

УДК 159.9

DOI: 10.34670/AR.2026.94.20.017

Риски применения больших языковых моделей в когнитивном моделировании социальных процессов

Смирнов Олег Аркадьевич

Кандидат физико-математических наук, доцент,
Российский государственный университет им. А.Н. Косыгина,
115035, Российская Федерация, Москва, ул. Садовническая, 52/45;
e-mail: smirnovoleg1952@mail.ru

Елагина Анна Сергеевна

Кандидат экономических наук, доцент,
кафедра экономических дисциплин,
Еврейский университет,
127273, Российская Федерация, Москва, ул. Отрадная, 6;
e-mail: yelagina.anna@gmail.com

Аннотация

В работе представлен систематический концептуальный анализ рисков, сопряжённых с применением больших языковых моделей в когнитивном моделировании социальных процессов. Стремительное внедрение данных вычислительных систем в исследовательскую практику открывает новые эвристические возможности для обработки текстовых массивов, выявления скрытых статистических закономерностей и генерации гипотез, однако сопровождается комплексом эпистемологических, методологических, институциональных и этических проблем, требующих критического осмысления. В центре внимания находятся онтологический статус модельных конструкторов и сопряжённая с ним проблема реификации, методологические ограничения операционализации социальных конструкторов через языковые корреляты, корреляционная природа моделей, исключая каузальный анализ, а также систематические смещения, обусловленные структурой обучающих корпусов и их культурно-исторической спецификой. Дополнительно рассматриваются проблемы интерпретации результатов и смещения уровней теоретического обобщения, институциональные риски депрофессионализации и эрозии методологического скептицизма, сложности воспроизводимости и документирования исследований, а также этические аспекты, связанные с конфиденциальностью, информированным согласием и потенциальной дискриминацией. На основе проведённого анализа формулируются стратегии смягчения выявленных рисков, включая развитие культуры прозрачности, междисциплинарный диалог, триангуляцию методов, внедрение этических кодексов и сохранение фундаментальных исследовательских компетенций. Делается вывод о том, что языковые модели могут служить эффективным инструментом познания лишь при условии ответственного и критического подхода, сочетающего техническую компетентность с методологической рефлексией и этической рефлексивностью, а также при сохранении различия между вычислительными репрезентациями и эмпирической социальной реальностью.

Для цитирования в научных исследованиях

Смирнов О.А., Елагина А.С. Риски применения больших языковых моделей в когнитивном моделировании социальных процессов // Психология. Историко-критические обзоры и современные исследования. 2026. Т. 15. № 7А. С. 202-212. DOI: 10.34670/AR.2026.94.20.017

Ключевые слова

Большие языковые модели, когнитивное моделирование, социальные процессы, эпистемологические риски, операционализация конструкторов, каузальный анализ, внешняя валидность, культурное смещение, воспроизводимость, депрофессионализация, этика искусственного интеллекта, методологический скептицизм, междисциплинарность.

Введение

Прогресс в области искусственного интеллекта, выразившийся в создании больших языковых моделей, существенно трансформировал методологический ландшафт когнитивных наук [Giabbanelli, 2026; Mou et al., 2026]. Эти вычислительные системы, демонстрирующие способность обрабатывать и генерировать текстовые массивы беспрецедентного объёма и сложности, открывают новые перспективы для исследования фундаментальных механизмов человеческого мышления, поведенческих паттернов и социальных взаимодействий [Jiang et al., 2026; Larooij, Törnberg, 2025]. Вместе с тем, имплементация подобных инструментов в теоретически нагруженную область когнитивного моделирования социальных процессов сопровождается комплексом эпистемологических, методологических и этических проблем, требующих систематического критического анализа [Larooij, Törnberg, 2025; Wu et al., 2025]. Статья представляет собой концептуальное исследование указанных рисков, рассматриваемых в контексте применения больших языковых моделей для моделирования социальных феноменов, с акцентом на их онтологические основания, операциональные ограничения, проблемы валидности и институциональные последствия [Mou et al., 2026; Taillandier et al., 2025].

