1. ПРОБЛЕМНЫЕ И МЕТОДОЛОГИЧЕСКИЕ АСПЕКТЫ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

УДК 681.3

В.А. Артамонов, Е.В. Артамонова

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: УГРОЗА ИЛИ БЛАГО ДЛЯ ЧЕЛОВЕЧЕСТВА

Введение

История искусственного интеллекта (ИИ), как нового самостоятельного научного направления в области информационно-коммуникационных технологий (ИКТ), начинается в середине XX века. К этому времени уже было сформировано множество предпосылок его зарождения: среди философов давно шли споры о природе человека и процессе познания мира, нейрофизиологи и психологи разработали ряд теорий относительно работы человеческого мозга и мышления, экономисты и математики задавались вопросами оптимальных расчётов и представления знаний о мире в формализованном виде. И наконец, зародился фундамент математической теории вычислений – «теория алгоритмов», что привело к созданию первых компьютеров.

Возможности новых машин в плане скорости вычислений оказались больше человеческих, поэтому в учёном сообществе зародился вопрос: каковы границы возможностей компьютеров и достигнут ли машины уровня развития человека? В 1950 году один из пионеров в области вычислительной техники, английский учёный Алан Тьюринг пишет статью под названием «Может ли машина мыслить?», в которой описывает процедуру, с помощью которой можно будет определить момент, когда машина сравняется в плане разумности с человеком. Эта процедура получила название «тест Тьюринга». Далее, в 1956 г. ученый-информатик Джон Мак Карти ввел в обиход выражение «искусственный интеллект» (ИИ) для описания науки изучения разума путем воссоздания его ключевых признаков на компьютере. Создание разумной системы, с помощью рукотворного оборудования, вместо нашего собственного «оборудования» в виде клеток и тканей, должно было стать иллюстрацией полного понимания этой проблемы, и повлечь за собой практические применения в виде создания умных устройств или даже роботов.

Выдающиеся советские математики Андрей Николаевич Колмогоров и Владимир Игоревич Арнольд доказали в 1957 году теорему о том, что любая непрерывная функция нескольких переменных может быть представлена в виде комбинации конечного числа функций меньшего числа переменных, и именно это стало математическим обоснованием для построения нейронных сетей (НС). Было доказано, что соответствие между

зависимыми элементами различных множеств или функций может быть представлено нейросетью фиксированной размерности с прямыми связями с определенным количеством «нейронов» входного слоя, увеличенным числом «нейронов» каждого следующего скрытого слоя с определенными функциями активации и «нейронами» выходного слоя с неизвестными функциями активации. Причем, НС могут настраиваться или «обучаться». Для человека мало знакомого с математическими теориями всё звучит несколько сложно, но это имеет принципиальное значение для ответа на вопрос возможно ли создать искусственный интеллект.

С тех далёких времён, сейчас «почти былинных», ИКТ прошли большой путь, развиваясь по экспоненциальному закону, и ИИ также получил соответствующий эволюционный прогресс. Появились системы машинного обучения (МО), нейронные сети, системы поиска в больших данных (Big Data), интернет вещей (IoT), компьютерные игры, системы распознавания речи и образов, роботизированные комплексы, военный ИИ и др.

Однако, до настоящего времени нет единой концепции (парадигмы) анализа и синтеза систем ИИ, что породило массу мифов и догматических толкований этого научного направления.

Заблуждения и мифы относительно искусственного интеллекта

Миф 1: *Большинство исследований по архитектуре ИИ проводятся в предположении что, вычислительные мощности, доступные учёному на определённом отрезке времени, были бы постоянными и в данном случае использование человеческих знаний было бы одним из единственных способов повышения результативности научного поиска.*

Однако, через некоторое время, может быть даже несколько меньшее, чем нужно для типичного исследовательского проекта, по закону Мура, согласно которому производительность и вычислительная мощность компьютеров увеличивается в два раза каждые пару лет, учёным становятся доступными гораздо больше вычислительных ресурсов чем в начале исследований [1]. В поисках улучшений, которые могут помочь в краткосрочной перспективе, ученые пытаются использовать максимум человеческих знаний в этой области, но единственное, что имеет значение в долгосрочной перспективе – это нарастающее использование вычислительных мощностей. Эти два аспекта не должны идти вразрез друг с другом, но на практике идут. Время, потраченное на один из них, не равно времени, потраченному на другой. Есть некоторые психологические обстоятельства по инвестированию в тот или иной подход. А подход, основанный на знаниях человека, имеет тенденцию усложнять методы таким образом, что они становятся малопригодными для использования преимуществ методов, использующих крупномасштабные вычисления.

