ЕЛЕНА ЛАРИНА, ВЛАДИМИР ОВЧИНСКИЙ

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: ТЕХНОЛОГИЯ ТРОЙНОГО НАЗНАЧЕНИЯ

Сегодняшний бум искусственного интеллекта (ИИ) и робототехники создает у многих иллюзию, что мы имеем дело с каким-то новейшим открытием. Это не так.

Попытки скопировать человеческий интеллект и создать "думающие" машины предпринимаются уже несколько веков.

Некоторые источники утверждают, что первым создателем прообраза роботов был *Архит из Тарента* – древнегреческий математик и механик. Между 400 и 350 годами до нашей эры он построил *деревянного парового голубя*, который мог покрывать расстояние в 200 метров.

В 1495 году *Леонардо да Винчи* изобрел механического человека, напоминающего *германского рыцаря.* А несколько ранее, в 1478 году, он придумал *самоходную тележку*, которую многие современные эксперты считают первым в истории программируемым аппаратом. Вместо паровой энергии и двигателя внутреннего сгорания транспортное средство приводилось в движение заведённой пружиной. Тележка двигалась только вперёд, но "оператор" мог повернуть её колеса в определенные промежутки времени, поместив колышки в маленькие отверстия.

В 2004 году итальянские специалисты (дизайнеры, программисты, инженеры и плотники) собрались вместе, чтобы реконструировать автоматическую тележку Леонардо. У них это получилось.

*В 1770 году венгерский изобретатель Вольфган фон Кемпелен* создал автомат, который обыгрывал в шахматы всех, кто с ним состязался, и даже победил Наполеона Бонапарта. При этом для большего эффекта зрителям демонстрировались сложные механизмы машины. Позже изобретатель был разоблачен – внутри автомата сидел опытный шахматист, управляющий всем процессом.

В 1830-х годах *английский математик Чарльз Бэббидж* придумал концепцию сложного цифрового калькулятора – аналитической машины, которая, как утверждал разработчик, могла бы рассчитывать ходы для игры в шахматы. А уже в 1914 году *директор одного из испанских технических институтов Леонардо Торрес Кеведо* изготовил электромеханическое устройство, способное разыгрывать простейшие шахматные эндшпили почти так же хорошо, как и человек.

В 1954 году американский исследователь *Аллен Ньюэлл* решил написать программу для игры в шахматы. К работе были привлечены аналитики корпо-

рации RANDCorporation. В качестве теоретической основы программы был использован метод, предложенный основателем теории информации *Клодом Шенноном*, а его точная формализация была выполнена английским математиком и криптографом *Аланом Тьюрингом*.

С середины 30-х годов прошлого столетия, с момента публикации работ Тьюринга, в которых обсуждались проблемы создания устройств, способных самостоятельно решать различные сложные задачи, к проблеме ИИ в мировом научном сообществе стали относиться внимательно. Тьюринг предложил считать интеллектуальной такую машину, которую испытатель в процессе общения с ней не сможет отличить от человека. Тогда же появился *термин BabyMachine – концепция, предполагающая обучение искусственного разума на манер маленького ребенка, а не создание сразу "умного взрослого" робота.*

Летом 1956 года в Университете Дартмута в США прошла первая рабочая конференция с участием таких ученых, как *Маккарти, Минский, Шеннон, Тьюринг* и другие, которые впоследствии были названы *основателями сферы искусственного разума*. В течение 6 недель обсуждали возможности реализации проектов в сфере ИИ. Тогда и появился сам *термин artificial intelligence – искусственный интеллект*. После знаменитой конференции в Дартмуте ИИ получил впечатляющее развитие. Были созданы машины, которые могли решать математические проблемы, обыгрывать в шахматы, и даже первый прообраз чат-бота, который мог разговаривать с людьми.

Пожалуй, самым распространенным заблуждением тех лет было представление, будто для создания ИИ надо точно разобраться с тем, как работает человеческий мозг и как он связан с сознанием. Современные специалисты в области ИИ полагают, что компьютеры ближе к арифмометрам и калькуляторам, чем к человеческому мозгу. *Нейронные сети, о которых говорят компьютерщики, не имеют никакого отношения к нейронам человеческого мозга*. Наиболее совершенные нейронные сети имеют сегодня пять-шесть слоёв и минимум синапсов. В человеческом мозге таких слоёв сотни тысяч и миллионы. При этом в одних областях компьютеры, как программно-аппаратные комплексы, уже сегодня превосходят людей, а в других – безнадежно уступают. *Поэтому человек и компьютер – принципиально разные устройства, также как человек не похож на компьютер, так и компьютер не похож на человека*.

За истекшие 60 лет сложилось три основных направления определения ИИ. Эти различные понимания – не отвлеченные рассуждения. На программы, построенные на каждом из них, потрачены миллиарды долларов.

Обычно, когда дело касается ИИ, все вспоминают знаменитый *тест Тьюринга*. С тестом связано первое направление определения ИИ. Суть теста в следующем: *если при общении с компьютером посредством анонимного канала связи нельзя понять, с кем идет беседа – с человеком или машиной – то такой компьютер можно считать ИИ*. Грубо говоря, ИИ – это интеллект, похожий на человеческий, по результатам действий, т. е. по поведению. Долгое время всех удовлетворяло такое понимание ИИ. Собственно, знаменитый Watson – это и есть реализация на практике программно-аппаратного комплекса, способного пройти тест Тьюринга. Watson, кстати, породил нынешний бум ботов. Боты призваны вести элементарную беседу. Они используются сегодня во многих странах повсеместно – от торговых площадок до больниц, от полицейских участков до справочных служб.

