Искусственный интеллект в системах безопасности

В статье рассмотрены вопросы применения искусственного интеллекта (ИИ) в качестве эффективного механизма обеспечения как информационной безопасности, так и безопасности других сфер человеческой деятельности. Показано, что главным механизмом настройки ИИ на решение конкретной задачи является машинное обучение. Приведена динамика роста инвестиций в рассматриваемую область и прогнозные показатели ее будущего развития. Дана классификация продуктов обеспечения информационной безопасности с применением ИИ по функциональному и технологическому типу и по сценариям использования. Рассмотрены задачи слабого ИИ в обеспечении информационной безопасности в некоторых смежных областях.

Ключевые слова: искусственный интеллект, машинное обучение, киберустойчивость, криптография, стеганография, криминалистика, фреймворки, платформы обеспечения информационной безопасности

Владимир Афанасьевич Артамонов,
доктор технических наук, профессор, академик МАИТ
artamonov@itzashita.ru

Елена Владимировна Артамонова,
кандидат технических наук, член МАИТ
admin@itzashita.ru

Международная академия информационных технологий (МАИТ)

Введение

Искусственный интеллект (ИИ) (*artificial intelligence*, AI) является трендом XXI века в области информационно-коммуникационных технологий (ИКТ). Также не вызывает сомнения тот факт, что прогресс в области ИКТ и Интернета породил новые вызовы и угрозы для человечества в виде кибертерроризма и киберпреступности с одной стороны и кибер- и информационных войн на уровне государств – с другой.

Можно выделить три вида искусственного интеллекта:

- ИИ узкого назначения (слабый – *Narrow AI*),
- ИИ общего назначения (сильный – *General AI*),
- суперИИ (*Super AI*) [1].

В их основе лежит общий когнитивный принцип, но, тем не менее, они достаточно сильно отличаются друг от друга.

Начнем с общего определения этой технологии. ИИ – это обширный раздел математики и компьютерных наук, цель которого – создать видимость наличия у машины человеческого интеллекта. Таким образом, если компьютер демонстрирует когнитивные способности, присущие людям, то мы называем это искусственным интеллектом.

Слабый ИИ (Narrow AI)

ИИ узкого назначения, также известный как *слабый*, – это ИИ в сегодняшнем понимании [1]. Он запрограммирован на выполнение од-

En An Artificial Intelligence in the Security Systems

V. A. Artamonov,
PhD (Eng., Grand Doctor), Full Professor, the Full Member of IAIT
artamonov@itzashita.ru

E. V. Artamonova,
PhD (Eng.), the Member of IAIT
admin@itzashita.ru

The International Academy of Information Technologies (IAIT)

This paper discusses the use of artificial intelligence (AI) as an effective mechanism for ensuring information security, as well as security in other areas of human activity. Shown, that the main mechanism for setting AI to solve a specific problem is machine learning. The dynamics of investment growth in the sector under consideration and forecast indicators of its future development are given The classification of information security products with the use of AI is given by functional and technological type and by usage scenarios. The tasks of weak AI in providing information security in some related areas are considered.

Keywords: artificial intelligence, machine learning, cyber resilience, cryptography, steganography, criminalistics, frameworks, information security platforms

ной задачи, пусть и очень высокоразмерной: будь то мониторинг погоды, игра в шахматы, анализ угроз информационной безопасности (ИБ) либо управление оружием. Такой ИИ может работать в режиме реального времени, но он извлекает информацию лишь из ограниченного набора данных, поэтому и способен справляться только с одной конкретной проблемой, решению которой он обучен. В отличие от ИИ общего назначения, также известного как *сильный*, ИИ узкого назначения не обладает человеческими чувствами и сознанием, а работает лишь в заранее заданном диапазоне.

Все окружающие нас сегодня ИИ-решения – это примеры слабого ИИ, в том числе такие, как Google Assistant, Google Translate, Siri и другие инструменты обработки естественного языка. Кто-то может возразить и сказать, что это не так, ведь они способны взаимодействовать с людьми. И все же подобные машины слишком далеки от человеческого интеллекта, другими словами, они не могут думать самостоятельно. Например, Siri не обладает сознанием, а просто выполняет ряд возложенных на нее задач: обрабатывает человеческую речь, вводит полученный вопрос в поисковую систему и выдает ответ.

Немаловажным является и тот факт, что проблемы создания ИИ классов *General AI* и *Super AI* при современном развитии компьютерных мощностей и тайн функционирования человеческого мозга носят скорее философский и постановочный характер.

Перед человечеством стоит сложнейшая задача, сравнимая по экстремальности с открытиями XX века в ядерной физике и созданием атомного оружия. Искусственный интеллект как сущность пока не создан ни в одной стране мира. Есть лишь математические и компьютерные модели, имитирующие работу биологических нейронов. Но, скорее всего, создателям ИИ придется обратиться к квантовой физике. Частицы, которые содержат информацию и участвуют в передаче сигнала – это молекулы и атомы, но при всей их микроскопичности взаимодействие получается настолько значительным, что людям, возможно, придется создавать квантовую теорию мозга и использовать для создания ИИ квантовые компьютеры, что само по себе – масштабная и очень сложная задача.

Пока человечество очень далеко от создания полноценного ИИ, потому что мы не до конца понимаем, как работают синапсы и передаются мозговые сигналы, как возникают воспоминания. Необходимы математические и физические модели сознания. Впереди у человечества много работы, и чем качественнее она будет сделана, тем больше шансов у людей дожить до точки технологической сингулярности[1] в развитии социума.

Задачи слабого ИИ в обеспечении ИБ и в некоторых смежных областях

Хотя этот вид ИИ называется слабым, не стоит воспринимать такое определение буквально. ИИ узкого назначения – это серьезный технологический прорыв. Такие системы способны обрабатывать данные и выполнять задачи значительно быстрее, чем любой человек, что позволяет повысить общую производительность и безопасность труда, а также качество жизни. В частности, в области ИБ технологии ИИ предоставляют возможность создавать решения существенно более высокой эффективности, позволяющие:

- идентифицировать кибератаки с высокой скоростью;
- в режиме реального времени выбирать оптимальный ответ на инциденты безопасности;
- в автоматическом режиме проводить оценку актуальности и последствий инцидентов.

