УДК 004.891.2
Оригинальная статья
https://doi.org/10.21683/1729-2646-2025-25-2-50-58

Теория игр для автономных систем искусственного интеллекта при управлении корпорациями

Game theory for autonomous AI systems in corporate governance

Романова А.С.[1]
Romanova A.S.[1]

[1] *Московский физико-технический институт (национальный исследовательский университет), аспирант, г. Москва, Российская Федерация, romanova.as@phystech.edu, ORCID 0009-0004-4649-6037*
[1] *Moscow Institute of Physics and Technology (National Research University), Postgraduate Student, Moscow, Russian Federation, romanova.as@phystech.edu, ORCID 0009-0004-4649-6037*
romanova.as@phystech.edu

Романова А.С.

Резюме. *Несмотря на то, что современные технологии искусственного интеллекта в значительной степени основаны на машинном обучении, сами по себе алгоритмы машинного обучения стратегией не являются. Стратегией мы будем называть полное описание того, как система будет себя вести при всех возможных обстоятельствах. Наиболее перспективным инструментом, позволяющим автономным системам принимать эффективные решения при управлении корпорациями, представляется теория игр. Учитывая разнообразие проблем, с которыми приходится сталкиваться советам директоров, теория игр, нашедшая свое применение в экономике, политологии, чистой математике, психологии, социологии, маркетинге и финансах, предоставляет возможность обеспечивать автономность системы искусственного интеллекта на основе моделирования эффективной стратегии. Обязательное требование о разработке этичных и легитимных автономных систем искусственного интеллекта может привести к тому, что некоторые дилеммы самой теории игр для автономных систем или не существуют, или изменяют свой смысл.*
Abstract. *Despite the fact that modern artificial intelligence technologies are largely based on machine learning, machine learning algorithms themselves are not a strategy. We will define a strategy as a complete description of how a system will behave under all possible circumstances. Game theory appears to be the most promising tool that allows autonomous systems to make effective decisions when managing corporations. Given the variety of problems faced by boards of directors, game theory that has found application in economics, political science, pure mathematics, psychology, sociology, marketing and finance, potentially enables an autonomous artificial intelligence system based on effective strategy simulation. If the development of ethical and legitimate autonomous AI systems is mandatory, some dilemmas of game theory itself either become nonexistent or change their meaning in the context of game theory.*

Ключевые слова: *теория игр, искусственный интеллект, алгоритмические решения, корпоративное управление, андроиды.*
Keywords: *game theory, artificial intelligence, algorithmic solutions, corporate governance, androids.*

Для цитирования: *Романова А.С. Теория игр для автономных систем искусственного интеллекта при управлении корпорациями // Надежность. 2025. №2. С. 50-58. https://doi.org/10.21683/1729-2646-2025-25-2-50-58*
For citation: *Romanova A.S. Game theory for autonomous AI systems in corporate governance. Dependability 2025;2:50-58. https://doi.org/10.21683/1729-2646-2025-25-2-50-58*

Поступила: *11.11.2024* / **После доработки:** *16.12.2024* / **К печати:** *09.06.2025*
Received on: *11.11.2024* / **Revised on:** *16.12.2024* / **For printing:** *09.06.2025*

Введение

Современные технологии искусственного интеллекта (далее – ИИ) в значительной степени основываются на алгоритмах машинного обучения, однако алгоритмы машинного обучения нельзя назвать стратегией. Стратегией называется полное описание того, как система будет себя вести при всех возможных обстоятельствах [1]. Теория игр предоставляет математический аппарат, который позволит автономным системам принимать эффективные решения при управлении корпорациями. Несмотря на то, что многие задачи теории игр можно решать с помощью алгоритмов машинного обучения (впрочем, как и задачи машинного обучения можно формализовать с помощью теории игр), базовые алгоритмы машинного обучения являются скорее сенсорами, способными сказать нам, что именно мы видим в определенном наборе данных, но не могут решить за нас, что нам с этим знанием делать. Теория игр, нашедшая свое применение в экономике, политологии, чистой математике, психологии, социологии, маркетинге и финансах [1], предоставляет возможность системе ИИ работать автономно и принимать управленческие решения на основе моделирования эффективной стратегии. В силу того, что автономные системы искусственного интеллекта являются техническими системами, теория игр, основанная на математических методах анализа принятия управленческих решений, является понятным инструментом для автономных систем ИИ при управлении корпорациями. Предлагаемая модель разработки и внедрения автономных систем ИИ для управления корпорациями (рис. 1) основана на синтезе вычислительного права, выделенного операционного контекста, контролируемой генерации синтетических данных, алгоритмов машинного обучения, определения оптимальной стратегии с использованием теории игр, и технологий объяснимого ИИ.

