
ЦИФРОВАЯ СОЦИОЛОГИЯ: НАПРАВЛЕНИЯ ИССЛЕДОВАНИЙ

Опыт применения больших языковых моделей для анализа социологических данных, полученных в результате интервью о восприятии студентами предпринимательской деятельности

УДК 316.354.2

DOI 10.26425/2658-347X-2024-7-3-4-14

Получено 29.05.2024

Доработано после рецензирования 19.07.2024

Принято 26.07.2024

Ашихмин Евгений Георгиевич

Аспирант

ORCID: 0000-0001-9193-4535

E-mail: e.ashikhmin@icloud.com

Селеткова Гюзель Ильясовна

Ст. преп. каф. социологии и политологии

ORCID: 0000-0003-3402-3473

E-mail: guzal.ka@mail.ru

Левченко Валерий Витальевич

Д-р психол. наук, зав. каф. социологии и политологии

ORCID: 0000-0002-7627-9162

E-mail: levv66@mail.ru

Пермский национальный исследовательский политехнический университет, г. Пермь, Россия

АННОТАЦИЯ

В современном обществе наблюдается цифровая трансформация различных сфер, связанная с развитием искусственного интеллекта и больших данных. Внедрение больших языковых моделей (англ. large language model, далее – LLM) в научные исследования открывает новые возможности, но и ставит ряд вопросов, в связи с чем актуальным становится изучение особенностей их применения для качественного анализа данных в социологии. Цель – изучить, как большие языковые модели могут влиять на методологию и практику социологических исследований, выявить преимущества и недостатки их применения. Авторы опираются на использование большой языковой модели Calude-3 для качественного анализа эмпирических данных социологического исследования восприятия студентами предпринимательской

деятельности. Раскрыты возможности LLM в анализе качественных данных: оценка тональности, построение логических выводов, классификация, кластеризация и формирование типологий. Показаны преимущества использования LLM: увеличение скорости обработки данных, экономия времени и ресурсов. Применение LLM становится инструментом для оптимизации исследовательского процесса в социологии, позволяя углубить анализ качественных данных, но имеет и ряд ограничений: социальная и политическая предвзятость, трудности с галлюцинациями. Необходимы повышение прозрачности моделей, улучшение их интерпретируемости и объяснимости и уменьшение их социальной, политической предвзятости, а также этическое и юридическое регулирование использования моделей LLM.

Ключевые слова

Большие языковые модели, LLM, цифровые инструменты, качественный анализ данных, анализ интервью, методы социологических исследований, цифровая трансформация, оценка тональности, алгоритм предобработки данных, кластеризация

Для цитирования

Ашихмин Е.Г., Левченко В.В., Селеткова Г.И. Опыт применения больших языковых моделей для анализа социологических данных, полученных в результате интервью о восприятии студентами предпринимательской деятельности // Цифровая социология. 2024. Т. 7. № 3. С. 4–14.

© Ашихмин Е.Г., Левченко В.В., Селеткова Г.И., 2024.

Статья доступна по лицензии Creative Commons «Attribution» («Атрибуция») 4.0. всемирная (<http://creativecommons.org/licenses/by/4.0/>).



DIGITAL SOCIOLOGY: RESEARCH DIRECTIONS

Experience in applying large language models to analyse sociological data obtained as a result of interviews on students' perception of entrepreneurial activity

Received 29.05.2024

Revised 19.07.2024

Accepted 26.07.2024

Evgenii G. Ashikhmin

Postgraduate Student

ORCID: 0000-0001-9193-4535

E-mail: e.ashikhmin@icloud.com

Valery V. Levchenko

Dr. Sci. (Psy.), Head of the Sociology and Political Science Department

ORCID: 0000-0002-7627-9162

E-mail: levv66@mail.ru

Perm National Research Polytechnic University, Perm, Russia

Gyuzel' I. Seletkova

Senior Lecturer at the Sociology and Political Science Department

ORCID: 0000-0003-3402-3473

E-mail: guzal.ka@mail.ru

ABSTRACT

Modern society is experiencing a digital transformation of various spheres associated with the development of artificial intelligence and big data. The introduction of large language models (hereinafter referred to as LLM) into scientific research opens new opportunities, but also raises a number of questions, which makes it relevant to study the peculiarities of their application for qualitative data analysis in sociology. The purpose of this article is to explore how LLM can influence the methodology and practice of sociological research, and to identify the advantages and disadvantages of their application. The authors rely on the use of the Calude-3 LLM to qualitatively analyse empirical data from a sociological study of students' perception of entrepreneurship.

The possibilities of LLM in the analysis of qualitative data are revealed: analysis of sentiment, construction of logical conclusions, classification, clustering, and formation of typologies. The advantages of using LLM are shown: increased data processing speed, saving time and resources. The application of LLM becomes a tool to optimise the research process in sociology, allowing to deepen the analysis of qualitative data, but it also has a number of limitations: social and political bias, difficulties with hallucinations. It is necessary to increase the transparency of models, improve their interpretability and explainability and reduce their social and political bias as well as ethical and legal regulation of the use of LLM models.

Keywords

Large language models, LLM, digital tools, qualitative data analysis, interview analysis, sociological research methods, digital transformation, sentiment analysis, data preprocessing algorithm, clustering

For citation

Ashikhmin E.G., Levchenko V.V., Seletkova G.I. (2024) Experience in applying large language models to analyse sociological data obtained as a result of interviews on students' perception of entrepreneurial activity. *Digital sociology*. Vol. 7, no 3, pp. 4–14. DOI: 10.26425/2658-347X-2024-7-3-4-14

© Ashikhmin E.G., Levchenko V.V., Seletkova G.I., 2024.