Другое направление ИИ связывается со *способностью программ к самосовершенствованию*. Не случайно, что о нейронных сетях, глубоком обучении и ИИ заговорили одновременно. На самом деле известны они примерно те же 60 лет. Главная проблема была в дороговизне "железа", т. е. самих компьютеров, способных выполнять эти программы. Нейронная сеть может эффективно решать конкретные задачи, но при этом никогда не пройдет теста Тьюринга.

Большинство практиков использует третье понимание ИИ. Это – *программно-аппаратный комплекс, работающий с использованием нейронных сетей, глубокого обучения и способный общаться с человеком на естественном языке, в том числе посредством голоса*.

Если подходить с инженерной точки зрения, то необходимо понять, где компьютеры сильнее людей, и что нам от них нужно. Посмотрим на эту проблему на примере анализа ФБР провалов и успехов ИИ, связанных с борьбой с криминалом.

На сегодняшний день успехи достигнуты там, где имеются огромные массивы Больших Данных (БД), ограниченное время для их анализа и возможность написать программу анализа. Грубо говоря, **компьютер превосходит человека там, где имеет место огромная комбинаторика, т. е. наличие множества вариантов, короткое время исполнения и возможность вести анализ чего-либо путем выполнения последовательных операций, т. е. возможность написать алгоритм.**

Где на сегодняшний день отмечены наибольшие прорывы? В анализе БД, распознавании образов, поиске незаметных на первый взгляд связей и закономерностей. Отсюда возникает простое заключение. Если бы человек имел бесконечное время на решение той или иной задачи, был дисциплинирован и имел неограниченный объём памяти, то он бы успешно решал все задачи, где компьютер уже сегодня первенствует над ним. Самые знаменитые достижения компьютеров, подаваемых как ИИ, связаны с победой в играх – от шахмат до го, от покера до "бесконечных шашек". Любая игра имеет правила. А там, где есть правила, путь к успеху лежит в комбинаторике и написании алгоритмов.

Приведенные соображения позволили информационным подразделениям ФБР совместно с Лабораторией искусственного интеллекта корпорации Google выработать следующее инженерное определение ИИ. Именно оно положено в разработку концепции архитектуры и перечня программных решений ФБР: *ИИ – это программно-аппаратный комплекс, обеспечивающий поддержку и/или принятие результативных решений в динамичной, неустойчивой среде в установленное время, на основе заведомо неполной, нечёткой и не имеющий полной доказательной базы информации.* Применительно к одним задачам ИИ самостоятельно принимает решения, но в большинстве случаев является элементом гибридного интеллекта, взаимодействуя с человеком.

Успехи ИИ связаны с тремя основными факторами. *Во-первых*, с использованием новой высокопроизводительной элементной базы. *Во-вторых*, с применением новых программных решений, базирующихся на сложной комбинаторике и машинном обучении. *В-третьих*, с широким использованием робототехники как периферийных устройств ИИ, аналогичным периферийным устройствам человека, типа рук, ног, по отношению к мозгу.

### ИИ как технология тройного назначения

ИИ – это технология тройного назначения. ИИ может быть использован как для *гражданских*, так и для *военных* целей. Отдельное направление использования ИИ – *мафиозно-террористическое*. Поскольку некоторые задачи, требующие интеллекта, являются доброкачественными с точки зрения права, а другие – нет, то ИИ обладает свойством тройного использования, также как и человеческий интеллект.

О гражданском, мирном использовании ИИ СМИ сообщают буквально каждый день. Но самое активное использование ИИ наблюдается в **военных целях**.

Например, Министерство обороны США изучает множество разнообразных направлений использования ИИ. Эта работа ведется в основном в рамках *DARPA* (Управление перспективных исследовательских проектов Минобороны США) и *IARPA* (Агентство передовых исследований в сфере разведки). Разработкой стратегии использования ИИ в сфере национальной безопасности и координации исследований занимается Канцелярия помощника министра обороны по исследованиям и инженерии, а сам помощник несет личную ответственность перед министром обороны, администрацией Президента и Конгрессом за максимально эффективное использование ИИ в интересах национальной безопасности*.

В апреле 2017 года под руководством заместителя министра обороны США по разведке создана и начала активно работать междисциплинарная и многофункциональная команда по разработке стратегии и тактики *алгоритмических войн*, а также их программно-аппаратному обеспечению со стороны ИИ. Работа этой команды известна как *проект Maven*. Главная цель проекта Maven состоит в максимально быстром внедрении ИИ в оборонительные

---

* См.: Artificial Intelligence and National Security. Congressional Research Service. 26.04.2018.

и наступательные системы в сфере военного, финансово-экономического и поведенческого противоборства. Проект призван продемонстрировать огромный потенциал технологий ИИ. В рамках проекта на период до 2020 года поквартально расписаны цели и ресурсы.

В начале 2018 года Директор проекта **Maven** заявил: "**Maven** предназначен для того, чтобы быть пилотным проектом. Он призван продемонстрировать неисчерпаемый потенциал ИИ в сфере алгоритмических войн, а конкретно кибер-, финансово-экономических и поведенческих конфликтов и противоборств, а также в сфере управления и прогнозирования конфликтов на пяти полях боя: на земле, в воздухе, в космосе, под водой и в киберсреде".

