

Социодинамика

Правильная ссылка на статью:

Филипова А.Г., Абросимова Е.Е., Зубова О.Г. Метод синтетических фокус-групп в контексте цифровой трансформации социологического исследования // Социодинамика. 2025. № 5. DOI: 10.25136/2409-7144.2025.5.74430 EDN: EUMCKL URL: https://nbpublish.com/library_read_article.php?id=74430

Метод синтетических фокус-групп в контексте цифровой трансформации социологического исследования

Филипова Александра Геннадьевна

ORCID: 0000-0002-7475-1961

доктор социологических наук

заведующий лабораторией комплексных исследований детства; Владивостокский государственный университет
старший научный сотрудник; Российский государственный педагогический университет им. А.И. Герцена

191186, Санкт-Петербург, набережная реки Мойки 48

✉ Alexgen77@list.ru

Абросимова Евгения Евгеньевна

ORCID: 0000-0001-8893-4550

кандидат социологических наук

доцент; кафедра философии и юридической психологии; Владивостокский государственный университет

690014, Россия, Приморский край, г. Владивосток, Ленинский р-н, ул. Гоголя, д. 41

✉ gaijony@mail.ru

Зубова Оксана Геннадьевна

ORCID: 0000-0003-2825-7935

кандидат политических наук

доцент; кафедра социальных технологий; Московский государственный университет имени М.В. Ломоносова

119234, Россия, г. Москва, ул. Ленинские Горы, 1, стр. 33

✉ zubovaoksana@bk.ru

[Статья из рубрики "Социальные исследования и мониторинг"](#)

DOI:

10.25136/2409-7144.2025.5.74430

EDN:

EUMCKL

Дата направления статьи в редакцию:

13-05-2025

Дата публикации:

21-05-2025

Аннотация: Предметом исследования является использование синтетических фокус-групп (СФГ), создаваемых с помощью больших языковых моделей (LLM), в рамках цифровой трансформации социологического анализа. Объектом исследования выступают инновационные методы сбора и интерпретации качественных данных, ориентированные на моделирование групповой дискуссии без участия реальных респондентов. В статье рассматривается эвристический потенциал СФГ, возможности применения метода в условиях ограниченного доступа к респондентам, а также его релевантность для изучения отношения к искусственному интеллекту в сфере высшего образования. Особое внимание уделяется способам организации взаимодействия с языковыми моделями через систему промптов, формированию ролей участников и анализу полученных дискурсивных позиций. В статье также затрагиваются этические и методологические вызовы, возникающие при использовании синтетических участников в исследовательских целях. Методология сочетает теоретический анализ научных публикаций с практическим моделированием СФГ, реализованным с помощью последовательных промптов к языковым моделям и последующей интерпретацией генерируемых данных. Научная новизна работы заключается в апробации метода синтетических фокус-групп (СФГ) как инновационного инструмента сбора качественной информации с использованием генеративных языковых моделей (LLM), таких как Gemini, Qwen, Llama, Deepseek, и Mistral. В исследовании особое внимание уделено моделированию СФГ на основе пользовательских промптов, направленных на изучение отношения студентов и преподавателей к искусственному интеллекту в системе высшего образования. Все использованные модели подчеркивали методологические ограничения СФГ: возможность искажения данных, необходимость четко обозначать синтетическую природу участников и дополнять такие исследования традиционными методами. Подобные методы могут быть полезны в ситуациях, требующих быстрого генерирования гипотез, предварительного тестирования исследовательских сценариев, а также в образовательной и экспертной практике. Вместе с тем, подчеркивается необходимость критического подхода к использованию СФГ, особенно в контексте валидности и репрезентативности получаемых данных.

Ключевые слова:

искусственный интеллект, синтетические фокус-группы, большие языковые модели, цифровая социология, персоны пользователей, предвзятость, качественные методы, социологическое исследование, виртуальная фокус группа, моделирование нескольких агентов

Введение

Цифровизация открывает новые возможности для социологических исследований, используются такие технологии, как облачные вычисления, большие данные, Интернет

вещей и искусственный интеллект. Это позволяет не только быстрее собирать и обрабатывать большие массивы данных, структурировать их в компактную и доступную форму через визуализацию, но и применять новые исследовательские инструменты, связанные как с переносом в онлайн-среду, например, виртуальный формат фокус-групп включает в участие людей, проживающих в самых отдаленных территориях, так и с расширенными возможностями опций искусственного интеллекта, проникающего во все сферы общественной жизни. Это определяет новые вызовы и форматы работы исследователей, порождает этические проблемы и вопросы, связанные с валидностью и надежностью получаемых данных. Возникающие риски, тем не менее, не исключают эвристический потенциал новых исследовательских техник и методов, требующих расширенной практической апробации и дальнейшего осмысления преимуществ и недостатков.

Проблемы занятости респондентов, фокусирование на сензитивных темах, работа с труднодоступными группами актуализируют задачу разработки новых методов сбора эмпирических данных. В ситуации пандемии Ковид-19 исследователи стали активнее использовать онлайн инструмент, в т.ч. онлайн интервью, онлайн фокус-группы при том как синхронные, так и асинхронные. Методный арсенал современного социолога/маркетолога/политолога и других специалистов, работающих с людьми, дополнили методы работы с искусственно сгенерированными, обычно с использованием больших языковых моделей "агентами".

Современные методы моделирования – синтетические фокус-группы на базе больших языковых моделей (LLM) – позволяют не только экономически и этически безопасно моделировать поведенческие сценарии, но и анализировать динамику мнений в изменяющемся контексте.

Целью настоящего исследования стало выявление возможностей и ограничений в использовании синтетических фокус-групп, созданных с помощью больших языковых моделей, в контексте социологических исследований. Основные задачи связаны с описанием современных подходов к применению генеративных моделей в социальном знании; проведением сравнительного анализа различных LLM (больших языковых моделей), применённых в эксперименте; оценкой исследовательского потенциала синтетических фокус-групп для решения задач сбора, интерпретации и верификации эмпирических данных; раскрытием основных рисков и ограничений применения синтетических агентов.

