Артамонов В. А.

д.т.н., профессор, академик МАИТ

Артамонова Е. В.

к.т.н. (PhD), член МАИТ

Международное научное общественное объединение «Международная академия информационных технологий» (МНОО МАИТ) рег. ООН, г. Минск

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И БЕЗОПАСНОСТЬ: ПРОБЛЕМЫ, ЗАБЛУЖДЕНИЯ, РЕАЛЬНОСТЬ И БУДУЩЕЕ

Ключевые слова: *информационно-коммуникационные технологии (ИКТ), математическая теория вычислений, искусственный интеллект (ИИ), машинное обучение (МО), нейронные сети, большие данные (big data), интернет вещей (IoT), закон Мура, компьютерные шахматы, игра ГО, распознавание речи и образов, технологическая сингулярность, постчеловеческий мир, угрозы ИИ, цифровой колониализм, парадоксы теории множеств, гипотеза континуума, кибербезопасность.*

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 году 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 нейросетей. 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 нейросети 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.

Со времён Тьюринга и Мак Карти 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, 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.

Сейчас так много говорят и пишут про искусственный интеллект, что начинает казаться, что он уже давно создан и присутствует повсюду. На самом деле это не так. Хотя автоматизация давно уже стала частью производственных и управленческих процессов, а компьютеры и научились распознавать речь и лица, управлять автомобилями и анализировать гигантские массивы данных. Распознавание образов или автоматические переводчики относятся к технологиям машинного обучения, использующим методологию искусственного интеллекта и отбор накопленных результатов при решении сходных задач. И всё же, существует принципиальный вопрос: искусственный интеллект – это искусство или наука

Искусственного интеллекта как целостного продукта ИКТ пока нет, как нет и фундаментальной теории его построения, их ещё только предстоит создать, а областей применения ИИ действительно очень много. Экономика и оборонная сфера не являются исключением. Развитие искусственного интеллекта становится задачей *обеспечения национальной безопасности и устойчивого развития государства.*

Искусственный интеллект и кибербезопасность

Работа в киберпространстве и отслеживание постоянно возникающих киберугроз требует большого количества высококвалифицированных специалистов. Искусственный интеллект также мог бы взять часть их работы на себя, поскольку значительно быстрее может находить уязвимости и генерировать коды и машинные алгоритмы. Выявление угроз и уязвимостей достигло огромных масштабов и становится проблемой для средств защиты периметров и данных систем ИКТ, контролируемых человеком. Кибератаки становятся намного более сложными и очень разрушительными, а также возрастают риски попадания таких опасных технологий в руки злоумышленников и конкурентов.

В связи с бурным развитием интернета вещей (IoT), «облачных вычислений», центров обработки данных (ЦОД) и соответственно технологий обработки «больших данных» (big data) произошла смена парадигмы информационной безопасности (ИБ) — от защиты периметра корпоративной сети, «облака» или ЦОДа к защите самих данных. Традиционно для безопасного доступа к ресурсам или объектам информационных технологий используются виртуальные частные сети (VPN) в качестве туннеля на базе криптографических методов защиты информации. Но использование исключительно VPN сопряжено с риском для безопасности. Проблема заключается в том, что VPN использует подход к безопасности на основе периметра. Пользователи подключаются через VPN-клиент, но оказавшись внутри периметра, они часто получают широкий доступ к информационным ресурсам и управлению объектом. Каждый раз, когда устройство или пользователь автоматически получает такое доверие, это создает угрозу для данных, приложений и интеллектуальной собственности организации. Помимо проблем, связанных с использованием VPN для удаленного доступа, сетевые операторы ищут оптимальный способ защиты приложений. То, что часть приложений размещены в облаке, а часть — локально, затрудняет предоставление общего метода контроля и его применение, особенно когда одни пользователи работают в офисе, а другие — удаленно. Развертывание приложений в облаке создает возможность зондирования со стороны нежелательных субъектов, что существенно повышает риск. Выход из этой ситуации предлагается в использовании новой технологии выхода за пределы VPN — *доступ с нулевым*