This is an open access article under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).



ВВЕДЕНИЕ / INTRODUCTION

В современном обществе можно наблюдать примеры цифровой трансформации, связанной с развитием искусственного интеллекта, машинного обучения, с ростом объема данных, собираемых и хранимых в различных областях. Происходящие изменения приводят к изменению научных и прикладных исследований в разных науках, в том числе и в социологии. Вступление в эпоху больших данных (англ. Big Data) позволило социологам обогатить свои методологические инструменты, в том числе и через применение больших языковых моделей (англ. large language model, далее – LLM). Языковые модели, основанные на архитектуре generative pre-trained transformer (англ. генеративный предобученный трансформер, далее – GPT), направлены на решение задач, связанных с обработкой естественного языка, генерированием и обобщением текста, с анализом настроений, дискурса, извлечением информации из различных источников данных.

В настоящее время большие языковые модели нашли применение в самых разных областях: чат-боты и виртуальные помощники; создание различных текстов: посты для социальных сетей, статьи, электронные письма; генерация программного кода, машинный перевод; семантический поиск [Кублик, Сабу, 2023]. LLM начинают применять как один из автоматизированных инструментов для анализа текста [Markowitz, 2024], тональности [Krugmann, Hartmann, 2024], в маркетинговых коммуникациях компаний [Аржанова, Писклакова, 2023], при анализе веб-аналитики [Алексеев, 2024]. Также рассматриваются перспективы и опыт применения LLM в сфере образования [Раицкая, Ламбовская, 2024], управления персоналом [Старовойтов, 2024], политики и в социальной сфере [Джибилова, Побываев, 2024] и др.

Учитывая возможности больших языковых моделей и тот факт, что они становятся все более доступными, представляется актуальным изучение особенностей применения больших языковых моделей как инструмента качественного анализа в социологических исследованиях. Цель данной статьи – изучить, как большие языковые модели могут влиять на методологию и практику социологических исследований, выявить преимущества и недостатки их применения в данных исследованиях.

ОБЗОР СУЩЕСТВУЮЩИХ РЕШЕНИЙ / REVIEW OF EXISTING SOLUTIONS

Для анализа качественных данных социологи используют два основных подхода: ручное

кодирование и программные средства. Преимущества ручного кодирования в том, что исследователь непосредственно погружается в данные, может уловить тонкие смысловые нюансы. Однако в целом применение в социологическом исследовании качественной методологии, предполагающей сбор информации о социальных явлениях и процессах с помощью различных методов, например интервью, сталкивается с рядом ограничений: они трудоемки, требуют больших временных затрат, так как необходимо производить ручное кодирование, интерпретацию. При этом для преодоления проблем субъективности, непоследовательности, влияния контекста, ситуативности, исследователя при кодировании и интерпретации проводится процедура триангуляции, то есть перекрестная интерпретация текста тремя исследователями, работающими в одной команде [Turner, Reynolds, 2012].

Программные средства (например, NVivo, MAXQDA, Atlas.ti, QDA Miner, HyperRESEARCH) предлагают инструменты управления и организации данных в процессе анализа данных. Функция программных средств заключается в помощи в этом процессе. Они не заменяют исследователя в нем, поэтому контроль за процессом анализа всегда должен оставаться у исследователя, так как программа не может самостоятельно анализировать качественные данные, но может помочь ученому при этом.

Развитие программных средств, применяемых для анализа качественных данных, сопровождается развитием использования языка программирования Python с различными библиотеками для обработки данных и машинного обучения (scikit-learn, NLTK, spaCy и др.). Применение этих библиотек позволяет реализовать практически любые задачи анализа текстов: от простого поиска ключевых слов до сложных алгоритмов машинного обучения для кластеризации и классификации. Однако использование Python требует достаточно высокой квалификации программиста и значительных временных затрат на разработку кода для решения исследовательских задач. Кроме того, часто требуются предварительная структуризация и формализация текстовых данных. Таким образом, несмотря на широчайшие возможности, применение Python для анализа качественных данных может быть достаточно сложным и трудоемким для социолога-исследователя.

К использованию программных средств можно отнести и большие языковые модели, которые способны предложить ряд преимуществ по сравнению с традиционным качественным анализом данных. При этом остаются открытыми вопросы,

связанные с этикой [Коваль, 2023; Летов, 2024; Середкина, Мезин, 2023] и плагиатом [Dehouche, 2021]. Также исследователи отмечают, что результаты работы LLM, в частности ChatGPT, обязательно должны проверяться профессионалом, в связи с чем актуализируются вопросы, сопряженные с влиянием больших языковых моделей на процесс принятия решения исследователем. Ученые отмечают, что доверие или зависимость от LLM, ментальная модель пользователя и характеристики обработки информации являются значимыми аспектами, влияющими на процесс принятия решений с помощью LLM [Eigner, Händler, 2024].

Большая языковая модель – модель, состоящая из нейронной сети со множеством параметров (обычно миллиарды весовых коэффициентов и более), обученной на большом количестве неразмеченного текста с использованием обучения без учителя [Manning, 2022].