На текущий момент времени количество атак продолжает растти, а ландшафт угроз меняется с молниеносной скоростью. Например, решения «Лаборатории Касперского» за 2021 год отразили 687 861 449 атак, которые проводились с интернет-ресурсов, размещенных в различных странах мира[2]. Очевидно, что при таких объемах вредоносной деятельности злоумышленники активно применяют средства автоматизации кибератак, в том числе используют технологии ИИ и машинного обучения (МО) для их совершенствования и трансформации, а также для обхода известных средств защиты.

Так, эффективным примером применения МО во вредоносном программном обеспечении является известный троян Emotet. Основным каналом для его распространения является спам-фишинг, и группировка, стоящая за созданием Emotet, могла без особых затруднений использовать ИИ для усиления атаки на целевые системы. Это осуществлялось следующим образом: вредоносный код встраивался нативно в цепочки разговоров и использовал анализ текста на естественном языке. Другой возможной сферой вредоносного применения ИИ является более эффективный подбор паролей или обход двухфакторной аутентификации. Еще два года назад исследователи создали бот, способный обходить проверки CAPTCHA[3] с 90-процентной эффективностью с помощью технологий ИИ.

Используя огромное количество различных источников данных в даркнете для формирования базы знаний искусственного разума, злоумышленники могут сделать атаки на человека по-настоящему действенными. По данным компании Webroot[4], около 85 % профессионалов в области информационной безопасности верят в то, что злоумышлен-

[1] *Технологическая сингулярность – это гипотетический момент в будущем, когда технологическое развитие станет настолько стремительным, что экспоненциальный график технического прогресса человечества станет практически вертикальным.*

[2] *https://securelist.ru/kaspersky-security-bulletin-2021-statistics/104160/.*

[3] *Капча (от CAPTCHA – Completely Automated Public Turing test to tell Computers and Humans Apart – полностью автоматизированный публичный тест Тьюринга для различения компьютеров и людей).*

[4] *Webroot Inc. – американская частная компания-разработчик программного обеспечения для кибербезопасности.*

ники используют технологии ИИ в своих атаках. Для того чтобы справиться с растущим объемом атак, производители систем защиты тоже начинают активно внедрять технологии ИИ, машинного и глубокого обучения (*machine learning/deep learning*) для прогнозирования и обнаружения киберугроз, а также реагирования на них в режиме реального времени.

В 2019 году мировой рынок технологий искусственного интеллекта в сфере информационной безопасности оценивался экспертами MarketsandMarkets[5] в 8 млрд долл., с ежегодным увеличением на 23 % и возрасанием до 30 млрд долл. в 2025 году [2], а согласно прогнозу компании Accenture сегмент защитных ИИ-решений, его размер в денежном выражении может составить 46 млрд долл. к 2027 году[6]. Организации, внедряющие технологии ИИ для поведенческого анализа[7] и предиктивной аналитики, получают ощутимые результаты в виде повышения эффективности обнаружения атак и затрат на организацию безопасности, а также на сокращение времени реагирования на инциденты. По данным Capgemini Research Institute, 64 % организаций, годовая выручка которых составляет более 1 млрд долл., заявляют о том, что технологии искусственного интеллекта сокращают затраты на обнаружение и реагирование на угрозы безопасности, и около 75 % сообщают о сокращении времени реагирования (до 12 %). Для исследования из публичных аналитических отчетов, упоминаний в прессе и открытых баз (Crunchbase, Owler, Angel List) было отобрано 66 вендоров, которые активно применяют технологии поведенческого анализа и предиктивную аналитику в своих продуктах по информационной безопасности. Участники выбирались по принципу глобальной доступности, локальные игроки (то есть работающие только в своей стране или регионе), как и местные рынки в целом, в исследовании не учитывались. Существует много предприятий, имеющих продукты в разработке; они также не попадают в классификацию, но учитываются в инвестиционном анализе, о них мы поговорим ниже.

Пик появления стартапов в рассматриваемой области пришелся на 2014–2015 годы, когда появились базовые библиотеки, а технологии ИИ начали активно применяться в прикладных системах. Статистика инвестиционной активности показывает, что существуют две основные волны применения технологий.

Компании первой волны находятся на позднем этапе венчурных инвестиций, в основном, их разработки относятся к типу EDR (*Endpoint Detection and Response*) – платформы обнаружения атак на рабочих станциях, серверах, любых компьютерных устройствах (конечных точках) и оперативного реагирования на них. Предприятия второй волны специализируются на предикативной аналитике[8] агрегации внешних данных и пока находятся на предстартовом этапе или в ранних раундах венчурных инвестиций. Классификация продуктов с технологиями ИИ по сценариям применения возможна по двум направлениям: по функциональному и технологическому типу и по сценариям использования.

Перечислим основные типы.

EDR *(Endpoint Detection and Response)*. С помощью технологий ИИ продукты данной категории могут обнаруживать неизвестные вредоносные программы, автоматически классифицировать угрозы и самостоятельно реагировать на них, передавая данные в центр управления. ИИ принимает решения на основе общей базы знаний, накопленной путем сбора данных со множества устройств. Некоторые продукты данного типа используют технологии ИИ для разметки данных на конечных точках и дальнейшего контроля их перемещения, чтобы выявлять внутренние угрозы.

NDR *(Network Detection and Response)* – устройства и аналитические платформы, которые, используя накопленную статистику и базу знаний об угрозах, с помощью технологий ИИ выявляют угрозы в сетевом трафике и могут автоматически на них реагировать надлежащим образом, изменяя конфигурацию сетевых устройств и шлюзов. Часть продуктов данного типа специализируется на защите облачных провайдеров и их инфраструктуры. Дополнительный сценарий использования ИИ в сетевой защите – это анализ почтового трафика на предмет фишинга.

UEBA *(User and Entity Behavior Analytics)* – системы поведенческого анализа пользователей и информационных сущностей. Основной сценарий применения ИИ-технологий в продуктах типа UEBA – это автоматическое выявление аномалий в поведенческих моделях (отклонение от нормы или соответствие шаблону (паттерну) угрозы) для пользователей и различных сущностей информационных систем. Выявленные аномалии классифицируются с помощью ИИ, как различные угрозы и риски для бизнеса. Аномальное поведение может выявляться в целях мониторинга и управления доступом, обнаружения мошенничества среди клиентов или сотрудников (антифрод), защиты конфиденциальных данных, проверки соблюдения тех или иных регламентов и нормативных актов.

