

Современные методы сбора и обработки информации о рынке труда и направления их использования в практике управления человеческими ресурсами

И. Н. Калиновская *Витебский государственный технологический университет, Республика Беларусь*

Аннотация. В статье представлены результаты исследования методов получения данных о рынке труда из веб-источников. Актуальность исследований обусловлена необходимостью поиска инновационных методов сбора и анализа данных о рынке труда, обеспечивающих получение полной, достоверной и своевременной информации в условиях цифровой трансформации общества.

Цель исследования – сравнение методов сбора и анализа информации, выявление их преимуществ и недостатков. Методологическая база исследований опирается на междисциплинарный подход, объединяя методы науки о данных, машинного обучения, обработки естественного языка и управления человеческими ресурсами. Эмпирическая база – результаты экспертного опроса руководителей белорусских IT-компаний, данные платформы Artificial Analysis.

Определено, что большие языковые модели (LLM) превосходят другие методы по скорости, качеству и простоте использования. Представлен рейтинг топ-10 моделей, в котором GPT-4o признана самой эффективной. GPT-4o определена как приоритетная для бизнес-задач и экономического анализа. Выявлены преимущества и недостатки методов сбора и анализа данных.

Научная новизна исследований состоит в получении новых теоретических и эмпирических результатов, раскрывающих потенциал использования LLM для получения данных о рынке труда, а также в определении приоритетных направлений развития методологии и инструментария в этой области.

Практическая значимость заключается в разработке методики сравнительного анализа эффективности методов сбора и анализа информации о рынке труда с веб-источников; обосновании перспективных направлений развития методов сбора информации, открывающих возможности для создания инновационных решений в области управления человеческими ресурсами на базе искусственного интеллекта и обработки больших данных.

Исследование выполнено при финансовой поддержке «Белорусского республиканского фонда фундаментальных исследований» в рамках научного проекта «Теоретические подходы и методическое обеспечение анализа рынка труда в Республике Беларусь с применением больших данных» № Г24-013.

Ключевые слова: большие языковые модели, парсинг веб-страниц, API, RSS-каналы, сбор данных, анализ рынка труда.

Информация о статье: поступила 27 мая 2024 года.

Modern methods of collecting and processing information about the labour market and the directions of their use in the practice of human resource management

Iryna N. Kalinouskaya *Vitebsk State Technological University, Republic of Belarus*

Abstract. The article presents the results of a study on methods for obtaining labour market data from web sources. The relevance of the research stems from the need to explore innovative approaches to collecting and analyzing labour market data, ensuring comprehensive, reliable, and timely information in the context of the digital transformation of society.

The purpose of the study is to compare the methods of collecting and analyzing information, to identify their advantages and disadvantages. Methodologically, an interdisciplinary approach combines the methods of data science, machine learning, natural language processing and human resource management. Empirically, the study draws from expert surveys conducted with heads of Belarusian IT companies and data sourced from the Artificial Analysis platform.

Key findings reveal that large language models (LLM) outperform other methods in terms of speed, quality and ease of use. Among the top 10 models, GPT-4o is recognized as the most effective, prioritized for business tasks and economic

analysis. The advantages and disadvantages of data collection and analysis methods are revealed. The scientific novelty of the research consists in obtaining new theoretical and empirical results that reveal the potential of using LLM to obtain data on the labour market, as well as in identifying priority areas for the development of methodology and tools in this area. The practical significance lies in the development of a methodology for comparative analysis of the effectiveness of methods for collecting and analyzing information about the labour market from web sources; substantiation of promising directions for the development of information collection methods that open up opportunities for creating innovative solutions in the field of human resource management based on artificial intelligence and big data processing.

Financial support for the research was provided by the Belarusian Republican Foundation for Basic Research within the framework of the scientific project No Г24-013 "Theoretical approaches and methodological support for labour market analysis in the Republic of Belarus using big data".

Keywords: large language models, web page parsing, API, RSS feeds, data collection, labour market analysis.

Article info: received May 27, 2024.

Введение

Быстрые изменения в структуре экономики, технологические сдвиги и растущая глобальная конкуренция делают критически важным наличие у всех участников рынка труда (работодателей, соискателей, государственных органов) своевременной и достоверной информации о трендах в требованиях к навыкам и компетенциям. От способности оперативно выявлять и реагировать на эти тренды зависит экономическая эффективность и конкурентоспособность организаций, сбалансированность спроса и предложения труда в экономике, успешность профессиональных и карьерных траекторий работников.

Традиционные методы анализа рынка труда, основанные на данных государственной статистики, опросах работодателей и экспертных оценках, зачастую не позволяют в полной мере уловить стремительные изменения в требованиях к навыкам, особенно в высокотехнологичных и инновационных отраслях. Это приводит к негативным последствиям для экономики: организации испытывают дефицит необходимых специалистов и теряют в производительности, система образования готовит кадры с неактуальными компетенциями, увеличиваются расходы государства на переобучение безработных, тормозится внедрение новых технологий.

В этом контексте огромный потенциал для повышения качества и оперативности анализа рынка труда представляют новые подходы, основанные на применении технологий обработки больших данных и искусственного интеллекта, такие как использование больших языковых моделей (LLM). LLM позволяют извлекать актуальную информацию о требованиях к навыкам из огромных массивов данных о вакансиях, резюме,

профилях сотрудников в реальном времени, выявлять скрытые закономерности и тренды, недоступные для традиционных методов анализа.

Широкое внедрение подходов на базе LLM в практику анализа рынка труда может дать значительный экономический эффект. Организации получают возможность быстро адаптировать свою кадровую политику и инвестиции в обучение под меняющиеся требования, повышая продуктивность и экономя на затратах на найм и удержание работников. Государство сможет повысить эффективность политики занятости и образования, снизить дисбалансы на рынке труда и расходы на социальную поддержку безработных. Индивиды смогут делать более информированный выбор профессии и траектории развития своих навыков, повышая свою конкурентоспособность и доходы.

Актуальность темы сбора и анализа информации из онлайн-источников о рынке труда обусловлена следующими факторами:

- быстрые изменения на рынке труда: технологический прогресс, глобализация и экономические сдвиги приводят к постоянным изменениям в структуре занятости и требованиях к навыкам и компетенциям сотрудников организаций;
- необходимость принятия обоснованных кадровых решений: доступ к актуальной информации о рынке труда позволяет работодателям принимать более взвешенные решения при найме, обучении и развитии сотрудников;
- важность для соискателей: понимание текущих тенденций и требований рынка труда помогает соискателям выбирать перспективные направления для развития карьеры и приобретения востребованных компетен-

ций;

– большой объем данных в интернет-пространстве: онлайн-источники (сайты по поиску работы, профессиональные сообщества, социальные сети и др.) содержат огромное количество информации о вакансиях, требованиях работодателей и профессиональных профилях.

Целью данного исследования является изучение и сравнение различных методов сбора и анализа информации из онлайн-источников для проведения эффективной оценки рынка труда, выявление их преимуществ и недостатков.

Для достижения поставленной цели определены следующие задачи:

1. Изучить и проанализировать методы сбора и анализа информации из онлайн-источников, включая парсинг веб-страниц, использование API, работу с RSS-каналами и применение больших языковых моделей.
2. Выявить преимущества и недостатки каждого метода в контексте задач анализа рынка труда с учетом факторов, имеющих первостепенное значение при выборе метода получения информации.
3. Провести сравнительный анализ больших языковых моделей на основании критериев оценки публичной платформы ArtificialAnalysis.
4. Определить перспективные направления развития методов сбора и анализа информации о рынке труда и потенциальные области для дальнейших исследований.

Решение данных задач позволит получить комплексное понимание возможностей и ограничений различных методов сбора и анализа информации из онлайн-источников о рынке труда, что, в свою очередь, даст возможность работодателям, соискателям и специалистам по управлению человеческими ресурсами принимать более обоснованные решения и эффективно адаптироваться к изменяющимся условиям рынка.

Гипотеза исследования: использование больших языковых моделей (LLM) является наиболее эффективным и перспективным методом сбора и анализа информации о рынке труда по сравнению с традиционными подходами, такими как парсинг веб-страниц, использование API и RSS-каналов.

