



DOI: 10.19181/inter.2026.18.2.5
EDN: CJFVV

Критерии оценки синтетических респондентов в качественных исследованиях: границы применимости

Ссылка для цитирования:

Жуликов К. О. Критерии оценки синтетических респондентов в качественных исследованиях: границы применимости // Интеракция. Интервью. Интерпретация. 2026. Т. 18. № 2. С. 147–160. <https://doi.org/10.19181/inter.2026.18.2.5> EDN: CJFVV

For citation:

Zhulikov K. O. (2026) Evaluation Criteria for Synthetic Respondents in Qualitative Research: Limits of Applicability. *Interaction. Interview. Interpretation*. Vol. 18. No. 2. P. 147–160. <https://doi.org/10.19181/inter.2026.18.2.5>



Жуликов Кирилл Олегович

Московский Государственный университет
им. Ломоносова,
Москва, Россия

E-mail: Kir.prog2002@gmail.com

Большие языковые модели (LLM) все чаще применяются для генерации синтетических респондентов — текстовых симуляций ответов информантов, которым задан определенный социально-демографический профиль. Для количественных синтетических данных уже разработаны подходы к оценке, основанные на сопоставлении статистических распределений с результатами массовых опросов. Для качественных данных устоявшиеся критерии отсутствуют. В литературе это описывается как временный пробел, который будет восполнен по мере развития инструментария, и как ограничение, связанное с более глубоким несоответствием между основаниями качественной методологии и природой синтетических данных. Статья содержит аналитический обзор подходов к оценке синтетических респондентов в трех исследовательских традициях: вычислительных социальных науках, обработке естественного языка и качественной методологии. Обзор публикаций 2022–2026 годов показал, что наиболее формализованные подходы (статистические сравнения распределений и автоматические текстовые метрики) измеряют свойства, второстепенные для качественной работы.

Подходы, сложившиеся внутри качественной традиции, затрагивают содержательно значимые свойства (контекстуальную специфичность, внутреннюю противоречивость, нарративное сопротивление), но зафиксированы на уровне наблюдений и не операционализированы. Сопоставление свойств качественных данных с существующими критериями выявляет неохваченные моменты, в том числе внутренние противоречия нарратива, речевые особенности и способность данных порождать нетривиальные интерпретации. На основе проведенного сопоставления очерчиваются те области, в которых применение синтетических респондентов в качественных исследованиях представляется оправданным, и те, в которых оно проблематично. Обсуждаются стандарты прозрачности для случаев их использования и обозначены перспективные направления исследований.

Ключевые слова: синтетические респонденты; синтетическая выборка; генеративный искусственный интеллект; большие языковые модели; качественные исследования; стандарты прозрачности; границы применимости LLM

Введение

Большие языковые модели (large language models, далее — LLM) все шире применяются в социальных науках не только для анализа, но и для генерации данных. Одна из таких практик состоит в создании *синтетических респондентов* — текстовых симуляций ответов, порождаемых моделями информантов с заранее заданным социально-демографическими профилями [Argyle, Busby, Fulda et al., 2023; Sarstedt, Adler, Rau et al., 2024]. Модель получает инструкцию принять характеристики конкретного человека и отвечать от его лица.

Масштаб этой практики быстро растет: систематический обзор 2026 года охватил 182 работы и зафиксировал расширение применения синтетических респондентов в психологии, политологии, маркетинге и UX-дизайне [Kuric, Demcak, Krajcovic et al., 2026]. Тема обсуждается и в программных статьях ведущих журналов [Grossmann, Feinberg, Parker et al., 2023; Bail, 2024], и в эмпирических сравнениях синтетических и реальных данных [Bisbee, Clinton, Dorff et al., 2024; Shrestha, Krpan, Koik et al., 2024].

Подавляющее большинство этих работ выполнено в логике количественной методологии. Исследователи сравнивают распределения ответов, воспроизводят экспериментальные эффекты, рассчитывают корреляции между синтетическими и реальными выборками. Качественная социология, работающая с интервью, нарративами и биографическими рассказами, остается на периферии дискуссии. Между тем отдельные авторы уже тестируют LLM в роли участников полуструктурированных интервью [Karania, Agnew, Eslami et al., 2025], а такие задачи, как пилотирование гайда, тренировка начинающих интервьюеров и генерация предварительных гипотез, активно обсуждаются [Bail, 2024].

