

ПАНОРАМА

УДК 004.896 + 17.037 + 51.77

DOI: 10.32326/2618-9267-2023-6-3-145-161

АВТОМАТИЗИРОВАННОЕ ПРИНЯТИЕ РЕШЕНИЙ В СОЦИАЛЬНОЙ СФЕРЕ: ПРОБЛЕМА ВЫНУЖДЕННОГО ДОВЕРИЯ

Кеосиди Игорь Игоревич – аспирант. Институт подготовки кадров высшей квалификации, Пятигорский государственный университет. Российская Федерация, 357532, г. Пятигорск, пр. Калинина, д. 9;
e-mail: keosidi10@gmail.com;
ORCID 0009-0003-6702-4365

Кеосиди Юрий Игоревич – аспирант. Институт подготовки кадров высшей квалификации, Пятигорский государственный университет. Российская Федерация, 357532, г. Пятигорск, пр. Калинина, д. 9;
e-mail: keosidi@yahoo.com;
ORCID 0009-0009-0550-5431

В XXI в. общественная жизнь претерпевает значительные изменения. Ежедневно генерируется гигантское количество информации о каждом из жителей планеты, которая затем используется различными алгоритмами с целью управления социальными процессами. Современные способы управления социальными процессами в большинстве случаев опираются на алгоритмы. Развивается нейросетевой подход, основанный не на полном описании алгоритма с помощью кода, а на машинном обучении. Такой подход позволяет значительно ускорить рост качества управления социальными процессами, поскольку не требует эмпирического познания. Наряду с преимуществами, этот процесс несёт в себе и угрозы. Сама модель GPT-3 в ответ на прямой вопрос об опасностях, которые она несёт, называет четыре ключевых компонента: предрасположенность к уже принятым ранее решениям, поскольку обучается на основе исторических данных; замену человеческого труда; стирание границ между личной и общедоступной информацией; отсутствие эмпатии, лишь сухое выполнение поставленной задачи. Они и формируют цель данной статьи – получить ответ на вопрос: может ли человек на современном этапе развития технологий доверять решениям, принятым алгоритмами, не видя всех причинно-следственных связей, преобразующих входящие данные? Эмерджентность – наличие у системы свойств, не присущих её компонентам по отдельности, – с одной стороны, является той отличительной особенностью, которая позволяет современным системам алгоритмического принятия решений быть достаточно точными и надёжными, но с другой стороны, именно эмерджентность усложняет понимание механизма работы автоматизированной системы. Это приводит к противоречию: степень доверия к алгоритмически принятым решениям, как мы полагаем, снижается тогда, когда увеличиваются области их применения.

Ключевые слова: информационное общество, социальные процессы, алгоритм, алгоритмизация общества, искусственный интеллект, нейросеть

Цитирование: Кеосиди И.И., Кеосиди Ю.И. Автоматизированное принятие решений в социальной сфере: проблема вынужденного доверия // Цифровой ученый: лаборатория философа. 2023. Т. 6. № 3. С. 145-161. DOI: 10.32326/2618-9267-2023-6-3-145-161

Рукопись получена: 1 августа 2023

Пересмотрена: 10 октября 2023

Принята: 12 октября 2023

ALGORITHMIC DECISION-MAKING IN THE SOCIAL SPHERE: THE PROBLEM OF FORCED TRUST

Igor I. Keosidi – PhD student.
Institute for Training of Higher
Qualifications, Pyatigorsk State
University, 9, Kalinina ave.,
357532 Pyatigorsk, Russian
Federation;
e-mail: keosidi10@gmail.com;
ORCID 0009-0003-6702-4365

Yuri I. Keosidi – PhD student.
Institute for Training of Higher
Qualifications, Pyatigorsk State
University, 9, Kalinina ave.,
357532 Pyatigorsk, Russian
Federation;
e-mail: keosidi@yahoo.com;
ORCID 0009-0009-0550-5431

In the 21st century, social life is undergoing significant changes. A huge amount of information is generated daily about each inhabitant of the planet, which is then used by various algorithms to manage social processes. Modern methods of managing social processes mostly rely on algorithms. The neural network approach is developing, based not on a complete description of the algorithm, but on machine learning. This approach allows to significantly accelerate the growth of the quality of management of social processes, since it does not require empirical knowledge. Along with advantages, this process also carries threats. The GPT-3 model itself, in response to a direct question about the dangers it poses, names four key components, each of which forms the goal of the article - to get an answer to the question: can a person trust decisions made by algorithms at the current stage of technological development without seeing all the cause-and-effect relationships that transform incoming data? Emergence behavior – the ability of a system to demonstrate properties that are not inherent in its components separately, is a distinctive feature that allows modern systems of algorithmic decision-making to be quite accurate and reliable. However, on the other hand, it is emergent nature that complicates the understanding of the mechanism of automated system operation. This leads to a contradiction: the extent of confidence in algorithmic decisions, as we believe, decreases inversely proportional to the growth of their application area.

Keywords: information society, social processes, algorithm, algorithmization of society, artificial intelligence, neural network

How to cite: Keosidi, I.I., Keosidi, Yu.I. (2023). Algorithmic decision-making in the social sphere: the problem of forced trust, *The Digital Scholar: Philosopher's Lab*, 6 (3): 145-161. DOI: 10.32326/2618-9267-2023-6-3-145-161 (In Russian)

Received: 1 August 2023

Revised: 10 October 2023

Accepted: 12 October 2023

Введение

С развитием технологий появилась система алгоритмического принятия решений, которая принесла не только повышение эффективности управления, но и проблему доверия к принятым решениям. В случае, когда решение принимается живым человеком, группа лиц, на которых оно направлено, очевидно, считает его ответственным за принятое решение. А что делать, если решение принято машиной? Существует три возможных способа распределения ответственности при передаче подготовки решения алгоритму:

- всю ответственность несёт человек;
- ответственность делится пополам;
- всю ответственность несёт машина.

