

5.4.1

Теория, методология и история социологии
Theory, Methodology and History of Sociology

DOI: 10.33693/2223-0092-2026-16-1-179-186

УДК: 316

ГРНТИ: 04.15.07

EDN: LRYLSI



Границы понимания синтетических респондентов: новый контур производства знаний, методологические риски и стандарты проверки данных

М.И. Незговорова ©

Online Market Intelligence (OMI),
г. Москва, Российская Федерация

Фонд Future Research – «Исследования Будущего»,
г. Москва, Российская Федерация

E-mail: mnezgovorova@omirussia.ru

Аннотация. Синтетические респонденты постепенно переходят из статуса экспериментальной техники в регулярный инструмент социальных и маркетинговых исследований. В статье анализируется этот переход и фиксируются методологические риски, границы применимости и стандарты проверки данных в новых исследовательских контурах. Показано, что синтетика способна ускорять разведочные этапы, расширять вариативность качественного анализа, а также поддерживать статистически контролируемое восстановление и усиление количественных массивов, но одновременно усиливает эпистемические ограничения, связанные с зависимостью выводов от источников данных, вариативности профилей, формулировки исследовательского вопроса и процедур валидации. Предложена типология шести разновидностей синтетических респондентов и описана логика переходов между ними: от разговорных персон к моделированию групповой динамики, восстановлению анкет, мультимодальной интерпретации стимулов и построению «цифровых двойников». Сделан вывод, что прикладные практики индустрии (Ipsos; Liverpanel) и решения разработчиков (Yabble; Synthetic Users; Lakmoos), а также академические эксперименты (включая Стэнфордские исследования генеративных агентов) формируют общий язык валидации и ограничений. Отдельно показано, что в российском контуре синтетика институционализируется в UX и продуктовых исследованиях как ускоритель разведки и подготовки гипотез, тогда как в количественной инфраструктуре OMI/Liverpanel фокус смещается к «дополненным синтетическим респондентам» и машинному дозаполнению анкет, обеспечивающему масштабируемую работу с крупными опросными массивами и статистически контролируемую достройку сложных целевых аудиторий с высокой точностью и воспроизводимостью измерений.

© Издательский дом «Юр-ВАК»
© Все права защищены

Ключевые слова: синтетические респонденты, маркетинговые исследования, социальные науки, большие языковые модели, машинное обучение, моделирование поведения, валидизация, цифровые двойники, UX-исследования

ОБРАЗЕЦ ЦИТИРОВАНИЯ: 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: 10.33693/2223-0092-2026-16-1-179-186. EDN: LRYLSI

DOI: 10.33693/2223-0092-2026-16-1-179-186

Boundaries of Understanding for Synthetic Respondents: A New Knowledge-Production Framework, Methodological Risks, and Data Validation Standards

M.I. Nezgovorova ©

Online Market Intelligence (OMI),
Moscow, Russian Federation

Future Research Foundation (“Future Research”),
Moscow, Russian Federation

E-mail: mnezgovorova@omirussia.ru

Abstract. Synthetic respondents are progressively shifting from an experimental technique to a routine instrument in social and marketing research. This article examines this transition and delineates the methodological risks, conditions of applicability, and data-verification standards that define emerging research pipelines. It is shown that synthetic approaches can accelerate exploratory phases, broaden the range of qualitative analysis, and enable statistically controlled reconstruction and augmentation of quantitative datasets; simultaneously, they intensify epistemic constraints arising from dependence on data sources, profile variability, research-question formulation, and validation procedures. The paper proposes a coherent typology of six types of synthetic respondents and explicates the transition logic among them, from conversational personas to simulations of group dynamics, survey reconstruction, multimodal interpretation of stimuli, and the construction of “digital twins.” It concludes that industry practices (Ipsos; Livepanel) and developer solutions (Yabble; Synthetic Users; Lakmoos), together with academic experiments (including Stanford studies on generative agents), are converging on a shared language of validation and articulated limitations. Russian cases further indicate that institutional uptake is strongest in UX and product research as an accelerator of exploratory work and hypothesis formulation, whereas within the quantitative infrastructure of OMI/Livepanel the emphasis shifts toward “augmented synthetic respondents” and machine-based survey imputation, supporting scalable use of large survey datasets and statistically controlled enrichment of complex target audiences with high accuracy and measurement reproducibility.

