



Искусственные общества. 2013-2021

ISSN 2077-5180

URL - <http://artsoc.jes.su>

Все права защищены

Выпуск 1 Том 16. 2021

Искусственные интеллектуальные системы и решение социальных задач: проблема доверия

Гуров Олег Николаевич

*Институт философии РАН, ИОМ РАНХИГС
Российская Федерация, Москва*

Аннотация

Мы ожидаем прорыв в совершенствовании и последующем широком применении моделей искусственного общества, на которых станет возможным тестировать управленческие и экономические решения. Такие технологии уже применяются в качестве инструмента апробации управленческих решений, и очевидно, что их использование для решения социальных задач требует всестороннего и трезвого внимания, поскольку затрагивает интересы общества и отдельного человека. Общим знаменателем всех аспектов этой деятельности является феномен доверия. Для решения сложных социальных задач с помощью искусственных интеллектуальных систем необходимо обеспечить доверие между всеми заинтересованными участниками. Прозрачность работы, высокие стандарты и доверие позволят достичь эффективный результат, значение которого невозможно переоценить, поскольку цена вопроса - успешное формирование социогуманитарной цивилизации.

Ключевые слова: искусственный интеллект, искусственная интеллектуальная система, этика, цифровая этика, цифровой двойник, компьютерная модель, доверие, социальные задачи

Дата публикации: 22.03.2021

Ссылка для цитирования:

Гуров О. Н. Искусственные интеллектуальные системы и решение социальных задач: проблема доверия // Искусственные общества. – 2021. – Т. 16. – Выпуск 1 [Электронный ресурс]. URL: <https://artsoc.jes.su/s207751800014095-3-1/> (дата обращения: 24.03.2021). DOI: 10.18254/S207751800014095-3

как мы впоследствии назовем 2020-е годы, ответил, что это десятилетие пройдет под знаком искусственного интеллекта. Технологии ИИ расширят возможности человека и вместе с этим станут применяться во всех сферах общественной жизни на качественно новом уровне [6]. Представляется, что сфера управления как экономическими, так и политическими процессами, не станет исключением, и это касается не только развития функционала цифрового государства и «Умного города», но и самой логики государственного управления. В частности, можно ожидать прорыва в совершенствовании и последующем широком применении моделей искусственного общества, на которых станет возможным тестировать управленческие и экономические решения. В нашей стране такой цифровой двойник общества, или иначе, демографическая компьютерная агент-ориентированная модель, разрабатывается в Центральном экономико-математическом институте (ЦЭМИ) РАН [8].

2 Хотя достижение предсказательной точности социальных явлений в такой же мере, как в физических науках, пока маловероятно, в первую очередь, вследствие сложности социальной системы и отсутствия достаточного количества эмпирических количественных данных, определяющих индивидуальное и групповое поведение, в последние годы идеи приближения общества к законам статистической физики стали реализовываться в виде конкретных исследовательских проектов [3].

3 Чтобы проиллюстрировать социальные задачи, которые можно решать с помощью таких технологий, приведем актуальный пример, связанный с текущими реалиями нашей жизни – пандемией коронавируса. При борьбе с COVID-19 в разных странах принимаются собственные политические решения, и проводится своя социальная политика. Установление того, какие меры оказываются более эффективными, на данный момент представляет объект дискуссии. При этом цена вопроса высока – на кону жизнь миллионов людей. Недавно немецкие ученые презентовали отчет о разработке агент-ориентированной модели для анализа эффективности различных стратегий по борьбе с коронавирусом «Пандемия в искусственном обществе». Значение именно такой симуляционной модели обусловлено тем, что ситуация очень непростая: определенные политические решения, связанные с ограничениями, могут нанести вред социально-экономической ситуации при обеспечении минимума заражений. С другой стороны, отсутствие ограничений подвергает население высокому риску заражения, и необходимо найти баланс между требованиями социально-экономических категорий и необходимостью наименее болезненного сценария формирования коллективного иммунитета. Ситуация осложняется тем, что данные и характеристики, связанные с вирусом и вызываемым им заболеванием, различаются для разных регионов и государств. Поэтому такое моделирование, позволяющее преобразовывать многоуровневую общественную систему в искусственное общество, способно предложить ответы на насущные вопросы и оценить эффективность различных мероприятий по борьбе с пандемией [18].