Сегодня наибольшей популярностью пользуются следующие большие языковые модели (LLM): GPT-3 и GPT-4 (разработчик – OpenAI)¹, Gemini (разработчик – Google)², Claude-3 (разработчик – Anthropic)³. В России свои модели представили «Яндекс» (YandexGPT 2)⁴ и «Сбер» (GigaChat)⁵. В данной статье раскрывается пример использования большой языковой модели Calude-3 (разработчик – Anthropic) для качественного анализа данных.

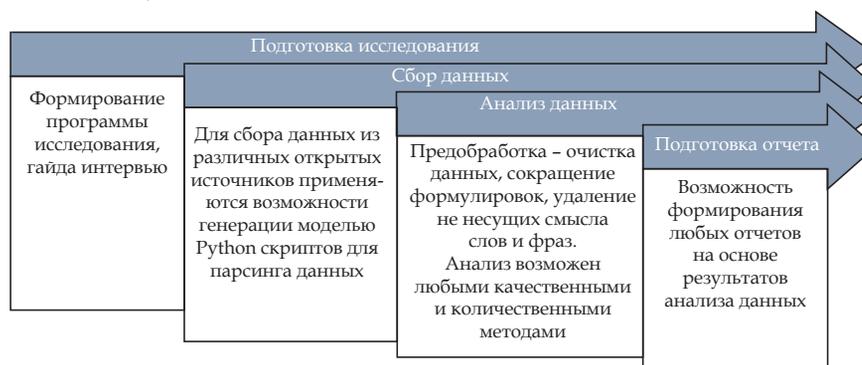
Опираясь на анализ общих стратегий социологического исследования, описанных в книге В.А. Ядова [Ядов, 2007], особенностей качественной

методологии, представленных в работе В.В. Семеновой [Семенова, 1998], в процессе качественного исследования в социологии выделяют несколько этапов: подготовку исследования, сбор информации, анализ данных и подготовку отчета (рис. 1).

На этапах подготовки исследования, анализа данных и подготовки отчета применяются большие языковые модели. При подготовке исследования они могут использоваться для формирования программы исследования и гайда интервью. Традиционно данные, полученные с помощью неформализованного интервью, анализируются с помощью качественных методов. Качественные методы в социологическом исследовании направлены на изучение особенностей индивидуальных социальных практик, через исследование которых можно изучать и более массовые социальные проблемы, касающиеся социальных общностей.

На этапе сбора данных, а именно для интервью, большие языковые модели не применяются, но для сбора данных из различных открытых источников используются возможности генерации моделью Python скриптов для парсинга данных. В ходе интервью мы получили эмпирические неструктурированные свидетельства (текстовые записи интервью), которые в дальнейшей работе требовали их структурирования.

Анализ качественных социологических данных требует применения разнообразных аналитических и интерпретативных процедур, начиная от описания и комментирования текстов до их кодировки и категоризации. Остановимся более подробно на этапе анализа данных. Под качественными социологическими данными понимаются разнообразные тексты – записи интервью, фокус-групп, открытые ответы в анкетах и т.д., которые содержат ценную информацию о мнениях, установках, ценностях исследуемых представителей социальных групп. Однако



Составлено авторами по материалам исследования / Compiled by the authors on the materials of the study

Рис. 1. Этапы качественного исследования

Fig. 1. Stages of qualitative research

в силу своей неструктурированности их анализ традиционно требовал значительных временных и трудовых затрат.

При использовании больших языковых моделей перед анализом текста была проведена его предобработка, то есть очистка данных, сокращение формулировок, удаление не несущих смысла слов и фраз. Стоит отметить, что предобработка приводит к тому, что теряется контекст. При ее применении необходимо помнить о цели исследования, нужно понимать, важно ли для нас знать, когда интервьюируемый затруднялся с ответом, как быстро или долго формулировал свою мысль, был ли у него готовый ответ, смеялся ли он.

При интерпретации данных важными моментами выступают фокус исследователя, а также многозначность и контекст, вопросы кодирования и сравнения выделенных фрагментов. С какой точки зрения интерпретируются полученные данные в больших языковых моделях? Исследователи отмечают проблему политической предвзятости [Durmus, Nguyen, Liao, Schiefer, Askell, Bakhtin et al., 2024].

Применение больших языковых моделей позволяет частично автоматизировать процесс анализа таких данных. С помощью этих моделей возможно решить множество задачи обработки качественных социологических данных: выявлять шаблоны, темы, тенденции в текстовых данных, автоматически кодируя и классифицируя отрывки текста на основе их семантического контекста, а также выполнять классификацию и кластеризацию текстов, анализ тональности, выявление ключевых тем и сущностей, построение логических умозаключений и т.д.

Таблица 1. Операционализация категории «отношение к предпринимателю»

Table 1. Operationalisation of the category "attitude to the entrepreneur"

Вопрос	Показатель
Как вы думаете, какие функции выполняют предприниматели, что делают?	Осведомленность о предпринимательской деятельности
Как вы думаете, какие в обществе есть социальные и культурные ожидания от предпринимательской деятельности?	Стереотипы о предпринимательской деятельности
Как вы думаете, выбор предпринимательской деятельности будет примером успешной карьеры?	Оценка предпринимательской деятельности как хорошей карьеры
Какие предприниматели добиваются успеха?	Оценка необходимых качеств для достижения успеха
Как предприниматель выстраивает отношения с другими предпринимателями?	Восприятие отношений предпринимателя с другими предпринимателями
Как предприниматель выстраивает отношения с клиентами?	Восприятие отношений предпринимателя с клиентами
Как предприниматель выстраивает отношения с подчиненными?	Восприятие отношений предпринимателя с подчиненными
Как вы относитесь к предпринимателю, у которого не получилось развить свой бизнес?	Восприятие неудачного опыта

