

Применение ИИ для разработки категорий в качественном исследовании: что возможно в текущем моменте

Ольга Борисовна Савинская
Национальный исследовательский университет "Высшая школа экономики",
Москва, Россия, osavinskaya@hse.ru

Аннотация. Представлен обзор опыта применения искусственного интеллекта для анализа качественных данных, и, в частности, разработке социологических категорий и понятий в качественном исследовании на примере интервью с женщиной, пережившей беременность во время студенчества. Опираясь на логику кодирования для последующей концептуализации данных, предлагается систематизация заданий для искусственного интеллекта (большой языковой модели) для использования его как помощника в анализе данных и аналитической триангуляции.

Ключевые слова: *качественный анализ, понятия и категории, тематическое кодирование, теоретическая чувствительность, Open AI, триангуляция с ИИ*

Статья подготовлена в ходе проведения исследования в рамках Программы фундаментальных исследований Национального исследовательского университета "Высшая школа экономики" (НИУ ВШЭ).

Для цитирования: Савинская О. Б. Применение ИИ для разработки категорий в качественном исследовании: что возможно в текущем моменте // Ойкумена. Регионоведческие исследования. 2025. Т. 19. № 3. С. 63–72.
<https://doi.org/10.63973/1998-6785/2025-3/63-72>

Original article
<https://doi.org/10.63973/1998-6785/2025-3/63-72>

Using AI to Develop Categories in Qualitative Research: What is Possible at the Moment

Olga B. Savinskaya
National Research University Higher School of Economics, Moscow, Russia, osavinskaya@hse.ru

Abstract. An overview of the experience of using artificial intelligence for the analysis of qualitative data, and, in particular, the development of sociological categories and concepts in qualitative research is presented coding the interview with a woman who experienced pregnancy during her student years. Based on the logic of coding for subsequent conceptualization of data, a systematization of tasks for artificial intelligence (a large language model) is proposed for using it as an assistant in data analysis and analytical triangulation.

Key words: *qualitative analysis, concepts and categories, thematic coding, theoretical sensitivity, Open AI, triangulation with AI*

The article was prepared as part of a research project conducted within the framework of the Basic Research Program of the National Research University Higher School of Economics (HSE).

For citation: Savinskaya O. B. Using AI to Develop Categories in Qualitative Research: What is Possible at the Moment // Ojkumena. Regional Researches. 2025. Vol. 19. No. 3. P. 63–72.
<https://doi.org/10.63973/1998-6785/2025-3/63-72>

Введение

В последние годы стремительно набирает обороты искусственный интеллект (ИИ), включаясь в решение многих задач, не только рутинного характера. Мощность искусственного интеллекта увеличивается год за годом, а практика применения расширяется во многих профессиональных сферах. Технологии в образовании и науке всё больше учитывают возможности ИИ, увеличивается спектр задач, которые может решать искусственный интеллект [3; 5]. Для развития социологического анализа и особенно для работы с текстовыми исходными данными, социологи обращаются к появившимся в последние годы хорошо натренированным LLM (large language model, далее "языковая модель"), которые позволяют генерировать осмысленный текст, основываясь на уже имеющихся в их распоряжении опубликованных текстах. Наиболее известная из них – это ChatGPT. Набирает также популярность и китайская нейросеть / LLM DeepSeek. "Языковые модели" всё больше приближаются к естественному языку и ментальности людей, помогая рассказать об их жизни, мышлении, установках. И это становится полезным инструментом для социологов-качественников.

Качественные данные могут быть полезными в разных сферах социального знания, в том числе сугубо прикладного знания. Например, маркетинг-

говые исследования потребительского поведения, когда важно понять логику потребления или увидеть обыденность потребителей в деталях для принятия практического решения при работе в сообществе, когда требуется буквально суммировать мнения информантов, оставаясь в поле обыденного языка. Иногда социологу, применяющему этнографический подход, также важно освоить язык, диалекты, ход мысли и взгляд на мир с точки зрения своих информантов. В этом случае погружение в обыденный язык также оказывается осознанной стратегией поискового исследования.

В академической социологии только этой стратегии часто недостаточно. От обыденных представлений, иногда нелогичных и непоследовательных, учёный переходит к научным понятиям и категориям, чтобы системно описать и объяснить феномен. Социолог в этом случае использует социологические теории и формирующие их понятия и категории. Категориальное мышление – основа качественного анализа в академической социологии: при работе с эмпирическими данными социолог-исследователь применяет теоретические концепции, чтобы эти данные объяснить. А если подходящей теории нет, то пытается её вырастить [14], развивая в себе теоретическую чувствительность [15]. При этом знания о социологических теориях всё равно остаются основой компетенций – новая частная (substantive) теория должна встраиваться в общее поле социологического знания.