4 Однако, существуют и отрицательные примеры использования технологий ИИ, бросающие тень на перспективы развития цифрового двойника общества. В частности, после скандала вокруг манипуляций политическими процессами, осуществленных Cambridge Analytica, в обществе распространяются опасения, что модели искусственного общества станут применяться для таких же целей в еще больших масштабах и с лучшей эффективностью. Кроме того, существуют опасения и другого рода: рабочая группа бостонского Центра интеллекта и культуры создала симуляционную модель, чтобы протестировать эффективность различных стратегий, направленных на борьбу с международной торговлей детьми. При этом руководитель группы У. Уайлдман отмечает, что международная мафия в свою очередь в состоянии нанять собственных специалистов, чтобы создать такую же имитационную модель для своих целей, и таким образом перевести борьбу в область технологической гонки [17].

5 Итак, сама технология не может быть панацеей и ее возможности налагают огромную ответственность за использование и результаты. Изучая примеры, как технологии ИИ применяются в качестве инструментов апробации управленческих решений, мы видим, что их использование для решения социальных задач требует всестороннего и взвешенного внимания, поскольку затрагивает интересы общества и отдельного человека. Актуальность вопроса лишь возрастает, поскольку люди сегодня уже живут в виртуальном обществе все более и более полноценной жизнью. Поэтому технологии, применяемые для создания и управления цифровыми двойниками по сути каждого гражданина, будут вызывать в обществе отдельное беспокойство по мере их распространения и масштабирования. При этом мы можем наблюдать, как ИИ с каждым годом все больше и больше применяется в повседневной жизни. По этому направлению, наряду с формированием новых возможностей для благополучия человека и поиска решений в области насущных социальных проблем, возникают и серьезные опасения, связанные с тем, как гарантировать принятие решений, которые люди будут воспринимать как справедливые и соответствующие человеческим ценностям. Так как инструментарий ИИ использует огромные объемы данных, также необходимо обеспечить понимание, как данные обрабатываются ИИ и как ими оперируют те, кто сопровождает эту деятельность. Важно осознать и то, что доверие к технологии должно быть дополнено доверием к тем, кто ее производит. Однако такое доверие может быть достигнуто лишь при условии, что компании обеспечивают прозрачность в отношении своих политик использования данных и процессов при создании новых продуктов. Поэтому необходимо гарантировать, чтобы тот субъект, чьи данные используются, имел полную информацию и осознавал, как его данные обрабатываются, в каком месте хранятся, и каким образом и для каких целей используются [11].

6 Преодоление проблем «черного ящика» и возможностей принятия дискриминационных решений, определение зон ответственности за ошибки ИИ – отдельные необходимые условия для обеспечения доверия к ИИ, без которого не удастся в максимальной степени использовать его положительные возможности. Согласно исследованию IBM Institute for Business Value, 82% всех предприятий рассматривают возможность внедрения ИИ или уже его внедряют в целях увеличения дохода, повышения качества обслуживания клиентов и сокращения издержек. Однако в то же время 60% этих компаний опасаются проблем, связанных с ответственностью за применение этих технологий [15]. Нам представляется, что общим знаменателем всего этого дискурса является феномен доверия. Неслучайно, что ключевые организации, задействованные в развитии и регулировании этих технологий, за последние годы сформулировали руководящие принципы, которыми следует руководствоваться при создании, внедрении и использовании технологий ИИ. Во всех этих документах сделан акцент на гуманитарной составляющей в качестве главной цели новых технологий. Отметим некоторые из наиболее важных документов в этой области:

- 7 • «Принципы доверия и прозрачности» IBM. В соответствии с этим документом целью ИИ является не замена или замещение человеческого интеллекта, а предоставление дополнительных возможностей человеку. При этом доверие представляет собой ключевую категорию, через призму которого оцениваются все принципы, в том или ином виде затрагивающие проблемы данных;
- «Принципы Google в отношении ИИ». Развитие технологий ИИ должен быть направлено на то, чтобы гарантировать конфиденциальность, и такие технологии обязаны соответствовать принципам социальной ответственности, в число которых входят безопасность и подотчетность человеку, а также не противоречить положениям, определяющим социальную справедливость;
 - «Асиломарские принципы ИИ». В 2017 году на конференции Beneficial AI около 4000 ведущих исследователей, экспертов и заинтересованных участников, включая И.

