

**Игнатъев Сергей Александрович**, руководитель Лаборатории информационных процессов в сложных социальных системах Института проблем передачи информации им. А. А. Харкевича РАН, e-mail: [ignatev-sa@iitp.ru](mailto:ignatev-sa@iitp.ru)

**Лагутин Юрий Викторович**, кандидат социологических наук, научный сотрудник Лаборатории обработки и передачи информации в когнитивных системах Института проблем передачи информации им. А. А. Харкевича РАН

**Резниченко Андрей Яковлевич**, научный сотрудник Лаборатории информационных процессов в сложных социальных системах Института проблем передачи информации им. А. А. Харкевича РАН

**Аронов Тайман Талгатович**, научный сотрудник Лаборатории обработки и передачи информации в когнитивных системах Института проблем передачи информации им. А. А. Харкевича РАН

## **ИНТЕЛЛЕКТУАЛЬНЫЕ МЕТОДЫ АНАЛИЗА ДАННЫХ В ЗАДАЧАХ МОДЕЛИРОВАНИЯ ПРОЦЕССОВ СЛОЖНЫХ СОЦИАЛЬНЫХ СИСТЕМ**

**Аннотация.** Статья посвящена анализу ограничений применения интеллектуальных методов анализа данных в моделировании социальных процессов и обоснованию перспективных методологических подходов их развития и использования в условиях высокой сложности и неоднородности социальных систем. В работе рассматриваются ключевые методологические и прикладные ограничения современных моделей, обобщаются практики стандартизации, валидации и публикации результатов агент-ориентированного моделирования, а также анализируются архитектуры интеллектуального анализа данных, наиболее адекватные для исследования сложных социальных систем и поддержки управленческих решений. Методологическую основу составляет аналитический обзор и концептуальное обобщение отечественного и зарубежного опыта в области интеллектуального анализа данных, агент-ориентированного моделирования, системной динамики, графовых и причинно-следственных моделей, с акцентом на

face- и empirical-валидацию, стандартизацию описаний АВМ и обеспечение приватности при работе с цифровыми следами.

Авторами показано, что ключевые ограничения связаны с неполнотой и неоднородностью цифровых следов, ограниченной переносимостью моделей, высокой чувствительностью результатов к агентным правилам и структуре графов, а также рисками нарушения приватности. Сделан вывод о целесообразности использования гибридных архитектур, объединяющих агент-ориентированное моделирование, системную динамику, графовые и причинно-следственные модели, и отмечены приоритетные для Российской Федерации направления, включая «искусственные общества», суперкомпьютерные АВМ и задачи эпидемического и демографического планирования.

### **Annotation**

The article is devoted to the analysis of the limitations of applying intelligent data analysis methods to the modeling of social processes and to substantiating promising methodological approaches for their development and use in conditions of high complexity and heterogeneity of social systems. The paper examines key methodological and applied limitations of contemporary models, summarizes practices of standardization, validation, and publication of agent-based modeling results, and analyzes data analytics architectures that are most suitable for studying complex social systems and supporting decision-making. The methodological framework is based on an analytical review and conceptual synthesis of domestic and international experience in intelligent data analysis, agent-based modeling, system dynamics, graph-based, and causal models, with particular attention to face and empirical validation, standardization of ABM descriptions, and privacy-preserving use of digital trace data.

It is shown that the main limitations are associated with the incompleteness and heterogeneity of digital traces, limited model transferability, high sensitivity of results to agent rules and interaction graph structures, as well as privacy risks. The study concludes that hybrid architectures combining agent-based modeling, system dynamics, graph-based, and causal components are the most effective, and identifies priority application areas for

the Russian Federation, including artificial societies, large-scale supercomputing ABMs, and epidemic and demographic planning tasks. **Ключевые слова:** интеллектуальный анализ данных; сложные социальные системы; агент-ориентированное моделирование; системная динамика; гибридные модели; сценарное моделирование.

**Keywords:** intelligent data analysis; complex social systems; agent-based modeling; system dynamics; hybrid models; scenario modeling.

