника и тренажером может быть задано функцией релевантности

$$p : U \times T \to \mathbb{R}, \quad (7)$$

а корректное ранжирование должно удовлетворять условию

$$\rho(u_i, q_{i,1}) \geq \rho(u_i, q_{i,2}) \geq ... \geq \rho(u_i, q_{i,K}). \quad (8)$$

Метрики качества и эксплуатационной пригодности

Поскольку выходом алгоритма является ранжированный список, основными являются метрики верхней части выдачи. Accuracy@1 показывает долю профилей, для которых первый рекомендованный тренажер совпал с основным экспертным вариантом:

$$\text{Accuracy}@1 = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left[q_{i,1} = q_{i,1}^{*}\right]. \quad (9)$$

Accuracy@3 отражает долю профилей, для которых основной экспертный тренажер попал в первые три рекомендации:

$$\text{Accuracy}@3 = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}\left[q_{i,1}^{*} \in \{q_{i,1}, q_{i,2}, q_{i,3}\}\right]. \quad (10)$$

Для оценки качества всего списка применяется NDCG@5. Если $\text{rel}_{i,k}$ – релевантность тренажера на позиции k для профиля u_i, то дисконтированный накопленный выигрыш определяется как

$$\text{DCG}@5_i = \sum_{k=1}^{5} \frac{\text{rel}_{i,k}}{\log_2(k+1)}. \quad (11)$$

Нормированное значение равно

$$\text{NDCG}@5_i = \frac{\text{DCG}@5_i}{\text{IDCG}@5_i}, \quad (12)$$

а среднее значение по выборке рассчитывается как

$$\text{NDCG}@5 = \frac{1}{N}\sum_{i=1}^{N} \text{NDCG}@5_i, \quad (13)$$

Дополнительно используется средний обратный ранг MRR@5

$$\text{MRR}@5 = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{\text{rank}_i}, \quad (14)$$

где rank_i – позиция первого появления основного экспертного тренажера в топ-5; если он отсутствует, вклад профиля принимается равным нулю.

Для прикладной системы важна не только содержательная точность, но и техническая пригодность ответа. Поэтому вводится структурированная валидность. Для профиля u_i она определяется как произведение трех условий: корректный JSON ($V_{\text{json},i}$), допустимые коды тренажеров ($V_{\text{codes},i}$) и правильная длина списка без повторов ($V_{\text{lenght},i}$)

$$V_i = V_{\text{json},i} \times V_{\text{codes},i} \times V_{\text{lenght},i}. \quad (15)$$

Агрегированная валидность равна

$$V = \frac{1}{N}\sum_{i=1}^{N} V_i. \quad (16)$$

Среднее время ответа модели оценивается как

$$\bar{\tau} = \frac{1}{N}\sum_{i=1}^{N} \tau_i, \quad (17)$$

где τ_i – время формирования рекомендации для профиля u_i.

Анализ базовой и гибридной реализации функций ранжирования

В базовой реализации LLM получает полный каталог тренажеров *T*, профиль сотрудника u_i и инструкцию вернуть пять рекомендаций в формате JSON. Запрос можно представить как

$$p_i = P(I, T, u_i, O), \tag{18}$$

где *I* – инструкция роли модели; *T* – каталог; u_i – профиль сотрудника; *O* – спецификация выходной структуры. Модель должна выбрать только существующие коды, упорядочить их по релевантности и для каждого элемента указать обоснование, адаптацию и замечания по безопасности.

Этот вариант реализации используется как исходная точка сравнения, поскольку показывает способность локальных LLM самостоятельно решать задачу семантического выбора.

Далее рассмотрим гибридный подход, объединяющий несколько компонентов, направленных на улучшение стабильности и качества работы LLM.

Первым компонентом гибридного подхода является детерминированный экспертно-признаковый скоринг. Его задача – оценить содержательную близость пары «сотрудник – тренажер» без обращения к экспертному эталону π_i^*. В обобщенном виде оценка задается как

$$S(u_i, t_j) = w_s S_s + w_n S_n + w_r S_r + w_x S_x - w_b B - w_c C - w_h H, \tag{19}$$

где S_s описывает совпадение отрасли; S_n – соответствие заявленным потребностям и дефицитам; S_r – связь должностной роли с назначением тренажера; S_x – текстово-семантическое пересечение профиля и описания тренажера; *B* – штраф за ограничения и противопоказания; *C* – штраф за несоответствие сложности уровню подготовки; *H* – штраф за повторное назначение уже успешно пройденного тренажера. Коэффициенты *w* задают относительную значимость компонентов и реализуются как экспертные эвристики.

На основе $S(u_i, t_j)$ для каждого профиля формируется лист возможных кандидатов

$$C_i = \mathrm{Top}_n \{S(u_i, t_i); t_i \in T\}, \quad n = 8. \tag{20}$$

Модель LLM в гибридном режиме получает не полный каталог, а только множество C_i. Это уменьшает число нерелевантных альтернатив, сокращает контекст, снижает вероятность галлюцинаций и предотвращает смещение к универсальным тренажерам, если профиль содержит явный специализированный риск.