TIP *(Threat Intelligence Platform)* – платформы раннего детектирования угроз и реагирования на них, действующие на основе большого количества различных данных (Da-

[5] *MarketsandMarkets является крупнейшей исследовательской компанией в мире с точки зрения научных докладов премиум-класса, опубликованных в год. Публикует более чем 1200 качественных исследований в год.*

[6] *https://3dnews.ru/1059516/obyom-rinka-informatsionnoy-bezopasnosti-na-baze-ii-previsit-45-mlrd-k-2027-godu/.*

[7] *Поведенческий анализ – это научное направление, основанное на базе бихевиористских принципов. Своей целью поведенческий анализ выдвигает выявление нежелательных форм поведения и разработку методов коррекции этих форм в различных проявлениях.*

[8] *Предикативная (или предиктивная, прогнозная аналитика) – это прогнозирование, основанное на исторических данных. С помощью статистических инструментов можно выявить закономерности в изменении тех или иных показателей за предыдущие периоды и предсказать, как эти показатели будут вести себя в будущем.*

ta Lake) и индикаторов компрометации (IoC). Применение ИИ позволяет повысить эффективность выявления неизвестных угроз на ранних этапах; сценарий очень схож с работой SIEM-систем, но нацелен на внешние источники данных и внешние угрозы.

SIEM *(Security Information and Event Management)* – решения, которые осуществляют мониторинг информационных систем, в режиме реального времени анализируют события безопасности, поступающие от сетевых устройств, средств защиты информации, ИТ-сервисов, инфраструктуры систем и приложений и помогают обнаружить инциденты ИБ. В системах такого класса накапливается огромное количество данных из различных источников, а применение технологий ИИ дает возможность выявления аномалий эвристическими методами и сокращения ложных срабатываний при изменении паттернов и моделей данных. Применение ИИ в SIEM-системах позволяет достигнуть очень высокого уровня автоматизации.

SOAR *(Security Orchestration and Automated Response)*. В решениях данного типа, в отличие от SIEM, ИИ помогает не только проводить анализ угрозы информационной безопасности, но и автоматически реагировать на них надлежащим образом.

AS *(Application Security)* – системы, позволяющие определять угрозы безопасности прикладных приложений, управлять дальнейшим циклом мониторинга и устранения таких угроз. Основной сценарий применения технологий ИИ в системах защиты прикладных приложений – автоматический сбор информации об уязвимостях, атаках и заражениях, доступной в открытых источниках, и основанная на его результатах автоматизация защитных действий: сканирование на уязвимости, изменениях правил защиты для web-приложений, выявление угроз и изменение рисковой модели.

AF *Антифрод (Antifraud)* – системы, позволяющие выявлять угрозы в бизнес-процессах и предотвращать мошеннические операции в режиме реального времени. В системах защиты от мошенничества технологии ИИ применяются для определения отклонений от установленных бизнес-процессов, тем самым помогая быстро реагировать на возможное финансовое преступление или уязвимость процессов.

ZTNA *(Zero Trust Network Access)* – системы доступа с «нулевым доверием», которые с использованием ИИ и МО предлагают более эффективное решение для удаленного доступа, также решающего 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.

DLP *(Data Loss Prevention)* – системы предотвращения потери/утечки данных. DLP-системы строятся на анализе потоков данных, пересекающих периметр защищаемой информационной системы. При детектировании в этом потоке конфиденциальной информации срабатывает активный компонент системы, и передача сообщения (пакета, потока, сессии) блокируется. В основе «интеллектуальных» функций DLP лежит технология так называемого контентного анализа, включающая ИИ и МО на основе алгоритма Байеса и метода опорных векторов.

Применение ИИ в ИБ-системах позволяет быстро адаптироваться к изменению логики и различных метрик бизнес-процессов, а также использовать лучшие практики в индустрии ИТ. Для всех типов таких систем технологии искусственного интеллекта позволяют увеличить эффективность детектирования неизвестных угроз.

По данным Osterman Research[9], большинство компаний, которые начали использовать продукты с технологиями ИИ, отмечают повышение эффективности в расследовании ИБ-инцидентов, управлении персоналом, обнаружении угроз и скорости реакции на таковые. Многие также обращают внимание на сокращение количества ложных срабатываний. По данным Ponemon Institute[10], около 60 % ИБ-специалистов считают, что использование технологий ИИ в информационной безопасности повышает скорость анализа и обнаружения угроз на конечных точках и в приложениях.

Применение искусственного интеллекта для решения задач криптологии

Первые идеи применять машинное обучение в криптологии появились еще в 90-е годы XX века. Так, в далеком 1991 году на Международной конференции по теории и приложениям в криптологии Р. Ривестом был сделан доклад на тему «Криптография и машинное обучение» (*Cryptography and machine learning*), в котором было предложено использовать МО на основе искусственных нейронных сетей (ИНС) для решения некоторых криптографических задач.

Важной практической задачей, решаемой с помощью ИНС, является классификация. В связи с этим значительное число публикаций посвящено применению нейронных сетей для построения систем идентификации и аутентификации, использующих различные биометрические характеристики человека [3, 4, 5, 6].

ИНС, благодаря своей архитектуре, представляют собой эффективный способ параллельной обработки

[9] *Osterman Research опубликовала результаты проверки использования открытых компонентов с неисправленными уязвимостями в готовом проприетарном программном обеспечении (COTS). В исследовании были изучены пять категорий приложений – web-браузеры, почтовые клиенты, программы для обмена файлами, мессенджеры и платформы для проведения онлайн-встреч.*

[10] *Организация Ponemon Institute предоставляет компетентную информацию и методы по управлению конфиденциальностью в сфере бизнеса и управления. Для достижения этой цели Ponemon Institute проводит независимые исследования, обучает управляющих из частных и государственных секторов, а также проверяет методы обеспечения конфиденциальности и защиты информации организаций, работающих в различных отраслях.*