Предположения гипотезы:

– LLM способны обрабатывать и анализировать большие объемы неструктурированных текстовых данных, таких как описания вакансий, резюме, отзывы сотрудников и другую информацию, присутствующую в открытых онлайн-источниках. Это позволяет получать

более полную и разностороннюю картину рынка труда по сравнению с использованием только структурированных данных, доступных через API или RSS-каналы;

– современные архитектуры LLM, такие как трансформеры и модели с механизмом внимания, обладают способностью глубоко понимать контекст, улавливать семантические нюансы и извлекать информацию из текстов, связанных с рынком труда. Это обеспечивает высокую точность и качество полученных данных, сопоставимые или превосходящие результаты, достигаемые с помощью традиционных методов сбора и анализа информации;

– унифицированный интерфейс взаимодействия с LLM и возможность тонкой настройки предобученных моделей под конкретные задачи анализа рынка труда делают их более простыми в использовании и внедрении по сравнению с разработкой и поддержкой специализированных инструментов для парсинга веб-страниц или интеграции с множеством разнородных API;

– масштабируемость LLM позволяет эффективно обрабатывать растущие объемы данных о рынке труда и адаптироваться к новым источникам информации без значительного увеличения вычислительных ресурсов или затрат на разработку. Это делает их более подходящими для долгосрочного использования в условиях постоянно меняющегося рынка труда;

– стоимость использования LLM для сбора и анализа данных о рынке труда может быть оптимизирована за счет применения предобученных моделей, трансферного обучения и возможности обработки неструктурированных данных из открытых источников. Это потенциально делает LLM более экономически эффективным решением по сравнению с затратами на приобретение доступа к платным API или разработку и поддержку специализированных инструментов парсинга.

Для проверки данной гипотезы планируется провести сравнительный анализ эффективности использования LLM и традиционных методов сбора и анализа информации (парсинг веб-страниц, API, RSS-каналы) на основе ряда критериев, таких как скорость, качество, простота использования, масштабируемость и стоимость. Также будет приведено ранжирование топ-10 современных LLM в соответствии с их показателями производительности.

Ожидается, что результаты исследования продемонстрируют превосходство LLM над традиционными методами сбора и анализа информации о рынке труда

по ключевым параметрам эффективности и обоснуют целесообразность их приоритетного использования в данной области. Подтверждение гипотезы будет иметь важное значение для исследователей, разработчиков и организаций, занимающихся анализом рынка труда, стимулируя дальнейшие исследования, разработки и внедрение LLM в практику принятия решений в сферах управления человеческими ресурсами, развития карьеры и формирования государственной политики занятости.

Проблемой сбора и анализа информации о рынке труда с использованием онлайн-источников занимаются многие исследователи:

– Ф. Стивен, Ф. Бреземан, М. Грэм (Stephany, Braesemann & Graham, 2020) применили методы веб-скрапинга и анализа больших данных для изучения влияния пандемии COVID-19 на рынок труда в США и Великобритании. При этом они обнаружили значительные различия в динамике спроса на навыки в зависимости от отрасли и региона, а также выявили потенциальные долгосрочные изменения в структуре занятости;

– М. Мамертино и Т. М. Синклер (Mamertino & Sinclair, 2016) использовали данные с сайта Indeed для анализа динамики рынка труда в США. Собранные данные позволили выявить изменения в структуре спроса на профессии, географическое распределение вакансий и влияние экономических факторов на рынок труда;

– Р. Боселли, М. Чезарини, Ф. Меркорио, М. Меццанзаника (Boselli, Cesarini, Mercorio & Mezzanica, 2018; Mezzanica & Mercorio, 2018) воспользовались методами машинного обучения и обработки естественного языка для классификации и анализа вакансий, собранных с онлайн-платформ по поиску работы в Италии. Результаты показали возможность автоматизированного извлечения ключевых характеристик вакансий и построения прогнозных моделей спроса на навыки;

– Н. Джайн, С. К. Шантакумар, А. Шарма, А. Арора, Р. Мутараджу (Jain, Shanthakumar, Sharma, Arora & Mutharaju, 2020) задействовали методы парсинга веб-страниц и обработки естественного языка для анализа вакансий на индийском рынке труда. Результаты доказали, что автоматизированные методы сбора и анализа данных позволяют эффективно извлекать информацию о требованиях к навыкам и образованию, а также выявлять востребованные компетенции в различных отраслях;

– Р. Верма, С. Ш. Рагхаван (Verma & Raghavan, 2017)

использовали API сайтов по поиску работы для сбора и анализа данных о вакансиях в сфере информационных технологий в Индии. Анализ собранных данных позволил выявить наиболее востребованные технические навыки, опыт и предлагаемый уровень заработной платы;

– С. Ю. Роцин и С. А. Солнцев (Роцин и Солнцев, 2019) воспользовались данными вакансий, извлеченных с помощью парсинга с крупных порталов по поиску работы в Российской Федерации;

– Е. В. Ванкевич и И. Н. Калиновская (Vankevich & Kalinouskaya, 2020, 2021; Калиновская, 2021, 2022) применили метод скрейпинга данных для извлечения из различных онлайн-источников информации о компетенциях, востребованных рынком труда Республики Беларусь, и их анализа в разрезе регионов, групп занятий, должностей.

Проведенные исследования демонстрируют растущий интерес к использованию методов сбора и анализа информации из онлайн-источников о рынке труда. Полученные результаты подтверждают эффективность этих методов для извлечения ценных данных о требованиях работодателей, востребованных навыках и тенденциях в различных отраслях и регионах. Несмотря на значительные достижения в использовании методов сбора и анализа информации из онлайн-источников, в работах авторов можно выделить несколько нерешенных проблем:

1. Ограниченность охвата данных – большинство исследований фокусируются на анализе данных только с крупных онлайн-платформ по поиску работы, таких как Indeed или Monster, однако, эти платформы не охватывают все сегменты рынка труда, особенно в развивающихся странах или специфических отраслях. Таким образом, необходимо расширить список источников данных для получения более полной картины рынка труда.

2. Стандартизация и качество данных – вакансии, размещенные на различных онлайн-платформах, часто имеют разную структуру и формат описания, что затрудняет их автоматическую обработку и сравнение, требуются дополнительные усилия по очистке и классификации данных для повышения точности анализа.

3. Учет неявных требований и компетенций – многие вакансии содержат не только явные требования к навыкам и квалификации, но и неявные ожидания работодателей, такие как коммуникабельность, умение работать в команде и т. д. Извлечение и анализ этих неявных требований остается сложной задачей, требующей применения более современных методов обработки

естественного языка.

4. Динамика и скорость изменений на рынке труда – рынок труда постоянно эволюционирует, и спрос на навыки быстро меняется под влиянием технологических и экономических факторов, поэтому необходимы методы, позволяющие отслеживать и прогнозировать изменения в режиме реального времени, для обеспечения актуальности анализа.

5. Интеграция с другими источниками данных – онлайн-вакансии предоставляют ценную, но не исчерпывающую информацию о рынке труда, при этом для получения более полной картины необходима интеграция данных из других источников, таких как опросы работодателей, государственная статистика и социальные сети.

6. Учет региональных и отраслевых особенностей – рынки труда могут сильно различаться в зависимости от географического региона и отрасли, поэтому необходимы методы, учитывающие эти особенности и позволяющие проводить сравнительный анализ различных сегментов рынка.

7. Интерпретация и применение результатов – получение данных и их анализ является только частью поставленной задачи, необходимо развивать методы эффективной визуализации и интерпретации результатов, а также разрабатывать практические рекомендации для работодателей, соискателей и государственных органов на основе полученных выводов.

Использование больших языковых моделей (LLM) для сбора и анализа данных о рынке труда является новой и развивающейся областью исследований. Изучение данного направления описано в работах:

– Э. Бриньолфсон, Т. Митчелл и Д. Рок использовали LLM, для анализа описаний вакансий и выявления трендов в требованиях к навыкам на рынке труда США. (Brynjolfsson, Mitchell & Rock, 2021);

– Д. Аутор и его коллеги применяли LLM для изучения влияния автоматизации на рынок труда и профессиональные задачи (Autor, Reynolds, Chin, Fikri, Fleming, Katz, Kearney, Murray & Veuger, 2020; Autor, 2015);

– Л. А. Кароли, О. Леухина, С. Айенгар и И. Де-Беккер использовали LLM для анализа описаний вакансий и выявления навыков, востребованных на рынке труда в условиях цифровой экономики (Karoly, Leukhina, Iyengar & Dew-Becker, 2021);

– исследовательская группа LinkedIn использует собственные данные и LLM для анализа тенденций на рынке труда, востребованных навыков и профессий¹.

Стоит отметить, что многие исследования данной области находятся на ранних стадиях, а применение LLM для анализа рынка является молодым и весьма перспективным направлением.

Решение проблем получения информации о рынке труда из веб-источников требует междисциплинарного подхода, объединяющего методы науки о данных, машинного обучения, обработки естественного языка и исследования в области управления человеческими ресурсами. Дальнейшее изучение этого направления позволит повысить точность, полноту и практическую ценность анализа рынка труда на основе онлайн-данных.