Вторым компонентом является промт, в котором явно задается логика выбора: профильный риск и должностная задача имеют приоритет над общим отраслевым сходством; конкретный дефицит компетенций важнее универсального модуля; общий тренажер не должен вытеснять специализированный, если в профиле указан конкретный риск; ограничения VR и безопасности должны снижать приоритет неподходящих сценариев; успешно пройденный тренажер не должен повторно назначаться как основной.

После ответа модели применяется корректирующий контур (Repair). Пусть LLM вернула список

$$\hat{\pi}_i = \left(\hat{q}_{i,1}, \hat{q}_{i,2}, ..., \hat{q}_{i,L}\right). \tag{21}$$

Операция коррекции формирует корректный список

$$\tilde{\pi}_i = R\left(\hat{\pi}_i, T, S\right), \tag{22}$$

удаляя недопустимые коды, исключая повторы и дополняя недостающие позиции лучшими кандидатами по экспертно-признаковому скорингу.

Наиболее важным компонентом является гибридное ранжирование (Hybrid). Для каждого кода q рассчитывается нормированная экспертно-признаковая оценка $\widetilde{S}_i(q)$ и позиционная оценка LLM:

$$L_i(q) = \frac{K - r_i(q) + 1}{K}, \tag{23}$$

где $r_i(q)$ – позиция кода q в обработанного выдаче LLM. Итоговая гибридная оценка задается как

$$H_i(q) = \alpha \widetilde{S}_i(q) + (1 - \alpha)\, L_i(q), \quad \alpha = 70. \tag{24}$$

Набор кандидатов формируется как объединение обработанного выдачи LLM и верхней части экспертно-признакового ранжирования

$$P_i = \widetilde{\pi}_i \bigcup C_i. \tag{27}$$

Затем элементы P_i сортируются по $H_i(q)$, и выбираются первые пять кодов. Такая схема сохраняет способность LLM предложить контекстно релевантный вариант, но защищает итоговую рекомендацию от случайного выбора универсальных или невалидных тренажеров.

Схема проведения экспериментов

Исследование включает две экспериментальные серии. В первой серии все модели получают полный каталог тренажеров и формируют топ-5 список без дополнительной постобработки. Эта серия позволяет оценить прямой подход и выявить его ограничения.

Вторая серия исследует гибридную реализацию. Для всех моделей по тем же профилям и тому же каталогу сравниваются режимы RawLLM, Repair и Hybrid. RawLLM соответствует исходному ответу модели без исправления. Repair устраняет ошибки структуры, но сохраняет порядок LLM настолько, насколько это возможно. Hybrid объединяет очищенную выдачу LLM с экспертно-признаковым скорингом и заново упорядочивает пул кандидатов. Дополнительно рассчитывается Rule Baseline – детерминированный экспертно-признаковый подход без участия LLM. Он необходим для понимания того, какой вклад вносит формализованная предметная модель и насколько полезно добавление LLM к этой модели.

Результаты

Описание датасета

Экспериментальный набор включает 20 цифровых профилей сотрудников (табл. 1) и каталог из 20 виртуальных тренажеров (табл. 2). Каждый профиль содержит идентификатор, должностную роль, отрасль, стаж, результат входного тестирования, описание потребностей обучения, ограничения, предпочтения и экспертный топ-5 список рекомендуемых тренажеров. Каталог содержит код тренажера, название, категорию, отраслевую область, режим реализации, сложность, длительность, формируемые компетенции, целевую аудиторию, ограничения и показатели оценки результата.

Таблица 1

Фрагмент цифровых профилей сотрудников

ID	Профиль и исходные данные	Потребности и ограничения	Экспертный топ-5
U001	Новый монтажник строительной организации. Отрасль: строительство; тест: 42; стаж: 0,5 г.	Работа на высоте; СИЗ; эвакуация. Ограничения: низкий опыт; тревожность при высоте	VR-HEIGHT-002 > > VR-OT-001 > > VR-EVAC-005 > > VR-FIRSTAID-006 > > VR-FIRE-004
U002	Электромонтер ремонтной службы. Отрасль: энергетика; тест: 61; стаж: 3,0 г.	LOTO; электробезопасность; аварийные переключения. Ограничения: ошибки в чек-листе отключения	VR-ELECTRO-003 > > VR-ENERGY-014 > > VR-OT-001 > > VR-FIRSTAID-006 > > VR-FIRE-004
U003	Оператор погрузчика на распределительном складе. Отрасль: склад; тест: 55; стаж: 1,0 г.	Управление погрузчиком; движение среди пешеходов; эвакуация. Ограничения: киберболезнь средняя	VR-FORKLIFT-010 > > VR-OT-001 > > VR-EVAC-005 > > VR-FIRSTAID-006 > > VR-FIRE-004