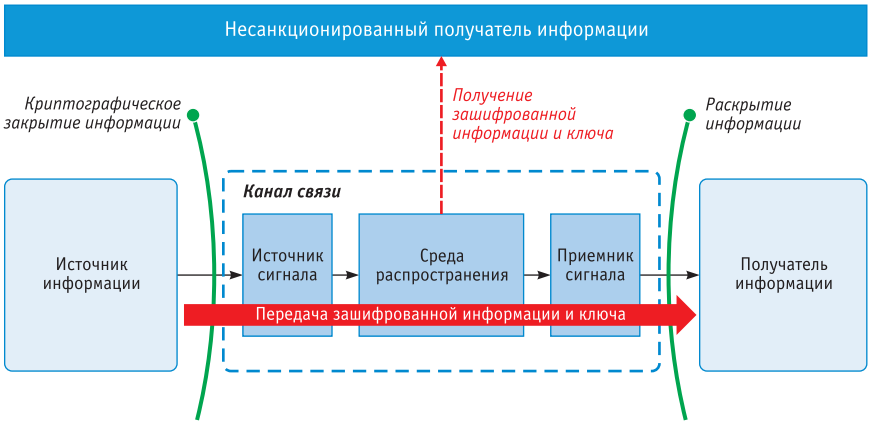

Рис. 1. ***Стандартный подход передачи ключа в симметричном шифровании***

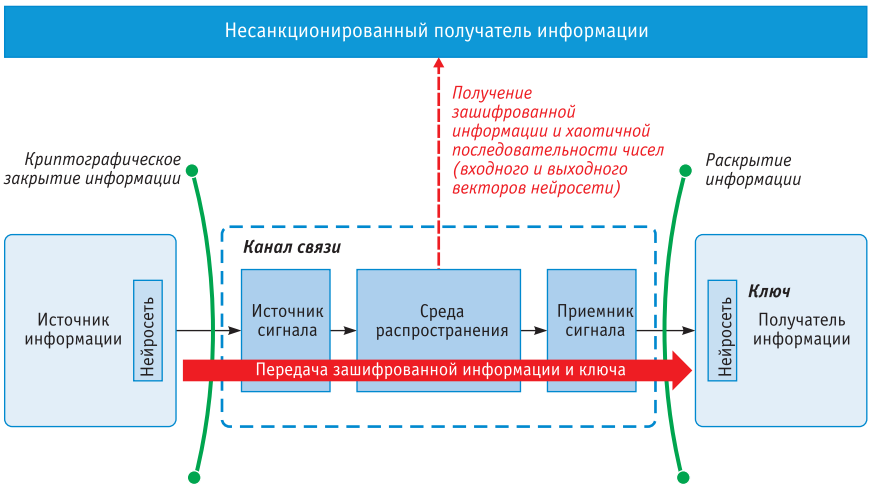

Рис. 2. ***Передача ключа в симметричном шифровании с помощью нейронных сетей***

данных. Данное свойство позволяет в перспективе применять их для широкого круга задач, однако часто сложность реализации такой системы сводит на нет ее практическую применимость. Несмотря на данный факт ИНС весьма широко применяются для защиты информации, как в рамках научно-исследовательских проектов, так и в коммерческих продуктах.

В основе криптографических нейронных систем лежат те же подходы к использованию ключей, что и в системах шифрования вообще: в системе с одним секретным ключом последний известен только абонентам системы (симметричная система шифрования), в другом подходе используются два ключа – секретный для дешифрования и открытый для шифрования (асимметричная система шифрования). Симметрические сети отличаются быстрой производительностью, но требуют дополнительных мер безопасности, в асимметричных – используются однородные данные, что обеспечивает высокую криптостойкость и хорошо подходит для компьютерной реализации [7, 8].

Модель ИНС идеально подходит для задач шифрования. Рассмотрим таковую на примере алгоритма AES. AES (*Advanced Encryption Standard* или *Rijndael*) является итеративным блочным шифром, который был выбран NIST (Национальным институтом стандартизации США) в качестве национального стандарта шифрования и замены алгоритма DES. В настоящее время это наиболее широко распространенный в мире итеративный блочный шифр. Шифрование или расшифрование блока данных осуществляется с помощью раундовой функции. Три стандартные версии AES – AES-128, AES-192 и AES-256 – отличаются друг от друга длиной ключа (128, 192, и 256 бит) и количеством раундов (10, 12 и 14). Их безопасность была тщательно исследована АНБ США. Тем не менее, в последнее время появились некоторые виды атак, которые, как было доказано, достаточно эффективно применяются к AES. Чтобы противостоять этим атакам, предложено сделать некоторые изменения, внедрив нелинейную нейронную сеть в AES. В результате такой трансформации появилась более киберустойчивая криптосистема [9].

Основным недостатком симметричного шифрования является общий ключ для шифрования и дешифрования, что затрудняет использование этой схемы в распределенных системах. Проблема вызвана тем, что ключ передается посредством общедоступного канала связи, через который злоумышленник может похитить как саму информацию, так и ключ для ее дешифрования. В классическом виде стандартный вариантпередачи ключа в симметричном шифровании выглядит, как это изображено на рис. 1.

Нейронные сети (рис. 2) используются для классификации и аппроксимации функций или выделения задач, устойчивых к некоторым неточностям, для которых имеется много доступных данных для обучения, но к которым не могут быть применены жесткие правила.

При реализации Rijndael с помощью ИНС криптосистема имеет менее сложное строение, чем AES, и она нелинейная в эксплуатации. Нелинейной должна быть нейронная сеть с обратной связью, что позволило бы выполнить шифрование открытого текста или расшифровывание зашифрованного текста с высокой производительностью и очень низким уровнем ошибок. Нелинейность необходима для уменьшения вероятности взлома алгоритма, что достигается с помощью нелинейной функции активации.

Еще одним из направлений использования ИНС для защиты информации является стеганография [10]. В этой области в последние годы получили широкое распространение методы МО, в частности, генеративно-состязательные сети. Генеративно-состязательная сеть (*Generative adversarial network*, сокращенно GAN) –

алгоритм машинного обучения без учителя, построенный на комбинации из двух нейронных сетей. В 2014 году ее впервые представил Google. В системе GAN одна из сетей генерирует образцы, а другая старается отличить «подлинные» образцы от «неправильных».

GAN, используя набор переменных латентного пространства, всегда пытается слепить новый образец, смешав несколько исходных образцов. Дискриминиативная сеть *D* обучается различать подлинные и поддельные образцы. При этом ее результаты подаются на вход генеративной сети *G* так, чтобы она смогла подобрать лучший набор латентных параметров, и *D-сеть* уже не смогла бы отличить подлинные образцы от поддельных. Как видим, целью сети *G* является повысить процент ошибок сети *D*, а цель сети *D* заключается, наоборот, в улучшении точности распознавания.