доверием (ZTNA), который с использованием ИИ и МО предлагает более эффективное решение для удаленного доступа, которое также решает проблемы, связанные с доступом к приложениям. Термин «нулевое доверие» следует понимать в буквальном смысле. Данная модель обеспечения безопасности предполагает, что ни один пользователь или устройство не считаются надежными, и ни одна транзакция не заслуживает доверия без предварительной проверки авторизации пользователя и устройства. Поскольку ZTNA исходит из идеи, что местоположение не подразумевает никакого уровня доверия, место работы пользователя перестает иметь значение. Этот подход с нулевым доверием применяется независимо от того, где физически находится пользователь или устройство. Поскольку любое устройство потенциально может быть заражено, и любой пользователь способен на вредоносное поведение, политика доступа ZTNA отражает эту реальность. В отличие от традиционного VPN-туннеля с неограниченным доступом, ZTNA предоставляет доступ каждому приложению и рабочему процессу на сеанс только после аутентификации пользователя и/или устройства. Прежде чем получить доступ, пользователи проходят верификацию и аутентификацию для доступа к приложению. Каждое устройство также проверяется при каждом доступе к приложению для обеспечения соответствия требованиям политики доступа к приложению. При авторизации используется различная контекстная информация, такая как роль пользователя, тип устройства, соответствие устройства требованиям, местоположение, время и способ подключения устройства или пользователя к сети или ресурсу. Если используется технология ZTNA, то после того, как пользователь указал соответствующие учетные данные, необходимые для многофакторной аутентификации и проверки конечной точки, и подключился, ему предоставляется доступ с ограниченными полномочиями. Пользователь может получить доступ только к тем приложениям, которые ему необходимы для эффективного выполнения своей работы, и ни к чему другому. Контроль доступа не заканчивается на точке доступа. ZTNA работает в контексте идентификации, а не защищает участок сети, что позволяет политикам безопасности (ПБ) отслеживать приложения и другие транзакции от начала до конца. Благодаря высокому уровню контроля доступа ZTNA является более эффективным решением для конечных пользователей и обеспечивает применение ПБ везде, где это необходимо. И хотя процесс аутентификации ZTNA предоставляет точки проверки подлинности, в отличие от традиционной VPN, он не определяет, как эта аутентификация происходит. По мере внедрения новых решений аутентификации их можно легко добавлять в стратегию ZTNA. Новые решения проверки подлинности могут помочь устранить проблемы, связанные со слабыми или украденными паролями и учетными данными,

решить проблемы, связанные с недостаточной безопасностью некоторых устройств Интернета вещей (IoT), или добавить дополнительные уровни проверки для доступа к конфиденциальной информации или важным ресурсам.

Обеспечение доступности объекта защиты является неотъемлемой частью обеспечения ИБ, поэтому системы мониторинга производительности и доступности являются обязательным инструментом при осуществлении мониторинга ИБ в ИТ-системах. Системы мониторинга производительности могут использоваться как отдельно, так и являться одним из источников событий для системы управления событиями — SIEM (англ. – Security Information Event Management). Такие системы предназначены для отслеживания состояния функционирования разнообразных сервисов сети и ее узлов (серверов, сетевого оборудования, приложений и других), в том числе подсистем ИБ, на основе различных критериев производительности и доступности. В решениях применяются следующие основные методы контроля функционирования с точки зрения обеспечения доступности систем:

