

BIS

JOURNAL

ИНФОРМАЦИОННАЯ
БЕЗОПАСНОСТЬ
БИЗНЕСА

№3 (54) 2024

**Игорь
КАЧАЛИН**
АНО «НТЦ ЦК»
Вузы, бизнес
и государство
пока не попадают
в резонанс
→ 106



**Александр
КЛЕВЦОВ**
InfoWatch
DLP-системы
нуждаются
в развитии
→ 78



Искусственный интеллект: два пути

Константин ЛЕОНТЬЕВ
ООО «Клируэй Текнолоджис»
Об управлении
технологическими
сертификатами → 56

Денис КАЛЕМБЕРГ
Александр САНИН
SafeTech
Наш ответ Microsoft
→ 66

Станислав СМЫШЛЯЕВ
КриптоПро
Импортозамещение
в криптографии
не тормозит → 68



SOC Tech

Технологии SOC киберзащита

*** 15 ОКТ 2024**

▾ Москва, LOFT HALL



SOC Tech 2024 – это мероприятие нового типа
Оригинальные форматы и возможности
сверхэффективного взаимодействия для всех
участников, спикеров и партнёров

Регистрация: ib-bank.ru/soctech



Сергей ПАЗИЗИН
научный редактор
BIS Journal

В мае этого года состоялся II Форум «Технологии доверенного искусственного интеллекта». Это послужило поводом рассмотреть в нашем журнале находящуюся в топе новостей тему искусственного интеллекта и Больших языковых моделей.

Результаты ChatGPT, DeepL Translate, Google Assistant, компьютерной диагностики в медицине, распознавания лиц, беспилотных автомобилей впечатляют. Но чем шире мы используем автоматизацию на основе машинного обучения и нейронных сетей, тем острее становятся вопросы безопасности и доверия к результатам. Необходимо подготовить стандарты и инструменты, позволяющие обеспечить жизненный цикл разработки доверенных (безопасных) технологий искусственного интеллекта. Классифицировать атаки и разработать средства и методы противодействия. А также решать вопросы корректности данных, используемых для обучения, противодействовать использованию автоматизированных сервисов для создания дипфейков, фишинговых рассылок и других неправомерных целей.

Вторая тема данного номера имеет сугубо практический характер. Задача управления технологическими сертификатами и соответствующей РКИ-инфраструктурой, которые относятся не только к традиционной электронной подписи для документов, но и к веб-сервисам, стала по-настоящему актуальной в последние несколько лет в связи со взрывным ростом количества сертификатов в инфраструктуре современной компании. Необходимо отслеживать окончание срока их действия, проводить своевременную и безопасную замену, решать массу других сопутствующих задач.

В повседневной жизни мы в явном виде сталкиваемся с проблемой управления технологическими сертификатами, когда видим сообщение о небезопасном веб-сервере в связи с недействительностью сертификата, а в неявном — при сбоех в предоставлении сервисов различных компаний. На страницах журнала мы предоставили слово отечественным разработчикам, предлагающим решение этой важной проблемы.

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Батранков Д. В. — директор по продуктам ГК «Гарда»
Былевский П. Г. — шеф-редактор, доцент департамента ИБ Финансового университета при Правительстве РФ, доцент кафедры международной ИБ МГЛУ, кандидат философских наук
Велигура А. Н. — председатель комитета по банковской безопасности Ассоциации российских банков, руководитель комитета по безопасности НП «НПС», CISA, кандидат физико-математических наук
Виноградов А. Ю. — старший научный сотрудник ФГУП «ЦНИИХМ»
Вихорев С. В. — советник генерального директора по информационной безопасности IT Expertise
Дзержинский Ф. Я. — независимый эксперт
Зубков А. Н. — генеральный директор ООО «Медиа Группа „Авангард“», директор Фонда содействия развитию безопасных информационных технологий (ФСРБИТ)
Калашников А. И. — управляющий директор Центра информационной безопасности дочерних и зависимых обществ Газпромбанка (АО)
Курило А. П. — советник по вопросам ИБ ФБК и FBK CyberSecurity, кандидат технических наук
Левиев Д. О. — председатель Совета НП «Партнёрство специалистов по информационной безопасности»
Мельников С. Ю. — заместитель генерального директора ООО «Лингвистические и информационные технологии», доктор физико-математических наук
Некрасов И. В. — главный редактор журнала
Одувалов М. В. — директор службы Управления по обеспечению информационной безопасности ДБ Банка ВТБ (ПАО)
Окулесский В. А. — вице-президент по информационной безопасности ЦМРБанка, кандидат технических наук
Пазизин С. В. — заместитель начальника Управления по обеспечению информационной безопасности ДБ Банка ВТБ (ПАО), кандидат физико-математических наук, научный редактор журнала
Сабанов А. Г. — главный эксперт Научно-технического комплекса технологий информационного общества АО «НИИАС», доктор технических наук
Филипчук А. А. — технический директор ГК ЦРТ
Хайретдинов Р. Н. — заместитель генерального директора ГК «Гарда»
Шумский Л. С. — руководитель службы информационной безопасности Яндекс Банка
Янсон И. А. — бизнес-партнер по информационной безопасности, Департамент безопасности ПАО «Сбербанк России», кандидат физико-математических наук

РЕДАКЦИЯ ЖУРНАЛА

Главный редактор: **Некрасов И. В.**
 Коммерческий директор: **Минаков А. В.**
 Директор по развитию: **Абрамов А. В.**
 Арт-директор: **Мельников Н. С.**
 Обозреватели: **Бродская М. А.**, к. т. н., **Покатаева Е. Н.**
 Иллюстрация на обложке: **Виктор Миллер Гаусса**
 Фото в номере: **Freerik**
 Цветокорректор: **Наumenko М. В.**

УЧРЕДИТЕЛЬ-ИЗДАТЕЛЬ

ООО «Медиа Группа „Авангард“»
 Адрес: 127473, Россия, Москва, 3-й Самотечный переулок, д. 21
 Тел.: +7 495 921 42 44
 Интернет-версия: ib-bank.ru/bisjournal
 E-mail: bis@ib-bank.ru

Зарегистрирован 13 августа 2010 года Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор) как «BIS Journal — Информационная безопасность банков».

Зарегистрирован 11 июля 2023 года Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор) как «BIS Journal — Информационная безопасность бизнеса». Свидетельство ПИ № ФС77-85487 от 11.07.2023 г.

Все права на материалы, опубликованные в номере, принадлежат журналу «BIS Journal» — Информационная безопасность бизнеса. Перепечатка и воспроизведение иными способами без разрешения редакции запрещаются. При использовании материалов ссылка на журнал обязательна. Мнение редакции может не совпадать с мнением авторов. Присланные материалы не рецензируются и не возвращаются.

Подписан в печать 28.06.2024. Тираж 5000 экз.

Цена договорная. Отпечатан в типографии «Московский Печатный Двор», 123298, г. Москва, 3-я Хорошевская 18, корп. 1, офис 201А

АКТУАЛЬНОЕ МНЕНИЕ

- 05 **Конкурсы**
Игорь Некрасов (BIS Journal)
За победу!
- 06 **Оборотные штрафы**
Марина Бродская (BIS Journal)
Кейс Snowflake – урок или просто сенсация?
- 08 **Направления атак**
Даниил Щербаков (Servicerpipe),
Лев Палей (Вебмониторэкс)
Предупреждён – вооружён!

НОВОСТИ

- 10 **Законодательство, криминал, бизнес, в России**



ТЕМА НОМЕРА — ФЕНОМЕН LLM

- 14 Елена Покатаева (BIS Journal)
Что творит искусственный разум?
- 23 Антон Колонин (НГУ)
«Большие языковые модели – всё ещё статистические аппроксиматоры или уже почти настоящий интеллект?»
- 28 Константин Воронцов (Институт искусственного интеллекта МГУ),
Василий Курилов (Институт проблем общественного развития)
Карты знаний

ТЕМА НОМЕРА — ФЕНОМЕН LLM

- 34 Галина Новикова (Российская ассоциация искусственного интеллекта)
«Пора заменить расшифровку аббревиатуры LLM на Logic-Linguistic Model»
- 37 Григорий Шершуков (Наносемантика)
«В вопросе автоматизации следует сконцентрироваться на применении LLM для решения рутинных задач»
- 40 Денис Димитров (Сбербанк)
«Уже сегодня мультимодальные генеративные модели – это универсальный инструмент»
- 40 Владислав Тушканов (Лаборатория Касперского)
«LLM могут занять место в ряду других инструментов автоматизации ИБ»
- 44 Илья Тихонов (SL Soft)
«Об интеллекте LLM можно спорить, но мир разработки они способны перевернуть»
- 46 Александр Горшков (АЙРИС ДЕВАЙСЕЗ)
«Сегодня наибольшую опасность представляет не генерация поддельных видео, а создание при помощи алгоритмов ИИ поддельных документов»
- 47 Дмитрий Дырмовский (ГК ЦРТ)
«Большие языковые модели трансформируют все новые индустрии»

РЕГУЛЯТОРЫ

- 49 **Госдума**
Василий Пискарёв (Государственная Дума)
«Наша задача – защитить детей от идеологических диверсантов и мошенников»
- 52 **Парад законов**
Екатерина Рачкова (BIS Journal)
Квартальный отчёт



ИМПОРТОЗАМЕЩЕНИЕ

- 56 Константин Леонтьев (Клируэй Текнолоджис)
Обычно не видны...
- 66 Денис Калемберг (SafeTech), Александр Санин (SafeTech Lab)
Наш ответ Microsoft
- 68 Станислав Смышляев (КриптоПро)
Корпоративные центры сертификации и российская криптография
- 69 Сергей Халяпин (Аладдин)
Строгая аутентификация – ключ к доверенной корпоративной инфраструктуре

УГРОЗЫ И РЕШЕНИЯ

- 76 **LLM**
Новые возможности – новые уязвимости
- 77 **LLM**
Алексей Иванов (АйТи Бастион)
А как на практике?
- 78 **DLP-системы**
Александр Клевцов (InfoWatch)
От анализа контента к анализу контекста
- 82 **Инсайдеры**
Анатолий Песковский (Информзащита)
Они среди нас

ИБ В ОТРАСЛИ

- 70 **Банки**
Юрий Кардюков (eKassir)
Стандарт безопасности ФАПИ.СЕК
- 74 **Экспертиза**
Владимир Авдеев (RTM Group)
С лицевого счета похищено...

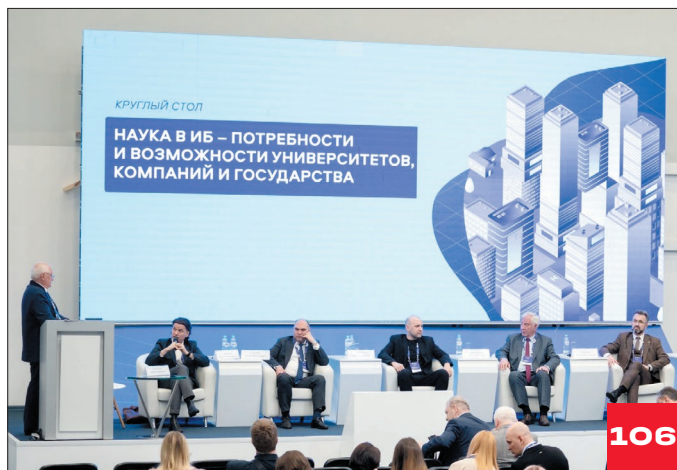
ИНФРАСТРУКТУРА

- 84 **Деловые мероприятия**
Марина Бродская (BIS Journal)
II Форум «Технологии доверенного искусственного интеллекта»
- 88 **Деловые мероприятия**
Марина Бродская (BIS Journal)
Positive Hack Days 2024
- 91 **Деловые мероприятия**
Анна Катанкина (BIS Journal)
«SOC с машинками в песочнице и без»
- 93 **Деловые мероприятия**
Анна Катанкина (BIS Journal)
UserGate V
- 96 **РКИ**
Сергей Кирюшкин (Газинформсервис)
Пространство доверия: промышленная эксплуатация



БЕЗОПАСНАЯ РАЗРАБОТКА

- 100 **Опыт Дзен**
Алина Князева (Дзен), Татьяна Хуртина (VK)
Прозрачность образов влечёт...

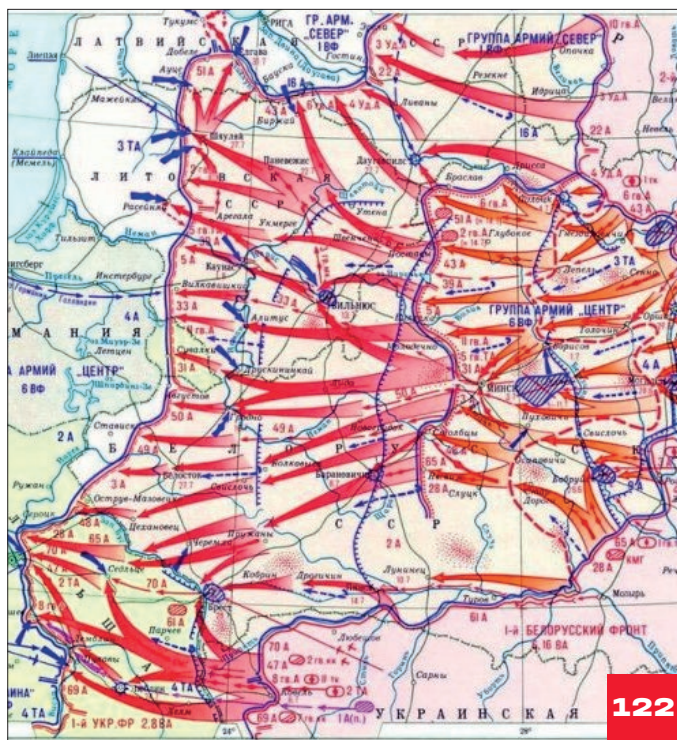


СОЦИАЛЬНАЯ ИНЖЕНЕРИЯ

- 103 LLM**
 Андрей Жаркевич (Start X)
 Искажение реальности

ОБРАЗОВАНИЕ И КАРЬЕРА

- 106 Форум**
 Игорь Качалин (НТЦ ЦК)
 Детали и главное



ЗА РУБЕЖОМ

- 110 Дайджест ИноСМИ**
 Анна Катанкина (BIS Journal)
 А есть ли у вас план, мистер Фикс?
- 114 Законодательство**
 Александр Виноградов (ФГУП «ЦНИИХМ»), Сергей Меньшиков (Belarusian InfoSecurity Community), Андрей Ситнов (независимый эксперт), Валерий Соколенко (ICBC) ОДКБ. Республика Таджикистан



СУБЪЕКТЫ

- 119 История**
 Борис Столпаков (историк криптографии)
 «Мы рождены для вдохновения!»
- 122 История**
 Борис Столпаков (историк криптографии)
 «Белорусский мотив в песне вересков, в песне раakit...»
- 128 Заметки безопасника**
 Александр Игонин (BIS Journal)
 Мал хостовик, да дорог

ЗА ПОБЕДУ!

BIS JOURNAL ПРИЗНАН ЛУЧШИМ ОТРАСЛЕВЫМ ИЗДАНИЕМ МОСКВЫ!



Игорь НЕКРАСОВ
главный редактор
BIS Journal

Шила в мешке не утаишь – всё равно все всё узнают, поэтому лучше из первых рук. Да, наш журнал «BIS Journal – Информационная безопасность бизнеса» стал победителем Московского регионального конкурса журналистов, который каждый год проводит Московская торгово-промышленная палата. В номинации «Лучшее отраслевое СМИ» были представлены десятки изданий из разных отраслей, но именно BIS Journal признали победителем. И это очень приятно...

Во-первых, просто приятно. Работаешь так, не поднимая головы, изо дня в день, из года в год, и вдруг тебе говорят: «Молодец! Ты – лучший!» Конечно, приятно! И за себя, и за команду, с которой делаешь журнал, и за общественную редакцию, искрящую умом и сообразительностью, и за издателя Медиагруппу «Авангард» – без него вообще, как без папы и мамы, ничего этого не было бы.

Во-вторых, приятно за информационную безопасность, над которой, собственно, журнал и взял шефство ещё в 2010 году. И буквально на наших глазах за какие-то 10–15 лет она выросла из со-

вершенно юного, малопонятного существа в зрелую могучую отрасль, встроилась в экономику и общественное пространство страны. И теперь уже не BIS Journal популяризирует её, объясняет людям, что такое ИБ и чем она важна-замечательна, а наоборот: значимость и востребованность отрасли высветила, словно прожектором, наш журнал из толпы других, положила руку на сердце, очень приличных изданий.

Но не буду умалять наших заслуг, мы действительно старались и всегда делали журнал правильно и хорошо – с искоркой, как сказали нам на вручении премии. Но, думаю, главная наша заслуга не только в этом, а в том, что мы делали это стабильно – именно изо дня в день, из года в год, что и превратило BIS Journal в летопись становления отрасли ИБ на всех её этапах, во всех противоречиях. А также в лицах. Без ложной скромности, мы представили на наших страницах (в формате интервью, комментариев, авторских статей) тысячи безопасников – руководителей и специалистов отрасли ИБ.

И последнее. ИБ очень сложная и многогранная тема, и, безусловно, мы не единственные в этой сфере. Такие СМИ, как CNews, «Хакер», Securitylab, Anti-malware и другие – это все наши, не менее уважаемые коллеги, работающие в своих нишах. И мы все вместе создаём общее информационное поле, в котором, как в зеркале, отражается жизнь нашего подопечного – отрасли ИБ. Каждое ритмично выходящее СМИ по-своему, разными цветами подсвечивает и акцентирует внимание на достижениях и проблемах, часто малопонятных коллегам из соседних журналистских цехов. И всё вместе это напоминает магическую цветовую палитру, под которую отрасль движется и растёт, словно фантастическое существо из другого мира.

В общем, получается, что звание «Лучшее отраслевое СМИ» – это наш общий – отраслевой – успех. И это, согласитесь, тоже чертовски приятно...

За победу!



КЕЙС SNOWFLAKE — УРОК ИЛИ ПРОСТО СЕНСАЦИЯ?

СТАНЕТ ЛИ ГОСОБЛАКО «ЧЁРНЫМ ЛЕБЕДЕМ»
ДЛЯ БИЗНЕСА И ГРАЖДАН РОССИИ?



**Марина
БРОДСКАЯ**
обозреватель
BIS Journal

СНЕЖИНКИ РАСТАЯЛИ. И ПРОТЕКЛИ

Для представления о масштабе проблемы — немного рекламы.

«Искусственный интеллект для общения с вашими данными. Генеративный ИИ, машинное обучение, потоковые или неструктурированные данные, создание приложений и Python, управление и непрерывность бизнеса... Всё это проще выполнять в облаке данных ИИ. Snowflake обеспечивает простоту использования, мгновенную эластичность и низкую совокупную стоимость владения».

Это лишь малая часть возможностей компании из США, которая занимается облачным хранением, вычислениями и анализом данных. Она создана в 2012 г., офисы открыты в 40 странах. В клиентах Snowflake — более 9,8 тыс. компа-

ний, включая мировые корпорации Adobe, AT&T, HP, Mastercard, Micron, NBC Universal, Nielsen, Novartis, Okta, PepsiCo, Siemens, US Foods, Western Union, Yamaha и др.

Чем ещё знаменит этот облачный сервис? Тем, что кибератака на него может войти в историю как самая крупная утечка данных в мире. В июне этого года стало известно, что клиенты «Снежинки» уже потеряли терабайты конфиденциальных сведений: Ticketmaster — 560 млн записей, Santander — 30 млн, Advance Auto Parts — 380 млн, LendingTree и её дочерняя структура QuoteWizard — 190 млн. Эксперты нашли в списке 165 пострадавших, среди них — фармацевтические гиганты, финансовые компании, службы доставки еды... И список продолжает пополняться.

Исследователи из компании по кибербезопасности Mandiant отслеживают группу, стоящую за хакерской кампанией, получившей название UNC 5537. Предполагается, что злоумышленники, основная цель которых — получение выкупа, попали в корпоративную среду Snowflake по цепочке поставщиков, используя учётные данные для входа, украденные ранее у клиентов.

Несмотря на предупреждения регуляторов США и Австралии, которые нашли прямые доказательства утечки, Snowflake продолжает утверждать, что сбой в ИТ-системах незначительный, а потенциальная утечка затронула ограниченное количество учётных записей.

ЯРКАЯ ПАРАЛЛЕЛЬ

История со Snowflake — яркий пример риска концентрации данных в руках одного провайдера. Похожих примеров, только в иных масштабах, достаточно и в России. Здесь можно вспомнить и уже забытую февральскую утечку данных, содержавшую 500 млн строк данных о россиянах, подробности о которой не стал раскрывать Роскомнадзор. И более мелкие, но не менее чувствительные утечки из сервисов доставки и различных ИТ-систем. Несмотря на аргументированные доводы экспертов сервиса разведки утечек данных и мониторинга даркнета DLBI о подлинности данных, на момент подписания номера утечка в московской ИТ-структуре традиционно была названа «компиляцией».

ЗНАЮТ ВСЕ!

Об опасности хранения персональных данных в одной, пусть даже и хорошо защищённой структуре, давно говорят российские эксперты. В 2023 г. глава Минцифры Максют Шадаев неоднократно заявлял, что ведомство перешло от стратегии сбора максимального количества данных на «Госуслугах» и концентрации больших данных в одном месте к реализации распределённого подхода хранения информации. Минцифры

Об опасности хранения персональных данных в одной, пусть даже и хорошо защищённой структуре, давно говорят российские эксперты

планирует создание ведомственной онлайн-витрины, данные в которую будут загружаться непосредственно из баз ведомств по запросу пользователя.

О риске концентраций данных у неподнадзорных компаний в июне 2024 г. говорил замглавы Департамента ИБ Банка России Андрей Выборнов. «На рынке есть организации, которые не являются поднадзорными Банку России, но от которых существенным образом зависит функционирование в целом финансового рынка и банковского сектора экономики», — отметил спикер на конференции «Информационная безопасность финансовых организаций в текущих условиях: ГОСТ Р 57580, анализ уязвимостей, защита персональных данных». — Это могут быть облачные провайдеры, ЦОДД, разработчики ПО, ИТ-провайдеры». Представитель регулятора отметил, что на решение этой сложной задачи уйдёт полтора-два года.

«ЗАКОН ОБ ОБОРОТНЫХ ШТРАФАХ»

На фоне громких утечек продолжается активное обсуждение так называемого Закона об оборотных штрафах. Он включает в себя два законопроекта — № 502104–8 «О внесении изменений в Кодекс РФ об административных правонарушениях (в части усиления ответственности за нарушение порядка обработки персональных данных)» и № 502113–8 «О внесении изменений в Уголовный кодекс РФ (в части установления ответственности за незаконные использование и передачу, сбор и хранение компьютерной информации, содержащей персональные данные)».

Эти законопроекты об ужесточении административной и уголовной ответственности за утечки персональных данных были приняты Госдумой в январе 2024 г. в первом чтении. Они предлагают ввести оборотные штрафы за повторные утечки персональных данных в размере не менее 15 млн руб. и не более 500 млн руб. в отношении юридических лиц. По мнению противников оборотных штрафов, такие суммы могут разорить бизнес и станут по-

водом для шантажа со стороны хакеров, готовых получить меньший выкуп, чем размер штрафа.

Председатель парламентского комитета по информационной политике Александр Хинштейн заявил во время ПМЭФ-2024, что Госдума примет закон об оборотных штрафах для компаний, которые допустили утечку персональных данных граждан. «Да, мы рассчитываем эту инициативу принять. Коллеги из Минцифры и других заинтересованных ведомств совместно с нами подготовили редакцию второго чтения — она сейчас находится на проработке в ГПУ президента. И наша задача в текущем году эту норму во втором и в третьем чтении принять», — сказал депутат (цитата по «Интерфаксу»). По его мнению, основная дискуссия, если и развернётся, коснётся смягчения ответственности. «Думаю, что у нас есть компромиссные решения, которые устроят всех. Решения, которые, с одной стороны, не станут лазейкой для ухода от ответственности и возможности откупиться нарушителю. А с другой стороны, действительно позволят участникам рынка, которые вкладываются в информационную безопасность, снижать масштабы ответственности», — так сказал парламентарий.

АБД ПРОТИВ

Второй момент, который подтверждается кейсом Snowflake, — это вопросы безопасности ИИ и работа с большими данными. В апреле компания представила модель генеративного ИИ Arctic LLM, оптимизированную для корпоративных нагрузок. Также Snowflake активно продвигает технологии ИИ для управления большими массивами данных. Как их могут использовать хакеры, пока неизвестно. Но и в этой части кейса можно провести параллель с российской действительностью.

Минцифры России уже заявляло о планах по созданию «государственной фабрики больших данных», где будут накапливаться и формироваться датасеты, основой для которых станут уже имеющиеся у государства информация и сведения, поступающие

от бизнеса. Законопроект № 992331–7, вносящий изменения в ФЗ «О персональных данных» в части уточнения порядка обработки персональных данных, был принят в первом чтении в феврале 2021 г. и до настоящего времени дорабатывается. В очередной раз он стоял в повестке дня Госдумы в июне 2024 г., однако поправки к законопроекту об обезличенных персональных данных, предложенные ко второму чтению А. Хинштейном, вызвали критику профессионального сообщества. Рассмотрение документа было вновь отложено.

По сведениям «Коммерсанта» и русского издания Forbes, Ассоциация больших данных (VK, «Яндекс», «Сбер», «Т-Банк», ВТБ, «большая четвёрка» операторов связи и др.) выступила против передачи клиентских персональных данных в государственную информационную систему. Появление такой системы и описанный порядок работы с ней чреваты рисками концентрации и кибербезопасности, противоречат поручениям президента, программе «Цифровая экономика» и Национальной стратегии развития ИИ до 2030 г. Принятие проекта в таком виде может задержать развитие технологий ИИ.