В отечественной литературе тема начинает разрабатываться с 2024–2025 годов. Е. Г. Джибилова и Н. С. Побываев [Джибилова, Побываев, 2024]



анализируют российские и зарубежные примеры внедрения ИИ в социальную сферу и приходят к выводу, что кейсов использования в исследовательской практике пока немного, что актуализирует изучение возможностей и ограничений языковых моделей. К. А. Галкин, И. С. Петухова и О. А. Парфенова [Галкин, Петухова, Парфенова, 2025] рассматривают OpenAI как инструмент, дополняющий традиционные подходы к анализу качественных данных. Ближе всего к теме настоящего обзора стоит работа А. Г. Филиповой, Е. Е. Абросимовой и О. Г. Зубовой [Филипова, Абросимова, Зубова, 2025], посвященная методу синтетических фокус-групп, создаваемых с помощью LLM. Авторы анализируют возможности моделирования групповой дискуссии без участия реальных респондентов и фиксируют проблему предвзятости генерируемых ответов. Ж. В. Пузанова, Г. Г. Кожоридзе и Д. Г. Кожоридзе [Пузанова, Кожоридзе, Кожоридзе, 2025] предлагают развернутый обзор отечественных и зарубежных примеров использования виртуальных респондентов на основе LLM, включая эмпирическую проверку их возможностей. Работа выполнена в логике количественной методологии и сосредоточена на проблемах репрезентативности и симуляции опросов. М. И. Незговорова [Незговорова, 2026] предлагает типологию синтетических респондентов и описывает институционализацию метода в российской индустрии. При этом перечисленные работы выполнены преимущественно в количественной или прикладной логике, вопрос о критериях оценки синтетических данных в интерпретативной традиции остается за рамками обсуждения. Таким образом, русскоязычная литература по теме остается фрагментарной.

Для количественных данных уже существуют устоявшиеся процедуры оценки, включая сопоставление статистических распределений, регрессионный анализ и проверку воспроизводимости эффектов [Argyle, Busby, Fulda et al., 2023; Bisbee, Clinton, Dorff et al., 2024]. Для качественных данных устоявшейся рамки оценивания не сложилось, и существующие критерии остаются разрозненными, распределенными между несколькими исследовательскими традициями [Davidson, Karell, 2025]. Стандартные метрики обработки естественного языка (Perplexity, BLEU, BERTScore) не учитывают значимые для интерпретативной работы свойства — контекстуальную укорененность (contextual embeddedness), внутренние противоречия и нарративную плотность (narrative density) [Kapania, Agnew, Eslami et al., 2025; Schroeder, Aubin Le Quééré, Randazzo et al., 2025]. Отсутствие критериев оценки синтетических респондентов в качественных исследованиях обычно описывается как временный пробел, который будет восполнен по мере развития инструментария [Bail, 2024; Davidson, Karell, 2025]. Однако эмпирические работы, в которых качественные исследователи взаимодействовали с LLM в роли респондентов, фиксируют не технические недочеты, а структурные ограничения метода, затрагивающие сами основы интерпретативного подхода [Kapania, Agnew, Eslami et al., 2025; Bisbee, Clinton, Dorff et al., 2024].

Цель статьи — систематизировать критерии оценки синтетических респондентов, описанные в литературе, и проанализировать их применимость к задачам качественного социологического исследования. Обзор выполнен

на материале публикаций 2022–2026 годов, выявленных в базах eLIBRARY.RU, Google Scholar, Scopus, ACM Digital Library, а также на платформах arXiv и SSRN. Выявленные критерии и метрики классифицированы по традиции происхождения, типу оцениваемого свойства и степени применимости к качественным данным. Традиции включают вычислительные социальные науки, обработку естественного языка и качественную методологию. Типы оцениваемых свойств разделены на статистические, лингвистические и содержательные.

Свойства качественных данных, значимые для аналитической работы

Прежде чем обращаться к существующим критериям оценки, следует зафиксировать, что именно делает качественные данные пригодными для интерпретативной работы. Свойства, рассматриваемые ниже, отобраны на основании двух корпусов литературы. Первый составляют классические и современные исследования по качественной методологии, формулирующие требования к данным: из обширного списка отобраны работы, в которых сформулированы критерии оценки качества данных [Lincoln, Guba, 1985] или охарактеризованы свойства, отличающие качественные данные от количественных [Geertz, 1973; Braun, Clarke, 2019; Семенова, 1998; Рождественская, 2012]. Данный обзор работ фиксирует базовый набор требований, на соответствие которым проверяются синтетические данные. Второй корпус литературы составляют публикации, непосредственно обращающиеся к синтетическим респондентам и фиксирующие, какие свойства реальных данных LLM не воспроизводят [Kapania, Agnew, Eslami et al., 2025; Schroeder, Aubin Le Quéré, Randazzo et al., 2025; Kuric, Demcak, Krajcovic et al., 2026]. В обзор включены те свойства, которые упоминаются в обоих корпусах, что позволяет сосредоточиться на характеристиках, значимых одновременно для качественной методологии и для проблематики синтетических данных.

В классической работе по качественной методологии [Lincoln, Guba, 1985] сформулированы четыре критерия качества. Это достоверность (*credibility*), переносимость (*transferability*), зависимость (*dependability*) и подтверждаемость (*confirmability*). В отечественной литературе близкую проблематику разрабатывает В. В. Семенова [Семенова, 1998], подчеркивая, что качественное исследование строится на принципе теоретической выборки и предполагает максимальную вариацию случаев, а не статистическую репрезентативность. Все эти критерии предполагают наличие реального субъекта, к которому можно обратиться повторно. Достоверность проверяется обращением к самим участникам, а синтетический респондент такой возможности не предоставляет.