Методы и средства исследований

Методологическая база исследований в области сбора и анализа информации о рынке труда с использованием онлайн-источников опирается на комплексный междисциплинарный подход, объединяющий методы и инструменты из различных областей, таких как наука о данных, машинное обучение, обработка естественного языка и управление человеческими ресурсами. Извлекаемые данные из онлайн-источников включают информацию о вакансиях (наименование организации, название должности, требования к навыкам и образованию, предлагаемая заработная плата), профили соискателей (опыт работы, навыки, образование, пол, возраст), а также метаданные (местоположение, отрасль, дата публикации). Для обеспечения репрезентативности и надежности результатов исследований применяются подходы к валидации и очистке данных – удаление дубликатов, обработка пропущенных значений и кросс-валидация моделей, применяются методы визуализации данных (графики, диаграммы, облака слов) для эффективного представления и интерпретации полученных результатов. Комбинирование различных методов и подходов позволяет проводить всесторонний и глубокий анализ рынка труда на основе данных из онлайн-источников, выявлять актуальные тренды, требования и закономерности, а также генерировать ценные идеи для принятия обоснованных решений всеми участниками рынка труда.

¹ <https://economicgraph.linkedin.com/>

Одним из ключевых методов получения информации из веб-источников является парсинг веб-страниц или веб-скрейпинг (web scraping) – автоматизированный сбор данных с веб-страниц, позволяющий извлекать информацию о вакансиях, требованиях к навыкам с крупных платформ по поиску работы, профессиональных социальных сетей, специализированных отраслевых сайтов и порталов, корпоративных сайтов компаний-работодателей, государственных порталов и баз данных о рынке труда.

Использование API (application programming interface) – это набор интерфейсов, предоставляемых онлайн-платформами для доступа к своим данным. API позволяют получать структурированную информацию о вакансиях и соискателях в формате JSON или XML, что упрощает процесс сбора и обработки данных.

RSS-каналы (really simple syndication) – это формат для синдикации и распространения часто обновляемого контента. RSS-каналы представляют собой XML-файлы, содержащие краткую информацию о новых публикациях на веб-сайте.

Большие языковые модели (large language models, LLM) – это класс моделей глубокого обучения, предназначенных для обработки и генерации естественного языка. Они способны извлекать информацию о вакансиях и резюме из веб-источников, используя различные методы обработки естественного языка (natural language processing, NLP).

Каждый из рассмотренных методов сбора и анализа информации имеет свои преимущества и ограничения. Выбор конкретного метода зависит от специфики задачи, доступности данных и технических возможностей. Для достижения наилучших результатов рекомендуется использовать комбинацию различных методов.

Эмпирическую базу выявления приоритетной модели сбора и анализа информации о рынке труда из веб-источников составили результаты экспертного опроса, который проводился в несколько этапов:

1. Формирование выборки экспертов. В опросе приняли участие 10 руководителей отделов по разработке данных (data engineering) и машинного обучения (machine learning) белорусских ИТ-компаний. Выбор экспертов осуществлялся целенаправленно, на основе их профессионального опыта и компетенций в области сбора и анализа больших данных, разработки систем

искусственного интеллекта. Все эксперты имеют опыт работы в этой сфере не менее 5 лет и занимают руководящие позиции в организациях.

2. Разработка опросника. Опросник состоял из трех блоков вопросов:

а) оценка преимуществ и недостатков каждого метода сбора и анализа информации (парсинг веб-страниц, использование API, анализ RSS-каналов, применение больших языковых моделей). Эксперты описывали сильные и слабые стороны методов в свободной форме;

б) оценка полноты охвата и качества данных, получаемых каждым методом, по 5-балльной шкале (где 1 – очень низкое качество/полнота, 5 – очень высокое);

в) сравнительная оценка методов по критериям: скорость получения информации, качество данных, простота использования, масштабируемость, стоимость. Для каждого критерия эксперты ранжировали методы от 1 (лучший) до 4 (худший).

3. Проведение опроса. Опрос проводился в формате индивидуальных структурированных интервью. Каждое интервью длилось около 60 минут и включало в себя заполнение опросника экспертом с последующим обсуждением и уточнением ответов с интервьюером. Это позволило получить не только количественные оценки, но и развернутые качественные комментарии экспертов.

4. Анализ результатов. Ответы экспертов были сведены в единую базу данных. Качественные оценки преимуществ и недостатков методов обобщены и систематизированы, выделены ключевые паттерны. Для количественных оценок рассчитаны средние значения, а также коэффициент конкордации Кендалла для оценки согласованности мнений экспертов. Значение коэффициента составило 0,721, что говорит о высокой согласованности экспертных оценок.

5. Интерпретация результатов. Полученные результаты были проанализированы и интерпретированы с точки зрения их значимости для решения задач сбора и обработки информации о рынке труда.

Эмпирическую базу определения приоритетной большой языковой модели для сбора и анализа информации о рынке труда из веб-источников составили данные платформы ArtificialAnalysis², в результате обзора которых выявлены наиболее важные показатели оценки больших языковых моделей, приведен топ-10 моделей.

² <https://huggingface.co/spaces/ArtificialAnalysis/LLM-Performance-Leaderboard>

Результаты исследований

С целью определения наиболее эффективного метода получения информации о рынке труда из веб-источников проведен опрос специалистов в области искусственного интеллекта, машинного обучения, сбора и обработки данных. В первой части опроса установлены преимущества и недостатки каждого метода извлечения информации, приведенные в таблице 1.

Выбор оптимального метода извлечения информации о рынке труда зависит от специфических требова-

ний и ограничений задачи исследования, таких как типы и форматы доступных данных, необходимость в структурированности и полноте извлекаемой информации, доступные ресурсы и бюджет.

Также экспертами проводилась оценка методов получения информации с точки зрения ее полноты и содержания:

1. У парсинга веб-страниц самый полный охват источников информации и содержания данных о рынке труда, так как данный метод позволяет собирать деталь-

*Таблица 1 – Достоинства и недостатки методов сбора и анализа информации о рынке труда из веб-источников
Table 1 – Advantages and disadvantages of methods for collecting and analyzing information about the labour market from web sources*

Метод	Преимущества метода	Недостатки метода
Парсинг веб-страниц	<ul style="list-style-type: none"> – возможность извлечения информации с любого веб-сайта; – гибкость в настройке для извлечения специфических данных; – не требует наличия специальных API или каналов данных от источника 	<ul style="list-style-type: none"> – трудоемкость разработки и поддержки для каждого отдельного веб-сайта; – риск нарушения правил использования сайтов и блокировки доступа при чрезмерной нагрузке; – зависимость от структуры и формата веб-страниц
API	<ul style="list-style-type: none"> – предоставление структурированных и стандартизированных данных; – обеспечение высокой скорости и надежности получения данных; – наличие документации и поддержки от провайдеров API 	<ul style="list-style-type: none"> – ограниченность доступных данных и функциональности, определяемых провайдером API; – необходимость оплаты за доступ к некоторым API или соблюдения лимитов использования; – зависимость от стабильности и доступности API-сервиса
RSS-каналы	<ul style="list-style-type: none"> – простота интеграции и обработки структурированных данных в формате XML; – возможность подписки на обновления и получения новых данных в режиме реального времени; – широкая поддержка стандарта RSS множеством веб-сайтов и платформ 	<ul style="list-style-type: none"> – ограниченность данных, предоставляемых через RSS; – зависимость от наличия и качества RSS-каналов у целевых веб-сайтов; – отсутствие гибкости в настройке и получении специфических данных, не включенных в канал
LLM	<ul style="list-style-type: none"> – возможность извлечения информации из неструктурированных текстовых данных; – способность понимать контекст и семантику текста, выявлять сущности и отношения; – универсальность применения к различным источникам данных и форматам 	<ul style="list-style-type: none"> – необходимость большого объема качественных обучающих данных для эффективного обучения модели; – вычислительная сложность и требовательность к ресурсам при обучении и использовании больших моделей; – риск генерации неточных или некорректных результатов; – сложность интерпретации и объяснения результатов

Источник: составлено автором по результатам опроса.

ные данные напрямую из первоисточников. Однако, при его использовании могут возникать проблемы с динамическими элементами страниц и блокировкой.

2. API имеет средний охват данных, поскольку они предоставляются в структурированном виде, но могут быть ограничены условиями использования и лимитами на количество запросов.

3. RSS-каналы обладают низким охватом данных, так как содержат только основную информацию и ссылки на источник. Данный метод полезен для отслеживания обновлений.

4. LLM при соответствующем обучении может собирать полные данные из источников, применяя методы парсинга и API, а также осуществлять обработку и анализ полученной информации.

Таким образом, для комплексного анализа рынка труда необходимо сочетание различных методов извлечения веб-данных, учитывая их сильные и слабые стороны. Парсинг является наиболее эффективным способом сбора и анализа детальной информации, но требует значительных технических усилий и может столкнуться с блокировками. API облегчают доступ к структурированным данным, но их использование ограничено условиями и лимитами провайдеров. RSS-каналы полезны для мониторинга обновлений, но не подходят как единственный источник информации из-за ее недостаточной глубины. Большие языковые модели открывают новые возможности по обработке, категоризации и обобщению собранной информации, позволяя извлекать дополнительные данные.

Для достижения наилучших результатов рекомендуется комбинировать извлечение первичных данных через парсинг и API с последующим применением LLM-техник для их анализа и интерпретации.

