

УДК 004.8**Искусственный интеллект и связанные с ним риски****Татаринов Константин Анатольевич**

Кандидат экономических наук, доцент,
Байкальский государственный университет,
664003, Российская Федерация, г. Иркутск, ул. Ленина, 11;
e-mail: tatarinov723@gmail.com

Аникиенко Николай Николаевич

Кандидат экономических наук, доцент,
Иркутский государственный аграрный университет,
664038, Российская Федерация, Иркутский район, п. Молодёжный, 1;
e-mail: anikienkonikolai@mail.ru

Савченко Инна Анатольевна

Кандидат экономических наук, доцент,
Иркутский государственный аграрный университет,
664038, Российская Федерация, Иркутский район, п. Молодёжный, 1;
e-mail: innasava2016@mail.ru

Музыка Сергей Михайлович

Кандидат биологических наук, доцент,
Иркутский государственный аграрный университет,
664038, Российская Федерация, Иркутский район, п. Молодёжный, 1;
e-mail: ignitmuz@mail.ru

Аннотация

Вычислительные машины имеют ряд преимуществ по сравнению с биологическим мозгом людей, который создавался эволюцией на протяжении миллиардов лет. Кремниевые микропроцессоры «срабатывают» и передают сигналы в миллионы раз быстрее нейронов, а суперкомпьютеры могут иметь размеры заводского цеха. Программное обеспечение можно легко редактировать и приумножать, чтобы использовать алгоритм несколькими способами одновременно. Большие базы данных позволяют получать актуальную информацию в любой момент времени. Однако, в вопросах изящной деградации, механической устойчивости и энергоэффективности человеческий мозг превосходит виртуальный. В ближайшие годы квантовые компьютеры нарушат имеющуюся зависимость между термодинамической эффективностью и сложностью на уровне обработки информации, и возникает вероятность того, что человеческий интеллект когда-то не будет превалировать над машинным. Поэтому для людей важно точно спрогнозировать и оценить, как и когда это случится, и каковы

последствия такого неблагоприятного сценария развития. В статье основное внимание уделяется не только вопросам неблагоприятных сценариев развития в традиционных сферах человеческой деятельности (производство, образование, медиа, потребление и государственный сектор), но «новым» рискам будущего суперинтеллекта. Акцент сделан на то, что преодолеть возможные угрозы человечеству со стороны машины можно только «привив» ей основополагающие этические ценности.

Для цитирования в научных исследованиях

Татаринов К.А., Аникиенко Н.Н., Савченко И.А., Музыка С.М. Искусственный интеллект и связанные с ним риски // Контекст и рефлексия: философия о мире и человеке. 2024. Том 13. № 9А. С. 75-82.

Ключевые слова

Этические ценности, манипулирование людьми, безработица, цифровая революция, суперкомпьютер, человеческий интеллект, интеллектуальная собственность, дискриминация.

Введение

Люди с древних времен мечтали об умных машинах, которые избавят их от опасной и утомительной работы, с которыми можно поговорить обо всем и которые будут им беспрекословно служить. В 1950-х годах искусственный интеллект приблизил человечество к осуществлению этой мечты. В третьем десятилетии XX века в любой области человеческой деятельности алгоритмы анализируют, классифицируют, коррелируют и накапливают данные, а также делают прогнозы относительно будущего. Поиск в Интернете, общение с чат-ботами, перевод напечатанного текста с помощью камеры смартфона, звонок в колл-центр, прогноз погоды, спортивные репортажи и новостные сюжеты, логистика складов, медицинские исследования и другие виды деятельности уже давно основаны на методах искусственного интеллекта. Многие даже говорят о невидимой цифровой революции, начало которой человечество только что пережило. Другие видят в искусственном интеллекте столь же большие опасения [Маркова, Пронин, 2022]. Новые технологии стали контролировать и манипулировать людьми, существующие IT-монополии еще больше упрочили свое влияние, увеличилась безработица из-за слияния искусственного интеллекта с роботизацией производства, прекратил действовать принцип солидарности в страховании и углубился социальный раскол в человеческом обществе [Романова, 2021]. Если смотреть на новую технологию сквозь очки научной фантастики, где человекоподобные автоматы развиваются и обращаются против своего создателя или, наоборот, люди влюбляются в робота, купленного в качестве домашнего помощника, то невозможно разглядеть индивидуальные методы лечения, новые модели обмена знаниями, рационально организованное сельское хозяйство и эффективно скоординированное электроснабжение. В последние годы новейшие IT-системы позволили создать мощные инструменты, которые уже давно превосходят человеческие способности в ряде областей. Поэтому важно с реалистичной точки зрения сделать оценку того, насколько можно доверять цифровому разуму, ведь от этого будет зависеть его контроль и достоверность результатов его вычислений.