Key words: synthetic respondents, marketing research, social sciences, large language models, machine learning, behavioral modeling, validation, digital twins, UX research

FOR CITATION: Nezgovorova, M.I. (2026). Boundaries of understanding for synthetic respondents: A new knowledge-production framework, methodological risks, and data validation standards. *Sociopolitical Sciences*, 16(1), 179–186. DOI: 10.33693/2223-0092-2026-16-1-179-186. EDN: LRYLSI

FOR CITATION: Nezgovorova M.I. “Boundaries of Understanding for Synthetic Respondents: A New Knowledge-Production Framework, Methodological Risks, and Data Validation Standards.” *Sociopolitical Sciences*. Vol. 16 no. 1 (2026): 179–186. DOI: 10.33693/2223-0092-2026-16-1-179-186. EDN: LRYLSI

ВВЕДЕНИЕ

Появление синтетических респондентов следует рассматривать как изменение исследовательской инфраструктуры: часть операций, ранее выполнявшихся через прямое взаимодействие с людьми, переносится в пространство моделирования и алгоритмического воспроизводства ответов. Это смещение не отменяет классических требований к валидности и репрезентативности, но перестраивает практику работы с данными, повышая значимость калибровки, прозрачности параметров и сопоставления с полевыми измерениями. Далее рассматривается, как разные разновидности синтетических респондентов формируют новый контур производства знаний и какие методологические риски и стандарты проверки данных оказываются ключевыми.

Синтетический респондент – это ИИ-модель, воспроизводящая поведение, ответы и мотивацию целевой аудитории на основе реальных данных и/или обученных языковых и ML-моделей, способная генерировать диалоговые, количественные и визуально-интерпретационные ответы, имитировать индивидуальные или групповые позиции и при необходимости калиброваться по реальным распределениям и панельным данным. Такое определение снимает упрощение, в котором синтетический респондент сводится к «чат-боту»: исследовательская ценность состоит не в форме интерфейса, а в воспроизводимой модели ответа как социального действия, где важны контекст, ограничения и проверяемость результата. По этой причине обсуждение синтетики в исследовательской среде закономерно ведет к вопросу о границах понимания: где модель действительно воспроизводит устойчивые паттерны аудитории, а где выдает правдоподобные формулировки без достаточной эмпирической опоры.

В социальных и маркетинговых исследованиях «понимание» исторически связывалось с доступом к контексту: к жизненной ситуации респондента, к порядку его аргументации, к прагматике высказывания и к динамике взаимодействия. Классические методы – интервью, наблюдение, фокус-группа, опрос – устроены как процедуры, которые не только извлекают содержание, но и контролируют условия его производства: кто говорит, в какой ситуации, в каких рамках нормы, с какими ожиданиями и рисками. Синтетические респонденты перераспределяют роли: вместо доступа к ситуации ответа исследователь получает доступ к симуляции ответа, сконструированной на основе данных, профиля и алгоритмических правил. Это создает выигрыш в скорости и стоимости на разведочных этапах, но требует усиления проверки, иначе «понимание» подменяется стилистической убедительностью.

МЕТОДОЛОГИЯ И МЕТОДОЛОГИЧЕСКАЯ БАЗА ИССЛЕДОВАНИЯ

Методологическая рамка статьи трактует синтетических респондентов как дополнительный слой моделирования, который встраивается в дизайн ис-

следования поэтапно и требует разведения режимов доказательности. В качестве рабочей схемы используется траектория «усиления опор»: от свободной разговорной симуляции к источниково-ограниченной генерации, затем к мультиагентным сценариям (спектр позиций), далее к количественной достройке и, при наличии устойчивой эмпирической базы, к персонализированным моделям. Такая поэтапность задает переход от правдоподобия формулировок к воспроизводимости и проверяемости результатов на уровне протокола и внешних измерений [7].

Разграничиваются *качественный* и *количественный* контуры. В качественном контуре синтетика используется для воспроизводимой генерации аргументаций и интерпретаций, а критерием качества выступают прозрачность допущений, контролируемость контекста и воспроизводимость выводов при повторении процедуры. В количественном контуре ключевыми становятся сопоставимость волн, воспроизводимость метрик и статистически контролируемая точность, что предполагает калибровку по опросным массивам и систематические тесты на срезах [3].