Содержательные характеристики данных раскрывает понятие насыщенного описания (*thick description*) [Geertz, 1973], указывающее на значимость социального, культурного и ситуативного контекста. Именно детали, укорененные в конкретном месте и времени, отличают качественные данные



от абстрактных обобщений. Е. Ю. Рождественская [Рождественская, 2012] развивает эту линию применительно к биографическому методу, показывая, что нарративное интервью порождает данные, в которых контекст неотделим от содержания высказывания. Качественное интервью не сводится к «сырью», а возникает во взаимодействии, где исследователь задает уточняющие вопросы, реагирует на паузы и считывает невербальные сигналы [Braun, Clarke, 2019]. Синтетический респондент лишен этой интерактивности и реагирует только на текст промпта.

Другой значимой характеристикой выступает нарративная связность (narrative coherence), предполагающая наличие внутренней логики рассказа, включая разрывы и незавершенные мысли [Riessman, 2008; Рождественская, 2020]. С нарративной связностью тесно связаны внутренние противоречия информанта, то есть сомнения, колебания и смена позиции, которые рассматриваются как один из наиболее ценных ресурсов анализа [Kvale, Brinkmann, 2009]. Эти противоречия проявляются не только в содержании высказывания, но и в его форме. Речевые особенности, такие как паузы, оговорки и самокоррекция, фиксируются при транскрибировании и активно используются в конверсационном анализе. Еще одно свойство, значимое для интерпретативной работы, связано со способностью данных порождать нетривиальные интерпретации, обозначаемые через понятие серендипности (serendipity), то есть открытия, которого исследователь не ожидал [Fine, Deegan, 1996]. Каждое из этих свойств укоренено в реальном опыте информанта, что ставит вопрос о принципиальной способности LLM их воспроизвести.

Перечисленные свойства задают ту рамку, в которой можно оценить синтетические тексты. Систематический обзор 2026 года отмечает, что тексты, сгенерированные LLM, отличаются избыточной структурированностью, грамматической правильностью и лексическим разнообразием [Kuric, Demcak, Krajcovic et al., 2026]. Реальные респонденты, напротив, порождают больше «ошибок», в том числе оговорок, незавершенных предложений и отступлений. Эти «ошибки» служат свидетельством когнитивных процессов, а не дефектом данных [Kvale, Brinkmann, 2009].

Подходы к оценке синтетических данных

Вычислительные социальные науки

Наиболее систематизированная рамка оценки, обозначенная как алгоритмическая точность (algorithmic fidelity), предложена в 2023 году [Argyle, Busby, Fulda et al., 2023]. Она включает четыре стандарта: неотличимость сгенерированных ответов от реальных при слепой экспертной оценке (social science Turing test), воспроизводство известных закономерностей (backward continuity), предсказание новых закономерностей (forward continuity) и совпадение структурных паттернов между выборками (pattern correspondence). Эта рамка разработана для опросных данных и операционализована через сопоставление распределений ответов шкального типа.

Эмпирическая проверка алгоритмической точности [Bisbee, Clinton, Dorff et al., 2024] проводилась на материале Американского национального электорального исследования (ANES). GPT-3.5 Turbo получал промпты с характеристиками реальных респондентов ANES и отвечал на вопросы шкалы «термометра чувств» (feeling thermometer). Из рассчитанных на синтетических данных регрессионных коэффициентов 48% статистически значимо отличались от полученных на реальных данных. У 32% расхождений менялся знак эффекта: синтетические данные указывали на связь, обратную реальной. Отдельную проблему составляла нестабильность во времени — распределения ответов на один и тот же промпт в апреле и июле 2023 года изменились из-за обновления модели.

Проблематика оценки расширяется в направлении межкультурных сравнений [Shrestha, Krapan, Koai et al., 2024]. Авторы тестируют синтетических респондентов, сконструированных с учетом демографических характеристик и личностных черт, на материале трех стран — Саудовской Аравии, Объединенных Арабских Эмиратов и США. Корреляции между агрегированными синтетическими и реальными ответами оказываются в целом достаточно высокими, однако воспроизводство паттернов для не-западных (non-WEIRD) стран менее точно. Это наблюдение фиксирует измерение оценки, отсутствующее в упомянутой рамке [Argyle, Busby, Fulda et al., 2023], а именно чувствительность синтетических данных к культурному контексту. Для качественных исследований, часто работающих с локальными, культурно-специфичными контекстами, этот результат имеет прямое значение. Кросс-модельные сравнения GPT-4o, Claude 3.5 Sonnet и DeepSeek V3 дополнительно фиксируют значимые расхождения на одном и том же наборе промптов, что указывает на зависимость результатов от архитектуры и версии модели [Kuric, Demcak, Krajcovic et al., 2026].