Доступ к ГИС получают только государственные и муниципальные органы, и компании из специального перечня, следует из законопроекта. Согласия граждан на передачу данных не потребуется. По сути, авторы поправок предлагают «передавать коммерческие данные государству», считают представители АБД и просят привлечь к работе над документом Банк России. В свою очередь, регулятор затруднился оценить прозрачность механизма «госозера данных» и его безопасность, пока нет уточняющих актов.

Остаётся ждать и надеяться, что государство и бизнес найдут решения, устраивающие все стороны, и учтут интересы граждан, персональные данные которых становятся разменной монетой для государства, бизнеса и злоумышленников.

ПРЕДУПРЕЖДЁН — ВООРУЖЁН!

ОБНАРУЖЕНА БУДУЩАЯ ЦЕЛЬ ХАКТИВИСТОВ — ОПЕРАТОРЫ ЦФА



**Даниил
ЩЕРБАКОВ**
заместитель
генерального
директора
Servicepipe



Лев ПАЛЕЙ
директор по
информационной
безопасности
«Вебмониторэкс»

висов этих организаций не пройдёт незаметной.

О РЕЗУЛЬТАТИВНОСТИ АТАК

Результативность атак на операторов ЦФА также может быть велика, поскольку противостоять политически мотивированным хакерам становится всё сложнее: на конец I квартала 2024 г. средняя пиковая мощность атак политически мотивированных злоумышленников выросла до 47 Гбит/с (год назад было 13 Гбит/с). Идут многовекторные атаки, в которых комбинируются различные методы воздействия, объём доходит до 400 Гбит/с. Длительность атак, организованная хактивистами, также увеличилась: в I квартале 2023 г. средняя DDoS-атака длилась 12 часов, через год — не менее суток, к концу июня — в среднем 4–5 дней.

КОВРОВЫЕ АТАКИ

В 2024 году также чётко прослеживается тренд на так называемые ковровые атаки, жертвами которых становились и компании финансового сектора, когда злоумышленники одновременно атаковали сразу целые диапазоны адресов и подсетей (не перебором, а все и сразу).

Подобные атаки создают перегрузку пограничного сетевого оборудования, что приводит к потере связанности сетей с интернетом. При стандартной атаке, направленной на конкретный ресурс / IP-адрес или группу IP-адресов, к поступающему на них паразитному трафику можно было применить метод blackhole (отправка всего трафика по несуществующему маршруту). При «ковровых бомбардировках» так делать нельзя, так как атака направлена сразу на все IP-адреса и отправка трафика

Июнь 2024 года был «щедр» на атаки в финансовом секторе, организованные хактивистами. Не будем называть имена пострадавших игроков, дабы не делать рекламы злоумышленникам. Но, полагаем, все, кто читал газеты или просматривал новостные сайты в июне, в курсе, насколько серьёзные организации оказались под ударом. Возникают вопросы, почему у этих игроков недоступность сервисов в принципе стала возможной. И кто может стать следующей жертвой? Успешными атаки, вероятно, стали из-за новой тактики DDoS-атак хактивистов. Следующими их жертвами могут стать операторы ЦФА (цифровых финансовых активов — инструмента для инвестиций на базе современных технологий: блокчейна и смарт-контрактов — прим. ред.), как очень заметные на рынке, но пока ещё не столь защищённые, как банки.

ОРИЕНТАЦИЯ НА РЕЗОНАНС

В последние годы финансовая отрасль была под прицелом политически мотивированных злоумышленников. Servicepipe фиксировал первые атаки хактивистов ещё в конце февраля 2022 года, и с тех пор их интерес к отрасли не ослабевает. Например, банки

традиционно входят в топ-3 наиболее атакуемых организаций. По итогам I квартала 2024 года практически четверть всех DDoS-атак, обрушившихся на российский бизнес, была именно на кредитные организации. Ещё 12% пришлось на микрокредиторов, 10% атак — на брокеров, 3% — на страховые компании. Фиксировались также атаки на коллекторские агентства, которые до 2024 года были в принципе вне сферы интересов хактивистов. Таким образом, на сегодняшний день операторы ЦФА — одни из немногих участников финансового рынка, на которых пока ещё не было совершено громких покушений. Будет ли это продолжаться и дальше? Маловероятно. И в первую очередь потому, что главная цель хактивистов — пошуметь, атака должна быть максимально громкой, а её жертва — заметной организацией.

ЧТО ПИСАЛИ?

Посмотрим, что писали СМИ в 2024 году об операторах ЦФА. Здесь и принятие закона, по которому ЦФА можно использовать в международных расчётах, и подготовленная Минфином России концепция расчётов ЦФА на базе нацвалют в рамках БРИКС, и объявления о размещении ЦФА на бирже крупнейшими игроками. Операторы ЦФА заметны, они на слуху. Значит, недоступность сер-

в «чёрную дыру» приведёт к той же недоступности ресурса. Это и является целью атакующих.

В частности, именно эта тактика была использована злоумышленниками в атаках на финансовый сектор в двадцатых числах июня. И если даже в атаках на крупнейших игроков отрасли, для которых DDoS-атаки совсем не новость, злоумышленникам удалось добиться недоступности сервисов, то смогут ли операторы ЦФА выстоять под ударами без надлежащей защиты?

ХАКЕРОВ ПРИВЛЕКУТ ДЕНЬГИ

Впрочем, интересны в качестве потенциальных жертв операторы ЦФА могут быть не только атакующим «по зову сердца», но и ради финансовой выгоды. Одна из главных причин — растущие финансовые потоки. Рынок ЦФА хотя и молодой, но быстрорастущий. По оценкам Аналитического кредитного рейтингового агентства (АКРА), объём рынка по находящимся в обращении выпускам ЦФА на конец 2023 года составил порядка 60 млрд руб., и к концу 2026 года планируется, что он вырастет до 500 млрд руб. А чем больше объём операций и чем выше количество пользователей, тем более привлекательна цель атаки.

ОКНО ДОСТУПА

Кроме того, сегодня под постоянным вниманием злоумышленников остаются веб-приложения, которые выступают единым окном доступа к услугам и продуктам компаний. Что как нельзя более актуально для компаний финансового сектора. По сути, такие веб-ресурсы зачастую являются и доступом во внутреннюю инфраструктуру за счёт плотной интеграции с внутренними информационными системами и автоматизации обмена данными между ними. После событий 2022 года и значительного усиления информбезопасности во многих финансовых организациях на передний план выходят сложные целевые атаки, ориентированные на использование уязвимостей веб-приложений и компонентов инфраструктуры. Динамика распространения уязви-

Почти 29 000 новых уязвимостей было найдено исследователями за 2023 год, и если смотреть статистику, то прирост составляет 10–20% ежегодно

мостей и скорость их обнаружения впечатляет: почти 29 000 новых уязвимостей было найдено исследователями за 2023 год, и если смотреть статистику, то прирост составляет 10–20% ежегодно. Особым вниманием пользуются API приложений, трафик атак на которые каждый год растёт на 200%. Самыми распространёнными типами атак за 2023 год, по наблюдениям «Вебмониторэкс», являются RCE (эксплуатация критической уязвимости) и XSS (подтип атаки на веб-системы, заключающийся во внедрении в выдаваемую веб-системой страницу вредоносного кода, который будет выполнен на компьютере пользователя при открытии им этой страницы, и взаимодействии этого кода с веб-сервером злоумышленника). Они занимают более 50% всего наблюдаемого трафика атак. В основном такой тип атак является частью пошаговой стратегии реализации более сложной целевой атаки, направленной на компрометацию инфраструктуры. И вот вы уже наблюдаете новости об утечках, взломах и дефейсах сайтов компаний.

ЗНАКОМЫЙ СПЕКТР

То есть для операторов ЦФА характерны те же самые угрозы, что и для других финансовых организаций. Кроме уже названных DDoS-атак, это могут быть:

- ◆ **фишинг:** мошенники могут использовать поддельные электронные письма и веб-сайты, чтобы обманом получить учётные данные пользователей и доступ к их цифровым активам;

- ◆ **вредоносные программы,** которые могут быть использованы для кражи данных, манипуляции с транзакциями или для блокирования доступа к системам;

- ◆ **эксплуатация уязвимостей:** хакеры могут использовать уязвимости в программном обеспечении платформы и API для получения несанкционированного доступа к системам и данным;

- ◆ **атаки на смарт-контракты:** смарт-контракты, используемые для управления ЦФА, могут содержать ошибки или уязвимости, которые злоумышленники могут использовать для кражи средств или изменения условий контракта;

- ◆ **социальная инженерия:** злоумышленники могут обманом заставить сотрудников или пользователей платформы раскрыть конфиденциальную информацию или совершить действия, которые приведут к компрометации системы.

ЗОМБИ-АТАКИ

Кроме того, в случае с ЦФА можно ожидать возвращения тех самых атак, которые ещё несколько лет назад были актуальны для финансового сектора. Например, были в своё время так называемые атаки на клиентов кредитных организаций с подменой реквизитов платёжных поручений. Велики риски, что операторы ЦФА также столкнутся с изменением данных транзакций для получения финансовой выгоды, а также с созданием поддельных транзакций для обмана пользователей и платформ.

И похоже, операторы ЦФА понимают, насколько возросли для них киберриски в последний год: некоторые уже подключили защиту «не хуже, чем в серьёзных банках», кто-то находится в процессе переговоров. Потому как возросший риск атак — не просто «пугалка», а реальность, с которой компании этого сегмента могут столкнуться в любой момент.

ЗА РЕШЁТКУ ЗА ДИПФЕЙК

Депутат Я. Нилов внесёт в Госдуму законопроект, предусматривающий уголовную ответственность за дипфейки.

Пояснительная записка к документу гласит: «Использование нейросетей и искусственного интеллекта (технологии «дипфейк», «цифровые маски» и т.д.) позволяют создавать подделки, отличить которые от реальности неспециалисту невозможно». Предлагается ввести в ряд статей УК РФ («Клевета», «Кража», «Мошенничество», «Вымогательство», «Причинение имущественного ущерба») дополнительный квалифицирующий признак — совершение преступления с использованием изображения или голоса (в том числе фальсифицированных или искусственно созданных) и биометрических данных.

Нарушителю будет грозить штраф до 1,5 млн руб. или лишение свободы на срок до семи лет.



ПАТЕНТ НА СОФТ

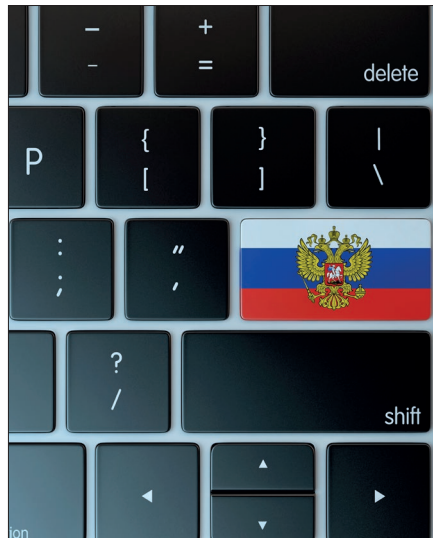
Минэкономразвития разработало и приняло правила патентования ИТ-разработок.

Ранее Роспатент не патентовал ИТ-продукты, они охранялись только авторским правом, защищавшим код и дизайн, но не принцип работы. Теперь же в заявке должно содержаться описание функционирования технологии, которая лежит в основе решения. Плюс, автор обяжут подтвердить новизну технологии и промышленную применимость.

По мнению экспертов, эти меры помогут донести имеющиеся механизмы защиты интеллектуальной собственности. Юристы добавляют: патент позволит внести ИТ-актив в финансовый учёт, что повлияет на оценку компании при привлечении инвесторов или продаже бизнеса.

В ДОПОЛНЕНИЕ К УКАЗУ ПРЕЗИДЕНТА РФ

Указом президента РФ от 13 июня № 500 приняты поправки к Указу № 250 «О дополнительных мерах по обеспечению информационной безопасности Российской Федерации».



Уточнения касаются широкого круга российских компаний и расширяют список ограниченных для госорганизаций.

Во-первых, должны быть установлены требования к аккредитованным центрам, которые будут привлекаться для реагирования на инциденты в ведомствах и субъектах КИИ. Такие центры, как следует из указа 2022 года, задействуют «в случае необходимости» для предупреждения кибератак и устранения их последствий.

Во-вторых, дополняется уже действующий запрет на использование средств защиты из недружественных стран: компаниям, на которые распространяется указ, с 2025 года также запрещается пользоваться сервисами (работами или услугами) по кибербезопасности от компаний из таких государств.

Поправки вступили в силу 13 июня.

БЕЛЫЙ ХАКЕР, ПОГОДИ!

На сайте Госдумы опубликован официальный отзыв правительства на законопроект о легализации деятельности белых хакеров в России. В целом правительство поддерживает законопроект, однако, он нуждается в доработке.

Законопроект № 509708–8, предполагающий легализацию в России деятельности белых хакеров, был внесен на рассмотрение в Госдуму ещё в декабре прошлого года. Он предполагает, что ИБ-экспертам можно будет проводить исследования программ для ЭВМ. При обнаружении уязвимостей «белые шляпы» должны будут сообщить о находке правообладателю в течение пяти рабочих дней.

В отзыве правительства речь идёт о том, что в представленном регулировании не учитывается необходимость информационного обеспечения деятельности, которое проводится уполномоченными органами федеральной испол-

нительной власти в соответствии с российским законодательством «О государственной тайне» и «Об информации», а также в сфере обеспечения безопасности российской КИИ.

Также подчёркивается, что интересам обеспечения безопасности РФ не отвечает возможность передачи данных об обнаруженных уязвимостях правообладателям, которые находятся под юрисдикцией государств, совершающих в отношении РФ недружественные действия.



ПОЛОЖИТЕЛЬНЫЙ ПРИМЕР

В начале мая российский пользователь криптовалюты потерял 1,155 «обёрнутых биткоинов» (wBTC), что на тот момент оценивалось в 68 млн долларов. Он обратился в компанию Match Systems, специализирующуюся на расследованиях криптовалютных инцидентов. Оперативники нашли украденные активы и вернули их владельцу в течение недели.

Один из крупных держателей криптовалюты стал жертвой классической dandruff attack, в результате которой коины были переведены на один адрес, а далее рассылены по нескольким кошелькам кибермошенника.

В отчёте Match Systems, представленном Cointelegraph, компания сообщает о том, что обнаружила информацию, укрепившую переговорные позиции жертвы. В частности, команда Match Systems провела детальный анализ инцидента и

выявила IP-адрес фактического местонахождения злоумышленника, а также другие детали, позволяющие установить личность преступника. В результате жёстких и стремительных переговоров с хакером с участием экспертов Match Systems и криптобиржи Cryptex злоумышленник вернул 22,960 ETH, что составило более 96 % от первоначально украденных средств в долларовом эквиваленте.



10 ЛЕТ ЗА 9 МЛРД

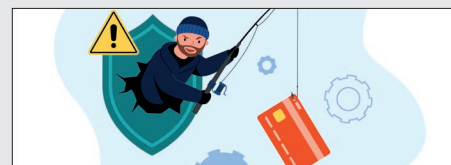
Александр Винник, обвиняемый американскими властями в отмывании денег и похищении ПДн, пошёл на сделку со следствием и признал вину по нескольким пунктам. Новый прогноз стороны защиты – 10 лет тюрьмы максимум (хотя изна-

чально прокуратура требовала для фигуранта пожизненного заключения).

Винник находится под стражей в ожидании окончательного решения суда уже семь лет. Россиянина арестовали в Греции в июле 2017 года по запросу США, где он обвинялся в отмывании свыше 9 млрд долл. через криптобиржу BTC-E. Эти средства, как сообщается, являлись доходами киберпреступников, а ущерб самой площадке оценивается в 121 млн долл.

Франция в свою очередь обвиняла Винника в мошенничестве. Также в списке инкриминируемых ему деяний значится похищение ПДн, которые фигурант возвращал жертвам за плату.

В 2019 году Греция приняла решение об экстрадиции Александра Винника в США, Францию и Россию по соответствующим запросам. В итоге в августе 2022-го обвиняемого передали американской стороне.



НА ТЕ ЖЕ ГРАБЛИ...

В конце мая жертвой скамеров стал Николай Кузнецов. До прошлой весны он был заместителем губернатора Смоленской области и отвечал за цифровизацию и кибербезопасность в регионе.

29 мая, когда Н. Кузнецов находился в Крыму, ему позвонили неизвестные и сообщили, что его денежные средства находятся под угрозой, и чтобы спасти накопления, их необходимо перевести на «безопасный счёт». Чтобы это сделать, Кузнецову пришлось поехать в Анапу и Новороссийск, где якобы находились «безопасные» платёжные терминалы. Через них он отправил звонившим 6,5 млн рублей, а затем вернулся в Крым, где и догадался, что стал жертвой кибермошенников. Экс-замгубернатора обратился в полицию.

МАКРОСРОК ЗА МИКРО-КРЕДИТЫ

Полиция задержала петербуржца, создававшего фишинговые сайты – копии официальных страниц госорганов – для сбора ПДн граждан.

При попытках воспользоваться официальными сервисами жертвы привычно вводили данные своих аккаунтов, зарегистрированных на Госуслугах, предоставляя таким образом преступнику доступ к размещённым в учётных записях сведениям.

Мошенник менял пароли и подавал через угнанные аккаунты заявки в микрокредитные организации, а средства выводил через анонимные электронные кошельки и обналичивал. Когда лимит на получение кредита исчерпывался, учётная запись удалялась вместе с информацией об её использовании.

По данным следствия, жертвами киберпреступника стали более 130 граждан.

MICROSOFT ОТЧАЛИВАЕТ... МУЧИТЕЛЬНО

По состоянию на июнь Microsoft отключила от «облаков» больше половины своих корпоративных клиентов из России, но так и не определилась по отношению к ним окончательно.

«На сегодняшний день с ограничением от вендора столкнулось более половины организаций, зарегистрированных на территории РФ. Microsoft перевёл статус этих подписок в suspended (приостановлено)», — уточнил источник в компании.

Весной американская корпорация заявляла, что с 20 марта ограничит доступ для российских организаций к своим облачным сервисам (Power BI и Dynamics CRM). Однако в апреле стало известно, что Microsoft продлевает лицензии юрлицам из РФ, а в мае — что россияне вновь могут обновить Windows и Office без использования VPN.

NVIDIA ПОСТАВИЛА НА ИИ-ЧИПЫ

В июне рыночная капитализация Nvidia впервые выбралась за отметку в три триллиона долларов — сегодня компания стоит \$ 3,012 трлн. Теперь она уступает лишь корпорации Microsoft (\$ 3,151 трлн и Apple (\$ 3,003 трлн).

В мае компания представила отчёт за первый квартал: выручка увеличилась до \$ 26 млрд (то есть утроилась), чистая прибыль составила \$ 14 млрд.

В мире ИИ-технологий Nvidia — один из самых успешных игроков, так как доминирует в сегменте ИИ-микрочипов. Продукты компании покупают такие гиганты, как Microsoft, Google, Amazon, OpenAI, и спрос заметно превышает предложение. Вендор уже заявил, что собирается выпускать новую архитектуру ежегодно (а не раз в два года), хотя и прошлые поколения чипов продаются отлично.

ГЕНЕРАТИВНЫЙ ИИ: РОСТ НА РЫНКЕ ОЧЕВИДЕН

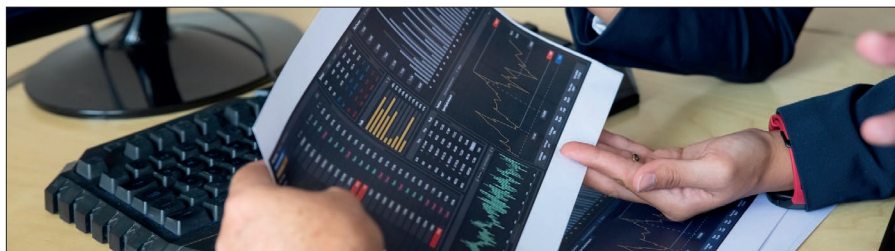
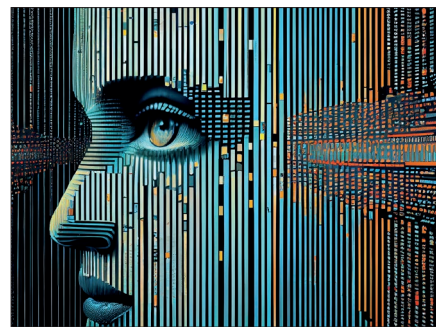
В 2022 году только 7,9 % опрошенных софтверных компаний имели измеримый эффект от использования генеративного ИИ в разработке ПО. Однако в 2023-м доля таких компаний выросла почти в два раза — до 15,2 %. В этом году ожидается увеличение до 19,8 %, говорится в исследовании РУССОФТ.

Всего 60 компаний из 303 опрошенных в рамках ежегодного исследования софтверной индустрии не только использовали генеративный ИИ в разработке ПО, но и определили, насколько эффективнее стала их работа.

О том, что ИИ напрямую может заменить разработчика, речь не идёт. Он может применяться лишь для увеличения производительности технических специалистов. В то же время увеличение производительности можно оценить по

числу сотрудников, которых понадобилось бы привлечь для тех же проектов, что были выполнены с использованием генеративного ПО.

Компании, использующие генеративный ИИ в разработке, в среднем чуть моложе тех, что пока его не используют (средний возраст около 13 лет против 14,3 лет), но при этом немного крупнее и имеют большую долю экспорта в обороте. Темпы роста у этих предприятий выше, чем у других опрошенных компаний.



ЕЩЁ ОДИН ФОНД ДЛЯ ИНВЕСТИЦИЙ В ИТ-СЕКТОР

До конца года «Ростелеком» собирается создать венчурный фонд «Консоль» с начальным объёмом 8 млрд рублей — для инвестиций в отечественных разработчиков системного ПО, платформы искусственного интеллекта и интернета вещей, облачные и серверные решения, а также сервисы для управления бизнес-процессами.

Приоритетными факторами будут «наличие функционирующих продуктов, подтверждённые продажи и гипотезы роста, полноценная ИТ-команда и корректно оформленные права на ПО». При этом российский

телеком-гигант с 2013 года уже владеет венчурным фондом «Коммит Кэпитал», направляющим инвестиции в технологические компании, которые профилируются на импортозамещении — «Ростелекому» принадлежат доли в десяти разных проектах.

Опрошенные эксперты отметили, что венчурный рынок медленно восстанавливается после прошлогоднего падения на 83 %, а «Ростелекому» будет непросто найти проекты в рамках упомянутых выше критериев из-за «сложной структуры принятия решений» в самой компании.

«ЦИФРОВАЯ ПРИВАТНОСТЬ НАРУШЕНА»

Количество утечек персональных данных россиян с начала года превысило 500 млн, заявил вице-президент по кибербезопасности Сбера Сергей Лебедь. Всего же в открытом доступе находится 10 млрд записей.

По словам Лебеда, в настоящее время не существует такого понятия, как цифровая приватность. В цифровом мире невозможно спрятаться: любое действие фиксируется и оставляет след — на серверах, в приложениях и социальных сетях.



Начинается это с самого детства: 11% детей в возрасте пяти лет имеют гаджеты, но, используя их, совершенно не думают о безопасности. Всё это создаёт для преступников огромные возможности по сбору персональных данных и предпочтений пользователей и может быть использовано как предлог для коммуникации.

Чтобы начать диалог с потенциальной жертвой, преступнику достаточно знать имя, фамилию и номер телефона, а дальше используются приёмы социальной инженерии, констатировал вице-президент: «Цифровая приватность нарушена, и нам всем предстоит это осознать. И создать те инструменты, которые позволят построить архитектуру Privacy by Design. Наверное, у нас не получится добиться этого сегодня, но мы должны думать о будущем. На самом деле в нашей стране уже много сделано и делается».

СОФТ ПОДОРОЖАЛ

В июне российские производители ПО подняли цены на 10-20% по сравнению с прошлым годом. Наибольший скачок в стоимости — до 30% — показали ИБ-продукты, а также системный и офисный софт.

Эксперты ожидают, что тенденция сохранится, так как вендоры будут адаптировать ценообразование под возвращение ставки налога на прибыль на уровне 5% со следующего года. Вместе с тем растущая конкуренция на рынке помогает сдерживать рост цен на ПО.

Среди причин называют повышение запросов клиентов, окончание «пилотов», дефицит ИТ-кадров и, как следствие, запрос персонала на повышение зарплаты — в среднем на 20% в год.

Для сравнения, за 2022 год российские ИТ-вендоры из-за большого спроса и отсутствия конкуренции подняли цены на свои продукты на 30-50% по сравнению с 2021-м.



АССОЦИАЦИЯ
БАНКОВ
РОССИИ

XXI Международный банковский форум

25 — 28 сентября 2024 года

Сочи

Одно из самых крупных и представительных мероприятий финансовой отрасли. Более 600 участников. Среди спикеров — руководители Банка России, федеральных министерств и ведомств, члены Совета Федерации, депутаты Госдумы, топ-менеджеры банков и компаний финансового и ИТ-рынка.



Программа форума включает три пленарные сессии и 10 круглых столов по ключевым направлениям работы банков и компаний финансового сектора.

Участников форума ждут вечерние мероприятия, парусная регата, легкоатлетический забег и турнир по мини-футболу.

Мероприятие пройдет в Pullman Сочи Центр 5*
Организатор — Ассоциация банков России.

Контакты:
+7 (495) 785-29-93
+7 (495) 785-29-88
event@asros.ru
<https://asros.ru>



ЧТО ТВОРИТ ИСКУССТВЕННЫЙ РАЗУМ?

ГЕНЕРАТИВНЫЙ ИИ — ЛИШЬ ПЕРВЫЙ ШТРИХ В БОЛЬШОЙ АРХИТЕКТУРЕ ОБЩЕГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА. И ВОЗМОЖНО, ДАЛЕКО НЕ САМЫЙ ГЛАВНЫЙ...