Натренированные в том числе, на социологических текстах языковые модели становятся всё более полезными для социологов не только в тематической группировке данных и конденсации смысла, но и в разработке социологических категорий. С 2023 г. "языковые модели" стали включаться в пакеты компьютерной обработки качественных данных. Так, компании, разрабатывающие пакеты ATLAS-ti и MAXQDA заключили соглашение с Open AI по интеграции ChatGPT в анализ данных. Однако возможности "языковых моделей" для социологического анализа текстовых данных в качественном исследовании остаются пока ограниченными. Разработка ясных алгоритмов для взаимодействия с "языковой моделью" и методическая дискуссия вокруг этого – актуальный этап современного развития ИИ и его помощи для решения социологических и социальных вопросов.

Данная статья направлена на обзор опыта привлечения "языковой модели" к анализу качественных данных и разработке социологических категорий на основе данных. Мы опишем возможности и ограничения применения "языковых моделей" в социологическом теоретизировании.

Теория и эмпирика в качественном исследовании: связь через категории

Если мы применяем качественные методы в научном, академическом исследовании, то оставаться в регистре обыденного языка недостаточно. Цель научного исследования – уметь работать в двух регистрах – обыденных представлений и научных концепций, объяснять через теорию человеческую рутину.

Методологи качественного анализа говорят о двух стратегиях работы с данными: движение от теории к данным (theory-driven) и от данных к теории (data-driven) [16, с. 7]. И та, и другая стратегии полезны для решения разного типа задач, но объединяет их одно – социолог, даже если он занимается сугубо эмпирическими исследованиями, должен с одной стороны хорошо владеть социологическими теориями, быть холодным, отстранённым и рациональным, с другой – быть всегда восприимчивым к обыденному языку и подготовленным к изучаемым практикам, применять в поле активно развивающиеся в последние десятилетия принципы включенности, инклюзивности, открытости к "другому" [20, 17]. И наконец, независимо от стратегии анализа данных, уметь рефлексивно описывать свой путь между этими полюсами, эксплицировать свой научный вывод.

Теоретическая модель анализируемого материала в качественной методологии достигается за счёт учёта широкого контекста изучаемого феномена. Это касается не только знаний об окружающей социальной реальности, но и теоретических и методологических знаний. А. Кларке подробно раскрывает это в ситуативном анализе как методологическом подходе: феномен происходит и развивается в определённой ситуации, которая становится условием и средой его существования [11, 12]. Её модель перекликается с парадигмальной

ной моделью её учителя А. Страусса [14]. К. Шармаз добавляет к классической версии обоснованной теории "теоретическое кодирование" как аналитическую процедуру, связывая её с "теоретической выборкой" и "теоретическим насыщением" [10]. В "обоснованной теории" хотя и делается акцент на "кодировании", этот процесс полностью предопределяется аналитическими усилиями исследователя и его "теоретической чувствительностью" в движении к построению своей частной (substantive) теории [4].

Начиная с Глезера и Страусса (1967), авторы, описывающие кодирование, предлагают свои определения видов кодирования, но едины в одном: кодирование проводится не единожды, а несколько раз, итеративно, и у каждого этапа кодирования свои цели, правила работы с кодами и критерии качества. Так, например, К. Шармаз [10] выделяет разные виды кодирования двигаясь от простого, изначального к фокусированному и осевому, и наконец, к теоретическому.

Методология анализа и теоретизирования в биографическом подходе также предполагает многоэтапную аналитическую работу: "анализ ситуации его генезиса, формальные характеристики материала, направление анализа, теоретическое дифференцирование вопроса, определение техник(и) анализа и выбор конкретной его модели, дефиниции единиц анализа, шаги анализа с помощью системы категорий: обобщение, истолкование и структурирование, проверка системы категорий на теории и на эмпирическом материале, интерпретация результатов в зависимости от основного вопроса" [7, с. 121].

Приведённые примеры указывают на то, что за кодированием в качественном исследовании скрывается менее алгоритмизированная работа аналитика, предполагающая знание не только теорий, но и разного рода социальных контекстов. Некоторые темы предполагают погружение в поле и рефлексия своей позиции в полевых взаимодействиях [2]. Обладание этими разными компетенциями приводит к результативному завершению исследования – не только построению детальной схемы кодов, но и теоретическому насыщению понятий и применению их в подробном описании и глубокой интерпретации феномена.

Описанные выше компетенции – характеристики естественно интеллекта, которым обладают исследователи с накопленным жизненным и профессиональным опытом. У искусственного интеллекта таких возможностей нет, но его потенциал становится всё более мощным. Тогда как на данный момент его развития, мы можем применять его в кодировании, разработке категорий и социологической концептуализации?