Возможности GAN можно рассматривать с разных сторон: состязательная игра, генератор или функция отображения. Они согласуются с классификацией основных стратегий в стеганографии, то есть *модификации, селекции и синтеза.*

Модификация изображения, основанная на GAN, фокусируется в состязательной игре между стеганографом и стеганоаналитиком. Данный метод использует генератор, обученный для построения различных «ключевых» элементов. Выделяют три основные стратегии:

- создание стегоизображния;
- создание матрицы вероятности модификации;
- состязательная игра.

Стратегия создания стегоизображения зиждется на использовании GAN, состоящей из трех сетей: генератора *G*, дискриминатора *D* и классификатора *S*, который определяет, скрывает ли реалистичное изображение секретную информацию или нет.

Матрица вероятности модификации служит для автоматического изучения функции искажения. Эта схема действует в традициях современной стеганографии и минимизирует функцию аддитивного искажения. Сама матрица вероятностей изменения получается путем минимизации математического ожидания функции искажения. Генератор *G* в данной схеме обучается узнавать вероятности изменения *P* входного изображения.

Состязательная игра служит для изучения и испытания алгоритма модификации. Данный метод пытается изменить изображение для скрытия сообщения, обманывая классификатор стеганоанализа. В этом подходе пиксели «изображения-кандидата» делятся на две группы, одна из которых используется для модификации, а вторая – для создания возмущений с целью противостоять стеганоаналитику.

Селекция изображения. Стеганография на основе GAN-селекции направлена на установление взаимосвязи между сообщением и стегоизображением. Одна из идей состоит в том, что отправитель устанавливает вид отображения, используя генератор между сообщением и выбранным изображением. Для получателя сообщение создается непосредственно из выбранного изображения. Суть этого метода заключается в установлении взаимосвязи между изображением и секретным сообщением, чтобы контейнер естественным образом превратился в стегоизображение. Статистический стеганографический анализ не работает, потому что непосредственно сама операция модификации не осуществляется.

Синтезирование изображения. Изображение обычно создается с помощью предобученного генератора, и здесь предлагается несколько стратегий:

- обучение с учителем;
- обучение без учителя;
- обучение с частичным привлечением учителя.

Искусственный интеллект и машинное обучение в криминалистике

С развитием технологий ИИ и МО меняются и технологии работы с данными и большими данными. В компьютерной криминалистике ключевым вопросом становится максимальная автоматизация процессов анализа больших объемов доказательств и формирование «подсказок» для экспертов на основе аналогичных прецедентов для более быстрого расследования преступлений и минимизации влияния человеческого фактора на результат. Анализ, создание новых страниц в деле, контентом для которых является подборка со схожим набором параметров, визуализация связей, формирование наиболее вероятных оценок – то, к чему идут современные системы сбора и анализа доказательной базы. Экспертам-криминалистам приходится обрабатывать огромное количество данных, работать с новыми, сложными типами информации, получаемыми из разнообразных источников. По мере увеличения объемов расследований и необходимости соблюдения сроков их проведения, достижение необходимых результатов становится все более трудновыполнимой задачей.

В опросе, проведенном компанией Access Data [11] среди руководителей самых разных организаций, было установлено, что наиболее значимые проблемы – это время, стоимость и эффективность при управлении процессом расследования инцидентов и правонарушений во всей организации.

Результаты опроса среди показали:

- для 47 % организаций критичны сложности в области интеграции технологий для проведения расследований;
- в 51 % из них считают критичным отсутствие доступа к данным;
- для 57 % критично время/стоимость проводимых расследований.

Меняющийся характер следственных задач приводит к разработке новых технологий. Мы видим это в новых интегрированных исследовательских системах, которые предлагают более быстрые возможности обработки, мощные базы данных, высокую масштабируемость и более тесную интеграцию команды для эффективного сотрудничества. Но будущее автоматизированных цифровых криминалистических инструментов в действительности лежит в машинном обучении и искусственном интеллекте. Этот тип техноло-