Привлекательность больших языковых моделей для когнитивного моделирования социальных процессов обусловлена их значительными вычислительными свойствами [Jiang et al., 2026; Roy et al., 2026]. Данные системы демонстрируют способность выявлять статистические закономерности в естественном языке, который выступает в качестве основного носителя социального знания, культурных норм, ценностных ориентаций и индивидуальных различий [Hu et al., 2025; Schwager et al., 2026]. Языковые модели могут применяться для экстракции семантических структур из обширных текстовых корпусов, реконструкции дискурсивных практик, симуляции коммуникативных актов и генерации правдоподобных поведенческих сценариев [Mou et al., 2026; Zhang et al., 2026]. Данные возможности создают впечатление, что подобные системы способны выполнять функции своеобразных «вычислительных феноменологов», реконструирующих субъективный опыт и социальные значения на основе их языковых репрезентаций [Choi, Young, Ferrara, 2026]. Однако именно это служит источником рисков, систематический анализ которых составляет основное содержание настоящей работы [Larooij, Törnberg, 2025; Wu et al., 2025].

Основная часть

Наиболее фундаментальный риск применения больших языковых моделей в когнитивном моделировании социальных процессов связан с их онтологическим статусом и природой тех

конструктов, которые они порождают [Larooij, Törnberg, 2025; Choi, Young, Ferrara, 2026]. В методологии когнитивного моделирования принципиально значимым является различие между формальной моделью социальной реальности и самой социальной реальностью, между вычислительной репрезентацией феномена и его эмпирическим содержанием [Giabbanelli, 2026]. Большие языковые модели представляют собой статистические распределения языковых единиц, аппроксимированные на значительных текстовых корпусах [Mou et al., 2026]. Данные системы не обладают интенциональностью, ментальными состояниями, аффективными переживаниями или социальными нормами в том смысле, в каком эти атрибуты присущи человеческим агентам [Choi, Young, Ferrara, 2026; Larooij, Törnberg, 2025]. Тем не менее, их способность генерировать связные и контекстуально релевантные высказывания порождает иллюзию понимания и квазисоциального интеллекта, что создаёт эпистемически опасную ситуацию [Wu et al., 2025].

Данная иллюзия может приводить к реификации модельных конструктов, то есть к приписыванию им онтологической реальности, которой они объективно не обладают [Larooij, Törnberg, 2025]. Исследователь, применяющий языковую модель для изучения социальных установок, аттитюдов или процессов коллективного принятия решений, подвергается риску интерпретировать выходные данные модели как прямое отражение социальной реальности, тогда как фактически он имеет дело со статистической рекомбинацией языковых паттернов, унаследованных из обучающей выборки [Choi, Young, Ferrara, 2026; Hu et al., 2025]. Данная проблема усугубляется непрозрачностью современных моделей, функционирующих как «чёрные ящики», чья внутренняя организация не поддаётся прямой интерпретации [Giabbanelli, 2026]. В этих условиях исследователь не может установить, каким образом входные данные преобразуются в результирующие выходы и какие промежуточные репрезентации возникают в процессе вычислительной обработки [Larooij, Törnberg, 2025]. Отсутствие прозрачности в сочетании с когнитивной иллюзией понимания создаёт эпистемически уязвимую среду, в которой метафоры и аналогии, лежащие в основе функционирования модели, начинают восприниматься как объяснительные принципы социальной реальности, что ведёт к концептуальной путанице и теоретическим заблуждениям [Wu et al., 2025; Mou et al., 2026].

Онтологический риск непосредственно связан с проблемой операционализации социальных конструктов в условиях использования языковых моделей [Hu et al., 2025; Schwager et al., 2026]. Когнитивное моделирование социальных процессов опирается на теоретически обоснованные конструкты, такие как социальное доверие, групповую идентичность, нормативную регуляцию, аттитюды, стереотипы, ценностные ориентации и другие [Mou et al., 2026]. Данные конструкты обладают сложной историей эмпирической операционализации, включающей разработку валидизированных опросников, экспериментальных парадигм и поведенческих индикаторов [Larooij, Törnberg, 2025]. Применение больших языковых моделей предлагает альтернативный способ операционализации, основанный на языковых коррелятах конструктов, а именно частоте и контексте употребления терминов, семантической близости к другим понятиям, тональности и эмоциональной окраске высказываний, содержащих соответствующие конструкты [Choi, Young, Ferrara, 2026; Jin et al., 2026]. Данный подход, однако, сопряжён с серьёзными методологическими ограничениями [Larooij, Törnberg, 2025].