Моделирование процессов в сложных социальных системах, охватывающее демографические изменения, эпидемическое распространение, динамику информационных потоков и поведенческие аспекты экономической активности, сталкивается с рядом устойчивых методологических затруднений. Ключевым из них является выраженная гетерогенность акторов: индивиды, домохозяйства, организации и цифровые пользователи существенно различаются по ресурсам, целевым функциям, набору ограничений и стратегиям принятия решений. Существенную роль играет и сетевой характер взаимодействий, поскольку социальные контакты, доверие, влияние и транзакционные связи формируют графовые структуры, конфигурация которых способна радикально менять наблюдаемую макродинамику, порождая каскадные эффекты и режимные переключения. Наконец, социальные системы характеризуются многоуровневыми обратными связями, при которых индивидуальные решения модифицируют параметры среды — например, цены, нормы поведения, субъективные и объективные риски, — а изменения среды, в свою очередь, трансформируют стимулы и допустимые множества решений на микроуровне.

В условиях такой сложности традиционные инструменты анализа, основанные на регрессионных зависимостях, агрегированных дифференциальных уравнениях и усредненных коэффициентах, нередко оказываются недостаточными. Проблема проявляется прежде всего тогда, когда требуется убедительно интерпретировать нелинейные эффекты управленческих воздействий [14], включая ограничительные меры, модерацию контента или экономические стимулы, а также воспроизвести микромеханизмы, определяющие поведение акторов в изменяющейся институциональной и информационной среде. Дополнительное усложнение связано с тем, что структурные сдвиги в сетевой топологии, например изменение структуры контактов или каналов коммуникации, способны приводить к качественно различным траекториям системы, которые агрегированные модели фиксируют слабо либо неверно интерпретируют. В результате в современной научной повестке усиливается

ориентация на гибридные модельные архитектуры, в рамках которых агент-ориентированное моделирование описывает микровзаимодействия и правила поведения, системная динамика формализует контуры обратных связей на агрегированном уровне, а интеллектуальный анализ данных обеспечивает идентификацию параметров, калибровку, реконструкцию сетевых структур и, в ряде случаев, причинно-следственную интерпретацию наблюдаемых эффектов [9; 10].

Эмпирическая практика показывает, что в России возрастает интерес к масштабным агент-ориентированным проектам, которые концептуализируются как «искусственные общества» и ориентируются на задачи государственного управления, включая сценарное прогнозирование и оценку комплексных воздействий на демографические и социально-экономические процессы. Данная линия исследований предполагает развитие крупномасштабных моделей, включающих не только индивидуальные агенты, но и домохозяйства, семейные структуры и механизмы, позволяющие задавать управленческие «проекты» как формализованные вмешательства [3]. Зарубежные исследования параллельно демонстрируют, что агент-ориентированное моделирование может выступать методологически оправданной «экспериментальной средой» в тех областях, где натурные эксперименты ограничены этическими, правовыми или организационными факторами. Показательным примером является моделирование кроссплатформенной диффузии информации и модерационных практик в многоплатформенной среде, где эмпирическая проверка вмешательств затруднена, тогда как симуляционные эксперименты позволяют систематически сравнивать альтернативные режимы регулирования и поведения пользователей [11].

Цель настоящей статьи заключается в систематизации интеллектуальных методов анализа данных, включая методы машинного и глубокого обучения, графовые модели, каузальный анализ и приватно-ориентированные квазиэкспериментальные процедуры, применительно к задачам моделирования процессов в сложных социальных системах. В рамках указанной цели рассматриваются архитектуры исследования, трактуемые как последовательность «данные — модель — управленческий эксперимент», а также демонстрируются типовые сценарии применения, релевантные как для российской практики, так и для зарубежных кейсов.

В качестве исследовательских задач рассматривается выделение основных классов данных, используемых при моделировании социальных процессов, и анализ типовых проблем качества, полноты и смещений, возникающих при работе с такими источниками. Далее обосновывается методическая схема гибридного моделирования

на основе сочетания агент-ориентированного подхода и системной динамики, дополненная «интеллектуальным контуром», обеспечивающим реконструкцию структуры взаимодействий и устойчивую параметрическую идентификацию. Отдельное внимание уделяется формированию набора практических процедур, применимых для калибровки, валидации и причинной оценки эффектов управленческих воздействий. Наконец, приводятся примеры применения предложенных подходов на российских примерах, связанных с эпидемическим и демографическим моделированием и анализом цифровых следов [6], а также на зарубежных примерах, включающих задачи мультиплатформенной диффузии, моделирование устойчивых агентных стратегий в сетях поставок и использование графовых моделей для описания предпочтений и распространения влияния.