Под ответственностью машины в данном случае понимаются полное отстранение человека и восхождение ответственности к самой сути алгоритма, описывающей правила его работы. Принятые решения не всегда удовлетворяют тех, на кого они направлены. И если ответственным за решение (хотя бы частично) назначен алгоритм (то есть нечто обезличенное в глазах пользователя), то может возникать чувство незащищённости и (или) неуверенности в таком способе взаимодействия

Цель статьи – получить ответ на вопрос: может ли человек на современном этапе развития технологий доверять решениям, принятым алгоритмами, не видя всех причинно-следственных связей, преобразующих входящие данные. В процессе написания статьи вопрос будет исследоваться в разных аспектах, например:

- есть ли способ или по крайней мере предпосылки существования способа гарантировать безошибочную работу алгоритма?
- какой доверительный интервал считать удовлетворительным?
- имеет ли смысл сравнивать результат, полученный с помощью алгоритма, с неким абстрактным идеальным показателем? Или нужно сравнивать с тем результатом, который будет достигнут путём решения той же задачи человеком или группой людей?

На основе трудов Ю. Хаяши, Й. Масуды (Masuda, 1981), Д. Белла (Bell, 1999) Ф. Уэбстера (Webster, 2001) мы сформировали определение информационного общества, от которого будем отталкиваться в данной статье. Информационное общество – это тип общества, характеризующийся высокой степенью зависимости качества управленческих решений от объёма и качества информации, описывающей отдельно взятые субъекты общества. Цифровое общество, являющееся подвидом информационного, можно охарактеризовать как общество, в котором оцифровка данных сильно влияет на каждый аспект жизни: как мы общаемся и взаимодействуем, как мы работаем, учимся, остаёмся здоровыми и участвуем в политике и экономике (Digital Society, web). Современное обще-

ство является одновременно информационным и цифровым, поскольку характеризуется высокой степенью зависимости управленческих решений от качества и количества доступной информации, а также высокой степенью цифровизации социальных процессов.

Предметом исследования является проблема вынужденного доверия общества к тем решениям, которые принимаются или непосредственно вычислительными технологиями, или людьми на основе результатов, полученных с помощью этих технологий – компьютеров, роботов, вычислительных машин.

Системы автоматизированного принятия решений: от инженерного к гуманитарному измерению

Системы алгоритмического принятия решений (АПР) основаны на сборе и анализе больших объёмов информации (Newell, Marabelli, 2015). Информация может быть получена из различных источников. Если для принятия решения в относительно изолированном пространстве (например, внутри одной семьи или небольшой группы лиц) достаточно информации, которую пользователи соберут и внесут самостоятельно, то для решений, затрагивающих большее количество лиц, используется *big data*. Этот термин включает множество значений, которые имеют общие начала. На основе положений книги (Mayer-Schönberger, Cukier, 2014) будем использовать термин *big data* для описания той информации, которая удовлетворяет следующим условиям:

- её объём настолько велик, что для сбора и анализа требуется автоматизированная система;
- она постоянно обновляется, а объём – увеличивается;
- количество респондентов достаточно велико, а их выборка разнообразна для того, чтобы делать выводы о больших группах населения, таких как страна, регион, континент, вся планета;
- она добывается без осознанного согласия людей (это условие не всегда является обязательным, но, как правило, применимо).

В статье Д. Лэйни (Laneу, 2001, web) описаны три V, сочетание которых необходимо для качественной *big data*: объём (*volume*), разнообразие (*variety*) и скорость получения (*velocity*). Впоследствии автор добавил ещё несколько V, однако для целей нашей статьи достаточно этих ключевых характеристик.

Решения, принятые с помощью АПР, обладают двумя характеристиками (Mayer-Schönberger, Cukier, 2014):

- основаны на *big data* или другом наборе информации, который по своим характеристикам схож с описанием *big data*;
- доступ лиц, принимающих решение, к правилам, по которым алгоритмы формируют свои «предложения», зачастую закрыт или игнорируется.

На наш взгляд, эти характеристики верны именно для систем АПР, основанных на анализе *big data*. Локальные АПР могут быть

основаны на ручном вводе входной информации. Все дальнейшие рассуждения будем проводить именно для *big data*, так как социальные процессы не могут быть локализованы и требуют её использования.

Как видно из определения и второй характеристики, правила работы современных систем АПР не могут быть полностью описаны человеком, что означает, что системы АПР являют собой так называемые чёрные ящики. Пользователи видят входящую информацию и результат, а сам процесс отдан алгоритму. АПР основано на истории. Чем больше информации доступно, тем больше статистики использует система для выведения общего правила. Это нейросетевой подход, который и дал новый толчок развитию алгоритмизации после двух десятилетий зимы искусственного интеллекта.

Мы будем придерживаться определения Оксфордского словаря относительно искусственного интеллекта (Oxford University Press, 2008, web). Искусственный интеллект – это теория и разработка компьютерных систем, способных выполнять задачи, обычно требующие человеческого интеллекта, такие как визуальное восприятие, распознавание речи, принятие решений и перевод с одного языка на другой.

Понятие алгоритма мы разработали самостоятельно на основе исследованной литературы (Алексеев, 2021; Брагилевский, 2019, web, с. 7). Алгоритм – это набор точных шагов для решения поставленной задачи, чёткое следование которым приведёт к желаемому результату.