Во-первых, языковые репрезентации социальных конструктов в обучающих корпусах исторически и культурно детерминированы; они отражают не универсальные социальные законы, а конкретные дискурсивные практики, сложившиеся в определённых обществах в конкретные исторические периоды [Hu et al., 2025; Choi, Young, Ferrara, 2026]. Модель

неизбежно воспроизводит эти исторически преходящие и культурно специфические значения, придавая им статус статистически устойчивых закономерностей [Larooij, Törnberg, 2025]. Во-вторых, социальные конструкты характеризуются значительной вариативностью и контекстуальной зависимостью: одно и то же понятие может приобретать существенно различные значения в зависимости от социального контекста, коммуникативной ситуации, статусно-ролевых позиций участников взаимодействия и других факторов [Schwager et al., 2026]. Языковая модель, усредняющая значения по всему корпусу, лишена чувствительности к ситуативным нюансам, нивелируя контекстуальную вариативность во имя статистической устойчивости [Wu et al., 2025]. В-третьих, многие социальные феномены имеют выраженный невербальный и экстралингвистический компонент, который принципиально отсутствует в текстовых данных, что делает репрезентации модели неизбежно усечёнными и потенциально искажёнными [Mou et al., 2026; Taillandier et al., 2025]. Совокупность данных факторов приводит к тому, что операционализация социальных конструктов через языковые модели оказывается систематически смещённой и теоретически неадекватной, ставя под сомнение валидность выводов, полученных на её основе [Larooij, Törnberg, 2025; Choi, Young, Ferrara, 2026].

Когнитивное моделирование социальных процессов в идеале должно быть ориентировано на выявление каузальных структур: понимание условий возникновения социальных феноменов, механизмов влияния одних факторов на другие и возможности прогнозирования развития социальных ситуаций [Giabbanelli, 2026; Mou et al., 2026]. Большие языковые модели, однако, по своей природе являются корреляционными системами [Larooij, Törnberg, 2025]. Они улавливают статистические ассоциации между языковыми элементами, но не обладают каузальными репрезентациями [Choi, Young, Ferrara, 2026]. Языковая модель может с высокой точностью предсказывать, что за словом «гнев» с высокой вероятностью следует «агрессия» или «несправедливость», однако данная предсказательная способность не эквивалентна пониманию каузальной связи между восприятием несправедливости, возникновением аффективного состояния и агрессивным поведением [Wu et al., 2025].

В социальных науках проблема дискриминации корреляции и каузации является одной из наиболее фундаментальных [Larooij, Törnberg, 2025]. Применение языковых моделей усугубляет данную проблему, поскольку их высокая предсказательная сила создаёт ложное впечатление каузального понимания [Giabbanelli, 2026]. Исследователь, наблюдая успешное предсказание социальных исходов на основе языковых маркеров, может прийти к ошибочному выводу о каузальной роли данных маркеров [Choi, Young, Ferrara, 2026]. Модель может предсказывать события, не понимая их причин, используя коррелятивные связи, не имеющие объяснительной ценности [Larooij, Törnberg, 2025]. Более того, языковые модели склонны воспроизводить и усиливать статистические корреляции, присутствующие в обучающих данных, включая те, которые не имеют эмпирических каузальных оснований [Hu et al., 2025]. Это может вести к закреплению ложных стереотипов, необоснованных обобщений и ошибочных выводов об устройстве социальной реальности, что особенно критично в прикладных областях, где результаты моделирования используются для принятия решений [Wu et al., 2025; Mou et al., 2026].

Следующая группа рисков связана с ограничениями обобщаемости и внешней валидности результатов, получаемых с использованием больших языковых моделей [Larooij, Törnberg, 2025; Schwager et al., 2026]. Когнитивное моделирование социальных процессов традиционно нацелено на получение знаний, допускающих экстраполяцию за пределы конкретных выборок, контекстов и исторических периодов [Mou et al., 2026]. В случае языковых моделей проблема

обобщения приобретает особую остроту ввиду структуры обучающих корпусов [Choi, Young, Ferrara, 2026]. Современные модели обучаются на текстовых массивах, статистически доминируемых материалами из определённых сегментов глобального интернета и книжных коллекций, преимущественно представляющих западные, индустриализованные, образованные, богатые и демократические общества [Hu et al., 2025; Larooij, Törnberg, 2025]. Данный дисбаланс порождает систематическое смещение, известное в научной литературе как WEIRD-проблема (Western, Educated, Industrialized, Rich, Democratic) [Larooij, Törnberg, 2025]. Модели, обученные на таких корпусах, оказываются адаптированными к репрезентации социального опыта, норм и ценностей, характерных для доминирующих групп, и демонстрируют существенно более низкую эффективность при анализе текстов, отражающих иные культурные контексты, исторические периоды или социальные позиции [Choi, Young, Ferrara, 2026; Hu et al., 2025].