Цель статьи — всесторонне рассмотреть искусственный интеллект, как базовую технологию цифровизации и риски его применения в сферах человеческой деятельности.

Основная часть

Тема искусственного интеллекта стала массово популярна с ноября 2022 года, когда десятки миллионов человек стали одновременно пользоваться текстовым чат-ботом ChatGPT и нейросетью Midjourney. Фактически же компьютерный разум используется в повседневной жизни уже несколько лет и зачастую делает это незаметно. Например, системы голосового (Alexa, Siri и Google Home) и навигационного управления (Google Maps) уже давно используются в повседневной взрослой жизни. Дети и подростки следуют рекомендациям неросетей в социальных сетях, которые предоставляют индивидуально подобранные результаты и оптимизируют их в зависимости от конкретных условий. На музыкальных порталах целые плейлисты создаются автоматически под вкусы меломана. Создаваемый таким образом «пузырь мнений» преодолевается только активным анонимным поиском по другим темам. Антропоморфизм (перенесение человеческого облика на неодушевлённые предметы) происходит не только в информационной сфере, это уже всеобъемлющее явление. Например, ChatGPT при выдаче результата учитывает этические и моральные ценности и отвечает на вопрос не сразу, а только постепенно. Чат-бот печатает и делает паузы для размышлений, а не публикует свои результаты за раз в блоке ответов [Ван, 2024]. Но ответы основаны на вероятностях, а не на ценностях и нормах общества. С каждой итерацией обновляются и поддерживаются только те записи, которые пользуются спросом. «Краевые участки» в кривой Гауса обрезаются и таким образом выдаются только средние ответы и не генерируются инновационные. Такое смещение приводит к предвзятости чат-бота и способствует проявлению предубеждений и изоляции. В дальнейшем в Интернете станет сгенерированных текстов больше, чем написанных человеком. В эпоху социальных сетей многие пользователи используют встроенного чат-бота «Му AI» не только для написания постов, но и для поддержания или установления реальных дружеских отношений. При таком аномальном использовании цифрового разума настоящие дружеские отношения будут игнорироваться.

В свете цифрового развития человечества термин «искусственный интеллект» стал наиболее распространенным. Однако используется он в совершенно разных контекстах и совершенно неясным становится то, что на самом деле стоит за этим понятием. Проблема в том, что в разные эпохи простые механизмы часто называют интеллектуальными. Например, для человека XIX века калькулятор будет «умным». Иными словами, это вопрос уже имеющегося опыта взаимодействия с машинами и будущих ожиданий от них. Сегодня ведь никто голосового помощника о погоде не назовет «мыслящим». Таким образом под искусственным интеллектом подразумевается то, как машины справляются с задачами, в решении которых люди пока превосходят их. В этой формулировке нет конкретной технологии или машинного обучения. Сегодня особенно хорошо люди справляются с совершенно новыми ситуациями, где требуется интуиция. «Победное шествие» цифрового разума в бизнесе началось в 2016 году, когда AlphaGo победил Ли Седоля в игре в го, а в 2017 году Libratus одержало победу над четырьмя лучшими игроками в покер. В настоящее время развитие цифрового разума идет по экспоненте, так как «большая пятерка» GAFAM вкладывает миллиарды в исследования в этой области. Сегодня людям стоит задать себе три вопроса: во-первых, что будет представлять из себя искусственный интеллект через несколько десятков лет; во-вторых, каким целевым установкам он будет следовать; в-третьих, какое влияние он окажет на людей. Самым непредсказуемым из путей развития машинного интеллекта является «глубокое обучение» [Прокофьев, 2023]. Ведь при этом информация обрабатывается в иерархических уровнях нейронных сетей, которые

имеют структуру аналогичную человеческому мозгу. Большая опасность состоит в том, что ИТ-система сама по себе разрабатывает модели мира на основе гигантских наборов данных [Паньшин, 2023]. Таким же образом поступают и дети до подросткового возраста. Одновременный рост объема данных и вычислительной мощности позволяет достичь «подросткового» восприятия мира в течение нескольких лет, что делает политический контроль над цифровым разумом невозможным. Поэтому вопрос о непредсказуемости будущих возможностей умной машины самый актуальный (рис.). Непредсказуемость в данном контексте означает то, что созданные искусственные модели мира могут сильно отличаться от имеющихся, что приведет к иным стратегиям действий нейросети.