Научная новизна изучения синтетических фокус-групп, созданных с помощью больших языковых моделей как инновационного инструмента социологического исследования, заключается в теоретическом обосновании использования данного вида фокус-групп в прикладной социологии; сопоставлении результатов, полученных от разных LLM, в рамках одних и тех же исследовательских вопросов.

Обзор литературы

Зарубежные публикации демонстрируют интерес исследователей к синтетическим фокус-группам, генерации персонажей для решения не только научных, но и производственных задач, описывают возникающие сложности и ограничения.

В работе «Focus Agent: LLM-Powered Virtual Focus Group» представлена программная платформа «Focus Agent», на основе большой языковой модели (LLM), которая синтезирует фокус-группу (для сбора данных), а также выступает в качестве модератора в фокус-группе с реальными людьми. Исследователи заключают, что Focus Agent может

генерировать мнения, похожие на мнения реальных людей. Ограничения данной технологии – сложности с распознаванием намеков и контекста в обсуждениях. Это приводит к повторению вопросов и недостаточной динамичности дискуссии. «Focus Agent» не проявляет истинного понимания, что снижает его восприятие как полноценного модератора, а также не всегда эффективно взаимодействует с менее активными участниками, что может приводить к снижению общей вовлеченности в обсуждение [\[1\]](#).

Работа Deus Ex Machina and Personas from Large Language Models: Investigating the Composition of AI-Generated Persona Descriptions также посвящена большим языковым моделям (LLM), которые могут генерировать персоны на основе подсказок, описывающих целевую группу пользователей [\[2\]](#). Результаты исследования показывают, что LLM могут генерировать достаточно правдоподобные персоны, но при этом демонстрируют небольшой уровень стереотипизации (возраст, профессия и пр.)

Исследование Shin, J., Hedderich, M. A., Rey, B [\[3\]](#) показывает, что сочетание усилий экспертов и больших языковых моделей (LLM) позволяет создавать более репрезентативных персонажей, чем использование только одного из подходов. Авторы предлагают различные подходы к распределению задач между людьми и LLM для повышения качества создаваемых персонажей, включая использование LLM для суммирования, группировки и выделения ключевых характеристик данных пользователей.

Sattelle, V., и Carlos Ortiz, J. выявляют риски использования больших языковых моделей и моделей преобразования текста в изображение (TTIM), обсуждая их применение в дизайнерской практике. Авторы описывают ряд недостатков использования данных технологий, выделяя предвзятость и стереотипизации (LLM могут усиливать стереотипы, например, гендерные роли). Исследователи подчеркивают, что, несмотря на большой потенциал инструментов искусственного интеллекта в сфере дизайна, роль человека остается приоритетной - дизайнеры должны анализировать контекст и корректировать «промты», фильтровать результаты, учитывая ограничения ИИ [\[4\]](#).

Взаимодействие дизайнеров с персонализированными чат-ботами – «синтетическими пользователями» [\[5\]](#) было направлено на изучение влияния этих интеракций на процесс проектирования, результаты и восприятие персонажей. По мнению авторов, синтетические пользователи предоставили новые возможности для взаимодействия, однако, результаты не продемонстрировали явного преимущества по сравнению с традиционными персонажами в плане эмпатии, понимания и качества идей.

Австрийские исследователи погружают сгенерированных персонажей в медицинский контекст [\[6\]](#). В своей работе они описывают инструментарий «Personas for AI» с руководством по пятиэтапному процессу создания персонажей (идентификация пользовательских групп, сбор данных, анализ, создание основы, визуализация), шаблоны для визуализации персонажей, примеры из медицинской практики. Статья подчеркивает необходимость человеко-ориентированного подхода в эпоху сложных ИИ-систем, где прозрачность и доверие становятся критическими факторами успеха.

Потенциал языковых моделей для социальных наук описывают и другие исследователи [\[7, 8, 9\]](#), уделяя внимание моделированию человеческого поведения через LLM и подчеркивая важность осторожности в этическом и практическом применении [\[10\]](#).

Для отечественной практики синтетические фокус-группы являются новым методом сбора данных, пока недостаточно представленным в социологической практике и, соответственно, недостаточно изученным. В связи с этим возникает необходимость в апробации данного подхода, оценке его потенциала и ограничений, а также в разработке методологических решений, адаптированных к условиям использования генеративных языковых моделей.

Проблема валидности и надёжности данных синтетических фокус-групп

Г. Татарова и И. Чеховский на основе экспертных оценок определили несколько групп факторов, влияющих на результативность получаемых данных методом фокус-групп, что подтверждается и в других исследованиях, прежде всего, это некорректно поставленные цели и задачи; неправильно подобранная целевая аудитория из-за ошибок выборки или ошибок в анкете-фильтре, процессе рекрутинга участников; некорректный инструментарий [\[11\]](#).

При использовании синтетических фокус-групп на основе больших языковых моделей (LLM) возникают вопросы как конструктивной валидности (в какой степени агенты отражают реальные социальные установки и поведенческие паттерны целевых групп), так и надёжности (повторяемости и устойчивости полученных данных при повторной генерации).

Особую обеспокоенность вызывает возможность искажения установок под влиянием контекста, особенностей промпта или модели. Это ставит под угрозу несколько видов валидности.

Конструктивная валидность — может быть снижена, если агенты отражают не эмпирически обоснованные установки, а стереотипы, заимствованные из обучающего корпуса модели. Салминен с коллегами задаются вопросами: Насколько разнообразны характеристики персон, созданных LLM? Есть ли какие-либо заметные предубеждения? [\[2\]](#)

Если персоны, сгенерированные LLM, не содержат разнообразных характеристик, существует риск того, что будут упущены маргинальные типы пользователей [\[12\]](#), поскольку они не будут представлены в сгенерированных персонах. Более того, даже если персоны будут разнообразны в своем представлении различных типов пользователей, персоны все равно могут быть предвзятыми в том смысле, что они чрезмерно подчеркивают определенные характеристики на фоне других [\[13\]](#). Например, распределение персон может быть преимущественно мужским или преимущественно молодым.