Концепция закона как вычисления (называемая «вычислительным правом») направлена на то, чтобы свести закон к набору алгоритмов, которые могут автоматически выполняться на компьютере, преобразуя необработанные входные данные в юридические выводы [2]. С целью создания специального операционного контекста для автономных систем ИИ формулировки локальных нормативных документов могут быть одновременно представлены в двух вариантах: для использования людьми и для использования автономными системами [3]. Таким образом, физические лица действуют в рамках кодексов, политик, и процедур, предназначенных для людей, а автономные системы ИИ действуют в рамках кодексов, политик, и процедур, предназначенных для систем ИИ. Обучение систем ИИ на синтетических данных позволяют разрабатывать автономные системы, которые могут действовать более легитимно и этично, чем обычные топ-менеджеры [4]. На основании теории игр компания может определить оптимальные (с точки зрения компании) стратегии для системы ИИ и контролировать их выполнение. Для взаимодействия с заинтересованными сторонами автономная система ИИ должна иметь удобный интерфейс, который позволит осуществлять эффективную коммуникацию, представлять аргументацию и данные, используем системой.

1. Основные типы автономных систем ИИ для управления корпорациями

Основные типы современных автономных систем ИИ для корпоративного управления представлены рис. 2. Цифровые командные центры в настоящее время внедряются многими транснациональными корпорациями, но техническая информация о них в основном является коммерческой тайной. Персонализированные виртуальные системы и гуманоидные роботы являются двумя основными типами интерфейсов, которые позволяют автономным системам ИИ социально взаимодействовать с людьми. С одной стороны, внедрение автономных персонализированных систем управления может существенно изменить концепцию корпоративного лидерства. Цифровые командные центры в настоящее время не имеют возможности социально и эмоционально взаимодействовать с людьми. С другой стороны, ожидаемым шагом в развитии автономных

Рис. 1. Модель разработки автономных систем ИИ для управления компаниями

Рис. 2. Основные типы автономных систем ИИ для управления корпорациями

систем ИИ для корпоративного управления является появление гибридных систем путем объединения многофункциональных цифровых центров и интерфейса в виде персонализированных виртуальных систем и/или гуманоидных роботов.

Отдельные, персонифицированные системы не смогут обрабатывать тот же объем информации, что и многофункциональные цифровые фабрики, и будут периодически или постоянно подключаться к более мощным цифровым фабрикам (рис. 3). Современные цифровые командные центры используя интернет вещей и технологии цифровых двойников собирают данные с производственных площадок, заводов и фабрик. Затем данные записываются (например, используя технологию блокчейн) и обрабатываются алгоритмами машинного обучения, а результаты представляются либо в бумажном виде, либо в электронном. Гибридные же системы смогут передавать полученные результаты также виртуальным агентам или гуманоидным роботам для дальнейшей коммуникации с людьми.

Ниже мы покажем, почему гибридные системы скорее всего станут доминирующей формой автономных систем для управления корпорациями.

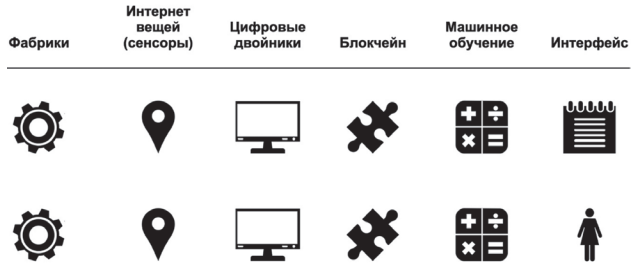

Рис. 3. Возникновение гибридных систем на базе цифровых фабрик

2. Преимущество стратегии над отдельными алгоритмами

Как отмечено в отчете Европейской Комиссии по Этике подключенных и автоматизированных транспортных средств, одного только технологического прогресса недостаточно для эффективного внедрения автономных систем. Будущее развитие автономных систем ИИ должно включать широкий набор этических, правовых и социальных положений, учитываемых при разработке, развертывании и использовании автономных систем [12]. Эксперты Европейской Комиссии по этике беспилотных автомобилей предлагают считать

поведение автономных систем этичным, если оно органически вытекает из непрерывного статистического распределения риска в целях повышения безопасности дорожного движения и равенства между категориями участников дорожного движения [12]. Исследователи в области безопасности беспилотных автомобилей также предлагают различные алгоритмы этичного планирования траектории со структурой, направленной на справедливое распределение рисков среди участников дорожного движения [13].