Цифровая трансформация в пяти сферах жизни общества (работа, образование, потребление, СМИ, госуправление) является темой оценки воздействия технологий и искусственный интеллект здесь играет «первую скрипку». Достижения в области алгоритмического интеллекта позволяют решать задачи, с которыми до сих пор не могли справиться традиционные компьютерные программы (генерация изображений и естественного языка, распознавание речи и текста, перевод с древних языков, диагностика заболеваний, управление транспортными средствами и т. д.). Основными проблемными областями, сосредоточенными на самой технологии, являются экономические последствия, конкуренция за большие данные и геополитические соображения из-за того, что США и Китай инвестируют огромные суммы в развитие данной технологии [Масовец, Скоробогатова, 2020]. Этот круг вопросов отражен не только в международных дебатах, но дополняется правовыми вопросами, касающихся законов об интеллектуальной собственности, защите данных и запрете на дискриминацию [Костенков, 2023]. В сфере труда широко обсуждается то, что роботизированный ум сделает труд человека устаревшим. Данный вывод основан на историческом прошлом, когда автоматизация производства также сопровождалась проблемами поляризации заработной платы и сокращением объемов занятости у неквалифицированных работников. Поэтому некоторые виды деятельности, осуществляемые людьми, должны субсидироваться.

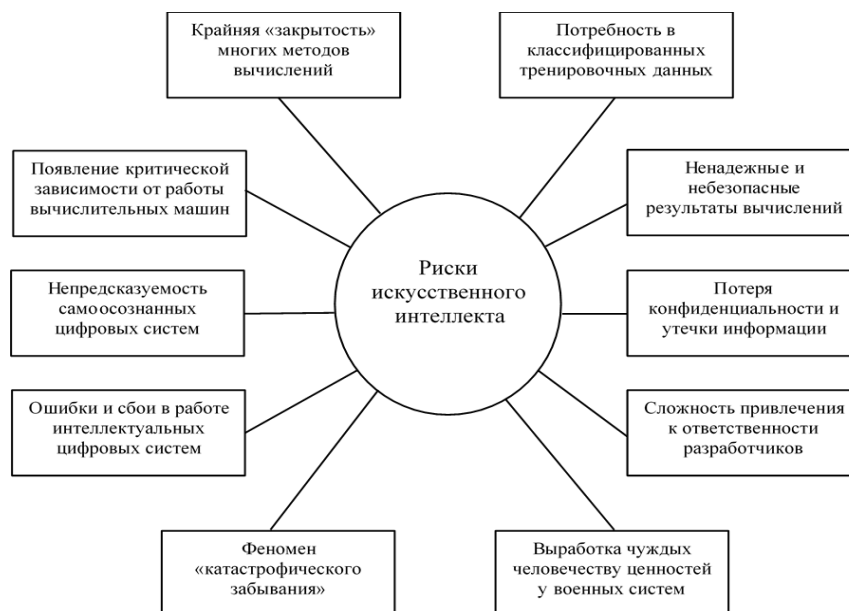


Рисунок 1 - Общекультурные риски искусственного интеллекта

В идеале, распространение среди работников инновационных технологий должно подразумевать безусловный базовый доход и отрицательный налог на прибыль. Кроме того, виртуальный разум оказывает влияние и на дизайн самой работы, отбор сотрудников, мониторинг их работы и планирование самой карьеры. Особую значимость приобретает непрерывное образование как мера, способствующая позитивным цифровым изменениям. В области потребления стали распространены приложения для персонализации, рекомендательные сервисы и цифровые помощники. Также в обществе обсуждается тема «Цифровой разум как черный ящик для покупателя» и вытекающие из этого вопросы доверия и анонимности, а также предотвращения создания цифровых олигополий. В СМИ «AI» подвергается анализу в контексте нестабильного информационного поведения, особенно с учетом того, что влияние прессы и телевидения за последние годы сильно уменьшилось. В центре внимания стали находиться эхо-камеры социальных сетей, которые приводят к унификации потребления информации целыми группами людей. Риск фейковых новостей возрастает в разы, из-за того, что новейшие технологии обладают возможностями подделывать аудио- и видеоконтент неотличимо от оригинала [Гуреева, Киреева, 2022]. В государственной сфере в центре анализа находятся вопросы к действиям государства на основе искусственного интеллекта (упреждающая полицейская деятельность, дискриминация, вынесение судебных решений и т. д.) [Михайлов, Кокодей, 2023]. К факторам риска можно отнести непрозрачность процедур принятия судьбоносных решений, угроза презумпции невиновности и недальновидность при разработке стратегических государственных решений [Некрасов, 2019].