На втором этапе опроса экспертов были выявлены показатели оценки эффективности метода получения информации:

- скорость получения информации – отражает, как быстро данные о рынке труда могут быть собраны и обработаны с помощью метода;
- качество данных – показывает, насколько собранные данные соответствуют запросам о ситуации на рынке труда;
- простота использования метода на практике – оценивает легкость и удобство применения метода;
- масштабируемость – показывает, насколько легко метод может быть расширен для обработки больших объемов данных из множества источников;
- стоимость применения – отражает финансовые и ресурсные затраты на использование метода.

В заключительной части опроса эксперты с помощью баллов оценивали методы сбора и анализа информации по установленным показателям (1 – лучший метод, 4 – худший). В таблице 2 представлены результаты ранжирования и выделен приоритетный метод.

Анализ полученных результатов сравнения методов сбора и анализа информации о рынке труда Республики Беларусь из веб-источников показал, что:

- по скорости получения информации приоритетным является LLM метод, так как обученная модель мо-

Таблица 2 – Компаративная матрица методов сбора и анализа информации о рынке труда Республики Беларусь из веб-источников

Table 2 – Comparative matrix of methods for collecting and analyzing information about the labour market of the Republic of Belarus from web sources

Метод	Скорость получения информации	Качество данных	Простота использования	Масштабируемость	Стоимость	Сумма баллов	Ранг
Парсинг веб-страниц	3	2	4	2	3	14	3
API	2	3	2	1	4	12	2
RSS-каналы	4	4	3	3	1	15	4
LLM	1	1	1	4	2	9	1

Источник: составлено автором по результатам опроса.

жет быстро обрабатывать и анализировать огромные объемы неструктурированных данных о вакансиях, резюме и отзывах сотрудников, обеспечивая актуальную информацию о рынке труда в режиме реального времени;

- высокое качество данных обеспечивается методом LLM, поскольку современные модели, обученные на большом объеме данных о рынке труда, способны глубоко понимать контекст, извлекать релевантную информацию и обеспечивать высокую точность анализа, сопоставимую или превосходящую другие методы;

- простым для использования на практике является LLM метод, так как унифицированный интерфейс взаимодействия с LLM позволяет легко интегрировать модель в различные приложения и системы анализа рынка труда, не требуя сложной настройки или специальных навыков, в отличие от разработки парсеров или интеграции с множеством разнородных API;

- самым масштабируемым признан метод API, позволяющий легко масштабировать сбор данных о вакансиях и соискателях за счет стандартизированного интерфейса;

- наиболее доступными по финансовым и ресурсным затратам отмечены RSS-каналы, предоставляемые на бесплатной основе и не требующие значительных ресурсов для обработки данных о рынке труда.

Таким образом, по сумме баллов, выставленных экспертами, наиболее приоритетным методом сбора и анализа информации о рынке труда Республики Беларусь из веб-источников определено применение больших языковых моделей.

Рынок больших языковых моделей активно развивается, и в него вовлечены как крупные технологические компании, так и исследовательские институты и сообщества разработчиков открытого программного обеспечения. Разнообразие доступных моделей и их уникальные особенности создают богатую экосистему инструментов для решения различных задач обработки естественного языка, включая анализ рынка труда. Исследователи и практики могут выбирать модели, наиболее подходящие для их конкретных потребностей и целей, и комбинировать их для достижения оптимальных результатов. Дальнейшее развитие методов сбора и анализа информации о рынке труда будет определяться сочетанием технологических инноваций, растущих потребностей в качественных данных и социально-экономических факторов, таких как изменения в структуре

занятости и требованиях к навыкам.

В ходе исследований представлен сравнительный анализ больших языковых моделей на основании ряда показателей для оценки производительности LLM, выбранных из отчетов Artificial Analysis на Hugging Face. Artificial Analysis – это публичная платформа, которая позволяет исследователям и разработчикам сравнивать производительность различных моделей машинного обучения, в частности, больших языковых моделей, на основе их результатов по выбранным показателям.

В качестве критериев оценки больших языковых моделей выбраны:

1. Chatbot Arena – представляет собой сравнительный тест, в котором языковые модели соревнуются в задачах генерации текста, диалога, решения проблем и т. д. Оценки выражаются в процентах или баллах, отражающих точность, релевантность и качество ответов моделей.

2. MMLU (massive multitask language understanding) – набор из 57 задач на понимание естественного языка из различных областей, таких как гуманитарные науки, социальные науки и др. Оценка производится по средней точности на всех задачах и выражается в процентах, представляя долю правильных ответов модели.

3. MT Bench (machine translation benchmark) используется для оценки качества машинного перевода языковых моделей, сравнивая их производительность с профессиональными переводами и определяя точность, плавность и адекватность перевода. Оценки MT Bench измеряются в баллах BLEU (bilingual evaluation understudy).

4. Human Eval представляет собой оценку, в которой люди проверяют и оценивают производительность языковой модели на задачах генерации текста, ответов на вопросы, ведения диалогов, чтобы определить, насколько модель соответствует человеческим ожиданиям и стандартам. Оценки выражаются в баллах или рейтингах, которые дают люди-оценщики, и представлены в различных форматах, таких как средний балл по шкале от 1 до 5, проценты, отражающие качество работы модели по различным критериям.

Выбор этих показателей обусловлен их способностью предоставить всестороннюю, объективную и надежную оценку производительности больших языковых моделей в различных аспектах их работы. Данные показатели применимы и для оценки LLM, как инструмента анализа рынка труда:

– модель, показывающая высокие результаты в Chatbot Arena, может эффективно собирать данные с помощью опросов и анкетирования соискателей и работодателей, способна выявлять скрытые тренды;

– модель с высокими оценками в MMLU может обрабатывать и анализировать данные из различных источников, таких как отчеты по занятости, экономические прогнозы и т. д., для формирования комплексного представления о рынке труда;

– модель, успешная в MT Bench, может переводить вакансии и резюме на различные языки, что позволяет анализировать рынок труда в разных странах и регионах, выявлять глобальные тренды и потребности;

– модель с высокими оценками в Human Eval сможет эффективно интерпретировать отзывы работников и работодателей, анализировать социальные сети и форумы для понимания настроений и предпочтений на рынке труда.

В таблице 3 представлено ранжирование топ-10 больших языковых моделей с оценками по основным критериям из Hugging Face.

Анализ данных топ-10 больших языковых моделей позволяет сделать следующие выводы:

– модель GPT-4o занимает первое место в рейтинге,

демонстрируя наивысшие оценки по всем параметрам: Chatbot Arena (1310), MMLU (88,7 %), MT Bench (9,32) и Human Eval (90,2). Это указывает на её универсальность и высокую производительность в проведенных тестах и решенных задачах;

– на втором месте находится Claude 3 Opus, с высокими показателями в Chatbot Arena (1256), MMLU (86,8 %), MT Bench (8,06) и Human Eval (71,2), что подтверждает её эффективность и надежность при работе с текстами, связанными с рынком труда;

– модель Mistral Medium занимает третье место, получив высокие оценки в Chatbot Arena (1148), MMLU (75,3 %) и MT Bench (8,61). Однако отсутствие данных по Human Eval может свидетельствовать о неполной оценке её возможностей;

– некоторые модели, такие как Llama 3 (70B) и Open Chat 3.5, демонстрируют хорошие результаты в одних тестах, но отсутствующие показатели в других. Например, Llama 3 (70B) имеет высокие оценки в Chatbot Arena (1082) и MMLU (68,9 %), но отсутствуют данные по Human Eval и MT Bench;

– для некоторых моделей, таких как DBRX, Arctic и DeepSeek-V2, отсутствуют оценки по нескольким критериям, что затрудняет их полное сравнение с другими мо-

Таблица 3 – Рейтинг топ-10 больших языковых моделей

Table 3 – Ranking of the top-10 large language models

Название LLM	Оценка				Рейтинг
	Chatbot Arena, балл	MMLU, %	MT Bench, балл	Human Eval, балл	
GPT-4o	1310	88,7	9,32	90,2	1
DBRX	-	73,7	-	70,1	9
Llama 3 (70B)	1082	68,9	6,86	-	6
Claude 3 Opus	1256	86,8	8,06	71,2	2
Arctic	-	67,3	-	-	10
DeepSeek-V2	-	78,5	-	-	8
Mistral Medium	1148	75,3	8,61	-	3
Open Chat 3,5	1075	64,3	7,81	68,9	7
Gemini 1.0 Pro	1122	71,8	-	63,4	4
Mixtral 8x7B	1114	70,6	8,3	-	5

Источник: составлено автором по данным Artificial Analysis.