На основе определений алгоритма и АПР можно заключить, что проблема доверия – это проблема, вызванная именно АПР, поскольку алгоритм в классическом понимании является скорее белым ящиком – системой правил, формализованных и описанных человеком. Однако системы АПР для достижения целей могут создавать собственные алгоритмы, которые уже являются чёрными ящиками. В дальнейшей работе под термином алгоритм будем понимать именно такие алгоритмы, которые не являются полностью формализованными человеком.

Мы будем опираться на понятие «алгоритмическое управление», данное Р. Питерсом (Schuilenburg, Peeters (eds.), 2021), которое обозначает замену человека в принятии решений, рациональные и расчётливые выводы с помощью алгоритмов так называемого чёрного ящика. По аналогии с понятием «информационное общество» мы определим «алгоритмическое общество» как тип общества, характеризующийся высокой степенью зависимости от решений, принятых вычислительными компонентами информационных технологий или на их основе.

Все абстрактные вычислительные компоненты информационных технологий обладают двумя общими характеристиками – упрощение и закрытие. Упрощение – это разделение процесса на множество более мелких подпроцессов, выполняемых последова-

тельно. Закрытие – это изоляция этих подпроцессов, в частности и всего процесса в целом, гарантирующая их выполнение без вмешательства извне (Kallinikos, 2005).

В инженерном ракурсе алгоритм является закрытой системой правил, которая принимает на вход заданные параметры, проводит их через все описанные действия и выдаёт результат. Для пользователя это выглядит как отображение одного множества в другое. Оно может быть инъективным или сюръективным, однако всегда является конечным процессом. Каждому набору входных параметров соответствует ровно один результат. Если алгоритм создан для прогнозирования, то результатом также является однозначный набор данных, однако применяется доверительный интервал, поскольку заданная ситуация была смоделирована, и ответ – это прогноз. В основе алгоритмического прогнозирования, как и любого прогнозирования, лежит теория вероятности и комбинаторика. Насчёт применения этих принципов в социальных процессах более корректно будет утверждать, что чаще используется наука, родственная теории вероятности – математическая статистика. Один из основных методов математической статистики – это метод максимального правдоподобия, который заключается в том, чтобы количество исследуемых элементов множества, попавшее в выборку, было достаточно для корректного вывода обо всём множестве. Очевидно, чем больший процент составляет количество элементов, попавших в выборку, тем точнее будет результат. Есть ограничение «снизу» – число, меньше которого выборка не будет иметь смысла. Однако всегда существуют и ограничения «сверху» в виде:

- доступности информации;
- стоимости получения информации;
- скорости обработки информации.

Как следствие, возникает задача достижения баланса, при котором, с одной стороны, прогнозирование достаточно точно для заданной цели, а с другой – затраты на весь процесс окупаются эффектом от достижения цели. При решении сугубо формальных математических задач, как нам кажется, указанный баланс всегда достижим, поскольку можно оцифровать важность задачи, цену ошибки и стоимость уточнения результата. В случае с социальными процессами таких однозначных ответов дать невозможно. Рассмотрим процесс управления и алгоритмизации в гуманитарном ракурсе.

М. Фуко в начале 1980-х гг. ввёл неологизм «правительственность» (governmentality) (Foucault, 1982; 2016), который он определил как совокупность методов реализации власти, включающую в себя, с одной стороны, развитие правительственных аппаратов, а с другой – определённый набор знаний об управлении обществом. Концепция «правительственности», по мнению М. Фуко, совершенно отличается от привычного концепта власти, представляющего собой совокупность политических идеологий и философий. Фуко в большей степени фокусировался на взаимосвязях

между правительством и «управляемыми», то есть на самих способах управления повседневной жизнью, включающих практики, техники, механизмы и аппараты управления, а также знания о субъектах управления. Он положил начало новой области исследований под названием «исследования правительственности», объектом которой стало, по обозначению Фуко, «поведение поведения» (Foucault, Sheridan, 2012). Фуко утверждал, что «правительство» – это особая форма власти, возникшая в Средние века. Центральное место в подходе Фуко занимает использование более древнего значения слова «правительство», которое возникло в отношении не государственных организаций, а практик управления. Под «правительством» он понимал «структурирование возможного поля действий других» или «рассчитанное направление поведения» (Foucault, 1982, С. 18). В таком контексте правлением можно называть деятельность не только глав государств, но и работодателей, учителей, родителей и прочих людей, своими действиями или правилами направляющих поведение других. Исследования правительственности Фуко – это возникновение рационализированных знаний о способах управления обществом. Предметом исследования становится сам процесс управления, отличный от политической философии, которая была темой большинства исследований общественного устройства, а технологии управления в них рассматривались как нейтральные инструменты.

Социологи П. Миллер и Н. Роуз (Miller, Rose, 2013) разделили правительственность на две составляющие: политическую рациональность и технологию управления. Объединяя эмпирические работы по управлению экономической, социальной и личной жизнью, они рассмотрели всю социальную сферу как совокупность взаимосвязанных процессов, изменение одного из которых влечёт за собой неизбежные изменения в других. Воздействуя на экономику, можно достичь желаемого результата, например, в показателях здоровья населения.

Мы рассмотрели работы о правительственности, чтобы продемонстрировать, с одной стороны, важность проблемы исследования способов управления обществом, а с другой стороны, невысокую степень разработки проблемы в современных реалиях, в период экспоненциального роста технологического прогресса.