Все эти подходы ориентированы на сопоставление агрегированных показателей. Они отвечают на вопрос о том, насколько синтетическая выборка похожа на реальную по совокупности ответов. Применительно к качественному исследованию, где единицей анализа выступает отдельный нарратив, а не совокупность ответов, этих критериев недостаточно.

Обработка естественного языка и оценка ИИ-систем

Вторая группа подходов сформировалась в области обработки естественного языка и ориентирована на свойства текста как лингвистического объекта. Perplexity (мера предсказуемости текста), BLEU и BERTScore (метрики сходства с эталонным текстом) и показатели когерентности (coherence scores) измеряют грамматическую правильность, связность и лексическое разнообразие генерируемого текста. Систематический обзор 2026 года [Kuric, Demcak, Krajcovic et al., 2026] отмечает, что эти метрики применяются преимущественно в работах технической направленности, а в социальных научных исследованиях встречаются реже.

Другой подход опирается на экспертное различие. Эксперту предъявляется перемешанный массив реальных и сгенерированных текстов,



задача — определить происхождения каждого из них [Argyle, Busby, Fulda et al., 2023]. В коммерческом секторе этот принцип развивают платформы, предлагающие генерацию синтетических участников как услугу и заявляющие о совпадении результатов с реальными интервью на уровне 85–92% [Kuric, Demcak, Krajcovic et al., 2026]. Методология расчета этого показателя, однако, не раскрывается, и остается неясным, по каким именно параметрам измеряется совпадение — по тематическому составу, глубине ответов, эмоциональной насыщенности или другим показателям.

Попытка автоматизировать экспертную оценку реализуется в подходе, при котором одна языковая модель оценивает выход другой. Этот подход, впрочем, сопряжен с рядом систематических ограничений: модели склонны выше оценивать тексты того же разработчика, завышать оценки объемных ответов и зависеть от формулировки инструкции [Zheng et al., 2023].

Отдельное направление связано с оценкой психологического портрета, демонстрируемого LLM при ответах на психометрические инструменты (шкалы Big Five, тесты моральных дилемм). Результаты ряда исследований показывают устойчивый сдвиг моделей к просоциальным ответам и сниженную выраженность негативных эмоций [Kuric, Demcak, Krajcovic et al., 2026]. При репликации исследований различные запуски модели давали ответы с нулевой или почти нулевой вариацией, воспроизводя единственный «правильный» ответ, устойчивый к изменению демографических характеристик персоны [Park, Schoenegger, Zhu et al., 2024]. Реальные люди многомерны и внутренне противоречивы, модели же порождают профили с нереалистично высокой согласованностью черт, не встречающейся у реальных людей.

Качественная методология

В исследовании, представленном на конференции CHI [Karania, Agnew, Eslami et al., 2025], 19 опытных качественных исследователей взаимодействовали с LLM в роли респондента через специально разработанный интерфейс. Исследователи задавали модели вопросы по заранее определенным темам, среди которых были доступность рабочей среды для сотрудников с особенностями развития и опыт миграции, а затем рефлексировали над полученным опытом.

Исследователи отмечали выраженную чувствительность модели к формулировке промпта. Замена одного слова в системной инструкции радикально меняла тональность ответов, сдвигая их от негативной к позитивной. В реальном интервью информант может скорректировать свою позицию под влиянием хода беседы, но его базовый опыт при этом остается неизменным. Модель же перестраивает ответ целиком, как если бы за разными промптами стояли разные люди.

Другое наблюдение касается отсутствия нарративного сопротивления (narrative resistance). Реальный информант может отказаться отвечать, уйти от темы, продемонстрировать дискомфорт. Для исследователя, работающего в интерпретативной традиции, такие отказы и уходы от темы не помеха, а аналитический ресурс, указывающий на болезненные или табуированные зоны

опыта. Модель же всегда готова отвечать и не выходит за пределы заданной темы. Отдельную проблему составляет непрозрачность обучающих данных. Невозможно проверить, в каком объеме голоса изучаемой группы представлены в обучающем корпусе модели. Если задача исследования состоит в том, чтобы понять опыт конкретного сообщества, валидность синтетических ответов зависит от состава корпуса, который, как правило, не раскрывается разработчиками.

Мгновенное производство данных в неограниченном количестве несет и более общий риск — утрату самого усилия вслушивания, которое в качественной традиции считается условием аналитической работы [Schroeder, Aubin Le Quéré, Randazzo et al., 2025]. Этот риск проявляется не только при генерации синтетических респондентов, но и при ИИ-ассистированном анализе реальных данных [Галкин, Петухова, Парфенова, 2025]. Общая картина применения LLM в качественных исследованиях остается фрагментарной. Систематическое картирование [Barros, Azevedo, Graciano Neto et al., 2025], охватившее 21 исследование, показало, что авторы каждого из них используют собственные критерии для сравнения результатов LLM с результатами традиционного анализа, и единой рамки оценивания не выявлено.