Составлено авторами по материалам исследования / Compiled by the authors on the materials of the study

При этом типологический метод представляет собой «совокупность оптимальных способов поиска общего в особенном» [Татарова, Бабиц, Бессокирная, Кученкова, 2023, с. 13], который также предполагает выделение существенных и несущественных признаков. Для адекватного описания социальной реальности и выделения общности социальных объектов требуется «одновременное сочетание многих существенных признаков» [Татарова, Бабиц, Бессокирная, Кученкова, 2023, с. 14]. Именно поиск таких сочетаний выступает задачей типологического метода. Возникают вопросы: каким образом большие языковые модели осуществляют данный поиск? Каковы критерии типологии, классификации?

ПРАКТИКА ПРИМЕНЕНИЯ LLM ДЛЯ АНАЛИЗА КАЧЕСТВЕННЫХ ДАННЫХ / PRACTICAL APPLICATION OF LLM FOR ANALYSING QUALITATIVE DATA

Рассмотрим особенности применения LLM в анализе данных, полученных в ходе исследования восприятия студентами предпринимательской деятельности, в рамках которого было проведено 20 интервью со студентами в январе-феврале 2024 г. в Перми.

Все участники интервью – девушки, обучающиеся на третьем курсе бакалавриата по направлению «Социология». Для определения идентификации предпринимателя был сформулирован вопрос: кто для вас предприниматель? Для изучения отношения к предпринимателю и предпринимательской деятельности были сформулированы следующие вопросы (табл. 1).

В ходе интервью были записаны аудиофайлы, далее проведена транскрибация текста и получены транскрипты. Для возможности анализа данных транскриптов с помощью LLM была выполнена предобработка данных. Алгоритм предобработки данных представлен на рис. 2. В зависимости от сути вопроса применялись различные комбинации шагов предобработки.

В ходе анализа полученных данных с помощью больших языковых моделей использовались следующие методы: оценка тональности, построение логических выводов, классификация, кластеризация и формирование типологий.

ОЦЕНКА ТОНАЛЬНОСТИ / SENTIMENT ANALYSIS

В анализе качественных данных важное значение имеет возможность оценки тональности текста. В зависимости от целей анализа различают два основных вида оценки тональности: объектную и безобъектную. Объектная оценка тональности нацелена на определение эмоционального отношения автора текста к конкретным объектам, личностям или событиям, которые упоминаются в нем. Например, при анализе отзывов о продуктах или услугах объектная оценка позволяет выявить, какие именно аспекты вызывают положительные или отрицательные эмоции у потребителей. С другой стороны, безобъектная оценка тональности направлена на выявление общего эмоционального состояния автора, независимо от каких-либо специфических объектов. Она применяется, когда необходимо определить

преобладающий эмоциональный фон текста, например для анализа настроений в социальных сетях или отслеживания общественного мнения.

В последние годы для задач оценки тональности текста активно применяются большие языковые модели, такие как BERT, GPT и др. Ниже приведен пример анализа тональности для определения настроения опрошенных студентов (на основе сводного транскрипта ответов интервьюируемых студентов-социологов).

Входной промт к LLM: «Представь, что ты – аналитик социологических данных. Перед тобой стоит задача выполнить анализ тональности текстов-ответов на вопрос о настроении респондентов. Тексты, тональность которых нужно оценить, представлены ниже. Выполни оценку тональности текстов, результат выдай в виде таблицы: количество позитивных, нейтральных, негативных ответов. У каждого респондента возможен только один вариант ответа».

Ответ LLM: «Позитивное – 14, нейтральное – 4, негативное – 4».

КЛАССИФИКАЦИЯ И ЛОГИЧЕСКИЕ ВЫВОДЫ / CLASSIFICATION AND LOGICAL INFERENCE

Большие языковые модели могут быть использованы для построения на основе семантики текста логических выводов различной сложности: первого, второго, *n*-го порядков.

К логическим выводам первого порядка относится построение выводов непосредственно из фактов, упомянутых в интервью. Логический



Составлено авторами по материалам исследования / Compiled by the authors on the materials of the study

Рис. 2. Алгоритм предобработки данных

Fig. 2. Data preprocessing algorithm

вывод – это процесс получения заключений на основе исходных посылок с помощью правил логики. Существует два основных вида логических выводов: дедуктивный и индуктивный. Дедуктивный вывод представляет собой процесс получения логически обоснованных заключений из имеющихся посылок. Если посылки истинны, то и заключение, сделанное путем дедукции, также будет истинным. Индуктивный вывод – это переход от частных случаев к общим утверждениям. На основе наблюдений за отдельными случаями делается обобщение, которое считается вероятным, но не абсолютно достоверным.

Участникам интервью задавался вопрос, целью которого было понять идентификацию предпринимателя, узнать, кто для интервьюируемых предприниматель. В процессе анализа к полученным ответам был применен метод оценки формирования логических выводов из текстовых сущностей (см. пример ниже).

Входной промт к LLM: «Представь, что ты – аналитик социологических данных. Твоя задача – выполнить классификацию ответов 20 респондентов на вопрос интервью. Тексты, классификацию которых нужно выполнить, представлены ниже. После того как классификация будет выполнена, реши задачу формирования аналитической сводки с применением методов формирования логических выводов по полученным результатам классификации. Аналитическая сводка должна быть написана простым и понятным языком».