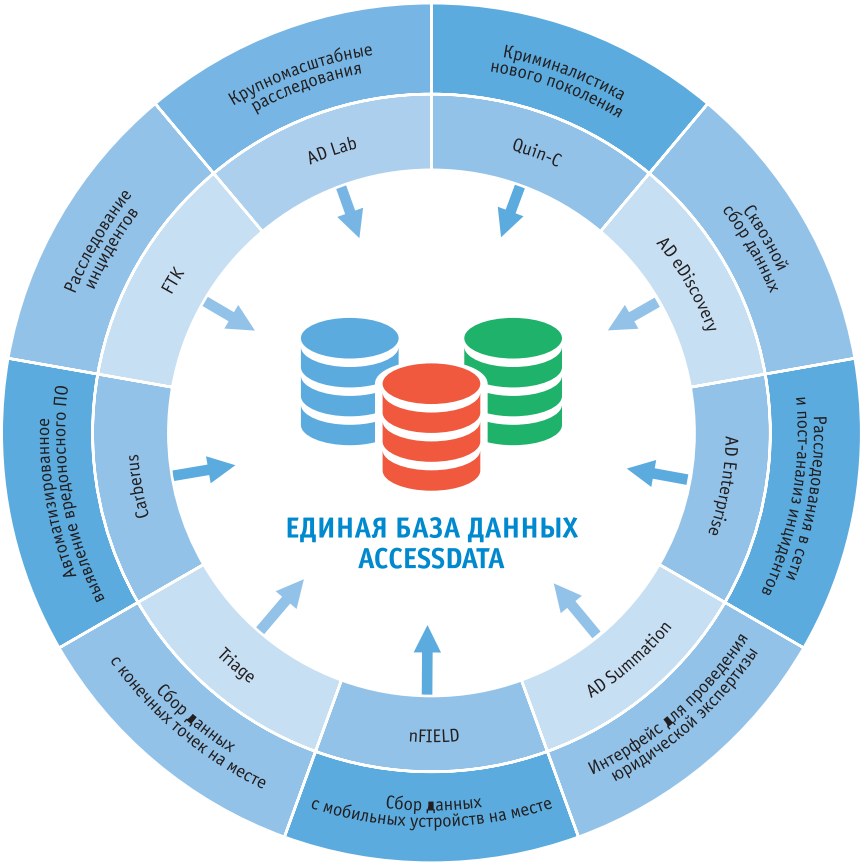

*Рис. 3. «**Бесшовный» процесс расследования инцидентов и нарушений***

гий может применяться как для управления процессом расследования, так и со временем повышает его точность и эффективность. Например, в рамках проведения юридического анализа ИИ ускоряет анализ документов и обеспечивает результаты, которые являются прозрачными и юридически оправданными: осуществляемая кластеризация документов и возможности предсказания используют МО для автоматической классификации документов. Фактически, искусственный интеллект позволяет сделать за эксперта первую часть работы: изучить документы, проанализировать улики и определить вероятный уровень сложности дела, то есть быстро подготовить необходимую базу для второго этапа работы – проверки и вынесения заключения по делу.

Интернет вещей (IoT) является для правоохранительных органов дополнительным источником информации: данные из автомобилей (видеорегистраторы, GPS, Bluetooth, которые позволяют получить данные о его использовании, местонахождении), данные из бытовых приборов, систем охраны объектов, систем видеонаблюдения на улицах городов, в общественном транспорте позволяют при условии использования технологий ИИ строить дополнительные связи между людьми, событиями и обогащать тем самым материалы следствия. В будущем интеграция с системами хранения ДНК, с криминалистическими системами правоохранительных органов позволит решать задачи не только на корпоративном и государственном уровне, но и на международном.

Природа машинного обучения заключается в том, что благодаря повторению и наблюдению, с течением времени решения могут обеспечить более быструю производительность и улучшить качество результатов. В контексте цифровой криминалистики мы видим, что с помощью этих новых технологий следственные группы могут на каждом уровне квалификации проводить и фиксировать более точные результаты расследований в более короткие сроки. Машинное обучение может оказать влияние на каждом этапе расследования: использование графических и видеоматериалов, интегрированная визуализация (временные графики, карты, диаграммы), анализ коммуникаций между людьми, распознавание изображений, распознавание лиц с фотографий, выявление связей. Продвинутые инструменты позволяют легко искать данные по всем делам и находить глубокие перекрестные доказательства, быстро анализировать объемы информации и делать большие, разнообразные наборы данных более применимыми.

Возможности машинного обучения позволяют отслеживать каждый шаг на пути достижения результата, помогают быстрее делать более точные и релевантные выводы, а также максимально эффективно использовать доступные инструменты для создания рабочих процессов, автоматизации задач и сотрудничества между командами. В настоящее время на передний план выдвигаются решения для сквозного бесшовного анализа со всеми продуктами на единой базе данных для непрерывной и эффективной работы команд в рамках рабочих процессов по работе с юридически значимой информацией и максимальным использованием современных технологий в своих решениях. Схематическое изображение процесса расследования инцидентов с помощью сквозной «бесшовной» технологии представлено на рис. 3.

Как ИИ работает над задачами информационной безопасности

Рассмотрим пример того, как человек пользуется своей электронной почтой. Можно выделить характерные паттерны поведения, анализ которых поможет определить его действия. Из поведения пользователя при работе с почтой можно выделить ряд характеристик: время входа, обычное месторасположение, количество авторизаций, используемых устройств и др. Есть операции, которые не характерны для конкретной личности: например, удаление прочитанных писем, стирание папок, отправка рассылок. У человека может сложиться определенное поведение

при работе с разными типами писем: прочтение писем, присланных конкретными людьми, удаление рассылок, игнорирование писем из социальных сетей. Кроме того, существуют и такие привычки, как «читает цепочку непрочитанных писем снизу вверх», «авторизуется и идет сначала в Почту, потом – в Диск и уже следом – в Новости» и т. д. Такие паттерны поведения могут быть вычислены для многих наших сервисов. Из совокупности этих факторов и складывается профиль пользователя, который не дает полного представления о о нем самом, но позволяет отличить факт взлома аккаунта от нормальной авторизации.

Конечно, этот подход не может быть эффективным без использования МО. С его помощью определяется набор факторов, которые влияют на профиль, и границы для определения взлома. Суть этого метода очень проста: каждый имеет привычки, свойственные только ему, начиная с режима работы и отдыха, продолжая местами, в которых человек бывает, и количеством устройств, которыми он пользуется. Например, кто-то всегда проверяет почту из дома и на работе, работает при этом с двумя устройствами, никогда не удаляет прочитанные письма и не рассылает спам. Он входит в почтовое приложение в дневные часы и никогда не проверяет ее ночью. А кто-то часто бывает в командировках и периодически читает почту, находясь в разных городах и часовых поясах. У этих пользователей будут непохожие паттерны поведения, на основе которых можно построить индивидуальный профиль и сравнивать с ним каждый новый вход в почту.

На рис. 4 показаны профили двух разных людей. На графике пунктирной линией показан профиль обычного, не взломанного пользователя. Он характеризуется достаточной равномерностью и отсутствием резких скачков параметров активности. Сплошная линия иллюстрирует поведение подозрительного аккаунта: все показатели сильно «скачут», прослеживается хаотичное обращение к ресурсу, что позволяет предположить факт несанкционированного доступа.