- сбор и агрегация разнообразных данных, показателей и счетчиков об использовании аппаратных ресурсов системы, как правило посредством устанавливаемых агентов на контролируемых узлах или с использованием протокола SNMP (уровень потребления CPU, память, жестких дисков, сетевых адаптеров и других данных);
- анализ и корреляция собранных данных для определения или упреждения достижения пороговых значений показателей производительности и доступности с целью реагирования или предотвращения нештатных ситуаций функционирования систем;
- автоматизированное выполнение заранее запрограммированных тестов, выполняющих проверку функционирования различных параметров сервисов по заданному сценарию. Успешное выполнение таких тестовых сценариев позволяет подтверждать доступность сервисов и систем на различных уровнях;
- автоматизированное реагирование системы в виде выполнения заданных скриптов, программ или задач при выявлении значимых отклонений показателей на этапе корреляции;
- генерации оповещений (уведомлений) о выявленных отклонениях в производительности и доступности систем. Оповещение может, как выводиться на экран мониторинга интерфейса системы, так и направлено в различные каналы оповещений: по электронной почте, на GSM-шлюз, в системы обмена мгновенными сообщениями (например, jabber) и другие;
- визуализация собираемых данных в виде диаграмм, помогающих идентифицировать аномалии или значимые отклонения, отличные от стандартного поведения систем. Так же визуализация включает в себя представление данных в виде отчетов;

- хранение собранных данных в базе данных.

Так как целью нашей статьи является описание методов мониторинга информационной безопасности в части использования ИИ и МО, мы не будем подробно рассматривать данные системы мониторинга с точки зрения их архитектуры и особенностей функционирования.

Ещё одной технологией обеспечения кибербезопасности с широким использованием функций ИИ и МО является **DLP**. Эта аббревиатура на английском расшифровывается как *Data Loss Prevention* (предотвращение потери данных) или *Data Leakage Prevention* (предотвращение утечки данных). Чаще всего для продуктов этого класса используется именно это сокращение. Но встречаются другие. Если вам попадается аббревиатура ILP, ILDP, EPS или CMF, скорее всего речь тоже идет о системе безопасности, обеспечивающей защиту от утечек. DLP-системы строятся на анализе потоков данных, пересекающих периметр защищаемой информационной системы. При детектировании в этом потоке конфиденциальной информации срабатывает активная компонента системы, и передача сообщения (пакета, потока, сессии) блокируется. В основе «интеллектуальных функций» DLP лежит технология так называемого контентного анализа, включающая:

Поиск по ключевым словам и словарям. Предполагает поиск точных совпадений текстовых строк. При этом в текстовых строках могут использоваться спецсимволы, обозначающие группы символов. Не стоит путать поиск по ключевым словам с технологией лингвистического анализа, которая также имеется в DLP и учитывает различные словоформы. Офицер безопасности может создавать собственные словари под конкретные тематики или воспользоваться одним из предустановленных словарей.

Машинное обучение. Собственно самообучающаяся технология DLP построена на основе алгоритма Байеса и метода опорных векторов. В процессе анализа различных документов технология самостоятельно выделяет признаки различных категорий конфиденциальных данных. Чем больше конфиденциальных документов увидит система на этапе обучения, тем выше будет её результативность в ежедневной работе.

Цифровые отпечатки Docu Prints. Технология основана на сравнении перехваченной информации с образцами конфиденциальных документов, в том числе оцифрованных документов по их характерным деталям на бланке. Технология особенно эффективна для контроля документов, объём которых значителен, а содержимое при использовании меняется незначительно. При этом DLP устойчива к модификациям исходных документов и максимально точно обнаруживает информацию, полностью или частично совпадающую с заданными конфиденциальными данными.

Анализ графических файлов (OCR). Технология распознаёт конфиденциальные данные, содержащиеся в скриншотах, фотографиях, отсканированных, рукописных и других графических документах. Интегрированные в DLP OCR-модули ABBYY FineReader и Google Tesseract извлекают из изображений текст для последующего анализа и проверки соответствия DLP-политикам безопасности.

Графические отпечатки. Технология применяется для обнаружения таких элементов в графических файлах, как подписи, печати, бланки, а также изображения с определённой структурой, например, сканы паспортов или водительских прав. Искажения цветов, наклон изображения, перекрывающие элементы (например, надписи и фоновая графика на печати) не мешают распознаванию образов. Кроме шаблонов и регулярных выражений DLP обнаруживает передачу структурированной информации — паспортных данных, адресов, номеров кредитных карт, банковских счетов, URL-адресов, номеров телефона и других, а также любых типов данных по заданным регулярным выражениям.