Вывод: *В исследовательских проектах нужно стараться сразу отбрасывать попытку решить задачу ИИ методом «мозгового штурма»,*

потому что пройдет некоторый период времени, и она решится гораздо быстрее и проще, благодаря росту мощности вычислений.

В недалёкой нашей истории можно привести много примеров, когда исследователи ИИ запоздало понимали этот горький урок. Рассмотрим некоторые из таких случаев.

В компьютерных шахматах методы, победившие чемпиона мира Каспарова в 1997 году, основывались на массивном глубоком поиске вариантов шахматных партий, когда либо сыгранных сильнейшими шахматистами мира. В то время к таким подходам с тревогой относилось большинство исследователей компьютерных шахмат, которые использовали методы, основанные на понимании человеком особой структуры этой игры. И тем не менее, когда более простой, основанный на поиске в Big Data подход со специальным аппаратным и программным обеспечением оказался намного более эффективным, исследователи, отталкивающиеся от человеческого понимания шахмат, не признали поражения. Они сказали: «В этот раз подход грубой силы, может быть, и победил, но он не станет общей стратегией, и уж точно люди не играют в шахматы таким образом». Эти исследователи хотели, чтобы методы, основанные на человеческом способе мышления, победили, и очень разочаровались, когда этого не произошло.

Аналогичная картина прогресса в исследованиях была замечена в игре ГО, превосходящей на порядок по сложности шахматы. Первоначально, огромные усилия направлялись для того, чтобы, используя человеческие знания и особенности игры, достигнуть победы над игроком-человеком. Однако все эти усилия оказались ненужными или даже вредными, как только исследователи эффективно применили поиск в больших данных и вычислительные мощности компьютеров. Также, важно было использовать машинное обучение в процессе самостоятельной игры, чтобы выявить ценностную функцию (как это было во многих других играх и даже в шахматах, хотя машинное обучение не играло большой роли в программе 1997 года, которая впервые обыграла чемпиона мира). Обучение игре с самим собой, обучение в целом, это как поиск, позволяющий применять огромные массивы вычислений. Поиск и обучение – это два самых важных класса технологий, которые задействуют огромные объемы вычислений в исследованиях ИИ.

В компьютерном противоборстве с человеком по игре в ГО [2], как и в компьютерных шахматах, первоначальные усилия исследователей были направлены на использование человеческого понимания (чтобы использовать меньше поиска в больших данных), и лишь много позже был достигнут гораздо больший успех – за счет использования поиска и МО.

В области распознавания речи в 1970-х годах был проведен конкурс, спонсируемый DARPA (Управление стратегических исследований министерства обороны США). Участники представляли различные методы,

которые использовали преимущества человеческого знания – знания слов или фонем, человеческого голосового тракта и так далее. По другую сторону баррикад были более новые методы, статистические по своей природе, и выполняющие больше вычислений, на основе скрытых моделей Маркова. И опять же статистические методы победили методы, основанные на знаниях человека. Это привело к серьезным изменениям во всей технологии по обработке естественного языка, и в итоге, статистика и вычисления начали доминировать в этой области. Недавний рост глубокого МО в области распознавания речи – это самый последний шаг в этом исследовательском направлении. Методы глубокого машинного обучения еще меньше полагаются на человеческие знания и используют всё больше вычислительных ресурсов, наряду с обучением на огромных наборах данных, и выдают потрясающие результаты при реализации систем распознавания речи и образов.

Как и в играх, ученые всегда пытались создавать системы, которые будут работать так, как они представляли этот процесс в своих головах, т.е. они пытались поместить свои знания в эти системы, однако, все это выходило крайне непродуктивно, ученые просто тратили время, до тех пор, пока вследствие закона Мура, им становились доступными более мощные компьютеры. Вследствие чего проблема решалась совершенно на другом уровне.

Похожая картина была и в области компьютерного зрения. Первые методы воспринимались как поиск неких контуров, обобщенных цилиндров, либо с применением возможностей SIFT (масштабно-инвариантной трансформации признаков). Но сегодня все это уже в прошлом. Современные нейронные сети с глубоким обучением используют понятие свертки и определенных инвариантов, это работает намного лучше.