Результаты моделирования социальных процессов могут быть систематически смещены в сторону конкретных культурных и социальных парадигм, что ставит под сомнение их универсальность [Larooij, Törnberg, 2025]. Проблема усугубляется исторической динамикой социальных процессов: социальные нормы, ценности, институты и формы взаимодействия изменяются во времени, и модели, обученные на корпусах, отражающих определённый временной срез, могут оказаться неадекватными для понимания современных или будущих социальных конфигураций [Hu et al., 2025]. Кроме того, социальные процессы часто характеризуются нелинейностью и пороговыми эффектами, когда незначительные изменения начальных условий приводят к качественным трансформациям системы [Jiang et al., 2026]. Языковые модели, основанные на линейной статистике распределений, плохо приспособлены к улавливанию подобных нелинейностей, что снижает надёжность их прогнозов в отношении социальных процессов [Larooij, Törnberg, 2025; Mou et al., 2026].

Помимо культурно-исторических смещений, существуют жанровые, стилистические и ситуативные искажения [Schwager et al., 2026]. Корпуса включают гетерогенный материал, такие как новостные статьи, научные публикации, блоги, форумные обсуждения, художественную литературу — каждый из которых обладает специфической стилистикой, лексикой и устойчивыми оборотами [Hu et al., 2025]. Модель, не дифференцирующая жанровые контексты, усредняет эти характеристики, формируя гипотетическое «нейтральное» языковое пространство, не соответствующее ни одному реальному коммуникативному контексту [Wu et al., 2025]. Моделирование социальных процессов, таким образом, осуществляется в сконструированной среде «усреднённого дискурса», которая не имеет прямого эмпирического референта [Larooij, Törnberg, 2025]. Особую проблему представляет тот факт, что многие социальные процессы, такие как коллективное принятие решений, формирование общественного мнения, эмоциональное заражение в группах, разворачиваются преимущественно в устной или мультимодальной коммуникации, слабо представленной в текстовых корпусах [Mou et al., 2026; Taillandier et al., 2025]. Языковая модель моделирует не столько сами социальные процессы, сколько их письменные репрезентации, которые могут существенно расходиться с неформальными, имплицитными и ситуативными аспектами социальной реальности [Choi, Young, Ferrara, 2026].

Значимой проблемой является интерпретация результатов моделирования и формулировка теоретических утверждений на их основе [Larooij, Törnberg, 2025; Giabbanelli, 2026]. Когнитивное моделирование социальных процессов традиционно предполагает не только получение эмпирических данных, но и их концептуальное осмысление в рамках определённых

теоретических систем [Mou et al., 2026]. Языковые модели как инструменты лишены теоретического аппарата; они являются средствами обработки данных, а не теориями [Choi, Young, Ferrara, 2026]. Однако в исследовательской практике часто происходит смешение данных уровней: результаты, полученные с помощью языковой модели, интерпретируются как прямое подтверждение или опровержение теоретических положений [Larooij, Törnberg, 2025]. Подобная интерпретация методологически некорректна [Wu et al., 2025].

Языковые модели не имеют доступа к теоретическим конструктам; они функционируют на уровне поверхностных языковых паттернов, и любое «теоретическое» осмысление их выходных данных осуществляется исследователем, который привносит в интерпретацию собственные допущения и категориальные схемы [Larooij, Törnberg, 2025]. Поскольку модели непрозрачны, исследователь часто не способен дифференцировать в их выходных данных статистические артефакты и потенциально значимые результаты [Giabbanelli, 2026]. Возникает ситуация, когда модель начинает восприниматься как «автономный теоретик», генерирующий новые концептуальные схемы и объяснения [Choi, Young, Ferrara, 2026]. Данная эпистемическая позиция особенно опасна в случаях, когда модель используется для генерации гипотез, которые затем не проходят строгой эмпирической верификации, а принимаются на основании того, что были «предложены» сложной вычислительной системой [Larooij, Törnberg, 2025; Wu et al., 2025]. Подобная практика ведёт к девальвации теоретической работы и критического мышления, замещая их механистическими процедурами [Mou et al., 2026].