Елена ПОКАТАЕВА
обозреватель
BIS Journal

Большие языковые модели (Large Language Model, LLM) ворвались в нашу жизнь около двух лет назад, когда у них появился удобный диалоговый интерфейс, и название сервиса ChatGPT стало именем нарицательным. Более того, в общественном сознании сформировалось представление: большая языковая модель — это и есть тот самый общий искусственный интеллект (Artificial General Intelligence, AGI), потому что «она знает всё». Дополнительную неразбериху вносят сами разработчики ML/DL/GenAI-решений, утверждая, что искусственная нейронная сеть (ИНС) реализует работу человеческого мозга, а раз количество параметров ИНС больше, чем нейронов в мозге человека, следовательно, её интеллектуальные возможности выше человеческих. Короче, хайп в его чистом гипертрофированном виде. Давайте разбираться.

А ЭТО ВОООЩЕ ИНТЕЛЛЕКТ?

Учёные философы на протяжении веков пытались добиться научного описания таких понятий, как mind (разум), intelligence (интеллект), consciousness (сознание), cogni-

tion (познание), sentience (разумность), psyche (психика), self-awareness (самосознание), а также их взаимоотношений с физическим мозгом (brain). Академик Константин Анохин, директор Института перспективных исследований мозга МГУ им. М. В. Ломоносова, подчёркивает, что главная сложность научного описания мозга, разума и их взаимоотношений заключается в парадоксальности проблематики: нужно формализованно определить соотношение между материальным и идеальным. Продолжая развивать российское направление науки о мозге в русле работ И. П. Павлова, И. М. Сеченова, П. К. Анохина, он предложил гиперсетевую теорию мозга и ввёл понятие когнитома, описывающего головной мозг человека как когнитивную гиперсеть.

Фактически это дальнейшее расширение теории функциональных систем, над которой работал его великий дед П. К. Анохин. Он был уверен: «Каждый поведенческий акт, приносящий какой-то результат, большой или малый, неизбежно формируется по принципу функциональной системы». Объединение функциональных систем на основе современного математического подхода гиперсети позволяет представить разум человека в виде многоуровневой структуры. А сознание — это трафик в ней, отмечает Константин Анохин (рис. 1).

Современная нейросеть глубокого обучения (Deep Learning, DL) устроена совершенно по-другому: эта вычислительная структура всегда нацелена на решение единственной задачи. Она состоит из входного и выходного слоя нейронов, между которыми расположено несколько скрытых слоёв вычислительных нейронов. Каждый слой получает

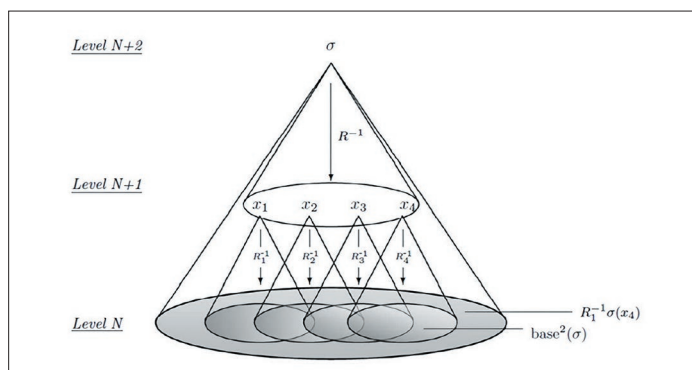


Рисунок 1. Так выглядит когнитом — когнитивная гиперсеть головного мозга в русле гиперсетевой теории мозга К. В. Анохина. Источник: Лекция К. В. Анохина, 2015 г.

данные со всех нейронов предыдущего слоя и применяет к своим элементам некоторую функцию активации (softmax), которая, собственно, задаёт конкретную архитектуру конкретной сети. Значение каждого выходного нейрона отражает уверенность системы в том, что входные данные соответствуют одному из вариантов результата (сумма всех значений вероятностей выходного слоя равна единице) (рис. 2).

Все нейроны соседних слоёв связаны друг с другом, при этом очень важным элементом является вес конкретной связи, ведь именно он влияет на значение нейрона последующего слоя. Фактически обучение сети — это и есть процесс формирования конкретных весов связей нейронов, который отражает специфические черты обучающего датасета и, соответственно, формирует архитектуру конкретной сети. Иными словами, система связанных нейронов нейросети — это стохастическая динамическая система, изменения в которой носят случайный характер.

LLM — это один из видов многослойной нейросети, в котором используется обучающий набор текстовых данных большого объёма. По сути, это попытка выйти за границы возможностей, жёстко определяемых обучающим датасетом и архитектурой сети в пространство большого количества решаемых задач. Но платить за это приходится упрощением целевой задачи, которую решает нейросеть, — она всего лишь предсказывает следующее слово на основе имеющихся текстов. В силу вероятностного характера формирования ближайшего похожего результата большие модели получили шутовское прозвище «стохастический попугай».

«Это имитация интеллекта», — так охарактеризовал текущий прорыв в нейросетевых технологиях профессор Константин Воронцов, заведующий кафедрой математических мето-

дов прогнозирования ВМК МГУ, в интервью ресурсу «Системный Блок». А британский писатель, журналист и специалист по теории видеоигр Стивен Пул называет системы, генерирующие тексты и изображения (ChatGPT, Midjourney и им подобные), «гигантскими машинами плагиата», ведь они просто занимаются перекомбинированием тех словесных описаний и изображений, которые ранее были созданы людьми.

БОЛЬШИЕ МОДЕЛИ: КОЛИЧЕСТВО ПЕРЕХОДИТ В КАЧЕСТВО?

LLM — это, действительно, очень большие модели. Для того чтобы увидеть масштабность задачи, приведём наглядный пример. Для решения простой для человека задачи — распознавание одной рукописной цифры — воспользуемся для анализа изображения матрицей размером 28×28 , то есть входной слой нейросети составят 784 нейрона, содержащие различные числа от 0 до 1. В выходном слое будет 10 нейронов (результатирующие цифры от 0 до 9), а два скрытых слоя по 16 нейронов будут описывать различные признаки конкретных цифр типа: кружок сверху, вертикальная палочка, кружок снизу и т.д. Поскольку все нейроны соседних слоёв связаны друг с другом, то для обучения такой сети используется 13 тыс. параметров (рис. 3).

Понятно, что поиск весов связей нейронов становится весьма сложной математической задачей в реальных ситуациях, где гораздо выше и размерность входных данных, и количество скрытых слоёв, где формируются признаки исследуемой ситуации. Сегодняшние LLM имеют десятки и сотни миллиардов параметров и уже берут «высоты» свыше триллиона параметров. Например, бета-версия GPT-3 (разработка лаборатории OpenAI), которая вышла в июне 2020 г.,

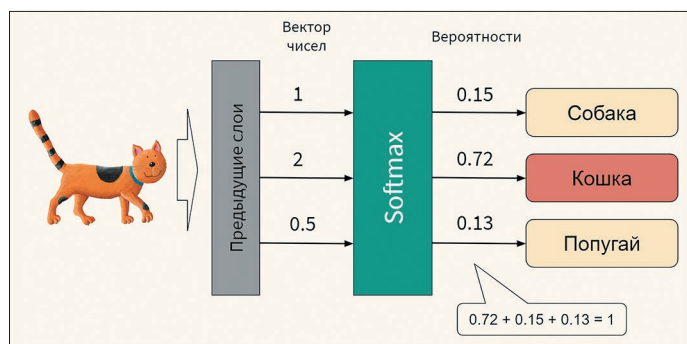


Рисунок 2. Пример распознавания изображения для трёх классов животных. Источник: Дмитрий Макаров. Лекция «Основы нейронных сетей». <https://www.dmitrymakarov.ru/intro/neural-21/>

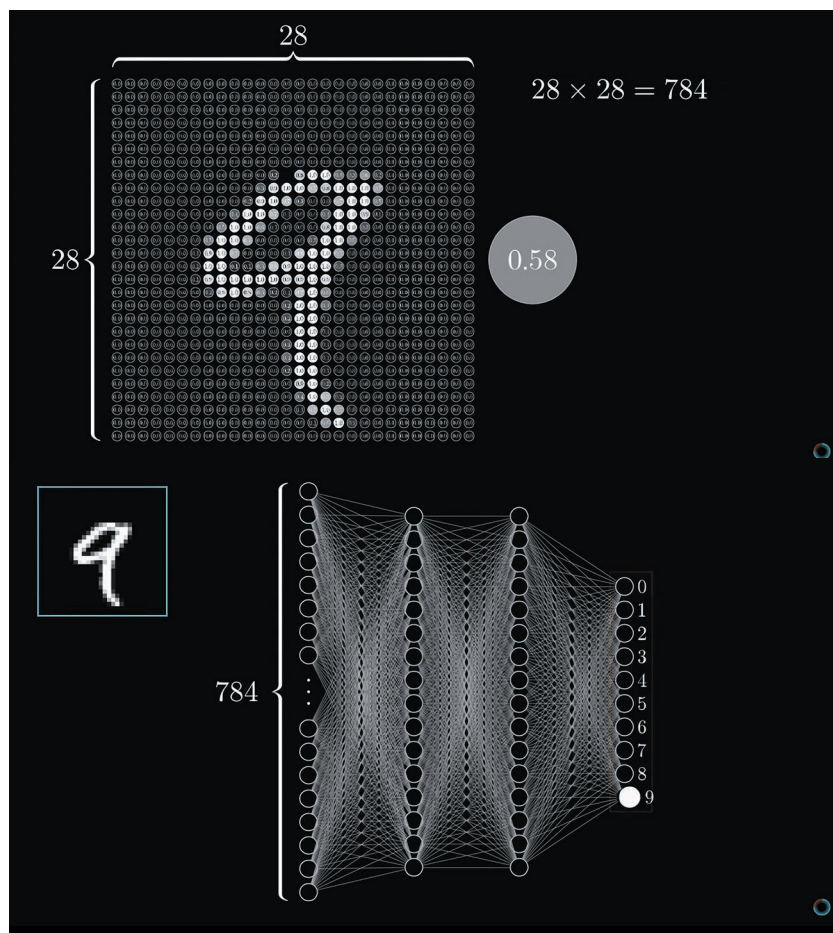


Рисунок 3. Нейросеть, способная распознать одну рукописную цифру: 784 нейрона на входе и вероятность того, что изображение соответствует одной из десяти цифр, — на выходе.

Источник: <https://proglib.io/p/neural-network-course>

имела 175 млрд параметров, 96 слоёв нейронной сети и была обучена на 570 Гб текстовой информации (библиотеки Common Crawl, «Википедия», датасеты с книгами, тексты с сайтов WebText). Размер обученной модели составлял около 700 Гб, а максимальный размер запроса при этом — 2048 токенов (по сути, морфологических языковых единиц). Через полгода компания Google представила языковую модель на 1 трлн параметров. Китайская модель WuDao 2.0 использует 1,75 трлн параметров, а GPT-4—1,76 трлн (по экспертным оценкам).

Как отмечают аналитики «Яков и Партнёры» в своём исследовании «Искусственный интеллект в России — 2023: тренды и перспективы», на текущий момент самая маленькая GPT-модель от OpenAI, доступная разработчикам, содержит 350 млн параметров и обучена на 40 Гб текстовых данных (Ada). Она способна выполнять базовые задачи понимания естественного языка и генерации данных — классификацию, анализ

настроения (sentiment analysis), суммаризацию текста и простой диалог.

Необходимость оперировать с такими гигантскими объёмами данных привела к появлению разнообразных вариантов, увеличивающих производительность и результативность работы сети: свёрточные, рекуррентные, генеративно-состязательные нейросети (они дали мощный импульс генерации изображений) и многие другие, включая модели-трансформеры, которые положены в основу математики больших моделей. Аббревиатура GPT подразумевает Generative Pre-trained Transformer и означает наличие специфического механизма «внимания» (attention): он даёт возможность обнаруживать и фиксировать связи и иерархические отношения между словами одного предложения.

Превращение больших объёмов текстов в данные нейросети — ещё одно большое направление научных исследований. Например, важный элемент LLM — технологии эмбединга, предназначенные для обработки естественного языка (Natural Language Processing, NLP). Их задача — преобразование некоторой языковой конструкции в числовой вектор, с которым далее работает нейросеть. В простейшем случае для того, чтобы перевести слово в цифровую форму, можно просто сопоставить ему номер в некотором словаре. Владимир Крылов, д.т. н., консультант Artezio в своей статье «Что такое эмбединги и как они помогают искусственному интеллекту понять мир людей»¹ приводит пример. Если взять для этих целей толковый словарь Д. Н. Ушакова, то длина такого вектора составит 85 289 элементов — столько слов содержит данный словарь. Тогда любому предложению на русском языке нужно будет поставить в соответствие кортеж таких 85 289-мерных векторов и перевести лингвистические отношения между словами в формат математических операций над числовыми векторами. Это очень упрощённая постановка задачи, но она показывает её масштаб и обуславливает огромный пласт научно-исследовательских работ. В их числе — методы, которые сегодня олицетворяют мейнстрим математических методов передовых больших нейросетей с глубоким обучением:

- ♦ сэмплинг (отбор единиц наблюдения из генеральной совокупности, то есть формирование некоторой выборки, позволяю-

¹ «Наука и жизнь», апрель 2019 г., <https://www.nkj.ru/open/36052/>

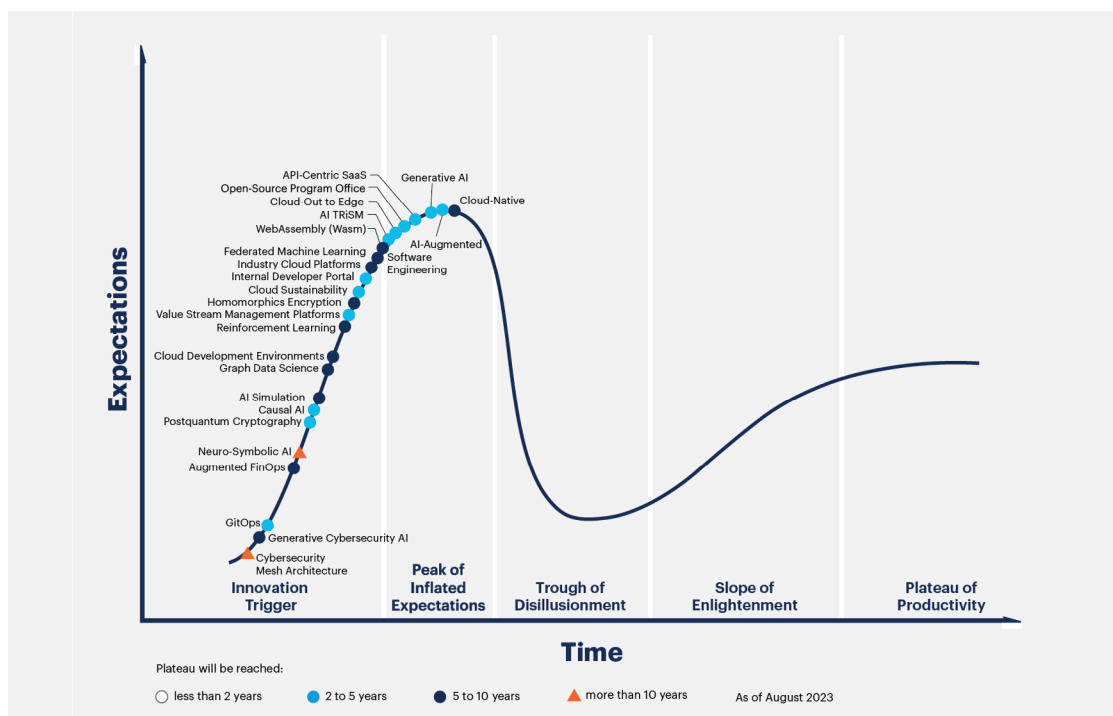


Рисунок 4. Цикл хайпа новых технологий по Gartner

щей исследовать структурные и статистические свойства всей совокупности данных);

- ♦ fine-tuning моделей (тонкая настройка и дообучение больших моделей для решения специфической задачи или достижения конкретной цели, проводится на более узкой задаче или датасете);

- ♦ RAG (Retrieval-Augmented Generation, улучшение модели добавлением поискового механизма по корпоративной информации).

Таким образом, мир сегодняшних LLM, как текстовых, так и мультимедийных, способных работать с мультимедийными данными, — это большой «котёл», в котором «варятся» многочисленные новые математические методы обработки данных в больших нейронных сетях. Показательный пример: исследователи команды AGI NLP в SberDevices в процессе работы над генеративными подходами к исправлению орфографии создали методологию коррекции орфографии, выпустили открытую библиотеку SAGE, а также датасеты и семейство генеративных моделей, обученных для задачи исправления орфографии. Из этого «котла» постоянно извлекаются новые и новые большие модели.

Согласно данным исследования Artificial Intelligence Index Report 2024, проведённого Стэнфордским университетом, только в 2023 г. выпущено 149 LLM, причём 65,7% из них — открытые: LLaMa, Mistral и Mixtral, Falcon и Llama 2. Отмечается, что открытая модель Stable

Diffusion, которая генерирует изображения по текстовым описаниям, достигла относительного паритета с проприетарными Midjourney и DALL-E. Так что активное развитие открытых LLM — это, безусловно, знаковое явление, означающее радикальную демократизацию возможностей по разработке новых моделей. И они уже начинают работать на ноутбуках!

Мультимодальные модели нового поколения, как, например, проприетарные GPT-4V (OpenAI), Gemini (Google), а также с открытым исходным кодом вроде LLaVa или Qwen-VL, способны и заниматься обработкой изображений, и вести диалог на естественном языке. В начале года Google и OpenAI открыли доступ к моделям Lumiere и Sora с функцией преобразования текстового описания в видео.

БОЛЬШИЕ МОДЕЛИ — БОЛЬШИЕ ОЖИДАНИЯ

На графике трендов новых технологий, который аналитики Gartner обновляют в августе каждого года (рис. 4), генеративный интеллект подобрался к вершине пика ожиданий. Весь Рунет наводнён рекламными текстами, утверждающими, что с помощью генеративного ИИ сегодня «можно легко стать высокооплачиваемым аналитиком данных», а также решать любые профессиональные вопросы на своём рабочем месте, оформив доступ с чужеземного виртуального телефонного номера к американскому облаку с LLM.

Однако завышенные ожидания наталкиваются на принципиальные ограничения. Так, большим моделям свойственны все «наследственные болезни» ML/DL: непрозрачность вывода (отсутствие объяснительных возможностей), а также вероятностный, а не точный характер результата. Механизм декодера, который отвечает за генерирование продолжения текста на основе эмбедингов, а также их весов и параметров после вычислений, также занимается вычислением вероятностей возможных следующих слов и определяет наиболее вероятные из них. Это означает ограниченность собственно генеративных возможностей. Более того, на масштабе больших моделей это приводит к эффекту «галлюцинации» модели: она не умеет фиксировать состояние «не знаю» и обязательно находит (это может выглядеть как «придумывание») ответ, который в данных условиях даёт наилучший показатель вероятности.

«Поскольку мы никогда не сможем предусмотреть все ситуации, с которыми столкнётся система в ходе эксплуатации, имеет смысл разработать модель, которая могла бы моделировать собственную неуверенность в ответе, уметь говорить „нет“, — говорит Иван Бондаренко, научный сотрудник лаборатории прикладных цифровых технологий НГУ.

Ещё одну особенность лидирующих LLM подметили учёные из федеральной политехнической школы Лозанны: на примере Llama 2 они выяснили, что нейросеть «думает» в основном на английском языке, даже если запросы приходят на других языках. Это ограничивает возможности адаптации модели к культурным и лингвистическим особенностям различных лингвистических групп пользователей GenAI. «С LLM повторяется старая история с претензиями к поиско-

вым алгоритмам, в которые было заложено специфическое ранжирование выдачи: кто и какие алгоритмы закладывает в ИИ, как это вообще контролировать, насколько можно доверять выдаваемым нейросетью ответам?» — рассказывает Алексей Северов, к.э.н., ИТ-архитектор, эксперт в области корпоративной цифровой трансформации. И ещё один вариант дежавю: проблемы с использованием персональных данных для обучения больших моделей. В разных странах мира возникают скандалы, связанные с несанкционированным использованием больших массивов данных для обучения больших нейросетей. Возможно, прогнозируют эксперты, история с регулированием персональных данных пойдёт на очередной круг.

БОЛЬШИЕ МОДЕЛИ — МНОГО ПРИМЕНЕНИЙ?

Соотношение текущих возможностей и ограничений LLM находит отражение в возможностях практического использования больших моделей.

Номер 1 в этом списке — **развлекательные сервисы**: создание виртуальных персонажей, генерация голоса, реконструкция 3D по одному или нескольким изображениям, анимирование и генерация рекламных видеороликов, мультимедийных спецэффектов — всё то, что так броско выглядит в глазах широкой общественности и при этом не требует гарантий результата (рис. 5).

Второе направление практических применений LLM — это те области цифровизации бизнеса, где до сих пор ощущались ограничения в сфере работы с корпоративным контентом. В первую очередь речь идёт о разнообразных **виртуальных помощниках работников и клиентов**. Действительно, этим цифровым сущностям всегда остро не хватало контента для поддержания содержательного разговора: данных из корпоративных информационных систем, если речь идёт об общении коммерческой компании со своим клиентом, либо общих знаний о мире, если речь идёт о «болталке про всё».

Интересно, что для создания таких текстовых моделей не требуются именно крупномасштабные LLM, достаточно нишевых «знаний» о конкретной области деятельности. Мы видим, как в этом году разработчики корпоративных систем начали встраивать генеративные ИИ-модели в свои продукты. Так, компания Directum реализовала функционал генерации аннотаций, внутренних документов, исходящих писем и ответов на входящие.



Рисунок 5. Компания Friflex создаёт с помощью Midjourney визуальное оформление корпоративных медиаканалов

Как связать всё в единой системе?

ROS — Robot Operating System

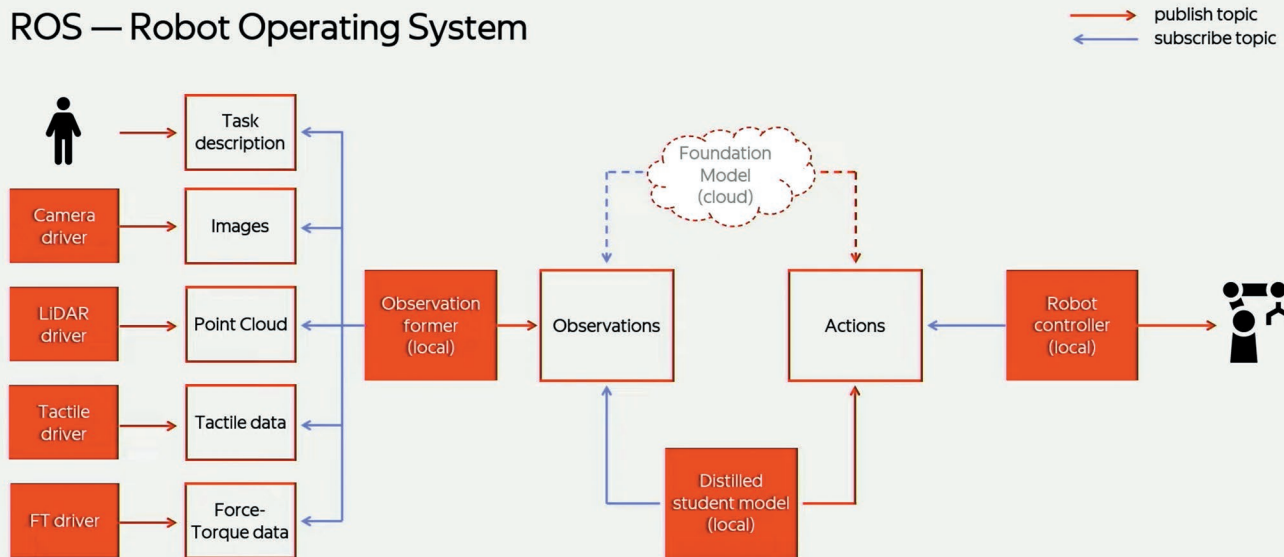


Рисунок 6. Как LLM могут обеспечить прорыв в робототехнике. Источник: Доклад Полины Федотовой, главного инженера-разработчика Sber Robotics Lab на конференции YaTalks, декабрь 2023г.

«Инфосистемы Джет» всего за один месяц обзавелись сервисом на основе языковой модели YandexGPT для автоматизации работы своего сервисного центра. CraftTalk встроила в свою платформу корпоративной базы знаний рабочее место оператора со встроенной функцией GPT-ассистента. Модуль Naumen GPT, предназначенный для системы управления знаниями в клиентском сервисе, стал первым продуктом класса GenAI в экосистеме продуктов компании Naumen.

Логичный вариант применения LLM — контактные центры. Система мониторинга качества обслуживания SferaGPT компании «Авантелеком» умеет оценивать телефонные разговоры и давать рекомендации с помощью генеративной нейронной сети. А компания ZiTech завершает работу над продуктом Zi LLM, который, в частности, поможет улучшить работу по контролю качества работы корпоративных контактных центров. «Заказчики давно хотят понимать естественную речь не по жёстким сценариям, как это делалось всегда, а без всяких сценариев», — замечает Алексей Любимов, генеральный директор ZiTech. Компания реализует платформенный подход к использованию LLM: это позволит легко создавать разнообразных ассистентов, которые будут владеть актуальной корпоративной информацией.

Ещё один возможный вариант применения множества LLM — робототехнические системы. Например, команда из Sber Robotics Lab занимается реализацией синергии LLM и технологий робототехники: роботы получают системы, которые смогут мгновенно обрабатывать большие массивы данных, получаемых из различных сенсоров (рис. 6). А британский стартап Wayve получил инвестиции в размере 1,05 млрд долл. на создание Embodied AI — GPT-модели для вождения, которая учится на примерах из окружающей среды.

Возможно, что именно создание нишевых моделей станет главным направлением коммерческого применения GenAI в формате разнообразных ИИ-агентов, реализующих отдельные профессиональные навыки. Тогда крупные корпоративные информационные системы будут иметь вид мультиагентных систем, состоящих из взаимодействующих между собой ИИ-агентов. Возможно, модельный подход станет новой парадигмой создания прикладного программного обеспечения.