делями. Например, DBRX имеет только оценки по MMLU (73,7 %) и Human Eval (70,1), что делает её девятой в общем рейтинге;

– модели, такие как Gemini 1.0 Pro и Mixtral 8x7B, показывают средние результаты по большинству параметров, что свидетельствует об их надежности, но недостаточной конкурентоспособности в сравнении с лидерами рейтинга. Например, Gemini 1.0 Pro имеет оценку Chatbot Arena (1122), MMLU (71,8 %) и Human Eval (63,4);

– разнообразие оценок по различным метрикам подчёркивает важность многогранного подхода к оценке производительности языковых моделей. Высокие результаты в одном тесте не всегда коррелируют с аналогичными успехами в других тестах.

В целом, таблица 3 показывает, что лидирующие модели, такие как GPT-4o и Claude 3 Opus, демонстрируют высокие результаты по всем критериям, тогда как другие модели имеют специализированные сильные стороны и ограничения. Это подчеркивает необходимость выбора языковых моделей в зависимости от конкретных задач и контекста их применения.

На основании данных Artificial Analysis на Hugging Face³ представлен топ-3 больших языковых моделей, имеющих лучшие результаты в области решения задач бизнеса и экономики (таблица 4, максимальный балл оценивания результатов – 1).

Приоритетной моделью при решении бизнес-задач (оценка 0,7858) и задач в области экономики (оценка 0,8080) установлена GPT-4o. Модель демонстрирует глубокое понимание соответствующих концепций, умение оперировать специфической терминологией и давать обоснованные ответы на поставленные вопросы. Уч-

тывая эти результаты, компаниям и специалистам, работающим над проектами и исследованиями в сферах бизнеса и экономики, рекомендуется в первую очередь рассматривать применение модели GPT-4o. Её возможности позволят повысить эффективность и качество решения соответствующих задач, что даст ощутимые преимущества по сравнению с использованием других языковых моделей.

Направления использования больших языковых моделей в анализе рынка труда

Применение метода больших языковых моделей способствует анализу такой важной составляющей рынка труда, как динамика требований к профессиональным навыкам и компетенциям в ИТ-индустрии. Он позволяет изучить, как меняется спрос на определенные навыки программирования в зависимости от развития технологий и запросов бизнеса. Традиционно для этого используются методы контент-анализа описаний вакансий на сайтах по поиску работы, опроса работодателей или анализа учебных программ ИТ-специальностей в университетах. Однако, эти подходы имеют ограничения: выборка вакансий может быть нерепрезентативной, опросы дают субъективные оценки, а учебные программы отстают от реальных требований рынка.

Альтернативой традиционных методов является применение LLM для анализа большого массива текстовых данных, связанных с ИТ-рынком: описаний вакансий, профилей специалистов на профессиональных платформах, дискуссий на форумах разработчиков, документации по новым фреймворкам и технологиям. Обученная на этих данных LLM, может автоматически выявлять упоминания различных языков программиро-

Таблица 4 – Рейтинг топ-3 больших языковых моделей, имеющих лучшие результаты в области решения задач бизнеса и экономики

Table 4 – Ranking of the top-3 large language models with the best results in solving business and economic problems

Название LLM	Оценка		Рейтинг
	задачи из области бизнеса	задачи из области экономики	
GPT-4o	0,7858	0,8080	1
Claudi – 3 – Opus	0,7338	0,7980	2
Gemini – 1,5 – Flash	0,6670	0,6943	3

Источник: составлено автором по данным Artificial Analysis.

³ <https://huggingface.co/spaces/TIGER-Lab/MMLU-Pro>

вания, инструментов, методологий и оценивать частоту и контекст их использования. Так, модель может обнаружить резкий рост упоминаний нового фреймворка для разработки мобильных приложений в вакансиях и обсуждениях на форумах за последние несколько месяцев. Это ранний сигнал того, что спрос на специалистов, владеющих этим фреймворком, вскоре может значительно вырасти. HR-службы организаций, получив такую информацию, могут оперативно скорректировать требования при найме разработчиков или организовать переобучение существующих сотрудников, чтобы не упустить новый технологический тренд. Или, например, LLM может выявить постепенное снижение частоты упоминания определенного языка программирования в профилях специалистов и вакансиях за последние годы. Это может свидетельствовать об устаревании технологии и необходимости для разработчиков, использующих этот язык, задуматься о расширении своего пучка навыков для сохранения конкурентоспособности на рынке труда.

Важно, что LLM позволяет анализировать не только формальное наличие или отсутствие упоминаний тех или иных навыков, но и контекст, тональность, взаимосвязи между ними в профессиональной коммуникации. Это дает более глубокое понимание реальных трендов и требований рынка. Конечно, такой анализ не заменяет полностью другие методы изучения динамики навыков, но дает ценное дополнение к ним, обеспечивая более оперативный и масштабный мониторинг ситуации. Он также подходит для выявления новых, зарождающихся трендов, которые могут быть не видны при использовании традиционных методов.

Несмотря на большие возможности и высокую производительность больших языковых моделей, исследователи, использующие LLM для анализа рынка труда, могут столкнуться с рядом проблем и ограничений:

1. Доступность и качество данных:

- обучение LLM требует большого объема качественных данных о рынке труда, таких как описания вакансий, профили соискателей и т. д.;

- сбор и подготовка данных может быть трудоемким и затратным процессом, особенно для специфических отраслей или регионов;

- данные могут быть неполными, неструктурированными или содержать шум, что затрудняет их эффективное использование для обучения моделей.

2. Интерпретируемость и объяснимость результатов:

- LLM воспринимаются как «черные ящики», и их внутренняя логика может быть сложной для понимания и интерпретации;

- исследователи могут столкнуться с трудностями при объяснении полученных результатов и выводов;

- необходимость обеспечения прозрачности и объяснимости моделей может ограничивать их сложность и эффективность.

3. Обобщаемость и переносимость моделей:

- LLM обучаются на конкретных наборах данных и могут не обобщаться на новые или отличающиеся контексты;

- модели, обученные на данных одной отрасли или региона, могут быть менее эффективны для анализа рынка труда в других условиях;

- адаптация моделей к новым задачам или областям может потребовать значительных усилий по переобучению и настройке.

4. Временная динамика и актуальность:

- рынок труда постоянно меняется, и модели, обученные на исторических данных, могут не учитывать последние тенденции и изменения;

- исследователям необходимо регулярно обновлять и переобучать модели, чтобы обеспечить их актуальность и точность;

- отслеживание и прогнозирование будущих трендов на рынке труда может быть сложной задачей, даже с использованием LLM.

5. Этические и правовые аспекты:

- использование LLM для анализа данных рынка труда может поднимать вопросы конфиденциальности, безопасности и этичного использования информации;

- исследователи должны соблюдать правовые нормы и этические принципы при сборе, хранении и анализе данных о соискателях и работодателях.

6. Вычислительные ресурсы и затраты:

- обучение и использование LLM, особенно моделей большого размера, требует значительных вычислительных ресурсов и может быть дорогостоящим;

- исследователям необходимо найти баланс между сложностью моделей, качеством результатов и доступными ресурсами;

- затраты на инфраструктуру и обслуживание моделей могут быть препятствием для организаций или исследовательских групп.

В целом, исследование показывает, что применение LLM является перспективным направлением для анализа рынка труда, несмотря на имеющиеся ограничения. Дальнейшие исследования должны быть направлены на преодоление выявленных проблем и развитие возможностей LLM в этой области.

Применение больших языковых моделей при анализе рынка труда, способно дать следующие измеримые результаты:

1. Повышение эффективности подбора персонала и снижение затрат на найм. Точное выявление требуемых навыков и компетенций с помощью LLM позволит работодателям быстрее находить подходящих кандидатов и сокращать время и ресурсы на заполнение вакансий. Потенциальный эффект может быть оценен через сокращение средней стоимости и длительности закрытия вакансии.

2. Оптимизация инвестиций в обучение и развитие персонала. Выявление актуальных трендов спроса на навыки даст организациям возможность более целенаправленно инвестировать в программы повышения квалификации сотрудников. При этом эффективность может быть измерена через повышение экономической эффективности инвестиций в обучение и развитие сотрудников (return on investment, ROI, – отношение прироста производительности к затратам на обучение).

3. Рост производительности труда. Как уже отмечалось, более эффективное развитие востребованных компетенций у работников будет способствовать повышению их индивидуальной продуктивности, на макроуровне это может привести к приросту производительности труда в экономике.

4. Сокращение затрат на социальную поддержку безработных и программы переобучения. Снижение структурной безработицы в результате более эффективного выявления несоответствий в требуемых и доступных навыках на рынке позволит государству оптимизировать расходы на пособия по безработице и программы переквалификации.

5. Прирост экспорта образовательных услуг и технологий анализа данных. Регионы, наиболее активно развивающие и применяющие передовые методы анализа рынка труда, смогут экспортировать свои образовательные программы и технологические решения, ориентированные на актуальные навыки.

Важно, что многие из этих эффектов будут иметь косвенное влияние на другие сферы экономики через

повышение общей эффективности использования трудовых ресурсов, ускорение инноваций и цифровизации. Это создаст потенциал для дополнительного экономического роста и повышения конкурентоспособности на глобальном уровне.