Маска и С. Хокинга, разработали и приняли принципы в отношении технологий ИИ. Особое внимание в документе придается тому, что развитие технологии должно осуществляться в соответствии с этическими и гуманитарными ценностями;

- «Принципы этичного ИИ» ВЭФ: среди главных пяти принципов выделяются справедливость и право каждого человека использовать ИИ для своего благополучия;
- «Общие принципы Института инженеров по электротехнике и радиоэлектронике»: принципы, направляющие развитие ИИ в сторону достижения человеческого благополучия, ответственности, ценности и этики [11].

8 Возвращаясь непосредственно к теме искусственных обществ, дополним представленный список проблем еще одной – вопросом точности. В 1931 году А. Коржибски заявил, что «Карта - это не территория», подразумевая, что объект не соответствует символу [14]. Это большая проблема для любой виртуальной модели. Насколько точно она способна симулировать реальную систему? Любое моделирование предполагает абстракцию и упрощение, и вместе с этим и потерю точности. Тем более, если речь идет о системе с нелинейным поведением, которое приводит к появлению новых свойств. Например, моделируя город с населением в несколько миллионов человек, становится сложным предсказать и обработать неожиданные результаты. Даже если модель постоянно обновляется в режиме реального времени и использует максимальные вычислительные возможности, новейшие технологии машинного обучения и ИИ, она никогда не может быть идеальной [12]. И с учетом такой данности решение лежит не в технической сфере, а снова в гуманитарной, в плоскости доверия.

9 Это было очевидно с самого начала развития концепции и технологий ИИ. Поэтому здесь мы только отметим, что об искусственном обществе впервые заговорили в 1950-е, сразу после того, как были созданы первые ЭВМ и впервые появились техническая возможность для нового способа познания, при котором ученые смогли не только изучать существующие явления, но и конструировать собственный мир [5]. Практически сразу же эти новые возможности вызвали дискуссию по поводу рисков и угроз, которые возникают перед человечеством в свете перспектив развития таких технологий.

10 В первую очередь «реакция» последовала от массовой культуры, которая является убедительным индикатором того, какие проблемы актуальны для общества. Более, художественные произведения часто точно прогнозируют появление и распространение таких явлений, и нередко эти предсказания удивительным образом сбываются в полной мере. То, что несколько десятилетий назад казалось невообразимой и сложной для восприятия фантазией, становится впоследствии важной частью общественных процессов, требующих не просто интеллектуального осмысления, но и всестороннего практического регулирования. Концепция моделирования жизни как один из способов исследования и познания мира для последующего преобразования окружающей действительности является убедительным примером такой рефлексии. Согласно определению, представленному в экономическом словаре, изданном около 20 лет назад, под термином моделирование понимается «...воспроизведение экономических объектов и процессов в ограниченных, малых, экспериментальных формах, в искусственно созданных условиях (натурное моделирование)... Моделирование служит предпосылкой и средством анализа экономики и протекающих в ней явлений и обоснования принимаемых решений, прогнозирования, планирования, управления экономическими процессами и объектами» [7]. За прошедшие два десятилетия мы стали гораздо смелее определять потенциал моделирования в первую очередь в отношении масштаба – как мы писали выше, сегодня благодаря вычислительным возможностям и большим данным моделирование может быть развернуто в масштабах экономической системы целой страны.