Ответ LLM: «Создание собственного бизнеса – 13, получение прибыли – 8, предпринимательская деятельность – 4, принятие рисков – 3, ответственность – 3, целеустремленность – 2, удовлетворение потребностей – 2, финансовая грамотность – 1, трудолюбие – 1, предпринимательское мышление / стиль жизни – 1».

В целом респонденты определяют предпринимателя как человека, который создает собственный бизнес и получает от него прибыль. Акцент делается на факторах создания бизнеса, получении дохода, принятия рисков и ответственности. Многие респонденты упоминают такие ключевые характеристики предпринимателя, как целеустремленность, финансовая грамотность, стремление к удовлетворению потребностей, трудолюбие. Один опрошенный отмечает предпринимательский стиль мышления и жизни. Таким образом, представления респондентов об образе предпринимателя соответствуют общепринятым определениям и включают как объективные критерии (создание бизнеса, получение прибыли), так и субъективные качества личности.

КЛАСТЕРИЗАЦИЯ / CLUSTERISATION

Кластеризация выполнялась по вопросу «кто для вас предприниматель?» в двух вариантах. При этом критерии кластеризации выбирались большей языковой моделью в автоматическом режиме.

Первый вариант кластеров строился на основе акцентирования ключевых аспектах предпринимательства, таких как создание бизнеса, своего предприятия; получение прибыли; предпринимательская деятельность и необходимые личные характеристики.

Второй вариант кластеров строился на основе следующих критериев: объективные аспекты (создание бизнеса, получение прибыли) и субъективные (личностные качества, мотивация).

Входной промт к LLM: «Представь, что ты – социолог-аналитик. Ты выполняешь анализ ответов 20 респондентов на вопросы интервью об их отношении к предпринимательской деятельности. Задача – предложить два варианта кластеризации (далее – типологизации) респондентов. Все необходимые для кластеризации (типологизации) критерии определи сам. Для формирования кластеров респондентов используй набор ответов на вопрос № 6. Ответ выдай в виде текста».

Ответ LLM:

«Предлагаю два варианта кластеризации.

Вариант 1. По акценту на ключевых аспектах предпринимательства.

Критерии: создание бизнеса, получение прибыли, личностные качества, предпринимательская деятельность. Типы: сфокусированные на создании бизнеса, сфокусированные на получении прибыли, сфокусированные на личностных качествах, сфокусированные на предпринимательской деятельности.

Вариант 2. По упору на объективные или субъективные аспекты.

Критерии: объективные аспекты (создание бизнеса, получение прибыли), субъективные аспекты (личностные качества, мотивация). Типы: упор на объективные аспекты, упор на субъективные аспекты, сбалансированное определение».

С опорой на феноменологический метод мы проанализировали, как студенты воспринимают предпринимателей. В данном контексте предпринимательство и предприниматель были проинтерпретированы с помощью структуры субъективных значений, выработанных самими респондентами, и тех смыслов, которые они вкладывают в функции предпринимателей [Шюц, 2004]. Из проведенной кластеризации мы видим, что студенты различным образом идентифицируют предпринимателя, давая ему разные определения. Одни

близки к пониманию предпринимателя как бизнесмена или того, кто организует свое предприятие ради получения прибыли. Другие склоняются к шумпетерианскому пониманию предпринимателя и связывают его деятельность с новшествами, инновациями. Третьи близки к психологической интерпретации предпринимателя: они акцентируют внимание на его личностных качествах. При этом при определении функций предпринимателя респонденты заостряют внимание на его инновационных, организационно-управленческих функциях.

Интересно, что среди необходимых качеств предпринимателей для достижения успеха профессиональные качества (англ. *hard skills*) практически не упоминаются – к ним из всех вариантов респондентов мы посчитали нужным отнести только варианты, связанные с образованием предпринимателя. При этом часто респонденты высказывают мнение о необходимости *soft-skills* (англ. *гибкие навыки*): целеустремленности и настойчивости; креативности и инновационности; трудолюбия и умения управлять; ответственности. Одним из распространенных мнений среди опрошенных был выбор среди нужных качеств предпринимателей для достижения успеха целеустремленности и настойчивости. Студенты подчеркивают важность упорства, готовности преодолевать трудности, стремления к целям для достижения успеха, также выделяют нестандартный подход к делу, способность к риску, творческий потенциал и важность умения принимать решения, управлять, трудолюбия.

РЕЗУЛЬТАТЫ АНАЛИЗА КАЧЕСТВЕННЫХ ДАННЫХ ПРИ ПОМОЩИ LLM / RESULTS OF QUALITATIVE DATA ANALYSIS WITH LLM

В результате анализа транскриптов интервью при помощи LLM был получен аналитический отчет, соответствующий общепринятым в социологической среде требованиям, часть выводов из которого представлена ниже. Следует отметить, что применение LLM для анализа транскриптов интервью позволило получить следующие преимущества (табл. 2).