Какие задания возможно дать "языковой модели": кейс с ChatGPT

Для того, чтобы показать возможности присвоения кодов "языковой моделью" и дальнейшее построение категорий с её помощью, обратимся к примеру. Для иллюстрации взят фрагмент интервью с молодой женщиной, имевшей опыт беременности и родов на первом курсе обучения в университете в 18 лет. Исследование проводилось в рамках поиска новых механизмов политики повышения рождаемости в РФ и было направлено на изучение современного опыта беременности в вузе и рефлексии этого опыта молодой мамой. Целью исследования была реконструкция событий во время беременности, выявление позитивного и негативного опыта, его типологизация и нахождение позитивных стратегий проживания опыта беременности в вузах. Нам важно было не только зафиксировать факты из жизни беременных, но и отношение к ним, выраженное в их рассказе. Анализ нарративов в последующем мог быть полезен для выявления сложностей ранних студенческих браков и разработки институциональных механизмов и языка поддержки.

Для методического разбора в данной статье мы отобрали фрагмент интервью с описанием завершённого события в развивающихся отношениях. На небольшом примере предлагаем разобратить, как возможно взаимодействие с искусственным интеллектом для разработки категорий. Здесь интервьюер просит рассказать о том, как складывались отношения с супругом (отцом ребенка), а информантка же после общей оценки отношений для её раскрытия, обратилась к одному из запомнившихся событий периода беременности.

"Инт: Он так же к тебе продолжал относиться, как к женщине – что женщина должна это всё делать? Или это было ещё больше, потому что ты сейчас мама – ты вообще должна..."

Вера: Нет, грубо говоря, требования ко мне те же и остались: то есть, ты женщина – ты должна то-то, то-то, то-то, то-то, вот... и это не менялось, ммм... При этом даже был один случай... э-э-э... случилось так, что я в марте, э-э-э... получается, у меня была 23–25 неделя беременности – я попала в больницу, вот. Ну, произошёл немножко несчастный случай – я упала с лестницы, вот, и-и-и просто из-за риска возможности прерывания беременности было решено лечь в больницу, просто понаблюдаться хотя бы. Меня привезли после того в больницу на осмотр, потому что, когда я упала, я сказала, что мне нужна скорая, потому что я беременна и как бы, ну мало ли, вдруг что, не дай Бог, вот... И когда меня просто привезли туда, то выяснилось, что всё в порядке и с ребенком всё тоже хорошо, но тем не менее посоветовали просто понаблюдаться какое-то время в больнице, вот. И наверное, да, неделю я там лежала. При этом я сообщила об этом и молодому человеку, и родителям – и тем, и тем – о том, что я остаюсь в больнице, вот. Рассказала всю эту ситуацию, ммм... короче говоря, дошло просто до того, что, наверное, на третий день, э-э-э... утром мне звонит молодой человек – говорит, мол, не знаю, делай, чё хочешь, типа, но мне дома – цитата: "жрать нечего" – типа, я сам себе приготовить не знаю что, типа, мол, я хочу есть. Я говорю: "Ром, ну ты сам понимаешь – я в больнице лежу. Когда меня выпустят, я не знаю." Вот, говорю, сходи к маме – может быть, у мамы поужинай – или, господи, купи себе сосисок, свари себе сосиски с макаронами, пельмени, я говорю, не знаю, те же нагетсы пожарь. То есть, ну что там – достал, бросил на сковородку и жди там, когда они пожарятся. Ну всё, ну элементарно. "Нет, и всё. Мне нужна, типа дома ты, мол, я сам – я не могу."

Но вот благо, наверное, там был выходной день и особо так сильно не контролировали – я отпросилась на время тихого часа. Я отпросилась у главной медсестры съездить домой под предлогом, что дайте себя хотя бы в порядок привести, то есть, там, душ принять, голову помыть – хоть что-то. "Да, хорошо, езжай, типа, но к 4м должна вернуться." Без проблем. Получается, где-то в 2, наверное, я приехала домой. Я приготовила ему ужин, всё это оставила дома. Я поехала туда обратно. Вот вечером он мне опять звонит, типа: "А я не понял, а чё ты не дома?" Я говорю: "А как ты себе это представляешь?" Я говорю, меня отпустили на 2–3 часа и то под обязательной подпиской, что я вернусь. Он говорит: "Ну вот, ну хоть спасибо еду оставила." То есть в таком плане. Ну то есть это про... ну, 1 или 2 раза ещё так приезжала туда домой. Вот, пока... Также родила ребенка. Ну я не скажу, что прямо как-то особо его отношение ко мне как-то там изменилось в лучшую или худшую сторону, ммм... Ну, мне кажется просто отношение – оно как было, так оно и осталось. То есть, да отношения у нас, ммм, продолжали быть тёплыми, но-о-о всё равно вот эти вот все требования: что ты женщина, ты должна – оно осталось. То есть никаких поблажек тоже, ну там, на все мои просьбы, что вот, пожалуйста, спина болит, там, или мне тяжело наклоняться, там, чтобы пол протереть или ещё что-то – ну помоги хоть чем-то. Нет, это всё не делалось".