Таблица 2

Фрагмент каталога виртуальных тренажеров

Код	Название	Компетенции и ограничения	Режим и параметры
VR-OT-001	Базовый инструктаж по охране труда и подбору СИЗ	Категория: охрана труда. Компетенции: выбор СИЗ, идентификация опасностей, безопасное поведение. Ограничения: нет	VR + интерактивные чек-листы 25 мин; сложность 0,25
VR-HEIGHT-002	Работа на высоте и применение страховочных систем	Категория: работа на высоте. Компетенции: осмотр привязи, выбор точки анкерования, порядок подъема. Ограничения: выраженная киберболезнь, паническая боязнь высоты без адаптационного режима	VR с контролем последовательности действий 35 мин; сложность 0,70
VR-FIRE-004	Пожарная безопасность и применение первичных средств тушения	Категория: пожарная безопасность. Компетенции: выбор огнетушителя, оценка класса пожара, безопасная дистанция. Ограничения: тяжелая киберболезнь при дымовых эффектах	VR с моделированием очага возгорания 30 мин; сложность 0,45

Сравниваемые модели

В эксперименте участвовали десять локальных LLM в формате GGUF (табл. 3). Все модели использовались в режиме inference-only, без дообучения на профилях сотрудников и без подстройки под конкретный каталог тренажеров. Такой режим позволяет оценить исходную способность моделей выполнять семантическое ранжирование по единому протоколу запроса.

Таблица 3

Локальные LLM, включенные в эксперимент

Обозначение	Базовая конфигурация	Квантование
DeepSeek-7B-Q4	DeepSeek-R1-Distill-Qwen 7B	Q4_K_M
Qwen3.6-35B-Q4	Qwen3.6-35B-A3B	Q4_K_S
LFM2-24B-Q4	LFM2-24B-A2B	Q4_K_M
Qwen3.5-9B-Q8	Qwen3.5-9B	Q8_0
Qwen3.5-9B-Q4	Qwen3.5-9B	Q4_K_M
Gemma-4B-Q8	gemma-4-E4B-it	Q8_0
GPT-OSS-20B	gpt-oss-20b	MXFP4
Ministral-14B-Q4	Ministral-3-14B-Reasoning	Q4_K_M
Nemotron-4B-Q8	NVIDIA-Nemotron-3-Nano	Q8_0
Qwen3.6-27B-IQ3	Qwen3.6-27B	IQ3_XXS

Результаты прямого LLM-ранжирования

В первой серии каждая модель получила полный каталог тренажеров и цифровые профили сотрудников. Результатом являлся топ-5 список тренажеров, возвращенный моделью. Сводные результаты приведены в табл. 4.

Полученные результаты показывают, что прямой LLM-подход решает задачу недостаточно устойчиво. Лучшая модель по NDCG@5 действительно выделяется на фоне остальных, однако абсолютный уровень качества остается ограниченным: даже у лидера первая рекомендация совпадает с экспертным основным тренажером менее чем в половине профилей. При этом Accuracy@3 заметно выше, но полученные результаты по-прежнему остаются неудовлетворительными.

Таблица 4

Результаты прямого LLM-ранжирования по полному каталогу

Модель	Acc@1	Acc@3	NDCG@5	MRR@5	Валидн.	Время, с
DeepSeek-7B-Q4	0,4500	0,7000	0,4627	0,5792	0,5500	13,93
Qwen3.6-35B-Q4	0,4000	0,5000	0,4165	0,4667	0,5000	15,47
LFM2-24B-Q4	0,3500	0,5500	0,3924	0,4625	0,6500	14,76
Qwen3.5-9B-Q8	0,4500	0,6000	0,3900	0,5167	0,4500	14,34
Qwen3.5-9B-Q4	0,4000	0,5500	0,3797	0,4583	0,5000	14,47
Gemma-4B-Q8	0,3000	0,5500	0,3725	0,4183	0,7500	13,42
GPT-OSS-20B	0,3500	0,5500	0,3688	0,4625	0,4500	13,74
Ministral-14B-Q4	0,3500	0,5000	0,3685	0,4500	0,5000	15,45
Nemotron-4B-Q8	0,3000	0,4500	0,3419	0,3892	0,4500	16,23
Qwen3.6-27B-IQ3	0,2500	0,4500	0,3272	0,3625	0,4500	15,20

Основная причина низкого результата связана не только с качеством отдельных LLM, но и с тем, что в полном каталоге присутствуют универсальные тренажеры – охрана труда, пожарная безопасность, эвакуация и т.д. – которые подходят многим профилям. При отсутствии жесткой рубрики модель нередко выбирает такой универсальный сценарий вместо специализированного. Кроме того, тренажеры имеют пересекающиеся компетенции: эвакуация связана с пожаром, первая помощь – с травмами и авариями, электробезопасность – с энергетикой и производственным обслуживанием.