Давайте сопоставим возможности этих двух последних нами рассматриваемых систем SIEM и DLP. Отметим, что одна система при этом не заменяет другую, они решают разные задачи. Простой пример: сотрудник несколько раз ввел неверный пароль. SIEM не только обнаружит эти действия, но и сопоставит факторы — сколько раз пароль введен неверно? в течение какого времени? Система выявит угрозу информационной безопасности — кто-то пытается подобрать пароль к учётным данным — и своевременно оповестит о ней. Тем не менее без глубокого анализа SIEM бесполезен. DLP в свою очередь позволяет детализировать данные и выяснить подробности инцидента. Такой симбиоз SIEM и DLP в разы повышает уровень информационной защиты организации и упрощает работу службе безопасности.

Сейчас DLP-разработчики активно интегрируют свои решения с популярными SIEM-системами. Процесс только набирает обороты и говорить о конкретных результатах еще рано. Для пользователей интеграция своих данных в SIEM — процесс небыстрый, ведь недостаточно отдать сторонней системе данные, необходимо понять, какие конкретно задачи способна решать такая схема, как именно поможет такой союз. Потребуется еще не одна итерация, чтобы данная связка начала решать прикладные задачи — из всего невообразимого массива данных выбирать нужные и представлять из себя удобный и функциональный инструмент для решения реальных проблем, а не просто конструктор.

Однако в чем основная проблема большинства SIEM-систем и DLP? В том, что мало кто понимает, как использовать огромный массив собранной информации (Big Data). Работа с системой требует специальных знаний и навыков (Data Science), если их нет, то SIEM и DLP

превращаются в дорогой и, практически, бесполезный для бизнеса инструмент. Потому задача разработчиков сделать так, чтобы симбиозная система была максимально дружелюбна к МО и интуитивно понятна ИБ-специалистам.

Задачи искусственного интеллекта в сфере национальной безопасности

Новая мировая гонка технологий в обозримом будущем приведет к внедрению самых современных инноваций в военную сферу. Этим будут заниматься все ведущие мировые державы, потому что любое отставание от соперников увеличивает уязвимость, которую будет очень сложно прикрыть обычными конвенциональными видами вооружений. К тому же появление новых технологий может привести к заметным изменениям в стратегиях, планировании и организации деятельности вооруженных сил. Предотвратить использование искусственного интеллекта в военных целях невозможно. Искусственный интеллект значительно расширит возможности сбора и анализа данных, что позволит получить определенные преимущества в быстроте и качестве обработки информации. В области военной разведки появится больше возможностей и разного рода источников информации, но и возможностей скрыть истину от противника также прибавится. Искусственный интеллект может дополнить информационное пространство большим объемом искусственно созданных данных, виртуальной истиной, что с одной стороны запутает потенциальных противников, но с другой стороны, может создать дополнительные политические риски.

Сейчас даже те технологии, что уже созданы в области машинного обучения и искусственного интеллекта, имеют значительный потенциал для обеспечения национальной безопасности. Существующая технология распознавания образов может обеспечить высокую степень автоматизации при анализе спутниковых снимков и данных радаров. Искусственный интеллект способен увеличить эффективность работы радиолокационных станций системы предупреждения о ракетном нападении и системы обработки информации на радиооптических комплексах распознавания. К тому же происходящая сейчас миниатюризация спутников и увеличение их количества на орбитах потребует технологии для быстрого распознавания.