Управление данными и проблема перехода количества в качество

Информация присутствовала в жизни людей всегда. Но если до XX в. информация являлась средством обмена данными или их накопления, то начиная с XX в. количество накопленной и регулярно создаваемой информации стало настолько большим, что сам процесс управления информацией стал отдельной функцией работы, например, предприятий. Информация стала активом, который может быть использован более или менее эффективно в зависимо-

сти от организации управления данными. О. Гиацинтов (Гиацинтов, 2019) выделяет корпоративную функцию управления данными как одну из ключевых в работе современных предприятий. В масштабах организаций, государств или органов управления социальными процессами задача эффективного управления имеющейся информацией нередко становится, на наш взгляд, важнее самого наличия информации.

Как и все автоматизированные процессы, алгоритм управления чем-либо должен быть полностью описан на математическом языке. Например, наиболее простой вид алгоритма – это последовательность действий, включающая несколько шагов с разветвлением на возможные варианты ответа. В языках программирования эти операции обычно носят характер «если..., то...» (*if... then*). Одно из важнейших свойств автоматизированного принятия решений – это возможность быстро перебрать большое количество вариантов, что значительно ускоряет получение результата.

Предпосылки возникновения искусственного интеллекта с точки зрения философии были заложены родоначальниками рационализма Б. Спинозой (Спиноза, 2018) и Р. Декартом (Декарт, 1989). Несмотря на разные взгляды относительно связи разума и тела, оба философа предполагали возможным создание такой системы символов, которая станет фундаментом того, что сегодня называется искусственным интеллектом. Технологические предпосылки возникли также в XVII в. с созданием арифмометра Г. Лейбница и машины Б. Паскаля. Позднее, в XVIII в., математики искали способ описания всех законов динамики исключительно аналитическим путём без использования ручного опыта или иллюстрирующих рисунков (Euler, 1984; Truesdell, Euler, 1960; Даламбер, 1950). Затем А. Тьюринг, определивший понятие «алгоритм» как всё, что умеет делать машина Тьюринга (Turing, 1937), предложил тест, призванный ответить на вопрос «может ли машина мыслить?» (Turing, 1950). Суть теста сводится к тому, сумеет ли машина в процессе коммуникации с живым человеком обмануть его и заставить поверить в то, что с ним взаимодействует человек, а не машина. Ни одна машина, работающая на классическом алгоритме белого ящика, не прошла тест.

В середине XX в. появилось новое направление развития искусственного интеллекта. Параллельно традиционным математическим методам описания действий алгоритмов учёные стали использовать нейронные сети, которые способны относительно самостоятельно обучаться на основе существующей информации. В частности, в 1962 г. было предложено обучение на примерах (Malsburg, 1986), а в 1965 г. сформулировано правило резолюций (Robinson, 1965), описывающее метод доказательства через противоречия, которое легло в основу создания языков программирования, решающих задачи методом перебора бесчисленного количества вариантов. Таким образом, стало возможным перейти от ко-

личества к качеству не путём математического описания, а с помощью погружения машины в реальный мир и её адаптации.

В конце XX в. наступили так называемые две «зимы» искусственного интеллекта, отделённые друг от друга пятилетним периодом в начале 80-х гг., во время которого вера в потенциал искусственного интеллекта возобновилась благодаря появлению возможности его использования в коммерческих целях. Сами «зимы» заключались в том, что учёные приостанавливали работу над изучением искусственного интеллекта, считая её бесперспективной. Причинами «зим» явились:

- исследования, которые показали экспоненциальный рост количества вариантов при незначительном увеличении входных параметров;

- несколько неудачных примеров запуска широко анонсированных государственных проектов различных стран (Марголин, Дубровская, 2018).

Начиная с середины 1990-х гг., подход к обучению и, как следствие, методам работы систем АПР изменился. Учёные отошли от идеи создать искусственный интеллект, который бы функционировал как человек. Вместо этого для удовлетворения точечных потребностей, таких как, например, распознавание лиц, поддержка пользователя в форме чат-бота, стали использоваться нейросети. Ключевое отличие нового подхода заключается в том, что авторы систем АПР больше не пытаются описать все возможные сценарии с помощью формального языка. Вместо этого они используют *big data* для обучения машины. Нейросеть обучается на основе входной информации, а не на основе правил. Чем больше информации доступно для обучения, тем выше точность работы системы. В основе работы лежит не закрытая, а вероятностная модель. Маchine больше не пытаются придать свойства человека. Она остаётся машиной, но при этом количество данных для обучения настолько велико, что точность результата достаточна для использования. *BERT* от *Google* и *GPT-2* от *OpenAI*, выпущенные в 2018 г., уже умели улавливать контекст разговора. Это означало, что ответ на один и тот же вопрос, заданный разными пользователями, отличался в зависимости от содержания диалога, предшествовавшего запросу. Эти модели в своё время казались прорывом, но всё же была очевидна их «машинность». В частности, она проявлялась в конечном количестве возможных вариантов ответа. Однако в 2023 г. с появлением *GPT-3* ситуация изменилась. Вторая версия была основна на 1,5 млрд параметров, а третья – на 175 млрд (“*GPT-2 vs. GPT-3, web*”), что и объясняет колоссальный прирост точности ответов, разнообразия действий и, как следствие, повышения производительности и «интеллектуальности». Именно увеличение количества входных параметров привело к росту эмерджентности. Таким образом, можно заключить, что количество перешло в качество.

Проблема, вызванная развитием эмерджентных свойств нейросетей, заключается в том, что этическая составляющая, недоступная для машины, теряется, поскольку чем больше действий машина выполняет самостоятельно, тем меньшее вовлечение возможно для человека. Классический алгоритм, представляющий собой белый ящик, остался символьным, что не позволяет ему достичь уровня нейросети. Нейросеть, являющаяся чёрным ящиком, даёт лучший результат с точки зрения выполнения поставленной перед ним задачи, но порождает этическую проблему, которую можно разложить на две ключевые составляющие.