Отдельным ограничением стала структурированная валидность. Даже когда ответ модели является синтаксически корректным JSON, он может содержать повторяющиеся коды, несуществующие обозначения или список неправильной длины. Поэтому прямое LLM-ранжирование не может использоваться как автономный механизм назначения тренажера. Оно требует программного контроля и экспертной проверки.

Сравнение методологических режимов

Во второй серии оценивались режимы RawLLM, Repair, Hybrid и Rule Baseline. Сравнение этих режимов позволяет отделить вклад языковой модели от вклада методологической обвязки. Результаты приведены в табл. 5.

Переход от исходных ответов LLM к Repair устраняет технические ошибки и повышает структурированную валидность до единицы. При этом рост качества ранжирования оказывается умеренным: Repair в основном исправляет непригодность формата, но не решает проблему слабого порядка рекомендаций. Это подтверждает, что одних технических проверок недостаточно. Если модель поставила универсальный тренажер выше специализированного, простое удаление повторов не исправит содержательную ошибку.

Наиболее существенный эффект дает гибридное ранжирование (Hybrid). Среднее значение NDCG@5 почти удваивается относительно прямого raw-подхода, а Accuracy@1 и Accuracy@3 также резко возрастают.

Rule Baseline показал высокий результат. Формализованная предметная модель, учитывающая отрасль, дефициты, роль, ограничения и историю подготовки, сама по себе дает большой объем информации для обработки, поэтому LLM не должна заменять экспертную логику. Более корректная архитектура состоит в их объединении: экспертно-признаковый модуль обеспечивает безопасность и базовую релевантность, а LLM интерпретирует слабоструктурированный контекст и может уточнять или объяснять выбор.

Гибридные результаты по всем моделям

Результаты Hybrid-режима по всем моделям приведены в табл. 6. В этом режиме итоговый топ-5 формируется не как прямой ответ LLM, а как результат объединения обработанной выдачи модели и экспертно-признакового ранжирования.

Таблица 5

Сравнение методологических режимов обработки результатов (в среднем по моделям)

Режим	Acc@1	Acc@3	NDCG@5	Валидность	Лучший NDCG
Raw LLM	0,3600	0,5400	0,3820	0,5250	0,4627
Repair	0,4100	0,6200	0,5013	1,0000	0,5563
Hybrid	0,6800	0,8700	0,7387	1,0000	0,7562
Rule Baseline	0,7000	0,8500	0,7726	1,0000	0,7726

Таблица 6

Результаты гибридной реализации на различных моделях

Модель	Acc@1	Acc@3	NDCG@5	MRR@5	Время, с
Rule Baseline	0,7000	0,8500	0,7726	0,7792	0,00
DeepSeek-7B-Q4	0,7000	0,9000	0,7562	0,8125	13,93
Qwen3.6-35B-Q4	0,6000	0,9000	0,7520	0,7625	15,47
Qwen3.5-9B-Q8	0,7500	0,8500	0,7444	0,8125	14,34
Qwen3.5-9B-Q4	0,6500	0,8500	0,7407	0,7625	14,47
Ministral-14B-Q4	0,7000	0,9000	0,7372	0,7917	15,45
Nemotron-4B-Q8	0,6500	0,8500	0,7366	0,7500	16,23
Gemma-4B-Q8	0,7500	0,8500	0,7355	0,7917	13,42
LFM2-24B-Q4	0,6500	0,9000	0,7349	0,7792	14,76
Qwen3.6-27B-IQ3	0,6500	0,8500	0,7273	0,7625	15,20
GPT-OSS-20B	0,7000	0,8500	0,7219	0,7850	13,74

В отличие от первой серии, качество всех моделей значительно улучшилось. Наиболее важный практический эффект состоит в том, что структурированная валидность Hybrid-режима становится полной. Следовательно, гибридный подход решает не только задачу повышения точности, но и задачу надежности вывода.

Обсуждение

Сопоставление двух экспериментальных серий показывает, что слабый результат прямого LLM-подхода показывает, что задача выбора иммерсивного тренажера является слишком многокритериальной для полностью автономного ранжирования по полному каталогу. Он хорошо извлекает смысловые связи, но плохо гарантирует строгий формат, устойчивый порядок и соблюдение всех ограничений.

Гибридный подход показывает намного более перспективные результаты за счет указания приоритетов выбора и снижения неоднозначности интерпретации, перевода ответа в машинно-обрабатываемую форму и объединения позиции LLM с экспертно-признаковой оценкой.

Полученный эффект важен тем, что в ходе сравнения выявлен наибольший эффект не просто от замены одной LLM на другую, а путем изменения архитектуры решения. Это означает, что при проектировании СППР нельзя ограничиваться сравнением моделей как «черных ящиков», необходимо задавать структуру предметной области, формализовать допустимые альтернативы, контролировать формат ответа и отделять содержательную релевантность от технической валидности.