Еще масштабнее задачи у комплексов обработки информации загоризонтных радиолокаторов, использующих принцип пространственного ионосферного распространения радиоволн длиной свыше 10 метров или дифракционного поверхностного распространения более коротких радиоволн. Эти радары «видят» все движущиеся объекты, включая гражданскую технику, поэтому стоит задача быстрого распознания среди всех поступающих

тысяч и даже миллионов образов именно военные объекты, а также необычное поведение на земле и в воздухе. Это колоссальные массивы информации и образов, с обработкой которых людям без помощи машин не справиться. К тому же военные получат так называемую «библиотеку целей», что поможет системам распознавания и наведения. Если для противодействия ПЗРК с инфракрасной головкой самонаведения достаточно с борта самолета или вертолета отстреливать ложные тепловые цели, а против радиолокационной ставить помехи, то системы с искусственным интеллектом, даже если он находится не в ракете, а в руках оператора, «видят» летательный аппарат целиком. А во-вторых у загоризонтных РЛС существует проблема несовместимости со стандартной системой радиолокационного опознавания «свой-чужой», поэтому при анализе воздушной обстановки помощь искусственного интеллекта будет весьма кстати. Искусственный интеллект также можно использовать для противодействия радарам противника, изучая его работу и подбирая методы подавления радиосигнала.

Возможности автономных систем пока ограничены. Несмотря на то, что такие системы «квазиавтономны», в них люди все равно должны быть всегда в контуре управления и непосредственно принимать решения на применение оружия. Но обычный человек в сравнении с возможностями современной военной техники — существо слабое, хрупкое и бестолковое, а в цепочке принятия боевых решений — еще и самое медленное звено. Искусственный интеллект призван полностью исключить человека из системы принятия решений, а заодно и сохранить жизни военнослужащим.

В боевых условиях преимущество у тех, кто примет решение быстрее и ударит первым, поэтому полностью автономные системы получат в будущем заметное развитие. Более того, уже появилась концепция «контравтономности», согласно которой искусственный интеллект, подвергшийся нападению, но при этом не уничтоженный, очень быстро обучится и сделает выводы, после чего нанесет противнику смертельный удар. Возможностей применения тактического оружия с искусственным интеллектом множество. Это и беспилотные летательные аппараты, и бронемашины, и ракетные катера, самостоятельно находящие цели и принимающие решения на их уничтожение. Сейчас происходит быстрое удешевление стоимости беспилотников и дронов, а их производство становится массовым. Использование искусственного интеллекта поможет объединять тысячи дронов в огромный управляемый «рой», способный к массовой атаке.

Искусственный интеллект также может быть включен в технологии государственного управления и укрепления власти и стать инструментом внутренней политики. Также он будет

помощником государственных органов при управлении экологическими рисками и предотвращении техногенных катастроф.

Прогресс в создании искусственного интеллекта окажет мощнейшее влияние на экономику и может привести к новой промышленной революции. Держава, которая первой его внедрит, обретет экономическое, информационное и возможно, военно-политическое превосходство над остальными государствами[1].

Известный бизнесмен Илон Маск заявил, что искусственный интеллект в конечном итоге уничтожит человечество, поэтому он и еще 116 экспертов, ученых и представителей компаний сектора новых технологий направили ходатайство в адрес ООН, в котором содержится призыв запретить разработку любых видов вооружений и автономных технологий, использующих искусственный интеллект. Группа бизнесменов и ученых утверждает, что введение автономных технологий будет равносильно «третьей революции в войне» после появления пороха и ядерного оружия, и в этом они, несомненно, правы. Но также очевидно, что начавшееся в ООН обсуждение конвенции о запрещении такого оружия не что иное, как попытка американцев и их союзников заблокировать, прикрываясь рассуждениями об общечеловеческих ценностях, создание их геополитическими соперниками, в первую очередь Россией и Китаем вооружений с искусственным интеллектом.

Развитие искусственного интеллекта становится стратегической задачей для сверхдержав в XXI веке. Крайне важно при этом ответить на вопрос: кого мы вырастим себе в помощники — циничного и бесчеловечного искусственного «Мефистофеля» или электронного «Ангела-хранителя»? Далее с опорой на достижения современных математиков попытаемся ответить на данный вопрос.