Переход от свободной симуляции к симуляции, ограниченной источниками, обосновывается работами по retrieval-augmented generation: связывание генерации с извлеченным контекстом повышает прослеживаемость аргументации и воспроизводимость результата, снижая долю необоснованного «общего знания» модели [5]. Архитектурная логика агентных систем (память, извлечение релевантных эпизодов, рефлексия, планирование) используется как основание для описания того, почему контекстная привязка меняет характер симуляции и условия ее валидации [7].

Мультиагентные сценарии трактуются как метод получения распределения позиций и конфликтов интерпретации, а не как «модель среднего мнения». Связка качественного профилирования и количественной проверки валидности задается исследованиями, где агенты, построенные на интервью реальных людей, тестируются на стандартизированных измерителях; валидность операционализируется через сопоставление с человеческой воспроизводимостью при повторном измерении¹.

Методологические риски и требования к протоколу опираются на результаты, показывающие, что синтетические ответы могут сохранять близость средних значений и при этом давать ненадежные выводы для статистического вывода, быть чувствительными к формулировкам и демонстрировать нестабильность распределений. Это задает обязательность фиксации параметров генерации, тестов чувствительности и регулярного сопоставления с внешней эмпирией [2].

Количественная «достройка» массивов описывается в логике статистики пропусков и импутации: корректность обеспечивается явной моделью восстановления,

¹ Park J.S. et al. Generative agent simulations of 1,000 people. 2024. arXiv:2411.10109. URL: <https://arxiv.org/abs/2411.10109> (data of accesses: 14.01.2026).

контролем смещений и учетом неопределенности, где множественная импутация выступает каноническим ориентиром для протоколов воспроизводимого дозополнения [6; 8]. В совокупности методология задает критерии перехода между контурами: каждый следующий шаг должен устранять структурный дефицит предыдущего (контекст, разнообразие, измеримость, связь с полем) и усиливать проверяемость результата в терминах выбранного режима доказательности [3; 7].

ТИПЫ СИНТЕТИЧЕСКИХ РЕСПОНДЕНТОВ И ГРАНИЦЫ ИХ ПРИМЕНИМОСТИ

Практика внедрения синтетики редко начинается с максимально строгих контуров. Типичный путь разворачивается как последовательность усиления опор: от свободной разговорной симуляции к симуляции, ограниченной источниками; от одиночного голоса к распределению позиций; от текстовой формы к количественной достройке; от интерпретации текста к интерпретации визуальных стимулов; от сегментного уровня к персонализированному уровню. Каждый следующий шаг стремится закрыть слабое место предыдущего: недостаток контекста, недостаток разнообразия, недостаток измеримости или слабую связку с «полем».

Внутри этой траектории важно различать два режима. На первом этапе формируются качественные синтетические респонденты: они воспроизводят речь, аргументацию и интерпретации (включая разногласия и альтернативные объяснения), но не претендуют на восстановление статистических распределений. На втором этапе появляются количественные синтетические респонденты: здесь синтетика начинает «держаться» распределения, сопоставимость волн и воспроизводимость метрик, а валидность проверяется через калибровку по опросным массивам и строгие тесты качества.

Первой в этой последовательности обычно становится *ИИ-персона* (качественный контур). Этот тип синтетического респондента представляет собой разговорную модель «типичного представителя» заданной аудитории, которому задаются различающие характеристики (пол, возраст, доход, география, жизненный цикл семьи), а также психографика и элементы потребительского поведения (отношение к категории, драйверы и барьеры, стиль выбора). В прикладной рамке ОМІ синтетический респондент описан как искусственный интеллект, имитирующий поведение представителя целевой аудитории, где большая языковая модель выступает «голосом», а знания подпитываются поведенческими сигналами, отзывами и обучением на материалах реальных исследований². Такой подход удобен как инструмент ранней диагностики: он помогает сформулировать гипотезы, уточнить язык аудитории и выявить потенциальные линии аргументации.