Одновременно результаты показывают, что Rule Baseline остается сильным ориентиром. Если детерминированная экспертная модель сопоставима с отдельными LLM, то они должны использоваться не как замена экспертного правила, а как контекстный компонент, способный работать со слабоструктурированными описаниями, формировать объяснения и дополнять строгое ранжирование там, где признаки профиля не сводятся к простым ключевым словам.

Заключение

В работе исследована задача выбора виртуальных тренажеров и иммерсивных средств профессиональной подготовки по цифровому профилю сотрудника. Выбор тренажера формализован как задача ранжирования ограниченного множества альтернатив, где объектами обработки являются слабоструктурированный профиль сотрудника, каталог тренажеров и экспертное эталонное ранжирование. В качестве интеллектуального компонента рассмотрены локальные большие языковые модели, развертываемые без передачи данных во внешние облачные сервисы.

Первая экспериментальная серия показала, что прямое применение LLM к полному каталогу тренажеров дает ограниченное качество. Модели способны находить тематически близкие сценарии, но часто ошибаются в первом выборе, смещаются к универсальным модулям и нарушают формальные требования к структуре ответа. Поэтому прямой LLM-подход не может рассматриваться как готовое решение для автономного назначения тренажеров.

Для устранения выявленных ограничений предложен гибридный подход. Экспериментальное сравнение показало, что такой подход существенно повышает качество рекомендаций и обеспечивает полную структурированную валидность итогового списка. Основной прирост связан не с изменением локальной модели, а с методологической организацией процесса выбора.

Научный результат работы состоит в формализации применения LLM как контролируемого ранжирующего компонента СППР, предложении воспроизводимого программного контура, который может быть использован при разработке систем организации профессиональной подготовки на базе иммерсивных технологий.

Ограничения исследования связаны с размером датасета и характером цифровых профилей. Полученные результаты следует рассматривать как методологическое обоснование и экспериментальную проверку подхода. Дальнейшее развитие работы состоит в расширении выборки, привлечении нескольких независимых экспертов к разметке релевантности, оценке качества объяснений LLM, учете результатов прохождения тренажеров и проверке гибридной схемы в составе действующей платформы иммерсивного обучения.

Таким образом, проведенное исследование показывает, что локальные LLM могут быть полезны для организации профессиональной подготовки на базе иммерсивных систем, но наибольший практический эффект достигается при их включении в гибридную архитектуру, где формализованные правила обеспечивают надежность, а языковая модель – интерпретацию слабоструктурированного контекста и поддержку экспертного выбора.

Работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ в рамках проекта «Разработка иммерсивной системы взаимодействия с виртуальной реальностью для профессиональной подготовки на основе всенаправленной платформы» (124102100628-3).

Список литературы

1. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda / J. Radianti, T. A. Majchrzak, J. Fromm, I. Wohlgenannt // Computers & Education. – 2020. – Vol. 147. – Art. 103778. doi: 10.1016/j.compedu.2019.103778

2. Системный анализ и формализация структуры адаптивных тренажерных комплексов эргатических систем / М. Н. Краснянский, Д. Л. Дедов, А. Д. Обухов, С. Ю. Алексеев // Вестник компьютерных и информационных технологий. – 2019. – № 4(178). – С. 45–52. doi: 10.14489/vkit.2019.04.pp.045-052

3. Virtual Reality for Safety Training: A Systematic Literature Review and Meta-Analysis / D. Scorgie, Z. Feng, D. Paes [et al.] // Safety Science. – 2023. – Vol. 171. – Art. 106372. doi:10.1016/j.ssci.2023.106372

4. Visualization Technology and Tool Selection Methods for Solving Adaptive Training Complex Structural-Parametric Synthesis Problems / M. N. Krasnyanskiy, D. L. Dedov, A. D. Obukhov, S. Yu. Alekseev // Journal of Computing and Information Science in Engineering. – 2020. – Vol. 20, No. 4. – P. 1–13. doi: 10.1115/1.4045289

5. Jensen, L. A review of the use of virtual reality head-mounted displays in education and training / L. Jensen, F. Konradsen // Education and Information Technologies. – 2018. – Vol. 23, No. 11. – P. 1515–1529. doi: 10.1007/s10639-017-9676-0

6. Система управления адаптивной беговой платформой на основе технологий компьютерного зрения / А. Д. Обухов, К. И. Патутин, Е. О. Суркова, А. Е. Архипов // Вестник Тамбовского государственного технического университета. – 2023. – Т. 29, № 3. – С. 363–374. doi: 10.17277/vestnik.2023.03.pp.363-374

7. Urdaneta-Ponte, M. C. Recommendation Systems for Education: Systematic Review / M. C. Urdaneta-Ponte, A. Mendez-Zorrilla, I. Oleagordia-Ruiz // Electronics. – 2021. – Vol. 10, No. 14. – Art. 1611. doi: 10.3390/electronics10141611