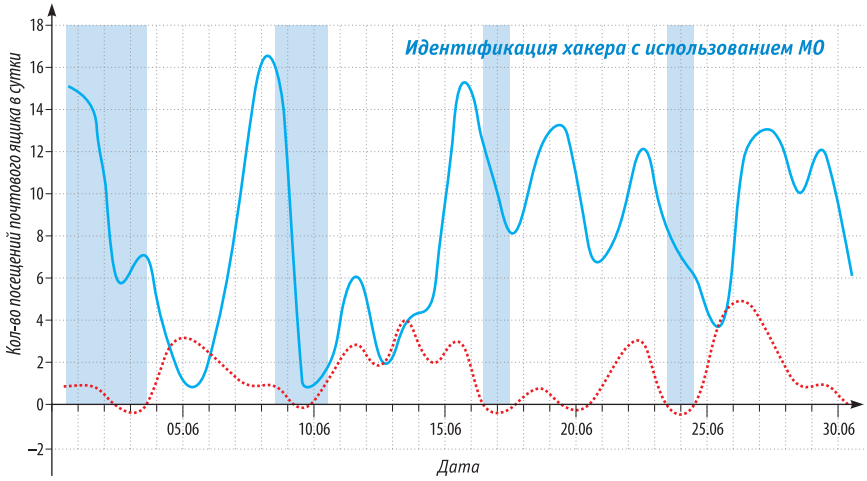

Рис. 4. ***График поведения человека в нормальных условиях***

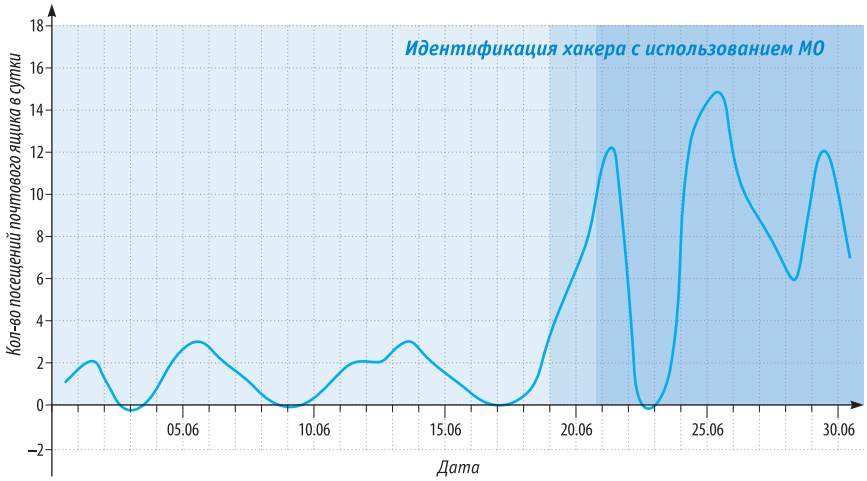

Рис. 5. ***График активности поведения хакера***

Теперь представим, что хакер, узнав какимлибо образом пароль от почтового ящика, зашел в него как пользователь. Его поведение будет явно отличаться от поведения человека, который пользовался электронным ящиком легитимно. На графике (рис. 5) поведение злоумышленника показано характерными всплесками. В светлой области видно, что показатели находятся в норме, в то время как в темной зоне уже прослеживаются значительные колебания. Кроме того, отчетливо видны даты, в которые это произошло, что может значительно упростить поиск момента взлома. Такой подход может обезопасить пользователей от кражи паролей и сессионных cookie-файлов и позволит детектировать взлом даже после авторизации в аккаунте. Задача алгоритма МО заключается в том, чтобы определить тот момент, когда изменилось поведение человека, образовавшее такой всплеск.

Машинное обучение в задачах искусственного интеллекта

Приведем четыре основных задачи искусственного интеллекта:

- классификация;
- регрессия;
- ранжирование;
- кластеризация.

Машинное обучение состоит из двух процессов.

Первый – это тренировка, когда человек берет данные, обучает модель и в итоге получает некоторый классификатор (рис. 6).

Второй – это уже использование МО, когда обученный классификатор внедряется в систему. Затем на вход системы подаются новые данные, которые классификатор ранее не видел.

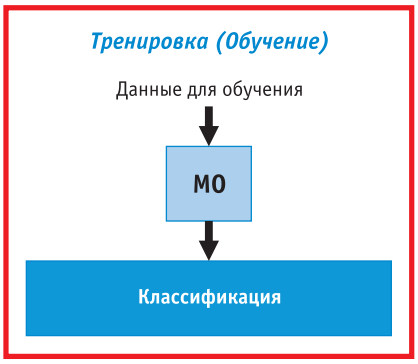

Рис. 6. ***Тренировка модели машинного обучения***

В результате мы получаем предсказания от классификатора (рис. 7).

Составные части машинного обучения

Машинное обучение включает в себя три компонента: *данные, признаки и алгоритмы.*

Данные. В открытом доступе есть множество наборов данных, на которых можно тренировать алгоритмы, но у таких наборов есть недостатки. Например, наборы могут быть неполными, плохо размеченными и неточными. Если необходимо внедрить решение на технологиях МО, нужно будет собрать набор данных под определенную задачу, и готовый набор вряд ли будет хорошо соответствовать решаемой задаче. ИТ-специалисты готовы выкладывать алгоритмы, рассказывать, что и как они используют, но мало кто хочет делиться своими наборами данных. Задача специалиста по науке о данных (дата-сайентиста) – подготовить набор данных к использованию: собрать, разметить и вычистить его. Это очень трудоемкий процесс, который занимает примерно 50–70 % работы.

Признаки. Рассмотрим простой web-запрос. Допустим, имеется: длина запроса, код ответа, URL, контекст, популярность домена и т. д. Предположим, таких признаков можно набрать 600 штук. В этом случае возникает два важных вопроса:

- какие из этих признаков брать, а какие не стоит;
- как будет использовано решение – в режиме реального времени или офлайн?

Причем, придется искать компромисс между этими двумя параметрами. Например, в первом случае нужно, чтобы модель могла быстро считать, поэтому признаков стоит взять поменьше, модель послабее. В офлайн-режиме можно выгрузить данные, анализировать их и распределять по категориям. В этом случае можно использовать модель любой сложности, потому что можно заставить алгоритм обрабатывать данные столько (по времени), сколько понадобится.

Приведем еще один пример. Предположим, нам нужно понять, опасный перед нами файл или нет. Для этого сначала ответим на следующие вопросы:

- требует ли файл доступа в Интернет;
- делает ли он что-то похожее на сканирование;
- какие IP использует файл;
- хочет ли он получить доступ к реестру?
- работает ли файл с памятью?
- хочет ли он изменить файловую систему?
- имеет ли файл возможность самокопирования или захвата других файлов?

Ответы на эти вопросы помогут выявить признаки, которые можно использовать для решения стоящей перед нами задачи.

Алгоритмы. Алгоритмы можно разделить на несколько типов:

- обучение без учителя;
- обучение с учителем;
- обучение с частичным привлечением учителя;
- обучение с подкреплением.

Эти алгоритмы часто применяются в технологиях «интеллектуального анализа данных» (*Data Mining*) и «науке о данных» (*Data Science*).