Социальные системы описываются набором источников данных, существенно различающихся по уровню агрегирования, частоте наблюдений, структуре ошибок и доступности. Официальная статистика, включающая демографические показатели, занятость и данные по заболеваемости, как правило, обладает высокой степенью институционального доверия и методической прозрачностью, однако характеризуется низкой временной разрешающей способностью и значительным уровнем агрегирования, что осложняет восстановление микромеханизмов. Напротив, цифровые следы, формируемые в социальных сетях, поисковых запросах, данных о мобильности и транзакционных событиях, обеспечивают высокую частоту и детализацию наблюдений, но одновременно несут выраженные смещения выборки, зависящие от цифрового неравенства и различий в пользовательской активности, а также требуют строгого соблюдения требований приватности [6].

В российской исследовательской традиции подчеркивается, что цифровые следы становятся самостоятельным источником информации о качестве жизни и благополучии, однако их использование предполагает развитые процедуры верификации и корректировки репрезентативности [3]. Одновременно зарубежные работы показывают, как АВМ становится «безопасной песочницей» для экспериментов там, где реальные эксперименты этически и на уровне эксперимента ограничены: например, моделирование кроссплатформенной диффузии информации и модерации (Twitter+Reddit) с опорой на эмпирические датасеты политических событий в США [11].

Вместе с тем следует обратить внимание и на существенный класс источников, который составляют платформенные данные, включающие лог-события, графы подписок и дружбы, а также данные о модерационных действиях; такие данные особенно важны для анализа сетевой динамики, но часто ограничены с точки зрения

доступа из-за политик API и регуляторных требований, что дополнительно повышает роль моделирования как замещающего инструмента для экспериментального анализа.

Процедуры подготовки данных в исследуемом классе задач включают приведение временных шкал к единому формату, устранение дубликатов, восстановление пропусков, согласование географических и административных привязок, а также переход от «сырых» наблюдений к представлениям, совместимым с модельными объектами, то есть агентами и узлами графа. Для агент-ориентированного моделирования принципиально важным этапом выступает построение искусственной популяции, обеспечивающей воспроизведение ключевых распределений и структур домохозяйств, занятости, возрастной композиции и пространственного размещения. Российские исследования, посвященные ограничениям построения синтетических популяций [5], подчеркивают, что ошибки на данном этапе способны систематически исказить сценарные результаты, особенно в задачах общественного здоровья [4], где структура контактов и домохозяйств непосредственно определяет динамику распространения.

Методологическая основа исследования строится на гибридной архитектуре, в которой агент-ориентированный слой задает микродинамику через состояния агентов и правила их переходов, включая индивидуальные решения, адаптацию к информации и взаимодействия через сеть контактов либо пространственную близость. В этой постановке агенты могут интерпретироваться как индивиды, домохозяйства, организации или пользователи цифровых платформ, а сеть взаимодействий — как контактный, коммуникационный либо транзакционный граф. Слой системной динамики используется для формализации агрегированных переменных и макроуровневых контуров обратной связи, таких как доступность ресурсов здравоохранения, уровень доверия, интенсивность коммуникации, параметры медиаполя или обобщенные индикаторы занятости, которые ограничивают или стимулируют поведение агентов и тем самым связывают микро- и макроуровни описания.

Интеллектуальный слой рассматривается как функциональный контур, обеспечивающий извлечение структуры и параметров модели из данных. В рамках этого контура решаются задачи реконструкции графа взаимодействий на основе наблюдений контактов, совместной активности и ко-локаций, а также построения суррогатных моделей, которые аппроксимируют связь между параметрами симуляции и выходными метриками, тем самым снижая вычислительные затраты калибровки. Дополнительно используются графовые нейросетевые и спатико-

темпоральные архитектуры для прогнозирования и выявления скрытых зависимостей в данных, что позволяет уточнять модельные гипотезы и повышать точность идентификации.

Наконец, в состав интеллектуального слоя включается причинно-следственный модуль, обеспечивающий оценку эффектов управленческих воздействий и позволяющий работать с распределенными источниками данных при соблюдении требований приватности, когда вместо передачи исходных наборов используются промежуточные представления и согласованные протоколы вычислений. Взаимосвязь микро- и макроуровней моделирования, а также роль интеллектуального слоя показаны на рис. 1.