8. Краснянский, М. Н. Математическое моделирование адаптивной системы управления профессиональным образованием / М. Н. Краснянский, А. И. Попов, А. Д. Обухов // Вестник Тамбовского государственного технического университета. – 2017. – Т. 23, № 2. – С. 196–208. doi: 10.17277/vestnik.2017.02.pp.196-208

9. Gao, Y. The effectiveness of traditional tools and computer-aided technologies for health and safety training in the construction sector: A systematic review / Y. Gao, V. Gonzalez, K. T. W. Yiu // Computers & Education. – 2019. – Vol. 138. – P. 101–115. doi: 10.1016/j.compedu.2019.05.003

10. Попов, А. И. Концептуальные подходы к формированию и оцениванию компетенций будущих специалистов в экстремальной деятельности / А. И. Попов, С. В. Карпушкин, А. Д. Обухов // Профессиональное образование в России и за рубежом. – 2020. – № 1(37). – С. 51–59.

11. Pellas, N. Immersive Virtual Reality in K-12 and Higher Education: A systematic review of the last decade scientific literature / N. Pellas, S. Mystakidis, I. Kazanidis // Virtual Reality. – 2021. – Vol. 25, No. 6. – P. 835–861. doi: 10.1007/s10055-020-00489-9

12. Makransky, G. The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive Virtual Reality / G. Makransky, G. B. Petersen // Educational Psychology Review. – 2021. – Vol. 33, No. 3. – P. 937–958. doi: 10.1007/s10648-020-09586-2

13. Makransky, G. Adding Immersive Virtual Reality to a Science Lab Simulation Causes More Presence But Less Learning / G. Makransky, T. S. Terkildsen, R. E. Mayer // Learning and Instruction. – 2017. – Vol. 60. – P. 225–236. doi: 10.1016/j.learninstruc.2017.12.007

14. A Critical Review of Virtual and Augmented Reality (VR/AR) Applications in Construction Safety / X. Li, W. Yi, H.-L. Chi, [et al.] // Automation in Construction. – 2017. – Vol. 86. – P. 150–162. doi: 10.1016/j.autcon.2017.11.003

15. Информационная система управления профессиональным становлением студента в процессе самостоятельной работы / М. Н. Краснянский, А. И. Попов, А. Д. Обухов, С. В. Карпушкин // Вестник Самарского государственного технического университета. Серия: Психолого-педагогические науки. – 2019. – № 1(41). – С. 75–92. doi: 10.17673/vsgtu-pps.2019.1.6

16. Lampropoulos, G. Virtual Reality and Gamification in Education: A Systematic Review / G. Lampropoulos, Kinshuk // Educational Technology Research and Development. – 2024. – Vol. 72, No. 3. – P. 1691–1785. doi: 10.1007/s11423-024-10351-3

17. Virtual Reality, Augmented Reality, and Mixed Reality in Experiential Learning: Transforming Educational Paradigms / H. Crogman, V. D. Cano, E. Pacheco, R. B. Sonawane, R. Boroon// Education Sciences. – 2025. – Vol. 15, No. 3. – Art. 303. doi: 10.3390/educsci15030303

18. Obukhov, A. Algorithm of Adaptation of Electronic Document Management System Based on Machine Learning Technology / A. Obukhov, M. Krasnyanskiy, M. Nikolyukin // Progress in Artificial Intelligence. – 2020. – Vol. 9, No. 4. – P. 287–303. doi: 10.1007/s13748-020-00214-2

19. Резаев, А. В. ChatGPT и искусственный интеллект в университетах: какое будущее нам ожидать? / А. В. Резаев, Н. Д. Трегубова // Высшее образование в России. – 2023. – Т. 32, № 6. – С. 19–37. doi: 10.31992/0869-3617-2023-32-6-19-37

Application of Large Language Models for Organizing Professional Training Based on Immersive Systems

© A. Yu. Sveshnikov[1] ✉, Yu. V. Nikitnikov[1], M. A. Shiltsyn[1], D. L. Dedov[2]

[1] *Department of Automated Decision Support Systems, hammer68@mail.ru; TSTU, Tambov, Russian Federation;*
[2] *Department of Fundamental and Applied Research, TSTU, Tambov, Russian Federation*

Keywords: large language models; virtual training simulators; immersive systems; professional training; alternative ranking; recommendation systems; decision support system; employee digital profile.

Abstract: The paper examines the potential of using local large language models (LLMs) to rank virtual training simulators and immersive professional training tools based on an employee's digital profile. A semantic ranking task is formulated, utilizing LLMs to analyze semi-structured user data and match it against a catalog of available training solutions. An experimental evaluation was conducted using a dataset of 20 employee digital profiles from various industries and a catalog of virtual simulators covering diverse areas (occupational health and safety, electrical safety, fire safety, evacuation procedures, first aid, etc.). Ten local LLMs were compared. Recommendation quality was assessed using various accuracy metrics; additionally, the analysis covered the structured validity of responses, the correctness of returned simulator codes, and average execution time. A conclusion is drawn regarding the applicability of local LLMs as an intelligent component of a decision support system for selecting immersive professional training tools.