Для оценки эффективности применения больших языковых моделей, предлагаются следующие методики и подходы к сбору и анализу необходимых статистических данных:

1. Методика оценки влияния LLM на эффективность подбора персонала:

- отслеживание динамики среднего времени заполнения вакансии и стоимости закрытия позиции до и после внедрения LLM в процесс рекрутинга. Сбор данных через системы трекинга рекрутмента и внутреннюю финансовую отчетность организации;

- проведение опросов среди рекрутеров и менеджеров по найму для оценки их удовлетворенности качеством подбора и эффективностью процесса после внедрения новых методов анализа;

- анализ корреляции между использованием LLM и показателями эффективности подбора персонала на основе данных от множества организаций.

2. Методика оценки влияния на ROI программ обучения и развития:

- отслеживание изменений в производительности, качестве работы, уровне удовлетворенности клиентов для сотрудников, прошедших обучение по программам, разработанным с учетом выявленных LLM трендов в требованиях к навыкам. Сбор данных через системы управления эффективностью персонала, опросы клиентов и внутреннюю отчетность;

- расчет показателя ROI как отношение приведенной стоимости дополнительных доходов или экономии от повышения производительности к затратам на обучение по новым программам. Сравнение с ROI для традиционных программ обучения;

- метаанализ результатов оценки ROI программ обучения, основанных на данных LLM, для организаций из разных отраслей и регионов.

3. Методика оценки влияния на производительность труда:

- разработка экономико-математических моделей, оценивающих вклад факторов, связанных с внедрением LLM (доли сотрудников, обученных по новым программам, основанным на данных LLM), в динамику производительности труда на уровне организации и отраслей.

Построение моделей на основе статистических данных, собираемых на микроуровне;

- проведение эконометрического анализа на основе панельных данных для организаций и отраслей, учитывающего внедрение методов анализа на основе LLM как один из факторов, влияющих на производительность. Сбор необходимых данных через государственные органы статистики, отраслевые ассоциации, финансовую отчетность организации;

- организация специальных статистических наблюдений и опросов компаний и работников по вопросам влияния новых методов анализа рынка труда на производительность.

4. Методика оценки влияния на затраты по социальной поддержке безработных:

- проведение эксперимента по внедрению программ переобучения и трудоустройства безработных на основе данных о востребованных навыках, полученных через LLM. Оценка затрат на эти программы и их эффективности в снижении длительности периода безработицы и повышении уровня трудоустройства по сравнению с контрольной группой. Сбор данных через службы занятости и социальной защиты;

- построение экономико-математических моделей, прогнозирующих изменение затрат государства на пособия и программы поддержки безработных при различных сценариях внедрения методов анализа на основе LLM и его влияния на структурную безработицу. Использование данных государственной статистики и результатов экспериментальных исследований;

- анализ международного опыта применения методов анализа рынка труда для адаптации программ поддержки безработных и оценка экономического эффекта этих инициатив. Сотрудничество с международными организациями, такими как МОТ, ОЭСР, для сбора и анализа данных.

5. Методика оценки влияния на экспорт образовательных услуг и технологий:

- мониторинг динамики экспортных доходов организаций, предоставляющих услуги по обучению навыкам и компетенциям, выявленным с помощью LLM как наиболее востребованные на глобальном рынке. Сбор данных через финансовую отчетность, опросы организаций.

- отслеживание роста продаж и экспорта программного обеспечения и технологических решений для анализа рынка труда на основе LLM. Получение данных через финансовую отчетность организаций, специали-

зированные исследования рынка;

- оценка кросс-эффектов между активностью стран и регионов во внедрении методов анализа на основе LLM и их позициями в глобальных рейтингах конкурентоспособности, инновационности, привлекательности для талантов. Использование данных международных организаций и рейтинговых агентств.

Важно отметить, что для полноценной реализации этих методик необходимо наладить тесное взаимодействие между различными стейкхолдерами – организациями, государственными органами, исследовательскими институтами, учреждениями образования. Предложенные методики и подходы к сбору и анализу данных могут стать основой для комплексной оценки экономической эффективности и целесообразности инвестиций во внедрение инновационных методов анализа рынка труда, содействуя принятию более обоснованных решений на уровне организаций, отраслей и государственной политики.

Анализ полученных результатов

Основываясь на результатах исследования методов сбора и анализа информации о рынке труда из веб-источников, можно сделать следующие выводы относительно выдвинутой гипотезы:

- гипотеза о том, что использование больших языковых моделей (LLM) является наиболее эффективным и перспективным методом сбора и анализа информации о рынке труда по сравнению с традиционными подходами, такими как парсинг веб-страниц, использование API и RSS-каналов, в целом подтверждается результатами исследования;

- экспертный опрос показал, что LLM превосходят другие методы по ключевым показателям эффективности, таким как скорость получения информации, качество данных и простота использования. Это согласуется с предположением гипотезы о преимуществах LLM в обработке больших объемов неструктурированных текстовых данных и извлечении информации;

- рейтинг топ-10 больших языковых моделей демонстрирует высокую производительность современных LLM, таких как GPT-4o, в решении различных задач обработки естественного языка. Это подтверждает предположение гипотезы о способности LLM обеспечивать высокую точность и качество анализа данных о рынке труда;

- выявленные проблемы и ограничения использования LLM, такие как необходимость большого объема

качественных данных для обучения, сложность интерпретации результатов и высокие требования к вычислительным ресурсам, частично опровергают предположения гипотезы о безусловном превосходстве LLM над традиционными методами. Эти ограничения указывают на необходимость дальнейших исследований и разработок для повышения эффективности и применимости LLM в анализе рынка труда.

В целом, результаты исследования в значительной степени подтверждают основные положения выдвинутой гипотезы, демонстрируя перспективность использования LLM для сбора и анализа информации о рынке труда. Однако, выявленные проблемы и ограничения свидетельствуют о том, что LLM не являются универсальным решением и требуют дальнейшего изучения и совершенствования. Для получения более полного представления о возможностях и ограничениях LLM в контексте анализа рынка труда необходимы дополнительные исследования, учитывающие специфику различных отраслей, регионов и типов данных.

Основываясь на полученных результатах, разработан план дальнейших исследований в области применения LLM моделей с целью анализа рынка труда. Первоочередной задачей стоит изучение возможностей интеграции LLM с другими методами и источниками данных о рынке труда. Для анализа возможностей интеграции LLM необходимо провести следующее исследование:

1. Определение источников данных:

- идентифицировать основные источники структурированных и неструктурированных данных о рынке труда, такие как базы данных вакансий, результаты опросов работодателей и соискателей, профили и активность пользователей в профессиональных социальных сетях;
- оценить доступность, полноту и качество данных из каждого источника, а также возможные ограничения и правовые аспекты их использования.

2. Разработка схемы интеграции:

- создать концептуальную схему интеграции данных из различных источников, учитывая их формат, структуру и семантику;
- определить роль LLM в процессе интеграции;
- разработать методы преобразования и нормализации данных для их совместного анализа и использования.

3. Сбор и предобработка данных:

- реализовать процедуры сбора данных из выбранных источников;

- провести предварительную обработку собранных данных, такую как очистка, структурирование и анонимизация персональной информации;

- подготовить данные для подачи на вход LLM.

4. Интеграция данных и анализ:

- применить LLM для извлечения сущностей, отношений и ключевых характеристик из неструктурированных текстовых данных;

- интегрировать извлеченную информацию со структурированными данными из других источников, используя разработанную схему интеграции;

- провести комплексный анализ интегрированных данных, применяя методы статистического анализа, машинного обучения и визуализации данных для выявления закономерностей, трендов и взаимосвязей.

5. Оценка качества и эффективности:

- оценить полноту, достоверность и согласованность информации, полученной в результате интеграции данных из различных источников с использованием LLM;

- сравнить результаты анализа интегрированных данных с результатами, полученными на основе отдельных источников данных и методов анализа;

- проанализировать вклад LLM в улучшение качества и полноты информации о рынке труда, а также в автоматизацию процессов сбора и анализа данных.

6. Выводы и рекомендации:

- обобщить результаты исследования и сделать выводы о потенциале комбинированных подходов, включающих LLM и другие методы и источники данных, для получения более полной и достоверной информации о рынке труда;

- определить наиболее эффективные стратегии интеграции данных и использования LLM в сочетании с другими методами анализа;

- сформулировать рекомендации по внедрению комбинированных подходов в практику анализа рынка труда и поддержки принятия решений в области управления человеческими ресурсами.

Проведение такого исследования позволит количественно и качественно оценить потенциал интеграции LLM с другими методами и источниками данных для получения более полной и достоверной информации о рынке труда. Результаты исследования могут быть использованы для разработки инновационных решений в области HR-аналитики, прогнозирования спроса на навыки и компетенции, оптимизации процессов рекрутинга и управления талантами.