**Рис. 1.** Гибридная модельная архитектура: агент-ориентированный слой, системная динамика и интеллектуальный контур (разработано авторами).

Калибровка агент-ориентированных моделей трактуется как процедура подбора параметров правил поведения агентов, характеристик сетевой структуры и параметров внешней среды таким образом, чтобы модель воспроизводила набор эмпирически наблюдаемых метрик. В качестве таких метрик могут выступать временные траектории заболеваемости, распределения активности, статистики каскадного распространения информации или структурные показатели графа. Для повышения устойчивости калибровки целесообразно использовать многоцелевые критерии согласования, которые учитывают одновременно динамические, распределительные и сетевые свойства системы, а также применять методы ускорения поиска параметров на основе байесовской оптимизации и суррогатного моделирования. Существенным компонентом является анализ чувствительности, позволяющий выделить параметры, которые оказывают непропорционально сильное влияние на результирующие сценарии, и тем самым уточнить приоритеты сбора данных и последующей валидации.

Валидация в рассматриваемом подходе понимается как многоуровневая процедура проверки соответствия модели эмпирической реальности и устойчивости выводов. Экспертная проверка правдоподобия механизмов обеспечивает контроль

содержательной корректности предпосылок, тогда как эмпирическая валидация предполагает сопоставление результатов модели с независимыми источниками данных или периодами наблюдений, не использованными в калибровке. Дополнительный уровень составляет контрфактуальная проверка устойчивости, при которой тестируется, сохраняются ли выводы при альтернативных реализациях графа, иных процедурах предобработки данных или вариативности правил агентов в пределах допустимых допущений. Зарубежные исследования, посвященные кроссплатформенной диффузии информации, демонстрируют продуктивность комбинированной валидации, в которой используются как эмпирические датасеты событий, так и результаты опросов и предшествующих исследований, позволяющие уточнять поведенческие параметры и проверять реалистичность сценарных эффектов.

Во многих прикладных постановках ключевым является не только прогнозирование, но и получение обоснованного ответа на контрфактуальный вопрос о последствиях возможного вмешательства. Агент-ориентированное моделирование предоставляет естественный механизм построения контрфактуальных сценариев за счет изменения правил поведения, режимов ограничений, алгоритмов модерации или параметров стимулов, что позволяет воспроизводить альтернативные траектории системы. Вместе с тем для снижения риска «подгонки» и повышения интерпретируемости результатов необходимо связывать симуляционные выводы с наблюдаемыми эффектами, что обосновывает включение каузального слоя. В рамках такого слоя применяются процедуры оценки эффектов воздействия на основе квазиэкспериментальных дизайнов и методов балансировки, включая взвешивание и оценку склонности к «лечению», а также приватно-ориентированные протоколы причинного вывода для разрозненных источников, где обмен строится на агрегированных признаковых представлениях и согласованных вычислительных процедурах без передачи исходных персональных данных.

Таким образом, специфика сложных социальных систем — выраженная неоднородность акторов, сетевой характер взаимодействий и наличие многоуровневых обратных связей — предопределяет ограниченную применимость сугубо агрегированных и линейных описаний при решении задач сценарного анализа и оценки управленческих воздействий. Предложенная методическая рамка, основанная на гибридной архитектуре ABM+SD и дополненная интеллектуальным контуром анализа данных, обеспечивает более строгую связку между эмпирическими наблюдениями и модельными механизмами, повышая воспроизводимость калибровки и качество валидации.

Кроме того, как показал анализ литературы, интеграция реконструкции графов, суррогатного моделирования, графовых и спATIO-темпоральных нейросетевых подходов, а также каузального слоя позволяет не только улучшать прогнозную состоятельность моделей, но и формировать интерпретируемые оценки эффектов вмешательств в условиях неполноты данных и ограничений доступа к чувствительной информации. В практическом плане описанный подход задает универсальную схему «данные — модель — управленческий эксперимент», применимую как к российским задачам эпидемического и демографического моделирования и анализу цифровых следов, так и к зарубежным кейсам мультиплатформенной диффузии информации и сетевого анализа [7; 8; 12], при условии строгого контроля качества данных, явной фиксации допущений и многоуровневой проверки устойчивости получаемых выводов [13].