Ожидается, что к 2020 году ИИ даст максимальный эффект в разведке для обработки и анализа больших данных, в том числе неструктурированных, зашумлённых и неполных. Одним из результатов проекта **Maven** стало *создание системы опережающего мониторинга и прогнозирования на основе разнообразных данных действий противника* (на примере борьбы с ИГИЛ). Система Cointer-ISIL-**Maven** начала эксплуатироваться с июля 2017 года, она включает в себя сложный программно-аппаратный комплекс, состоящий как из периферийных систем, так и центрального ИИ. В качестве периферийных систем используются автоматизированные дроны, оснащенные системами компьютерного оптического зрения. Среди принципиально новых модулей центрального ИИ, созданного в рамках проекта, необходимо отметить гибкие модифицированные блоки нейронных сетей с машинным обучением, позволяющих распознавать нечёткую оптическую информацию на уровне более высоком, чем наблюдатели-люди.

Помимо засекреченных, у разведывательного сообщества есть несколько публично рекламируемых исследовательских проектов в области ИИ. На начало 2018 года только в интересах ЦРУ осуществляется 137 публично финансируемых проектов, связанных с ИИ. Более 25 проектов связаны с использованием ИИ, в том числе в составе симбиотического интеллекта, совместно *с группами экспертов для прогнозирования будущих событий*, таких как террористические атаки, гражданские беспорядки, финансово-экономические, политические и военные кризисы и т. п.

IARPA в настоящее время финансирует крупнейший в истории США *проект по созданию человеко-машинной платформы симбиотического (гибридного – человек + ИИ) интеллекта для распознавания слабых сигналов в информационном шуме и прогнозирования маловероятных событий*. Также ИИ активно используется для разработки алгоритмов одновременного многоязычного распознавания речи и перевода акустической речи в тексты с уровнем, превосходящим применяющиеся в настоящее время системы машинного перевода.

В сентябре 2017 года Управление материально-технического снабжения сухопутных войск США подписало контракт с IBM на сумму 135 млн долларов для создания *персонального электронного помощника бойца штурмового отряда на базе ИИ*. Этот проект стал продолжением первого проекта, начатого в 2014-м и завершенного в 2016 году. В рамках первого проекта электронный индивидуальный помощник-эксперт был создан для работников полевых штабов дивизий быстрого развертывания на базе IBM Watson.

ВМС США заказали в 2017 году *версию Watson, предназначенную для разработки планов оптимального материально-технического снабжения военно-морских группировок и отдельных судов, находящихся в мировом океане, и контроля над их выполнением*. Командование сухопутной армии полагает, что использование логистического Watson в армии обеспечит ежегодную экономию 100 млн долларов за счет оптимального распределения логистических потоков и планов материально-технического обеспечения вооруженных сил.

Наиболее активно ИИ будет использоваться Министерством обороны США *в киберпространстве*.

В 2018 году Киберкомандование США разместило через DARPA заказы по использованию ИИ для мгновенного обнаружения аномалий и дыр в киберзащите. Представляется, что именно ИИ с его быстродействием позволит наиболее эффективно управлять боевыми киберплатформами на самой деликатной стадии киберпротивоборств – фазе проникновения в сети противника.

Вооруженные силы США стремятся максимально использовать ИИ *в области управления и контроля*. Наиболее продвинутая система создана в настоящее

время в ВВС США. Она в настоящее время доведена до уровня штабных работников командования ВВС. В период до 2019 года система охватит уровень авиационных полков и дивизий.

Как известно, одной из наиболее сложных в практическом плане задач является сохранение управляемости и поддержание взаимодействия командования и боевых единиц в ходе реальных военных действий, когда противник наносит удары не только на поле боя, но и по центрам командования. До настоящего времени ни в одной стране мира, насколько известно, не создана система регенерации командования и контроля в жёстких конфликтах. Регенеративная система должна быть организована таким образом, чтобы после выхода из строя тех или иных узлов и уровней командования система перестраивалась и в новой конфигурации сохраняла высокий уровень управления и координации. В настоящее время командование ВВС совместно с корпорацией Lockheed Martin и корпорацией Alphabet приступили к созданию такой системы на основе симбиотического интеллекта, используя традиционные командные центры и защищённый ИИ.

Все рода войск США в последние годы *имплантируют ИИ в различные типы автономных транспортных средств*. По сути, вооруженные силы ведут работу параллельно с бизнес-сектором по созданию транспортных средств с полным самообслуживанием. Военные подрядчики вооруженных сил, начиная с 2017 года ежегодно представляют такого рода автономные транспортные средства с использованием ИИ. С 2019 года Министерство обороны запускает проект стоимостью в 430 млн долларов по созданию систем, включающих центральный ИИ и роевые или стайные автономные транспортные средства, оснащённые датчиками и интерфейсами, позволяющими перейти от индивидуального к коллективному машинному обучению.

Исследовательская лаборатория ВВС завершила вторую фазу испытаний по программе “Недоверчивый Уигман”. В рамках программы впервые создан и проходит испытания *полноценный беспилотный истребитель пятого поколения*. В 2017 году тестовый вариант, реализованный на более дешёвом истребителе F16, прошёл испытание. В их ходе машина, оснащённая ИИ, автономно реагировала на события, которые не были включены в программу полетов и представляли собой непредвиденные препятствия и сложности для выполнения заданий. Из 17 испытательных заданий в 16, не считая самого первого, платформа справилась со всеми сложностями. Уже сегодня очевидно, что ИИ позволяет создавать полностью функциональные боевые истребители и самолёты-штурмовики, не уступающие, а по ряду параметров превосходящие такие же самолёты, пилотируемые людьми.