Первая часть проблемы заключается в том, что с развитием машинного обучения алгоритмически произведённые оцифрованные знания всё больше отделяются от человеческого опыта. Категории и действия алгоритмически генерируются в соответствии с логикой, которая определяется алгоритмом самостоятельно, а не задаётся человеком. Машинное обучение включает в себя установку многочисленных внутренних переменных и связей между ними, а также предоставление алгоритму обучающего входного материала в виде набора данных (например, имя, пол, возраст, уровень дохода, семейное положение, аккаунты в социальных сетях, фотографии) и выходной набор данных для обучения (например, кто из перечисленных людей получал Нобелевскую премию). Алгоритм неоднократно изменяет свои внутренние переменные и весовые коэффициенты между ними, чтобы наилучшим образом преобразовать входные данные в выходные. При этом алгоритм создаёт свои собственные правила, категории и оцифрованные знания для достижения наилучших ассоциаций между входными и выходными данными. Как результат, алгоритм выводит закономерности. Например, люди в возрасте от 35 до 50 лет с уровнем дохода от 10 до 15 тыс. долларов в месяц, замужем или женатые становятся лауреатами Нобелевской премии значительно чаще, чем все остальные. С одной стороны, это корректный вывод, а с другой – он может повредить качественному прогнозированию и исключить подходящих кандидатов из выборки. Это потенциальное заблуждение аналогично следующему. Большинство изобретений, открытий, научных достижений принадлежит мужчинам, из чего, на первый взгляд, можно сделать вывод о том, что мужчины сильнее в науке, чем женщины. Однако стоит принять во внимание следующие факты:

- долгое время женщинам либо запрещалось заниматься наукой, либо это не было принято;
- исторически женщины выполняли больше домашних обязанностей, чем мужчины (в силу обычаев, религии и др.), что оставляло им меньше свободного времени;
- летописи/архивы/исторические записи могли быть не беспристрастными.

Учитывая это, вывод перестаёт быть настолько очевидным. Для этого эффекта – как в жизни, так и в теории алгоритмов – есть тер-

мин «заблуждение/предубеждение, основанное на фактах», в английском – *bias*. Машинное обучение основано на исторических и (или) статистических данных. Эти данные могут быть интерпретированы алгоритмом как причинно-следственные связи, но на самом деле они таковыми не являются. Как следствие, все будущие суждения, основанные на неверном толковании прошлого, будут некорректны. Таким образом, первая часть этической проблемы заключается в том, что машинное обучение может быть предвзятым.

Вторая часть этической проблемы лежит в плоскости социотехнических систем и заключается в том, что любой результат – это, пусть и очень точная, но всё-таки аппроксимация. Если в случае с математическими моделями, это допустимо, то в случае с социальными процессами – нет. К примеру, если бы речь шла о математическом безэмоциональном процессе, то точность 95 %, которая приводит к сокращению необходимого количества работников в 20 раз, но зато допускает 5 % ошибок, могла бы считаться приемлемой. Однако когда ценой ошибки может стать жизнь человека, можем ли мы доверять «доверительному интервалу»? Примерами программ, работающих на базе нейросетей, могут являться:

- приложение «Родинка», определяющее злокачественные образования с помощью фото в домашних условиях;

- беспилотные такси.

Теперь перейдём к исследованию потенциальных рисков, вызванных использованием систем АПР в социальных процессах.

Автоматизированное принятие решений в социальной сфере (примеры, проблемы, методы решения, угрозы и риски)

Компьютеры в чистом виде исчезают, уступая место умным вещам (Чеклецов, 2015). Стационарная работа за компьютером, предполагающая чёткое разделение обычной и виртуальной жизни, уступает место непрерывному взаимодействию с продуктами информационных технологий, что и формирует проблему, которую частично решить может данная статья.

Предвосхищающее алгоритмическое управление выходит за рамки традиционного применения процедур и правил, а вместо этого вычисляет будущие реалии на основе вероятностной логики. Например, на основе введённых вами первых букв в поисковую строку, учитывая историю ваших прошлых поисков, геолокацию и т. д., алгоритмы *Google* выдают результаты, которые отличаются от результатов других людей, напечатавших такие же символы в строке поиска.

Приведём примеры внедрения алгоритмов на государственном уровне. Часто они используются в принудительном управлении наиболее маргинализированными слоями населения. Службы защиты детей и их социального обеспечения сравнительно рано начали использовать алгоритмы предвидения для классификации риска для детей в плане жестокого обращения или безнадзорности.

Ф. Гиллингем (Gillingham, 2019; 2006) изучал внедрение и эволюцию оценок на основе риска в Австралии, Новой Зеландии и Великобритании, начиная с форм компьютерных экранных потоков и заканчивая более автоматизированными системами. Гиллингем наблюдал, как такие системы поддержки принятия решений создают оцифрованные знания о будущей жизни ребенка на основе количественных обезличенных данных о нём. С одной стороны, результаты анализа были достаточно точны, но с другой – они одновременно бросали вызов более тонким и глубоким профессиональным знаниям работников системы социального обеспечения детей. Система автоматизированного принятия решений фактически становится основой для принятия решения об изъятии ребенка у родителей, что чревато серьёзными последствиями для детей в случаях как ложноотрицательных, так и ложноположительных результатов (Keddel, 2016).

Аналогичные модели используются и в уголовном правосудии для того, чтобы дополнить или заменить профессиональные человеческие суждения, например, работников полиции или судей. Развивается предиктивная полицейская деятельность (Smith et al., 2017). Алгоритмы используют аналитику данных, моделирование и машинное обучение для прогнозирования и предотвращения преступлений, а также для вынесения вердиктов о виновности или невиновности подсудимого на основе данных о нём и машинном обучении, использующем данные о людях, совершавших преступления в прошлом. Здесь также существует вероятность ошибочного суждения на основе *bias*, поскольку, очевидно, в таком случае, человек с одним набором антропологических данных (схожих с прошлыми преступниками) будет с большей вероятностью признан виновным, чем другой.