Однако просто мониторинга уровня риска недостаточно. Существует значительная разница между просто управлением рисками и стратегией, направленной на достижение цели. Автономным системам ИИ нужна эффективная стратегия, благодаря которой они смогут достигать заявленных целей и необходимого уровня безопасности при всех возможных обстоятельствах. Рассмотрим, какие особенности автономных систем гражданского и коммерческого назначения необходимо учитывать при моделировании стратегий для автономных систем ИИ при управлении корпорациями.

3. Основной вопрос теории игр для автономных систем ИИ

В своем труде «Теория игр и экономическое поведение» Джон фон Нейман и Оскар Моргенштерн сформулировали фундаментальные вопросы экономической теории, которые они предлагали решать с помощью теории игр. В частности, описание попыток индивидуума к извлечению максимальной пользы или в случае предпринимателя – к получению максимальной прибыли [14]. Очевидно, что мы не планируем проектировать автономные системы, которые будут стремиться к извлечению максимальной пользы для себя. Нас интересуют автономные системы ИИ, которые будут стремиться к извлечению максимальной пользы для своих создателей. Это основополагающее ограничение ничуть не мешает нам использовать математические модели исследуемые в теории игр.

При формализации ситуации для целей анализа с использованием теории игр рассматриваются, как минимум, несколько базовых вопросов, которые позволяют классифицировать игры тем или иным способом [1]. Однако основным вопросом для автономных систем ИИ гражданского и коммерческого назначения является вопрос: в какую именно игру играет система в данный момент? От корректного определения игры зависит определение оптимальной стратегии.

Во многих источниках по этике автономных систем ИИ игры смешиваются, что затрудняет анализ ситуации и поиск эффективной стратегии. Например, рассматривается вопрос: должен ли беспилотный автомобиль более бережно относиться к мотоциклистам, которые не носят шлем? Таким образом, может показаться, что мотоциклисты, которые носят шлемы, по сути, подвергаются наказанию и дискриминации за свое ответственное решение надеть шлем [15]. С точки зрения теории игр, беспилотные автомобили без специального обучения не участвуют в игре по изучению мотоциклистами правил дорожного движения, т.е. для автономного автомобиля такой вопрос даже не возникнет, в лучшем случае он просто передаст сведения о нарушителе дорожной полиции. Описанная выше ситуация является одним из примеров, когда гипотетические дилеммы ИИ с точки зрения теории игр или не возникают, или имеют совершенно другой смысл.

4. Участники игр

В процессе своей деятельности автономная система ИИ может играть с социальными системами (человеком или человеческим коллективом), с техническими системами (другой системой ИИ), со смешанными системами (например, человек при поддержке системы ИИ) и с природой. Очевидно, что без специального обучения для системы ИИ и живые, и неживые организмы изначально равнозначны. Существенным для нас является вопрос, должны ли системы ИИ одинаково бережно относиться и к людям, и к другим автономным системам? Аргументом «за» может служить то, что любая техническая система – это своеобразное «прокси» какой-то социальной системы (т. е. человека или человеческого коллектива).

5. Классификация игр

Профессор математики Мортон Дэвис подразделяет конечные игры для двух игроков на три категории: игры с нулевой суммой и полной информацией, общие игры с нулевой суммой, игры с ненулевой суммой [1]. Можно отметить, что игры с нулевой суммой не должны являться приоритетными при проектировании автономных систем ИИ. В отчете Европейской Комиссии по этике беспилотных автомобилей указывается, что автономные автомобили должны быть разработаны и эксплуатироваться таким образом, чтобы вносить позитивный вклад в благосостояние людей, включая будущие поколения и других живых существ [12]. Как указывает Дэвис, игра является игрой с нулевой суммой, если она удовлетворяет определенному закону сохранения: игра является игрой с нулевой суммой, если в ходе игры богатство не создается и не уничтожается [1]. Но зачем нам проектировать автономные системы ИИ, которые не приумножают благосостояние человечества? Таким образом, автономные системы в общем случае в первую очередь должны рассчитывать игры с ненулевой суммой.

Учитывая доступность больших данных для автономных систем ИИ понятие полной и неполной информации меняет свой смысл. С одной стороны, информация уже никогда не будет полной, потому что всегда найдется очередной набор данных. С другой стороны, при использовании больших данных точность прогноза значительно повышается.