Опишем основные этапы кодирования во взаимодействии с ИИ по правилам "обоснованной теории". Данный фрагмент был предложен для прочтения Chat-gpt 4.5 с последующим заданием "конденсировать смысл по ходу прочтения теста". Был получен следующий ответ:

"Вот по ходу интервью я выделяю основные смыслы и подтексты, которые передаёт молодая мама:

1) Отношения с молодым человеком во время беременности не претерпели значительных изменений. Прямой смысл: героиня отмечает, что "ничего сильно не поменялось". Подтекст: это говорит скорее не о стабильности, а о застое в отношениях – несмотря на беременность, её партнёр не стал более чутким, ответственным или заботливым.

2) Гендерные стереотипы: "ты женщина – ты должна". Это – центральная линия интервью: молодой человек, по словам героини, продолжает считать, что все заботы и обязанности – на женщине. Даже беремен-

ность и физические ограничения не вызвали сочувствия или пересмотра ролей. Он требует, чтобы она обслуживала его, даже когда сама находится в больнице.

3) Эпизод с больницей как показатель крайнего неравенства. Она попадает в больницу из-за падения, и её самочувствие – потенциально под угрозой. Вместо поддержки, партнёр жалуется на голод и требует, чтобы она приехала готовить еду, несмотря на госпитализацию. Ключевая цитата: "Мне дома жрать нечего" – это подчеркивает незрелость и эгоцентризм партнёра.

4) Самоотверженность и подавление своих потребностей. Она отправляется из больницы под надуманным предлогом, чтобы угодить его требованиям – хотя сама нуждается в заботе. Героиня демонстрирует жертвенное поведение, на фоне эмоционального и бытового давления.

5) После родов – всё то же. Отношение не изменилось, несмотря на появление ребенка. Никаких "поблажек" она не почувствовала: ни во время беременности, ни после. Физические трудности (боль в спине, невозможность наклоняться) игнорируются партнёром".

Предложенная конденсация смысла имеет хронологическую последовательность, но мы видим, что данным интервью нет подробного рассказа о том, что происходило после родов, однако "языковая модель" достраивает пятый пункт. "Боль в спине, невозможность наклоняться" относилась к опыту во время беременности, но после выписки из больницы. Этот пример показывает, что ИИ может давать основу для описания случая, но критическая перепроверка предложенной структуры нарратива должна начинаться с первых этапов работы с текстом и построения категориальной сетки. Это же касается и формулировки смыслов и акцентов. Например, мы видим фразу "самочувствие под угрозой", больше описано негативное поведение партнёра, нежели реконструированы мотивы и поведение самой информантки.

Следуя логике кодирования, мы шли пошагово, предлагая задания (промпты) по открытому (первичному), осевому (фокусированному) кодированию. Мы заметили, что ответы на запросы не всегда согласовывались между собой. Возможно, это связано с логикой работы "языковой модели". В алгоритме нет самостоятельного "редактирования", сомнения и скептицизма по предложенным кодам и смыслам. ChatGPT среагирует на новый запрос, если этот запрос будет уточнён, переформулирован другими словами. Алгоритм продолжает нашу реплику, начиная снова писать, базируясь на всех предыдущих текстах, а не редактирует свою. Также, в разные дни предлагая чату тот же фрагмент интервью и давая схожий промт / задание, мы получали различающиеся варианты ответов. В этом смысле языковая модель похожа на человеческое сознание, осуществляющее интерпретацию – она также как и человек не повторяет буквально порядок слов, а даёт новую версию, близкую по смыслу.

В ходе осевого кодирования мы предлагали осевые категории, основанные на теориях. Например, мы давали такие запросы: "Используй для кодирования и конденсации смысла феминистские теории гендерного режима" или "Используй для кодирования теории телесности для интерпретации проживания беременности".

Получили ответ модели:

"Добавим категорию "гендерного режима" в структуру анализа, с опорой на теории феминистской социологии (в частности, Рэйвена Коннелл), и встроим её в существующую систему категорий по логике обоснованной теории.

Сначала – краткое теоретическое обоснование, затем – включение в категориальную схему и взаимосвязи.

1. Теоретическая основа: гендерный режим.

Понятие "гендерного режима" (gender regime) предложено социологом Р. У. Коннелл для описания того, как конкретные институты, организации или практики воплощают и воспроизводят гендерный порядок.

Гендерный режим – это не просто совокупность норм, но и структурно закреплённый способ организации: труда и заботы; власти и подчинения; сексуальности и телесности; символических ожиданий и нормативных ролей.