В некоторых городах применяется алгоритмический расчёт количества полицейских, необходимых для патрулирования улиц в зависимости от времени суток, района, месяца и других параметров. В таком случае каждый гражданин не ощущает на себе прямую никакого воздействия со стороны государства, но участвует в общественной жизни, некоторые параметры которой определены алгоритмом. Человек, проживающий в статистически более благополучном районе, платит такие же налоги, как и проживающие в районах, острее нуждающихся в полиции, но получает это общественное благо в меньшей степени.

Обобщим рассмотренные в статье особенности использования систем алгоритмического принятия решений.

Механизм работы подобных алгоритмов являет собой чёрный ящик, что порождает следующие риски. Невозможно быть уверенным в правильности работы системы, поскольку машинное обучение основано на исторических данных, которые могут быть неверно интерпретированы. Однако с каждым днём алгоритмическим системам принятия решений делегируется всё больше ответственности, на основе их предложений могут быть сделаны очень серьёз-

ные, необратимые выводы. Таким образом, проблема вынужденного доверия действительно актуальна.

Заключение

Как мы видим, системы АПР прочно вошли в жизнь современного человека. С каждым шагом своего развития системы АПР становятся, с одной стороны, всё точнее, а с другой – всё автономнее и неподвластнее для понимания человеком их причинно-следственных связей. Рост вычислительной мощности даёт рост релевантности ответов, однако не решает проблему чёрного ящика, а, напротив, лишь усугубляет её, поскольку точность, удобство и эффективность вынуждают расширять использование АПР.

Приведённые примеры показывают, что качества алгоритмов (беспристрастность, последовательность, объективность), которые являются преимуществами относительно человека, в некоторых ситуациях могут оказаться недостатками. Это приводит к следующему вопросу, который, как нам кажется, задают себе исследователи во всех областях: как из многогранного процесса научиться извлекать лишь положительные стороны, отделяя их от побочных негативных эффектов?

На наш взгляд, при современном уровне развития технологий ответ на поставленный в начале статьи вопрос – нет, человек всё еще не может полностью доверять решениям, принятым алгоритмами, не видя всех причинно-следственных связей, преобразующих входящие данные. Мы придерживаемся мнения, что даже если на бесконечном количестве выборок (примеров, проб) машина будет выдавать верный ответ, люди не станут полностью доверять ей. Путь к доверию лежит в возможности на любом этапе посмотреть, как именно принято то или иное решение. Под конечным решением мы понимаем возможность вмешаться в работу алгоритма на любом этапе, иметь доступ ко всем входным данным и к самим логическим связям, заложенным в алгоритм. Иначе говоря, мы придерживаемся мнения, что на данный момент человек не может доверять чёрному ящику.

Данный вывод нами сделан на основе изучения теоретических основ алгоритмического управления и примеров его реализации в социальной сфере. Мы предполагаем, что на сегодняшний день в социальных процессах, требующих индивидуального подхода, алгоритмы могут и должны использоваться с целью повышения скорости, точности и, как следствие, эффективности, однако конечное решение должен принимать человек.

Информация о конфликте интересов

Автор заявляет об отсутствии конфликта интересов.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Список литературы

Алексеев, 2021 – *Алексеев В.В.* Теория алгоритмов. Учебно-методическое пособие. СарФТИ НИЯУ МИФИ, Саров, 2021. 98 с.

Брагилевский, 2019, web – *Брагилевский В.Н.* Лекции по теории алгоритмов. Конспект лекций. СПб, 2019. 113 с. URL: https://gsom.spbu.ru/images/cms/data/teoriya_algoritmov.pdf (дата обращения: 31.05.2023)

Гиацинтов, 2019 – *Гиацинтов О.* Управление данными для эффективной цифровой трансформации. М.: DIS Group, 2019. 51 с.

Даламбер, 1950 – *Даламбер Ж.Л.* Динамика / Пер. с франц. В.П. Егоршина. М.: Гостехиздат, 1950. 315 с.

Декарт, 1989 – *Декарт Р.* Сочинения в двух томах, Т.1, Философское наследие / Пер. с лат. и франц. В.В. Соколова. М: Мысль, 1989. 654 с.

Марголин, Дубровская, 2018 – *Марголин И., Дубровская Н.* Основные этапы развития искусственного интеллекта // Молодой ученый. 2018. Вып. 206. С. 23-26.

Спиноза, 2018 – *Спиноза Б.* Этика, доказанная в геометрическом порядке и разделённая на пять частей / Пер. с лат. Н. А. Иванцов. 2-е изд. М.: Академический проект, 2018. 334 с.

Чеклецов, 2015 – *Чеклецов В.* Динамические эмерджентные интерфейсы сложных социотехнических систем // Философские проблемы информационных технологий и киберпространства. 2015. № 9. С. 70-85. <https://doi.org/10.17726/philIT.2015.9.1.1.18>

Bell, 1999 – *Bell D.* The Coming of Post-industrial Society: a Venture in Social Forecasting. N.Y.: Basic Books, 1999. 507 p.

Dean, 2010 – *Dean M.* Governmentality: Power and Rule in Modern Society. L.: SAGE; Calif: Thousand Oaks, 2010. 294 p.

Euler, 1984 – *Euler L.* Elements of Algebra. N. Y.: Springer New York, 1984. 593 p., doi:10.1007/978-1-4613-8511-0

Foucault, 2016 – *Foucault M.* On The Government of the Living: Lectures at the Collège de France, 1979-1980. N.Y.: Picador, 2016. 260 p.