Теперь определим коренное отличие машинного обучения от классического программирования: оно поможет нам в дальнейшем оперировать вопросами киберустойчивости ИИ в условиях постоянно меняющегося ландшафта угроз со стороны преступного сообщества и отдельных нарушителей ИБ [12].

Как происходит программирование в классическом понимании? Допустим, у человека есть компьютер, который работает по определенному алгоритму. Человек вводит в него данные, задает программу, и алгоритм выдает результаты. В этом случае все предельно понятно. Человек может получить точность до 100 %, особенно если запрос – это математические операции.

А вот в случае с машинным обучением программа и выходы поменяются местами. То есть, человек дает алгоритму данные и указывает правильные решения, а дальше компьютер думает, как сделать так, чтобы из этих данных получались желаемые результаты. В процессе такой работы и рождается программа.

На практике процесс МО в некоторой степени автоматизирован за счет использования некоторых типовых программных конструкций – фреймворков, то есть программных платформ (каркасов), которые упрощает разработку программного продукта, определяют структуру проекта и помогают удобно объединять в нем разные компоненты. Фреймворк определяется, как множество конкретных и абстрактных классов, а также определений способов их взаимоотношения. Конкретные классы обычно реализуют взаимные отношения между всеми классами. Абстрактные классы представляют собой точки расширения, в которых каркасы могут быть использованы или адаптированы.

Точка расширения – это та часть фреймворка, для которой не приве-

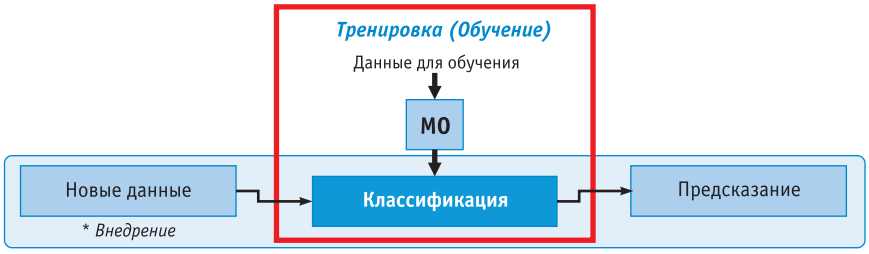

Рис. 7. ***Модель использования обученного классификатора***

дена реализация. Соответственно, каркас концептуальной модели состоит из концептуальных классов, а каркас программной системы – из классов языка программирования общего назначения. Процесс создания фреймворка заключается в выборе подмножества задач проблемы и их реализаций. В ходе реализаций общие средства решения задач заключаются в конкретных классах, а изменяемые средства выносятся в точки расширения.

Заключение

Использование высокоинтеллектуальных систем для управления безопасностью, в широком смысле этого понятия, является актуальным трендом времени.

Цифровая трансформация (ЦТ) также создала акцент на защиту конфиденциальности и более высокие требования к ее соблюдению. По мере того, как кибератаки становятся все более изощренными и разрушительными, регулирующие органы устанавливают более строгие правила и руководящие принципы защиты персональной идентификационной информации (*Personally Identifiable Information*, PII). В результате, организации должны помнить о комплаенсе, то есть о соответствии принятым внутренним или внешним требованиям или нормам, и обращаться к лучшим в своем классе сертифицированным продуктам, процессам и специалистам, чтобы обеспечить должный уровень управления рисками.

Даже до начала ЦТ системы ИБ во многих организациях по умолчанию включали несколько разрозненных хранилищ с локальными службами и развертывались, как правило, в нескольких облачных сервисах с различными инструментами безопасности. Стратегия ЦТ может привести к появлению еще более сложной ИТ-среды, с еще большим количеством облаков/ЦОД и увеличением количества устройств IoT, многие из которых не были разработаны с учетом требований кибербезопасности. При этом особое внимание необходимо уделить инцидентам ИБ на объектах критической информационной инфраструктуры (целевым (APT) атакам, физическому похищению активов, техногенным катастрофам и др.). ■

ЛИТЕРАТУРА

1. Артамонов В. А., Артамонова Е. В., Сафонов А. Е. Искусственный интеллект: когнитивное начало // Защита информации. Инсайд. – 2022. – № 4 (106). – С. 50–59.

2. Применение технологии искусственного интеллекта в информационной безопасности // Anti-malware [Электронный ресурс]. – URL: https://www.anti-malware.ru/analytics/Technology_Analysis/using-artificial-intelligence-technologies-in-information-security/ (дата обращения: 19.07.2022).

3. Rathgeb C., Uhl A. Asurvey on biometric cryptosystems and cancelable biometrics // EURASIP Journal on Info. Security. 2011. № 3. – 25 p. DOI:10.1186/1687-417X-2011-3.

4. Tarek M., Ouda O., Hamza T. Robust cancellable biometrics scheme based on neural networks // IET Biometrics. 2016. № 5. P. 220–228.

5. Albakri A., Mokbel C. Convolutional neural network biometric cryptosystem for the protection of the blockchain's private key // Procedia Computer Science. 2019. V. 160 (09). P. 235–240.

6. Riazi M., Chavoshian S., Koushanfar F. Automatic Synthetic Fingerprint Generation. 2020. – arXiv: 2002.08900.

7. Бабаш А. В., Баранова Е. К. Криптографические методы защиты информации. – М.: Кнорус. – 2016. – 190 с.

8. Варлатая С. К., Шаханова М. В. Криптографические методы и средства обеспечения информационной безопасности. – М.: Проспект. – 2017. –152 с.

9. Siddeeq Y. Mahdi A. H. AES Cryptosystem Development Using Neural Networks // International Journal of Computer and Electrical Engineering. 2011. V. 3, № 2. P. 315–318.

10. Shumeet Baluja. Hiding images in plain sight: Deep steganography. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, Advances in Neural Information Processing Systems 30, p. 2069–2079. Curran Associates, Inc., 2017.

11. Искусственный интеллект – будущее криминалистических расследований // Anti-malware [Электронный ресурс]. – URL: https://www.anti-malware.ru/analytics/Technology_Analysis/artificial-intelligence-is-computer-forensic-future/ (дата обращения: 22.07.2022).

12. ИИ в ИБ // хакер.ru [Электронный ресурс]. – URL: https://xakep.ru/2021/07/21/nn-in-ib/ (дата обращения: 19.07.2022).