В какую бы область мы ни заглянули, мы везде продолжаем совершать одни и те же ошибки. Чтобы увидеть это и эффективно побороть, нужно понять, почему эти ошибки так привлекательны. Мы должны усвоить «горький урок», состоящий в том, что построение нового, отталкиваясь от того, как мы думаем, не работает в долгосрочной перспективе.

Опыт, основанный на исторических наблюдениях, показывает, что исследователи ИИ часто пытаются встроить знание в своих «агентов» – это всегда помогало в краткосрочной перспективе и приносило ученым удовлетворение, но в долгосрочной перспективе все заходило в тупик и тормозило дальнейший прогресс. Прорывной прогресс неизбежно приходил с применением противоположного подхода, основанного на масштабировании вычислений за счет поиска в больших данных и машинного обучения. Успех иногда разочаровывал исследователей и зачастую не воспринимался полностью, потому что это был успех вычислений, а не успех ориентированных на человека подходов.

Второе, что следует извлечь из этого горького урока, состоит в том, что фактическое содержание человеческого ума чрезвычайно сложное и кажущееся иногда непознаваемым. Нам стоит перестать пытаться найти простые способы осмыслить содержание ума, похожие на простые способы осмысления пространства, объектов, множественных «агентов» или симметрий. Все они являются частью произвольно сложного внешнего мира. Нам не стоит пытаться от них отталкиваться, потому что их сложность бесконечна. Нам стоит строить свои стратегии научного поиска в области ИИ на мета-методах, которые могут находить и улавливать эту произвольную сложность. Эти методы могут находить хорошие приближения, но поиск их должен осуществляться новыми методами, а не нами умозрительно. Нам нужны «агенты» ИИ, которые могут открывать новое в мироздании, как это делают люди, а не содержать то, что мы уже открыли. Построение на наших открытиях только усложняет процесс познания мира и поиска новых сущностей.

Выводы: *1. Нужно опираться на масштабируемые вычисления и поиск в больших данных, а не пытаться воспроизвести человеческие размышления и догмы, в попытках объяснить сложные методы познания простыми схемами, ибо в долгосрочной перспективе сработает первое, а не последнее.*

2. Поиск в больших данных и машинное обучение, подпитанные вычислительной мощностью, намного превосходят попытки решить задачу «нестандартным подходом человеческого мышления».

Миф 2: *В результате экспоненциального роста производительности компьютеров наступит время «технологической сингулярности», когда вычислительная (логическая) мощность ИИ сравняется по интеллекту с человеческим разумом, и как поведёт себя этот искусственный разум в «постчеловеческом мире» невозможно предугадать.*

Для каждого периода времени развития человечества характерна своя трансформация, которую можно описать как некую совокупность промышленных технологий, позволяющих создать определенный качественный скачок в росте производительности труда. Это определение вписывается в широко принятую концепцию смены технологических укладов, где трансформация на базе ИКТ является одним из этапов (см. рис.1). Кривая изменения экономического прогресса (роста производительности труда) отображается в виде S-образной кривой с периодами зарождения (медленного роста), активного роста и зрелости (замедления роста). Совокупность технологических инноваций приводит к смене одного уклада на другой. Каждый из этапов экономического прогресса на рис.1 (включая стадию ИКТ) можно разделить на более мелкие части, и в каждой выделить свои трансформирующие технологии.

Осмысление экспоненциального роста технологий потребовало некоторого времени, прежде чем они получили полное признание всего

за несколько лет. Этой тенденции следуют самые разные области, такие как искусственный интеллект и машинное обучение в качестве одной из ветвей развития ИКТ, робототехника, военное дело, здравоохранение, электро- и самоуправляемые автомобили, образование, 3D-печать, промышленность и сельское хозяйство.