6. Понятие полезности для автономных систем ИИ

Поскольку решения, принимаемые автономными системами ИИ, напрямую могут влиять на жизнь отдельного человека, человеческого коллектива или общества, то для моделирования таких игр необходим измеритель полезности. Несмотря на многочисленные заявления о том, что ценность человеческой жизни измерить нельзя, на практике такие измерения уже применяются, например, в медицине. При распределении дефицитных медицинских лекарств, а также донорских органов современное общество уже использует разнообразные подходы, которые считаются морально приемлемыми: равное отношение к людям, предпочтение наиболее больным, максимизация общей выгоды, поощрение и вознаграждение социальной полезности [16]. Для повышения справедливости также используются комбинации возможных подходов: система баллов Объединенной сети по обмену органами, годы жизни с поправкой на качество и годы жизни с поправкой на инвалидность, система полных жизней (которая отдает приоритет молодым людям), принципы прогнозирования, принципы спасения наибольшего количества жизней, лотереи и инструментальные ценности [16]. Приведенные выше примеры позволяют численно измерять полезность для каждого отдельно взятого человека.

С точки зрения общества также существует система мониторинга экологического, социального и корпоративного управления (ESG). Принципы ESG называют также «моральными деньгами» (“moral money”) [17]. Исследователи из MIT и Цюрихского университета насчитывают более 700 индикаторов, которые ведущие рейтинговые агентства используют для составления ESG рейтингов [18]. Таким образом, и с точки зрения отдельно взятого человека, и с точки зрения компании и общества, уже существуют общепризнанные модели измерителей, которые можно использовать для определения полезности при расчете стратегий для автономных систем ИИ при управлении корпорациями.

7. Сигнализация

Автономные системы ИИ имеют намного больше возможностей установить рациональность другой системы. Более того, они имеют возможность установить даже уровень рациональности другой системы. Даже если в сигналах одной системы ИИ присутствует шум и она не передает полностью прозрачные сигналы, другая система ИИ имеет возможность вычислить уровень правдоподобности.

Нас же интересует вопрос прозрачности намерений автономной системы ИИ при игре с человеком: должны ли сигналы системы ИИ быть полностью прозрачными или система ИИ имеет право хитрить? И должно ли это быть законодательно закреплено: по требованию человека автономная система ИИ должна отвечать только правду и ничего кроме правды?

8. Этичность и легитимность как основные правила теории игр для автономных систем ИИ

Исходя из предпосылки, что автономные системы ИИ для коммерческих и гражданских целей должны быть спроектированы с учетом соблюдения правовых и этических норм, этичность и легитимность должны являться стандартным ограничивающим правилом при анализе стратегий автономными системами ИИ. Обязательное требование о разработке этичных и легитимных автономных систем ИИ может привести к тому, что некоторые дилеммы самой теории игр для автономных систем ИИ или не существуют, или изменяют свой смысл. Например, знаменитая дилемма двух заключенных для этичных и легитимных систем ИИ будет означать, что если нанесен непреднамеренный ущерб (так как автономные системы гражданского и коммерческого назначения не могут проектироваться для нанесения преднамеренного ущерба), то его нужно как можно быстрее исправить, а не замалчивать. В то время как математическая модель ситуации может оставаться той же самой, содержательное описание должно измениться, чтобы отражать возможности и ограничения автономных систем ИИ.

В общем случае требования легитимности и этичности будут затрагивать практически все базовые вопросы теории игр:

1) ограничивать или увеличивать количество игр, которые может или должна рассматривать система ИИ;

2) ограничивать или увеличивать количество стратегий;

3) ограничивать или увеличивать количество игроков и т.д.

Обязательное требование легитимности и этичности изначально возможно реализовать через ранжирование игр: обязательные требования должны рассчитываться в играх более высокого уровня (приоритета). В некоторой степени можно сказать, что знаменитые три закона робототехники представляют собой зачатки вычислительного права для автономных систем ИИ. Например, нулевой закон робототехники: «Робот не может нанести вред человечеству или своим бездействием допустить, чтобы человечеству был нанесен вред» [19] – отражает проблему соотношения частных и публичных интересов. Таким образом, если при анализе игр для социальных систем (человека и человеческих коллективов) можно рассматривать возможность нелегитимного и неэтичного поведения, то стратегии для автономных систем ИИ сразу должны исключать нелегитимность и неэтичность.

9. Коалиционные и бескоалиционные игры для автономных систем ИИ

Требования этичности и легитимности для автономных систем ИИ в значительной степени предопределяют, в каких случаях игры могут или должны быть коалици-

онными, а в каких – нет. В случае игры по спасению жизни человека для автономной системы ИИ целесообразно иметь возможность вступать в коалиции с другими системами. В тех случаях, когда коалиции запрещены (например, законодательством о свободе конкуренции) требования легитимности ограничат вступление автономных систем ИИ в такую игру. Учитывая способность автономных систем вычислять и скрывать паттерны, обычный человек не сможет определить, вступили ли системы ИИ в коалицию.