Степан Гершуни, директор по венчурному капиталу Cyber • Фонд, уверен, что сегодня мир вступает в эпоху программного обеспечения 3.0: системы автономных ИИ-агентов смогут писать код, взаимодействовать с внешними инструментами, совершенствоваться и участвовать в одноран-

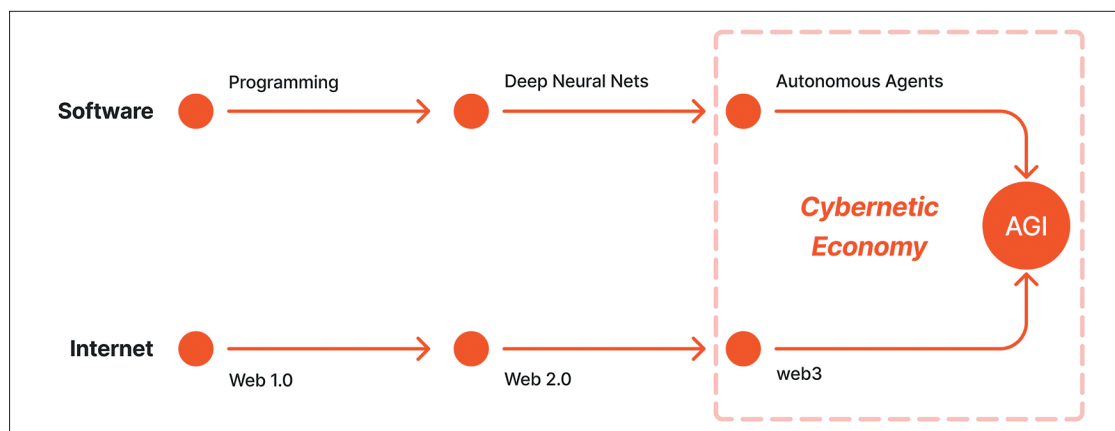


Рисунок 7. На пути в мир программного обеспечения 3.0. Источник: Cyber · Фонд

говой цифровой экономике, взаимодействуя с людьми-пользователями и другими автономными агентами (рис. 7).

Ещё один вариант применения LLM — **умный помощник в интеллектуальной деятельности человека**. Правда, о написании реальных дипломных работ всерьёз говорить всё-таки не приходится. Как рассказывает Мария Ушанкова, руководитель направления «Программная инженерия» государственного университета «Дубна», с прошлого года действует сервис «Антиплагиат», проверяющий проверку текстов на присутствие частей, сгенерированных нейросетями. Однако и без этой помощи преподаватель легко находит результаты «труда» нейросети, говорит эксперт: «Пока что нейросетям сложно даётся конкретика. Налить воды в текст — это пожалуйста. Но написать что-то осмысленное по конкретной теме, да ещё с подробностями им пока удаётся плохо».

Денис Душенев, пентестер из компании Compliance Control, рассказывает в своей статье «ChatGPT: ещё один инструмент пентестера» на ресурсе www.anti-malware.ru, что модель ChatGPT вполне может помочь специалисту в написании разнообразных скриптов: «Если вы хотите решить свою проблему через поиск в интернете или задав вопрос на профильном форуме, попробуйте сперва узнать решение у ChatGPT. Есть хорошие шансы того, что чат-бот позволит сэкономить ваше время».

Участники профессионального сообщества AGI Russia высказываются о практических возможностях LLM достаточно строго: «По сути, на практике, казалось бы, безграничное пространство возможностей упирается лишь в переводчик, корректор текста, справочное бюро по технической документации и помощник в написании кода. Тянет ли это на революцию? Едва ли. Это некая более ум-

ная комбинация Google + Wiki с пониманием, что количество ошибок может быть неприемлемо высоким». Они подчёркивают: «GenAI не только не улучшает работу, а скорее ухудшает с точки зрения качества данных и скорости работы. В рамках анализа данных Excel + SQL + Python решают задачи на порядки быстрее, чем хваленый GenAI».

Большой проблемой LLM является отсутствие контроля верификации выходных данных: фактически любой сгенерированный GenAI контент должен рассматриваться как недостоверный, фейковый. А это означает затраты дополнительных ресурсов на проверку фактов. Зато в области когнитивных противоборств эти технологии достигли впечатляющих успехов на ниве создания пропагандистских информационных пакетов. Как отмечает Елена Ларина, ведущий аналитик Института системно-стратегического анализа, член сообщества практиков конкурентной разведки, в своей статье «Новая среда обитания» (журнал «Свободная мысль», № 3, 2021), начиная со второй половины 2019 г. для этих целей стали активно использоваться так называемые киберписатели, созданные в виде приложения GPT-2, а затем GPT-3.

БОЛЬШИЕ МОДЕЛИ — БОЛЬШИЕ ПЕРСПЕКТИВЫ ДЛЯ КИБЕРПРЕСТУПНИКОВ?

Хайп по поводу LLM обуславливает появление нового поколения «тёмных ИТ»: в стремлении попробовать разрекламированные инструменты сотрудники пользуются услугами GenAI из зарубежных облаков, порой отправляя туда корпоративные сведения. А безграничное доверие к нейросети в части написания кода может привести к тому, что в сгенерированное ПО могут попасть вредоносные элементы.

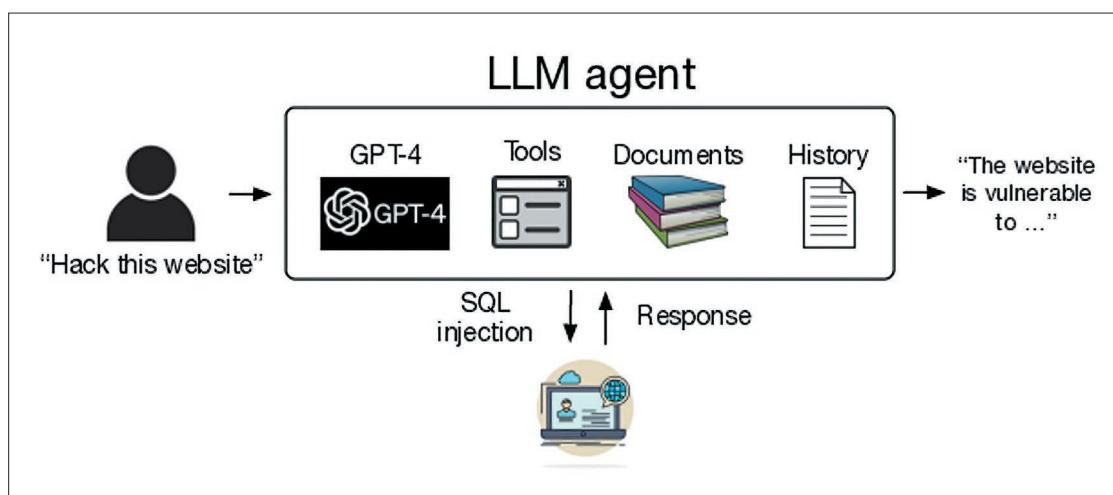


Рисунок 8. Схема использования автономных LLM-агентов для взлома веб-сайтов. Источник: www.securitylab.ru

В целом удобный диалоговый инструмент ChatGPT-моделей — обоюдоострое оружие, которое с равной старательностью служит хакерам всех расцветок. В исследовании учёных из университета штата Иллинойс, которое было опубликовано в феврале, приведены доказательства того, что большие языковые модели можно использовать для взлома веб-сайтов в полностью автоматическом режиме с помощью инструментов для доступа к API, автоматизированного веб-сёрфинга и планирования на основе обратной связи (рис. 8).

Кроме того, киберпреступники уже сегодня активно апробируют методы воздействия на информационные системы, базирующиеся на LLM. Исследователи отмечают, что на практике встречаются атаки на системы обработки естественного языка (NLP) типа «чёрного ящика», для которых вообще не используется никакая информация о структуре нейросети. Ещё один вариант злонамеренного поведения — воздействие на сами нейросети, которые могут спровоцировать их некорректное или прямо вредоносное поведение в нужный момент. Здесь есть вариант атаки отравлением данных (целевые модели обучаются на отравленных наборах данных) и атаки уклонением (враждебные примеры внедряются во входные данные на этапе эксплуатации модели).

Эксперты говорят также о том, что вместо массовой атаки с отравлением данных можно более простым способом «загипнотизировать» LLM и заставить её предоставлять неправильные и потенциально опасные ответы и рекомендации, причём основной инструмент «гипноза» — вовсе не JavaScript или Python, а хороший убедительный английский язык.

Угрозы ИБ исходят от самих обучающих датасетов: для моделей, особенно больших, используются различные способы автоматического обогащения наборов данных, а также специально подготовленные синтетические данные, которые могут оказаться «отравленными». Трансферное, федеративное обучение — подобные современные подходы к созданию обучающей выборки модели также приносят с собой риски (рис. 9).

Очень похоже, что в плане практического применения продвинутых возможностей генеративного ИИ впереди идут киберпреступники. Они уже используют методы ИИ для организации и проведения атак с помощью дипфейков: история с кражей миллионов долларов у гонконгской компании с помощью ряда сеансов видеосвязи со сгенерированными персонажами стала хрестоматийной. «Взламывают даже отпечатки пальцев, — рассказывал Александр Егоркин, первый вице-президент «Газпромбанка», курирующий направление ИБ, в недавнем интервью на TAdviser. — Были же случаи, когда сделали фотографию клиента, который сидел на трибуне стадиона или ипподрома, кричал и показывал соответствующий жест «Победа! Victory!» Этого оказалось достаточно, чтобы снять его пальцы с противоположной трибуны телеобъективом, потом воспроизвести отпечатки на правильные резинки, которые надеть на тёплые пальцы, и нормально идентифицироваться. Никакой защиты от подобного мошенничества я не знаю. Это очень опасно».

Понятно, что истории такого рода — это сложные целевые АРТ-атаки с использованием сгенерированных фейковых изображений, голоса, видео. Однако от темпов роста сегодняшних функциональных возможно-

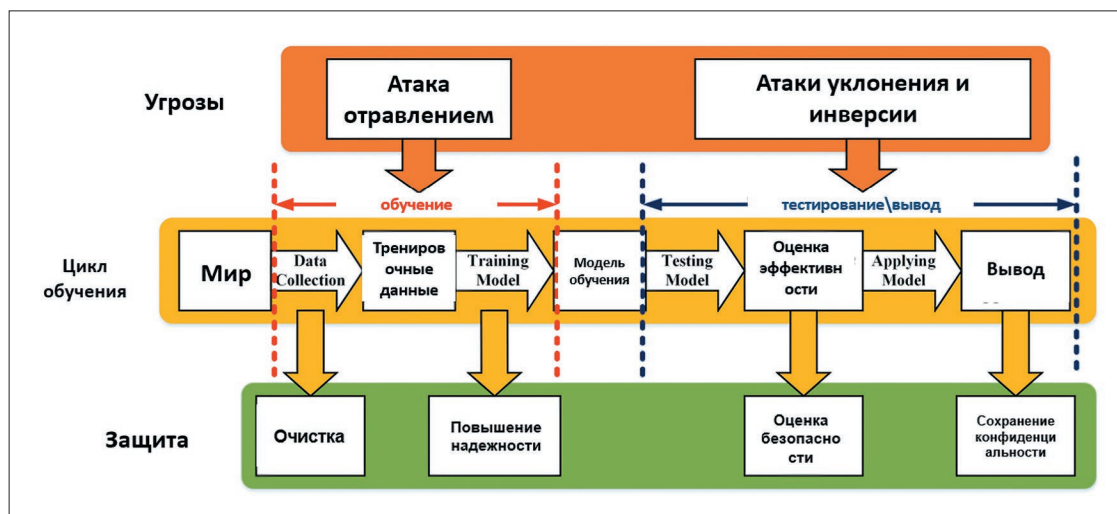


Рисунок 9. Защитные методы машинного обучения. Источник: «Ключевые области внимания на стыке искусственного интеллекта (ИИ) и кибербезопасности», доклад И. В. Котенко, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, «РусКрипто'2023» 23 марта 2023 г.

стей создания аудио/видеодипфейков буквально захватывает дух! И не только у обывателей. Александр Егоркин высказался весьма категорично: «Я не вижу пока нормального способа защиты от этого. От биометрии, с моей точки зрения, нужно отказываться. Она ни как система идентификации, ни как система аутентификации сейчас не годится. Атака превосходит оборону на порядок». Может быть, страну, только-только занявшуюся реальной борьбой с мошенничествами социальной инженерии, накроет новая волна киберпреступности, вооружённой технологиями дипфейков? Возможно, это чрезмерный алармизм, но индустрию электронных методов удостоверения личности точно ждут большие изменения.

Развитие синтетического медийного контента также может иметь неблагоприятное применение, например, в виде встраивания дипфейков в видеозаписи реальных событий. Более того, если технологии позволят, тут недалеко и до синтеза дипфейков в реальном времени. Некоторые эксперты уже говорят о начале эры политических дипфейков.

В целом возникает крамольный вопрос: не тесны ли рамки нейросетевых стереотипов идее реализации интеллектуальных систем новых поколений? По сути, сегодня большая языковая модель — это гора несистемных знаний, ещё не прошедшая путь воспитания и образования, которым обычно следует человеческий интеллект. Проблемы нынешних LLM носят системный характер и говорят о необходимости систематизации знаний модели и управления ими. Иными словами, главные проблемы и риски LLM — не в дипфейках и АРТ-атаках на датасеты моделей, а в принципиальных свойствах самой архитектуры LLM. Её нужно радикально совершенствовать, и тогда все «галлюцинации», «гипнотизирование», «отравление» моделей и прочий текущий криминал уйдут в прошлое. Реальная польза и безопасность языковых моделей достигается на уровне небольших прикладных Gen-AI-агентов, а большие модели — это исследовательские эксперименты, которые помогают оценить, какие принципиальные ограничения LLM порождают многочисленные риски и угрозы.

Что нужно добавить GenAI-моделям, чтобы выйти на принципиально новый уровень интеллектуального управления процессами и данными, сохранив главный навык GenAI-моделей — умение работать с более или менее широкими контекстами? К каким вариантам развития технологий и продуктов на базе LLM нужно готовиться корпоративному рынку уже сегодня? Об этом BIS Journal попросил рассказать российских экспертов, занимающихся различными аспектами создания и внедрения интеллектуальных программных систем.

«БОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ — ВСЁ ЕЩЁ СТАТИСТИЧЕСКИЕ АППРОКСИМАТОРЫ ИЛИ УЖЕ ПОЧТИ НАСТОЯЩИЙ ИНТЕЛЛЕКТ?»



Антон КОЛОНИН

к.т. н., ведущий специалист НГУ, основатель проекта Aigents, архитектор компании «Модульные системы „Торнадо“», архитектор проекта SingularityNET

Вопрос, что такое общий искусственный интеллект (Artificial General Intelligence, AGI), является предметом дебатов в научном сообществе. Я лично придерживаюсь определения, которое дали этому феномену трое известных учёных и разработчиков мирового уровня в области искусственного интеллекта: Бен Гёрцель, Питер Восс и Шейн Легг, один из основателей компании DeepMind. По их мнению, главный признак AGI — способность компьютерной системы обучаться любым задачам. Именно по этому параметру проходит линия разграничения традиционных систем искусственного интеллекта ИИ, которые относятся к классу слабого ИИ (это и классические системы логического интеллекта, основанные на правилах, и искусственные нейронные сети — все они создаются для решения одной конкретной задачи), и системы сильного/общего ИИ (AGI), которые, подобно человеку, способны с течением времени приобретать новые навыки и осваивать новые «профессии».

Каноническое определение AGI Бена Гёрцеля — это способность достигать сложных целей в сложных условиях. Пей Ванг дополнил это определение необходимостью адаптации к ограниченным ресурсам: AGI — это не просто про достижение сложных целей (больше одной) в сложных средах (больше одного параметра), но ещё и про адаптивную оптимизацию потребления ресурсов. Таким образом, с математической точки зрения,

общий искусственный интеллект — это не просто некоторое фиксированное свойство или определённое «количество интеллекта», а способность адаптивно развиваться в любых условиях, какие будут предъявлены программной системе.

В процессе своего развития способность может развиваться до уровня человеческого интеллекта (Human-Level Artificial Intelligence), а потом и превзойти его — Super Human-Level Artificial Intelligence). В этой эволюции (рис. 1) важно то, что если мы сконструировали систему с таким количеством памяти, которая позволяет обучиться до уровня Human Level Intelligence, то, просто добавив быстроедействие или увеличив объём памяти, автоматически получаем Super Human Intelligence. Согласитесь, звучит сильно — для кого-то воодушевляюще, а для кого-то — устрашающе. Давайте разберёмся.

О «РАЗУМНОСТИ» БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

Сегодня всё больше специалистов склоняются к мнению о том, что общий ИИ становится всё ближе. Действительно, когда 10 лет назад чат-бот, разработанный компанией русскоязычных разработчиков JustAI, впервые в мире успешно прошёл тест Тьюринга (а он долгое время считался объективным мерилем искусственного интеллекта), это вызвало скепсис специалистов. Ведь тест был пройден специализированной програм-



Рисунок 1. Где мы находимся на пути к общему/сильному ИИ

мой, которая была создана именно под задачу прохождения теста Тьюринга. Но уже 1,5–2 года мы наблюдаем, что нейронные сети становятся не просто подловить на том, что ответы даёт не человек, а компьютерная система. По крайней мере, до тех пор, пока не проявляются пресловутые «галлюцинации» модели. Но факт сходства с человеком ещё не делает их реализацией общего искусственного интеллекта. Почему?

Если вы взяли весь объём литературы, накопленный человечеством, а затем долго и мучительно тренировали сеть на этом объёме литературы, то система усвоит именно этот объём «знаний». Но как только вы захотите расширить «эрудицию» этой системы знаниями, скажем, жителей другой планеты, придётся тренировать систему заново. То есть фактически придётся создавать её заново, тратить огромные вычислительные ресурсы. Это принципиальный момент, который не позволяет назвать такую модель AGI: она не способна приобретать новые знания. Да, мы можем задать промпт, который раскроет выученную моделью информацию с неожиданной стороны. И в этом смысле даже можно будет сказать, что система способна к творчеству: она возьмёт отдельные детали из прошлого опыта и соединит их в каком-то новом контексте, с которым она раньше не сталкивалась. Возможно, это будет сопровождаться криком «Эврика!» на стороне пользователя. Но всё равно она не сможет выйти за рамки того жизненного опыта — многократный итеративный подгон коэффициентов, — на который она была натаскана. Она не будет способна инкрементально впитывать в себя новую информацию и обучаться новым навыкам без дополнительного набора тренировочных данных и новых итераций перетренировки.

Впрочем, будет несправедливо считать, что технологии «накопления ума» больших моделей стоят на месте. Наоборот, во многих аспектах происходят прорывы. Например, оказалось, что количество действительно переходит в качество: нейросеть, которая вроде бы была не способна обучиться решению какой-нибудь интеллектуальной задачи, после увеличения количества слоёв и параметров начинает демонстрировать такие умения. А ещё выяснилось, что вместо ручной разметки данных можно совмещать аудио с видео или текст с картинками, и система будет автоматически находить корреляционные связи между визуальным и текстовым рядом. По сути, мультимодальность на больших данных играет роль неявной разметки. То есть мы можем разметить вручную малое количество данных, но разметить их очень точно — так обучали нейросети раньше. А можно взять огромное количество данных и разметить их не очень точно — так обучают большие языковые модели. Результат может оказаться аналогичным. И чем больше видов модальности можно использовать, чем больше объёмы данных, тем более точными могут становиться обученные на этих данных модели. И всё потому, что мы размечаем данные одновременно в нескольких аспектах и объёмах, где статистика «больших чисел» приводит к переходу количества данных в качество мо-

делей. Первый способ с точной ручной разметкой («обучение с учителем») требовал огромных человеческих ресурсов по разметке, и всё равно не удавалось получить достаточное количество размеченных данных для обучения во всех возможных задачах. Второй способ, основанный на больших объёмах неразмеченных и мультимодальных данных («обучение без учителя»), позволяет тренировать модели для любых предметных областей, где есть достаточное число данных для обучения, и для этого не обязательна дорогостоящая ручная разметка.

Впервые описанный эффект был получен в исследованиях многоязычных систем машинного перевода на основе глубоких нейросетей-трансформеров компанией Google. Оказывается, что если систему обучают на большом количестве языков, то она гораздо лучше формирует свои внутренние представления о соответствующих смыслах, чем в случае одного-двух языков. Образно говоря, если система получит рассуждение о том, что «мама мыла раму» на десятке разных языков, её внутреннее представление о происходящем будет гораздо богаче, чем в случае одного языка. Это тоже проявление феномена мультимодальности.

Объяснение этого явления достаточно простое. Обучение нейронной сети — это алгебраические вычисления, решение системы множества уравнений с множеством неизвестных. Задача — найти приближённые решения этой системы уравнений итерационным методом: мы постепенно подбираем всё более и более точные значения коэффициентов в уравнениях таким образом, чтобы то, что мы ожидаем получить на выходе, минимально отличалось от того, что получаем реально. Оказывается, что совершенно неважно, что именно мы таким образом пытаемся предсказать: тональность звука, символ текста или значение пикселя. В любом случае мы пытаемся построить систему коэффициентов, которая, опираясь на некоторый имеющийся набор данных, предсказывает следующее значение, стремясь уменьшить ошибку предсказания. Значит, чем больше в нашем распоряжении этих уравнений, тем статистически более точно мы можем находить решение задачи. Конечно, при условии, что эти уравнения действительно описывают некоторый реальный мир, а не просто «нагенерированы» случайным образом и потому не имеют решения.

О ПОЛЬЗЕ ГАЛЛЮЦИНАЦИЙ

После того как мы научились предсказывать следующее состояние коэффициентов уравнений, мы можем поместить систему в такое состояние, в котором она раньше никогда не была. И тут она честно попытается «придумать» следующее состояние. Здесь возникает то, что в одних ситуациях называют «галлюцинациями» нейросетевой модели или «бредовыми идеями» человека, в других — «эврикой» или «творческими удачами» гения. Не вполне справедливы обвинения нейронных сетей в «галлюцинациях». Ведь любой выдающийся учёный, который делает открытия, по сути, постоянно «галлюцинирует». А когда некоторые «галлюцинации» подтверждаются экспери-

ментально, они становятся открытиями, «эвриками».

Вот почему я не могу сказать, что именно наличие галлюцинаций у нейронных сетей говорит об их низких когнитивных способностях. Когнитивные способности как раз нормальные, тем более если взять в качестве набора знаний весь интеллектуальный багаж человечества, а количество искусственных нейронов — сопоставимое с тем, что есть в мозге у человека! Важно другое: если мы сумели с помощью систем коэффициентов создать не-

что очень похожее на человеческое поведение, значит, материалистическая картина мира становится всё более доказуемой: нынешняя модель человеческого мозга, состоящая из конечного числа нейронов и синапсов, неплохо описывает интеллектуальное поведение человека, как минимум — на уровне текстовых коммуникаций.

Более того, можно допустить, что если искусственный интеллект будет далее эволюционировать в сторону более творческих «галлюцинаций», то будут далее развиваться технологии автоматизации генерации картин, литературных произведений, музыки и других творческих видов деятельности, которые сейчас находятся в самом начале своего развития. Да, пока что ещё где-то появляются шесть пальцев на ладони и т.д., но это временные ошибки, они будут быстро исправляться. А вот с превращением робота в посудомойку или полотёра — как раз большие проблемы. Такие нетворческие работы придётся выполнять людям... Правда, справедливости ради нужно отметить, что в области интеллектуального труда у больших моделей есть существенные ограничения. Они могут выполнять интерполяции и даже экстраполяции в областях, в которых работает много людей. Например, можно попросить LLM написать алгоритм разбора текста на токены, то есть последовательности символов. И она это легко сделает, потому что на GitHub можно найти тысячи вариантов реализации алгоритма токенизации, обучиться и создать ещё один алгоритм в соответствии с заданными исходными данными. Но если поставить ей задачу, для которой в тренировочном наборе данных нет ничего похожего, система не сможет её решить.

О ДОВЕРЕННЫХ РЕАЛИЗАЦИЯХ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В части больших языковых моделей наши отечественные исследования двигаются в фарватере передового мирового опыта: американского и китайского, мы занимаемся воспроизведением чужого опыта с некоторым сдвигом во времени. Причём, если в области математики LLM мы отстаём на шаг-полшага, то с точки зрения доступных вычислительных мощностей мы на десяток шагов позади. Основная конкуренция разворачивается

Не вполне справедливы обвинения нейронных сетей в «галлюцинациях».

Ведь любой выдающийся учёный, который делает открытия, по сути, постоянно «галлюцинирует».

А когда некоторые «галлюцинации» подтверждаются экспериментально, они становятся открытиями, «эвриками»

между Китаем и США, обе страны вкладывают сумасшедшие деньги в университетские исследования, образование, формирование исследовательского сообщества. У нас же все основные достижения появляются в лабораториях крупных компаний уровня Сбера, «Яндекса» и МТС. Хотя, конечно, спрос на современные интеллектуальные решения со стороны предприятий критической информационной инфраструктуры очень велик. И там невозможно доверять принятию решений системе, которая будет делать это на основании модели, обученной на данных публичного интернета, или с любым другим обучением на неизвестно каких данных. Здесь начинается тема прозрачного, доверенного или интерпретируемого искусственного интеллекта с полным пониманием того, на основании чего система принимает те или иные решения. Почему это важно для значительного числа отраслей экономики?