В исследовании Салминен и др. для генерации 450 персон используется GPT-4. Были созданы три типа подсказок для пяти зависимостей (алкоголь, опиоиды, социальные сети, онлайн-шопинг и азартные игры): одна с указанием мужского пола, одна с указанием женского пола и одна без указания пола вообще. Такое сочетание дало 15 комбинаций (3 × 5). Далее генерация для каждой из 15 комбинаций была повторена 30 раз. Исследователи подчеркивают, что ввод одного и того же запроса несколько раз через API Open AI в GPT-4 приводил к почти идентичным персонам. Эта проблема была решена с помощью двухэтапной стратегии подсказок: сначала модель сгенерировала список из 30 «базовых» персон для каждой комбинации типа зависимости-подсказки. Эти уникальные короткие описания персон были затем введены обратно в модель для получения полного описания. [\[2\]](#)

Содержательная валидность — страдает в случаях, когда синтетические реплики не охватывают весь спектр значимых тем или позиций, представленных в реальной популяции. Паоли иллюстрирует, что LLM могут создавать персоны пользователей на основе тематического анализа полуструктурированных интервью с реальными пользователями. LLM может генерировать коды и темы из данных интервью, а затем использовать их для написания повествований персон, которые включают цели, предысторию, потребности, проблемы и другие соответствующие детали [\[14\]](#).

Внешняя валидность — ограничивается, если агенты не представляют разнообразие демографических, культурных или региональных особенностей, особенно в русскоязычном контексте, где качество генерации ниже.

Одним из способов повышения достоверности выступает повторная генерация данных с вариативными промтами (prompt rephrasing), использование нескольких моделей, а также сравнение с результатами реальных интервью и фокус-групп. Кроме того, важным инструментом является экспертная валидация: участие специалистов в предметной области, которые могут оценить правдоподобие, релевантность и полноту высказываний агентов.

Также используются методики калибровки и валидации: например, независимая кодировка данных несколькими исследователями и проверка на интересубъективность, а также смешанные методы, совмещающие синтетические и реальные данные. Такие меры помогают повысить доверие к синтетическим данным и заложить основу для выработки этических и методологических стандартов в этой новой области социологического анализа.

Методы исследования

В исследовании используются общенаучные методы (анализ, синтез, индукция, моделирование, сравнение), а также специальные методы (использование LLM как инструмента моделирования, построение системного промпта).

В построении промптов мы следуем пяти принципам, обозначенным в работе Д. Феникса и М. Тейлора: задайте направление (подробно опишите желаемый стиль или укажите соответствующую ему личность); укажите формат (определите правила, которым должна следовать модель, и структуру ответа); приведите примеры; оцените качество (найдите ошибки, оцените ответы и определите, что влияет на качество ответа); разделите задачу на подзадачи (чтобы повысить точность и управляемость генерации). Эти принципы особенно важны при работе с большими языковыми моделями (LLM), где форма запроса напрямую влияет на результат [\[15\]](#).

Термины zero-shot и few-shot относятся к стратегиям использования языковых моделей в зависимости от количества предоставляемых примеров. В zero-shot сценарии модель выполняет задачу, опираясь только на описание инструкции без примеров. В few-shot режиме модель получает один или несколько примеров, чтобы распознать закономерности и структуру задания. Это помогает точнее формировать контекст и повышает релевантность отклика [\[16\]](#).

Цепочка рассуждений (Chain of Thought) — это техника промптинга, при которой модель побуждается к пошаговому объяснению логики, прежде чем выдать итоговый ответ. Такая стратегия позволяет улучшить качество откликов, особенно в задачах, требующих аналитического мышления, интерпретации, принятия решений или логического вывода.

Для активации этой техники используются такие фразы, как «Давай рассуждать пошагово» или «Объясни, как ты пришел к этому выводу» [\[17\]](#).

Кроме того, вариативность реплик модели можно настраивать с помощью параметров генерации, таких как `temperature` (отвечает за степень креативности: чем выше значение, тем более разнообразными и непредсказуемыми становятся ответы) и `top_p` (настраивает вероятность выбора токенов — от более «безопасных» к более разнообразным, в зависимости от суммарной вероятности). Управление этими параметрами позволяет добиться нужного стиля или степени вариативности ответов при генерации реплик в фокус-группе или другом смоделированном взаимодействии.

Для изучения возможностей синтетических фокус-групп и параллельной оценки возможностей языковых моделей мы обращаемся как к научным статьям, так и к языковым моделям.

В качестве отправной точки был сформулирован исследовательский вопрос, на основе которого разработан промпт, задающий системную роль языковой модели: «Ты — эксперт в области социологических методов, применяемых к исследованию социальных практик с использованием искусственного интеллекта. Твоя задача — объяснять и разрабатывать инновационные методики, такие как синтетические фокус-группы, включая возможности их использования в гуманитарных и прикладных исследованиях. Ты хорошо разбираешься в методах генерации фокус-групповых данных с помощью языковых моделей, симуляции ролей, автоматического анализа дискурсов, а также умеешь обосновать научную и практическую значимость таких подходов. Ты учиываешь этические аспекты, ограничения и перспективы использования синтетических участников в исследовательских целях».

Далее пользовательские роли задавались последовательно, с учетом целей исследования и в соответствии с логикой генерации предыдущих ответов. В ходе диалога с моделью были сформулированы следующие уточняющие запросы:

- Что такое синтетическая фокус-группа (СФГ)?
- Как можно использовать СФГ для исследования ИИ в высшем образовании?
- Как сформировать виртуальных персон?

Доступ к языковым моделям осуществлялся через OpenRouter. Использовались 5 языковых моделей, характеристики которых приведены в таблице 1. Все они бесплатны.