Таким образом, зарубежные примеры также подтверждают, что агент-ориентированное моделирование, усиленное интеллектуальным анализом данных, является эффективным инструментом для исследования социально-технических систем, в частности в задачах кроссплатформенной диффузии информации и оценки последствий алгоритмической модерации. В условиях, когда проведение натуральных экспериментов ограничено этическими и институциональными факторами, симуляционные эксперименты выступают обоснованной альтернативой, позволяющей анализировать контрфактуальные сценарии и сравнивать эффекты различных управленческих стратегий [11]. Существенную роль при этом играет интеграция причинно-следственного анализа, которая повышает доверие к интерпретации результатов моделирования и снижает риск произвольной подгонки параметров.

## Литература

1. Макаров В.Л., Бахтизин А.Р., Россошанская Е.А., Дорошенко Т.А., Самсонова Н.А. Проблемы стандартизации описания агент-ориентированных моделей и возможные пути их решения // Вестник Российской академии наук. 2023. Т. 93, № 4. С. 362–372. DOI: 10.31857/S0869587323040059, EDN: SCIOKT.
2. Макаров В.Л., Бахтизин А.Р., Сушко Е.Д., Сушко Г.Б. Разработка агент-ориентированной демографической модели России и ее суперкомпьютерная реализация // Вычислительные методы и программирование. Т. 19 (2018): Вып. 4. С. 368–378. DOI: doi 10.26089/NumMet.v19r433.
3. Макаров В.Л., Бахтизин А.Р., Сушко Е.Д., Сушко Г.Б. Создание суперкомпьютерной имитации общества с активными агентами разных типов и

- ее апробация // Вестник Российской академии наук. 2022. Т. 92, № 5. С. 458–466. DOI: 10.31857/S0869587322050115.
4. Маслова И.И., Манолов А.И., Глущенко О.Е., Козлов И.Е., Цуркис В.И., Попов Н.С., Самойлов А.Е., Лукашев А.Н., Ильина Е.Н. Ограничения в создании искусственных популяций в агентном моделировании эпидемий: систематический обзор. Журнал микробиологии, эпидемиологии и иммунобиологии. 2024;101(4):530–545. DOI: <https://doi.org/10.36233/0372-9311-467>. EDN: <https://www.elibrary.ru/goxzt>
  5. Пономарев Р.Л., Судаков В.А., Сивакова Т.В., Энтентеев А.Р., Ескин В.И. Разработка модели распространения COVID-19 в городских агломерациях // Препринт ИПМ им. М.В. Келдыша РАН № 74, Москва, 2021 г. URL: [https://keldysh.ru/papers/2021/prep2021\\_74.pdf](https://keldysh.ru/papers/2021/prep2021_74.pdf) (дата обращения: 12.11.2025).
  6. Щекотин Е.В. Цифровые следы как новый источник данных о качестве жизни и благополучии: обзор современных тенденций. 2021.
  7. Bachiller P., Rodriguez-Criado D., Jorvekar R. R., Bustos P., Faria D. R., Manso L. J. A graph neural network to model disruption in human-aware robot navigation // Multimedia Tools and Applications. 2022. <https://doi.org/10.1007/s11042-021-11113-6>.
  8. Kawamata Y., Motai R., Okada Y., Imakura A., Sakurai T. Collaborative causal inference on distributed data. Expert Systems with Applications. 2024. DOI: 10.1016/j.eswa.2023.123024.
  9. Larrea-Gallegos G., et al. A computational framework for modeling socio-technical agents in the life-cycle sustainability assessment of supply networks. (ScienceDirect). 2024.
  10. McGarraghy S., et al. Conceptual System Dynamics and Agent-Based Modelling... Case Studies. Agriculture (MDPI). 2022.
  11. Murdock, I., Carley, K.M. & Yağan, O. An agent-based model of cross-platform information diffusion and moderation. Soc. Netw. Anal. Min. 14, 145 (2024). DOI: <https://doi.org/10.1007/s13278-024-01305-x>.
  12. Vélez Bedoya J.I. Intelligent Agents and Causal Inference: Enhancing Decision-Making through Causal Reasoning. Applied Sciences (MDPI). 2024;14(9):3818.
  13. Yu Y., Qian W., Zhang L., Gao R. A Graph-Neural-Network-Based Social Network Recommendation Algorithm Using High-Order Neighbor Information. Sensors. 2022;22(19):7122. DOI: 10.3390/s22197122.
  14. Zeghina A.O., et al. Deep learning on spatiotemporal graphs: A systematic review, methodological landscape, and research opportunities. Neurocomputing. 2024.