Мифы и угрозы искусственного интеллекта и современные реалии

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

---

[1] Лосев А. Военный искусственный интеллект//Ж-л «Арсенал отечества» - 2017 - №6(32)

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[2]. В поисках улучшений, которые могут помочь в краткосрочном периоде, 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 психологические обязательства по инвестированию в тот или иной подход. А подход, основанный на знаниях человека, имеет тенденцию усложнять методы таким образом, что они становятся менее подходящими для использования преимуществ общих методов, использующих 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.*

Было много примеров, когда исследователи ИИ запоздало понимали этот горький урок. Рассмотреть некоторые из таких случаев. В компьютерных шахматах методы, победившие чемпиона мира Каспарова в 1997 году, основывались на массивном глубоком поиске. В то время к ним с тревогой относились большинство исследователей компьютерных шахмат, которые использовали методы, основанные на понимании человеком особой структуры шахмат. Когда более простой, основанный на поиске 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 это он не станет общей стратегией, и уж точно люди не играют в шахматы таким образом.» Эти ученые хотели, чтобы методы, основанные на человеческом способе мышления, победили, и очень разочаровались, когда этого не произошло.

---

[2] *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).*

Вывод. *Прямое решение проблемы с помощью компьютерной вычислительной мощности возьмет свое, рано или поздно.*

Аналогичная картина прогресса в исследованиях была замечена в игре ГО, превосходящей на порядок по сложности шахматы, но только с задержкой на 20 лет. 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 (*англ., Big data*) 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, только 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 техник, которые задействуют огромные объемы вычислений в исследованиях ИИ.

В компьютерном противоборстве с человеком по игре в ГО[3], 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.

Вывод. *Поиск и машинное обучение, подпитанные вычислительной мощностью, намного превосходят попытки решить задачу «нестандартным подходом мышления».*

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 обучения в области распознавания

---

[3] *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*

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 все более мощные компьютеры. В следствии чего, проблема решалась совершенно на другом уровне.

Вывод: *нельзя входить в одну и тоже реку дважды, учиться нужно на чужих, а не на своих ошибках.*

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 *нашими 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.

Вывод. *Нужно опираться на масштабируемые вычислениям и поиск, 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.*

Миф2*: В результате экспоненциального роста производительности компьютеров наступит время «технологической сингулярности», когда вычислительная 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.