² Незговорова М.И. Синтетические респонденты: от теории к практике: презентация // ОМІ (Online Market Intelligence). 2025. URL: <https://socforum.niioz.ru/wp-content/uploads/2024/11/1.pdf> (дата обращения: 24.01.2026).

Однако именно разговорная персона быстрее всего демонстрирует границы понимания. При слабом профиле и некорректно заданном вопросе результат становится «общим» и усредненным; при высоком доверии к гладкости текста возникает риск принять за эмпирию то, что является рефлексом языковой модели. ОМІ прямо подчеркивает необходимость критической оценки выдачи, предупреждает о возможных галлюцинациях и фиксирует требование последующей проверки на реальных респондентах³. Поэтому в производственном исследовательском контуре следующая задача обычно формулируется как снижение доли свободного рассуждения и усиление доказуемости.

Эту задачу решает *RAG-персона* (качественный контур). Разговорность сохраняется, но ответы ограничиваются корпоративной базой знаний и исследовательскими документами: модель должна опираться на извлеченные фрагменты источников, а не на общую эрудицию. На уровне отраслевых рамок Ipsos выделяет практики, связанные с библиотеками знаний и инструментами, которые соединяют скорость генеративных технологий с требованием доверенного вывода⁴. Методологически это переводит синтетического респондента в режим, где можно отследить, на что опирается аргументация, и где повышается воспроизводимость вывода при повторении запроса.

Сближение прикладных решений с академическими экспериментами особенно заметно в исследованиях, где построение поведения агента зависит от извлечения релевантного контекста. В работе «Generative agents: Interactive simulacra of human behavior» предложена архитектура, в которой агент использует память с последующим извлечением релевантных эпизодов, формирует обобщающие «рефлексии» и строит планы поведения [7]. По сути, это демонстрация того, почему контекстная привязка меняет характер симуляции: без целенаправленного отбора релевантного контекста (когда контекст не подключен вовсе, подключен без фильтрации либо отобран плохо) агент отвечает как усредненная языковая модель; при наличии отбора появляется шанс на контекстную устойчивость и проверяемость.

Тем не менее привязка к источникам не снимает рисков интерпретации. В социальной перспективе это означает, что «заземление» должно дополняться процедурами сопоставления с внешней эмпирией и проверкой переносимости: можно ли применять прошлые данные к новому вопросу, не изменился ли социальный контекст, не перепутаны ли уровни анализа (индивидуальный, сегментный, институциональный). Когда эта дисциплина сформирована, следующий исследовательский запрос касается не столько точности отдельного ответа, сколько разнообразия позиций и воспроизводимости интерпретаций внутри дискуссии.

Так возникает *мультиагентная качественная группа* (качественный контур) – набор синтетических

³ Там же.

⁴ Ipsos. Ipsos Views on AI Topics (directory): July 2025.

респондентов с разными профилями, которые взаимодействуют друг с другом и тем самым производят распределение аргументов, несогласий и линий оправдания, напоминающее фокус-группу или групповую дискуссию. В прикладном поле подобная логика поддерживается поставщиками решений, описывающими синтетических участников как ИИ-«людей», способных отвечать на вопросы и обсуждать тему в группе⁵. В методическом смысле эта форма полезна не как «модель среднего мнения», а как генератор спектра позиций и конфликтов интерпретации, который затем выносится на проверку полем.

В этом же качественном контуре расположен *визуальный синтетический респондент*: он интерпретирует изображение (упаковку, рекламный макет, интерфейс) и формулирует качественную оценку с аргументацией, указывая на смысловые триггеры, когнитивные перегрузы, потенциальные неверные чтения и альтернативные трактовки. Здесь речь не о предсказании опросных KPI, а о быстрой диагностике восприятия и формировании гипотез для последующего тестирования на людях. В российской практике этот сценарий закреплен как отдельная прикладная линия тестирования креативов и упаковки на ранних стадиях, включая предсказательные модели, которые позиционируются как инструмент ускорения качественной оценки (например, «Шагал» Profi Online Research)⁶.

Мультиагентность, однако, не является гарантией валидности. Без строгого протокола она легко превращается в «театрализацию» различий, где роли задаются стилистически, а не эмпирически. Поэтому ключевыми становятся правила задания профилей, ограничения дискуссии, способы фиксации расхождений и алгоритмы агрегации аргументов. Когда же исследовательский контур смещается от реконструкции аргументации к измерению и сопоставимости метрик, появляется иной тип мультиагентности – количественный.