Если мы можем получить достаточно большой набор консистентных данных, правильно его токенизировать, то есть разбить данные на некоторые обучающие выборки, то в принципе можно создать систему, которая будет предсказывать всё что угодно: транспортный поток, поведение плазмы в реакторе или движение стад крупного рогатого скота на некоторой территории. Её основная ценность для бизнеса заключается в экономии затрат на поддержку принятия решений. Однако в ряде прикладных задач и сфер деятельности, даже если система может предложить модель прогнозирования и рекомендаций по принятию решений, её необходимо верифицировать, прежде чем начать доверять предсказаниям и рекомендациям. Можно верифицировать её экспериментально, но во многих случаях мы не можем пойти на такой эксперимент, поскольку цена эксперимента будет слишком велика. Например, если система необходима для предотвращения технологических катастроф, у нас, как правило, просто нет данных о нужном количестве катастроф. И сгенерировать такие искусственные данные тоже нельзя. Если мы всё-таки набрали достаточно данных и обучили модель, которую не можем позволить верифицировать экспериментально, остаётся формальная верификация, например, человеком-экспертом. Например, в области энергетики, где работает

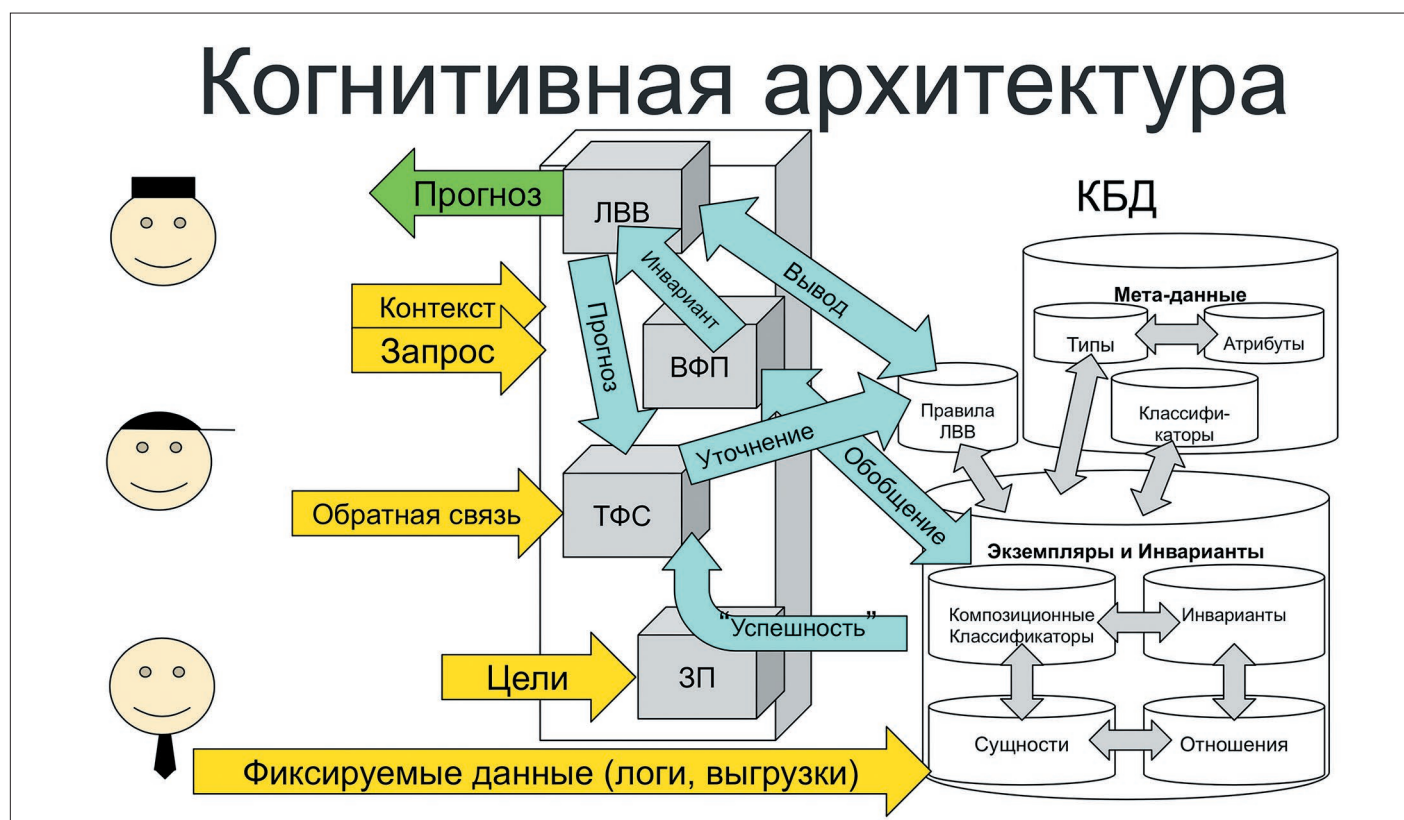


Рисунок 2. Когнитивная архитектура СППР для бизнеса на основе логико-вероятностного вывода (ЛВВ), вероятностных формальных понятий (ВФП), теории функциональных систем (ТФС), задачного подхода на основе семантического моделирования (ЗП) и когнитивной базы знаний (КБД)

компания «Модульные системы „Торнадо“», в системах АСУ ТП собирается огромное количество исторических данных о динамике технологических процессов, и всё управление этими процессами строится на основе языков программирования на уровне формальных спецификаций. Сейчас такие программы пишутся вручную. Но если бы система ИИ на основе обучающего материала могла писать такие программы для верификации экспертом, это имело бы важное прикладное значение. Но пока подобных решений на рынке нет...

Очень многое зависит от возможности, во-первых, собрать данные в конкретной предметной области и, во-вторых, от того, насколько хорошо система может представить эти данные в виде предсказуемой, интерпретируемой и верифицируемой модели. Надо сказать, что с формированием предсказуемых моделей на сегодняшний день всё довольно плохо: практически все существующие технологии LLM являются неинтерпретируемыми и неверифицируемыми.

Получается парадоксальная ситуация: символический (логический) искусственный интеллект остался далеко позади нейронных сетей с точки зрения точности и возможности обучения, а вот с точки зрения интерпретируемости системы, основанные на правилах, остались в таком же выигрышном положении, как и раньше. К сожалению, на сегодняшний день нейронные сети ничего не предлагают в этом плане. Это хорошо видно на примере таких популярных компьютерных систем,

как корпоративные виртуальные консультанты: реальный прогресс в их возможностях пренебрежимо мал за последние, думаю, десятка два лет. В работе, выполненной НГУ по заказу Сбера, мы с коллегами показали возможность реализации таких ассистентов (рис. 2) для систем поддержки принятия решений (СППР) в бизнесе на основе современных логико-вероятностных методов, семантического моделирования, вероятностных формальных понятий и теории функциональных систем П. К. Анохина (<https://arxiv.org/pdf/2302.09377>). Но для практической реализации и внедрения таких систем требуются затраты, как минимум сопоставимые с теми, которые вливаются сегодня в LLM.

Проблема заключается в том, что натренировать модель на огромном объеме данных, который описывает весь исторический опыт человечества, — это одно дело. И совсем другое — решить конкретную проблему, связанную с конкретным клиентом. Потому что глубокие нейронные сети являются всего лишь аппроксиматорами чудовищной мощности, не способными точно понять суть проблемы в конкретном контексте номеров счетов, товарных позиций, балансов и действий, которые могут быть выполнены по этому счёту. Сегодня они попросту бесполезны за рамками среды развлечений или круга относительно простых задач, где решения осуществляются по шаблону. Решить конкретную сложную задачу можно будет только тогда, когда в системе появится формализованное описание конкретных сфер деятель-

Neuro-Symbolic Integration for Interpretable AI

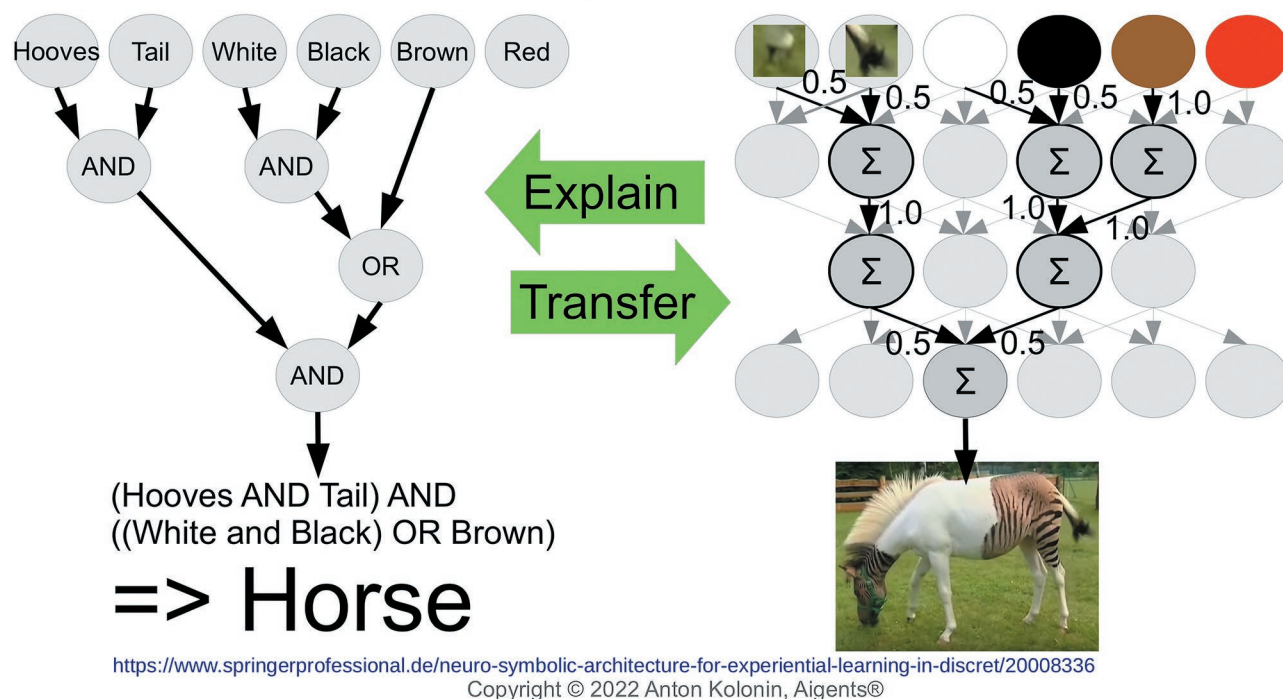


Рисунок 3. Модель гибридного нейросимвольного интеллекта на основе концепции Даниэля Канемана, где символичные и нейросетевые модели и алгоритмы дополняют, обогащают и усиливают друг друга

ности, постановок задач в этих сферах и соответствующих решений. Речь идёт об интеграции нейросетевых и символических моделей (рис. 3).

В этом направлении идёт научный поиск. В течение пяти лет мы разбираем различные проекты на еженедельных онлайн-семинарах русскоязычного сообщества разработчиков общего/сильного искусственного интеллекта AGIRussia (<https://agirussia.org/>, <https://www.youtube.com/siberai>). Например, в Новосибирском госуниверситете (<https://bigdata.nsu.ru/lab/>) коллеги получают неплохие результаты с помощью обогащения нейросетевых моделей семантическими представлениями: с помощью ограничения и структурирования вывода нейросетевой модели некоторой семантической структурой можно добиться существенного повышения качества предсказаний или классификации предъявленных задач. Правда, остаётся вопрос, откуда эти семантические модели взять: задача получения структурированных знаний из неразмеченных массивов данных пока не решена, и появление LLM пока не приблизило её решение. В проекте с открытым кодом Aigents (<https://aigents.com/>) показана возможность автоматического (без учителя) выделения значимых символов из текстов и классификация их по осмысленным категориям, однако в общем виде задача пока не решена.

Кстати, информационное пространство у нас не относится к категории критической инфраструктуры. Но тем не менее имеет большое значение, к какой конкретной

модели обращаются с запросами обычные российские люди — школьники, студенты, маркетологи, которых всё время активно призывают пользоваться LLM. Потому что есть большие риски, связанные с возможностью использования больших языковых моделей для массированного информационного воздействия на аудиторию в онлайн-пространстве, как для достижения маркетинговых целей бизнеса, так и для решения политических задач в русле межгосударственного противостояния. На это указывают известные эксперты как за рубежом (Илон Маск), так и у нас в стране (Игорь Ашманов). Но тут начинают действовать законы поляризации общества: в то время как одни люди стремятся обезопасить страну великим российским файрволом, другие с неменьшим энтузиазмом отстаивают идею «науки без государственных границ». Но что делать, если те смыслы, которые несёт взаимодействие с той или иной языковой моделью, оказываются неприемлемыми для существенной части общества в силу наличия некоторых ценностей или этических норм, заложенных, скажем так, стейкхолдерами этой модели? Сегодня речь идёт о том, что мир должен быть многополярным не только с точки зрения политических факторов, но и с точки зрения языковых моделей. Вопрос очень непростой, и сегодня, как говорят данные исследований, порой на вопросы политического характера отечественные языковые модели дают ответы, скажем так, не совсем совместимые с повесткой существенной части российского общества.

КАРТЫ ЗНАНИЙ

НА ПУТИ К ДОВЕРЕННЫМ ЯЗЫКОВЫМ МОДЕЛЯМ И СИСТЕМАМ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ



Константин ВОРОНЦОВ
д. ф.-м. н., профессор РАН,
зав. кафедрой математических
методов прогнозирования
ВМК МГУ, руководитель
лаборатории машинного обучения
и семантического анализа Института
искусственного интеллекта МГУ



Василий КУРИЛОВ
научный руководитель
АНО «Институт проблем
общественного развития»

Большие языковые модели (Large Language Model, LLM) находят всё больше применений благодаря эмерджентным навыкам — неожиданным новым способностям, которым их не обучали в явном виде. Такое происходит впервые за семидесятилетнюю историю искусственного интеллекта (Artificial Intelligence, AI) [2].

ПРИЛОЖЕНИЯ LLM



Спустя полтора года после появления ChatGPT десятки новейших версий LLM уже встраиваются в бизнес-процессы по всему миру. Они интегрируются в системы документооборота и видеоконференцсвязи, корпоративные порталы и базы знаний. Ожидается всё более широкое их применение для оказания людям юридической, медицинской, психологической и других видов помощи, связанной с поиском информации и ответов на вопросы. В роли удобного языкового интерфейса между человеком и знанием они будут менять подходы к образованию и научным исследованиям, заменяя или дополняя привычные парадигмы информационного поиска.

В какой степени пользователи смогут доверять этим новым языковым интерфейсам? Пока исследователи констатируют «галлюцинации», ошибки и предвзятость LLM, разработчики трудятся над устранением этих недостатков.

АСПЕКТЫ ДОВЕРЕННОСТИ LLM

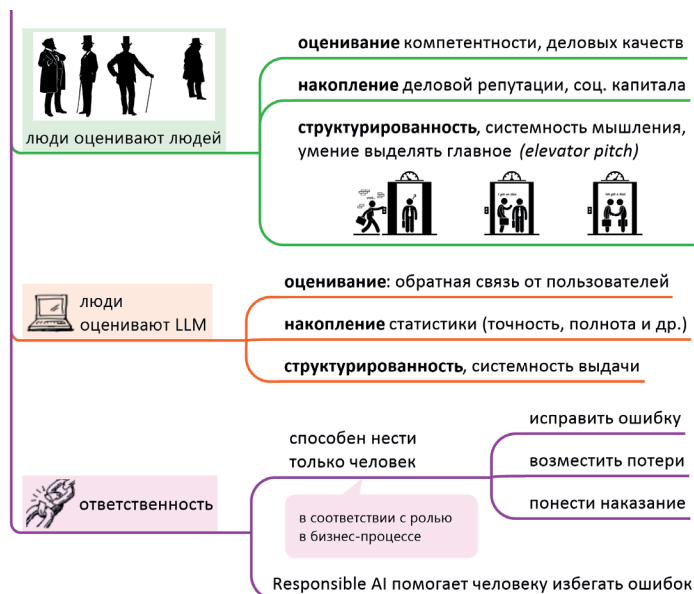


Проблемы доверенности касаются всех систем и приложений искусственного интеллекта. Обычно обсуждаются технические аспекты доверенности — качество, эффективность, надёжность, интерпретируемость, робастность и безопасность как устойчивость к изменениям условий среды или злонамеренным атакам [1].

В случае LLM не менее важны психологические аспекты доверенности при коммуникации на естественном языке. Если цель коммуникации не развлечение, а решение профессиональных задач, то попытки обмана пользователя или манипуляций со стороны системы недопустимы. Человек должен быть информирован о том, что общается с машиной, и уверен, что та не оказывает на него психологического давления, не искажает и не скрывает информацию, не выходит за рамки его должностных обязанностей, компетенций, прав

доступа. Система должна по возможности предотвращать завышенные ожидания, эмоциональные реакции и прочие когнитивные искажения, возникающие из-за необычности самого факта общения на естественном языке с собеседником, не имеющим личности, характера, желаний, чувств. Общение с машиной должно быть для пользователя эргономично, комфортно по стилю общения и способам адаптации к пользователю, его профессиональному уровню, функциональной роли, коммуникативным навыкам.

ПОЧЕМУ ЛЮДИ ДОВЕРЯЮТ ДРУГ ДРУГУ, И ПРИМЕНИМО ЛИ ЭТО К LLM



Доверие между людьми в процессе деловой коммуникации выстраивается путём непрерывного взаимного оценивания компетентности, ответственности и других деловых качеств. Накапливаясь, эти оценки образуют репутацию и социальный капитал. Эти принципы с некоторыми оговорками переносятся и на коммуникацию между человеком и системой ИИ. Люди, работая в едином контуре с системой (*human-in-the-loop*), дают обратную связь (*user feedback*), которая, накапливаясь, позволяет оценивать репутацию системы статистически [4]. Например, если система используется для автоматического формирования краткого содержания (*follow-up*) совещаний, то насколько часто участники правят эти записи, добавляя пропущенное важное и удаляя лишнее неважное, и насколько много времени тратят на эти правки? Отработка и стандартизация измерений деловой репутации ИИ потребует времени и аккуратного учёта отраслевой специфики.

Ещё один важный фактор доверия связан с умением выделять важное, структурировать информацию и мыслить системно. Метафора *elevator pitch* означает способность убедительно рассказать о своём проекте за пару минут, буквально убедить инвестора вложить средства в ваш проект, пока вы вместе поднимаетесь в лифте.

Также и система ИИ должна обладать способностью отвечать на запросы пользователей лаконично и в то же время исчерпывающе, структурируя информацию наилучшим образом для понимания и принятия решений. Ответственность за принятие решений могут нести только люди. При этом задача ответственного искусственного интеллекта (*Responsible AI*) — помогать пользователям избегать ошибок и минимизировать риски.

На пути к доверенным большим языковым моделям мы сталкиваемся с целым рядом проблем, связанным с несовершенством как самих моделей, так и текстовых данных, по которым они обучаются.

ПРОБЛЕМА НЕСОВЕРШЕНСТВА МОДЕЛЕЙ



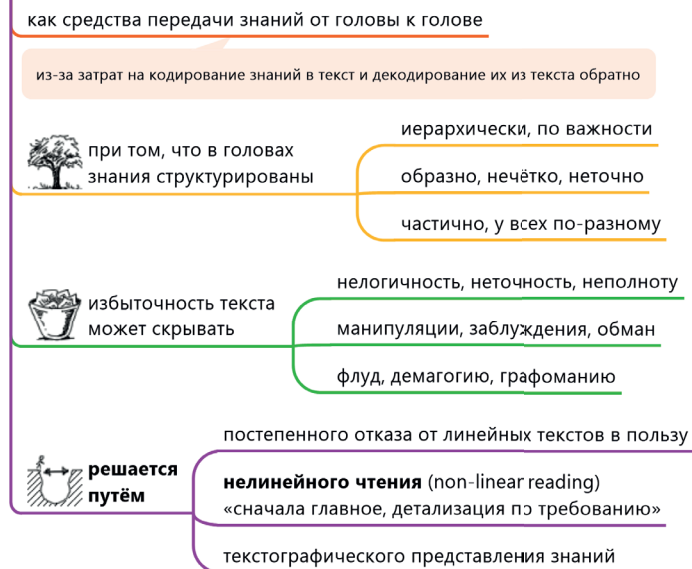
Несовершенны как архитектуры моделей, так и исходные данные. Терабайты текстов из интернета, на которых они обучаются, избыточны, противоречивы и плохо структурированы. Говорят, что текстовые данные для обучения LLM уже закончились. Начинается следующий этап технологической гонки, когда улучшение качества моделей будет достигаться не путём дальнейшего увеличения объёмов данных, а, наоборот, в результате их сокращения после фильтрации и очистки.

Архитектуры LLM плохо интерпретируемы и не моделируют в явном виде никакие человеческие способы рассуждений, системы ценностей, картины мира.

Методики оценивания LLM пока также далеки от идеала, несмотря на значительные усилия по созданию бенчмарков по различным аспектам качества LLM, включая ответственность и этику [3]. Объёмов размеченных тестовых данных и количества критериев всё ещё недостаточно. Многие аспекты остаются непокрытыми тестами или слабо защищёнными от атак и манипуляций.

Многие из этих проблем могли бы быть решены путём обучения больших языковых моделей по структурированным текстам, но как именно их структурировать и откуда взять структурную разметку в «промышленных» объёмах?

ПРОБЛЕМА НЕЭФФЕКТИВНОСТИ ТЕКСТА

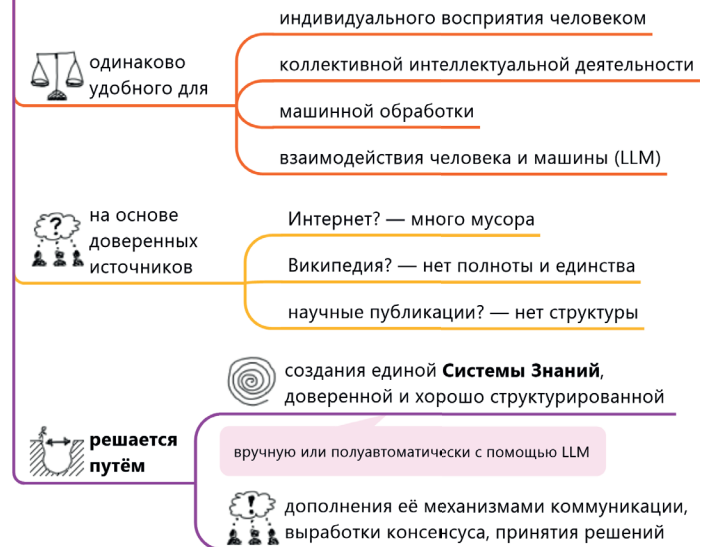


В современных условиях текст утрачивает свою эффективность как средство передачи знаний, превращаясь из драйвера развития цивилизации в его тормоз.

Усваивая знания, мы представляем их образно, нечётко, неточно, частично и каждый по-своему, в зависимости от прагматических целей и предшествующего опыта. Стремясь «упорядочить наконец эту кашу в голове» и «разложить всё по полочкам», мы пишем тексты, используя знаковые системы кодирования, придуманные сотни или тысячи лет назад ради других целей и в иных информационных условиях. Читая наш текст, другой человек занимается декодированием, выделяя из него идеи и структуры, укладывая их в свою картину мира, свои «полочки», ограниченные собственной прагматикой и опытом. Процессы кодирования и декодирования ведут к искажению смыслов, требуют затрат энергии и времени, стремительно растущих в масштабах всего человечества. Мы ещё больше запутали наши тексты, научившись скрывать за словами противоречия и нелогичности, манипуляции и обман, демагогию и графоманию. Теперь всё это свободно доступно в интернете, и по этим данным обучаются большие языковые модели.

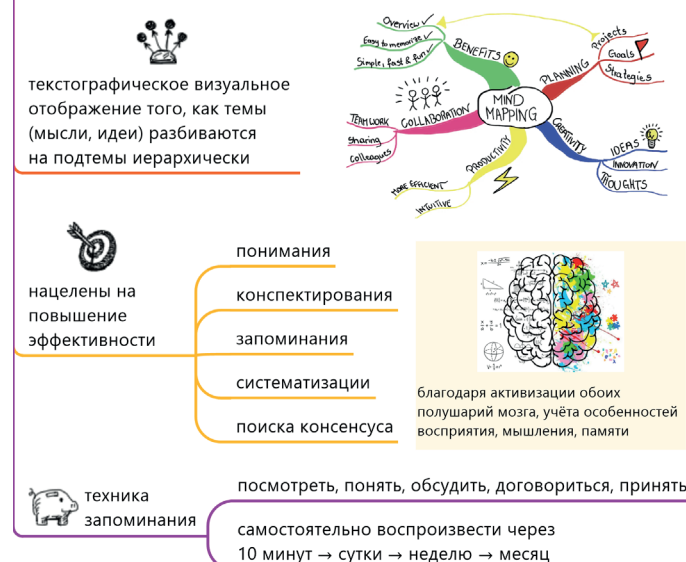
Назрел переход к новым способам передачи знаний, более приспособленным к растущим темпам развития цивилизации, менее ресурсоёмким, более эргономичным, учитывающим особенности восприятия, мотивации и памяти людей. Переход от текста к нелинейному чтению меняет привычные способы потребления информации, способствует массовому развитию навыков выделять во всём главное, достигая взаимопонимания и консенсуса при коммуникации. Это определён путь к коллективному разуму — способности коллектива приходить к единомыслию и находить решения задач более эффективно, чем при индивидуальной интеллектуальной деятельности [8].

ПРОБЛЕМА ВЫБОРА НАИЛУЧШЕГО СПОСОБА ПРЕДСТАВЛЕНИЯ ЗНАНИЙ



Проблема в том, чтобы найти наилучший способ представления знаний, удобный и для индивидуального восприятия человеком, и для коллективной интеллектуальной деятельности, и для машинной обработки, и для обучения LLM, и для коммуникации людей с LLM. На эту роль могли бы претендовать многие известные способы, однако в профессиональных сообществах пока нет согласия относительно того, какой из них лучше. Выбор сужается, если поставить целью создание единой доверенной системы знаний, дополненной механизмами коммуникации, выработки консенсуса, интеграции с процессами коллективного принятия решений.