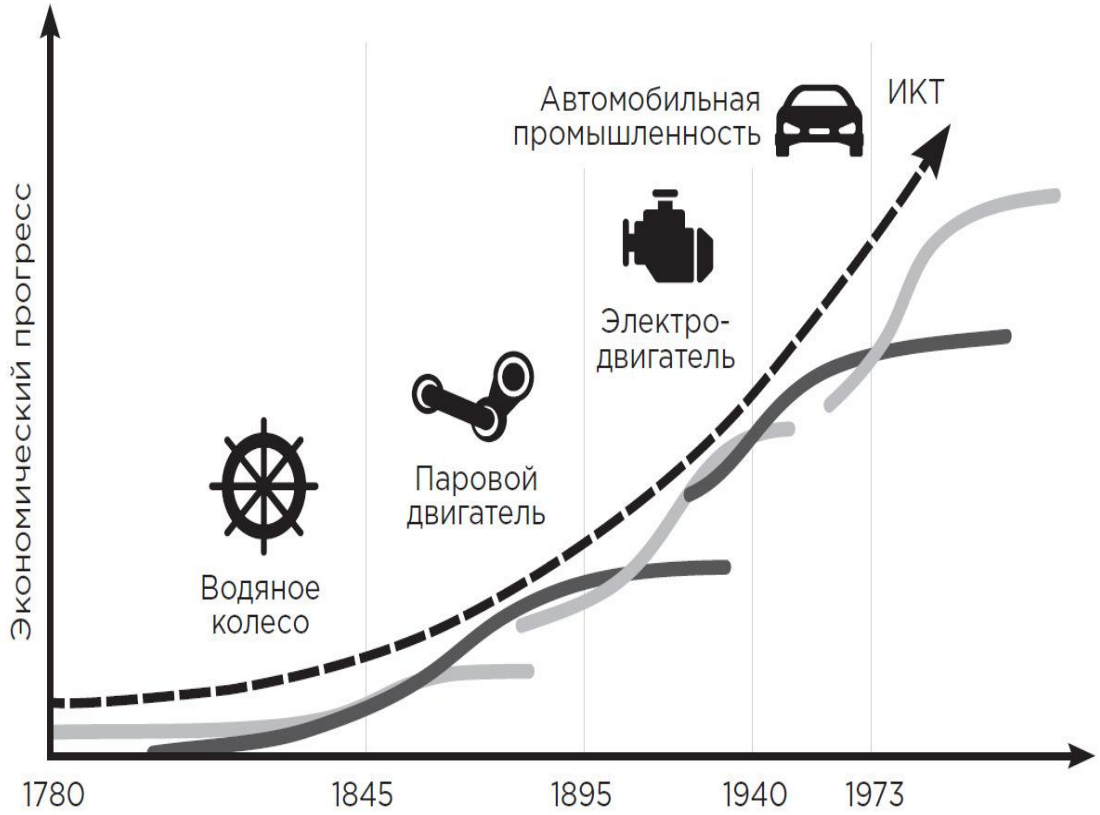

Рис. 1. Трансформирующие технологии и технологические уклады, (источник: M. Hilbert, University of California).

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, здравоохранение, электро- и самоуправляемые автомобили, образование, 3D-печать, промышленность и сельское хозяйство.

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, то сингулярность произойдет из-за прогресса в области искусственного интеллекта, интеграции человека с ИКТ или других методов увеличения разума.[4] 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

[4] *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).*

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[5] и того же Винжа, придерживающихся концепции сингулярности, она должна наступить около 2030 года 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 не открытых 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.*

Миф3. *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»[6]. Новые технологии формируют новые формы антиутопии. И общество пока не понимает, как адаптироваться к меняющейся реальности.

Харари вывел формулу предстоящего глобального кризиса**:**

---

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

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

$$B * C * D = HH$$

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 правительства 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.

Общество подвержено взлому на всех уровнях, но больше всего Харари пугает биологический: «Эксперты по ИИ могут общаться с философами. С историками – да, пожалуйста. С литературными критиками – замечательно. Но меня пугает их общение с биологами», – признал он в интервью изданию Wired. Тем не менее, исследователь подчеркивает, что ИИ обладает и массой преимуществ. Особенно это касается медицины. Харари подчеркивает, что никто не станет препятствовать внедрению технологии – ведь она способна принести столько пользы людям.

Распространение ИИ в комплекте с биотехнологическими открытиями породит два возможных сценария антиутопии.

Первый: *надзор-капитализм* – даст алгоритмам полную власть над людьми. Машинный интеллект и МО решат за нас, где жить, работать, с кем встречаться и за кого голосовать.

Второй: укрепление *тоталитаризма и диктатуры*, при котором каждый житель Земли – это объект непрерывной слежки. Особую роль в этом процессе сыграют

биометрические и видео-системы, которые не дадут гражданину скрыться от всевидящего ока государства.

Историк подчеркивает, люди могут даже не заметить, как оказались во власти ИИ и МО. Большинство не сможет понять, как работают механизмы алгоритмов и как именно нами манипулируют. Человечество привыкло к традиционным формам объяснения и повествования, но машинный интеллект работает со статистическими данными и оперирует другими понятиями.

Харари считает, что *чрезмерное усложнение систем – одна из главных актуальных проблем*. Из-за этого, например, ученым все сложнее объяснять свои теории и доносить до аудитории суть открытий.

Важный побочный эффект этого – расцвет теорий заговора. По этой причине сейчас возникает все больше антиглобалистов и тех, кто не верит в глобальное потепление. То же касается и сферы финансов – с каждым годом она усложняется, и некоторые концепции можно объяснить, только если потратить 10 лет на изучение экономики и математики. «В этом тоже выражается философский кризис», – отмечает Харари.