В академической среде наиболее содержательным экспериментом, демонстрирующим именно *количественный поворот мультиагентности*, является работа «Generative agent simulations of 1,000 people» (Стэнфорд), где агенты строятся на основе двухчасовых качественных интервью реальных людей и затем проверяются на способности воспроизводить установки и ответы исходных людей на стандартизированных измерителях⁷. Здесь важно развести уровни: исходные интервью – это качественный материал для построе-

ния профиля, но проверка результата проводится в количественной логике (сопоставление ответов на шкалы и вопросы, входящие в Всеобщий социологический опрос граждан США и оценка совпадений). Авторы показывают, что ответы синтетических агентов на вопросы Всеобщего социального опроса США совпадают с ответами реальных людей примерно так же часто, как совпадают ответы этих же людей при повторном опросе через две недели. Также они отмечают, что если агенты построены на материалах интервью, то точность получается более ровной для разных расовых и идеологических групп, чем у агентов, заданных только набором демографических признаков⁸. Для социальных наук это важно тем, что предлагает язык измерения валидности симуляции и аккуратно задает границу между «правдоподобием» и «измеримой воспроизводимостью».

Таким образом, возникает *мультиагентная панель* (количественный контур) – не «симуляция фокус-группы», а ансамбль профилированных агентов, который используется для генерации распределений ответов на стандартизированные вопросы и для проверки устойчивости выводов при повторных измерениях и вариациях постановки задачи. Это переходная форма между качественными групповыми симуляциями и статистическими контурами усиления данных: она сохраняет профилирование и контекстность, но ставит целью воспроизводимость метрик.

Инструментально эта задача решается через механизм *дополненных синтетических респондентов* (количественный контур), где методы машинного обучения позволяют восстанавливать недостающие ответы, сохраняя как внутреннюю логику исследовательского инструмента, так и устойчивость итоговых распределений.

В материалах Livepanel описан конвейер, включающий отбор качественных анкет, разделение на обучающую и тестовую выборки, автоматический поиск большого числа признаков (250+), подбор множества моделей по каждому вопросу, балансировку редких классов и проверку качества на большом числе срезов⁹. Параллельно в близком количественном контуре работает и OMI (в связке с Livepanel), развивая практики восстановления и достройки опросных массивов на основе машинного моделирования ответов, калибровки распределений и многоуровневой валидации по срезам. В материалах OMI/Livepanel этот класс решений позиционируется как инфраструктурный элемент количественных исследований: он направлен на повышение полноты и сопоставимости данных, включая реконструкцию пропусков и уточнение сложных целевых аудиторий при сохранении требований к воспроизводимости метрик и контролю смещений.

⁸ Там же.

⁹ Материалы по восстановлению данных и синтетическим ответам (конвейер машинного дозополнения анкет): презентационные материалы // Livepanel. 2025. URL: <https://n-infinite.livepanel.ai/Survey%20augmentation%20with%20Synthetic%20Responses.%20N-Infinite%20by%20Livepanel%202024%20Benchmark.pdf> (дата обращения: 24.01.2026).

⁵ Описание продукта и подхода к синтетическим участникам исследований // Synthetic Users. URL: <https://www.syntheticusers.com/> (дата обращения: 14.01.2026).

⁶ Предсказательная модель... «Шагал» (описание проекта «Шагал») // Profi Online Research. URL: <https://profiresearch.ru/tpost/axtrpheng1-proekt-goda-2024> (дата обращения: 14.01.2026). Основание: сообщение компании о модели «Шагал» как предсказательном инструменте для оценки рекламных материалов/дизайнов/упаковки.

⁷ Park J.S. et al. Generative agent simulations of 1,000 people. 2024. arXiv:2411.10109. URL: <https://arxiv.org/abs/2411.10109> (data of accesses: 14.01.2026).

Существенной особенностью названа чувствительность к ветвлению анкеты («эффект домино»), когда ошибка на раннем вопросе может изменить траекторию последующих предсказаний¹⁰.