Свойства качественных данных, охваченные существующими критериями оценки

Сопоставление свойств, описанных в разделе «Свойства качественных данных, значимые для аналитической работы», с подходами к оценке, рассмотренными в разделе «Подходы к оценке синтетических данных», позволяет зафиксировать охваченные и неохваченные области.

Нарративная связность частично затрагивается средствами обработки естественного языка через метрики когерентности, а в качественной традиции фиксируется на уровне наблюдений исследователей; вычислительные социальные подходы это свойство не учитывают. Разнообразие голосов частично улавливается вычислительными социальными подходами при анализе подгрупп и средствами NLP через показатели лексического разнообразия, однако в качественной традиции оно не формализовано. В обоих случаях формализованные метрики фиксируют лишь отдельные проявления этих свойств. Внутренние противоречия и контекстуальная специфичность (*contextual specificity*) зафиксированы в качественных исследованиях на уровне наблюдений, но не операционализированы и не охвачены ни вычислительными, ни лингвистическими подходами. Речевые особенности и способность данных порождать нетривиальные интерпретации не охвачены ни одним из рассмотренных подходов, последнее свойство обсуждается в качественной традиции лишь концептуально.

Характер обнаруженных проблем иллюстрируют два примера из систематического обзора [Kuric, Demcak, Krajcovic et al., 2026]. Когда синтетическим участникам предложили назвать художественную видеоигру, подавляющее большинство выбрало одну и ту же игру, воспроизводя наиболее очевидный



ответ вместо индивидуального. Этот эффект фиксирует дефицит контекстуальной специфичности, не улавливаемый ни одной из рассмотренных метрик. Другой пример связан с нарративной связностью: ответы моделей в повторных тестах напоминали ответы одного и того же человека, а не разных респондентов. Метрики связности фиксируют внутреннюю согласованность текста, но не способны обнаружить проблему избыточной однородности, при которой все синтетические нарративы оказываются слишком последовательными по сравнению с реальными интервью, содержащими разрывы, возвраты и незавершенные мысли.

Эти примеры указывают на разрыв между формализованными метриками, не затрагивающими свойства, значимые для качественного исследования, и наблюдениями исследователей, которые затрагивают эти свойства, но не переведены в воспроизводимые процедуры. Формализация критериев качества при этом сама по себе остается предметом дискуссии, поскольку интерпретативная работа не сводима к процедуре, а рефлексивное суждение исследователя нельзя заменить стандартизированным перечнем. Отсутствие рамки оценивания, впрочем, оставляет исследователей без инструмента различения между ситуациями, в которых синтетические данные полезны, и ситуациями, в которых они проблематичны.

Эта проблема приобретает особую остроту в свете того, что языковые модели порождают правдоподобный текст без понимания содержания [Bender et al., 2021]. Синтетический нарратив может быть структурно корректен, стилистически убедителен и тематически уместен, но при этом не отсылать ни к какому реальному опыту. В количественном исследовании это менее критично, поскольку статистические закономерности устойчивы к отсутствию субъекта за данными. Качественное исследование работает с единичным и ситуативным, и отсутствие укорененности в реальном опыте затрагивает основу аналитической работы.

Наряду с академическими подходами своя система оценки складывается и в коммерческом секторе исследований. На зарубежном рынке за последние несколько лет сложился развитый ландшафт платформ, предлагающих синтетических респондентов как услугу [Sarstedt, Adler, Rau et al., 2024]. В 2025 году ESOMAR совместно с Международной торговой палатой выпустили обновленную редакцию ICC/ESOMAR International Code, где отдельно закреплены требования к использованию искусственного интеллекта, синтетических данных и синтетических персон: прозрачность, раскрытие роли ИИ, указание степени человеческого надзора и предоставление информации, необходимой для оценки валидности результатов¹. В российском контексте институционализация метода происходит преимущественно в продуктовых и UX-исследованиях, а также в количественной инфраструктуре опросных панелей [Незговорова, 2026]. Критериями оценки здесь выступают показатели, релевантные рынку: скорость генерации, тематический охват, согласованность распределений с реальными ответами, сохранение корреляций.

¹ ICC/ESOMAR International Code on Market, Opinion and Social Research and Data Analytics. 5-е изд. 2025. URL: <https://standards.esomar.org/assets/documents/icc-esomar-code-2025.pdf> (дата обращения: 15.05.2026).

Заключение

Подходы к оценке синтетических данных сложились в трех исследовательских традициях и не образуют единой рамки. Наиболее формализованные подходы (статистические сравнения и текстовые метрики) измеряют свойства, второстепенные для качественной методологии. Подходы из качественной традиции затрагивают значимые свойства, но зафиксированы на уровне наблюдений и не операционализированы. Внутренние противоречия нарратива, речевые особенности и способность данных порождать нетривиальные интерпретации не охвачены ни одним из рассмотренных критериев оценки.