Риск депрофессионализации и эрозии методологического скептицизма представляет собой одно из наиболее широких социально-институциональных последствий внедрения языковых моделей в когнитивную науку [Larooij, Törnberg, 2025; Giabbanelli, 2026]. Когда доступный вычислительный инструмент становится рутинно применяемым, возникает тенденция к некритическому полаганию на него как на основной или исключительный источник данных, что сопровождается ослаблением способности к самостоятельной формулировке исследовательских проблем, сбору и анализу данных традиционными методами [Giabbanelli, 2026]. Депрофессионализация проявляется на уровне институциональных практик: если продуктивность научной работы начинает оцениваться по количеству и сложности применённых моделей, а не по глубине теоретического осмысления и методологической строгости, формируются неверные стимулы, поощряющие поверхностное использование языковых моделей без должного критического анализа [Larooij, Törnberg, 2025; Wu et al., 2025].

В этой ситуации утрачивается фундаментальный компонент научного метода — систематическое сомнение и проверка исходных предположений [Mou et al., 2026]. Языковые модели должны служить средством расширения познавательных возможностей, а не их замещением [Choi, Young, Ferrara, 2026]. Однако в реальной практике наблюдается смещение акцента с развития исследовательских компетенций на освоение технических интерфейсов, что создаёт поколение исследователей, технически компетентных, но испытывающих дефицит навыков самостоятельного теоретического мышления и критического обобщения [Larooij, Törnberg, 2025; Giabbanelli, 2026].

Воспроизводимость является фундаментальным принципом научного метода: результаты должны быть проверяемы другими исследователями при использовании тех же методов и данных [Giabbanelli, 2026; Larooij, Törnberg, 2025]. Однако в случае больших языковых моделей воспроизводимость становится серьёзной проблемой [Mou et al., 2026]. Модели постоянно обновляются, изменяются их версии, архитектурные параметры, способы предобработки и методы обучения [Hu et al., 2025]. Незначительные модификации в конфигурации могут

приводить к существенно различным выходным данным [Choi, Young, Ferrara, 2026]. Многие модели являются проприетарными, и их внутренняя организация не раскрывается разработчиками, что делает полное документирование исследовательского процесса невозможным [Larooij, Törnberg, 2025]. В этих условиях воспроизведение результатов становится крайне затруднительным или невозможным [Wu et al., 2025].

Отсутствие воспроизводимости ставит под сомнение объективность научного знания, особенно в области когнитивного моделирования социальных процессов, где выводы могут иметь практические импликации для управления, образования и политики [Larooij, Törnberg, 2025; Mou et al., 2026]. Это создаёт условия для манипуляций, когда результаты могут быть выборочно представлены в поддержку определённых решений без возможности независимой проверки [Giabbanelli, 2026].

Этический аспект использования больших языковых моделей в когнитивном моделировании социальных процессов заслуживает отдельного рассмотрения [Larooij, Törnberg, 2025; Wu et al., 2025]. Модели обучаются на данных, часто содержащих информацию о людях, не дававших информированного согласия на использование их текстов в исследовательских целях, что нарушает базовые этические принципы конфиденциальности и автономии личности [Hu et al., 2025]. Языковые модели могут применяться для анализа текстов с целью выявления социальных, демографических и психологических характеристик людей и групп, что создаёт серьёзные риски дискриминации и социального контроля [Choi, Young, Ferrara, 2026]. Подобные применения открывают возможность профилирования населения, прогнозирования поведения и принятия решений, затрагивающих судьбы людей [Larooij, Törnberg, 2025]. В условиях непрозрачности и слабой контролируемости моделей эти применения представляют угрозу автономии и достоинству человека [Wu et al., 2025]. Социальное моделирование оказывается не только теоретической или методологической, но и глубоко этической проблемой, требующей разработки адекватных регулятивных механизмов [Mou et al., 2026].

Из представленного анализа не следует вывод о неприемлемости использования больших языковых моделей в когнитивном моделировании социальных процессов [Larooij, Törnberg, 2025]. Напротив, данные модели обладают значительным эвристическим потенциалом: они позволяют обрабатывать массивы данных, недоступные для ручного анализа, выявлять скрытые паттерны и генерировать гипотезы для последующей эмпирической проверки [Jiang et al., 2026; Mou et al., 2026]. Однако реализация этого потенциала требует ответственного и критического подхода [Giabbanelli, 2026].