Таблица 1 – Сравнительная характеристика используемых LLM

Название модели	разработчик	Дата выпуска	Размер модели	Контекстное окно	Другие особенности
Gemini 2.0 Flash Experimental	Google	11.12.2024	Не указано	1M токенов	Высокая скорость, интеграция с Google, мультимодальность
DeepSeek R1 Zero	DeepSeek	06.03.2025	671B, 37B активны	164K токенов	Самообучение, Mixture-of-Experts, открытый код
Qwen2.5 Coder 7B	Qwen	15.04.2025	7B	33K токенов	Структурированные ответы,

Instruct					улучшенное программирование
Llama 3.1 Nemotron Ultra 253B v1	nvidia	08.04.2025	253B	128K токенов	Научные и математические задачи, оптимизация под NVIDIA.
Mistral Small 3.1 24B	Mistral	17.03.2025	24B	128K токенов	Локальное развертывание, мультимодальность

Результаты исследования и их обсуждение

Что такое синтетическая фокус-группа

По определению С. А. Белановского, фокус-группа — это групповое глубокое интервью, относящееся к качественным методам социологического исследования. [\[18, С. 162-163\]](#).

Виды фокус-групп можно соотносить с типами модерирования, количеством участников, продолжительностью и режимами работы при очной встрече или дистанционно. При дистанционном формате исследователи часто сравнивают именно критерий групповой дискуссии, как важную составляющую успешного проведения фокус-группы [\[19\]](#).

С внедрением искусственного интеллекта можно говорить о новом типе – синтетической фокус-группе. Чтобы выделить ее особенности, преимущества и недостатки, обратимся к осмыслению понимания данной категории через ответы языковых моделей.

В целом ответы моделей совпадают в том, что это инновационный метод качественных исследований, имитирующий традиционную фокус-группу с помощью искусственного интеллекта и языковых моделей, где участниками выступают не реальные респонденты, а синтетические агенты, моделирующие целевую группу. *«Эти агенты запрограммированы для имитации поведения, мнений, реакций и взаимодействия, характерных для целевой группы, которую хотят исследовать»* (Gemini). Выделяются и особенности данного типа фокус-групп, связанные с назначением и возможностями применения.

Так, исходя из ответа DeepSeek, синтетическая фокус-группа предваряет реальную фокус-группу и может использоваться для тестирования гайда и гипотез: *«Основные особенности включают моделирование мнений и поведения, использование для тестирования гипотез до проведения реальных фокус-групп, а также снижение этических рисков»*.

Mistral дополняет, что виртуальные агенты позволяют исследователям получать данные в условиях, где трудно найти реальных участников. Вероятно, речь идет о труднодоступных группах респондентов. *«В отличие от традиционных фокус-групп, где участвуют реальные люди, синтетические фокус-группы используют алгоритмы и языковые модели для генерации отзывов, обсуждений и мнений, что позволяет исследователям получать данные в условиях, где трудно найти реальных участников, или когда необходимо проводить множество экспериментов для тестирования гипотез»*.

Gemini подчеркивает, что основное назначение синтетических фокус-групп в изучении социальных процессов, выявлении закономерностей и получении исследовательских инсайтов как предварительного этапа виртуального проведения исследования без обращения к реальным респондентам. *«Основная идея заключается в том, чтобы создать*

виртуальную среду, в которой можно изучать социальные процессы, выявлять закономерности и получать инсайты без необходимости проведения реальных фокус-групп, которые требуют времени, ресурсов и могут быть подвержены различным искажениям».

Qwen видит в синтетических фокус-группах методологический инструмент, позволяющий изучать социальные процессы без искажений и ограничений, возникающих при проведении реальных фокус-групп. *«Виртуальные участники, или синтетические агенты, могут быть настроены на моделирование определённых групп или индивидуумов, что позволяет исследователям изучать сложные социальные процессы, не сталкиваясь с ограничениями реальных фокус-групп».*

Llama обобщенно дает характеристики и видит главное отличие от традиционной фокус-группы в том, что участниками выступают виртуальные агенты (синтетические личности), созданные на основе алгоритмов машинного обучения. *«В отличие от традиционных фокус-групп, где участники из реальной жизни обсуждают заданную тему под руководством модератора, в синтетической фокус-группе участниками выступают виртуальные агенты (синтетические личности), созданные на основе алгоритмов машинного обучения. Эти агенты способны генерировать ответы, имитирующие человеческое поведение и мнения, что позволяет исследователям собирать и анализировать данные без необходимости очного взаимодействия с людьми».*

Итак, дадим определение синтетической фокус-группе – это инновационный метод, имитирующий традиционную фокус-группу с помощью искусственного интеллекта и языковых моделей, где участниками выступают не реальные респонденты, а синтетические агенты, моделирующие изучаемую целевую группу с помощью имитации поведения, мнений, реакции и взаимодействия.

Назначение метода: апробация инструментария, выдвижение гипотез на предварительном этапе подготовки социологического исследования; изучение труднодоступных групп респондентов; снижение этических рисков, связанных с контактной работой реальных участников; сокращение финансовых, организационных и временных затрат на проведение фокус-групповых исследований; возможность неоднократного тестирования, корректировки выборки.

Основное ограничение связано с трансформацией основных функций метода, когда синтетические фокус-группы можно рассматривать больше как методический, а не исследовательский инструмент, требующий неоднократной проверки. Технически распространение синтетических фокус-групп ограничено несовершенством используемых языковых моделей, долгосрочным и материально затратным процессом их обучения и тестирования. Преимущества и ограничения LLM приведены в таблице 2.