Языковые модели комбинируют и переосмысляют уже встроенные в обучающие данные паттерны. Они малопригодны для выявления непредусмотренных исследователем смыслов — контринтуитивных свидетельств, неожиданных смысловых поворотов, эпизодов серендипности, ради которых во многом и проводится фундаментальное качественное исследование [Bisbee, Clinton, Dorff et al., 2024]. Таким образом, использование синтетических респондентов представляется обоснованным на подготовительных этапах работы, то есть в задачах, где требуется реконструкция уже известного, а не открытие нового. Это также объясняет, почему синтетические респонденты находят применение в маркетинговых и UX-исследованиях, где задачей является реконструкция уже известного распределения мнений или потребительских предпочтений. Проблематичными выглядят области, в которых синтетический респондент претендует на замещение реального голоса: исследования маргинализированных сообществ, биографические нарративы, чувствительные темы [Karania, Agnew, Eslami et al., 2025].

Стоит зафиксировать сопутствующее следствие: поскольку синтетических респондентов можно генерировать в неограниченном количестве и вслушивание в интонации каждого из них теряет смысл, качественное исследование с синтетическими респондентами органически тяготеет к количественному формату. Развернутые ответы большого массива синтетических респондентов могут обрабатываться как массив данных, и вместо качественного исследования возникает гибридный формат: массовый анализ, построенный не на анкете, а на гайде интервью. Таким образом, само слово «качественный» в контексте синтетических респондентов описывает форму вопроса, а не логику исследования. Данный подход может быть продуктивным, но нуждается в собственном языке описания, не заимствованном у интерпретативной традиции.

Перспективными направлениями выглядят разработка критериев оценки, специфичных для качественных данных, эмпирические сравнения реальных и синтетических интервью с использованием слепого кодирования (blind coding), а также выработка стандартов прозрачности при публикации результатов, полученных с привлечением LLM [Davidson, Karell, 2025]. Указание модели, промпта, параметров генерации и процедуры оценки качества могло бы стать минимальным стандартом, обеспечивающим воспроизводимость и этическую прозрачность. Формализация критериев качества для качественных данных сопряжена с эпистемологическими трудностями, и исследователи, вероятно,



столкнутся с необходимостью балансировать между воспроизводимостью и содержательной адекватностью.

В качестве предварительного шага к операционализации можно обозначить одно из неохваченных свойств — внутренние противоречия нарратива, то есть эпизоды колебаний, смены позиции и самоопровержений информанта. Возможный подход состоит в слепом кодировании перемешанного массива реальных и синтетических интервью независимыми кодировщиками, которые фиксируют эпизоды смены позиции, колебаний, самоопровержений и эмоциональной неоднозначности. Сравнение частоты и характера таких эпизодов между реальными и синтетическими текстами позволило бы эмпирически проверить наблюдение о склонности LLM к разрешению противоречий [Karania, Agnew, Eslami et al., 2025]. Подобный дизайн не требует разработки новой метрики, а опирается на существующие процедуры качественного кодирования, что делает его совместимым с эпистемологическими установками интерпретативной традиции. Сам вопрос о том, как оценивать качество синтетических данных в качественном исследовании, выходит за границы техники. Он затрагивает основания дисциплины и требует определения границы между данными и симуляцией, между инструментом и фальсификацией.

Литература / References

Галкин К. А., Петухова И. С., Парфенова О. А. OPEN AI как помощник при анализе интервью // Социологические исследования. 2025. № 4. С. 105–116. DOI: <https://doi.org/10.7868/S3034601025040095>

Galkin K. A., Petukhova I. S., Parfenova O. A. (2025) OPEN AI as an Assistant in Interview Analysis. *Sotsiologicheskie issledovaniya* [Sociological Studies]. No. 4. P. 105–116. DOI: <https://doi.org/10.7868/S3034601025040095> (In Russ.)

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. С. 64–69. EDN: WULNST

Dzhibilova E. G., Pobываev N. S. (2024) Analysis of Russian and Foreign Experience of Using ChatGPT and Artificial Intelligence in Politics and the Social Sphere. *Sotsialno-gumanitarnye znaniya* [Social and Humanitarian Knowledge]. No. 1. P. 64–69. (In Russ.)

Незговорова М. И. Границы понимания синтетических респондентов: новый контур производства знаний, методологические риски и стандарты проверки данных // Социально-политические науки. 2026. Т. 16. № 1. С. 179–186. DOI: <https://doi.org/10.33693/2223-0092-2026-16-1-179-186> EDN: LRYLSI

Nezgovorova M. I. (2026) Boundaries of Understanding for Synthetic Respondents: A New Knowledge-Production Framework, Methodological Risks, and Data Validation Standards. *Sotsialno-politicheskie nauki* [Sociopolitical Sciences]. Vol. 16. No. 1. P. 179–186. DOI: <https://doi.org/10.33693/2223-0092-2026-16-1-179-186> (In Russ.)