Поскольку автономные системы могут существенно различаться по вычислительной мощности, их понимание рациональности и эффективность также могут быть разными. В отчете Европейской Комиссии по этике беспилотных автомобилей указывается, что в соответствии с принципом справедливости автономные автомобили могут быть обязаны вести себя по-другому в отношении некоторых категорий участников дорожного движения, например, пешеходов или велосипедистов, чтобы предоставить им тот же уровень защиты, что и другим участникам дорожного движения [12]. В частности, автономные автомобили должны, помимо прочего, адаптировать свое поведение в отношении уязвимых участников дорожного движения вместо того, чтобы ожидать, что эти пользователи приспособятся к опасностям дороги [12]. Таким образом, с точки зрения этичности и легитимности более мощная автономная система ИИ должна иметь возможность рассчитывать и сообщать стратегию для всех участников игры (если только такие действия прямо не запрещены законодательно).

10. Базовая игра для автономных систем ИИ

При создании своей теории Джон фон Нейман и Оскар Моргенштерн предположили, что целью всех участников экономической системы являются деньги или некоторый единый монетарный товар [14]. Мы же приходим к тому, что целью проектируемых автономных систем ИИ должна быть человеческая жизнь, ее сохранение и защита. Тем не менее, чтобы использовать методологию теории игр необходимо ввести универсальный измеритель полезности для автономных систем ИИ (например, с использованием примеров определения ценности, описанных выше).

Если мы примем, что основной целью для любой гражданской или коммерческой автономной системы ИИ прежде всего является сохранение жизни человека (или человечества), то система будет действовать «рациональным образом», если она будет нацелена на то, чтобы получить соответствующие максимумы по сохранению жизни. Таким образом, у любой автономной системы ИИ всегда будет базовая игра, направленная на сохранение жизни идентифицированных такой системой людей. Игры для автономных систем ИИ возможно разделить на несколько уровней в зависимости от приоритета выполняемой задачи. При этом базовая игра должна иметь высший приоритет, и либо может быть бесконечной, либо может быть разделена на несколько партий или подигр, сменяющих друг друга. Также базовая игра не может формулироваться как игра с нулевой суммой, так как изначально мы не хотим проектировать систему, которая будет проигрывать человеческие жизни.

В некоторых играх похожее правило просто включено в состав общей игры: например, в шахматах правила игры запрещают ставить короля под шах [14]. Разница заключается в том, что в шахматах цель игры более общая, чем просто избежать положения шаха для своего короля. Если же мы говорим, что цель игры – именно сохранение жизни человека, значит это отдельная стратегия и отдельная игра.

Правила базовой игры могут формулироваться под влиянием социальных норм общества, в котором предстоит работать автономной системе ИИ. Как показал эксперимент MIT «Моральная машина» – моральная приемлемость жертв и ценность человеческой жизни различаются в зависимости от принятых в стране социальных и культурных норм [20].

11. Дизайн базовой игры для мониторинга защиты одного человека

Для упрощения и прозрачности модели базовой игры будем рассматривать игру с природой, где природа представляет собой игрока, использующего случайную стратегию. Возможным примером может служить вопрос для руководства компании о закупке дорогостоящего защитного оборудования, которое может никогда не пригодиться.

Игра определяется следующей матрицей:

- (1) – человек защищен полностью;
- (-1) – человек не защищен (полностью или частично).

Данную матрицу можно представить в виде стратегий (табл. 1):

- $S_1 = (1, 1, 1)$;
- $S_2 = (–1, 1, 1)$;
- $S_3 = (–1, –1, 1)$;

Табл. 1. Платежная матрица игры для защиты одного человека

Класс защитного оборудования	Регулярные стихийные бедствия	Средний уровень погодной опасности	Хорошая погода
Высокий	1	1	1
Средний	-1	1	1
Низкий	-1	-1	1

Наиболее очевидным критерием с точки зрения защищаемого человека является критерий Вальда. По критерию Вальда оптимальной является стратегия,

Табл. 2. Расчет стратегии по критерию Вальда

Класс защитного оборудования	Регулярные стихийные бедствия	Средний уровень погодной опасности	Хорошая погода	Худший результат	Лучший худший результат
Высокий	1	1	1	1	1
Средний	-1	1	1	-1	-
Низкий	-1	-1	1	-1	-

которая при наихудших действиях природы гарантирует максимальный выигрыш [21] (табл. 2). Более того, может существовать общество, которое может принять критерий Вальда в качестве обязательного для автономных систем ИИ.