Необходимым условием является развитие культуры прозрачности и документирования [Larooij, Törnberg, 2025]. Любое исследование с использованием языковых моделей должно содержать исчерпывающее описание всех этапов работы: выбора модели и версии, параметров предобработки, способов промптинга, методов анализа выходных данных [Giabbanelli, 2026]. Без такой документации исследование становится непроверяемым и утрачивает научную ценность [Larooij, Törnberg, 2025]. Предпочтение следует отдавать открытым моделям с доступной архитектурой и опубликованными весами, а также использовать открытые репозитории для хранения кода, данных и протоколов [Hu et al., 2025; Mou et al., 2026].

Существенное значение имеет развитие методологического скептицизма и критической рефлексии [Larooij, Törnberg, 2025]. Выходные данные модели не следует рассматривать как прямое отражение социальной реальности; они должны интерпретироваться с учётом ограничений модели, культурной и жанровой специфичности, корреляционной природы,

неспособности к каузальному анализу, систематических смещений [Choi, Young, Ferrara, 2026; Wu et al., 2025]. Исследовательская практика должна включать триангуляцию результатов с использованием независимых методов: опросов, экспериментальных процедур, качественного анализа [Mou et al., 2026].

Необходим междисциплинарный диалог, объединяющий не только когнитивных учёных и компьютерных лингвистов, но и социологов, антропологов, философов, историков, специалистов в области этики [Larooij, Törnberg, 2025; Giabbanelli, 2026]. Социальные процессы являются сложными многомерными феноменами, требующими интеграции различных дисциплинарных перспектив, что позволяет преодолеть дисциплинарные предубеждения и избежать редукционизма [Mou et al., 2026].

Заключение

Большие языковые модели представляют собой двойственный феномен в контексте когнитивного моделирования социальных процессов [Larooij, Törnberg, 2025]. С одной стороны, они открывают беспрецедентные возможности для обработки данных, выявления скрытых закономерностей и генерации гипотез [Jiang et al., 2026; Mou et al., 2026]. С другой стороны, они несут существенные риски: онтологическую путаницу, методологические ограничения операционализации, каузальную слепоту, культурную и жанровую смещённость, проблемы воспроизводимости, депрофессионализацию и этическую уязвимость [Larooij, Törnberg, 2025; Choi, Young, Ferrara, 2026]. Данные риски не являются фатальными, но требуют систематического осмысления и институциональных мер по их минимизации [Giabbanelli, 2026].

Научное сообщество стоит перед необходимостью выработки общих стандартов, практик и норм, позволяющих использовать вычислительные возможности языковых моделей без ущерба для качества, глубины и объективности научного познания [Larooij, Törnberg, 2025; Wu et al., 2025]. Только сбалансированный подход, сочетающий техническую компетентность с методологической рефлексией и этической ответственностью, позволит реализовать эвристический потенциал языковых моделей, избегая их искажающего влияния на понимание социальной реальности [Mou et al., 2026; Giabbanelli, 2026]. В конечном счёте, именно ответственность исследователей, их методологическая грамотность и этичность будут определять, станут ли языковые модели подлинным инструментом познания или источником систематических заблуждений [Larooij, Törnberg, 2025].

Когнитивное моделирование социальных процессов представляет собой область, в которой исследовательские ставки особенно высоки, поскольку её выводы затрагивают фундаментальные аспекты человеческого существования, а любые искажения здесь могут иметь серьёзные практические последствия [Mou et al., 2026; Choi, Young, Ferrara, 2026]. Поэтому критическое и ответственное отношение к языковым моделям является не просто методологическим требованием, но этическим императивом современной науки [Larooij, Törnberg, 2025; Wu et al., 2025].

Библиография

1. Choi E. C., Young L. E., Ferrara E. Overstating attitudes, ignoring networks: LLM biases in simulating misinformation susceptibility // Proceedings of the International AAAI Conference on Web and Social Media. – 2026. – Т. 20. – №. 1. С. 520-547.