В контексте интервью молодой мамы гендерный режим проявляется как устойчивая структура взаимоотношений в паре, где: женщина (даже беременная) несёт полную ответственность за быт и заботу; мужчина освобождён от этих обязанностей и не видит причин менять поведение; беременность усиливает асимметрию, а не смягчает её.

2. Категория в системе обоснованной теории.

Название категории: Гендерный режим домашнего уклада. Примеры кода: Нормализация неравенства: "Он так же ко мне относился... ничего не изменилось". Институционализация женской ответственности: "Мне нужна ты дома... Я не могу сам". Отсутствие пересмотра ролей после беременности и родов. Неприсутствие мужчин в заботе: "Сама себе приготавливаю, я не знаю как".

Описание категории: "тот гендерный режим формирует устойчивую модель домашней жизни, в которой женщина – основной ресурс ухода, эмоционального труда и физического обеспечения, независимо от её состояния. Мужчина – бенефициар системы, в которой женская забота воспринимается как "естественная" и "бесконечная". Беременность усиливает, а не ослабляет эти механизмы".

В зависимости от предложенных заданий, варианты предлагаемых кодов ожидаемо различались. Предложенный "языковой моделью" текстовый материал становился поводом для дальнейших размышлений: соотношения предложенных категорий с текстом, обсуждения достаточности материала в тексте для принятия категории, её редактирование для описания более точного смысла. Так, например, ChatGPT предлагал категорию "эмоциональное обесценивание" беременной её партнёром, хотя буквально об этом не шла речь во фрагменте интервью. Или сконструировал не используемую в сообществе исследователей категорию "гендерное ожидание".

Двигаясь далее, мы запросили на основе данного фрагмента сконструировать парадигмальную модель по Б. Глезеру и А. Страуссу (1967). ChatGPT предложил вот такую версию:

"1) Явление (Central Phenomenon): Переживание женщиной беременности и материнства в условиях гендерного неравенства и эмоционального обесценивания со стороны партнёра.

2) Каузальные условия (Causal Conditions): Воспитание в патриархальных культурных установках: "женщина обязана заботиться". Низкий уровень эмпатии и зрелости у партнёра. Закреплённая модель: "забота – это женская обязанность вне зависимости от состояния". Отсутствие опыта взаимного партнёрства.

3) Контекст: Беременность (физиологически и психологически уязвимое состояние). Госпитализация из-за угрозы прерывания беременности. Совместное проживание с партнёром, зависимость в быту. Отсутствие поддержки со стороны семьи / окружения.

4) Интервенирующие условия (Intervening Conditions): Образ "хорошей женщины / матери", навязанный социальной нормой. Чувство вины за "неудовлетворённость" партнёра. Ограниченный доступ к альтернативным моделям отношений. Ожидание благодарности за минимальные действия ("спасибо, что оставила еду").

5) Стратегии действия / взаимодействия (Action/Interaction Strategies): Подчинение и сохранение отношений любой ценой. Самообесценивание: отказ от признания собственной боли. Выполнение эмоционального и физического труда, несмотря на медицинские показания. Рационализация поведения партнёра (поиск оправданий)".

6) Последствия (Consequences): Закрепление неравного положения женщины в семье. Повышенная нагрузка и стресс в период беременности. Отсутствие полноценной заботы в важнейший период жизни. Продолжение циклов обесценивания и эксплуатации. Риск психологических и физических осложнений".

Данная парадигмальная модель отличается от нашей тем, что в ней не все логики, заявленные информанткой, отражены. Подчёркнута излишняя жертвенность информантки, нет интерпретации её активизма, желания поддержать отношения согласно традиционным ценностям, вписаться в требования общества и соответствовать им. Посредством своего приезда в дом,

она хотела угодить мужу, показать свою энергичность, желание поддержать "тёплые отношения". Ещё одна линия – это её почти подростковая молодость, желание "играть в выполнение заданий", не соотнося со своими силами. В рамках данного исследования мы предполагали показать разные логики действия беременных в процессе беременности и их собственное понимание, а языковая модель давала только самые распространённые в мировой науке интерпретации. Возможно, для развития осевого кодирования стоит давать разные задания и потом соотносить их для получения более полной картины.

В нашем примере модель построена только по фрагменту интервью, то есть по отдельной ситуации, которая является частью прожитого опыта за время беременности и в интервью также рассматривали и другие стороны жизни информантки – отношения со своими родителями, с родителями мужа, обучение в университете, отношения с преподавателями и со студентами. Однако, для описания только приведённого во фрагменте случая важно учитывать контекст всего интервью.