Foucault, 1982 – *Foucault M.* The subject and power // Critical Inquiry. 1982. Vol. 8. No. 4. Pp. 777-795.

Foucault, Sheridan, 2012 – *Foucault M., Sheridan A.* Discipline and Punish: the Birth of the Prison. N.Y.: Vintage, 2012. 333 p.

Gillingham, 2019 – *Gillingham P.* Decision support systems, social justice and algorithmic accountability in social work: a new challenge // Practice. 2019. No. 31. Pp. 277-290, doi:10.1080/09503153.2019.1575954

Gillingham, 2006 – *Gillingham P.* Risk assessment in child protection: problem rather than solution? // Australian Social Work. 2006. No. 59. Pp. 86-98, doi:10.1080/03124070500449804

GPT-2 vs. GPT-3, web – GPT-2 vs. GPT-3: Which is Better for Language Modeling? URL: <https://ts2.space/en/gpt-2-vs-gpt-3-which-is-better-for-language-modeling/> (дата обращения: 31.05.2023)

Kallinikos, 2005 – *Kallinikos J.* The order of technology: complexity and control in a connected world // Information and Organization 2005. No. 15. Pp. 185-202, doi:10.1016/j.infoandorg.2005.02.001

Keddell, 2016 – *Keddell E.* Substantiation decision-making and risk prediction in child protection systems // *Policy Quarterly*. 2016. Vol. 12. No. 2. Pp. 46-56, doi:10.26686/pq.v12i2.4587

Laney, 2001, web – *Laney D.* 3-D data management: controlling data volume, velocity and variety // *Application Delivery Strategies META Group Inc.* 6 February 2001. URL: <https://studylib.net/doc/8647594/3d-data-management--controlling-data-volume--velocity--an...> (дата обращения: 05.05.2023).

Masuda, Y., 1981 – *Masuda Y.* The Information Society as Post-industrial Society. Tokyo: Institute for the Information Society, 1981. 35 p.

Mayer-Schönberger, Cukier, 2014 – *Mayer-Schönberger V., Cukier K.* Big Data: a Revolution That Will Transform How We Live, Work, and Think. Boston: Mariner Books, Houghton Mifflin Harcourt, 2014. 252 p.

Miller, Rose, 2013 – *Miller P., Rose N.S.* Governing the Present: Administering Economic, Social and Personal Life. Cambridge Malden: Polity Press, 2013. 246 p.

Newell, Marabelli, 2015 – *Newell S., Marabelli M.* Strategic opportunities (and challenges) of algorithmic decision-making: a call for action on the long-term societal effects of “datification” // *The Journal of Strategic Information Systems*. 2015. No. 24. Pp. 3-14, doi:10.1016/j.jsis.2015.02.001

Oxford University Press. 2008. URL <https://global.oup.com/academic/?lang=en&cc=us> (дата обращения: 05.05.2023).

Robinson, 1965 – *Robinson J.A.* A Machine-oriented logic based on the resolution principle // *Journal of the ACM*. 1965. No. 12. Pp. 23-41, doi:10.1145/321250.321253

Schuilenburg, Peeters (eds.), 2021 – *Schuilenburg M., Peeters R.* (eds.). The Algorithmic Society: Technology, Power, and Knowledge. L.; N.Y.: Routledge/Taylor & Francis Group, 2021. 203 p.

Smith, Bennett Moses, Chan, 2017 – *Smith G.J.D., Bennett Moses L., Chan J.* The challenges of doing criminology in the big data era: towards a digital and data-driven approach // *The British Journal of Criminology*. 2017. Vol. 57. No. 2. Pp. 259-274, doi:10.1093/bjc/azw096

The innovative research programme of the Dutch universities // *Digital Society*. URL: <https://www.thedigitalsociety.info/> (дата обращения: 31.05.2023)

Truesdell, Euler, 1960 – *Truesdell C., Euler L.* The Rational Mechanics of Flexible or Elastic Bodies: 1638-1788. Turici: Auctoritate et impensis Societatis Scientiarum Naturalium Helveticae: Venditioni exponunt O. Füssli, 1960. 435 p.

Turing, 1950 – *Turing A.* Computing machinery and intelligence // *Mind, New Series*. 1950. No. 59. Pp. 443-460.

Turing, 1937 – *Turing A.M.* On computable numbers, with an application to the Entscheidungsproblem // *Proceedings of the London Mathematical Society*. 1937. Vol. s2-42. No. 1. Pp. 230-265, doi:10.1112/plms/s2-42.1.230

Malsburg, 1986 – *Malsburg C., van der.* Frank Rosenblatt: principles of neurodynamics: perceptrons and the theory of brain mechanisms // *G. Palm, A. Aertsen* (eds.), *Brain Theory. Proceedings of the First Trieste Meeting on Brain Theory*. Berlin; Heidelberg: Springer Berlin Heidelberg, 1986. Pp. 245-248, doi:10.1007/978-3-642-70911-1_20

Webster, 2001 – *Webster F.* Information society // *International Encyclopedia of the Social & Behavioral Sciences*. Elsevier, 2001. Pp. 7464–7468, doi:10.1016/B0-08-043076-7/04334-5