Некоторые эксперты предупреждают человечество о рисках совершенно другого масштаба, в которых машины, работающие на сверхбыстрых компьютерах, перехватят контроль над мировой экономикой. Например, социолог и экономист Робин Хэнсон считает правдоподобным, что в конце этого столетия произойдет перенос логико-математической системы из человеческого мозга в IT-систему [Хэнсон, www...]. При наличии достаточного количества быстрых, дешевых и параллельных компьютеров эмуляция мозга будет работать во много раз быстрее своего биологического аналога. Результатом станет «демографический взрыв», когда появятся множество чрезвычайно эффективных ученых во многих областях. В настоящее время человечеству очень далеко до первых IT-стимуляций мозга, не говоря уже о входных данных для иммерсивной реальности. Однако, реализация этого сценария развития искусственного интеллекта влечет за собой этические риски [Головкина, Валебникова, Чупина, 2023]. Во-первых, квалифицированные люди будут заменены сложными симуляциями и станут безработными. Во-вторых, если «эмуляторы» будут обладать феноменальной эрудицией и субъективными предпочтениями, то будут ли они эмоционально страдать от вынужденной трудовой деятельности? В-третьих, если роботы будут знать потаенные желания, то они смогут психологически манипулировать человеком [Саенко, 2024].

Как рациональный субъект, цифровой разум стремиться к максимизации заданной целевой функции, поэтому действовать этично и не противоречить интересам людей и других живых организмов — это вопрос, открытый [Ратнер, 2023]. Наделить машину человеческими ценностными стандартами не означает «порабощение». Ведь люди волей-неволей формируют ценности для своих биологических детей и это не происходит неэтичным образом. Если у человечества есть сильная этическая обязанность прививать детям основополагающие этические ценности, то это следует делать и с искусственным интеллектам. Только таким образом, машинный сверхразум даже без наличия сознания не сможет приносить вреда человеку с этической точки зрения.

Заключение

В заключение следует сказать, что в фундаментальных исследованиях электронного интеллекта у ученых растет интерес к результатам нейробиологии (как человек мыслит и воспринимает этот мир). Для рутинных процессов людям нужны системы, которые не ошибаются, быстрее человека, не знают чувств и превосходят нашу способность запоминать. Ведь никому не нужны эмоциональная поисковая система или складской робот. Но как только роботы выходят за пределы идеальных условий, то становится ясно, что интеллектуальные системы слишком примитивны и сверхспециализированы. Между заданными и четкими правилами игры и беспорядочным реальным миром огромная пропасть. Человек может ошибаться при умножении друг на друга двузначных чисел, но он может постоянно и успешно решать совершенно новые для него задачи. Поэтому уже сейчас становится ясно, что еще большее количество данных и еще более мощные компьютеры не приблизят нас к гибкому человеческому интеллекту. Нужны новые исследования, ориентированные на принципы человеческого познания [Яковлев, Андрейчук, Степанян, 2020]. Однако это может привести к тому, что «новые» риски будущего суперинтеллекта (научное творчество, здоровый «человеческий разум» и социальные навыки), как и другие возможные нарушения основных прав и ценностей человека, будут совершенно другими и сегодня их точно оценить нельзя.