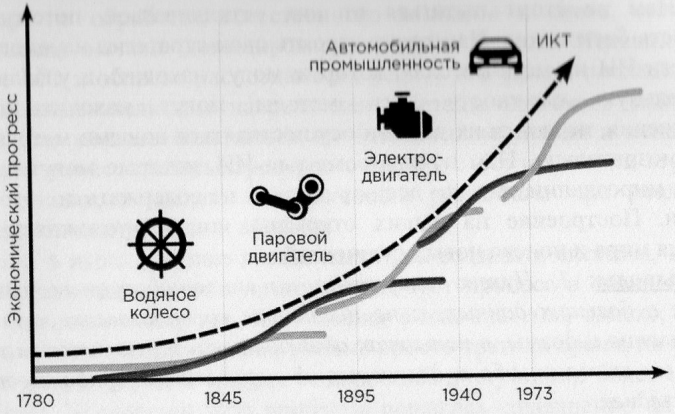

Рисунок 1 – Трансформирующие технологии и технологические уклады, (источник: M. Hilbert, University of California)

Добро пожаловать в 4-ю промышленную революцию. Добро пожаловать в Экспоненциальный Век. Эта концепция была предложена Вернором Винжем, который предположил, что если мы сумеем избежать гибели цивилизации до этого самого века, то сингулярность произойдет из-за прогресса в области искусственного интеллекта, интеграции человека с ИКТ или других методов увеличения разума [3]. Усиление разума, по мнению Винжа, в какой-то момент приведет к положительной обратной связи: более разумные системы могут создать еще более интеллектуальные системы и сделать это быстрее, чем первоначальные их конструкторы – люди. Эта положительная обратная связь, скорее всего, окажется столь сильной, что в течение очень короткого промежутка времени (месяцы, дни или даже всего лишь часы) мир преобразится больше, чем мы сможем это представить, и внезапно окажется населен сверхразумными созданиями.

По мнениям некоторых учёных футурологов [4] и того же Винжа, придерживающихся концепции сингулярности, она должна наступить около 2030 г. и даже по самому пессимистическому сценарию не позднее середины этого века, т.е. в 2050 году. Сторонники теории технологической сингулярности считают, что если возникнет принципиально отличный от человеческого разум (*постчеловек*), дальнейшую судьбу цивилизации невозможно предсказать, опираясь на человеческую логику. С понятием

сингулярности часто связывают идею о невозможности предсказать, что будет после нее.

Постчеловеческий мир, который в результате появится, возможно, будет столь чуждым для нас, что сейчас мы не можем знать о нем абсолютно ничего. Единственным исключением могут быть фундаментальные законы природы, но даже тут иногда допускается существование еще неоткрытых законов (у нас пока нет теории квантовой гравитации) или не до конца понятых следствий из известных законов (путешествия через пространственные «дыры», рождение "вселенных-карликов", путешествия во времени и т.п.), с помощью которых *постлюди* смогут делать то, что мы привыкли считать физически невозможным.

Вывод: *Вопрос предсказуемости важен, поскольку, не имея возможности предсказать хотя бы некоторые последствия наших действий, нет никакого смысла в том, чтобы пытаться направить развитие в желательном направлении.*

Миф 3: *Угрозы ИИ и кризис человечества.*

Человечество стоит на пороге не только технологического, но и философского кризиса, считает историк Юваль Харари, автор книги «Sapiens: Краткая история человечества» [5]. Новые технологии формируют новые формы антиутопии. И общество пока не понимает, как адаптироваться к меняющейся реальности.

Харари вывел формулу предстоящего глобального кризиса:

$$B * C * D = HH$$

В данном случае **B** – это познания в биологии, **C** – это вычислительная мощность, а **D** – это данные. Если помножить их друг на друга, появится возможность взламывать людей (**HH** – hack humans).

Под взломом исследователь подразумевает возможность управлять человеком на глубинном уровне, то есть контролировать его желания и стремления. Харари опасается, что власти и корпорации скоро изучат людей настолько, что смогут с легкостью регулировать их мысли.

Технологии отдаленно будут напоминать таргетированную рекламу, только их действие будет более точным, а эффект – стопроцентным.

Ранее исследователь отмечал, что в сложившихся обстоятельствах привычные философские концепции отмирают. Это касается свободы воли и свободы выбора. Люди ошибочно полагают, что контролируют ситуацию, но на самом деле это не так.

Главное следствие масштабного внедрения искусственного интеллекта – *это утрата человеком автономии и авторитета.* При этом ИИ не обязательно выходить на один интеллектуальный уровень с людьми и обладать сознанием. Алгоритмам МО достаточно будет изучить личность досконально, чтобы найти самую слабую точку и запустить процесс манипуляций.

связано с парадоксами, открытыми австрийским математиком Гёделем в 1930-х годах [7].