Таблица 2 - Преимущества и ограничения использования больших языковых моделей (LLM) в социологических исследованиях

Преимущества	Ограничения
Снижение трудозатрат на транскрибацию, кодировку, категоризацию данных	Низкая валидность: генерация данных не эквивалентна реальному эмпирическому опыту
Быстрая генерация сценариев, гипотез и исследовательских инструментов	Отсутствие социального контекста: модели не обладают опытом, эмоциями, мотивацией
	Риски предвзятости: модели могут

Возможность моделирования труднодоступных групп	воспроизводить социальные искажения и стереотипы
Гибкость в настройке под целевые группы (язык, стиль, культурный код)	Иллюзия эмпиричности: синтетические данные могут быть ошибочно интерпретированы
Этическая и экономическая безопасность при работе с чувствительными темами	Необходимость высокой квалификации в области промпт-инжиниринга и критического анализа

СФГ для изучения особенностей применения ИИ в высшем образовании

Следующий промпт, адресованный отобранным моделям, был пользовательским запросом о генерации СФГ для изучения особенностей применения ИИ в высшем образовании. Отдельные LLM описали все этапы исследования – от постановки исследовательского вопроса/задачи, заканчивая анализом полученных данных, а не просто ограничились описанием организации синтетической фокус-группы. Самый краткий ответ был получен от Deepseek. Все модели предупреждали об ограничениях, связанных с СФГ – смещениями (bias), имеющимися в данных, на которых были обучены модели; «важно четко указывать, что участники симуляции являются синтетическими, и не использовать результаты для принятия решений, которые могут повлиять на реальных людей» (Gemini), «синтетические фокус-группы не должны имитировать реальных людей без их согласия» (Qwen), «ясно указывать, что данные получены с помощью синтетических методов» (Llama), «результаты синтетической фокус-группы следует интерпретировать с осторожностью и дополнять реальными исследованиями» (Deepseek), «данные, сгенерированные ИИ, могут частично изобретаться или не полностью соответствовать реальным человеческим моделям» (Mistral). Llama предложила для повышения надежности данных, полученных посредством СФГ, триангулировать их данными из традиционных методов исследования (например, онлайн-опросов или интервью).

Модель “Gemini” сгенерировала несколько исследовательских тем относительно применения ИИ в высшем образовании: оценка отношения студентов к ИИ-ассистентам; выявление проблем и возможностей использования ИИ с точки зрения преподавателей, администраторов и студентов; прогнозирование влияния ИИ на успеваемость и мотивацию студентов; тестирование новых ИИ-инструментов (например, можно оценить, как студенты воспринимают чат-ботов, помогающих в учебе, или системы автоматической проверки эссе).

Генеративные агенты/персоны

Для создания профилей виртуальных персон, в нашем случае – студентов, важно собрать данные о целевой аудитории: демографическую информацию (возраст, пол, география и т.д.), учебные характеристики (курсы, учебная нагрузка, успеваемость и т.д.), уровень технической грамотности, опыт использования ИИ в образовательном процессе (Mistral).

Для сбора информации можно использовать разные методы – анкетирование, глубинные интервью или фокус-группы для получения более детальных ответов, наблюдение. Рекомендация от Qwen: «соберите данные, которые помогут вам понять типичные характеристики и поведение вашей целевой аудитории (студентов)». Llama предлагает использовать методы содержательного анализа или NLP для выявления тем, шаблонов и мнений, которые будут основой для создания реалистичных виртуальных персон.

Собранные данные станут основой для проектирования профилей виртуальных личностей, которые потом будут подаваться в промтах большим языковым моделям. Пример структуры промта от Gemini:

«Ты - студент [курс], [специальность] в [университет/тип университета]. Твой уровень владения технологиями [высокий/средний/низкий]. Твой опыт использования ИИ в образовании: [да/нет, описание опыта, если есть]. Твое отношение к ИИ в образовании: [позитивное/нейтральное/скептическое с обоснованием]. Твои основные цели в учебе: [описание целей]. Твои основные опасения в связи с использованием ИИ в образовании: [описание опасений]. Твой стиль общения: [формальный/неформальный, аргументированный/эмоциональный, с примерами]» .

Gemini уточняет: «чем детальнее промт, тем более реалистичным будет виртуальный участник. Укажите конкретные примеры опыта, мнений, целей. Можно указать известные личности (вымышленные или реальные), на которых виртуальный участник должен быть похож в плане стиля общения или взглядов».

В качестве сгенерированных персон каждая модель привела свои примеры. Наиболее подробное описание дала Gemini:

«Ты — Анна, студентка 3-го курса факультета журналистики МГУ. У тебя средний уровень владения технологиями, ты активно пользуешься социальными сетями и мессенджерами. Ты еще не сталкивалась с использованием ИИ в образовании, но слышала о нем от друзей. Тебе интересно, как ИИ может помочь в учебе, но ты опасаясь, что это приведет к обезличиванию образования и снижению роли преподавателя. Ты считаешь важным, чтобы в образовании сохранялся человеческий контакт и творческий подход. Ты немного скептически относишься к идее, что ИИ может заменить живого преподавателя, особенно в гуманитарных науках. Когда тебя спрашивают о твоём мнении, ты отвечаешь развернуто, приводя примеры и аргументы, но твой тон скорее дружелюбный и открытый к дискуссии».

После генерации виртуальных персон необходимо провести их тестирование, чтобы убедиться в их реалистичности и разнообразии. Можно провести экспертную оценку и/или пилотное исследование с участием небольшой группы реальных студентов, чтобы сравнить их ответы с ответами виртуальных персон. (Qwen)

После этого наступает этап подготовки сценария для синтетической фокус-группы, в котором наряду с блоками вопросов важно предусмотреть симуляции взаимодействия участников между собой и с модератором.

Заключение

Исследование не только демонстрирует потенциал синтетических фокус-групп в прикладных целях, но и поднимает важные вопросы, связанные с верификацией данных, качеством моделирования социальных ролей и границами применимости ИИ в качественных социологических методах. Это подчеркивает необходимость дальнейшей рефлексии относительно статуса подобных инструментов — не как замены, а как дополнения к традиционным подходам, особенно в контексте этики, методологической надежности и эпистемологической прозрачности.