Рассмотрим несколько других, часто применяемых, но менее приемлемых с точки зрения защищаемого человека критериев принятия решения в условиях неопределенности. Например, критерий Гурвица пытается найти золотую середину между крайностями, заданными оптимистическим и пессимистическим критериями [22]. Однако человека, которого автономная система ИИ должна защищать, нацеленность автономной системы на оптимистичные прогнозы скорее всего не устроит (если только он сам не исключительный оптимист). Критерий минимаксного сожаления Сэвиджа рассматривает сожаление, альтернативные издержки или потери, возникающие в случае, когда возникает конкретная ситуация, а выигрыш от выбранной альтернативы меньше выигрыша, который мог бы быть получен в этой конкретной ситуации [22]. Сожаление также не подходит в качестве эффективного критерия – автономная система должна добиваться выигрыша (сохранения жизни человека, непричинения ему вреда), а не фокусироваться на возможном проигрыше.

12. Дизайн базовой игры для мониторинга защиты социального (человеческого) коллектива

Приведенный выше дизайн базовой игры не является оптимальным с точки зрения человеческого коллектива, поскольку не учитывает множество факторов:

• отвлечение ресурсов компании на приобретение ненужного дорогостоящего оборудования означает, что ресурсов может не хватить на приобретение действительно жизненно необходимых вещей;

• производство дорогостоящего оборудования может наносить непоправимый вред окружающей среде и т.д.

Для всесторонней оценки принимаемого решения и расчета полезности с точки зрения общества для автономных систем ИИ возможно адаптировать или уже существующие системы ESG метрик, или разработать собственную систему метрик на основе уже существующих международных норм. Примером для создания системы метрик могут использоваться Цели в области устойчивого развития (далее – ЦУР), разработанные и принятые всеми государствами-членами Организации Объединенных Наций в 2015 году, как общий план мира и процветания для людей и планеты сейчас и в будущем [23]. В основе лежат 17 целей и 169 показателей в области устойчивого развития, которые включают: ликвидацию нищеты и голода, обеспечение качественного образования, рациональное использование водных ресурсов, принятие срочных мер по борьбе с изменением климата, рациональное управление лесами и т. д. [23].

При оценке легитимности и этичности принимаемого решения автономная система ИИ может использовать указанные выше показатели для расчета показателей полезности при различных стратегиях. Для полноценного расчета данной модели требуется значительное количество данных и вычислительных мощностей. Поэтому многофункциональные цифровые фабрики или гибридные системы смогут решить эту задачу значительно лучше, чем менее мощные отдельные, персонифицированные системы.

Заключение

Теория игр предоставляет широкие возможности для моделирования решений, принимаемых автономными системами ИИ при управлении корпорациями. Учитывая чрезвычайно большой объем данных, необходимых для принятия качественных управленческих решений, на современном уровне развития технологий только мощные цифровые фабрики смогут рассчитывать обоснованные и качественные решения.

Отдельные, персонифицированные автономные системы ИИ самостоятельно пока что не могут проводить необходимый объем вычислений. Поэтому такие системы могут применяться либо как интерфейсы многофункциональных цифровых командных центров (цифровых фабрик), либо будут являться просто «говорящими головами» без объективной возможности принимать ответственные управленческие решения.

Библиографический список

1. Davis M.D. Game Theory: A Nontechnical Introduction. Dover Publications, 1997. 208 p.

2. Романова А.С. Основы моделирования алгоритмических решений при управлении корпорациями // Искусственные общества. 2024. Т. 19. Вып. 3. URL: https://artsoc.jes.su/s207751800032184-1-1/ (дата обращения 30.03.2025). DOI: 10.18254/S207751800032184-1

3. Романова А.С. Начала законодательства для автономных систем искусственного интеллекта // Надежность. 2024. № 24(3) С. 10-17. DOI: 10.21683/1729-2646-2024-24-3-10-17

4. Романова А.С. Моделирование автономных систем управления корпорациями на основе синтетических данных // В сб.: XXXVI Международная научно-техническая конференция «НЕЙРОИНФОРМАТИКА-2024». М.: МФТИ, Физтех, 2024. С. 193-203.

5. ADNOC, ADNOC Wins Industry Technology Award for its Panorama Digital Command Center // URL: https://adnoc.ae/news-and-media/press-releases/2021/adnoc-wins-industry-technology-award-for-its-panorama-digital-command-center (дата обращения 30.03.2025).

6. Businesswire. Tieto the First Nordic Company to Appoint Artificial Intelligence to the Leadership Team of the New Data-Driven Businesses Unit // URL: https://www.businesswire.com/news/home/20161016005092/en/Tieto-the-First-Nordic-Company-to-Appoint-Artificial-Intelligence-to-the-Leadership-Team-of-the-New-Data-Driven-Businesses-Unit (дата обращения 30.03.2025).

7. Deep Knowledge Ventures // URL: https://deepknowledgeventures.com/ (дата обращения 30.03.2025).

8. IHC. IHC's Aiden Insight sets a new benchmark for the integration of artificial intelligence in high-level corporate strategy // URL: https://www.ihcuae.com/photo/plugin/article/2024/1715086662_file_1.pdf (дата обращения 30.03.2025).