Парадоксы – это формально-логические противоречия, которые возникают в теории множеств и формальной логике при сохранении логической правильности рассуждения. Парадоксы возникают тогда, когда два взаимоисключающих (противоречащих) суждения оказываются в равной мере доказуемыми.

С точки зрения математики, вопрос «обучаемости» сводится к тому, сможет ли алгоритм извлечь шаблон из ограниченных данных. Ответ на этот вопрос связан с парадоксом, известным как вышеупомянутая континуум-гипотеза (проблема континуума или 1-я проблема Гильберта) и разрешенным в 1963 г. американским математиком Полом Коэном [7].

Решение оказалось весьма неожиданным: то, что утверждается в гипотезе континуума, нельзя ни доказать, ни опровергнуть, исходя из аксиом теории множеств. Гипотеза континуума логически независима от этих аксиом. Неспециалисту довольно трудно понять, почему утверждения такого рода играют для математики столь большую роль и ставятся на первое место в списке важнейших проблем. Отметим лишь, что на самом деле речь идет о вещах принципиальных и фундаментальных, так как континуум – это, по сути, базовая математическая модель окружающей нас физической, пространственно-временной реальности (частью которой являемся и мы сами), а в математике континуум – еще и синоним совокупности всех действительных чисел, также центрального понятия математики и ее рабочего инструмента.

По сути, Гёдель и Коэн доказали, что континуум-гипотеза не может быть доказана ни как истинная, ни как ложная, начиная со стандартных аксиом, утверждений, принятых как истинные для теории множеств, которые обычно принимаются за основу всей математики.

Иными словами, утверждение не может быть ни истинным, ни ложным в рамках стандартного математического языка.

Вывод: *Математически доказано, что возможности ИИ не беспредельны. И какими бы огромными вычислительными ресурсами не обладал человек, машинное обучение никогда не приведёт к победе искусственного разума над человеческим.*

В пользу данного доказательства говорят и последние исследования нейробиологов в области исследования структуры и возможностей человеческого мозга.

Так учёные Стэнфордского университета потратили несколько лет, разрабатывая новый способ 3D-сканирования мозга. Они совместили объёмную компьютерную томографию (*англ., array tomography* – техника «антенных решёток» из радиоастрономии) и специально разработанный софт, чтобы получить объёмную и реалистичную 3D-модель. Такую,

по которой можно перемещаться, масштабировать и вращать её в разных измерениях.

Изучив полученную картину, учёные пришли к выводу, что синапсы (соединительные ткани нервных клеток) устроены гораздо сложнее, чем предполагалось раньше. Здоровый человеческий мозг содержит около 200 млрд нервных клеток, которые соединяются друг с другом сотнями триллионов синапсов. От каждой нервной клетки могут отходить десятки тысяч синапсов. В одной только коре больших полушарий человека находится около 125 трлн. синапсов – в 1500 раз больше, чем звёзд в нашей галактике. По результатам визуальной реконструкции данных учёные обнаружили, что каждый синапс содержит около 1000 молекулярных «переключателей», на подобие аналоговых транзисторов. То есть отдельный синапс можно сравнить с микропроцессором. Получается, что количество «транзисторов» в человеческом мозге теперь нужно увеличить на три порядка. Их больше, чем транзисторов во всех компьютерах на планете и маршрутизаторах вместе взятых [9].

Вывод: *Получается что один человеческий мозг по сложности примерно равен всей мировой ИТ-инфраструктуре, а учитывая тот факт, что возможности человеческого мозга задействованы человечеством максимум на 20%, говорить о победе ИИ над человеческим разумом не приходится даже в отдалённой перспективе.*

Заключение

Проблемам информационной безопасности ИКТ и защищённости человеческого социума от негативного воздействия ИИ и МО уделено достаточно много внимания в ряде исследований.

Выделяются основные проблемы: нарушение работоспособности технического и программного обеспечения, распространение информационного оружия, придание традиционному оружию новых качеств, непрерывное усложнение информационных и коммуникационных систем, возможность концентрации информационных средств в руках небольшой группы собственников, использование во вред информационных данных, манипулирование сознанием, использование технологического воздействия на психическую деятельность [10].