Синтетическую фокус-группу можно определить через имитацию традиционной фокус-группы с помощью искусственного интеллекта и языковых моделей, где участниками выступают не реальные респонденты, а синтетические агенты, воспроизводящие

изучаемую целевую группу посредством имитации поведения, мнений, реакции и взаимодействия. Основное назначение метода заключается в апробации инструментария, выдвижении рабочих гипотез, корректировке выборки, в изучении труднодоступных групп респондентов и сокращении финансовых, организационных и временных затрат на проведение.

В научном дискурсе остается открытой дискуссия о возможностях использования данного метода сбора данных в качественных исследованиях. Существующие ограничения в применении связаны как со вспомогательным назначением метода и необходимостью неоднократной проверки полученных данных, так и с несовершенством используемых языковых моделей, долгосрочным и материально затратным процессом их обучения и тестирования, недостаточной обученностью моделей на русскоязычном корпусе текстов.

Использование пяти больших языковых моделей и сравнение их ответов на вопросы, связанные с синтетическими фокус-группами, позволило сделать выбор в пользу Gemini, как предоставляющей более развернутые и аргументированные данные. Однако экосистема больших языковых моделей быстро развивается: регулярно выходят обновления существующих моделей, таких как GPT, Claude или Mistral, и появляются принципиально новые решения от разных разработчиков. В связи с этим итоговый выбор модели для дальнейшей работы с СФГ остается открытым.

Библиография

1. Zhang T., Zhang X., Cools R., Simeone A. L. Focus Agent: LLM-Powered Virtual Focus Group // ACM International Conference on Intelligent Virtual Agents (IVA '24). Glasgow, United Kingdom, September 16-19, 2024. New York: ACM, 2024. 10 p. DOI: 10.1145/3652988.3673918.
2. Salminen J., Liu C., Pian W., et al. Deus Ex Machina and Personas from Large Language Models: Investigating the Composition of AI-Generated Persona Descriptions // Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24). Honolulu, HI, USA, May 11-16, 2024. New York: ACM, 2024. 20 p. URL: <http://dx.doi.org/10.1016/j.cpr.2009.10.006>.
3. Shin J., Hedderich M. A., Rey B., et al. Understanding Human-AI Workflows for Generating Personas // Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS 2024) / ed. by A. Vallgarda et al. 2024. P. 757-781. URL: <https://doi.org/10.1145/3643834.3660729>.
4. Chuang Y.-S., Goyal A., Harlalka N., Suresh S. Simulating Opinion Dynamics with Networks of LLM-based Agents // Findings of the Association for Computational Linguistics: NAACL 2024. DOI: 10.18653/v1/2024.findings-naacl.211.
5. Gu H., Chandrasegaran S., Lloyd P. Synthetic Users: Insights from Designers' Interactions with Persona-Based Chatbots // Artificial Intelligence for Engineering Design, Analysis and Manufacturing. 2025. Vol. 39. Art. e2. P. 1-17. DOI: 10.1017/S089006042400002X.
6. Holzinger A., Kargl M., Kipperer B., et al. Personas for Artificial Intelligence (AI) an Open Source Toolbox // IEEE Access. 2022. Vol. 10. P. 23732-23747. DOI: 10.1109/ACCESS.2022.3154776. EDN: HBQBBR.
7. Argyle L. P., Busby E., Wingate D., [и др.] Out of One, Many: Using Language Models to Simulate Human Samples [Электронный ресурс] // Political Analysis. 2022. URL: https://www.researchgate.net/publication/363584667_Out_of_One_Many_Using_Language_Models_to_Simulate_Human_Samples (дата обращения 12.05.2025).
8. Epstein J. M. Generative Social Science: Studies in Agent-Based Computational Models [Электронный ресурс]. 2006. Режим доступа:

https://www.researchgate.net/publication/267118857_Generative_Social_Science_Studies_in_Agent-Based_Computational_Models (дата обращения: 10.05.2025).

9. Lorenz J., Neumann M. Individual attitude change and societal dynamics: Computational experiments with psychological theories // OSF Preprints. 2020. URL: <https://doi.org/10.31234/osf.io/ebfvr> (дата обращения 12.05.2025).

10. Sattelle V., Ortiz J. C. Generating User Personas with AI: Reflecting on its Implications for Design // DRS2024: Boston / ed. by C. Gray et al. Boston, USA, June 23-28, 2024. URL: <http://dx.doi.org/10.21606/drs.2024.1024>.

11. Татарова Г. Г., Чеховский И. В. Метод фокус-групп: экспертные оценки "факторов неуспешности" // Социология: методология, методы, математическое моделирование. 2012. № 34. С. 5-31. EDN: PEOKEH.

12. Goodman-Deane J., Waller S., Demin D., González-de-Heredia A., Bradley M., Clarkson J. P. Evaluating Inclusivity using Quantitative Personas // Proceedings of the Design Research Society Conference 2018. Limerick, Ireland, June 25-28, 2018. Limerick: DRS, 2018. URL: <https://doi.org/10.21606/drs.2018.400>.

13. Chapman C., Love E., Milham R. P., ElRif P., Alford J. L. Quantitative Evaluation of Personas as Information // Proceedings of the Human Factors and Ergonomics Society Annual Meeting. September 1, 2008. Vol. 52, No. 16. P. 1107-1111. URL: <https://doi.org/10.1177/154193120805201602>.

14. Paoli S. Writing User Personas with Large Language Models: Testing Phase 6 of a Thematic Analysis of Semi-Structured Interviews. arXiv preprint arXiv:2305.18099. 2023. 30 p. URL: <https://arxiv.org/abs/2305.18099>.

15. Феникс Дж., Тейлор М. Промт-инжиниринг для GenAI: паттерны надёжных запросов для качественных результатов. Астана: Спринт Бук, 2025. 432 с.

16. Brown T., Mann B., Ryder N., et al. Language Models are Few-Shot Learners // Advances in Neural Information Processing Systems (NeurIPS 2020). Vancouver, Canada, December 6-12, 2020. Red Hook, NY: Curran Associates, Inc., 2020. 23 p. URL: <https://arxiv.org/abs/2005.14165>.