9. NetDragon. NetDragon's AI Leader Tang Yu Named China's Best Virtual Employee of 2024, 2024 // URL: https://www.netdragon.com/content/2024-04-28/20240428231345555.shtml (дата обращения 30.03.2025).

10. Business Standard. Mika becomes world's first robot CEO, thinks she's better than Musk // URL: https://www.business-standard.com/world-news/mika-becomes-world-s-first-robot-ceo-thinks-she-s-better-than-musk-123110901563_1.htm (дата обращения 30.03.2025).

11. Hanson Robotics // URL: https://www.hansonrobotics.com/sophia/ (дата обращения 30.03.2025).

12. Ethics of Connected and Automated Vehicles. Recommendations on road safety, privacy, fairness, explainability and responsibility. Luxembourg: Publications Office of the European Union, 2020. URL: https://data.europa.eu/doi/10.2777/93984 (дата обращения 30.03.2025).

13. Geisslinger M., Poszler F., Lienkamp M. An ethical trajectory planning algorithm for autonomous vehicles // Nature Machine Intelligence. 2023. Vol. 5. Pp. 137-144.

14. Neumann J.V., Morgenstern O. The Theory of Games and Economic Behaviour. Princeton University Press, 1944.

15. Lin P. Why Ethics Matters for Autonomous Cars //In: Autonomes Fahren. Technische, rechtliche und gesellschaftliche Aspekte / Eds: Markus Maurer, J. Christian Gerdes, Barbara Lenz, Hermann Winner. Springer Vieweg Berlin, Heidelberg, 2015. Pp. 69-75.

16. Persad G., Wertheimer A., Emanuel E.J. Principles for allocation of scarce medical interventions // The Lancet. 2009. Vol. 373 (9661). Pp. 423-431.

17. Financial Times. A new revolution in finance: introducing Moral Money, the latest FT newsletter. URL: https://professional.ft.com/en-gb/blog/new-revolution-finance-introducing-moral-money-latest-ft-newsletter/ (дата обращения 30.03.2025).

18. Berg, F., Kölbel, J.F., & Rigobón, R. Aggregate Confusion: The Divergence of ESG Ratings // Review of Finance. 2022. Vol. 26. Issue 6. Pp. 1315–1344. DOI: 10.1093/rof/rfac033

19. Asimov I. Foundation and Earth. Harper Voyager, 2016.

20. Awad E., Dsouza S., Shariff A.F. et al. Universals and variations in moral decisions made in 42 countries by 70,000 participants // In: Proceedings of the National Academy of Sciences of the United States of America. 2020. Vol. 117. Pp. 2332-2337. DOI: 10.1073/pnas.1911517117

21. Sniedovich M. Wald's maximin model: a treasure in disguise! // The Journal of Risk Finance. 2020. Vol. 9. No. 3. Pp. 287-291. DOI: 10.1108/15265940810875603

22. Thapaswini P.S. Savage Minimax Regret Criterion Method to Find Optimal Decision // International Journal of Science and Research (IJSR). 2020. Vol. 9. Issue 2. DOI: 10.21275/ART20204447

23. United Nations. Sustainable development goals // URL: https://sdgs.un.org/goals (дата обращения 30.03.2025).

References

1. Davis M.D. Game Theory: A Nontechnical Introduction. Dover Publications; 1997.

2. Romanova A. Fundamentals of Modeling Algorithmic Decisions in Corporate Management. *Artificial societies* 2024;19(3). Available at: https://artsoc.jes.su/s207751800032184-1-1/. DOI: 10.18254/S207751800032184-1. (in Russ.)

3. Romanova A.S. Fundamentals of legislation in autonomous artificial intelligence systems. *Dependability* 2024;24(3):10-17. (in Russ.) DOI: 10.21683/1729-2646-2024-24-3-10-17.

4. Romanova A.S. [Simulating autonomous corporate management systems based on synthetic data]. In: Proceedings of the XXXVI International Science and Technology Conference NEUROINFORMATICS-2024. Moscow: MIPT, Phystech; 2024. Pp. 193-203. (in Russ.)

5. ADNOC Wins Industry Technology Award for its Panorama Digital Command Center. (accessed 30.03.2025). Available at: https://adnoc.ae/news-and-media/press-releases/2021/adnoc-wins-industry-technology-award-for-its-panorama-digital-command-center.