11 Массовая культура же осознала значение и перспективы развития моделей искусственного общества гораздо раньше. Эта тема стала основой сюжета ряда значимых научно-фантастических произведений более полувека назад. В частности, в 1960-е С. Лем в цикле «Кибериада» описал математическое моделирование и бесконечные возможности «оцифровки» для создания идеального общества, а в одном из рассказов «Из воспоминаний Ийона Тихого» показана бесконечная череда реальностей, в каждой из которых моделируется последующая. В романе Д.Ф. Галуэе «Симулакрон-3» (1964 г.) показан виртуальный город, созданный для маркетинговых целей. Эта симуляция настолько совершенная, что ее обитатели имеют полноценное сознание, и при этом не осознают, что являются кодами в программе. Первая экранизация романа состоялась спустя десятилетие после выхода книги: в 1973 году немецкий режиссер Р.В. Фасбиндер снял фильм «Мир на проводе», в котором цифровой двойник общества был создан и использовался для прогнозирования экономического развития. Режиссер показывает полноценных личностей, живущих в симуляции, и вносит в нее человеческие ценности – любовь, теплоту, человечность и доверие. Отметим, что это первый фильм, в котором демонстрируется киберпространство (сам термин введет У. Гибсон лишь спустя десятилетие). В 1999 году выходит новая экранизация романа - «Тринадцатый этаж» режиссера Й. Руснака, в котором акцент делается на этической составляющей – определении того, что допустимо и недопустимо в отношении искусственного мира и его обитателей. Кроме этого, сюжет фильма строится вокруг рефлексии главного героя по поводу смысла его жизни в симуляции и его взаимоотношений с создателями из реального мира. Таким образом, массовая культура в течение многих десятилетий поднимала целый спектр вопросов, связанных с развитием моделей искусственного общества - таких как ответственность в отношении новых технологий и коммерциализации в ущерб общественному благу, размывание границы между реальным и виртуальным миром, экзистенциальные проблемы, связанные с определением подлинности существования и поиска смысла жизни.

12 Однако вернемся к практической сфере. В философском смысле проводимые исследования и разработки в области создания цифровых обществ предполагают достижение результата, который обеспечит максимальное соответствие реального и виртуального. Конечной целью создания и применения таких технологий является принятие верных управленческих решений для создания наилучших условий человеческой деятельности и достижения максимально комфортного и благополучного индивидуального и общественного развития. При этом, поскольку мы говорим о создании моделей для принятия решений в общественной сфере, формирование виртуального мира, максимально идентичного физическому миру, является не столько технической, сколько социальной проблемой. Поэтому важным становится не только создать, но и обеспечить функционирование верно сконструированной и корректно работающей системы, которая оперирует данными и в которой автономные агенты формируют виртуальное общество. Для этого необходимо учитывать не только очевидные требования безопасности и конфиденциальности, но также правовые аспекты и этические нормы, в комплексе формирующие сложный комплекс условий.

13 Цифровой двойник представляет собой набор данных, среди которых большая доля относится к личным данным. Поскольку целью является создание цифрового двойника, максимально соответствующего оригиналу из физического мира, значение конфиденциальных персональных данных сложно переоценить. В то же время такие ресурсы, как данные гаджетов и поведенческие данные из Интернета, требуют очень осторожного обращения, поскольку представляют собой чувствительную сферу, любой просчет в использовании которых способен подорвать доверие к технологии и дискредитировать работу по созданию общественного блага [2].

14 В этой области можно выделить следующие проблемы. В первую очередь, это уже упомянутые вопросы безопасности и конфиденциальности – защита от утечек и доступность в

той мере, которая действительно необходима. Этот блок, многократно описанный в последние годы, становится более сложным, если рассматривать его через призму обеспечения целостности цифрового двойника. Иными словами, представляется ли формирующий цифровой двойника объем данных достаточным для решения той или иной прогностической задачи, с одной стороны. И не собираются ли данные избыточно, без необходимости, с другой стороны. В данном случае задача состоит в том, чтобы обеспечить безопасный сбор и использование необходимых и точных данных [16].

15 К этой проблеме примыкает снова философский вопрос, каким образом и в какой степени человек, являющийся «оригиналом» цифрового двойника, способен в целом осознать сам факт существования собственного виртуального двойника и принять цели создания и использования цифровых двойников общества. Данные, используемые для создания цифрового двойника человека, генерируются различными источниками, принадлежащими государству, технологическим компаниям, банкам, и даже другими цифровыми двойниками. Факт того, что цифровой двойник представляет собой коллективный труд различных субъектов, может стать экзистенциальной проблемой для человека, приверженного концепции автономии личности, и способен привести к отчуждению и неуверенности в реальности существования [9].