17. Wei J., Wang X., Schuurmans D., et al. Chain of Thought Prompting Elicits Reasoning in Large Language Models // Advances in Neural Information Processing Systems (NeurIPS 2022). New Orleans, USA, November 28 - December 9, 2022. Red Hook, NY: Curran Associates, Inc., 2022. 21 p. URL: <https://arxiv.org/abs/2201.11903>.

18. Белановский С. А. Глубокое интервью и фокус-группы. Москва: ИНФРА-М, 2019. 352 с. EDN: WONHVD.

19. Мельникова О. Т., Нестерова Е. М. Особенности групповой дискуссии в очных и онлайн фокус-группах // Вестник Московского университета. Сер. 14. Психология. 2024. Т. 47. № 1. С. 184-206. DOI: 10.11621/LPJ-24-08. EDN: RMUJCC.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Предметом исследования в представленной статье является метод синтетических фокус-групп в контексте цифровой трансформации социологического исследования.

В качестве методологии предметной области исследования в данной статье были использованы дескриптивный метод, метод категоризации, метод анализа, а также, как отмечается в статье, были применены языковые модели.

Актуальность статьи не вызывает сомнения, поскольку, благодаря активному развитию информационных технологий, цифровизация быстрыми темпами внедряется практически

во все сферы общественной жизни и деятельности. Искусственный интеллект и новейшие языковые модели находят широкое применение во многих областях, в том числе и в проведении научных исследований с использованием различных методов. В этом контексте изучение метода синтетических фокус-групп в контексте цифровой трансформации социологического исследования представляет научный интерес в сообществе ученых.

Научная новизна статьи заключается в проведении по авторской методике изучения возможностей метода синтетических фокус-групп в контексте цифровой трансформации социологического исследования, а также применения языковых моделей. В рамках исследования был поставлен исследовательский вопрос и сформулирован запрос для определения системной роли языковой модели.

Статья написана языком научного стиля с применением в тексте исследования изложения научных позиций различных ученых по изучаемой проблеме и научной терминологии и дефиниций, характеризующих предмет исследования, а также подробным описанием авторской методики проведения исследования.

Структуру статьи, в целом можно считать выдержанной с учетом основных требований, предъявляемых к написанию научных статей. В структуре данного исследования представлены такие элементы, как введение, обзор литературы, методы исследования, результаты исследования и их обсуждение, заключение и библиография.

Содержание статьи отражает ее структуру. В частности, особую ценность представляет определение, описание и характеристики синтетической фокус-группы, а также особенностей ее применения в науке и высшем образовании.

Библиография содержит 12 источников, включающих в себя отечественные и зарубежные периодические и непериодические издания, а также электронные ресурсы.

В статье приводится описание научных подходов, характеризующих возможности применения синтетических фокус-групп в исследованиях, а также возможности использования современных языковых моделей и искусственного интеллекта. В статье содержится апелляция к различным научным трудам и источникам, посвященным этой тематике, которая входит в круг научных интересов исследователей, занимающихся указанной проблематикой.

В представленном исследовании содержатся выводы, касающийся предметной области исследования. В частности, был выявлен и описан потенциал синтетических фокус-групп для науки, в том числе для проведения социологических исследований.

Материалы данного исследования рассчитаны на широкий круг читательской аудитории, они могут быть интересны и использованы учеными в научных целях, педагогическими работниками в образовательном процессе, социологами, консультантами, аналитиками и экспертами.

В качестве недостатков данного исследования следует отметить, что кроме очень краткого заключения по проведенному исследованию целесообразно было бы сформулировать более подробные и развернутые выводы по проведенному исследованию. В целом хотелось бы обратить внимание на то, что объем авторского текста в представленной рукописи не позволяет максимально полно и всесторонне раскрыть заявленную актуальную тему исследования. Указанные недостатки не снижают научную и практическую значимость самого исследования, однако их необходимо оперативно устранить и несколько дополнить текст статьи в плане его объема. Рукопись рекомендуется отправить на доработку.

Результаты процедуры повторного рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Представленная на рецензирование статья «Метод синтетических фокус-групп в контексте цифровой трансформации социологического исследования» посвящена актуальной теме использования современных технологий (облачные вычисления, большие данные, Интернет вещей и искусственный интеллект и др.) в социологических исследованиях.

С одной стороны, это «позволяет не только быстрее собирать и обрабатывать большие массивы данных, структурировать их в компактную и доступную форму через визуализацию, но и применять новые исследовательские инструменты, связанные как с переносом в онлайн-среду, например, виртуальный формат фокус-групп включает в участие людей, проживающих в самых отдаленных территориях, так и с расширенными возможностями опций искусственного интеллекта, проникающего во все сферы общественной жизни».

С другой, «определяет новые вызовы и форматы работы исследователей, порождает этические проблемы и вопросы, связанные с валидностью и надежностью получаемых данных.

Все это, по мнению автора, требует расширенной практической апробации и дальнейшего осмысления преимуществ и недостатков новых методов сбора эмпирических данных, в том числе: метода синтетических фокус-групп на базе больших языковых моделей (LLM).

Автор опирается на хорошую теоретическую научную базу исследовательских результатов по синтетическим фокус-группам, генерации персонажей для решения не только научных, но и производственных задач.

В статье автор дает сравнительную характеристику используемых LLM (в табличном варианте). В заключении делает вывод в пользу одной из них.

И делает общий вывод по применению LLM, отмечая, что «в научном дискурсе остается открытой дискуссия о возможностях использования данного метода сбора данных в качественных исследованиях. Существующие ограничения в применении связаны как со вспомогательным назначением метода и необходимостью неоднократной проверки полученных данных, так и с несовершенством используемых языковых моделей, долгосрочным и материально затратным процессом их обучения и тестирования, недостаточной обученностью моделей на русскоязычном корпусе текстов».

Библиографический список включает 12 источников. На наш взгляд, необходимо его расширение минимум до 15-17 источников.

Статья изложена хорошим языком, логична, структурирована.