Этот пример показывает, что для включения "языковой модели" в аналитический процесс, нужно реализовывать итерационное многоэтапное взаимодействие с "языковой моделью", разрабатывать подробные формулировки для каждого этапа кодирования, принимая тот факт, что алгоритм начнёт писать, реагируя буквально на наши слова в запросе. Поэтому мы можем менять формулировки, изучая, как меняется ответ / реплика используемой "языковой модели". Не стоит предлагать "сложных" заданий для модели, чтобы избежать "опротечивых" реакций. И наконец, процесс верификации остается полностью под ответственностью исследователя, который перепроверяет каждый предложенный код и категорию, перепроверя сначала достаточность данных для их принятия и точность переданных смыслов.

Сопоставим результаты нашего взаимодействия с ChatGPT с опытом коллег, апробировавших его для качественного анализа ранее. Для того, чтобы показать, как различаются ручное кодирование и кодирование языковой моделью Д. Морган представил в своей статье несколько примеров сравнительного анализа кодов и их интерпретации [18]. Его результаты также показали, что используемый им ChatGPT лучше справился с описанием и выделением тем, нежели с их последующей интерпретацией и связыванием с теоретическими понятиями. Опыт российских коллег [1] показал, что языковая модель может провести объективный семантический анализ больших текстовых данных, то есть выделить наиболее часто встречающиеся слова, на основании которых можно зафиксировать тематику разговора или изучаемого медийного поля, и провести сентимент-анализ (тональность текстов). Однако эти выполненные "языковой моделью" задания оказываются подготовительными этапами для дальнейшего теоретического осмысления текстового материала. Все примеры указывают на то, что обращение к "языковой модели" должно быть осмысленным и управляемым со стороны исследователя, а её мощности дополнять имеющиеся у исследователя знания и давать почву для новых идей и новых заданий / промтов.

Учитывая ранее описанные типы кодирования для естественного интеллекта, но предполагающую пошаговую работу от сырых данных к социологической концептуализации, логика написания промтов / заданий для "языковой модели" – может быть следующей:

Задания на поиск данных, если речь идет об открытых данных в медиа.

1) По имеющимся ключевым словам, найти ссылки на данные, где много об этом говорится.

2) Поиск статей, в которых есть теории, касающиеся какой-то социологической категории, допустим "доверия".

Задания на группировку частей / секвенций текста.

3) Группировка текста и кодирование – разметка текста по темам, которые предложит сама "языковая модель" (что-то похожее на первичное кодирование), или разметка текста по предложенным темам (близкое к осевому кодированию).

4) Уточнение группировки текста за счёт конкретизации задания – добавления деталей, фокусов тем, фактов, логики проявления изучаемого феномена (доработка осевого кодирования).

5) Пошаговая работа в каждом кодом / категорией при уточнении заданий (что-то близкое к выборочному, фокусированному кодированию).

Аналитические задания.

6) Свободное "размышление" о найденных данных по определенному коду, категории.

7) Более фокусные задания на размышление: проведение сравнительного анализа разных цитат из текста, помеченных как релевантные коду (предложить найти сходства и различия), выявление разных логик присвоения кодов цитатам, поиск цитат, которые можно было бы объединить по общей логике действия, произошедших событий, имеющихся ресурсов, похожих условий и т. д.

8) Задания на размышления, которые обосновывают применение социологического понятия для объяснения сгруппированных данных.

9) Задания на поиск релевантных статей, где в похожем ключе применяется социологическое понятие или социологическая теория для объяснения собранных данных.

10) Визуализация данных, ментальная карта описанных слов, образов и т. д.

Работа с "языковой моделью" как аналитическая триангуляция

Как показывает опыт, тест Тьюринга на идентификацию себя как профессионала-социолога "языковые модели" пока ещё сдать не могут. Как отмечает К. Бейл, "LLM, по-видимому, присваивают более точные коды одним темам, чем другим, что может быть артефактом способа их обучения" [9, p. 5], что связано с тем, что задания универсального характера машина выполняет чаще, чем социологические, и быстрее тренируется. Во-вторых, "одним из существенных недостатков количественного анализа текстов выступает отсутствие специализированных социологических словарей, которые могли бы эффективно работать со сложными социальными значениями слов" [1, с. 113] и помогать "языковым моделям" точнее идентифицировать как задания, так и данные, на которых нужно их выполнить. "Языковые модели", как и люди, проходят "социализацию", что влияет на выборочность знаний и умений, убеждения и готовность "галлюцинировать" ради выполнения задания.

В своё время Б. Глезер и А. Страус [14] практиковали приглашение в свою аналитическую команду свежих сил, не являющихся социологами, которые могли выдвигать самые разные, в том числе и "странные" идеи для их дальнейшего критического осмысления. В современном партисипаторном подходе также важно видеть разные стороны осмысления социальной проблемы. И поскольку мы живём в цифровом мире, то к процессу аналитической триангуляции [13] или методу длинного стола [8] мы приглашаем искусственный интеллект, чтобы он внёс свой вклад. Триангуляция предполагает, что каждый вносит свою лепту в первичный материал для дальнейшего сравнения. Это хорошо согласуется со стратегией применения "языковой модели" для выполнения тематического анализа, которую предлагают М. Наим, Т. Смит и Л. Томас, советуя исследователям выполнять ручное кодирование на всех этапах и сравнивать его с машинным [19], контролируя смысловые смещения, заданные несовершенством алгоритма [21].