По сути, это представляет собой следующий шаг *после массового внедрения в военную практику дронов – робототехнических летательных комплексов с ограниченными огневыми возможностями.* Кроме того, по заданию ВВС в настоящее время завершается отработка комплексных авиационных звеньев, которые предусматривают патрулирование и ведение боевых действий группой самолетов, один из которых пилотируется человеком, а несколько – системами с ИИ. В этом случае человек может в определённых случаях изменить команды ИИ. Данная система разрабатывается ВВС, поскольку командование авиацией, по крайней мере, в настоящее время и в ближайшем будущем не готово доверить решение о применении тактического ядерного оружия, которым оснащены многоцелевые истребители-бомбардировщики, роботам.

Сухопутные войска и Корпус морской пехоты США испытали прототипы автономных транспортных средств, а том числе оснащённых средствами огневого поражения. В ходе действий Сил специального назначения США в Ираке, Афганистане и Сирии в 2017 году сухопутные войска уже активно применяли в боевых условиях роботизированные автономные эвакуационные машины.

Корпус морской пехоты в 2018 году начнёт принимать на вооружение *многофункциональный универсальный роботизированный тактический транспорт*. Роботизированное с элементами ИИ транспортное средство грузоподъемностью от одной до трёх тонн будет следовать за ротами и взводами морских пехотинцев по местности с любым рельефом и любой сложности. Средства предназначены для транспортировки любых грузов – от запасных патронов и снарядов до пищи и одеял. Несколько аналогичных средств в настоящее время разрабатываются и для сухопутных вооружённых сил.

Скорее всего, сама логика развития ИИ ведёт к тому, что возможности тройного использования будут постоянно увеличиваться. Многие задачи, которые в настоящее время автоматизируются, являются по своей сути тройственными. Например, ИИ, анализирующий программное обеспечение на уязвимости, может выполнять функции киберзащиты, кибердиверсий и киберпреступлений. Все то же самое относится к стаям дронов, находящихся под управлением ИИ. Машинное обучение в одном случае будет, например, повышать качество доставки медикаментов в отдаленные районы, в другом – боеприпасов в зоны конфликтов, а в третьем случае – наркотиков, минуя полицейские службы.

### Криминальные угрозы ИИ

*Развитие ИИ происходит в условиях увеличения анонимности.* Многие задачи включают в себя общение с другими людьми, наблюдение за ними, принятие решений, воздействующих на их поведение. Согласно экспертным оценкам, до 90% подобных задач могут быть автоматизированы и на горизонте пяти лет переданы ИИ. При этом соединение ИИ и интернета создаёт поистине безграничные возможности для злонамеренной деятельности, вплоть до убийств, которые правоохранительным органам весьма трудно квалифицировать именно как убийство. Проблема мира интернета в том, что в нём любое целенаправленное действие может быть замаскировано либо под отказ, либо нерегламентную работу оборудования. Разделить их никто не сможет.

*Развитие ИИ открывает широчайшие перспективы для преступности.* Уже сегодня злоумышленники могут без каких-либо препятствий приобрести как программные, так и аппаратные компоненты самых мощных систем ИИ. Более того, *широкое использование открытого кода в ИИ позволяет преступникам без каких-либо затрат средств получать доступ к последним разработкам ведущих компаний.*

В настоящее время алгоритмы ИИ создаются в течение месяцев, а не лет. Это достигается за счёт того, что над ними работают не закрытые коллективы специалистов, а открытые сообщества программистов. Соответственно, преступники имеют возможность бесплатно получать новейшие разработки. Министерство внутренней безопасности США пыталось законодательно ограничить распространение определенных разработок в области ИИ, но потерпело поражение.

Сегодняшние системы ИИ страдают от ряда новых неурегулированных уязвимостей. К ним относятся *атаки через фиктивные данные*, используемые при обучении, а также *злонамеренное вмешательство в работу нейронных сетей.* Эти уязвимости не имели аналогов в прошлом, поэтому не очень понятно, как с ними справляться. Есть основания полагать, что в будущем атаки на ИИ будут более эффективными, более точными, более сложными для атрибуции и будут использовать в основном уязвимости в программном обеспечении ИИ.

Проведенные эксперименты показывают, что более 70% атак на ИИ осуществляется с использованием ИИ. Соответственно, в будущем, по мере широкого распространения ИИ, следует ожидать нарастания количества подобных атак экспоненциальными темпами. Кроме того, при широком внедрении ИИ неизбежно усиление дифференциации в мощи и функциональных возможностях различных ИИ. Соответственно с уверенностью можно сказать, что *более сильные ИИ будут использоваться для атак, в том числе в целях установления программного контроля над более слабыми ИИ.*

Наверняка следует ожидать *атак на ИИ с использованием фишинга.* В обозримой перспективе основным типом ИИ будут системы, постоянно взаимодействующие с человеком: с программистами, аналитиками или администраторами. Поскольку именно *человек является наиболее уязвимым звеном в системах гибридного интеллекта*, есть основания полагать, что злоумышленники будут предпочитать дешёвые фишинговые атаки дорогостоящим атакам непосредственно на ИИ.

Самые современные фишинговые атаки предполагают *высокий уровень социального инжиниринга** и хорошее знание психологических особенностей

---

* Социальный инжиниринг – система управления поведением человека с использованием методов социологии и психологии.

того человека, против которого осуществляется атака. В конечном счете, всё делается для того, чтобы он кликнул на ссылку, которая позволит злонамеренному софту проникнуть не только в компьютер человека, но и в ИИ, с которым он взаимодействует.

Прогресс, несомненно, приведёт к увеличению масштабов и эффективности злоумышленников. Этот вывод следует из того простого факта, что *ИИ способствует повышению анонимности взаимодействующих с ним акторов*. Если актор знает, что его атака не может быть отслежена, или даже в худшем случае обнаружение не может надежно идентифицировать его с атакой, ему гораздо легче решиться на нападение, чем хакерам в прошлом. По мнению многих практиков в сфере кибербезопасности, распространение ИИ приведёт к значительным изменениям в психологии потенциальных злоумышленников и сформирует у них *установку на безнаказанность*.