References

1. Radianti J., Majchrzak T.A., Fromm J., Wohlgenannt I. A Systematic Review of Immersive Virtual Reality Applications for Higher Education: Design Elements, Lessons Learned, and Research Agenda, *Computers & Education*, 2020, vol. 147, art. 103778. doi: 10.1016/j.compedu.2019.103778

2. Krasnyanskiy M.N., Dedov D.L., Obukhov A.D., Alekseev S.Yu. [System Analysis and Formalization of the Structure of Adaptive Training Complexes of Ergatic Systems], *Vestnik komp'yuternykh i informatsionnykh tekhnologiy* [Herald of Computer and Information Technologies], 2019, no. 4(178), pp. 45-52. doi: 10.14489/vkit.2019.04.pp.045-052 (In Russ., abstract in Eng.)

3. Scorgie D., Feng Z., Paes D. [et al.] Virtual Reality for Safety Training: A Systematic Literature Review and Meta-Analysis, *Safety Science*, 2023, vol. 171, art. 106372. doi:10.1016/j.ssci.2023.106372

4. Krasnyanskiy M.N., Dedov D.L., Obukhov A.D., Alekseev S.Yu. Visualization Technology and Tool Selection Methods for Solving Adaptive Training Complex Structural-Parametric Synthesis Problems, *Journal of Computing and Information Science in Engineering*, 2020, vol. 20, no. 4, pp. 1-13. doi: 10.1115/1.4045289

5. Jensen L., Konradsen F. A Review of the Use of Virtual Reality Head-Mounted Displays in Education and Training, *Education and Information Technologies*, 2018, vol. 23, no. 11, pp. 1515-1529. doi: 10.1007/s10639-017-9676-0

6. Obukhov A.D., Patutin K.I., Surkova Ye.O., Arkhipov A.Ye. [Control System of an Adaptive Running Platform Based on Computer Vision Technologies], *Transactions of the Tambov State Technical University*, 2023, vol. 29, no. 3, pp. 363-374. doi: 10.17277/vestnik.2023.03.pp.363-374 (In Russ., abstract in Eng.)

7. Urdaneta-Ponte M.C., Mendez-Zorrilla A., Oleagordia-Ruiz I. Recommendation Systems for Education: Systematic Review, *Electronics*, 2021, vol. 10, no. 14, art. 1611. doi: 10.3390/electronics10141611

8. Krasnyanskiy M.N., Popov A.I., Obukhov A.D. [Mathematical Modeling of an Adaptive Professional Education Management System], *Transactions of the Tambov State Technical University*, 2017, vol. 23, no. 2, pp. 196-208 doi: 10.17277/vestnik.2017.02.pp.196-208 (In Russ., abstract in Eng.)

9. Gao Y., Gonzalez V., Yiu K.T.W. The Effectiveness of Traditional Tools and Computer-Aided Technologies for Health and Safety Training in the Construction Sector: A Systematic Review, *Computers & Education*, 2019, vol. 138, pp. 101-115. doi: 10.1016/j.compedu.2019.05.003

10. Popov A.I., Karpushkin S.V., Obukhov A.D. [Conceptual Approaches to the Formation and Assessment of Competencies of Future Specialists in Extreme Activities], *Professional'noye obrazovaniye v Rossii i za rubezhom.* [Professional Education in Russia and Abroad], 2020, no. 1(37), pp. 51-59. (In Russ., abstract in Eng.)

11. Pellas N., Mystakidis S., Kazanidis I. Immersive Virtual Reality in K-12 and Higher Education: A systematic review of the last decade scientific literature, *Virtual Reality*, 2021, vol. 25, no. 6, pp. 835-861. doi: 10.1007/s10055-020-00489-9

12. Makransky G., Petersen G.B. The Cognitive Affective Model of Immersive Learning (CAMIL): A Theoretical Research-Based Model of Learning in Immersive Virtual Reality, *Educational Psychology Review*, 2021, vol. 33, no. 3, pp. 937-958. doi: 10.1007/s10648-020-09586-2

13. Makransky G., Terkildsen T.S., Mayer R.E. Adding Immersive Virtual Reality to a Science Lab Simulation Causes More Presence But Less Learning, *Learning and Instruction*, 2017, vol. 60, pp. 225-236. doi: 10.1016/j.learninstruc.2017.12.007

14. Li X., Yi W., Chi H.-L. [et al.] A Critical Review of Virtual and Augmented Reality (VR/AR) Applications in Construction Safety, *Automation in Construction*, 2017, vol. 86, pp. 150-162. doi: 10.1016/j.autcon.2017.11.003

15. Krasnyanskiy M.N., Popov A.I., Obukhov A.D., Karpushkin S.V. [Information system for managing the professional development of a student in the process of independent work], *Vestnik Samarskogo gosudarstvennogo tekhnicheskogo universiteta.*

Seriya: Psikhologo-pedagogicheskiye nauki [Bulletin of Samara State Technical University. Series: Psychological and Pedagogical Sciences], 2019, no. 1(41), pp. 75-92. doi: 10.17673/vsgtu-pps.2019.1.6 (In Russ., abstract in Eng.)