В этом же статистическом контуре возможна визуальная компонента, но она устроена иначе, чем качественная визуальная интерпретация. Креативы декомпозируются средствами распознавания изображений и извлечения признаков (класс решений компьютерного зрения, включая распространенные пайплайны на базе TensorFlow), после чего эти признаки используются как вход для обучения модели, предсказывающей значения опросных KPI. Валидность такого подхода держится на парном соответствии «креатив – опросный замер»: каждому креативу должен соответствовать набор анкетных данных, а KPI-вопросы должны быть одинаковыми для сопоставляемых кейсов (например, привлекательность, понятность, связь с брендом и т.п.). Тогда задача формулируется как предсказательная: обучить модель на визуальных признаках так, чтобы она воспроизводила ожидаемые опросные оценки без повторного опроса по каждому новому креативу, а затем подтверждать переносимость на новых материалах через контроль ошибок и калибровку.

Методологически это наиболее «жесткий» класс синтетики: здесь меньше пространства для правдоподобия и больше пространства для измеримости.

Граница понимания проходит по линии распределений: синтетические дозаполнения способны повысить устойчивость оценок и расширить возможности анализа, но одновременно могут незаметно смещать редкие сегменты, усиливать перекосы исходных данных или «сглаживать» локальные аномалии. Поэтому требование калибровки под реальные распределения и проверки на множестве срезов становится условием корректности вывода¹¹.

Наконец, наиболее амбициозный и одновременно наиболее требовательный тип – *цифровой двойник*. Это персонализированная модель респондента или сегмента, откалиброванная на полевых и панельных данных так, чтобы можно было задавать новые вопросы без возврата в поле и получать ответы в числовом, категориальном и открытом формате с обоснованием. В отраслевых обсуждениях цифровых двойников (в частности, в материалах Ipsos о ИИ-двойниках) центральным становится вопрос о границе между воспроизведением установок/решений и глубиной/эмпатией как компонентой социального понимания¹².

Стэнфордский протокол построения агентов на основе интервью можно рассматривать как академический аналог цифровых двойников: при наличии эм-

пирической базы появляются измеримые показатели точности и смещений, а сравнение ведется с человеческой воспроизводимостью¹³. Однако соблазн «заменить поле симуляцией» создает серьезный риск: симуляция может выглядеть убедительно, но приводит к иной структуре выводов. В статье «Synthetic replacements for human survey data? The perils of large language models» показано, что даже при близости средних оценок к опросным средним модель может менять выводы в объяснительных моделях и быть чувствительной к формулировке роли и параметрам запроса [2]. Это делает необходимыми строгие протоколы проверки переносимости и отчетность о параметрах генерации.

Российские практики демонстрируют, что синтетика институционализируется прежде всего в двух зонах. Первая – UX и продуктовые исследования как ускорение разведочных этапов. Материал Interforums описывает практический кейс применения синтетических респондентов в проверке продуктовой гипотезы (Точка Банк), включая симуляцию интервью, сопоставление синтетических и реальных ответов и фиксирование областей совпадений и расхождений¹⁴. Вторая – встраивание синтетики в количественную инфраструктуру через восстановление и усиление массивов; здесь показателен конвейер OMI/Livepanel с многоуровневой проверкой качества¹⁵.

ЗАКЛЮЧЕНИЕ

Таким образом, несколько типов синтетических респондентов образуют не перечень «функций», а карту компромиссов и переходов. ИИ-персоны, RAG-персоны, мультиагентные качественные группы и визуальные интерпретаторы сильны как инструменты генерации гипотез и ранней диагностики; мультиагентные панели переводят симуляцию в режим воспроизводимых метрик; дозаполнения анкет обеспечивают статистически контролируемое усиление данных; цифровые двойники обещают расширение поля вопросов при наличии строгой эмпирической базы. Граница понимания проходит там, где заканчивается контролируемая реконструкция и начинается подмена эмпирики правдоподобием. Следовательно, синтетический респондент в социальной науке должен пониматься как инструмент интерполяции в явных условиях применимости, а не как автономный носитель человеческого смысла.

¹³ Park J.S. et al. Generative agent simulations of 1,000 people. 2024. arXiv:2411.10109. URL: <https://arxiv.org/abs/2411.10109> (data of accesses: 14.01.2026).