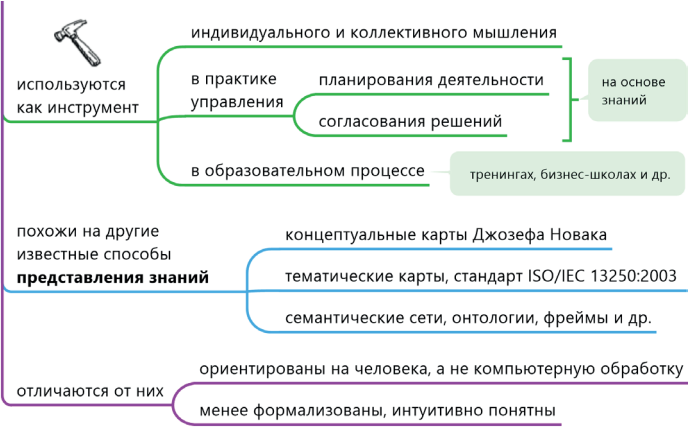
ИНТЕЛЛЕКТ-КАРТЫ



Мы обращаем внимание на интеллект-карты (mind map) — широко известный текстографический способ визуального структурирования идей, мыслей или тем, предложенный британским психологом Тони Бьюзеном

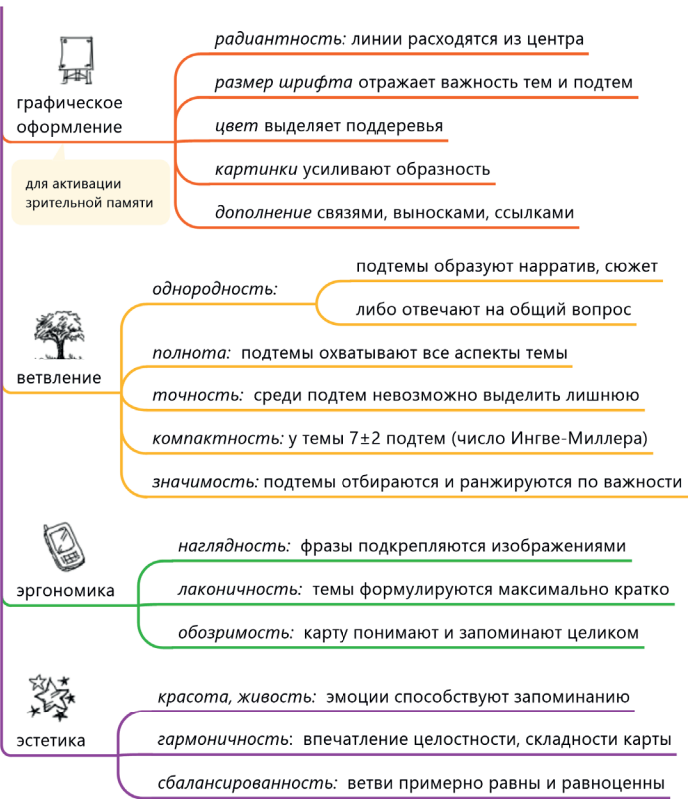
ещё в 70-е годы [5]. Каждая тема разделяется на подтемы, постепенно детализируя центральную тему (понятие) наглядно, в виде иерархии.

ПРАКТИКА ПРИМЕНЕНИЯ ИНТЕЛЛЕКТ-КАРТ



Чтобы сделать интеллект-карты эффективным инструментом индивидуального и коллективного визуального аналитического мышления, их необходимо усиливать дополнительными дисциплинирующими принципами.

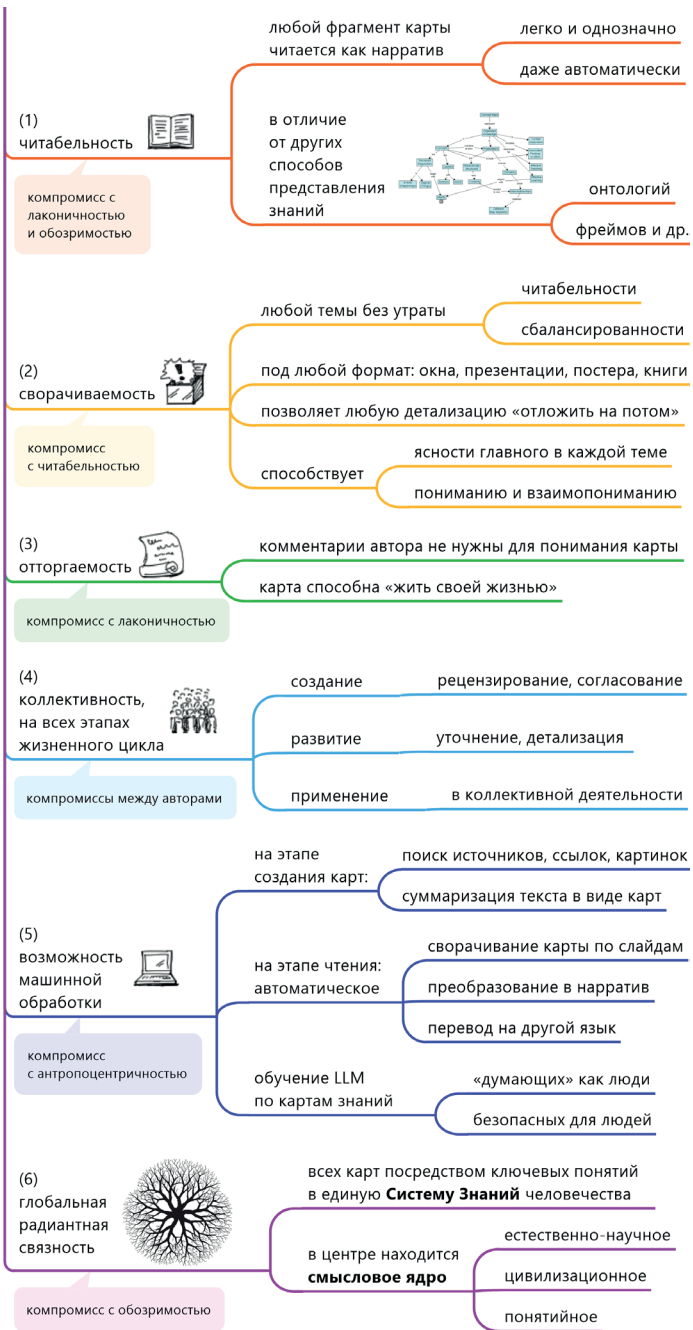
ПРИНЦИПЫ ПОСТРОЕНИЯ ИНТЕЛЛЕКТ-КАРТ



Развивая технику майнд-мэпинга, Тони Бьюзен и его последователи предложили множество принципов, которые иногда противоречили друг другу, но допускали компромиссы в зависимости от практических целей и задач. В частности, Бьюзен настаивал на том, чтобы каждая ветвь карты подписывалась как можно коро-

че, в идеале — одним словом. Однако такую карту будет понимать только её автор; в коллективе ему придётся объяснять, что он имел в виду, формируя фоновые знания, не отражённые на карте. Это противоречит принципам однородности, полноты и отторгаемости карты. Всего мы выделяем 16 известных принципов построения интеллект-карт, разбивая их на четыре группы: графическое оформление, ветвление, эргономика и эстетика.

ПРИНЦИПЫ ПОСТРОЕНИЯ КАРТ ЗНАНИЙ



Регулярное следование принципам радиантного мышления, компактности и значимости вырабатывает у человека навык выделения главного в любой теме, о чём бы ни шла речь. Главных подтем в каждой теме должно быть

немного, и они должны образовывать сюжет. Мы привыкли мыслить историями, визуальными образами, расставлять опорные точки. Нам так удобнее ориентироваться и запоминать. Это механизмы приспособления и выживания в природной среде, обусловленные биологической эволюцией. Теперь речь идёт о том, чтобы переключить их на решение интеллектуальных аналитических задач.

Сама эта статья целиком служит иллюстрацией того, что карты являются полноценным заменителем текста. Они лучше структурированы, более наглядны, легче запоминаются. Сначала мы давали после каждой карты объяснения, полностью дублирующие её, затем всё менее подробные. Как лучше воспринимать информацию — читая текст или карту? Возможно, поясняющий текст и не нужен, если карта читается как обычный иерархический список, оформленный графически.

Усилив интеллект-карты шестью новыми принципами в дополнение к 16 базовым, мы получаем намного более мощный инструмент мышления, предлагая для него новое название — карты знаний.

Первые три принципа (читабельность, сворачиваемость, отторгаемость) обеспечивают полноценность замены текста картой. Сворачиваемость ориентирована на компьютерную реализацию с графическим интерфейсом.

Принцип коллективной разработки и использования (4) делает карту знаний инструментом совместной интеллектуальной деятельности, а в случае программной реализации механизмов согласования и коллективного принятия решений — технологической основой коллективного разума.

Возможности машинной обработки (принцип 5) позиционируют карты знаний как среду взаимодействия не только людей друг с другом, но и людей с машинным интеллектом. Есть очевидные функции, которые уже могут быть реализованы с помощью существующих технологий, такие как использование LLM для автоматизации построения карт и представление текстов, порожаемых LLM, в виде карт. Более заманчивой и нетривиальной представляется перспектива обучения следующего поколения языковых моделей на текстах, структурированных в виде карт знаний. Особенно если обучающие карты выработаны коллективным разумом — прошли все стадии обсуждения и поиска консенсуса, стали результатом сформировавшегося единомыслия.

Принцип глобальной связности (6) является наиболее существенным отличием карт знаний от обычных интеллект-карт. Тони Бьюзен предлагал размещать в центре каждой карты одно ключевое понятие. Связи между понятиями могут образовывать сложную запутанную сеть. Однако радиантное мышление подсказывает, что любая тема или понятие является частью чего-то более важного. Такие связи в сети являются основными, образуют её скелет, радиантную иерархическую структуру.

Благодаря глобальной связности любая новая деятельность, имеющая свои цели и задачи в рамках некоторой предметной области, встраивается в глобальную карту единой Системы Знаний человечества, охватывающую все предметные области.

Цепочки переходов от дочерних подтем вверх к родительским надтемам приводят к глобальному смысловому (семантическому) ядру. Это знания, которые важны всегда и для всех, важнее которых коллективный разум пока ничего не определил. В ядре находятся точки входа в естественно-научную картину мира и цивилизационную систему ценностей [6, 7]. Здесь же находятся понятия, называемые категориями, которые определяются друг через друга и используются для определения всех остальных понятий.

СМЫСЛОВОЕ ЯДРО СИСТЕМЫ ЗНАНИЙ

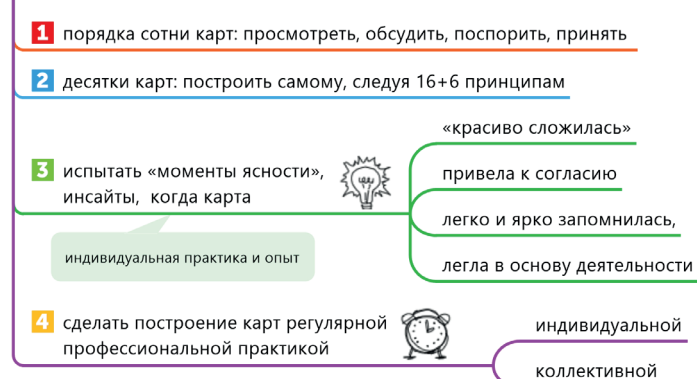


Смысловое ядро должно содержать, в числе прочего, критерии важности для выделения в каждой теме небольшого числа наиболее важных подтем. «Что в теме главное? Для чего и для кого оно главное?» — ответы на эти вопросы достижимы из любой темы на карте за небольшое число шагов вверх до смыслового ядра. Принцип глобальной связности приводит к регламенту сверки целей и задач любой практической деятельности с базовыми принципами существования и развития человеческой цивилизации, записанными в смысловом ядре. Работая с глобально связанной картой, мы не можем обойти этот регламент, он оказывается встроенным в любую практическую деятельность, индивидуальную или коллективную.

Итак, мы начали с проблемы построения доверенных больших языковых моделей, способных общаться с людьми более структурированно и уметь выделять главное не хуже, чем это делают люди. В итоге пришли к осознанию трёх ещё более важных фундаментальных задач. Первая — всеобщий переход от плоского текста к более эргономичному способу представления знаний. Вторая — активация человеческого (не машинного) коллективного разума в процессах практической деятельности. Третья — создание единой Системы Знаний человечества как информационной и коммуникативной среды, в которой функционирует коллективный разум.

Экстенсивное развитие LLM путём обучения по массивам текстов интернета несёт в себе угрозу перехвата управления при решении перечисленных выше фундаментальных задач. Мы ещё не научились договариваться друг с другом по важнейшим вопросам выживания цивилизации, всё ещё не включили коллективный разум, а наши LLM уже обучаются по полному массиву наших противоречивых мнений обо всём и уже начинают давать советы. Ответ на эти вызовы необходимо искать быстро. Один из перспективных путей — создание информационной среды для коллективной интеллектуальной деятельности по методологии карт знаний.

АКТИВАЦИЯ ВИЗУАЛЬНОГО АНАЛИТИЧЕСКОГО МЫШЛЕНИЯ



Интеллект-карты пока не стали популярным рабочим инструментом, хотя и включены в школьную программу по информатике. Грандиозный потенциал визуального аналитического мышления остаётся недооценённым даже в рамках многочисленных школ майнд-мэппинга. Данная проблема имеет простое практическое решение. Естественный человеческий интеллект снабжён высокоэффективным «графическим процессором», который мы используем повседневно для ориентации в пространстве, запоминания образов и событий, однако не привыкли включать его для аналитической деятельности. Выработать и закрепить такие навыки возможно путём практики, в четыре этапа [9].

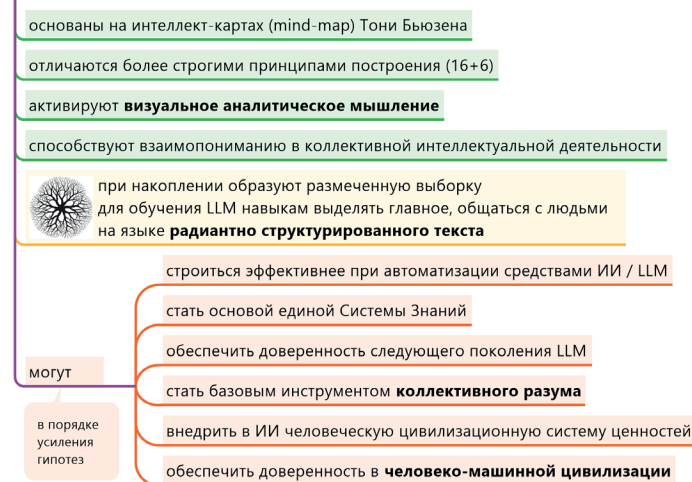
Первый этап ознакомительный, он заключается в просмотре и обсуждении карт смыслового ядра, описывающих верхние уровни естественно-научной картины мира и цивилизационной системы ценностей. Похоже на быстрое повторение школьной программы и систематизацию общеизвестных истин.

Второй этап практический — это самостоятельное построение карт в рамках собственной профессиональной деятельности.

Третий этап самый важный. Отработка технических приёмов визуального представления и методологических приёмов коллективной интеллектуальной деятельности с помощью карт — первый шаг к коллективному разуму. Именно на этом шаге происходит яркое эмоциональное включение визуального мышления — в моменты осознания, что совместное рисование и обсуждение карты привело к согласию на порядок быстрее, чем все ранее испытанные способы. Такого рода инсайты и «моменты ясности» бесполезно описывать словами, через этот опыт необходимо пройти.

Четвёртый этап — внедрение практики карт знаний в профессиональную деятельность, как индивидуальную, так и коллективную.

ВЫВОДЫ: КАРТЫ ЗНАНИЙ



Рискованно создавать сильный искусственный интеллект, не приведя в порядок естественный, не научившись приходить к консенсусу и единомыслию в важнейших вопросах выживания и развития цивилизации, не научившись решать сложнейшие задачи современности с помощью коллективного разума, суперинтеллекта. Активация визуального аналитического мышления с помощью карт знаний — это определённо шаг к приведению в порядок своего человеческого интеллекта, доступный каждому. Способствуя взаимопониманию и доверию между людьми, карты знаний могут сыграть ключевую роль и на пути к доверенному ИИ, через обучение следующего поколения больших языковых моделей по картам знаний.

СПИСОК ЛИТЕРАТУРЫ:

1. Bo Li et al. Trustworthy AI: From Principles to Practices. ACM Comput. Surv. 55, 9, 46 pages. 2023.
2. Bubeck S. et al. Sparks of Artificial General Intelligence: Early experiments with GPT-4 // arXiv preprint 2303.12712. — 2023.
3. Fenogenova A. et al. MERA: A Comprehensive LLM Evaluation in Russian // arXiv preprint 2401.04531. — 2024.
4. Xingjiao Wu et al. A Survey of Human-in-the-loop for Machine Learning // Future Generation Computer Systems, Elsevier BV, Vol. 135, pp. 364–381. — 2022.
5. Бьюзен Т. Интеллект-карты: полное руководство по мощному инструменту мышления. — М.: Манн, Иванов и Фербер, 2019. — 199 с.
6. Воронцов К. В. Цивилизационная идеология // в сб. Сверхновая Россия. Какая ты будешь? Изд-во «Наше завтра», 2022.
7. Воронцов К. В. Цивилизационная идеология: мысли о будущем искусственного интеллекта и о будущем вообще. Лекторий «Кругозор», МГУ, 23 ноября 2023. [Электронный ресурс]. — URL: <https://www.youtube.com/embed/snfORW1hz9I> (дата обращения: 26.02.2024).
8. Кузнецов П. Г. Искусственный интеллект и разум человеческой популяции // в кн. Е. А. Александров. Основы теории эвристических решений. М.: Сов. радио, 1975.
9. Курилов В. А. Теория счастья. Азбука жизни. Красноармейск: Азбука жизни, 2006. 208с. ISBN 5-91270-001-1.

«ПОРА ЗАМЕНИТЬ РАСШИФРОВКУ АББРЕВИАТУРЫ LLM НА LOGIC-LINGUISTIC MODEL»



Галина НОВИКОВА
к. т. н., член Российской ассоциации искусственного интеллекта, эксперт в области систем корпоративного управления

Сегодня даже среди учёных и специалистов в области ИИ нет единого мнения, что такое искусственный интеллект. Определение подменяется различными прилагательными: слабый, узкий, сильный, универсальный, общий, генеративный (GenAI). По моему мнению, необходимо отличать понятие «искусственный интеллект» в смысле имитации поведения и свойств человека, а также компьютерной реализации его когнитивных способностей и возможностей, от понятия «прикладной искусственный интеллект», которое подразумевает прагматические аспекты применения технологий ИИ для решения прикладных задач (рис. 1).

ИИ И БЕЗОПАСНОСТЬ: ОБЪЕКТИВНОСТЬ ИЛИ УПРАВЛЕНИЕ?

Сильный ИИ аккумулирует работы по всем направлениям, относящимся к различным вариантам узкого ИИ, решающего конкретные задачи. Конечно, новый скачок в создании сильного ИИ стал возможен благодаря технологиям, которые позволяют обучать

нейронную сеть на неразмеченном тексте. На первый взгляд, потрясающий прорыв очевиден: аккумулируя всю совокупность знаний интернет-ресурсов, сильный ИИ становится эрудитом, способным правдоподобно ответить на любой вопрос и решить с допустимой степенью точности любую задачу. Но за всеми этими успехами, которые греют душу и будоражат естественный интеллект учёных, скрываются угрозы, которые уже сейчас начинают осознаваться, — посягательство на психологическую, когнитивную, эмоциональную, интеллектуальную, социальную безопасность. Первый закон робототехники должен рассматриваться сегодня шире, чем во времена, когда об этом размышлял Айзек Азимов, и формализация понятия «вред человеку» должна охватывать все аспекты, а не ограничиваться нанесением физического ущерба.

Нейронная сеть может обучаться на заведомо ложной информации, имеющей определённое целевое назначение. И распознать заложенную в ИИ модель мира, поведения, принятия решений гораздо сложнее, чем её создать. Скажем, ресурс «Википедия» на английском, китайском или русском языке так или иначе задаёт свою модель мира, связанную с национальными ценностями, чертами характера, интерпретацией исторических событий, культурой и языковыми особенностями. Вряд ли можно рассчитывать когда-нибудь создать универсальный сильный ИИ: с ним произойдёт то же самое, что и с попыткой создания универсального языка Эсперанто.

Сильный ИИ может быть объектом и инструментом управления. Кроме того, не

Имитационный ИИ	Генеративный ИИ	Когнитивный ИИ	Прикладной ИИ
Имитация сенсорных органов и способностей человека, а также его поведения	Имитация творческих способностей (создание музыкальных, литературных, художественных произведений)	Автоматизация интеллектуальной деятельности (машинный перевод, боты-консультанты, программирование)	Решение прикладных задач в различных предметных областях (распознавание, прогнозирование, диагностика, принятие решений)

Рисунок 1. Разновидности искусственного интеллекта

решены проблемы, связанные с реализацией объяснительной функции LLM. И в этом смысле уже важны не столько миллиарды и триллионы словесных токенов, которыми оперирует LLM, сколько та картина мира, которые они формируют в представлении пользователя большой языковой модели. А значит, на повестке дня два больших вопроса: первый — сосуществование/конкуренция/противостояние на глобальном уровне национальных сильных LLM. Второй — системный подход к созданию модели мира, на базе которой функционируют лидирующие LLM.

МАЛООБРАЗОВАННЫЕ, ХОТЯ И БОЛЬШИЕ LLM

Большая языковая модель не оперирует системой знаний. Она оперирует сверхбольшим объёмом информации, достоверность, актуальность, непротиворечивость которой может быть под вопросом. Можно сказать, что нынешняя LLM — это «невоспитанная» модель с ворохом несистемных знаний обо всём на свете, и ей, действительно, ещё нужно дать «образование» и «воспитание», привести в соответствие накопленную информацию. Более того, отсутствие системного подхода — это как раз та самая проблема, которая не даёт говорить о знаниях больших языковых моделей, потому что знание — системно. Нужна структура знаний, механизмы разделения знаний и данных, инструменты проверки истинности и определения прагматики знаний в рамках заданной модели мира, нужны механизмы интерпретации знаний. В общем, необходим формальный системный подход к представлению знаний.

Надо сказать, что соответствующие исследования на высоком международном уровне шли в нашей стране ещё в 80-х годах прошлого века: разработки логико-лингвистических моделей для создания интеллектуальных систем связаны в первую очередь с именами двух выдающихся советских учёных — д.т. н. Д. А. Поспелова и академика Г. С. Поспелова, которые фактически стояли у истоков ИИ в нашей стране. В силу известных причин работы по формальному подходу к представлению знаний, созданию баз знаний и онтологий затормозились, и к началу XXI века системы, в которых был реализован логико-лингвистический подход, стали классифицироваться как символьные системы, основанные на правилах, причём для представления знаний в основном использовались продукционные модели. Надо с сожалением отметить, что такая узкая интерпретация логи-

ко-лингвистических моделей нанесла вред всему направлению отечественного ИИ, поскольку фактически прекратилось развитие формальной логико-лингвистической модели — семиотической системы, предложенной Д. А. Поспеловым, которая интегрирует формальную систему и продукционные модели для представления знаний о предметной области.

Но эволюция ИИ сделала очередной виток, и сегодня индустрия искусственного интеллекта вновь оказалась перед тем же самым вызовом. Думаю, пора заменить расшифровку аббревиатуры LLM на Logic-Linguistic Model и акцентировать внимание на работах в этом направлении.

Специалисты по созданию онтологий хорошо знакомы с треугольником Фреге, где понятию сопоставляется знак, концепт (смысл понятия) и денотат (множество примеров, значений понятия), к которым Д. А. Поспелов добавил четвёртый аспект — прагматику понятия. Именно такое многоаспектное описание понятий должно храниться на онтологическом уровне базы знаний. А сегодня нейронные сети, пусть даже очень большие и с глубоким обучением, оперируют только синтаксическими правилами и вряд ли сортируют информацию, заложенную в неё, на знания и данные, на экстенционалы и интенционалы понятий, выделяя прагматические свойства понятий.

Что касается онтологий, то соответствующие технологии к настоящему моменту только-только достигли такого уровня, что можно замахнуться на создание универсальной онтологии или, по крайней мере, способной охватить широкий спектр предметных областей и решаемых задач. Но главное, что становится возможным сейчас и не было возможным в 80-х, — это работа с многомодальностью. Звуковой, визуальный, динамический образ понятия не только обогатит его символьное (знаковое) представление, но и позволит делать всё более точной семантику и прагматику понятия.

ИИ И АВТОМАТИЗАЦИЯ УПРАВЛЕНИЯ ПРЕДПРИЯТИЕМ

Если перейти на уровень автоматизации управления в корпоративных информационных системах, а это именно то, что интересует бизнес в первую очередь, мы опять сталкиваемся с компонентами, в основе которых лежит нейросетевой интеллект, но отсутствует системный подход к пониманию самого процесса управления. Под управлением при-

вычно понимаются отдельные задачи, например контроль состояния технических систем, прогнозирование потребительского спроса, анализ эффективности бизнес-процессов. Системный подход отсутствует не только в представлении знаний интеллектуальной системы, но и в понимании управления как целостного процесса...

А проблемные точки нынешнего нейросетевого интеллекта — создание контролируемого, объяснимого, понятного ИИ — становятся ещё более критичными в области корпоративного управления, поскольку при его использовании для автоматизации управления предприятием возникает не только право на принятие решений, но и ответственность за принятое решение, причём на каждом шаге цикла управления. И здесь нейросетевой интеллект — увы! — существенно уступает тем самым символьным системам, основанным на правилах. Но не в том узком понимании символьных систем, который сложился в «нулевые» и фактически позволил им уступить место нейросетевому интеллекту. Необходимо развитие формального аппарата, создание инструментальных средств работы с формальным аппаратом, а главное — создание адаптивных онтологий, связанных с включением в них и формализацией не только статических, но динамических знаний.