Большую угрозу создаёт тот факт, что *особенностью настоящего времени является создание комбинированных систем, включающих ИИ и роботизированные устройства*. Известно, что ряд стран начал эксперименты по созданию обучающихся роев боевых дронов, связанных с ИИ, выступающим как их центр управления. Причем такие работы ведутся как государственными, так и негосударственными (террористическими и иными деструктивными) структурами.

Прогресс ИИ породит *новые угрозы*. Уже есть признаки того, что *начали совершаться киберпреступления и проведены хакерские атаки, которыми управлял не человек, а ИИ*.

Кроме того, системы ИИ активно используются *для распространения дезинформации и фейков*. В настоящее время дезинформация, как правило, разоблачается на основе анализа фотографического материала. Поскольку ИИ позволяет не только синтезировать любое фотоизображение, но и создать практически не отличимую от реальности фальсифицированную аудио- и видеозапись, можно ожидать, что в самое ближайшее время появятся технически сложные фейки, подкрепленные синтетическими фотографиями, аудио- и видеозаписями. Для того, чтобы доказать их поддельность, потребуются огромные финансовые средства, мощные технические возможности и усилия высококвалифицированного персонала.

В ближайшем будущем существующие угрозы будут дополнены новыми, связанными с развитием ИИ. Типовые угрозы станут гораздо более технически сложными и изощренными.

В ближайшие годы скачкообразно увеличится сложность кибератак и кибертерроризма. Типичными станут не привычные атаки, связанные с фишингом, заражением компьютеров и т. п., а гораздо более *высокотехнологичные атаки, нацеленные на овладение информационными массивами атакуемых компьютеров и перехват управления ими*. С другой стороны, более широкое распространение получат *целевые атаки*. В настоящее время типичная кибератака со стороны высокотехнологичных преступников ориентирована на компьютеры, обслуживаемые тем или иным провайдером, расположенные в той или иной местности и т. п. При подключении к киберпреступным атакам ИИ можно будет проводить предварительную селекцию не самих технических средств, а их обладателей по полу, возрасту, профессиональным занятиям и т. п. Соответственно в этом случае атаки будут ориентированы не на регионы или провайдеров, а на те или иные группы населения, либо компании, обладающие определенными характеристиками. Это будет *киберпреступностью принципиально нового типа*.

В 2017 году антитеррористические подразделения Израиля успешно провели испытания дрона, который атаковал в многотысячной толпе строго определенных лиц. В качестве эксперимента одежда этих лиц была обрызгана определенной краской. Однако никто не мешает вместо краски использовать отравляющее вещество либо просто пулю. Мы имеем дело с объединением дрона с ИИ, способного в потоковом видео опознать лицо среди тысяч субъектов.

В подавляющем большинстве докладов по теме ИИ львиная доля внимания, связанного с угрозами, приходится на риски попадания ИИ в детские руки либо в руки террористов и т. п. Что касается детей и подростков, то разрушительный эффект их деятельности подчас оказывается сопоставимым с ударами со стороны экстремистских радикалов, вооруженных гаджетами.

В 2017 году в Польше на протяжении двух дней было парализовано все городское движение просто потому, что одному 13-летнему "таланту" захотелось проверить свои расчеты относительно того, можно или нет проложить трамвайную линию не внутри города, а между городами. Эта шалость обошлась Польше почти в 30 млн злотых, и почти 10 человек, пострадавших в авариях, были доставлены в больницы.

На сегодняшний день накоплено достаточно материала, чтобы изложить классификацию вредоносного использования элементов ИИ по недосмотру, ошибке и т. п., приводящего к негативным последствиям.

В максимально грубом приближении можно выделить *три типа угроз, связанных с использованием ИИ добропорядочными акторами*.

*Первая группа* объединяет ИИ с подавляющим большинством других сложных машин, созданных человеком. Речь идёт о банальных отказах.

*Вторая группа* угроз сопряжена с особенностями программного обеспечения ИИ. На сегодняшний день и, видимо, в период ближайших пяти лет, алгоритмическим ядром ИИ будут выступать нейронные сети вкупе с машинным обучением. По сути, нейронные сети – это программная поисковая среда, которая постоянно меняется за счёт перенормирования удельных весов, определённых программой, в зависимости от фактически полученных результатов.

Если в 2015–2017 годах ИИ использовал простые нейронные сети, а соответственно разработчики и аналитики хорошо понимали значение перенормировок на каждой итерации расчётов, то *нынешние глубокие сети оказываются для человека чёрным ящиком*. Фактически *возникает ситуация, когда машины делают выводы, которые в подавляющем большинстве являются точными, но как и почему они делаются, люди не понимают.*

В научных и политических дискуссиях, которые ведутся вокруг модели "ИИ как чёрный ящик", прежде всего, в США, а также Великобритании и Израиле, на первый план выступает стремление сделать этот чёрный ящик прозрачным и понятным для аналитиков. Однако, если посмотреть статистику фактических инцидентов с ИИ, то заботиться надо не о вскрытии чёрного ящика, а о явном задании времени оптимизации.