2. Giabbanelli P. J. A guide to large language models in modeling and simulation: From core techniques to critical challenges //arXiv preprint arXiv:2602.05883. – 2026.
3. Hu Z. et al. Population-aligned persona generation for llm-based social simulation //arXiv preprint arXiv:2509.10127. – 2025.
4. Hua Y. et al. Towards Proactive and Collaborative Industrial AI Agents within the Human-Agent Interaction (HAI): A Review and Perspective //Available at SSRN 6076349.
5. Jiang S. et al. Large language models for spreading dynamics in complex systems //Physics Reports. – 2026. – T. 1171. C. 1-50.
6. Jin K. et al. A scalable framework for building and validating knowledge-driven synthetic consumer agents with large language models //Journal of Industrial Information Integration. – 2026. – C. 101183.
7. Larooij M., Törnberg P. Validation is the central challenge for generative social simulation: a critical review of LLMs in agent-based modeling //Artificial Intelligence Review. – 2025. – T. 59. – №. 1. – C. 15.
8. Lin Y. Z. et al. Exploring silicon-based societies: An early study of the moltbook agent community //arXiv preprint arXiv:2602.02613. – 2026.
9. Mou X. et al. From individual to society: A survey on social simulation driven by large language model-based agents //ACM Computing Surveys. – 2026. – T. 58. – №. 11. – C. 1-41.
10. Orlando G. M. et al. Validating generative agent-Based modeling in social media simulations through the lens of the friendship paradox //Information Processing & Management. – 2026. – T. 63. – №. 4. – C. 104636.
11. Roy P. K. et al. (ed.). Social Network Analytics: Empowering Data Engineering with Deep Learning and Large Language Models. – CRC Press, 2026.
12. Schwager N. et al. Towards simulating social media users with LLMs: Evaluating the operational validity of conditioned comment prediction //The Proceedings for the 15th Workshop on Computational Approaches to Subjectivity, Sentiment Social Media Analysis (WASSA 2026). – 2026. – C. 208-221.
13. Taillandier P. et al. Integrating llm in agent-based social simulation: Opportunities and challenges //arXiv preprint arXiv:2507.19364. – 2025.
14. Wu Z. et al. Llm-based social simulations require a boundary //arXiv preprint arXiv:2506.19806. – 2025.
15. Zhang Y. et al. Sense, Think, Act, Reflect: Distilling Fast and Interpretable Decision Functions from LLM-Driven Crowds //ACM Transactions on Modeling and Computer Simulation. – 2026.
16. Zhang Y., Jiang J., Tang X. Toward Agent-Based Educational Science: Rethinking Educational Research in the Age of AI //Educational Psychology Review. – 2026. – T. 38. – №. 1. – C. 113.
17. Zhou P. et al. Large Language Models Meet Graph Neural Networks: A Comprehensive Survey //Available at SSRN 6225922.
18. Zhou Q., Sheu J. B. The use of Generative Artificial Intelligence (GenAI) in operations research: Review and future research agenda //Journal of the Operational Research Society. – 2026. – T. 77. – №. 7. – C. 1675-1695.

Risks of Applying Large Language Models in Cognitive Modeling of Social Processes

Oleg A. Smirnov

PhD in Physics and Mathematics, Associate Professor,
Kosygin Russian State University,
115035, 52/45, Sadovnicheskaya str., Moscow, Russian Federation;
e-mail: smirnovoleg1952@mail.ru

Anna S. Elagina

PhD in Economics, Associate Professor,
Department of Economic Disciplines,
Jewish University,
127273, 6, Otradnaya str., Moscow, Russian Federation;
e-mail: yelagina.anna@gmail.com

Abstract

The paper presents a systematic conceptual analysis of the risks associated with the application of large language models in the cognitive modeling of social processes. The rapid introduction of these computational systems into research practice opens up new heuristic possibilities for processing text arrays, identifying hidden statistical patterns, and generating hypotheses; however, it is accompanied by a complex of epistemological, methodological, institutional, and ethical problems requiring critical reflection. The focus is on the ontological status of model constructs and the related problem of reification, the methodological limitations of operationalizing social constructs through linguistic correlates, the correlational nature of models that excludes causal analysis, as well as systematic biases conditioned by the structure of training corpora and their cultural-historical specificity. Additionally, the problems of interpreting results and mixing levels of theoretical generalization, the institutional risks of deprofessionalization and erosion of methodological skepticism, the difficulties of reproducibility and documentation of research, as well as ethical aspects related to confidentiality, informed consent, and potential discrimination are considered. On the basis of the conducted analysis, strategies for mitigating the identified risks are formulated, including the development of a culture of transparency, interdisciplinary dialogue, triangulation of methods, the introduction of ethical codes, and the preservation of fundamental research competencies. It is concluded that language models can serve as an effective instrument of cognition only under the condition of a responsible and critical approach combining technical competence with methodological reflection and ethical reflexivity, as well as while preserving the distinction between computational representations and empirical social reality.