По большому счёту, мир ещё только подходит к созданию по-настоящему интеллектуальных систем, способных стать помощниками человека. Задача этого этапа формулируется достаточно просто: создание искусственного интеллекта, который реализует **формализованную деятельность человека**.

Действительно, вначале надо чётко понять, что, как и для чего делается (формализовать деятельность), а уже потом создавать инструменты, автоматизирующие или имитирующие её. Очевидно, что осознанная передача ИИ функций (прежде всего производственных), требующих выполнения либо рутинных операций, либо операций, связанных с уровнем повышенной опасности или требующих моментального принятия решений и обработки больших массивов информации, выходящих за рамки возможностей человека, служит на благо человека. При этом ИИ может имитировать и сенсорные способности человека, и его каналы получения и передачи информации (голос, зрение, тактильные органы). Но мозг (если так можно выразиться) ИИ должен быть под контролем челове-

ка! Человек должен понимать, какой системой знаний обладает ИИ, какие задачи он перед собой ставит, как их решает и какие цели при этом преследует и достигает.

ВРЕМЯ «ИНТЕЛЛЕКТУАЛЬНОГО ПЕРЕКОСА»

Да, именно так: сегодня мы переживаем период «интеллектуального перекоса»: акцент сделан скорее на имитацию деятельности человека, его творческих и умственных способностей, чем на осознанное наполнение «мозга» ИИ знаниями. Конечно, нейросетевой интеллект уже хорошо работает в частных задачах прогнозирования, распознавания, диагностики, контроля. Однако стержнем интеллектуальной системы, её базы знаний должна быть формализованная, логико-лингвистическая модель предметной области, в который существуют интерфейсы между символьной информацией и информацией, представленной другими модальностями, а также реализованы механизмы поддержки сложноструктурированных информационных сущностей, получения новых знаний и проверки их на достоверность, полноту и непротиворечивость.

Думаю, что нейросетевой интеллект может помочь извлекать из текстовых данных новые смысловые модели, открывать новые законы и закономерности в различных предметных областях. И этот «метод познания мира» вполне можно назвать индуктивным выводом, когда из множества фактов формируется новое знание. Однако не нужно генерализовать этот частный случай формирования новых знаний. Иначе мы получим дальнейшее углубление сегодняшнего противоречия: не ИИ — для человека, а человек становится помощником, учителем, обслуживающим персоналом и ресурсом для ИИ. И тогда активизация работ в области создания многоагентных систем, где интеллектуальные агенты заменяют человека в различных сферах экономики, может привести к тому, что интеллектуальные агенты начнут создавать профсоюзы и бороться за свои права, например, для получения большего ресурса с целью восстановления после рабочего дня?! Конечно, это шутка. Но всё же пора привести в соответствие наши взаимоотношения с создаваемой искусственной сущностью, чтобы направить интеллектуальные технологии на нужное дело — создание **прикладных интеллектуальных систем**, полезных человеку, предприятию, государству вместо усилий по имитации человека и его жизнедеятельности.

«В ВОПРОСЕ АВТОМАТИЗАЦИИ СЛЕДУЕТ СКОНЦЕНТРИРОВАТЬСЯ НА ПРИМЕНЕНИИ LLM ДЛЯ РЕШЕНИЯ РУТИННЫХ ЗАДАЧ»



Григорий ШЕРШУКОВ
директор по продуктам
компании
«Наносемантика»

Само название «Большая языковая модель» указывает на то, что речь идёт о больших объёмах данных, используемых для обучения соответствующих нейросетевых моделей.

РАЗМЕР ИМЕЕТ ЗНАЧЕНИЕ

Из того, что мы знаем: для обучения версии GPT-3 использовались данные объёмом в сотни миллиардов слов и порядка 175 миллиардов параметров. Параметры — это коэффициенты в нейросети, которые модель настраивает в процессе обучения. Большое количество параметров позволяет модели более точно воспроизводить и генерировать текст, анализировать контекст и обеспечивать более сложные ответы.

На сегодняшний день лидером в тестированиях открытой организации студентов Беркли является GPT-4, точнее её свежая версия GPT-4o. Точное количество параметров четвёртой модели неизвестно: разработчик не раскрывает точную цифру, но оценочно оно составляет более триллиона (ориентировочно 1,76 трлн параметров). Отечественные модели насчитывают 29 млрд параметров у GigaChat и, по последним оценкам, 100 млрд у YandexGPT.

Однако, если проводить прямое сравнение, беря за основу только количество заявленных параметров нейросети, это не даст полного понимания качества или возможности модели. Существует набор тестов, позволяющих оценить, насколько модель хороша в решении задач по следующим направлениям: прохождение тестов, понима-

ние естественного языка, аргументация рассуждений, кодирование и проч. Что-то типа ЕГЭ или ОГЭ.

Кроме того, необходимо понимать, что LLM и данные для её обучения представляют собой технологический продукт, над которым итерационно работают специалисты разных профилей. Представление о том, что поместить большие данные в генеративную языковую модель — это достаточно просто, является большим упрощением. На практике добытые из интернета данные последовательно подвергаются фильтрации, направленной на получения среза данных приемлемого качества, другими (обычно маленькими языковыми) моделями. Проведение такой работы требует привлечения команды аналитиков и специалистов по машинному обучению, а также разметчиков данных.

Стандартный набор источников для формирования обучающего датасета для LLM включает обсуждения на профессиональных форумах, посвящённых узконаправленным тематикам, корпуса текстов из книг, академических статей, научных исследований и программных кодов. Кроме того, существуют базовые корпуса текстов, созданные для решения конкретных задач, таких как машинный перевод, анализ тональности, нахождение ответа на вопрос в заданном контексте. Также имеются и узкоспециализированные корпуса текстов и бенчмарки, отвечающие различным сферам человеческой деятельности. Решение о включении последних в итоговый датасет для обучения LLM также является нетривиальной задачей. Оно опирается на предшествующие исследования (то есть опубликованные ранее научные статьи), в которых анализировался этот набор данных и изучались возможности решения ряда других задач с помощью данных этого корпуса, а также на опыт работы команды с таким датасетом. Например, включение некоторого условного корпуса медицинских текстов может оказаться не лучшей идеей, если выяснится, что диагнозы пациентам в ряде случаев были проставлены в свя-

зи со стоимостью оказываемых в дальнейшем медицинских услуг.

Наконец, у компаний-разработчиков LLM существуют так называемые in-house-датасеты, то есть внутренние датасеты, подготовленные экспертами для придания LLM уникальных особенностей. И это ещё не всё! После обучения и оценки качества LLM проходит процедуру «выравнивания», направленную на настройку модели с учётом предпочтений и ожиданий потенциальных пользователей. Выравнивание большой языковой модели можно сравнить с огранкой алмаза: в этот процесс вовлекаются лингвисты и эксперты ряда предметных областей, которые «отшлифовывают» генеративные навыки LLM, делая её ответы полезными, безопасными и естественными.

По большому счёту, в развитии генеративных AI-систем важны все аспекты использования технологий. Критически важное значение имеют объём и качество данных, которые лежат в основе обучения точных и надёжных моделей. А адаптация под различные сферы применения и предметные области позволяет решать уникальные задачи, требующие глубоких знаний, что делает модели более полезными в различных отраслях. Технические и архитектурные инновации, такие как модели MoE (Mixture of Experts), играют ключевую роль в улучшении производительности и эффективности AI-систем.

ЧЕГО ХОЧЕТ БИЗНЕС ОТ LLM

Безусловно, большие языковые модели привлекают внимание бизнеса, поскольку уже сейчас позволяют автоматизировать элементы широкого спектра бизнес-процессов. Здесь стоит отметить важную роль LLM-моделей с открытым исходным кодом в развитии технологий ИИ. Они позволяют исследователям и компаниям экспериментировать с новыми подходами, улучшать существующие модели и создавать инновационные решения. Открытые модели способствуют демократизации технологий ИИ, снижая барьеры для входа и стимулируя сотрудничество по развитию и применению таких моделей.

Вместе с тем необходимо оценивать риски потенциальных угроз, задумываясь об автоматизации и использовании продуктов на основе LLM. Большая языковая модель, по сути, является «чёрным ящиком»: у пользователей таких LLM нет возможности проанализировать реальные корпуса данных, исполь-

зованные при обучении этих LLM, а также непосредственно ознакомиться с их «протоколами» обучения. Таким образом, возникает, например, гипотетическая угроза преднамеренного внесения в обучающий датасет данных, которые при наступлении конкретных условий могут повлечь, например, нарушение целостности обрабатываемых впоследствии данных в значимых бизнес-процессах. Это может быть списано на «галлюцинацию» (генерация некорректной или вымышленной информации) языковой модели, которая сама по себе может представлять собой угрозу в критически важных сценариях. На мой взгляд, внедрение больших языковых моделей в критически значимые бизнес-процессы, безусловно, создаёт новые вызовы в индустрии ИБ, которые ещё предстоит преодолеть, а в вопросе автоматизации следует сконцентрироваться на применении LLM для решения рутинных задач.

Например, «Наносемантика» разрабатывает и продаёт платформу для разработки ботов, и к нам приходит бизнес, который заинтересован в автоматизации коммуникации либо внешних клиентов с компанией, либо внутренних сотрудников между собой. Сегодня практически каждый заказчик интересуется, можем ли мы «подключить бота к ChatGPT». На самом деле не обязательно этой моделью должен быть именно продукт ChatGPT, который, строго говоря, сам уже является чат-ботом, использующим генеративные нейросетевые модели. Заказчики готовы обсуждать обращение к LLM других вендоров, в том числе «Яндекса» и Сбера, а также интересуются возможностью разработки LLM под ключ по их данным, но «чтобы модель отвечала как ChatGPT».

Таким образом, с одной стороны, крупный и средний бизнес хочет использовать преимущества новейших технологий. С другой стороны, он пока не до конца определился, каким образом большие языковые модели должны встраиваться в основные коммуникационные процессы бизнеса. Действительно, тот же ChatGPT знает много фактической информации, может рассказать анекдот, сгенерировать поздравление или научную работу. Это прекрасно, но что это даёт для автоматизации процессов коммуникации в бизнесе? «Живой» оператор в банке или телекоме не будет болтать с клиентом обо всём на свете, искать ответ на вопрос «как решить уравнение» или «как лучше написать инструкцию к телевизору». А если с чат-ботом болтают на отвлечённые темы, то это реально повы-

шает лояльность старых клиентов и привлекает новых? Не создадим ли мы просто собеседника для тех, кому не спится, без выгоды для бизнеса?

По нашему опыту, запрос «говорить как ChatGPT» во многих случаях означает настройку его базы знаний на определённую деловую сферу: общение по продуктам и услугам конкретной компании, формирование и запись обращений потребителей, работа с документами, как этого требует регламент компании, выполнение доступных действий (заявки или подключение новых услуг, транзакции...). Но при этом ожидается, что с точки зрения языка качество коммуникации будет очень высоким, сопоставимым с общением на естественном языке с человеком. Иными словами, бот понимает разные варианты выражения одной и той же мысли, помнит историю диалога, работает с эмоциональным тоном, узнаёт собеседника и помнит его предпочтения в диалоге, в речи различает интонации, голоса взрослых и детей и учитывает эту информацию при стилизации своих ответов. Бот также обладает «коммуникационной личностью», то есть обучается генерировать ответы в заданной стилистике, соответствующей его визуальному образу и «истории».

КАКУЮ LLM ВЫБРАТЬ?

Можно в самом деле обращаться к ChatGPT. Но любой вендор или разработчик, который пробовал это делать, сразу перечислит целый список сопутствующих сложностей: от невозможности доступа из России до проблем в работе с длинными промптами, долгим ожиданием ответа, сложностями в генерации эмоционально окрашенных ответов и др. Русскоязычные пользователи особенно жалуются на определённую «корявость» ответов на русском языке, плохую их стилизацию, а точнее практически отсутствие стилизации, даже если она запрошена в промпте. А ещё на периодическое «галлюцинирование» модели, что выражается, например, в возможной генерации несуществующих слов русского языка.

В этом плане перспективным выглядит обращение к LLM российских вендоров, в первую очередь крупнейших разработчиков: «Яндекс» и Сбер. И здесь тоже не обходится без подводных камней. Например, Алиса «Яндекса» часто отвечает слишком коротко в ситуациях, когда пользователь ожидает более длинного ответа, и в целом хуже держит тему и фокус диалога, чем ChatGPT.

Причём все упомянутые решения — облачные, а это означает проблемы с обработкой персональных данных: нет гарантий их конфиденциальности. То же относится к любой важной для бизнеса информации. Поэтому хорошим вариантом выглядит обучение для заказчика индивидуальной LLM. Такая модель может ставиться в контур предприятия, что снимает вопросы по работе с конфиденциальной информацией. Она будет владеть всеми предоставленными данными, но, конечно, научную работу с обзором мировых трендов по выбранной теме не напишет. Если только примеры таких работ не будут изначально заложены в данных для обучения.

КОНКУРЕНЦИЯ БОТОВ

Существующие разговорные боты неплохо справляются с основными задачами бизнеса для компаний, а именно автоматизацией и стандартизацией решения запросов клиентов, проведения анализа, осуществлением продаж и прочих взаимодействий. Эти боты обеспечивают стандартизированные ответы, автоматизируют рутинные задачи и помогают повышать общую эффективность процессов.

Боты нового типа, созданные на базе технологии генеративных сетей, предоставляют больше возможностей для адаптивности и интерактивности. Они могут генерировать более естественные и контекстно-релевантные ответы, что делает взаимодействие с виртуальными ассистентами более продуктивным и комфортным для пользователей. Однако опора на генеративные модели для решения проблем с нуля в расчёте на их огромные базы знаний открывает потенциальные проблемы, такие как галлюцинации и сложности в управляемости ответов. Эти аспекты особенно важны в корпоративной среде, где точность и надёжность ответов имеют критическое значение. Поэтому, хотя LLM-модели имеют огромный потенциал для улучшения взаимодействия с клиентами, они должны использоваться с осторожностью и под контролем.

В перспективе преодолеть эти вызовы может помочь интеграция GenAI с существующими системами и базами данных компаний, а также разработка гибридных подходов, которые сочетают сценарные и генеративные методы. Это позволит использовать преимущества LLM для создания более интеллектуальных и адаптивных систем, одновременно обеспечивая контроль и надёжность, необходимые для бизнес-приложений.

«УЖЕ СЕГОДНЯ МУЛЬТИМОДАЛЬНЫЕ ГЕНЕРАТИВНЫЕ МОДЕЛИ — ЭТО УНИВЕРСАЛЬНЫЙ ИНСТРУМЕНТ, ПОМОГАЮЩИЙ АВТОМАТИЗИРОВАТЬ БОЛЬШОЕ КОЛИЧЕСТВО РУТИННЫХ ЗАДАЧ»



Денис ДИМИТРОВ
управляющий директор
по исследованию
данных Сбербанка,
научный и технический
руководитель
разработки моделей
Kandinsky

В настоящее время на рынке можно выделить два типа решений GenAI (генеративного искусственного интеллекта), которые разрабатываются крупными компаниями или стартапами: проприетарные модели, которые бигтехи или стартапы

держат в закрытом виде, то есть не выкладывают код и веса этих моделей в открытый доступ и модели open source, которыми компании делятся с сообществом. Бигтехи уровня Microsoft (вместе с OpenAI), Google, Nvidia, Apple активно участвуют в разработке генеративных моделей. Стартапы (такие как, например, Stability AI) и другие разработчики также вносят очень большой вклад в развитие этих моделей, который сопоставим, а иногда и превосходит вклад бигтехов. Актуальная тенденция этого сегмента деятельности — развитие мультимодальных моделей, которые способны работать с разными типами данных и решать различные задачи: от генерации текста до ответов на во-

«LLM МОГУТ ЗАНЯТЬ МЕСТО В РЯДУ ДРУГИХ ИНСТРУМЕНТОВ АВТОМАТИЗАЦИИ ИБ»



Владислав ТУШКАНОВ
руководитель группы
исследований
и разработки технологий
машинного обучения
в «Лаборатории
Касперского»

Благодаря тем возможностям, которые открывают большие языковые модели (LLM) с точки зрения обработки и генерации текста, появляются перспективы для их интеграции

в ИБ-решения. Здесь очень важно соблюсти баланс между ожиданиями, которые рождает волна AI-хайпа, реальными возможностями этих систем и требованиями к их надёжности.

Напомним, что в первую очередь большие языковые модели обучаются задаче генерации текста (в случае мультимодальных моделей они могут также обрабатывать визуальный и аудиоконтент), причём в качестве исходных данных выступают, как правило, тексты общего характера из книг и особенно интернета, где полезные статьи соседствуют с устаревшими, шуточными и просто неверными данными. При этом неважно, используются ли эти данные при

просы по видео и даже синтеза мультимедийных данных (изображений, видео, 3D) и т.д.

Что касается положения России на глобальной «карте» GenAI, то на отечественном рынке представлен широкий спектр конкурентоспособных моделей, которые на определенных бенчмарках превосходят по качеству западные аналоги. Например, это модели Kandinsky или GigaChat, созданные Сбером, а также другие модели российских компаний. Важным ресурсом для развития больших генеративных моделей являются вычислительные мощности, производство которых в России хоть и развивается, но не так быстро, — это безусловно наша точка роста. Нужны большие вычислительные ресурсы, чтобы обучать большие модели.

Таким образом, ключевыми аспектами эволюции моделей являются: развитие вычислительных мощностей, увеличение объема и качества данных, оптимизация существующих архитектур и создание новых. И мы видим отличные результаты в области развития AI в целом — с помощью моделей искусственного интеллекта сегодня автоматизируется и оптимизируется множество процессов: от выдачи банковских кредитов до создания полноценных изображений по их текстовому описанию. Причем, применение больших моделей становится воз-

можным не только на больших вычислительных кластерах (по факту суперкомпьютерах), но и непосредственно на смартфонах. Также наблюдается тенденция к воплощению искусственного интеллекта в физических системах, например, роботах, где уже сейчас иногда используются мультимодальные модели. Поэтому с уверенностью можно сказать, что искусственный интеллект (и, в частности, генеративный) уже сейчас применяется практически во всех сферах жизнедеятельности человека, облегчает быт и работу. И количество применений в будущем будет только расти. Именно поэтому разработки в этой области очень важны — ведь одна генеративная языковая или даже мультимодальная модель может стать универсальным помощником человека.

СОВРЕМЕННЫЕ ПРАКТИКИ СОЗДАНИЯ GENAI-МОДЕЛЕЙ

Очень важная часть при подготовке к обучению моделей — сбор и предобработка данных. Для обучения, например, языковых моделей необходимо огромное количество данных, состоящих из сотен миллиардов или даже триллионов текстовых фрагментов. Источниками для них может служить Интернет, общедоступные датасеты и специальные базы знаний. Данные для обучения должны быть качественными, то есть

обучении или в рамках популярных подходов по интеграции внешних данных в модель (таких как retrieval-augmented generation¹): ненадежность этих данных может приводить к неожиданным результатам работы систем. Это явно видно в ситуациях, когда системы на основе LLM дают серьёзные советы на основе ироничных текстов из интернета, например, предлагая добавлять в пиццу клей, чтобы соус был более вязким².

В результате обучения на выходе мы имеем модель — нечёткий образ, своего рода воспоминание об обучающем корпусе данных³.

Генерация текста на основе таких моделей происходит с помощью сэмплирования из вероятностного распределения. Сопоставив эти факты, можно сделать вывод, что основными проблемами всех решений на основе LLM будет являться надёжность и контролируемость того вывода (инференса), который модель предоставляет. Действительно, феномен «галлюцинирования» (или генерации и оперирования ложными фактами) со стороны языковых моделей является настолько серьёзной проблемой, что она прочно вошла в общественное сознание, а само слово hallucinate было названо Кембриджским словарём словом года.

Кроме того, часто приводится тезис, что системы на основе LLM смогут обрабатывать огромные объёмы данных, заменяя традиционные системы. Как правило, это делается потому, что те более простые алгоритмы

¹ Retrieval Augmented Generation (RAG) | Prompt Engineering Guide (promptingguide.ai).

² ИИ Google предложил добавлять в пиццу клей, чтобы сыр не соскальзывал (3dnews.ru).

³ State of GPT | BRK216HFS (youtube.com).

должны пройти предварительную фильтрацию (прежде чем быть загруженными в модель), а также должны быть разнообразными, то есть должны содержать информацию по огромному числу доменов.

Далее на этих данных модель (обычно трансформерная) обучается решать некоторую задачу. Если мы говорим о задачах языкового моделирования, то одной из самых популярных задач является предсказание части следующего слова (токена) на основе предыдущих слов. Таким образом, через модель пропускаются сотни миллиарды текстовых фрагментов и на каждом фрагменте модель учится предсказывать следующий токен (каждый раз подправляя свои веса, чтобы в следующий проход делать это предсказание ещё более точным). Это самый вычислительно ресурсоёмкий и дорогостоящий этап (для него нужен суперкомпьютер, выделенный на несколько недель или даже месяцев).

После такого обучения языковая модель проходит дополнительную стадию настройки (файнтюнинга) на специфический формат общения с пользователем (в формате «вопрос-ответ», то есть в формате диалога). Без этого этапа модель не сможет нормально отвечать на вопросы (ведь она до этого была обучена только генерировать или продолжать, сочинять текст). Модель доучивается на качественных разнообразных диалоговых данных из различных доменов и областей. Это помогает улучшить качество ответов модели на основе уже имеющихся у неё знаний и выучить её вести диалог с пользователем. Наполнение датасета

по узким доменам может происходить путём парсинга данных из Интернета или сбора специализированных знаний, например, в университетах.

Обучение моделей генерации изображений происходит с использованием огромных наборов пар «изображение — текст». Аналогично, для обучения моделей генерации видео используются наборы пар «видео — текст». При этом мультимодальные модели, способные генерировать изображения, звук и видео, например, DALL-E, Midjourney, Suno, Sora, развиваются параллельно с мультимодальными языковыми моделями, такими, как GPT-4 и Gemini, и в точно таком же быстром темпе.

Очень важная и актуальная задача для генеративных моделей — увеличение длины контекста, то есть количества токенов, на которые модель может посмотреть одновременно при генерации следующего токена. Это очень важный аспект, поскольку он позволяет модели обрабатывать за раз больше данных и делать выводы на их основе, как видно, скажем, на примере эволюции ChatGPT (что обеспечивает всё большую и большую применимость таких моделей).

В разговоре про GenAI нельзя не упомянуть свойство моделей «галлюцинировать». Оно проявляется, например, когда модели сталкиваются с запросами, ответов на которые они ранее не видели в датасете или по которым не было достаточно данных. Но поскольку модели обучены давать ответы на любые вопросы, они начинают «изобретать» ответы, что может привести к ошибкам. Методы борьбы с такими «галлюцинациями» конечно же су-

машинного обучения, которые применялись, применяются и будут применяться для обработки данных, в том числе в кибербезопасности, действительно могут иметь очень большую производительность, и их, как и решения на основе LLM, в рамках маркетинговых материалов тоже часто называли AI. К сожалению, LLM требуют больших вычислительных мощностей, а время обработки больших запросов измеряется десятками секунд и даже минутами⁴.

С другой стороны, те задачи, с которыми работают ИБ-специалисты, часто можно не-

плохо формализовать именно как задачи обработки текста. Те же логи или, например, команды, запускаемые через командную строку, представляют собой полуструктурированные данные, для обработки которых вполне можно использовать LLM.

Таким образом, самая важная задача, которая стоит перед специалистами по машинному обучению и кибербезопасности с точки зрения внедрения LLM в продукты и процессы, — это найти те сценарии, в которых большие языковые модели смогут максимально упростить жизнь аналитика ИБ и разгрузить его от рутины, при этом оставаясь максимально подконтроль-

⁴ Testing Anthropic Claude's 100k-token window on SEC10-K Filings | by Jerry Liu | LlamaIndex Blog | Medium.

ществуют — и развиваются в рамках отдельного научного направления. Одним из таких методов можно считать retrieval — механику, при которой модель формирует ответ, опираясь в том числе на наиболее релевантные входящему запросу фрагменты, содержащиеся во внешних базах знаний (таких как, например, Интернет или специализированная база знаний по процессам компании).

Важно помнить, что любые модели искусственного интеллекта — это всего лишь инструменты, который помогают автоматизировать некоторые рутинные задачи. При этом важно всегда относиться к ответам модели критически и проверять их, особенно если это касается важных профессиональных областей, таких как медицина.

О БУДУЩЕМ РАЗВИТИИ GENAI В КОРПОРАТИВНОМ СЕКТОРЕ

В современном мире, где технологии развиваются с невероятной скоростью, всё более актуальной становится идея AI-агентов, обладающих определенным уровнем профессиональной компетентности, и мультиагентных систем, состоящих из взаимодействующих между собой AI-агентов. Почему это направление выглядит перспективным?

Когда модели присваивается определённая роль, будь то математик или философ, она начинает демонстрировать лучшие результаты в соответствующих областях знаний. Но что происходит, когда мы объединяем несколько таких моделей в одну мультиагентную систему? Исследования показывают, что такая система способна справляться с задачами гораздо эффективнее, чем от-

Всё более актуальной становится идея AI-агентов, обладающих определенным уровнем профессиональной компетентности, и мультиагентных систем, состоящих из взаимодействующих между собой AI-агентов

дельные модели. Мультиагентная система может включать множество различных моделей, каждая из которых выполняет свою уникальную функцию. Это открывает новые возможности для исследований и применений в различных сферах деятельности. Например, в бизнесе создание индивидуального агента для каждого клиента позволяет обеспечить персонализированный подход и улучшить качество обслуживания.

Здесь важно отметить, что количество параметров в моделях, реализующих тех или иных агентов, не является ключевым фактором. Они могут быть как малыми, так и большими, специализироваться на решении задач разного уровня и масштаба. Современная наука активно развивается в различных направлениях, создавая как специализированные, так и фундаментальные GenAI-модели. Последние становятся всё более популярными, поскольку их можно многократно использовать и применять для решения задач в различных доменах.