Многие исследователи опасаются, что компьютер при решении той или иной задачи построит программу, в которой оптимизироваться должно то, что оптимизируемым с точки зрения человеческого общества ни в коем случае быть не может. Грубо говоря, существует перезагруженный авиационный маршрут. Число желающих осуществить перелёт намного превышает возможности авиакомпании. Компьютер, рассмотрев различные способы решения этой проблемы, пришёл к выводу, что лучшим вариантом будет серьезная авария без смертельных случаев, но с большим числом раненых самолета данной авиакомпании на данном маршруте. Модель показала падение числа желающих до нормативного уровня. У математиков эта ситуация известна как *отсутствие запрета на скрытую оптимизацию*.

Данный пример показывает не только появление принципиально новых угроз, но и принципиальное различие в подготовке, анализе и принятии решения у человека и компьютера. Человек отказался бы от подобной оптимизации на самой ранней стадии разработки темы. А компьютер выбрал её, как основную.

Еще одна группа угроз связана, как это ни парадоксально, с *притуплением внимания и снижением ответственности лиц, принимающих решения, чьим советником является ИИ*. В отличие от триллеров и фантастических блокбастеров, лица, принимающие решения, это, в подавляющем большинстве, обычные, зачастую даже по интеллектуальным способностям средние люди. Они находятся под прессингом воздействия социальных СМИ, телевидения, интернета, которые изо дня в день вот уже на протяжении двух-трёх лет рассказывают о всемогуществе ИИ. Соответственно, даже в тех случаях, когда окончательные решения остаются за человеком, то, как показали эксперименты в университетах Йокогамы (Япония) и Ванкувера (Канада), лица, принимающие решения на уровне полицейских управлений городов, более чем в 98% случаев солидаризировались с рекомендациями ИИ и принимали те решения, которые де-факто выработал ИИ.

В одном случае опыт проводился для 70 ситуаций, в которых принимали участие три полицейских начальника, а в другом – для 300 ситуаций, где работали пять начальников. Самым удивительным итогом эксперимента стало

следующее. ИИ дали неправильные ответы по оценке ситуации для Японии примерно в 20% случаев, для Канады – в 17%. Начальники же в тех примерно 10% случаев, где приняли решение вопреки ИИ, правы оказались лишь в Канаде в 5%, а в Японии – ни в одном. Данные выкладки показывают, что тема гибридного или человеко-машинного интеллекта чрезвычайно сложна. В конечном счете мы пытаемся соединить то, в чём мы вообще ничего не понимаем, – человеческое сознание, с тем, что является техникой в первом поколении ИИ, и надеемся на базе этого соединения успешно решать все проблемы.

### Сценарии злонамеренного использования ИИ

Ключевой угрозой является *автоматизация социальной инженерии*. С помощью ИИ на человека, являющегося целью социальных инженеров, собирается досье. При этом особое внимание обращается на его непроизвольные автоматические реакции, которые и будут использоваться социальными инженерами при фишинговых атаках, использовании телефонии и т. п. По мере развития ИИ в целях обеспечения анонимности возможно использование социальными инженерами чат-ботов, которые будут вести разговоры с жертвами. Наряду с автоматизацией социальной инженерии следует ожидать использования ИИ для улучшения выбора целей и определения приоритетов в злонамеренных атаках. Автономное программное обеспечение, внедренное в атакуемую сеть, будет в течение долгого времени обеспечивать ИИ необходимой информацией.

Следует ожидать, что уже к 2020 году практически на всех континентах будут широко использоваться *беспилотные летательные аппараты, управляемые ИИ*. Наряду с бизнесовым и личным использованием дронов, следует ожидать их массовой закупки преступниками и террористами, например, для транспортировки наркотиков и доставки взрывчатых веществ.

Недооцененным, но чрезвычайно важным обстоятельством является *использование элементов ИИ как способа компенсации низкой квалификации людей*. Например, активное применение наводящихся с использованием ИИ снайперских винтовок дальнего действия, с джойстиками для управления, заметно снижает требования к профессиональной подготовке боевиков. Вооружённый подобной винтовкой малообразованный и не имеющий боевого опыта человек по эффективности действий может соответствовать бывалому бойцу спецназа.

*ИИ позволит перейти к принципиально новым типам боевых действий*. ИГИЛ в ходе боевых действий 2017 года активно использовал шахид-мобили. В среднем цели достигала пятая часть шахид-мобилей. Если предположить, что шахид-мобили будут иметь связь с ИИ, который одновременно будет соединен с дронами, показывающими топографию местности и боевую ситуацию, то результативность шахид-мобилей может быть увеличена в два-три раза.

### ИИ и преступность будущего

Сегодня аналитики, писатели и представители СМИ основное внимание при анализе будущего преступности уделяют ИИ.

По оценкам аналитиков ФБР США, в ближайшие годы главный ущерб американским домохозяйствам и корпорациям будут наносить хакеры и обычная традиционная беловоротничковая преступность. Однако это не означает, что не надо смотреть в будущее. Надо просто здраво оценить те направления, где преступники будут в первую очередь использовать систему ИИ.

Существуют три причины, по которым ИИ в ближайшие годы вызовет жгучий интерес у криминала в США. Первая причина – несомненное первенство Америки в области информационно-коммуникационных технологий. Уровень оснащенности программно-аппаратными комплексами самого различного типа домохозяйств, корпораций и государственного сектора является беспрецедентным в мире. По усреднённым оценкам ведущих исследовательских центров, США в области ИКТ в сфере разработок опережают страны Большой семерки на пять-семь лет, а в фазе применения – на три-четыре года. Однако столь высокий уровень оснащенности имеет и оборотную сторону. Сегодня телекоммуникационные сети и компьютерные устройства всех типов представляют собой наиболее важную инфраструктуру США, превосходящую по своему значению коммунальную, транспортную и энергетическую.