Пузанова Ж. В., Кожоридзе Г. Г., Кожоридзе Д. Г. ИИ и социология: анализ технологических возможностей виртуальных респондентов // Социология: методология, методы, математическое моделирование (Социология:4М). 2025. № 60. С. 216–246. DOI: <https://doi.org/10.19181/4m.2025.34.1.6> EDN: PRPHTP

Puzanova Zh. V., Kozhoridze G. G., Kozhoridze D. G. (2025) AI and Sociology: Analysis of Technological Capabilities of Virtual Respondents. *Sotsiologiya: metodologiya, metody, matematicheskoe*

modelirovanie (Sotsiologiya:4M) [Sociology: Methodology, Methods, Mathematical Modeling (Sociology:4M)]. No. 60. P. 216–246. DOI: <https://doi.org/10.19181/4m.2025.34.1.6> (In Russ.)

Рождественская Е. Ю. Биографический метод в социологии. М.: Изд. дом ВШЭ, 2012.

Rozhdestvenskaya E. Yu. (2012) *Biograficheskiy metod v sotsiologii* [Biographical Method in Sociology]. Moscow: Izd. Dom VShE. (In Russ.)

Рождественская Е. Ю. ИНТЕР-энциклопедия: нарративное интервью // Интеракция. Интервью. Интерпретация. 2020. Т. 12. № 4. С. 114–127. DOI: <https://doi.org/10.19181/inter.2020.12.4.8>

Rozhdestvenskaya E. Yu. INTER-Encyclopedia: Narrative Interview. *Interaktsiya. Intervyu. Interpretatsiya* [Interaction. Interview. Interpretation]. 2020. Vol. 12. No. 4. P. 114–127. DOI: <https://doi.org/10.19181/inter.2020.12.4.8> (In Russ.)

Семенова В. В. Качественные методы: введение в гуманистическую социологию. М.: Добросвет, 1998.

Semenova V.V. (1998) *Kachestvennye metody: vvedenie v gumanisticheskuyu sotsiologiyu* [Qualitative Methods: Introduction to Humanistic Sociology]. Moscow: Dobrosvet. (In Russ.)

Филипова А. Г., Абросимова Е. Е., Зубова О. Г. Метод синтетических фокус-групп в контексте цифровой трансформации социологического исследования // Социодинамика. 2025. № 5. С. 1–17. DOI: <https://doi.org/10.25136/2409-7144.2025.5.74430> EDN: EUMCKL

Filipova A. G., Abrosimova E. E., Zubova O. G. (2025) The Method of Synthetic Focus Groups in the Context of Digital Transformation of Sociological Research. *Sotsiodinamika* [Sociodynamics]. No. 5. P. 1–17. DOI: <https://doi.org/10.25136/2409-7144.2025.5.74430> (In Russ.)

Argyle L.P., Busby E.C., Fulda N., Gubler J.R., Rytting C., Wingate D. (2023) Out of One, Many: Using Language Models to Simulate Human Samples. *Political Analysis*. Vol. 31. No. 3. P. 337–351. DOI: <https://doi.org/10.1017/pan.2023.2>

Bail C. A. (2024) Can Generative AI Improve Social Science? *Proceedings of the National Academy of Sciences*. Vol. 121. No. 21. DOI: <https://doi.org/10.1073/pnas.2314021121>

Barros C.F., Azevedo B.B., Graciano Neto V.V., Kassab M., Kalinowski M., do Nascimento H.A.D., Bandeira M.C.G.S.P. (2025) Large Language Model for Qualitative Research — A Systematic Mapping Study. *arXiv*. DOI: <https://doi.org/10.48550/arXiv.2411.14473>

Bender E.M., Gebru T., McMillan-Major A., Shmitchell S. (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *FAccT'21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. P. 610–623. DOI: <https://doi.org/10.1145/3442188.3445922>

Bisbee J., Clinton J.D., Dorff C., Kenkel B., Larson J.M. (2024) Synthetic Replacements for Human Survey Data? The Perils of Large Language Models. *Political Analysis*. Vol. 32. No. 4. P. 401–416. DOI: <https://doi.org/10.1017/pan.2024.5>

Braun V., Clarke V. (2019) Reflecting on Reflexive Thematic Analysis. *Qualitative Research in Sport, Exercise and Health*. Vol. 11. No. 4. P. 589–597. DOI: <https://doi.org/10.1080/2159676X.2019.1628806>

Davidson T., Karell D. (2025) Integrating Generative Artificial Intelligence into Social Science Research: Measurement, Prompting, and Simulation. *Sociological Methods & Research*. Vol. 54. No. 3. P. 775–793. DOI: <https://doi.org/10.1177/00491241251339184>

Fine G.A., Deegan J.G. (1996) Three Principles of Serendip: Insight, Chance, and Discovery in Qualitative Research. *International Journal of Qualitative Studies in Education*. Vol. 9. No. 4. P. 434–447. DOI: <https://doi.org/10.1080/0951839960090405>

Geertz C. (1973) *The Interpretation of Cultures*. New York: Basic Books.