Внедрение новых методов анализа рынка труда, таких как большие языковые модели, может оказать существенное влияние на функционирование экономики в целом. Более эффективное выявление трендов спроса на навыки с помощью LLM способно привести к следующим макроэкономическим эффектам:

1. Трансформация системы образования и профессиональной подготовки. Своевременное выявление изменений в требованиях к навыкам позволит учебным заведениям адаптировать свои программы под актуальные запросы рынка труда. Это снизит риск подготовки специалистов с неактуальными компетенциями и повысит качество человеческого капитала в экономике. В результате возрастет соответствие между структурой подготовки кадров и реальными потребностями бизнеса, что положительно скажется на общей эффективности экономики.

2. Рост производительности труда. Точная информация о востребованных навыках даст работникам возможность целенаправленно развивать свои компетенции в соответствии с требованиями рынка. Это повысит их производительность и ценность для работодателей. Организации, в свою очередь, смогут эффективнее подбирать и обучать сотрудников, что также положительно повлияет на производительность. На макроуровне это будет способствовать росту общей производительности труда в экономике, что является одним из ключевых факторов долгосрочного экономического роста.

3. Сокращение структурной безработицы. Структурная безработица возникает из-за несоответствия навыков безработных требованиям доступных вакансий. Применение LLM для анализа рынка труда поможет лучше выявлять такие несоответствия и вырабатывать меры по их устранению, например, через программы переобучения или стимулы для развития определенных профессий. Это будет способствовать снижению структурной безработицы и более эффективному использованию трудовых ресурсов в экономике.

4. Ускорение технологических инноваций и цифровой трансформации. Выявление трендов спроса на навыки с помощью LLM позволит точнее прогнозировать направления технологического развития и потребности в специалистах для их реализации. Это даст бизнесу и государству возможность концентрировать ресурсы на поддержке наиболее перспективных инноваций, в том числе связанных с цифровизацией экономики. Ускорение технологического прогресса, в свою очередь, будет

стимулировать рост производительности, создание новых продуктов и бизнес-моделей.

5. Повышение инвестиционной привлекательности и конкурентоспособности экономики. Регионы, активно применяющие передовые методы анализа рынка труда, будут иметь преимущество в качестве человеческого капитала и скорости адаптации к новым технологическим трендам. Это сделает их более привлекательными для инвесторов, стимулируя приток капитала и создание высокотехнологичных производств. Выявление уникальных трендов спроса на навыки поможет развивать конкурентные преимущества на глобальном рынке.

В то же время необходимо учитывать и потенциальные риски внедрения методов на основе искусственного интеллекта для анализа рынка труда. Среди них – вероятность ошибочных прогнозов из-за несовершенства моделей, возможное усиление неравенства на рынке труда из-за неравномерного доступа к таким инструментам, риски высвобождения части аналитиков рынка труда в результате автоматизации их функций.

Грамотное регулирование и политика государства по поддержке равномерного внедрения новых технологий анализа рынка труда позволят снизить эти риски и реализовать положительные макроэкономические эффекты от их применения. Тесное взаимодействие бизнеса, образовательных учреждений и государства на базе получаемых аналитических данных будет способствовать созданию эффективного и инклюзивного рынка труда, отвечающего потребностям динамично развивающейся экономики.

Выводы

В результате проведенных исследований определен наиболее эффективный метод сбора и анализа данных о рынке труда из веб-источников – применение больших языковых моделей (LLM). LLM превосходят другие методы по скорости получения информации, качеству данных и простоте использования.

Выявлены преимущества и недостатки методов сбора и анализа информации о рынке труда из веб-источников: парсинг веб-страниц обеспечивает самый полный охват данных, но трудоемок в разработке и поддержке; API предоставляют структурированные данные, но ограничены условиями использования; RSS-каналы полезны для отслеживания обновлений, но содержат мало информации; LLM могут обрабатывать неструктурированные данные и проводить глубокий анализ.

Приведен рейтинг топ-10 больших языковых моделей по производительности на основе данных Artificial Analysis платформы Hugging Face. Модель GPT-4o от OpenAI заняла первое место, продемонстрировав высокие результаты по всем показателям. Также модель GPT-4o определена в качестве приоритетной для решения бизнес-задач и задач в области экономики по результатам анализа данных Artificial Analysis. Компаниям и специалистам рекомендуется использовать GPT-4o для повышения эффективности и качества решения соответствующих задач.

Выявлены проблемы и ограничения использования LLM: необходимость большого объема качественных данных, сложность интерпретации и объяснения результатов, полученных с помощью LLM, ограниченная обобщаемость моделей на новые контексты и области, необходимость регулярного обновления моделей для учета динамики рынка труда, этические и правовые аспекты использования данных о соискателях и работодателях, высокие требования к вычислительным ресурсам и затратам на обучение моделей.

Результаты исследования могут иметь высокую прикладную ценность для различных участников рынка труда:

1. Работодатели могут применять результаты анализа рынка труда с помощью LLM для оптимизации процессов найма и управления персоналом:

- быстрая адаптация требований в вакансиях под актуальные тренды в навыках, выявленные с помощью LLM, позволит привлекать кандидатов с востребованными компетенциями и сократит время и затраты на заполнение позиций;

- использование данных о востребованных навыках для оптимизации программ обучения и развития сотрудников. Организации смогут сфокусировать инвестиции в обучение на развитии наиболее важных для бизнеса компетенций, повышая продуктивность человеческих ресурсов и отдачу от этих инвестиций;

- применение выводов LLM для планирования будущих потребностей в персонале и развития кадрового резерва. HR-службы получат возможность заблаговременно выявлять потенциальные разрывы в навыках и компетенциях и принимать меры для их устранения;

- интеграция данных о трендах в навыках в системы мотивации и оценки эффективности персонала позволит поощрять сотрудников за развитие востребованных компетенций и привязать компенсацию к их ценности на

рынке, повышая удержание ключевых специалистов.

2. Государственные органы могут использовать результаты анализа на основе LLM для совершенствования политики занятости и развития человеческого капитала:

- корректировка государственных программ профессионального обучения и переподготовки с учетом выявленных трендов на рынке труда. Переориентация этих программ на развитие наиболее востребованных компетенций повысит трудоустраиваемость их выпускников и эффективность госрасходов;

- адаптация системы образования под требования новой экономики. Учет выводов LLM при разработке образовательных стандартов и учебных программ обеспечит подготовку специалистов с актуальными для рынка компетенциями, сократит дисбаланс в структуре спроса и предложения труда;

- использование данных о трендах в навыках для привлечения инвестиций и поддержки развития перспективных отраслей. Демонстрация инвесторам наличия кадров с востребованными навыками или госпрограмм по их развитию повысит инвестиционную привлекательность этих отраслей;

- применение выводов LLM для совершенствования службы занятости и программ поддержки безработных. Переобучение безработных с учетом реальных требований рынка ускорит их трудоустройство, снизит расходы на пособия и повысит эффективность использования трудовых ресурсов.

Соискатели могут применять результаты анализа рынка труда с помощью LLM для развития своей карьеры:

- выбор программ обучения и повышения квалификации, нацеленных на освоение наиболее востребованных работодателями навыков. Это повысит ценность соискателя на рынке труда и его шансы на трудоустройство на высокооплачиваемые позиции;

- планирование карьерных траекторий с учетом долгосрочных трендов в требованиях к навыкам. Понимание, какие компетенции будут цениться в отрасли через 5-10 лет, позволит соискателю делать правильный стратегический выбор работодателя, должности, программ развития;

- проведение самооценки своих компетенций в сравнении с актуальными требованиями рынка, выявленными с помощью LLM. Определение своих сильных сторон и областей для развития позволит соискателю

более эффективно позиционировать себя на рынке и инвестировать в свой человеческий капитал;

– использование данных о востребованных навыках для обоснования своей ценности для работодателя на переговорах о зарплате и условиях найма. Опора на объективные данные о трендах рынка усилит переговорную позицию соискателя.

Таким образом, внедрение методов анализа рынка труда на основе LLM открывает широкие возможности

для повышения эффективности экономических решений, принимаемых всеми участниками рынка.

Исследование выполнено при финансовой поддержке «Белорусского республиканского фонда фундаментальных исследований» в рамках научного проекта «Теоретические подходы и методическое обеспечение анализа рынка труда в Республике Беларусь с применением больших данных» № Г24-013.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

Калиновская, И. Н. (2021). Теоретические аспекты подбора кадров с применением технологий искусственного интеллекта. *Право. Экономика. Психология*, № 1 (21), С. 48–64.

Калиновская, И. Н. (2022). Анализ представленных на рынке труда компетенций, извлеченных из цифровых источников с помощью искусственного интеллекта. *Экономика и общество: международный научно-практический журнал*, № 04 (22), С. 29–42.

Рощин, С. Ю. и Солнцев, С. А. (2019). Индекс напряженности на российском рынке труда. *Мониторинг экономической ситуации в России: тенденции и вызовы социально-экономического развития*, № 1 (84), С. 15–18.