Однако, за здравый смысл и результат исследования отвечает только исследователь. Искусственный интеллект пока оказывается той "свежей головой" непрофессионала, который предлагает новые идеи, иногда необоснованные, требующие со стороны исследователя перепроверки и осмысления, критического мышления и творческого развития идей. С другой стороны, необычные идеи могут заложить новые тематические векторы анализа, снять "замыленность" исследовательского взгляда. Как в своё время отмечал Б. Глезер, "странные" данные никогда не смогут стать источником помех при опровержении гипотезы, полученной в ходе дедукции, но они дают отличную возможность с точки зрения того, что они могут привнести в результате сравнения в теоретическое расширение, оттачивание и обогащение возникающей теории" [15, p. 37–38; 4, с. 127]. Описанные им процедуры индукции, дедукции в верификации в процессе сравнительного анализа, даже если мы его делаем вместе с "языковой моделью", помогают делать шаги к научному выводу, эксплицировать его логику.

Заключение

Применение "больших языковых моделей" на данном этапе их развития даёт основания проблематизировать их потенциал для качественного анализа в социологии, формировать правила и этику применения. Во-первых, важно декларировать и чётко обозначать использование ИИ: в каких конкретно процедурах был использован, для чего это нужно было исследователю, как применение ИИ вписывалось в общую логику анализа, какие запросы к "языковой модели" осуществлялись и как исследователь контролировал несовершенства алгоритма. Во-вторых, этап применения ИИ должен быть итерационным, то есть должна производиться пошаговая серия запросов, уточняющая и развивающая предыдущий, и триангуляционным – ИИ только предлагает идеи, а их осмысление, обозначение вектора дальнейшего развития, корректировка и верификация должны осуществляться исследователями. Исследователь контролирует процесс "выращивания" теории – связки предложенных кодов/категорий с социологическими теориями, даёт интерпретацию полученных категорий, вдохновившись или удивившись сгенерированным текстом ИИ, и остается главным ответственным лицом за полученные результаты.

Литература / References

1. Давыдов С. Г., Матвеева Н. Н., Адемуклова Н. В., Вичканова А. А. Искусственный интеллект в российском высшем образовании: текущее состояние и перспективы развития // Университетское управление: практика и анализ. 2024. Т. 28. № 3. С. 32–44.
Davydov, S. G., Matveeva, N. N., Ademukova, N. V., Vichkanova, A. A. Artificial intelligence in Russian higher education: Current state and development prospects // University Management: Practice and Analysis. 2024. Vol. 28. No. 3. P. 32–44. (In Russ.).
2. Галкин К. А., Петухова И. С., Парфенова О. А. OPEN AI как помощник при анализе интервью // Социологические исследования. 2025. № 4. С. 105–116. <https://doi.org/10.31857/S0132162525040095>
Galkin, K. A., Petukhova, I. S., Parfenova, O. A. OPEN AI as an assistant in interview analysis // Sotsiologicheskie issledovaniya. 2025. No. 4. P. 105–116. <https://doi.org/10.31857/S0132162525040095> (In Russ.).
3. Готлиб А. С. Автоэтнография (разговор с самой собой в двух регистрах). (Часть I) // Социология: методология, методы, математическое моделирование (Социология: 4М). 2016. № 18. С. 5–16.
Gotlib A. S. Autoethnography (a conversation with myself in two registers). (Part I). // Sociology: Methodology, Methods, Mathematical Modelling (Sociology: 4M). 2016. No. 18. P. 5–16. (In Russ.).
4. Забаев И. В. Логика анализа данных в обоснованной теории: версия Б. Глезера // Социология: методология, методы, математическое моделирование (Социология: 4М). 2011. № 32. С. 124–142.
Zabaev I. V. The logic of data analysis in grounded theory: B. Glaser's version. // Sociology: Methodology, Methods, Mathematical Modelling (Sociology: 4M). 2011. No. 32. P. 124–142. (In Russ.).
5. Моисеев С., Стаф М. "There's an AI for that": возможности ChatGPT для работы с открытыми источниками данных // Социодиггер. 2023. Т. 4. Вып. 7–8(27). URL: <https://sociodigger.ru/articles/articles-page/algorithmy-iskusstvennogo-intellekta-v-prikladnykh-sociologicheskikh-issledovaniyakh> (дата обращения: 13.07.2025)
Moiseev, S., Staf, M. "There's an AI for that": Opportunities of ChatGPT for working with open data sources. // Sociodigger. 2023. Vol. 4. Issue 7–8(27). URL: <https://sociodigger.ru/articles/articles-page/algorithmy-iskusstvennogo-intellekta-v-prikladnykh-sociologicheskikh-issledovaniyakh> (accessed 13.07.2025) (In Russ.).
6. Практики анализа качественных данных в социальных науках : учебное пособие / отв. ред. Е. В. Полухина. М.: Издательский дом НИУ ВШЭ, 2023. 383 с.
Practices of qualitative data analysis in the social sciences: A textbook / Ed. E. V. Polukhina. Moscow: HSE Publishing House, 2023. 383 p. (In Russ.).
7. Рождественская Е. Ю. ИНТЕР-энциклопедия: нарративное интервью // Интеракция. Интервью. Интерпретация. 2020. Т. 12. № 4. С. 114–127. <https://doi.org/10.19181/inter.2020.12.4.8>
Rozhdestvenskaya, E. Yu. INTER-Encyclopedia: Narrative interview. // Interaction. Interview. Interpretation. 2020. Vol. 12. No. 4. P. 114–127. <https://doi.org/10.19181/inter.2020.12.4.8> (In Russ.).
8. Штейнберг И. Е. Метод "длинного стола" в качественных полевых социологических исследованиях. М: ВЦИОМ, 2021. 300 с.
Shteinberg I. E. The "long table" method in qualitative field sociological research. Moscow: VCIOM, 2021. 300 p (In Russ.).