Однако, несмотря на то, что в научном сообществе нет однозначного ответа на вопрос: *ИИ – это искусство или наука*, технологии искусственного интеллекта рассматриваются как одно из самых действенных средств в области информационных технологий, автоматизации производственных процессов, транспорта, национальной безопасности, включая кибербезопасность – сейчас и в будущем.

Почему ИИ – это будущее кибербезопасности? [11].

Обнаружение мошенничества, обнаружение вредоносных программ, обнаружение вторжений, оценка риска в сети и анализ поведения

пользователя/машины – это пятёрка самых актуальных способов применения ИИ для улучшения кибербезопасности. ИИ реально меняет привычные аспекты кибербезопасности. Он улучшает способность предвидеть и предотвращать киберпреступления, защищает устройства с нулевым уровнем доверия, может контролировать даже устаревание паролей! Таким образом, искусственный интеллект действительно необходим для обеспечения безопасности периметров любых объектов хозяйственной или финансовой деятельности.

Поиск взаимосвязей между угрозами и анализ вредоносных файлов, подозрительных IP-адресов или необычную деятельность сотрудника длится считанные секунды или минуты. Уже сейчас ИИ помогает человеку обеспечивать кибербезопасность. А в дальнейшем его возможности будут только расширяться, делая участие человека в процессе защиты чисто номинальным.

В банках, благодаря ИИ, антифрод-системы станут работать надёжнее и быстрее, что позволит сэкономить доверие и деньги как клиентов финансовых учреждений, так и самих банковских служащих. А по мнению компании Dell, занимающейся разработкой подобных продуктов, ИИ способен защитить, контролировать и отслеживать данные в гибридных средах, а также предотвращать 99% атак вредоносного ПО.

Кроме того, ИИ вполне можно сделать облачным. Это позволит ему автоматически масштабироваться при резком повышении нагрузки (например, если хакеры пытаются «атаковать» сервер или замаскировать свою активность под лавиной типовых действий в другом направлении). «Облако» позволит расширить безопасный периметр компании, если еще и вся носимая электроника (гаджеты) будет подключена к контролируемой ИИ среде.

Литература

1. G.E. Moore, No exponential is forever: but "Forever" can be delayed! [Электронный ресурс]. – URL: https://ieeexplore.ieee.org/document/1234194// (дата обращения: 25.04.2020).

2. Человек отстал от компьютера // Российская газета. вып. № 54 (6922) 2016 г. – URL: https://rg.ru/2016/03/15/chempion-mira-po-go-proigral-kompiuternoj-programme.html.

3. Vinge, V. 1993. "The Coming Technological Singularity" [Электронный ресурс]. – URL: http://www-ohan.sdsu.edu/faculty/vinge/misc/singularity.html// (дата обращения: 25.04.2020).

4. А. Новоселов. Технологическая сингулярность как ближайшее будущее человечества. [Электронный ресурс]. – URL: http://andrzej.virtualave.net/Articles/singularity.html// (дата обращения: 25.04.2020).

5. Харари Ю. «Sapiens: Краткая история человечества» [Электронный ресурс]. – URL:http://www.labirint.ru/books/498309// (дата обращения: 25.04.2020).

6. Colors collective [Электронный ресурс]. – URL:https://www.quanta-magazine.org/mathematicians-measure-infinities-find-theyre-equal-20170912// (дата обращения: 25.04.2020).

7. Демидов С.С. «Математические проблемы» Гильберта и математика XX века // Историко-математические исследования. – М.: Янус-К, 2001. – № 41 (6). – С. 84-99.

8. Ben-David S., Hrubes P., Moran S., Shpilka A., Yehudayoff A. Learnability can be undecidable. Nature Machine Intelligence, 1, pp. 44-48.

9. В человеческом мозге столько же «транзисторов», сколько их в мировой ИТ-инфраструктуре. [Электронный ресурс]. – URL:https://www.cell.com/neuron/fulltext/S0896-6273(10)00766-X // (дата обращения: 25.04.2020).

10. Артамонов, В.А. Безопасность информационно-коммуникационных технологий в контексте устойчивого развития социума / В.А. Артамонов, Е.В. Артамонова, Л.А. Кулак // Цифровая трансформация. – № 2 (7), 2019. – С. 36-45.

11. Кибербезопасность, будущее и ИИ. [Электронный ресурс]. – URL:https://www.securitylab.ru/contest/500573.php // (дата обращения: 25.04.2020).

МНОО «МАИТ», г. Минск