For citation

Smirnov O.A., Elagina A.S. (2026) Riski primeneniya bol'shikh yazykovykh modeley v kognitivnom modelirovanii sotsial'nykh protsessov [Risks of Applying Large Language Models in Cognitive Modeling of Social Processes]. *Psikhologiya. Istoriko-kriticheskie obzory i sovremennye issledovaniya* [Psychology. Historical-critical Reviews and Current Researches], 15 (7A), pp. 202-212. DOI: 10.34670/AR.2026.94.20.017

Keywords

Large language models, cognitive modeling, social processes, epistemological risks, operationalization of constructs, causal analysis, external validity, cultural bias, reproducibility, deprofessionalization, ethics of artificial intelligence, methodological skepticism, interdisciplinarity.

References

1. Choi E.C., Young L.E., Ferrara E. (2026) Overstating attitudes, ignoring networks: LLM biases in simulating misinformation susceptibility. *Proceedings of the International AAAI Conference on Web and Social Media*, 20(1), 520-547.
2. Giabbanelli P.J. (2026) A guide to large language models in modeling and simulation: From core techniques to critical challenges. *arXiv preprint arXiv:2602.05883*.
3. Hu Z. et al. (2025). Population-aligned persona generation for LLM-based social simulation. *arXiv preprint arXiv:2509.10127*.
4. Hua Y. et al. (n.d.). Towards Proactive and Collaborative Industrial AI Agents within the Human-Agent Interaction (HAI): A Review and Perspective. *Available at SSRN 6076349*.
5. Jiang S. et al. (2026) Large language models for spreading dynamics in complex systems. *Physics Reports*, 1171, 1-50.
6. Jin K. et al. (2026) A scalable framework for building and validating knowledge-driven synthetic consumer agents with large language models. *Journal of Industrial Information Integration*, 101183.

7. Larooij M., Törnberg P. (2025). Validation is the central challenge for generative social simulation: a critical review of LLMs in agent-based modeling. *Artificial Intelligence Review*, 59(1), 15.
8. Lin Y.Z. et al. (2026) Exploring silicon-based societies: An early study of the Moltbook agent community. *arXiv preprint arXiv:2602.02613*.
9. Mou X. et al. (2026) From individual to society: A survey on social simulation driven by large language model-based agents. *ACM Computing Surveys*, 58(11), 1-41.
10. Orlando G.M. et al. (2026) Validating generative agent-based modeling in social media simulations through the lens of the friendship paradox. *Information Processing & Management*, 63(4), 104636.
11. Roy P.K. et al. (eds.) (2026) *Social Network Analytics: Empowering Data Engineering with Deep Learning and Large Language Models*. CRC Press.
12. Schwager N. et al. (2026) Towards simulating social media users with LLMs: Evaluating the operational validity of conditioned comment prediction. *The Proceedings for the 15th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2026)*, 208-221.
13. Taillandier P. et al. (2025). Integrating LLM in agent-based social simulation: Opportunities and challenges. *arXiv preprint arXiv:2507.19364*.
14. Wu Z. et al. (2025). LLM-based social simulations require a boundary. *arXiv preprint arXiv:2506.19806*.
15. Zhang Y. et al. (2026) Sense, Think, Act, Reflect: Distilling Fast and Interpretable Decision Functions from LLM-Driven Crowds. *ACM Transactions on Modeling and Computer Simulation*.
16. Zhang Y., Jiang J., Tang X. (2026) Toward Agent-Based Educational Science: Rethinking Educational Research in the Age of AI. *Educational Psychology Review*, 38(1), 113.
17. Zhou P. et al. (n.d.). Large Language Models Meet Graph Neural Networks: A Comprehensive Survey. *Available at SSRN 6225922*.
18. Zhou Q., Sheu J.B. (2026) The use of Generative Artificial Intelligence (GenAI) in operations research: Review and future research agenda. *Journal of the Operational Research Society*, 77(7), 1675-1695.