Он также считает, что сегодня человек борется не с отдельными людьми, а с государствами и корпорациями. Перед лицом таких мощных соперников шансов на успех мало. Более того, влияние некоторых стран выходит за географические рамки. Историк обвиняет развитые государства и крупные корпорации в *цифровом колониализме*.

Вывод. *По мнению некоторых учёных-футурологов, после наступления технологической сингулярности, человечество ожидает технологический и философский кризис. Социум погрузится в эпоху цифрового колониализма.*

Нарисованная выше учёными-футурологами довольно пессимистическая картина мира, после достижения человеческой цивилизацией временной точки технологической сингулярности, скрашивается последними исследованиями учёных-математиков: *возможности ИИ оказались небеспредельными*[7]. Подобно человеческому разуму, ИИ ограничен парадоксами теории множеств.

До сих пор считалось, что самой фундаментальной проблемой развитии технологий ИИ является *необъяснимость* принимаемых им решений.

---

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

В январе 2019 к этой проблеме добавилась еще одна фундаментальная проблема – это принципиальная *непредсказуемость*, какие задачи ИИ может решить, а какие нет.

На пути триумфального развития технологий машинного обучения, как казалось, способных при наличии большого объема данных превзойти людей в чем угодно – в играх, распознавании, предсказаниях и т.д. – встала первая из 23 проблем, поставленных в докладе Давида Гильберта на международном математическом конгрессе в Париже еще в 1900-м году[8].

Первой в списке этих 23 проблем, решение которых до сих пор считается высшим достижением для математика, была так называемая *гипотеза континуума (континуум-гипотеза или 1-я проблема Гильберта)*, которую выдвинул и пытался решить (но потерпел неудачу) еще сам создатель теории множеств Георг Кантор.

И вот сейчас, на исходе второго десятилетия XXI века гипотеза континуума, будучи примененная к задачам машинного обучения, стала холодным отрезвляющим душем для всех техно-оптимистов ИИ.

Машинное обучение оказалось не всесильно. И что еще хуже,– в широком спектре сценариев обучаемость ИИ не может быть ни доказана, ни опровергнута.

Первая же научная сенсация 2019 года оказалась совершенно неожиданной. Опубликованная 7-го января того же года в *Nature Machine Intelligence* статья «Обучаемость может быть неразрешимой» (*англ., Learnability can be undecidable)* устанавливает предел возможностей машинного обучения – ключевого метода вычислений, на коем стоит весь современный ИИ[9].

Этот научный вывод столь важен, что журнал Nature сопроводил статью еще двумя популярно её разъясняющими статьями «Недоказуемость приходит в машинное обучение» (*англ., Unprovability comes to machine learning*) и «Машинное обучение приводит математиков к неразрешимой задаче» (*англ.*, *Machine learning leads mathematicians to unsolvable problem*).

Суть всех этих статей в следующем. Обнаружены сценарии, в которых невозможно доказать, может ли алгоритм машинного обучения решить конкретную проблему. Этот вывод может иметь огромное значение, как для существующих, так и для будущих алгоритмов

---

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

[9] *Ben-David S., Hrubeš P., Moran S., Shpilka A., Yehudayoff A. Learnability can be undecidable. Nature Machine Intelligence, 1, pp. 44-48.*

обучения. Обучаемость ИИ не может быть ни доказана, ни опровергнута с использованием стандартных аксиом математики, поскольку это 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[10].

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 и защищённости 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, 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[11].

---

[10] *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).*

[11] *Артамонов, В. А. Безопасность информационно – коммуникационных технологий в контексте устойчивого развития социума / В. А. Артамонов, Е. В. Артамонова, Л. А. Кулак //Цифровая трансформация. – 2019. -- №. – С.36 - 45.*

Однако, вместе с этим, технологии искусственного интеллекта рассматриваются как одно из самых действенных средств в области кибербезопасности сейчас и в будущем.

Почему ИИ — это будущее кибербезопасности[12].

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 компаний 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 банкиров. 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.

---

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