Таблица 2. Сравнение трудозатрат при анализе интервью с LLM и без LLM

Table 2. Comparison of labour costs in the analysis of interviews with and without LLM

Преимущество	Без LLM, ч	С LLM, ч
Погружение в тему исследования	2	0

Окончание табл. 2

Преимущество	Без LLM, ч	С LLM, ч
Сокращение трудозатрат на предобработку данных	16	1
Сокращение трудозатрат на анализ данных	8	1
Сокращение трудозатрат на формирование отчета	8	0,5
Итого трудозатрат	34	2,5

Составлено авторами по материалам исследования /
Compiled by the authors on the materials of the study

ЧАСТЬ ВЫВОДОВ ИЗ АНАЛИТИЧЕСКОГО ОТЧЕТА / PART OF THE FINDINGS FROM THE ANALYTICAL REPORT

1. Респонденты демонстрируют хорошее понимание сущности предпринимательской деятельности, ее функций и качеств, необходимых для достижения успеха. Они связывают предпринимательство с созданием собственного бизнеса, получением прибыли, готовностью к риску, а также с такими личностными чертами, как целеустремленность, трудолюбие, ответственность, креативность.

2. В обществе присутствуют как позитивные, так и негативные стереотипы относительно предпринимателей. С одной стороны, они воспринимаются как успешные, обеспеченные люди, ведущие роскошный образ жизни. С другой – от них ожидают социальная ответственность, создание качественных продуктов и услуг на благо общества.

3. Отношение к предпринимательству как к успешной карьере является неоднозначным. Часть респондентов безоговорочно считает это видом деятельности – примером успеха, другие указывают на зависимость успеха от ряда факторов, третьи сомневаются из-за высоких рисков и непредсказуемости.

4. Отношения предпринимателей с другими субъектами рынка (конкурентами, клиентами, подчиненными) рассматриваются преимущественно через призму рациональности и прагматизма. Конкуренция присутствует, но не доминирует. Основной акцент делается на партнерство, взаимовыгодное сотрудничество, на удовлетворение запросов клиентов и выстраивание с ними доверительных отношений.

5. В управлении подчиненными преобладают директивный стиль с высоким фокусом на задачи и относительно низким – на отношения, а также стиль наставничества, сочетающий высокие ориентации и на задачи, и на взаимоотношения.

6. Неудачный предпринимательский опыт в целом встречает нейтральное или даже положительное

отношение. Он рассматривается как неотъемлемая часть предпринимательства, возможность учиться на ошибках. Резко негативных оценок или пренебрежения к потерпевшим неудачу предпринимателям не наблюдается.

В целом результаты демонстрируют достаточно реалистичное и многогранное восприятие предпринимательства респондентами с осознанием его достоинств, рисков и особенностей взаимодействия с различными субъектами рынка.

ЗАКЛЮЧЕНИЕ / CONCLUSION

Использование моделей LLM дает ряд преимуществ: увеличение скорости обработки данных, автоматизация аналитических процессов, экономия времени и ресурсов, повышение точности анализа, обнаружение скрытых закономерностей и аномалий, поддержка принятия решений. Большие языковые модели выполняют анализ данных с гораздо большей скоростью, чем человек. Это позволяет существенно ускорить процесс исследования, освободив время и ресурсы для более глубокого анализа результатов. Использование больших языковых моделей может служить важным инструментом для принятия решений, открыть новые возможности для исследования. Технология способна обработать разнообразные типы данных и выявить новые аспекты изучаемых явлений, что способствует расширению границ знаний в социологии. Применение больших языковых моделей становится хорошим инструментом для социологов, позволяя углубить анализ и оптимизировать исследовательский процесс.

Итак, применение больших языковых моделей для анализа социологических данных предоставляет определенные преимущества и возможности, но также имеет и ряд ограничений. С одной стороны, большие языковые модели обучены на огромных объемах текстовых данных, в результате чего формируется широкий объем знаний по различным темам, существует возможность посмотреть на проблему с разных точек зрения. Однако при этом выходные данные LLM являются функцией данных, используемых в процессе ее обучения, и искажения, имеющиеся в данных обучения, вероятнее всего, сохранятся и в модели LLM, которая может отражать определенные предубеждения. Так, проявление социальной предвзятости в результатах LLM является важной проблемой при ее использовании. Социальная предвзятость может привести к тому, что модель LLM будет отражать мнение большинства, что, вероятно, повлечет систематическое игнорирование меньшинства, в результате чего

результаты моделей LLM могут быть неосознанно, незаметно искажены и усиливать существующие стереотипы и предубеждения [Motoki, Neto, Rodrigues, 2024]. Поэтому вопросы социальной предвзятости нужно обязательно учитывать при анализе полученных результатов.

Выделяют два типа социальной предвзятости: первый тип – это социальная предвзятость при отборе и в отчетности, второй – смещение групповой атрибуции или расширение наблюдаемой неявной предвзятости [Linegar, Kocielnik, Alvarez, 2023]. И то и другое связано со склонностью к стереотипизации групп меньшинств, тем самым сохраняются предубеждения против этих групп [Abid, Farooqi, Zou, 2021].

Также отмечают проблему чувствительности LLM к вариациям семантически нерелевантных входных данных [Linegar, Kocielnik, Alvarez, 2023], что затрудняет распознавание последствий, а также сложности, связанные с галлюцинациями (когда модели генерируют ложные данные из-за высокой вероятности, приписываемой не соответствующим действительности утверждениям), которые часто основаны на обучающих данных модели, и поэтому они могут сильно зависеть от контекста и их трудно предсказать, учитывая их вероятностный характер.

При принятии решений на основе результатов больших языковых моделей актуализируются вопросы справедливости. Существуют различные определения справедливости (индивидуальная, групповая, гибридный подход) в использовании больших языковых моделей, что затрудняет ее достижение на практике [Linegar, Kocielnik, Alvarez, 2023]. При применении к анализу данных социологических исследований важно учитывать аспекты справедливости, поскольку социальная предвзятость, присущая моделям LLM, может иметь негативные последствия для реализации справедливости на практике.