¹⁴ Синтетические респонденты: конец классических UX-исследований? Практический кейс Точка Банк // Interforums. 28.11.2025. URL: <https://interforums.ru/blog/sinteticheskie-respondentyi-konecz-klassicheskix-ux-issledovanij> (дата обращения: 14.01.2026).

¹⁵ Материалы по восстановлению данных и синтетическим ответам (конвейер машинного дозаполнения анкет): презентационные материалы // Livepanel. 2025. URL: <https://n-infinite.livepanel.ai/Survey%20augmentation%20with%20Syntetic%20Responses.%20N-Infinite%20by%20Livepanel%202024%20Benchmark.pdf> (дата обращения: 24.01.2026).

¹⁰ Материалы по восстановлению данных и синтетическим ответам (конвейер машинного дозаполнения анкет): презентационные материалы // Livepanel. 2025. URL: <https://n-infinite.livepanel.ai/Survey%20augmentation%20with%20Syntetic%20Responses.%20N-Infinite%20by%20Livepanel%202024%20Benchmark.pdf> (дата обращения: 24.01.2026).

¹¹ Там же.

¹² Ipsos. Ipsos views on AI topics (directory): July 2025.

Эпистемологическая специфика синтетического респондента состоит в том, что он производит ответы в модальности «как если бы», но без участия субъекта, который несет ответственность за высказывание. В классической методологии это соответствует различию между интерпретацией и доказательством: интерпретация допустима как гипотеза, тогда как доказательство требует процедуры сборки фактов и контроля условий их получения. Синтетика усиливает интерпретационную мощь исследования (быстрее перебирает объяснения, помогает выявить конкурирующие рамки), но не отменяет необходимости доказательного контура, особенно там, где выводы используются для управленческих решений.

Применение синтетических респондентов рационально рассматривать как включение дополнительного слоя моделирования в общий дизайн исследования. Такой слой может выполнять четыре функции: генерацию гипотез; предварительную проверку инструментария; расширение пространства сценариев; поддержку интерпретации. Каждая функция требует собственной формы отчетности: фиксировать профиль и параметры генерации, указывать источники контекста, документировать тесты качества и описывать случаи расхождения с полем. Без этого синтетические результаты не являются воспроизводимыми и не могут быть корректно включены в научный дискурс.

Вопрос валидности синтетики следует отличать от вопроса правдоподобия. Правдоподобие – это соответствие ожиданиям читателя и стилистическая связность; валидность – это проверяемое соответствие внешней эмпирии или устойчивым закономерностям. Стэнфордские эксперименты показывают, что валидность можно операционализировать через сравнение с устойчивостью при повторном измерении у людей и через сравнение с известными шкалами и исходами¹⁶. В прикладном количественном контуре валидность выражается через калибровку распределений и контроль ошибок на срезах¹⁷.

Смещения и каскадные ошибки являются структурными рисками. В опросных симуляциях без калибровки модель склонна сглаживать редкие позиции и усиливать наиболее частотные паттерны. В дозаполнениях анкет существует каскад ветвления: ранняя ошибка меняет траекторию последующих предсказаний¹⁸. Эти риски вынуждают включать в протокол отдельную проверку редких сегментов и чувствительность результатов к формулировке вопросов. Работа в журнале «Политический анализ» показывает, что чувствитель-

ность результатов к настройкам и формулировкам может менять выводы в объяснительных моделях, даже если средние показатели и распределения на первый взгляд выглядят близкими [2].

Российские кейсы показывают, что наиболее жизнеспособной оказывается стратегия «композиции методов»: синтетика используется как ускоритель и расширитель, а поле – как контур окончательной проверки. Кейс Точка Банка, описанный на Interforums, демонстрирует именно такую композицию: синтетика помогает подготовить гипотезы и вопросы, а затем сопоставляется с интервью реальных пользователей¹⁹. В количественном контуре конвейер OMI/Livepanel показывает, как синтетика может быть встроена в инфраструктуру измерения при наличии строгой проверки качества²⁰.