16. Lampropoulos G., Kinshuk. Virtual Reality and Gamification in Education: A Systematic Review, *Educational Technology Research and Development*, 2024, vol. 72, no. 3, pp. 1691-1785. doi: 10.1007/s11423-024-10351-3

17. Crogman H., Cano V.D., Pacheco E., Sonawane R.B., Boroon R. Virtual Reality, Augmented Reality, and Mixed Reality in Experiential Learning: Transforming Educational Paradigms, *Education Sciences*, 2025, vol. 15, no. 3, art. 303. doi: 10.3390/educsci15030303

18. Obukhov A., Krasnyanskiy M., Nikolyukin M. Algorithm of Adaptation of Electronic Document Management System Based on Machine Learning Technology, *Progress in Artificial Intelligence*, 2020, vol. 9, no. 4, pp. 287-303. doi: 10.1007/s13748-020-00214-2

19. Rezaev A.V., Tregubova N.D. ChatGPT and artificial intelligence in universities: what future should we expect?, *Vyssheye obrazovaniye v Rossii* [Higher Education in Russia], 2023, vol. 32, no. 6, pp. 19-37. doi: 10.31992/0869-3617-2023-32-6-19-37 (In Russ., abstract in Eng.)

Anwendung großer Sprachmodelle zur Organisation professioneller Schulungen auf Basis immersiver Systeme

Zusammenfassung: Der Artikel untersucht die Möglichkeit, lokale große Sprachmodelle (LLM) zur Bewertung virtueller Simulatoren und immersiver beruflicher Schulungstools auf Basis des digitalen Profils von Mitarbeitern einzusetzen. Das Ziel ist die semantische Bewertung mittels LLM auf Grundlage der Analyse semistrukturierter Nutzerdaten und deren Vergleich mit einem Katalog verfügbarer Simulatorlösungen. Die experimentelle Evaluierung ist mit 20 digitalen Profilen von Mitarbeitern verschiedener Branchen und einem Katalog virtueller Simulatoren aus unterschiedlichen Bereichen (Arbeitssicherheit, elektrische Sicherheit, Brandschutz, Evakuierung, erste Hilfe usw.) durchgeführt. Zehn lokale LLM wurden im Vergleich verwendet. Die Qualität der Empfehlungen wurde anhand verschiedener Genauigkeitsmetriken bewertet. Zusätzlich wurden die strukturierte Validität der Antwort, die Korrektheit der zurückgegebenen Simulatorcodes und die durchschnittliche Ausführungszeit analysiert. Abschließend ist die Anwendbarkeit lokaler LLM als intelligente Komponente des Systems zur Unterstützung der Auswahl immersiver beruflicher Schulungstools bewertet.

Application des grands modèles linguistiques pour la formation sur les systèmes immersifs

Résumé: Est envisagée la possibilité d'utiliser les grands modèles linguistiques locaux (MLL) pour classer les simulateurs virtuels et les outils de la formation immersive sur le profil numérique du collaborateur. La tâche de classement sémantique à l'aide de MLL est basée sur l'analyse des informations sur l'utilisateur et leur comparaison avec un catalogue des solutions de la simulation disponibles. L'évaluation pilote est réalisée sur un ensemble de 20 profils numériques d'employés de différents

secteurs et un catalogue des simulateurs virtuels couvrant de différents domaines (santé au travail, sécurité électrique, sécurité incendie, évacuation, premiers secours, etc.). Dix MLL locaux ont participé à la comparaison. La qualité des recommandations est évaluée selon diverses métriques d'exactitude, et l'on a également analysé la validité structurée des réponses, la correctitude des codes de simulateurs retournés ainsi que le temps moyen d'exécution. Est faite une conclusion sur l'applicabilité des MLL locales en tant que composant intellectuel d'un système de soutien de choix pour les moyens d'apprentissage professionnel immersif.

Авторы: ***Свешников Артем Юрьевич*** – старший преподаватель кафедры «Системы автоматизированной поддержки принятия решений», ***Никитников Юрий Витальевич*** – ассистент кафедры «Системы автоматизированной поддержки принятия решений»; ***Шильцын Максим Александрович*** – ассистент кафедры «Системы автоматизированной поддержки принятия решений», ФГБОУ ВО «ТГТУ», Тамбов, Российская Федерация; ***Дедов Денис Леонидович*** – кандидат технических наук, старший научный сотрудник, Управление фундаментальных и прикладных исследований, ФГБОУ ВО «ТГТУ», Тамбов, Российская Федерация.