Использование моделей LLM в качественном анализе данных социологических исследований связано с этическими соображениями. Применение этих моделей часто требует доверия к алгоритмам типа «черного ящика». Это означает, что любые искажения в данных обучения могут быть непреднамеренно внесены в исследование, искажая его результаты. Здесь важно понимать и помнить, что большие языковые модели – это не окончательный источник истины, а скорее, представление данных, на которых они обучались. Таким образом, необходимы повышение прозрачности моделей, улучшение их интерпретируемости и объяснимости и уменьшение их социальной, политической предвзятости, а также юридическое регулирование использования моделей LLM.

СПИСОК ЛИТЕРАТУРЫ

- Алексеев Д.В.* Использование генеративного ИИ при анализе данных веб-аналитики. *Universum: технические науки.* 2024;2(119):4–9. <http://doi.org/10.32743/UniTech.2024.119.2.16815>
- Аржанова К.А., Писклакова Л.Д.* Особенности использования искусственного интеллекта в маркетинговых коммуникациях компаний. *Цифровая социология.* 2023;6(4):4–12. <https://doi.org/10.26425/2658-347X-2023-6-4-4-12>
- Джибилова Е.Г., Побываев Н.С.* Анализ российского и зарубежного опыта применения ChatGPT и искусственного интеллекта в политике и социальной сфере. *Социально-гуманитарные знания.* 2024;1:64–69.
- Коваль Е.А.* Большие данные и большие вызовы социальному воображаемому. *Социальные нормы и практики.* 2023;3:54–63. <http://doi.org/10.24412/2713-1033-2023-3-54-63>
- Кублик С., Сабу Ш.* GPT-3. Руководство по использованию API Open AI. Перев. с англ. В.С. Яценкова. М.: ДМК Пресс; 2023. 172 с.
- Летов О.В.* Этические аспекты в области разработки искусственного интеллекта. *Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 3. Философия.* 2024;1:34–44. <http://doi.org/10.31249/rphil/2024.01.03>
- Раицкая Л.К., Ламбовска М.Р.* Перспективы применения ChatGPT для высшего образования: обзор международных исследований. *Интеграция образования.* 2024;1(28):10–21. <https://doi.org/10.15507/1991-9468.114.028.202401.010-021>
- Семенова В.В.* Качественные методы: введение в гуманистическую социологию: учебное пособие для студентов вузов. М.: Добросвет; 1998. 292 с.
- Середкина Е.В., Мезин Е.А.* Как может повлиять ChatGPT на культуру диалога и образование? *Научковедческие исследования.* 2023;3:74–79. <https://doi.org/10.31249/scis/2023.03.04>
- Старовойтов А.Г.* Исследование возможностей чат-бота ChatGPT и оценка перспектив их практического применения специалистами службы управления персоналом. *Journal of Monetary Economics and Management.* 2024;1:22–29. <https://doi.org/10.26118/2782-4586.2024.75.98.003>
- Татарова Г.Г., Бабич Н.С., Бессокирная Г.П., Кученкова А.В.* Типологический анализ в социологии как диагностическая процедура: монография. М.: Федеральный научно-исследовательский социологический центр Российской академии наук; 2023. 358 с.
- Шюц А.* Обыденная и научная интерпретация человеческого действия. В кн.: *Избранное. Мир, светящийся смыслом.* Пер. с нем. и англ. В.Г. Николаева. М.: РОССПЭН; 2004. С. 7–50.
- Ядов В.А.* Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности. 3^е изд., перераб. и доп. М.: Омега-Л; 2007. 567 с.
- Abid A., Farooqi M., Zou J.* Large language models associate Muslims with violence. *Nature Machine Intelligence.* 2021;3:461–463. <http://dx.doi.org/10.1038/s42256-021-00359-2>
- Dehouche N.* Plagiarism in the age of massive generative pre-trained transformers (GPT-3). *Ethics in Science and Environmental Politics.* 2021;21:17–23. <http://dx.doi.org/10.3354/esepp00195>
- Durmus E., Nguyen K., Liao Th.I., Schiefer N., Askill A., Bakhtin A. et al.* Towards measuring the representation of subjective global opinions in language models. *ArXiv.* 2024. <https://doi.org/10.48550/arXiv.2306.16388>
- Eigner E., Händler Th.* Determinants of LLM-assisted decision-making. *ArXiv.* 2024. <https://doi.org/10.48550/arXiv.2402.17385>
- Krugmann J.O., Hartmann J.* Sentiment analysis in the age of generative AI. *Customer Needs and Solutions.* 2024;11. <http://dx.doi.org/10.1007/s40547-024-00143-4>
- Linegar M., Kocielnik R., Alvarez R.M.* Large language models and political science. *Frontiers in Political Science.* 2023;5. <http://dx.doi.org/10.3389/fpos.2023.1257092>
- Manning Ch.D.* Human language understanding & reasoning. *Daedalus.* 2022;2(151):127–138. http://dx.doi.org/10.1162/daed_a_01905
- Markowitz D.M.* Can generative AI infer thinking style from language? Evaluating the utility of AI as a psychological text analysis tool. *Behavior Research Methods.* 2024;56:3548–3559. <http://dx.doi.org/10.3758/s13428-024-02344-0>
- Motoki F., Neto V.P., Rodrigues V.* More human than human: measuring ChatGPT political bias. *Public Choice.* 2024;198:3–23. <http://dx.doi.org/10.1007/s11127-023-01097-2>
- Turner J.C., Reynolds K.J.* Self-categorization theory. In: *Handbook of theories of social psychology. Volume 2.* New York: SAGE Publications; 2012. Pp. 399–417.