16 Здесь мы снова вплотную подходим к теме, заявленной в названии статьи – к проблеме доверия. Как может человек довериться своему цифровому двойнику, который к тому же является частью искусственной системы? Положительное отношение может быть достигнуто при обеспечении объективности, объяснимости и прозрачности всех процессов – и эта проблема сейчас плодотворно исследуется в контексте этики ИИ [1]. Однако доверие имеет еще одно немаловажное измерение – социальное. Является ли автономный агент прикладным инструментом, представителем человека, олицетворением его сущности, и следовательно, этично ли вообще использовать цифровых двойников? Дискуссии по этому поводу наиболее ярко проявляются в произведениях массовой культуры, которые мы упоминали выше.

17 Таким образом, положительный потенциал интеллектуальных систем может быть обеспечен только при доверии к ним со стороны общества. Необходимо его укрепление, основанное на уважении прав человека и гражданина, для чего требуется обеспечить прозрачность, подотчетность, справедливость и регулирование – особенно это важно при применении таких систем для принятия решений в социальной сфере в целях достижения общественного блага. Если нет доверия, то не может быть и пользы – и в этом сегменте доверие представляется не просто абстрактным конкурентным преимуществом, но и необходимым условием. По словам исследователя Р. Боцман, доверие представляет собой «социальный клей», позволяющий человечеству прогрессировать благодаря взаимодействию друг с другом и с окружающей средой, включающей технологии [13].

18 Еще одним фактором, необходимым для достижения доверия общества к искусственному обществу – это возможность контроля. Во многом этот аспект связан со страхом появления «суперинтеллекта», который, превзойдя человеческие способности, сможет взять под контроль человечество. Необходимо обеспечить, чтобы в самой основе технологий были заложены гуманитарные ценности и человеческие потребности, и донести это положение для всего общества. Здесь гарантией того, что технология служит на благо человека и ее развитие остается под контролем, может стать включение соответствующих специалистов в мониторинг всех этапов процесса – принятия решений, анализ, проверки моделей и перепроверки результатов решений ИИ. Для достижения объективных результатов в такой работе может потребоваться привлечение межфункциональных групп, в состав которых войдут профессиональные эксперты, инженеры, администраторы, исследователи пользовательского опыта, юристы и др., чтобы качественно оценить возможные

непредвиденные последствия [10].

19 Таким образом, чтобы искусственные интеллектуальные системы содействовали решению сложных социальных задач, необходимо обеспечить доверие между всеми заинтересованными участниками. Прозрачность работы, высокие стандарты и доверие позволят достичь эффективный результат при применении сложных технологий для решения проблем, значение которых сложно переоценить, и цена которых колоссальна. Главное при этом – не вынести за скобки человека, его сущность и его ценности, то ключевое, что составляет суть формирования социогуманитарной цивилизации, которая базируется на решении проблемы становления субъектности развития человечества [4].

Библиография:

1. Апресян Р.Г. Этика и дискуссии об искусственном интеллекте / XI международная конференция «Теоретическая и прикладная этика: Традиции и перспективы - 2019. К грядущему цифровому обществу. Опыт этического прогнозирования (100 лет со дня рождения Д. Белла - 1919-2019)». Санкт-петербургский Государственный Университет, 21-23 ноября 2019 г. Материалы конференции / Отв. ред. В. Ю. Перов. СПб.: ООО «Сборка», 2019.
2. Денисова О. Ю., Мухутдинов Э.А. Большие данные - это не только размер данных // Вестник технологического университета. 2015, Т.18, № 4.
3. Конькова (Кураева) Т.А. Агентное моделирование процесса голосования // Искусственные общества. 2018, Т. 13, № 4. URL: <https://artsoc.jes.su/s207751800000138-0-1/>
4. Лепский В.Е. От техногенной к социогуманитарной цивилизации // Проектирование будущего. Проблемы цифровой реальности: труды 1-й Международной конференции (8-9 февраля 2018 г., Москва). М.: ИПМ им. М.В.Келдыша, 2018.
5. Макаров В.Л. Искусственные общества и будущее общественных наук. СПб. : Изд-во СПбГУП, 2009. (Избранные лекции Университета; Вып. 99).
6. Митио Какую: Искусственный интеллект превратит нас в волшебников // ФГБУ «Редакция «Российской газеты» URL: <https://rg.ru/2020/11/20/mitio-kaku-iskusstvennyj-intellekt-prevratit-nas-v-volshebnikov.html>
7. Райзберг Б.А., Лозовский Л.Ш., Стародубцева Е.Б. Современный экономический словарь. 2-е изд., испр. М.: ИНФРА-М. 1999.
8. «Управленческие эксперименты лучше опробовать на искусственном обществе» / Академик РАН Валерий Макаров — о том, как работает цифровой двойник социума // Коммерсантъ URL: <https://www.kommersant.ru/doc/4433177>
9. Фромм Э Бегство от свободы. Человек для себя / Эрих Фромм; пер. с англ. М.: АСТ: АСТ МОСКВА, 2006.
10. Applying artificial intelligence for social good // McKinsey Global Institute URL: <https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good>
11. Building Trust In Artificial Intelligence // Journal of International Affairs URL: <https://jia.sipa.columbia.edu/building-trust-artificial-intelligence#2>

12. Digital Twins Are Useful, But Can You Trust Them? // Miller Klein Associates Ltd URL: <https://miller-klein.com/2019/03/18/digital-twins-are-useful-but-can-you-trust-them/>
13. How-To Build Trust in Artificial Intelligence Solutions // MEDIUM URL: <https://towardsdatascience.com/how-to-build-trust-in-artificial-intelligence-solutions-83ca20c39f0>
14. Korzybski A. A Non-Aristotelian System and its Necessity for Rigour in Mathematics and Physics. Доклад в Американском математическом обществе (Новый Орлеан, штат Луизиана, США). 28 декабря, 1931 года. Перепечатано в Science and Sanity. 1933
15. Shifting toward Enterprise-grade AI: Resolving data and skills gaps to realize value». IBM Institute for Business Value, (IBM, 2018).
16. Social Issues with Digital Twin Computing // NTT Technical Review URL: <https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202009fa5.html>
17. The digital power to simulate an entire society, and make good guesses at our motivations, is nearly here. What do we do with it? // The Alternative UK URL: <https://www.thealternative.org.uk/dailyalternative/2019/10/5/simulatesociety-whattodo>
18. The Pandemics in Artificial Society: Agent-Based Model to Reflect Strategies on COVID-19 // Munich Personal RePEc Archive URL: https://mpra.ub.uni-muenchen.de/102075/1/MPRA_paper_102075.pdf

Artificial Intelligent Systems and Social Problems Solving: Challenge of Trust

Oleg Gurov

*RAS Institute of Philosophy, IOM RANEP
Russian Federation, Moscow*

Abstract

We expect a breakthrough in the improvement and subsequent widespread use of artificial society models, on which it will be possible to test large-scale managerial and economic decisions. Such technologies are already used as a tool for approbation of managerial decisions, and it is obvious that its use for solving social problems requires comprehensive and balanced attention, since it affects the interests of society and the individual. The common denominator of all aspects of this activity is the phenomenon of trust. To solve complex social problems with the help of artificial intelligent systems, it is necessary to ensure trust between all the stakeholders. Transparency, high standards and trust will allow achieving an effective result, which importance can be hardly overestimated, since the price of the issue is the successful formation of a socio-humanitarian civilization.

Keywords: artificial intelligence, artificial intelligence system, ethics, digital ethics, digital twin, computer model, trust, social objectives

Date of publication: 22.03.2021

Citation link:

Gurov O. Artificial Intelligent Systems and Social Problems Solving: Challenge of Trust // Artificial societies. – 2021. – V. 16. – Issue 1 [Electronic resource].

URL: <https://artsoc.jes.su/s207751800014095-3-1/> (circulation date: 24.03.2021). DOI: 10.18254/S207751800014095-3

Код пользователя: 20068; Дата выгрузки: 24.03.2021; URL - <http://artsoc.jes.su/s207751800014095-3-1/> Все права защищены.