Не будет преувеличением сказать: существование США возможно только при работающих компьютерных сетях.

Любые сети нуждаются в защите. Все люди, обладающие минимальной компьютерной грамотностью, знают о стандартах информационной безопасности. Это уровень информационной безопасности, гарантирующий домохозяйства, корпорации или государственные учреждения от проникновения в сети хакеров, организации информации или даже перехват ими управления.

Если взять третью по вкладу в ВВП отрасль США — финансы, то выясняется удручающая картина. Для того чтобы американские банки и инвестиционные институты смогли модернизировать системы информационной безопасности, они должны при условии безубыточной работы и отсутствия инфляции начать выдавать кредиты под 7–9%. Это убьет американскую экономику.

Иными словами, в ближайшие годы и США, и другим странам предстоит жить в опасном мире.

Вторая причина притягательности ИИ для криминального сообщества — это характер киберпространства. Если при должной работе Министерства внутренней безопасности, миграционных служб, полиции и т. п. США, по оценкам ФБР, может положить конец нелегальной миграции, то в киберпространстве, увы, стену не построишь.

Между тем, динамика такова, что из года в год доля компьютерных преступлений, а точнее преступлений, осуществлённых в киберпространстве, в общем объёме преступности неуклонно растет. Если в 2000 году на долю компьютерной преступности приходилось не более 5% от общей криминальной добычи и ущерба, в 2007-м — 12%, в 2010-м — 25%, то в настоящее время, по оценкам Центра изучения компьютерной преступности Северо-восточного университета в Чикаго — не менее 45%, а по данным ФБР — около 30%. Это — экспоненциальный рост. При этом *доля открываемых уголовных дел по компьютерным преступлениям в США в пять-шесть раз ниже, чем по традиционным видам преступности*.

Преступность необратимо уходит и действует через киберпространство. При этом необязательно, чтобы само преступление совершалось в виртуале. С появлением интернета *киберпространство всё чаще используется для совершения традиционных преступлений при помощи нетрадиционных орудий и методов*.

Как было заявлено на Всемирном экономическом форуме (ВЭФ) в январе 2017 года, растущая взаимная киберзависимость инфраструктурных сетей является одним из ключевых факторов риска в мировом масштабе. *В докладе ВЭФ "Глобальные риски, 2017 год"* говорится о том, что кибератаки, дефекты программного обеспечения и другие факторы могут привести к системным сбоям, которые способны "каскадно распространяться по сетям, влияя на общество самым неожиданным образом".

Отчет Совета национальной разведки США (январь 2017 года), посвящённый глобальным тенденциям, также содержит предостережение о том, что общество стоит перед "*надвигающимся" риском киберразрушения* — потенциально в массовом масштабе и с "летальными последствиями" — ввиду уязвимости критически важной инфраструктуры.

В исследовании компании PwC (2018 год) подчеркнуто, что большинство коммерческих структур, ставших жертвами, утверждает, что не в состоянии точно установить лиц, совершивших кибератаки. Лишь 39% участников глобального опроса заявили, что уверены в том, кому вменять киберпреступление. *В России уверенность ещё ниже, лишь 19% респондентов могут определить источники кибератак.*

В мае 2017 г. лидеры стран "*Большой семерки*" взяли на себя обязательство сообща и вместе с другими партнёрами работать над противодействием кибератакам и снижением их воздействия на критически важную инфраструктуру и общество. Спустя два месяца лидеры стран "Большой двадцатки" вновь признали необходимость обеспечения кибербезопасности и повышения доверия к цифровым технологиям.

Как отмечается *в Отчете о Глобальном индексе кибербезопасности за 2017 год Международного союза электросвязи ООН*, глобальное межсетевое взаимодействие способно подвергнуть киберрискам "что угодно и кого угодно", и "всё что угодно, от критически важной государственной инфраструктуры до базовых прав человека может оказаться под угрозой".

В уже упомянутом Исследовании глобальных тенденций информационной безопасности на 2018 год (PwC) руководители организаций, использующих системы автоматизации и роботизации, отмечают осознание значительности потенциальных негативных последствий кибератак. В качестве основного возможного результата кибератаки 40% участников опроса в мире и 37% в России называют нарушение операционной деятельности, 39% – утечку конфиденциальных данных (48% – в России), 32% – причинение вреда качеству продукции (27% – в России); 29% – нанесение ущерба материальному имуществу (30% – в России) и 22% – причинение вреда человеческой жизни (21% – в России).

В интернете нет границ. Никакого отдельно американского, британского, китайского или русского интернета не существует. Соответственно юридическая база для работы правоохранительных органов, сложившаяся для обычного пространства, не подходит для киберпространства.

Как данные ФБР и правоохранительных органов, так и результаты исследовательских тем, открытых в ведущих университетах США, позволяют говорить, что *в настоящее время отсутствуют признаки целенаправленных усилий ОПГ по созданию собственных разработок в области ИИ*.

Успешные преступники, работающие по-крупному в таких сферах, как финансы, крупномасштабная контрабанда, нелегальная купля-продажа интеллектуальной собственности и т. п., – люди предельно рациональные. На данном уровне разработок в области ИИ у них нет необходимости привлекать внимание, вербуя в свои ряды команды наиболее продвинутых стартапов, за которыми охотятся военное и разведывательное сообщества, крупнейшие корпорации. Сегодня это не нужно. Почему?

Прежде всего, стремясь минимизировать издержки и привлечь к развитию собственного продукта максимальное количество внешних, в значительной степени бесплатных, разработчиков, большинство ведущих производителей платформ ИИ уже выпустили платформы с открытым кодом.