REFERENCES

- Abid A., Farooqi M., Zou J.* Large language models associate Muslims with violence. *Nature Machine Intelligence.* 2021;3:461–463. <http://dx.doi.org/10.1038/s42256-021-00359-2>

- Alekseev D.V.* Using generative AL in analyzing web analytics data. *Universum: Technical Sciences*. 2024;2(119):4–9. (In Russian). <http://doi.org/10.32743/UniTech.2024.119.2.16815>
- Arzhanova K.A., Pisklakova L.D.* Features of using artificial intelligence in companies' marketing communications. *Digital Sociology*. 2023;6(4):4–12. (In Russian). <https://doi.org/10.26425/2658-347X-2023-6-4-4-12>
- Dehouche N.* Plagiarism in the age of massive generative pre-trained transformers (GPT-3). *Ethics in Science and Environmental Politics*. 2021;21:17–23. <http://dx.doi.org/10.3354/esep00195>
- Durmus E., Nguyen K., Liao Th.I., Schiefer N., Askill A., Bakhtin A. et al.* Towards measuring the representation of subjective global opinions in language models. *ArXiv*. 2024. <https://doi.org/10.48550/arXiv.2306.16388>
- Dzibilova E.G., Pobyvaev N.S.* Analysis of Russian and foreign experience in the application of ChatGPT and AI in politics and social sphere. *Social and humanitarian knowledge*. 2024;1:64–69. (In Russian).
- Eigner E., Händler Th.* Determinants of LLM-assisted decision-making. *ArXiv*. 2024. <https://doi.org/10.48550/arXiv.2402.17385>
- Koval E.A.* Big data and big challenges to the social imaginary. *Social norms and practices*. 2023;3:54–63. (In Russian). <http://doi.org/10.24412/2713-1033-2023-3-54-63>
- Krugmann J.O., Hartmann J.* Sentiment analysis in the age of generative AI. *Customer Needs and Solutions*. 2024;11. <http://dx.doi.org/10.1007/s40547-024-00143-4>
- Kublik S., Saboo Sh.* GPT-3. The ultimate guide to building NLP products with Open AI API. *Trans. from Eng. V.S. Yatsenkov*. Moscow: DMK Press; 2023. 172 p. (In Russian).
- Letov O.V.* Ethical considerations in artificial intelligence development. *Social sciences and humanities. Domestic and foreign literature. Series 3. Philosophy*. 2024;1:34–44. (In Russian). <http://doi.org/10.31249/rphil/2024.01.03>
- Linegar M., Kocielnik R., Alvarez R.M.* Large language models and political science. *Frontiers in Political Science*. 2023;5. <http://dx.doi.org/10.3389/fpos.2023.1257092>
- Manning Ch.D.* Human language understanding & reasoning. *Daedalus*. 2022;2(151):127–138. http://dx.doi.org/10.1162/daed_a_01905
- Markowitz D.M.* Can generative AI infer thinking style from language? Evaluating the utility of AI as a psychological text analysis tool. *Behavior Research Methods*. 2024;56:3548–3559. <http://dx.doi.org/10.3758/s13428-024-02344-0>
- Motoki F., Neto V.P., Rodrigues V.* More human than human: measuring ChatGPT political bias. *Public Choice*. 2024;198:3–23. <http://dx.doi.org/10.1007/s11127-023-01097-2>
- Raitskaya L.K., Lambovska M.R.* Prospects for ChatGPT application in higher education: a scoping review of international research. *integration of education*. 2024;1(28):10–21. (In Russian). <https://doi.org/10.15507/1991-9468.114.028.202401.010-021>
- Schütz A.* Common-sense and scientific interpretation of human action. In: *Favourites. A world glowing with meaning*. *Trans. from Germ. and Eng. V.G. Nikolaev*. Moscow: ROSSPEN; 2004. Pp. 7–50. (In Russian).
- Semyonova V.V.* Qualitative methods: an introduction to humanistic sociology: textbook for university students. M.: Dobrosvet; 1998. 292 p. (In Russian).
- Seredkina E.V., Mezin E.V.* How can ChatGPT affect the culture of dialogue and education? *Science Studies*. 2023;3:74–79. (In Russian). <https://doi.org/10.31249/scis/2023.03.04>
- Starovoitov A.G.* Research of ChatGPT chatbot capabilities and assessment of prospects for their practical application by HR specialists. *Journal of Monetary Economics and Management*. 2024;1:22–29. (In Russian). <https://doi.org/10.26118/2782-4586.2024.75.98.003>
- Tatarova G.G., Babich N.S., Bessokirnaya G.P., Kuchenkova A.V.* Typological analysis in sociology as a diagnostic procedure: monograph. Moscow: Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences; 2023. 358 p. (In Russian).
- Turner J.C., Reynolds K.J.* Self-categorization theory. In: *Handbook of theories of social psychology*. Volume 2. New York: SAGE Publications; 2012. Pp. 399–417.
- Yadov V.A.* Strategy of sociological research. Description, explanation, understanding of social reality. 3rd ed., revised and enlarged. Moscow: Omega-L; 2007. 567 p. (In Russian).