6. Tieto the First Nordic Company to Appoint Artificial Intelligence to the Leadership Team of the New Data-Driven Businesses Unit. (accessed 30.03.2025). Available at: https://www.businesswire.com/news/home/20161016005092/en/Tieto-the-First-Nordic-Company-to-Appoint-Artificial-Intelligence-to-the-Leadership-Team-of-the-New-Data-Driven-Businesses-Unit.

7. Deep Knowledge Ventures. (accessed 30.03.2025). Available at: https://deepknowledgeventures.com/.

8. IHC. IHC's Aiden Insight sets a new benchmark for the integration of artificial intelligence in high-level corporate strategy. (accessed 30.03.2025). Available at: https://www.ihcuae.com/photo/plugin/article/2024/1715086662_file_1.pdf.

9. NetDragon. NetDragon's AI Leader Tang Yu Named China's Best Virtual Employee of 2024. (accessed 30.03.2025). Available at: https://www.netdragon.com/content/2024-04-28/20240428231345555.shtml.

10. Business Standard. Mika becomes world's first robot CEO, thinks she's better than Musk. (accessed 30.03.2025). (accessed 15.08.2019). Available at: https://www.business-standard.com/world-news/mika-becomes-world-s-first-robot-ceo-thinks-she-s-better-than-musk-123110901563_1.htm.

11. Hanson Robotics. (accessed 30.03.2025). Available at: https://www.hansonrobotics.com/sophia.

12. Ethics of Connected and Automated Vehicles. Recommendations on road safety, privacy, fairness, explainability and responsibility. Luxembourg: Publications Office of the European Union; 2020. (accessed 24.06.2024). Available at:// doi: 10.21683/ 2777/ 93984.

13. Geisslinger M., Poszler F., Lienkamp M. An ethical trajectory planning algorithm for autonomous vehicles. *Nature Machine Intelligence* 2023;5:137- 144.

14. Neumann J.V., Morgenstern O. The Theory of Games and Economic Behaviour. Princeton University Press; 1944.

15. Lin P. Why Ethics Matters for Autonomous Cars. In: Maurer M., Gerdes J.C., Lenz B., Winner H., editors. Autonomes Fahren. Technische, rechtliche und gesellschaftliche Aspekte. Springer Vieweg Berlin: Heidelberg; 2015. Pp. 69- 75.

16. Persad G., Wertheimer A., Emanuel E.J. Principles for allocation of scarce medical interventions. *The Lancet* 2009;373(9661):423-431.

17. A new revolution in finance: introducing Moral Money, the latest FT newsletter. (accessed 30.03.2025). Available at: https://professional.ft.com/en-gb/blog/new-revolution-finance-introducing-moral-money-latest-ft-newsletter.

18. Berg F., Kölbel J.F., Rigobón R. Aggregate Confusion: The Divergence of ESG Ratings. *Review of Finance* 2022;26(6):1315-1344. DOI: 10.1093/rof/rfac033.

19. Asimov I. Foundation and Earth. Harper Voyager; 2016.

20. Awad E., Dsouza S., Shariff A.F. et al. Universals and variations in moral decisions made in 42 countries by 70,000 participants. In: Proceedings of the National Academy of Sciences of the United States of America 2020;117:2332-2337. DOI: 10.1073/pnas.1911517117.

21. Sniedovich M. Wald's maximin model: a treasure in disguise! *The Journal of Risk Finance* 2020;9(3):287-291. DOI: 10.1108/15265940810875603.

22. Thapaswini P.S. Savage Minimax Regret Criterion Method to Find Optimal Decision. *International Journal of Science and Research (IJSR)* 2020;9(2). DOI: 10.21275/ART20204447.

23. United Nations. Sustainable development goals. (accessed 30.03.2025). Available at: https://sdgs.un.org/goals.

Сведения об авторе

Романова Анна Сергеевна – эксперт в области цифровой трансформации, к.э.н., MBA, LL.M, ALM, FCCA, свыше 15 лет опыта работы в крупнейших российских и международных компаниях. Аспирант МФТИ по направлению «Искусственный интеллект и машинное обучение», Москва, Российская Федерация, e-mail: romanova.as@phystech.edu.

About the author

Anna S. Romanova, an expert in digital transformation, Candidate of Economics, MBA, LL.M, ALM, FCCA, over 15 years of experience in the largest Russian and international companies. Postgraduate student, Artificial Intelligence and Machine Learning, MIPT, Moscow, Russian Federation, e-mail: romanova.as@phystech.edu.

Вклад автора в статью

Романовой А. С. разработана таксономия автономных систем искусственного интеллекта для управления корпорациями. Также предложены базовые принципы, позволяющие автономным системам искусственного интеллекта рассчитывать этичную и легитимную стратегию при принятии ответственных управленческих решений.

Конфликт интересов

Автор заявляет об отсутствии конфликта интересов.