ЛИТЕРАТУРА/REFERENCES

1. Argyle L.P., Busby E.C., Fulda N. et al. Out of one, many: Using language models to simulate human samples. *Political Analysis*. 2023. Vol. 31. No. 3. Pp. 337–351. DOI: 10.1017/pan.2023.2.
2. Bisbee J., Clinton J.D., Dorff C. et al. Synthetic replacements for human survey data? The perils of large language models. *Political Analysis*. 2024. Vol. 32. Pp. 401–416. DOI: 10.1017/pan.2024.5.
3. Deville J.-C., Särndal C.-E. Calibration estimators in survey sampling. *Journal of the American Statistical Association*. 1992. Vol. 87. No. 418. Pp. 376–382. DOI: 10.1080/01621459.1992.10475217.
4. Jansen B.J., Jung S.-g., Salminen J. Employing large language models in survey research. *Natural Language Processing Journal*. 2023. Vol. 4. Art. 100020. DOI: 10.1016/j.nlp.2023.100020.
5. Lewis P., Perez E., Piktus A. et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*. 2020. Vol. 33. arXiv: 2005.11401.
6. Little R.J.A., Rubin D.B. *Statistical analysis with missing data*. 2nd ed. Hoboken: Wiley, 2002. 408 p. ISBN: 978-0-471-18386-0.
7. Park J.S., O'Brien J.C., Cai C.J. et al. Generative agents: Interactive simulacra of human behavior. Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23). 2023. DOI: 10.1145/3586183.3606763.
8. Rubin D.B. *Multiple imputation for nonresponse in surveys*. New York: Wiley, 1987. 284 p. ISBN: 978-0-471-08705-2.
9. Valenzuela S., Winter S., Rivera S. Using large language models for survey research in communication: Opportunities and challenges. *Communication and Change*. 2025. Vol. 1. Art. 14. DOI: 10.1007/s44382-025-00014-z.

¹⁶ Park J.S. et al. Generative agent simulations of 1,000 people. 2024. arXiv:2411.10109. URL: <https://arxiv.org/abs/2411.10109> (data of accesses: 14.01.2026).

¹⁷ Материалы по восстановлению данных и синтетическим ответам (конвейер машинного дозаполнения анкет): презентационные материалы // Livepanel. 2025. URL: <https://n-infinite.livepanel.ai/Survey%20augmentation%20with%20Syntetic%20Responses.%20N-Infinite%20by%20Livepanel%202024%20Benchmark.pdf> (дата обращения: 24.01.2026).

¹⁸ Там же.

¹⁹ Синтетические респонденты: конец классических UX-исследований? Практический кейс Точка Банк // Interforums. 28.11.2025. URL: <https://interforums.ru/blog/sinteticheskie-respondentyi-konecz-klassicheskix-ux-issledovanij> (дата обращения: 14.01.2026).

²⁰ Материалы по восстановлению данных и синтетическим ответам (конвейер машинного дозаполнения анкет): презентационные материалы // Livepanel. 2025. URL: <https://n-infinite.livepanel.ai/Survey%20augmentation%20with%20Syntetic%20Responses.%20N-Infinite%20by%20Livepanel%202024%20Benchmark.pdf> (дата обращения: 24.01.2026).

Статья проверена программой Антиплагиат. Оригинальность – 88,15%

Рецензент: *Шашкин А.В.*, кандидат социологических наук; генеральный директор; Online Market Intelligence (OMI)

Статья поступила в редакцию 27.01.2026, принята к публикации 25.02.2026

The article was received on 27.01.2026, accepted for publication 25.02.2026

СВЕДЕНИЯ ОБ АВТОРЕ

Незговорова Мария Игоревна, кандидат социологических наук; директор по инновациям, Online Market Intelligence (OMI); г. Москва, Российская Федерация; член совета; фонд Future Research – «Исследования Будущего»; г. Москва, Российская Федерация. РИНЦ Author ID: 937774; SPIN-код: 7775-4760; E-mail: mnezgovorova@omirussia.ru

ABOUT THE AUTHOR

Maria I. Nezhgovorova, Cand. Sci. (Sociol.); Innovation Director, Online Market Intelligence (OMI); Moscow, Russian Federation; Member of the Board; Future Research Foundation (“Future Research”); Moscow, Russian Federation. RSCI Author ID: 937774; SPIN-code: 7775-4760; E-mail: mnezgovorova@omirussia.ru