Autor, D. (2015). Why are there still so many jobs? The history and future of workplace automation. *The Journal of Economic Perspectives*, № 29 (3), pp. 3–30.

Autor, D., Reynolds, E., Chin, A., Fikri, K., Fleming, W. B., Katz, L. F., Kearney, M. S., Murray, S. C., Siegel, D. and Veuger, S. (2020). The Work of the Future: Building Better Jobs in an Age of Intelligent Machines. *MIT Work of the Future*, [Online], URL: <https://workofthefuture.mit.edu/wp-content/uploads/2021/01/2020-Final-Report4.pdf>, [Accessed: 23.05.2024].

Karoly, L. A., Leukhina, O., Iyengar, S. and Dew-Becker, I. (2021). The Digital Workforce: Developing Skills for an AI-Driven Economy. *RAND Corporation*, [Online], URL: https://www.rand.org/pubs/research_reports/RRA1533-1.html, [Accessed: 24.05.2024].

Mamertino, M. and Sinclair, T. M. (2016). Online Job Search and Migration Intentions Across EU Member States. *Institute for International Economic Policy Working Paper Series, IIEP-WP-2016-4*, [Online], URL: <https://www2.gwu.edu/~iiep/assets/docs/papers/2016WP/MamertinoSinclairIIEPWP2016-4.pdf>, [Accessed: 24.05.2024].

Boselli, R., Cesarini, M., Mercurio, F. and Mezzanzanica, M. (2018). Classifying online Job Advertisements through Machine Learning. *Future Generation Computer Systems*, vol. 86, pp. 319–328. <https://doi.org/10.1016/j.future.2018.03.035>.

Brynjolfsson, E., Mitchell, T. and Rock, D. (2021). What Can Machines Learn, and What Does It Mean for Occupations and the Economy? *AEA Papers and Proceedings*, vol. 111, pp. 43–47. DOI: 10.1257/pandp.20211019.

Jain, N., Shanthakumar, S. K., Sharma, A., Arora, A. and Mutharaju, R. (2020). Analyzing and Visualizing the Skill Demand in Indian Job Market using Web Crawling and Machine Learning. *In 2020 International Conference on Intelligent Engineering and Management (ICIEEM)*, pp. 303–308. <https://doi.org/10.1109/ICIEEM48762.2020.9160247>.

Mezzanzanica, M. and Mercurio, F. (2018). Big Data Enables Labor Market Intelligence. *Encyclopedia of Big Data Technologies*, pp. 1–11. https://dx.doi.org/10.1007/978-3-319-63962-8_276-1.

Stephany, F., Braesemann, F. and Graham, M. (2020). Coding together – coding alone: the role of trust in collaborative programming. *Information, Communication & Society*, vol. 24, pp. 1944–1941. <https://doi.org/10.1080/1369118X.2020.1749699>.

Vankevich, A. and Kalinouskaya, I. (2020). Ensuring sustainable growth based on the artificial intelligence analysis and forecast of in-demand skills. *First Conference on Sustainable Development: Industrial Future of Territories (IFT 2020)*, Vol. 208, <https://doi.org/10.1051/e3sconf/202020803060>.

Vankevich, A. and Kalinouskaya, I. (2021). Better understanding of the labour market using Big Data. *Ekonomia i prawo. Economics and law*, vol. 20, № 3, pp. 677–692.

Verma, R. and Raghavan, S. S. (2017). Extrapolating Trends in Demand and Supply of IT Professionals: *The Indian Scenario*. In *International Conference on Research into Design*, Springer, Singapore, pp. 751–760. https://doi.org/10.1007/978-981-10-3518-0_65.

REFERENCES

- Kalinouskaya, I. (2021). Theoretical aspects of recruitment using artificial intelligence technologies [Teoreticheskie aspekty podbora kadrov s primeneniem tekhnologij iskusstvennogo intellekta]. *Pravo. Ekonomika. Psihologiya = Right. Economy. Psychology*, № 1 (21), pp. 48–64 (In Russian).
- Kalinouskaya, I. (2022). Analysis of the competencies presented on the labor market, extracted from digital sources using artificial intelligence [Analiz predstavlenykh na rynke truda kompetencij, izvlechennykh iz cifrovyykh istochnikov s pomoshch'yu iskusstvennogo intellekta]. *Ekonomika i obshchestvo: mezhdunarodnyj nauchno-prakticheskij zhurnal = Economics and Society: an international scientific and practical journal*, № 04 (22), pp. 29–42 (In Russian).
- Roshchin, S. Yu. and Solntsev, S. A. (2019). The index of tension in the Russian labor market [Indeks napryazhennosti na rossijskom rynke truda]. *Monitoring ekonomicheskoy situacii v Rossii: tendencii i vyzovy social'no-ekonomicheskogo razvitiya = Monitoring the Economic situation in Russia: Trends and Challenges of socio-economic development*, № 1 (84), pp. 15–18 (in Russian).
- Autor, D. (2015). Why are there still so many jobs? The history and future of workplace automation. *The Journal of Economic Perspectives*, № 29 (3), pp. 3–30.
- Autor, D., Reynolds, E., Chin, A., Fikri, K., Fleming, W. B., Katz, L. F., Kearney, M. S., Murray, S. C., Siegel, D. and Veuger, S. (2020). The Work of the Future: Building Better Jobs in an Age of Intelligent Machines. *MIT Work of the Future*, [Online], URL: <https://workofthefuture.mit.edu/wp-content/uploads/2021/01/2020-Final-Report4.pdf>, [Accessed: 23.05.2024].
- Karoly, L. A., Leukhina, O., Iyengar, S. and Dew-Becker, I. (2021). The Digital Workforce: Developing Skills for an AI-Driven Economy. *RAND Corporation*, [Online], URL: https://www.rand.org/pubs/research_reports/RRA1533-1.html, [Accessed: 24.05.2024].
- Mamertino, M. and Sinclair, T. M. (2016). Online Job Search and Migration Intentions Across EU Member States. *Institute for International Economic Policy Working Paper Series, IIEP-WP-2016-4*, [Online], URL: <https://www2.gwu.edu/~iiep/assets/docs/papers/2016WP/MamertinoSinclairIIEPWP2016-4.pdf>, [Accessed: 24.05.2024].
- Boselli, R., Cesarini, M., Mercorio, F. and Mezzanzanica, M. (2018). Classifying online Job Advertisements through Machine Learning. *Future Generation Computer Systems*, vol. 86, pp. 319–328. <https://doi.org/10.1016/j.future.2018.03.035>.
- Brynjolfsson, E., Mitchell, T. and Rock, D. (2021). What Can Machines Learn, and What Does It Mean for Occupations and the Economy? *AEA Papers and Proceedings*, vol. 111, pp. 43–47. DOI: 10.1257/pandp.20211019
- Jain, N., Shanthakumar, S. K., Sharma, A., Arora, A. and Mutharaju, R. (2020). Analyzing and Visualizing the Skill Demand in Indian Job Market using Web Crawling and Machine Learning. In *2020 International Conference on Intelligent Engineering and Management (ICIEM)*, pp. 303–308. <https://doi.org/10.1109/ICIEM48762.2020.9160247>.
- Mezzanzanica, M. and Mercorio, F. (2018). Big Data Enables Labor Market Intelligence. *Encyclopedia of Big Data Technologies*, pp. 1–11. https://dx.doi.org/10.1007/978-3-319-63962-8_276-1.
- Stephany, F., Braesemann, F. and Graham, M. (2020). Coding together – coding alone: the role of trust in collaborative programming. *Information, Communication & Society*, vol. 24, pp. 1944–1941. <https://doi.org/10.1080/1369118X.2020.1749699>.
- Vankevich, A. and Kalinouskaya, I. (2020). Ensuring sustainable growth based on the artificial intelligence analysis and forecast of in-demand skills. *First Conference on Sustainable Development: Industrial Future of Territories (IFT 2020)*, Vol. 208, <https://doi.org/10.1051/e3sconf/202020803060>.
- Vankevich, A. and Kalinouskaya, I. (2021). Better understanding of the labour market using Big Data. *Ekonomia i pravo. Economics and law*, vol. 20, № 3, pp. 677–692.

Verma, R. and Raghavan, S. S. (2017). Extrapolating Trends in Demand and Supply of IT Professionals: *The Indian Scenario*. In *International Conference on Research into Design*, Springer, Singapore, pp. 751–760. https://doi.org/10.1007/978-981-10-3518-0_65.

Информация об авторах

Information about the authors

Калиновская Ирина Николаевна

Кандидат технических наук, доцент кафедры «Экономика и электронный бизнес», Витебский государственный технологический университет, Республика Беларусь.
E-mail: i-kalinovskaya@yandex.by

Iryna N. Kalinouskaya

Candidate of Sciences (in Engineering), Associate Professor of the Department "Economics and E-business", Vitebsk State Technological University, Republic of Belarus.
E-mail: i-kalinovskaya@yandex.by