Библиография

1. Ван Ю. Риски и стратегии противодействия, связанные с продвижением цифровых образовательных изменений с помощью искусственного интеллекта типа ChatGPT / Ван Ю. // Перспективы науки. 2024. № 2 (173). С. 184-192.
2. Головкина С.И. Риски и этические аспекты искусственного интеллекта / С.И. Головкина, Н.В. Валебникова, А.Д. Чупина // Российский экономический интернет-журнал. 2023. № 3.
3. Гуреева А.Н. Цифровые платформы как субъекты конфликтогенной коммуникации: особенности, эффекты, риски / А.Н. Гуреева, П.А. Киреева // Вопросы теории и практики журналистики. 2022. Т. 11. № 4. С. 753-771.
4. Костенков К.А. Искусственный интеллект и права человека: риски взаимодействия / К.А. Костенков // Теория государства и права. 2023. № 4 (33). С. 224-234.
5. Маркова Е.С. Искусственный интеллект: риски и угрозы / Е.С. Маркова, К.М. Пронин // Инновационная экономика и право. 2022. № 2 (21). С. 30-38.
6. Масовец А.А. Цифровые технологии: основные направления развития, эффективность и возможные риски / А.А. Масовец, Ю.А. Скоробогатова // Global and Regional Research. 2020. Т. 2. № 2. С. 34-40.
7. Михайлов М.А. Риски злонамеренного использования искусственного интеллекта и возможности их минимизации / М.А. Михайлов, Т.А. Кокодей // Всероссийский криминологический журнал. 2023. Т. 17. № 5. С. 452-461.
8. Некрасов В.Н. Искусственный интеллект в уголовном праве: за и против / В.Н. Некрасов // Baikal Research Journal. 2019. Т. 10. № 4. С. 20.
9. Паньшин Б. Результативная цифровая культура как фактор снижения рисков эмерджентности искусственного интеллекта / Б. Паньшин // Наука и инновации. 2023. № 11 (249). С. 13-21.
10. Прокофьев О.В. Принятие ответственных решений и риски искусственного интеллекта / О.В. Прокофьев // XXI век: итоги прошлого и проблемы настоящего плюс. 2023. Т. 12. № 3 (63). С. 26-31.
11. Ратнер Н.П. Этика искусственного интеллекта: вызовы, риски и решения / Н.П. Ратнер // Вестник науки. 2023. Т. 1. № 10 (67). С. 308-318.
12. Романова И.Н. Внедрение технологий искусственного интеллекта: анализ вероятных рисков и возможных угроз / И.Н. Романова // Материалы Ивановских чтений. 2021. № 4 (35). С. 15-18.
13. Саенко А.Ю. Отношение к технологиям искусственного интеллекта как к источнику глобальных рисков / А.Ю. Саенко // Человеческий фактор: Социальный психолог. 2024. № 2 (50). С. 117-124.
14. Хэнсон Р. Что произойдет, если мы загрузим наш мозг в компьютеры? / Р. Хэнсон Режим доступа: <https://ideanomics.ru/lectures/13827> (26.06.2024).
15. Яковлев К.С. Обзор современных и перспективных направлений исследований и глобальных правовых рисков в области искусственного интеллекта и робототехники / К.С. Яковлев, А.А. Андрейчук, А.Ж. Степанян // Международный правовой курьер. 2020. № 1-2 (35-36). С. 18-27.

Artificial Intelligence and Associated Risks

Konstantin A. Tatarinov

PhD in Economics, Associate Professor,
Baikal State University,
664003, 11 Lenina str., Irkutsk, Russian Federation;
e-mail: tatarinov723@gmail.com

Nikolai N. Anikienko

PhD in Economics, Associate Professor,
Irkutsk State Agrarian University,
664038, 1 Molodezhny Village, Irkutsk District, Russian Federation;
e-mail: anikienkonikolai@mail.ru

Inna A. Savchenko

PhD in Economics, Associate Professor,
Irkutsk State Agrarian University,
664038, 1 Molodezhny Village, Irkutsk District, Russian Federation;
e-mail: innasava2016@mail.ru

Sergei M. Muzyka

PhD in Biological Sciences,
Associate Professor,
Irkutsk State Agrarian University,
664038, 1 Molodezhny Village, Irkutsk District, Russian Federation;
e-mail: ignitmuz@mail.ru

Abstract

Computing machines have a number of advantages over the biological human brain, which was created by evolution over billions of years. Silicon microprocessors "trigger" and transmit signals millions of times faster than neurons, and supercomputers can be the size of a factory workshop. Software can be easily edited and multiplied to use an algorithm in several ways simultaneously. Large databases allow access to relevant information at any time. However, in terms of graceful degradation, mechanical resilience, and energy efficiency, the human brain surpasses the virtual one. In the coming years, quantum computers will disrupt the existing relationship between thermodynamic efficiency and complexity at the level of information processing, and there is a possibility that human intelligence will no longer prevail over machine intelligence. Therefore, it is important for humans to accurately predict and assess how and when this will happen and what the consequences of such an unfavorable development scenario will be. The article focuses not only on the issues of unfavorable development scenarios in traditional spheres of human activity (production, education, media, consumption, and the public sector) but also on the "new" risks of

future superintelligence. The emphasis is on the fact that the possible threats to humanity from machines can only be overcome by "instilling" fundamental ethical values in them.