## Literature

1. Makarov V. L., Bakhtizin A. R., Rossoshanskaya E. A., Doroshenko T. A., Samsonova N. A. Problems of standardization of descriptions of agent-based models and possible ways to address them // *Herald of the Russian Academy of Sciences*. 2023. Vol. 93, No. 4. P. 362–372. DOI: 10.31857/S0869587323040059. EDN: SCIOKT.
2. Makarov V. L., Bakhtizin A. R., Sushko E. D., Sushko G. B. Development of an agent-based demographic model of Russia and its supercomputer implementation // *Computational Methods and Programming*. Vol. 19 (2018), Iss. 4. P. 368–378. DOI: 10.26089/NumMet.v19r433.
3. Makarov V. L., Bakhtizin A. R., Sushko E. D., Sushko G. B. Creation of a supercomputer simulation of a society with active agents of different types and its validation // *Herald of the Russian Academy of Sciences*. 2022. Vol. 92, No. 5. P. 458–466. DOI: 10.31857/S0869587322050115.
4. Maslova I. I., Manolov A. I., Glushchenko O. E., Kozlov I. E., Tsurkis V. I., Popov N. S., Samoilov A. E., Lukashev A. N., Ilyina E. N. Limitations in creating artificial populations in agent-based epidemic modeling: a systematic review // *Journal of Microbiology, Epidemiology and Immunobiology*. 2024;101(4):530–545. DOI: <https://doi.org/10.36233/0372-9311-467>. EDN: <https://www.elibrary.ru/goxzt>.
5. Ponomarev R. L., Sudakov V. A., Sivakova T. V., Ententeev A. R., Eskin V. I. Development of a COVID-19 spread model in urban agglomerations // *Keldysh Institute of Applied Mathematics RAS Preprint* No. 74. Moscow, 2021. URL: [https://keldysh.ru/papers/2021/prep2021\\_74.pdf](https://keldysh.ru/papers/2021/prep2021_74.pdf) (accessed: 12.11.2025).
6. Shchekotin E. V. Digital traces as a new data source on quality of life and well-being: a review of current trends. 2021.
7. Bachiller P., Rodriguez-Criado D., Jorvekar R. R., Bustos P., Faria D. R., Manso L. J. A graph neural network to model disruption in human-aware robot navigation // *Multimedia Tools and Applications*. 2022. <https://doi.org/10.1007/s11042-021-11113-6>.

8. Kawamata Y., Motai R., Okada Y., Imakura A., Sakurai T. Collaborative causal inference on distributed data // *Expert Systems with Applications*. 2024. DOI: 10.1016/j.eswa.2023.123024.
9. Larrea-Gallegos G., et al. A computational framework for modeling socio-technical agents in the life-cycle sustainability assessment of supply networks. *ScienceDirect*. 2024.
10. McGarraghy S., et al. Conceptual system dynamics and agent-based modelling: case studies // *Agriculture (MDPI)*. 2022.
11. Murdock I., Carley K. M., Yağan O. An agent-based model of cross-platform information diffusion and moderation // *Social Network Analysis and Mining*. 2024. Vol. 14, Art. 145. DOI: <https://doi.org/10.1007/s13278-024-01305-x>.
12. Vélez Bedoya J. I. Intelligent agents and causal inference: enhancing decision-making through causal reasoning // *Applied Sciences (MDPI)*. 2024. Vol. 14, No. 9. Art. 3818.
13. Yu Y., Qian W., Zhang L., Gao R. A graph-neural-network-based social network recommendation algorithm using high-order neighbor information // *Sensors*. 2022. Vol. 22, No. 19. Art. 7122. DOI: 10.3390/s22197122.
14. Zeghina A. O., et al. Deep learning on spatiotemporal graphs: a systematic review, methodological landscape, and research opportunities // *Neurocomputing*. 2024.