По мнению аналитиков ФБР, использование ИИ криминалом в США в течение ближайших пяти лет будет иметь место в нескольких приоритетных сферах. **Их объединяет наиболее благоприятное для криминала соотношение трёх переменных: полученный преступный доход, совокупные приведенные издержки на подготовку, совершение и сокрытие преступления и уровень риска.**

1. *Использование ИИ для компроментации и имплантации вредоносного софта в действующие платёжные системы*, в основном использующие протокол блокчейн.

Большая четверка (Visa, MasterCard, AmericanExpress и DinnerClub) вложила огромные деньги в создание инфраструктуры информационной безопасности. Тем не менее, преступники кладут в свой карман не менее одной десятой доходов процессинговых компаний. Одноранговые платёжные системы вытесняют процессинговые компании. Прежде всего, за счёт экономии издержек для клиентов. При этом по состоянию на 2017 год из без малого 30 платёжных сервисов, построенных на блокчейне, действующих в США, лишь семь удовлетворяют требованиям компьютерной безопасности. Соответственно подсоединение к платёжным сервисам и добавление к каждой транзакции порядка 0,1–0,3% принесет миллиардные доходы преступникам при отсутствии какого-либо риска.

Программы ИИ в данном случае крайне важны. Они позволяют использовать методы глубокого обучения нейронных сетей для взлома и перепрограммирования платежных протоколов, построенных на блокчейне. Эксперименты, проведенные в университете Санта-Фе и Дармутском университете, показали, что программы ИИ справляются с этой задачей эффективнее, чем люди-программисты. Уязвимость заключена в *блокчейне*. Он, как любой код, базируется на правилах и алгоритмах. Именно на них построены игры – от шахмат до покера, где ИИ победил человека.

2. На долю *высокотехнологичного киберкриминала, извлекающего прибыль из торговых операций крупнейших финансовых институтов*, приходится 40–50 млрд долларов ежегодно. Это – наиболее прибыльная, хотя и достаточно рискованная сфера организованной киберпреступности.

Поскольку в последние несколько лет развернулась настоящая гонка финансовых вооружений, выражающаяся в совершенствовании всеми крупней-

шими финансовыми институтами своих платформ на основе ИИ, преступникам даже для того, чтобы хотя бы сохранить долю доходов, необходимо участвовать в этой гонке. В этой связи использование преступными группами ИИ для операций на финансовых рынках путем проникновения и компрометации торговых платформ не оставляет для криминала другой возможности, как использовать лучшие решения ИИ с открытым кодом. В отличие от ситуации в платёжном бизнесе, где в 2017–2020 годах следует ожидать резкого увеличения размеров и доли преступных доходов в обороте платёжных систем, в алгоритмическом трейдинге в краткосрочной перспективе доля преступников будет снижаться. Вряд ли в ближайшие годы им удастся не только вырваться вперёд, но и просто сохранить паритет в оснащённости программами с ИИ по сравнению с ведущими финансовыми институтами.

3. Есть основания полагать, что по мере развертывания технологической гонки *интерес киберпреступников к интеллектуальной собственности* будет только нарастать. Известно также, что для вскрытия сегодняшних мощных систем корпоративно-информационной безопасности всё шире используются многофункциональные программы, в основе которых лежат самосовершенствующиеся алгоритмические модули. Подобные модули — это ключевой элемент ИИ.

Эксперты ФБР констатируют, что США оказались не готовы к отпору хакерским группировкам, нацелившимся на интеллектуальную собственность, принадлежащую корпорациям, федеральному правительству и университетам.

Средний срок пребывания хакерского софта в корпоративных сетях в тех случаях, когда он в итоге бывает всё-таки обнаружен, увеличился с 2014-го по 2016 год с примерно 150 до 230 дней. При этом, по оценке экспертов ФБР, удаётся обнаружить примерно 30–40% от общего числа активных проникновений в корпоративные сети. И это — в крупнейших компаниях.

4. *Использование ИИ для разведывательной деятельности организованной преступностью против полиции и ФБР.*

*Колумбийский наркокартель Кали* ещё в начале 90-х годов прошлого века приобрел мощную компьютерную систему IBMAS/400, стоившую в те времена полтора миллиона долларов, и обзавелся штатом сисадминов и программистов, разрабатывающих специализированный софт для datamining.

Техника была нужна для того, чтобы прочесать краденые базы данных с рабочими и домашними телефонами сотрудников американских спецслужб и дипломатических работников в Колумбии, сопоставить их с полным списком всех телефонных звонков, которые совершаются в стране, и выявить потенциальных информаторов, подлежащих ликвидации.

Если ОПГ не могут уничтожить базы правоохранителей, то они, очевидно, пойдут другим путём. В любой системе самый уязвимый фактор — это человек. В течение 2016-2017 годов ФБР внимательно отслеживало *попытки купить на чёрном рынке те или иные базы изображений с видеокамер, установленных в кафе, торговых центрах, рядом с полицейскими участками, зданиями ФБР* и т. п. Это наводит нас на мысль, что *преступники начали создание собственной базы* данных с использованием примерно тех же решений ИИ, что и правоохранительные органы. Первая — это *база агентов под прикрытием и осведомителей*. В делах, относящихся к компетенции ФБР, более чем в 70% случаев успех был связан с работой секретных сотрудников.

Также ожидаются попытки создания криминалом *баз данных на сотрудников информационных центров полиции штатов и ФБР*. То есть людей, допущенных в святая святых. С учетом того, что у каждого, даже самого преданного и отважного правоохранителя есть уязвимые места, создание подобных баз могло бы иметь гибельные последствия.