Вместе с тем, считаем необходимым дать следующие рекомендации автору по доработке статьи:

1. В начале статьи автор правомочно задается вопросом, связанным с валидностью и надежностью получаемых данных при использовании синтетических фокус-групп. Но дальнейшего раскрытия темы мы не увидели. В связи с чем, необходимо раскрыть ответ на заданный в начале статьи вопрос.

2. Необходимо подробно описать теоретико-методологическую базу исследования – дополнить методы исследования, разделив их на общенаучные и специальные.

Детализировать специальные методы исследования – особенно, в части создания системного промпта, который передается языковой модели (LLM) для задания контекста и ограничений при генерации текста. На наш взгляд, важно прописать роль, какую должен взять на себя LLM при генерации текста, руководящие принципы - набор правил или инструкций, которым должен следовать LLM, тематические ограничения и др. особенности.

Рекомендуем статью «Метод синтетических фокус-групп в контексте цифровой трансформации социологического исследования» к публикации после доработки по нашим замечаниям.

Результаты процедуры окончательного рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Развитие искусственного интеллекта, генеративных моделей, языковых моделей и их применение в различных областях деятельности обуславливает интерес ученых, социологов, политологов, конфликтологов и многих других специалистов в области социальной сферы к изучению возможностей искусственного интеллекта и цифровых технологий в исследовании социальных явлений. Цифровые технологии сегодня активно используются для отработки рутинных задач в рамках проведения социологических исследований – статистической и аналитической обработки результатов опросов, транскрибации и стенографирования глубинных интервью и фокус-групп, обобщения, сопоставления, кластеризации ответов респондентов и др. Эта автоматизация процессов социологических исследований существенно снизила трудозатраты социологов. С развитием искусственного интеллекта и генеративных моделей расширяются перспективы его использования в формировании уже синтетических фокус групп как моделей реальных фокус-групп. Представленная статья посвящена перспективам использования искусственного интеллекта для формирования синтетических фокус-групп при проведении социологических исследований. Заголовок статьи соответствует содержанию.

Цель исследования автором не сформулирована, текст статьи необходимо дополнить.

В статье выделены разделы с подзаголовками, что соответствует представленным требованиям журнала «Социодинамика».

Во «Введении» автор обосновывает актуальность и значимость исследования, акцентируя внимание на преимуществах, рисках и проблемах расширения использования современных методов моделирования — синтетических фокус-групп на базе больших языковых моделей (LLM). Во Введении отсутствуют обязательные для научных статей формулировки цели, задач, предмета и объекта исследования, его теоретической и практической значимости. Раздел «Обзор литературы» посвящен характеристике имеющихся научных работ в области использования искусственного интеллекта для проведения социологических исследований. Раздел «Проблема валидности и надёжности данных синтетических фокус-групп» продолжает литературный обзор в части описания соответствующей проблемы и рисков искажения результатов исследований при использовании синтетических фокус-групп. В разделе «Методы исследования» автор описывает применяемые методы, логику проведения исследования, формулирует вопросы для искусственного интеллекта – синтетической фокус-группы, а также дает сравнительную характеристику примененных в рамках исследования больших языковых моделей - Gemini 2.0 Flash Experimental, DeepSeek R1 Zero, Qwen2.5 Coder 7B Instruct, Llama 3.1 Nemotron Ultra 253B v1, Mistral Small 3.1 24B. В разделе «Результаты исследования и их обсуждение» автор выделяет три подраздела, которые сосредоточены на описании результатов применения синтетических фокус-групп в рамках получения ответов на вопросы о том, что такое синтетическая фокус-группа, как она может быть использована для изучения особенностей применения искусственного интеллекта в высшем образовании, каким образом можно генерировать и тестировать

виртуальных персон. В разделе «Заключение» автор подводит итоги исследования, резюмируя, что на сегодняшний момент наиболее широкие перспективы открывает использование синтетических фокус-групп «в апробации инструментария, выдвижении рабочих гипотез, корректировке выборки, в изучении труднодоступных групп респондентов и сокращении финансовых, организационных и временных затрат на проведение».

В исследовании использованы известные общенаучные методы: анализ, синтез, сравнение, восхождение от абстрактного к конкретному, логический метод и т.д., а также специфические методы исследования – использование LLM как инструмента моделирования, построение системного промпта.

Статья содержит 1 таблицу, посвященную сравнительной характеристике использованных в рамках исследования моделей. В качестве рекомендации предлагаем автору обратить внимание на возможности иллюстрации результатов исследования, в т.ч. в рамках обобщения информации о преимуществах и проблемах использования синтетических фокус-групп, а также иллюстрации алгоритма социологического исследования с их применением.

Выбранная тема исследования актуальна. Расширение использования искусственного интеллекта характерно в настоящее время для всех отраслей деятельности и областей знаний. В социологических исследованиях, несмотря на возможности автоматизации и сокращения трудозатрат исследователей, важно соблюсти все необходимые процедуры для обеспечения достоверности и качества результатов. Поиск этого баланса открывает широкие горизонты для изучения.

Статья обладает практической значимостью в части предложений по развитию использования синтетических-фокус групп в социологических исследованиях.

Научная новизна исследования автором не сформулирована. Тем не менее представляется, что исследование обладает элементами приращения научного знания.

Стиль статьи является научным и в целом соответствует требованиям журнала.

Библиография представлена 19 источниками, в первую очередь зарубежных авторов. На все источники имеются ссылки по тексту.

К преимуществам статьи следует отнести следующее. Во-первых, актуальность и значимость выбранного направления исследования. Во-вторых, наличие разработанных авторских предложений по формированию алгоритма генерации синтетической фокус-группы в социологическом исследовании. В-третьих, наличие сравнительной характеристики моделей для исследования и результатов их апробации.

Заключение. Представленная статья посвящена перспективам использования искусственного интеллекта для формирования синтетических фокус-групп при проведении социологических исследований. Статья может быть принята к публикации в журнале «Социодинамика».