9. Bail C. A. Can Generative AI improve social science? // *Proceedings of the National Academy of Sciences*. 2024. No. 21 (121). e2314021121. <https://doi.org/10.1073/pnas.2314021121>
10. Charmaz K. *Constructing Grounded Theory: A Practical Guide Through Qualitative Analysis*. Thousand Oaks (CA): Sage, 2006. 224 p.
11. Clarke A. *Situational Analysis: Grounded Theory after the Postmodern Turn*. Thousand Oaks (CA): Sage, 2005. 365 p.
12. Clarke A. Situational Analyses: Grounded Theory Mapping After the Postmodern Turn // *Symbolic Interaction*. 2003. 26 (4). P. 553–576. <https://doi.org/10.1525/si.2003.26.4.553>
13. Denzin N. K. *The research act: A theoretical introduction to sociological methods* (2nd ed.). New York, NY: McGraw Hill, 1978. 379 p.
14. Glaser B. G., Strauss A. L. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine, 1967. 271 p.
15. Glaser B. G. *Theoretical Sensitivity: Advances in the Methodology of Grounded Theory*. Mill Valley (CA): Sociology Press, 1978. 164 p.
16. Guest G., MacQueen K. M., Namey E. E. *Applied Thematic Analysis*. Thousand Oaks: SAGE Publications, Inc. 2012. 320 p. <https://doi.org/10.4135/9781483384436>
17. Ledwith M., Springett J. *Participatory practice. community based action for transformative change*. Bristol: Polity Press. 2010. 312 p.
18. Morgan D. L. Exploring the Use of Artificial Intelligence for Qualitative Data Analysis: The Case of ChatGPT // *International Journal of Qualitative Methods*. 2023. No. 22 (6). P. 1–10. <https://doi.org/10.1177/16094069231211248>
19. Naeem M., Smith T., Thomas L. Thematic Analysis and Artificial Intelligence: A Step-by-Step Process for Using ChatGPT in Thematic Analysis // *International Journal of Qualitative Methods*. Vol. 2023. No. 24 (7). P. 1–18. <https://doi.org/10.1177/16094069251333886>
20. *The SAGE Handbook of action research: Participative inquiry and practice*. London, UK: Sage, 2001. 720 p.
21. Ukanwa K. Algorithmic bias: Social science research integration through the 3-D Dependable AI Framework // *Current Opinion in Psychology*. 2024. Vol. 58. 101836. <https://doi.org/10.1016/j.copsyc.2024.101836>



Информация об авторе

Ольга Борисовна Савинская, канд. соц. наук, доцент кафедры методов сбора и анализа социологической информации, ведущий научный сотрудник Международной лаборатории исследований социальной интеграции, Национальный исследовательский университет "Высшая школа экономики", Москва, Россия, e-mail: osavinskaya@hse.ru

Information about the author

Olga B. Savinskaya, Candidate of Sociology, Associate Professor, Department of Methods of Collecting and Analyzing Sociological Data, leading research fellow of the International Laboratory for Social Integration Research, National Research University Higher School of Economics, Moscow, Russia, e-mail: osavinskaya@hse.ru

Поступила в редакцию 16.07.2025

Received 16.07.2025

Одобрена после рецензирования 07.08.2025

Approved 07.08.2025

Принята к публикации 25.08.2025

Accepted 25.08.2025