For citation

Tatarinov K.A., Anikienko N.N., Savchenko I.A., Muzyka S.M. (2024) *Iskusstvennyi intellekt i svyazannye s nim riski [Artificial Intelligence and Associated Risks]. Kontekst i refleksiya: filozofiya o mire i cheloveke [Context and Reflection: Philosophy of the World and Human Being]*, 13 (9A), pp. 75-82.

Keywords

Ethical values, manipulation of people, unemployment, digital revolution, supercomputer, human intelligence, intellectual property, discrimination.

References

1. Wang Yu. Risks and counteraction strategies associated with the promotion of digital educational changes using artificial intelligence such as ChatGPT / Wang Yu. // *Prospects of Science*. 2024. No. 2 (173). pp. 184-192.
2. Golovkina S.I. Risks and ethical aspects of artificial intelligence / S.I. Golovkina, N.V. Valebnikova, A.D. Chupina // *Russian Economic Online Journal*. 2023. № 3.
3. Gureeva A.N. Digital platforms as subjects of conflictogenic communication: features, effects, risks / A.N. Gureeva, P.A. Kireeva // *Questions of the theory and practice of journalism*. 2022. Vol. 11. No. 4. pp. 753-771.
4. Kostenkov K.A. Artificial intelligence and human rights: risks of interaction / K.A. Kostenkov // *Theory of State and Law*. 2023. No. 4 (33). pp. 224-234.
5. Markova E.S. Artificial intelligence: risks and threats / E.S. Markova, K.M. Pronin // *Innovative Economics and Law*. 2022. No. 2 (21). pp. 30-38.
6. Masovets A.A. Digital technologies: main directions of development, efficiency and possible risks / A.A. Masovets, Yu.A. Skorobogatova // *Global and Regional Research*. 2020. Vol. 2. No. 2. pp. 34-40.
7. Mikhailov M.A. The risks of malicious use of artificial intelligence and the possibility of minimizing them / M.A. Mikhailov, T.A. Kokodey // *All-Russian Journal of Criminology*. 2023. Vol. 17. No. 5. pp. 452-461.
8. Nekrasov V.N. Artificial intelligence in criminal law: pros and cons / V.N. Nekrasov // *Baikal Research Journal*. 2019. Vol. 10. No. 4. P. 20.
9. Panshin B. Effective digital culture as a factor in reducing the risks of emergence of artificial intelligence / B. Panshin // *Science and Innovation*. 2023. No. 11 (249). pp. 13-21.
10. Prokofiev O.V. Making responsible decisions and risks of artificial intelligence / O.V. Prokofiev // *XXI century: results of the past and problems of the present plus*. 2023. Vol. 12. No. 3 (63). pp. 26-31.
11. Ratner N.P. Ethics of artificial intelligence: challenges, risks and solutions / N.P. Ratner // *Bulletin of Science*. 2023. Vol. 1. No. 10 (67). pp. 308-318.
12. Romanova I.N. Introduction of artificial intelligence technologies: analysis of probable risks and possible threats / I.N. Romanova // *Materials of the Ivanovo readings*. 2021. No. 4 (35). pp. 15-18.
13. Sayenko A.Yu. Attitude to artificial intelligence technologies as a source of global risks / A.Yu. Sayenko // *Human factor: Social psychologist*. 2024. No. 2 (50). pp. 117-124.
14. Hanson R. What happens if we upload our brains to computers? / R. Hanson Access mode: <https://ideanomics.ru/lectures/13827> (06/26/2024).
15. Yakovlev K.S. Review of modern and promising research areas and global legal risks in the field of artificial intelligence and robotics / K.S. Yakovlev, A.A. Andreychuk, A.Zh. Stepanyan // *International Legal Courier*. 2020. No. 1-2 (35-36). pp. 18-27.