References

- Alexeev, V.V. (2021). *Theory of Algorithms*. Textbook. Sarov Institute of Physics and Technology of National Research Nuclear University MEPhI Publ. (In Russian)
- Bell, D. (1999). *The Coming of Post-industrial Society: a Venture in Social Forecasting*, special anniversary ed. ed. Basic Books.
- Bragilevsky, V.N. (2019). *Lectures on the Theory of Algorithms*: Lecture Notes. Retrieved May 31, 2023, from https://gsom.spbu.ru/images/cms/data/teoriya_algoritmov.pdf (In Russian)
- Giatsintov, O. (2019). *Data Management for Effective Digital Transformation*. DIS Group Publ. (In Russian)
- D'Alembert, J.L.R. (1950). *Dynamics* / transl. from French by V.P. Egorshin. State Publishing House of Technical and Theoretical Literature "GosTekhizdat" Publ. (In Russian)
- Dean, M. (2010). *Governmentality: Power and Rule in Modern Society*, 2nd ed. SAGE; Thousand Oaks.
- Descartes, R. (1989). *Essays in Two Volumes*, vol.1, Philosophical heritage / transl. from Latin and French by V.V. Sokolov. Mysl Publ. (In Russian)
- Euler, L. (1984). *Elements of Algebra*. Springer New York, doi:10.1007/978-1-4613-8511-0
- Foucault, M. (2016). *On The Government of the Living: Lectures at the Collège de France, 1979-1980 and Oedipal Knowledge*. Picador.
- Foucault, M. (1982). The Subject and Power. *Critical Inquiry* 8 (4), 777-795.
- Foucault, M., Sheridan, A. (2012). *Discipline and Punish: the Birth of the Prison*. Vintage.
- Gillingham, P. (2019). Decision Support Systems, Social Justice and Algorithmic Accountability in Social Work: A New Challenge. *Practice* 31, 277-290, doi.org:10.1080/09503153.2019.1575954
- Gillingham, P. (2006). Risk Assessment in Child Protection: Problem Rather than Solution? *Australian Social Work* 59, 86-98, doi:10.1080/03124070500449804
- GPT-2 vs. GPT-3: Which Is Better for Language Modeling? Retrieved May 31, 2023, from <https://ts2.space/en/gpt-2-vs-gpt-3-which-is-better-for-language-modeling/>
- Chekletov, V. (2015). Dynamic emergent interfaces of complex sociotechnical systems. *Philosophical Problems of Information Technologies and Cyberspace* 9, 70-85, doi:10.17726/phillIT.2015.9.1.1.18 (In Russian)
- Kallinikos, J. (2005). The order of technology: Complexity and control in a connected world. *Information and Organization* 15, 185-202, doi:10.1016/j.infoandorg.2005.02.001
- Keddell, E. (2016). Substantiation decision-making and risk prediction in child protection systems. *Policy Quarterly* 12 (2), 46-56, doi: 10.26686/pq.v12i2.4587
- Laney, D. (2001). 3-D Data Management: Controlling Data Volume, Velocity and Variety. *Application Delivery Strategies META Group Inc.* File 949. 6 February. Retried May 5, 2023, from <https://studylib.net/doc/8647594/3d-data-management--controlling-data-volume--velocity--an...>
- Malsburg, C., van der. (1986). Frank Rosenblatt: principles of neurodynamics: perceptrons and the theory of brain mechanisms, in: Palm, Aertsen, A. (Eds.), *Brain Theory* (245-248). Springer Berlin Heidelberg, doi: 10.1007/978-3-642-70911-1_20
- Margolin, I., Dubrovskaya, N. (2018). The main stages in the development of artificial intelligence. *Young Scientist* 206, 23-26. (In Russian)
- Masuda, Y. (1981). *The Information Society as Post-industrial Society*. Institute for the Information Society.

Mayer-Schönberger, V., Cukier, K. (2014). *Big Data: a Revolution That Will Transform How We Live, Work, and Think*, 1st Mariner Books ed. Mariner Books, Houghton Mifflin Harcourt.

Miller, P., Rose, N.S. (2013). *Governing the Present: Administering Economic, Social and Personal Life*. Polity Press.

Newell, S., Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification.' *The Journal of Strategic Information Systems* 24, 3-14, doi:10.1016/j.jsis.2015.02.001

Oxford University Press (2008). Retrived May 5, 2023, from <https://global.oup.com/academic/?lang=en&cc=us>

Robinson, J.A. (1965). A Machine-Oriented Logic Based on the Resolution Principle. *Journal of the ACM* 12, 23-41. <https://doi.org/10.1145/321250.321253>

Schuilenburg, M., Peeters, R. (Eds.) (2021). *The Algorithmic Society: Technology, Power, and Knowledge*. Routledge/Taylor & Francis Group.

Smith, G.J.D., Bennett Moses, L., Chan, J. (2017). The Challenges of Doing Criminology in the Big Data Era: Towards a Digital and Data-driven Approach. *The British Journal of Criminology* 57 (2), 259-274, doi:10.1093/bjc/azw096

Spinoza, B. (2018). *Ethics Proved in Geometrical Order and Divided into Five Parts* / transl. from Latin by N.A. Ivantsov. Academic Project Publ. (In Russian)

The innovative research programme of the Dutch universities Digit. Society. Retrieved May 31, 2023, from <https://www.thedigitalsociety.info/>

Truesdell, C., Euler, L. (1960). *The Rational Mechanics of Flexible or Elastic Bodies: 1638-1788*: Introduction to Leonhardi Euleri opera omnia vol. X et XI, seriei secundae. Auctoritate et impensis Societatis scientiarum naturalium Helveticae, Turici.

Turing, A. (1950). Computing machinery and intelligence. *Mind, New Series* 59, 443-460.

Turing, A.M. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society* s2-42 (1), 230-265, doi:10.1112/plms/s2-42.1.230

Webster, F. (2001). Information Society, in: *International Encyclopedia of the Social & Behavioral Sciences* (7464-7468). Elsevier, doi:10.1016/B0-08-043076-7/04334-5