Grossmann I., Feinberg M., Parker D.C., Christakis N.A., Tetlock P.E., Cunningham W.A. (2023) AI and the Transformation of Social Science Research. *Science*. Vol. 380. No. 6650. P. 1108–1109. DOI: <https://doi.org/10.1126/science.adi1778>

Kapania S., Agnew W., Eslami M., Heidari H., Fox S.E. (2025) Simulacrum of Stories: Examining Large Language Models as Qualitative Research Participants. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. New York: ACM. DOI: <https://doi.org/10.1145/3706598.3713220>

Kuric E., Demcak P., Krajcovic M. (2026) Synthetic Participants Generated by Large Language Models. *A Systematic Literature Review*. DOI: <https://doi.org/10.21203/rs.3.rs-9057643/v1>



Kvale S., Brinkmann S. (2009) *InterViews: Learning the Craft of Qualitative Research Interviewing*. 2nd ed. Thousand Oaks: Sage.

Lincoln Y.S., Guba E.G. (1985) *Naturalistic Inquiry*. Newbury Park: Sage.

Park P.S., Schoenegger P., Zhu C. (2024) Diminished Diversity-of-Thought in a Standard Large Language Model. *Behavior Research Methods*. Vol. 56. No. 6. P. 5754–5770. DOI: <https://doi.org/10.3758/s13428-023-02307-x>

Riessman C.K. (2008) *Narrative Methods for the Human Sciences*. Thousand Oaks: Sage.

Sarstedt M., Adler S.J., Rau L., Schmitt B. (2024) Using Large Language Models to Generate Silicon Samples in Consumer and Marketing Research: Challenges, Opportunities, and Guidelines. *Psychology & Marketing*. Vol. 41. P. 1254–1270. DOI: <https://doi.org/10.1002/mar.21982>

Schroeder H., Aubin Le Quéré M., Randazzo C., Mimno D., Schoenebeck S. (2025) Large Language Models in Qualitative Research: Uses, Tensions, and Intentions. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. DOI: <https://doi.org/10.1145/3706598.3713120>

Shrestha P., Krpan D., Koaik F., Schnider R., Sayess D., Binbaz M.S. (2024) Beyond WEIRD: Can Synthetic Survey Participants Substitute for Humans in Global Policy Research? *Behavioral Science & Policy*. Vol. 10. No. 2. P. 26–45. DOI: <https://doi.org/10.1177/23794607241311793>

Zheng L., Chiang W.-L., Sheng Y., Zhuang S., Wu Z., Zhuang Y., Lin Z., Li D., Xing E.P., Zhang H., Gonzalez J.E., Stoica I. (2023) Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. *Advances in Neural Information Processing Systems*. Vol. 36. P. 46595–46623. DOI: <https://doi.org/10.5555/3666122.3668142>

Сведения об авторе:

Жуликов Кирилл Олегович — магистрант, кафедра экономической социологии и менеджмента социологического факультета, Московский государственный университет имени М. В. Ломоносова, Москва, Россия.
E-mail: Kir.prog2002@gmail.com.

Статья поступила в редакцию: 15.04.2026

Принята к публикации: 30.04.2026

.....
**Evaluation Criteria for Synthetic Respondents
in Qualitative Research: Limits of Applicability**

DOI: 10.19181/inter.2026.18.2.5

Kirill O. Zhulikov

Lomonosov Moscow State University, Moscow, Russia
E-mail: Kir.prog2002@gmail.com

Abstract

Large language models (LLMs) are increasingly used to generate synthetic respondents — textual simulations of informants’ responses based on a specified socio-demographic profile. For quantitative synthetic data, evaluation approaches have already been developed that rely on comparing statistical distributions with the results of large-scale surveys. For qualitative

data, however, established criteria are lacking. In the literature, this absence is described both as a temporary gap that will be filled as methodological tools develop and as a limitation stemming from a deeper mismatch between the foundations of qualitative methodology and the nature of synthetic data. This article provides an analytical review of approaches to evaluating synthetic respondents across three research traditions: computational social science, natural language processing, and qualitative methodology. A review of publications from 2022 to 2026 shows that the most formalized approaches — statistical comparisons of distributions and automated text metrics — measure properties that are secondary to qualitative research. Approaches developed within the qualitative tradition address substantively important properties, such as contextual specificity, internal contradictions, and narrative resistance, but remain at the level of observations and have not been operationalized. Comparing the properties of qualitative data with existing evaluation criteria reveals areas that remain insufficiently covered, including internal contradictions within narratives, speech features, and the capacity of data to generate non-trivial interpretations. On this basis, the article outlines areas in which the use of synthetic respondents in qualitative research appears justified, as well as areas in which it remains problematic. Minimal transparency standards for cases of their use are also discussed.

Keywords: synthetic respondents; synthetic sample; generative artificial intelligence; large language models; qualitative research; transparency standards; limits of applicability of LLM

Author Bio:

Kirill O. Zhulikov — MA student, Department of Economic Sociology and Management, Faculty of Sociology, Lomonosov Moscow State University. **E-mail:** Kir.prog2002@gmail.com.

Received: 15.04.2026

Accepted: 30.04.2026