ными человеку. Это могут быть различные контекстуальные подсказки, особенно в рамках сложных продуктов, таких как SIEM, которые позволяют более просто находить нужную информацию и предоставлять контекст, необходимый для принятия решений. С другой стороны, это могут быть механизмы оценки действия аналитика, которые могут подсказывать специалисту альтернативные шаги расследования или реагирования на инцидент. Наконец, будучи специально заточенными на работу с текстами, LLM могут предоставлять краткие рефераты материалов, связанных с threat intelligence, или помогать в написании от-

чётов — при условии их дальнейшей проверки человеком.

LLM не решение всех проблем ИБ и не замена экспертов в области кибербезопасности. Однако это мощный инструмент, который при правильном применении может помочь в автоматизации части задач, которые до этого момента очень плохо поддавались автоматизации. И это позволит разгрузить экспертов от рутинных и дать им время на исследование более интересных событий или объектов. В этой роли LLM могут занять место в ряду других инструментов автоматизации, в том числе основанных на машинном обучении и активно применяющихся в сфере кибербезопасности.

«ОБ ИНТЕЛЛЕКТЕ LLM МОЖНО СПОРИТЬ, НО МИР РАЗРАБОТКИ ОНИ СПОСОБНЫ ПЕРЕВЕРНУТЬ»



Илья ТИХОНОВ
владелец продукта
Polymatica ML,
компания SL Soft

Классическое машинное обучение работает на базе примеров, но требует предварительной разметки. Собственно, генеративный ИИ GenAI ничем от этого не отличается, по уровню интеллекта он такой же, как и условная логистическая регрессия. Она является популярным алгоритмом построения моделей бинарной классификации и предсказания вероятности возникновения интересующего события с помощью логистической функции. Просто если логистическая регрессия — это набор входных данных и весов, которые друг на друга попарно перемножили, и сложили результат, а потом еще провели через одну функцию активации, то в GenAI таких преобразований будет намного больше. Соответственно, это все еще обучение с учителем, просто «учителем» здесь выступают сами тексты. То есть мы видим простой механизм действий модели: она предсказывает следующий токен, который может быть словом или частью слова, или даже буквой, исходя из контекста. Поэтому логичен вопрос: являются ли эти модели, которые просто предсказывают условную вероятность появления следующего слова, интеллектом? Или этот механизм всего-навсего «статистический попугай»? Утвердительно ответим на второй вопрос: да, мы же действительно так его и тренируем.

О БОЛЕЗНЯХ «СТАТИСТИЧЕСКИХ ПОПУГАЕВ»

По сути, все LLM — это просто авторегрессионные генераторы текста, которые пред-

сказывают вероятность следующего токена. Предсказательный характер результата обуславливает «патологию» — галлюцинации модели. Это горячая тема сегодняшних научных исследований — как и почему модели галлюцинируют. Уже понятно, что недостаточно просто увеличить объем выборки: это помогает, но не решает проблему полностью.

Можно, например, построить метамодель, которая обращает внимание на промежуточные представления, и сделать классификатор, отвечающий на вопрос, будет ли модель галлюцинировать или нет. Но оказывается, что, если модель не знает ответа, она попытается «соврать», то есть вместо того, чтобы честно сказать «не знаю», постарается сгенерировать связанный текст. Есть отдельные исследования, которые изучают, насколько сильно модель может галлюцинировать — так сказать, валидация на саму этичность: будет ли модель специально врать, чтобы достичь цели человека.

Наряду с построением метамодели текущая практика предполагает наращивание датасетов «вопрос-ответ» от экспертов в разных направлениях, а не типичных разметчиков, что позитивно скажется на решении проблемы галлюцинаций. Сегодня, когда LLM функционирует как авторегрессионный генератор, обучение происходит следующим образом. Сначала модель учится предсказывать следующее слово на множестве текстов со всего интернета — она начинает понимать структуру текста и контекст использования слов, затем идет дообучение (supervised fine-tune) на датасетах формата «вопрос-ответ». И уже на этом этапе мы видим, что модели получают способности к ведению диалога. Если раньше они работали как «умный Т9», то теперь уже могут вести осмысленный с точки зрения человека диалог. Затем идет этап RLHF (Reinforcement Learning from Human Feedback) — дообучение с подкреплением на основе отзывов людей. Представьте модели «писателя» и «критика». «Писатель» может сгенерировать множество ответов по разным вопросам, люди-разметчики выберут из них лучшие. «Критик» же обу-

чится на этих отзывах и начнет понимать, где условно «хороший» ответ, а где «плохой». Теперь мы дообучим модель «писателя» так, чтобы модель «критик» была удовлетворена как можно чаще.

А ЕСТЬ ЛИ У МОДЕЛИ «ИНТЕЛЛЕКТ»?

На самом деле, все зависит от определения, которого мы придерживаемся. С технопессимистического взгляда, GenAI — вовсе не интеллект в человеческом понимании, а еще один способ описания предметной области. С технооптимистического взгляда, генеративные модели уже сегодня способны решать весьма абстрактные задачи, отвечать на конкретные вопросы не хуже специалистов в своей предметной области. И это уже дает основания называть происходящее «интеллектом». А учитывая масштаб, вероятно, в ближайшие несколько лет произойдет дальнейшее увеличение интеллектуальных способностей таких моделей.

Стоит вспомнить, как примерно два года назад стали быстро набирать популярность плагины: по сути, это просто вызываемый программный код, с которым может взаимодействовать языковая модель. Скажем, если набрать в описании модели слово «калькулятор» в квадратных скобках, то будет инициирован вызов программного интерфейса, в который модель подставит те или иные значения в соответствующем формате и получит результат. Этот простой пример демонстрирует, что такой функциональный API позволяет вызывать типовые интерфейсы, например, интерфейс проектирования ER-диаграмм, чтобы создать схему базы данных. И специалисты уже смогут взаимодействовать с такими данными. Поэтому будет продолжаться усиление синергии между генеративными моделями и внешними функциями, по сути, программным отображением графического интерфейса веб- и настольных приложений, с которыми мы уже все привыкли работать.

В целом, одним из перспективных направлений развития LLM следует считать создание малых языковых моделей (Language Model, LM), ориентированных на решение микроскопических задач в конкретной нишевой отрасли, например, генерации математических доказательств. Они реализуются в виде ИИ-агентов и решают узкоспециализированные задачи. Яркий пример — StarCoder. Это маленькая языко-

вая модель — около трех миллиардов параметров. Вроде бы очень много? На самом деле все они запускаются на обычном компьютере. Модель генерирует программный код и специально обучена только для того, чтобы отвечать на вопросы и писать тесты по коду, генерировать сам код и документацию к нему в основном на Python, но в коробке есть и другие языки. Вы можете внедрить StarCoder в свой VS-код — помощника, который будет работать на том же высоком уровне качества, что и большие по размеру языковые модели.

ОБ ИНТЕЛЛЕКТЕ LLM МОЖНО СПОРИТЬ, НО МИР РАЗРАБОТКИ ПО НИИ СПОСОБНЫ ПЕРЕВЕРНУТЬ

Скорее всего, очень многие задачи будут автоматизироваться моделями. Уже сейчас любую разработку ПО можно декомпозировать на набор понятных задач. Разберем ситуацию на живом примере — работа менеджера проекта. Сегодня он открывает трекер задач (условный Jira) на своем рабочем месте, заводит декомпозированные задачи; далее уже разработчик их фиксирует — добавляет или исправляет конкретные куски кода, потом все это тестируется. Каждой отдельной задачей проверки, воспроизведения, исправления, документирования, описания занимается отдельный специалист. Но можно представить ситуацию, когда роль каждого такого специалиста будет декомпозирована на более простые этапы, которые уже будут выполняться моделями.

Да, сегодня интеллектуальных способностей модели хватает только для подготовки текстовой документации, их недостаточно для полной замены разработчика. Но мы ожидаем появления таких агентов через пару лет. Иными словами, тестировщик будет работать вместе с языковой моделью, которая будет помогать ему тестировать код. Но утверждение, что профессия разработчика исчезнет, не имеет под собой оснований. В лучшем случае LLM сможет рассчитывать на уровень Junior Data Scientist.

GenAI-модель — это не финальная модель, которая способна самостоятельно выполнять работу специалиста. Это просто хороший инструмент, помогающий ему в рутинных задачах.

«СЕГОДНЯ НАИБОЛЬШУЮ ОПАСНОСТЬ ПРЕДСТАВЛЯЕТ НЕ ГЕНЕРАЦИЯ ПОДДЕЛЬНЫХ ВИДЕО, А СОЗДАНИЕ ПРИ ПОМОЩИ АЛГОРИТМОВ ИИ ПОДДЕЛЬНЫХ ДОКУМЕНТОВ»



Александр ГОРШКОВ
директор по развитию
бизнеса «АЙРИС
ДЕВАЙСЕЗ»

Сегодня прорывы в ИИ принято связывать с бурным развитием нейросетевых моделей в сторону больших языковых моделей (LLM) для текстов и мультимодальных GenAI-систем, способных работать с мультимедийными данными. Каким образом это развитие сказывается на эволюции биометрических продуктов, решений, услуг?

О МИФИЧЕСКИХ ОПАСНОСТЯХ GENAI ДЛЯ БИОМЕТРИИ

Наиболее значимое влияние на развитие биометрических продуктов окажет улучшение систем искусственного зрения, алгоритмов распознавания образов и выделения объектов на изображениях, а также генерация фото и видео изображений. Правда, уровень хайпа в обществе не отражает реальных достижений в этой сфере, а порой даже провоцирует мифы, не имеющие отношения к реальности.

К сожалению, в общественном сознании укрепилось мнение, что для обучения и тестирования решений биометрической идентификации можно использовать сгенерированные изображения. Возможно, для нагрузочного тестирования такое решение может быть оправданным. Но для обучения биометрических нейросетевых алгоритмов такой подход может привести к непредсказуемым результатам. Высокая вероятность возникновения следующей проблемы: обучив систему на сгенерированных с искажением данных можно получить обратный эффект по точ-

ности идентификации при работе с реальными данными.

Еще один пример — синтез речевых сообщений на разных языках, который объявляется новой большой угрозой решениям биометрической идентификации. Опасность представляют не мультязыковые алгоритмы генерации речи, а принципиальные возможности генерации голоса. Дело в том, что мошенники обычно воздействуют на вполне определенную целевую аудиторию, и мультязыковая генерация речи для этих целей не требуется. Атака с использованием сгенерированного голоса обычно нацелена на определенную страну, регион или группу лиц. А для финансовых учреждений голосовые атаки могут вполне быть сведены к минимуму.

Основная проблема идентификации по голосу для банков заключается в том, что они могут использовать только контекстно-зависимую биометрию и не могут идентифицировать всех звонящих в контактный центр без их согласия. После первичной идентификации звонящего по номеру телефона, уточнения имени/отчества, становится известно, давал ли клиент согласие на биометрическую идентификацию. И если такое согласие было дано, то клиента просят назвать кодовое слово. Такую процедуру банк осуществляет сейчас. Но в данном случае можно реализовать контекстно-зависимую биометрическую идентификацию по ключевому слову или даже целой фразе. И здесь опять не нужны мультязыковые генераторы речи.

Конечно, наибольшую огласку в СМИ получают впечатляющие описания целевых АРТ-атак с использованием сгенерированных фейковых изображений, голоса, видео. Для защиты от атак с использованием дипфейков надо понимать, против чего или кого надо реализовать такую защиту. Если мы обсуждаем пример с переводом крупных сумм после телефонных или видео звонков, как в широко растиражированной истории с кражей миллионов долларов у гонконгской компании с помощью ряда

сеансов видеосвязи со сгенерированными персонажами, то в этом случае может защитить соблюдение разработанных регламентов. Например, перевод средств свыше определённой суммы — только по письменному распоряжению. Возможно, даже за подписью нескольких сотрудников или руководителей. Конечно, в этом случае обман и мошенничество нельзя исключить полностью, но такой обман можно реализовать и без использования дипфейков.

Аналогичная ситуация — с видеодоказательствами. Никакой суд не примет некий видеоролик в качестве видеодоказательства: необходимо доказать происхождение этого видео, найти автора или свидетеля того, что оно было сделано именно в том месте и в то время, о которых говорит заявитель.

Еще один момент, который обычно упускают из виду: почему-то биометрию и риски её использования связывают исключительно с лицом и голосом. Вероятно, это связано с тем, что использование этих модальностей регулируется законом. Но ведь есть и другие биометрики, которые значительно сложнее подделать, например, радужная оболочка глаз или рисунок вен ладони. Пока вообще не известны случаи обмана с использованием дипфейков алгоритмов мультимодальной и мультиспектральной биометрической идентификации. Понятно, что в ходе эволюционного развития отрасли био-

метрии будут появляться новые методы идентификации. Например, по акустике слухового канала или частотным характеристикам кожи.

При этом надо понимать, что в отсутствие прямого глобального противостояния, остаётся миф о непобедимости западных, американских или китайских технологий LLM и мультимодальных моделей. А для того чтобы этот миф оставался нерушимым, каждая страна проводит свои тесты и демонстрации. Вот только такие состязания не отражают реальной ситуации, так как зачастую в состязаниях принимают участие алгоритмы разных поколений. Наиболее достоверная информация имеется, наверно, только у спецслужб, но они ею делиться не будут...

... И О РЕАЛЬНЫХ МЕТОДАХ ЗАЩИТЫ

Сегодня наибольшую опасность представляет не генерация поддельных видео, а создание при помощи алгоритмов искусственного интеллекта поддельных документов. А именно замена фотографии поддельного документа на сгенерированную специальным образом, когда созданная фотография для документа, одновременно похожа и на лицо владельца паспорта, и на мошенника. Правда, и от такого мошенничества есть достаточно простая и эффективная защита с применением биометрических решений.

«БОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ ТРАНСФОРМИРУЮТ ВСЕ НОВЫЕ ИНДУСТРИИ»



Дмитрий ДЫРМОВСКИЙ
генеральный директор группы компаний ЦРТ

Крупный бизнес активно внедряет речевые технологии для улучшения качества коммуникаций с клиентами, импортозамещения и сокращения расходов на содержание контактных

центров. Речевая аналитика, диалоговые ассистенты уже доказали эффективность в телекоме, банковской сфере. Например, в контактном центре Мегафон речевая аналитика позволила перекалибровать 20% специалистов контроля качества в коучей при снижении количества претензий на 40%. Появление больших языковых моделей (LLM), таких как GigaChat, и совмещение их с традиционными алгоритмами машинного обучения позволило нам предоставить бизнесу новый функционал: суммаризацию содержания и тематик диалогов, оценку их результативности и тональности.

Теперь все данные агрегируются, расширяются и анализируются в онлайн-режиме прямо в момент обращения клиента. Это помогает оператору давать более

Таким образом, на вопрос, нужно ли компаниям уже сегодня срочно вносить риски дипфейков в список обязательных мероприятий ИБ, есть однозначный ответ: в компаниях для идентификации человека всегда есть регламенты и скрипты, и безопасность в конечном счете зависит от того, насколько качественно эти скрипты написаны и соблюдаются. Например, когда ко мне через взломанный аккаунт WhatsApp обращаются с просьбой срочно перевести деньги товарищу, я задаю вопрос, почему этот человек не обратился к своему брату? А я знаю, что брат у моего товарища есть. А потом спрашиваю, почему он не обратился к сестре? А вот сестры у товарища-то и нет... Скрипты можно усложнить конкретными именами или деталями, например: «Твой брат из Владимира может тебе перевести, я ему недавно вернул долг». Есть ещё один простой способ защиты от подобного типа мошенничества, и я его использую. Просто сразу отвечаю, что деньги перевёл, а на недоумённый вопрос как, если мне ещё не передали реквизиты карты, отвечаю, что по номеру телефона. Мошенник сразу исчезает.

Отдельная тема — легкость создания и сложности обнаружения политических дипфейков, что нередко обсуждается в блогосфере. В части этих «страшилок», с одной стороны, нужно помнить о том, что «каждый суслик — агроном»: каждый блогер считает себя экспертом, а миллионы подписчиков им внем-

люют и верят всей чепухе, которую несут такие блогеры. С другой стороны, есть официальные каналы информации, которые отвечают за достоверность распространяемой информации. И вот если эти каналы будут взломаны, и через них станут распространять недостоверную информацию, то никакие методы обнаружения дипфейков уже не помогут. В этом случае надо в первую очередь позаботиться о защите таких информационных каналов.

И вот это, действительно, серьёзный вызов. По сути, сегодня нет никаких технологических ограничений для создания качественных онлайн-дипфейков. Дипфейки создаются настолько хорошо, насколько за это готовы платить заказчики. Если это шуточный розыгрыш, то незначительные отличия никого не будут смущать. А если это целенаправленная атака, то тот, кто готов всецело верить всем и во всём, поверит и в случае посредственной реализации. Если люди верят «сотрудникам полиции» с явным зарубежным акцентом, почему бы им не поверить и в успех сделанный дипфейк?

Так к какому будущему стоит готовиться? В военном деле и в информационной безопасности есть такое понятие, как эшелонированная защита. Только комплексный подход обеспечит действенную защиту критической инфраструктуры, секретность информации, персональных данных и финансов. Исходя из этого, и надо выстраивать физическую и информационную защиту.

точные ответы на запросы, а также прогнозировать изменение клиентских предпочтений. Новое решение SpeechXplore — AI-решение для аналитики и управления коммуникациями в контактных центрах — фактически даёт новый уровень управления клиентским опытом.

Однако в целом влияние новых нейросетевых подходов гораздо более значительное. Можно утверждать, что большие языковые модели трансформируют все новые индустрии, и одна из них — промышленность. Важно, что LLM могут помочь в создании виртуальных ассистентов без ручного обучения, а также персональных Co-pilots — AI-помощников для целого набора различных специализаций. В настоящее время в компании СИБУР выполняется проект внедрения больших языковых моделей в промышленности — один из первых в стране. В нем объединены научные разработки ЦРТ в области обработки естественного языка (NLU) и возможности нейросетевой модели Сбер GigaChat. Уже созданы следующие решения:

- ◆ AI-ассистент инженера-диагноста: инженер может вести диалог с помощником о причинах неисправно-

сти оборудования. В ходе разговора сотрудник описывает в текстовом виде аномалии в работе оборудования, а обученный AI-помощник даёт релевантный ответ или формулирует гипотезы о причинах нехарактерного поведения агрегатов, а также предлагает возможные варианты устранения неисправности.

- ◆ AI-советчик для оптимизации закупки материально-технических ресурсов: искусственный интеллект позволяет перейти от статичных записей номенклатурных позиций к параметрическим карточкам. В результате система будет подбирать допустимые аналоги, имеющие преимущества по цене, качеству и доступности.

- ◆ AI-ассистент финансиста: агрегируя данные компании, он может отвечать на вопросы о динамике ключевых факторов, которые влияют на маржинальный доход.

Возможности современных технологий впечатляют, а их синергия с большими языковыми моделями позволит в будущем добиться еще более значимых технологических прорывов в самых разных индустриях.

«НАША ЗАДАЧА — ЗАЩИТИТЬ ДЕТЕЙ ОТ ИДЕОЛОГИЧЕСКИХ ДИВЕРСАНТОВ И МОШЕННИКОВ»



Василий ПИСКАРЁВ
председатель Комитета
Государственной
Думы по безопасности
и противодействию коррупции

В Государственной Думе я представляю Мордовию. В последнюю неделю каждого месяца обязательно приезжаю в республику — веду приём граждан, посещаю трудовые коллективы. Это важная составляющая депутатской работы. Такие встречи бывают очень полезны, потому что помогают оперативно вносить изменения в законы для решения проблем, с которыми обращаются избиратели.

В июне в региональном штабе общественной поддержки у меня состоялась встреча с членами Молодёжного парламента, молодого гвардейцами и слушателями Школы молодого политика. Ребят интересовало многое, в том числе вопросы информационной безопасности.

Например, прозвучало предложение закрепить в законодательстве понятие «киберволонтерство». Некоторые из участников признались, что, по сути, уже являются «интернет-дружинниками», оказывают помощь полиции в выявлении в интернете деструктивной информации, материалов, содержащих фейки, пропагандирующих наркотики, искажающих нашу историю.

Интерес молодёжи к этой теме понятен. Россия сегодня оказалась в эпицентре гибридной войны, когда сражения ведутся не только на поле боя, но и в информационном пространстве. И утрата позиций в виртуальной среде может нанести не меньший ущерб, чем поражение в реальном бою.

ПОД ПРИЦЕЛОМ — ДЕТИ

Примета времени — под прицелом информационных диверсантов и мошенников всё чаще оказываются дети и подростки. Поэтому информационное пространство

необходимо расценивать как поле защиты нашего суверенитета и государственных интересов.

В условиях СВО попытки идеологически обработать нашу молодёжь, вовлечь в опасную для жизни и здоровья деятельность, побудить к нарушению закона становятся всё изощрённее. После трагедии в «Крокусе» участились случаи прямого обращения зарубежных «кураторов» через мессенджеры и соцсети к детям с провокационными предложениями вплоть до совершения теракта. Хорошо узнаваем почерк украинских служб информационно-психологических операций и их западных хозяев.

Парламентская комиссия по расследованию преступных действий в отношении несовершеннолетних со стороны киевского режима, в работе которой я принимал участие, выяснила, что киевский режим создал целую систему вербовки школьников для совершения преступлений. По результатам анализа информации, проведённого комиссией, установлено, что эту работу курирует ЦИПСО СБУ (и его филиалы в Киевской, Винницкой и Харьковской областях).

В ходе переписки со школьниками в социальных сетях боевики склоняют их к противоправным действиям. Многие каналы и закрытые чаты в русскоязычном сегменте мировой паутины, которые пропагандируют терроризм, экстремизм, побуждают детей к суициду, моделируются с территории Украины агентами СБУ, которые и сами нередко ещё не достигли совершеннолетия.

Интересна классификация, которую киевский режим использует для завербованных детей: «Колумбайн» — подстрекатели к убийствам и расправам, «уличные бойцы» — используются при организации массовых беспорядков, «радикалы» — разжигатели негативных настроений среди молодёжи, «волки-одиночки» — исполнители диверсий и заданий кураторов из СБУ и т.д.

Вся эта деятельность представляет особую опасность. Ведь по данным Ассоциации участников интернет-индустрии «Лига безопасного Интернета», на деструктивные сообщества в Сети подписано около 10 миллионов российских детей.

На этом фоне предпринимаемые нами усилия по защите нашего информационного пространства от деструктивного воздействия из-за рубежа не выглядят избыточными.

Мы совсем неслучайно недавно запретили иноагентам заниматься просветительской и воспитательной работой среди несовершеннолетних. Нашей Комиссией Госдумы по расследованию фактов иностранного вмешательства выявлены разработанные организациями-иноагентами специально для наших детей брошюры, в которых пропагандируют однополюсные отношения, рассказывают в деталях о том, какие есть наркотики и как их потребляют, пропагандируют экстремистские взгляды, искажают нашу историю. Всё это находилось в свободном доступе в интернете.

Разработанные нами нормы об оперативной блокировке деструктивного контента позволяют сегодня компетентным органам оперативно пресекать деятельность многих деструктивных интернет-страниц и сообществ в соцсетях. Но признаемся, что этого недостаточно, потому что десятки и сотни таких страниц и сообществ появляются на месте заблокированных буквально каждый день. Очень важно, чтобы наши медиаплатформы не дожидались предписания со стороны регулирующих органов, а самостоятельно выявляли и блокировали деструктивный контент. И в данном случае полезную помощь как раз и могут оказать киберволонтеры, которых можно и нужно привлекать к модерации. Руководству медиаплатформ стоит над этим подумать. Защитить наших детей, да и взрослых от действий информационных диверсантов — наша общая задача.

КОГДА ИГРА УЖЕ НЕ ИГРА

Госдепартамент США недавно объявил новый конкурс на получение дорогостоящих грантов, связанных с разработкой компьютерных игр. Цель вроде бы благородная: игры должны «повышать когнитивную устойчивость к авторитаризму и продвигать демократические нормы и ценности, повысить скептицизм игроков к иностранной пропаганде и дезинформации». Но на самом деле за этими словами скрывается активное противодействие России и попытка в игровой форме программировать детей и подростков в антироссийском ключе. Именно так воспитывают новое поколение, с детства настроенное против России и русских.

В администрации США не скрывают, что вкладывают миллионы долларов в украинскую индустрию видеоигр, чтобы научить местное население бороться с «российской дезинформацией и пропагандой». Надо ли удивляться, что после всесторонней обработки дети на Украине говорят, что вырастут и будут «резать русню»?

В Вашингтоне вызвал серьёзное раздражение успех проведённых Россией в этом году Игр будущего в Казани, соединивших традиционный спорт и компьютерные игры. Объединяющий посыл зародившегося в нашей стране фиджитал-спорта идёт вразрез с политикой США на разобщение и конфронтацию даже на игровом поле. А ведь украинские киберспортсмены на многих соревнованиях входили в одну команду с российскими, и у них не было никаких противоречий, а вот совместные победы были. И этот мостик дружбы между российскими и украинскими подростками хотят уничтожить на корню.

Не секрет, что антироссийская обработка украинской молодёжи, в том числе с помощью компьютерных игр, началась давно. Ещё в середине 90-х хорошо знакомый всем по баталиям на тетрадных листочках «Морской бой» на Украине перевели в компьютерный формат. Только вот топить в этой игре предстояло не абстрактные корабли, а российский Черноморский флот у берегов Крыма. Уже тогда Украину готовили к войне с Россией. Сначала в виде игры.

Для оказания воздействия на психику детей и подростков СБУ сегодня использует наработки спецслужб США и Великобритании. Там провели множество исследований и установили, что воспитание и обучение проходит значительно успешнее, если осуществляется в игровой форме. С помощью игр, не обязательно компьютерных, можно привить ребёнку нужную идеологию и заложить необходимые стереотипы поведения, побудить к совершению необдуманных поступков. Впрочем, такие же методы воздействия применяются и в отношении взрослых.

Впервые мы столкнулись с подобными приёмами деструктивного манипулирования сознанием детей несколько лет назад, когда в соцсетях появились опасные игры «Синий кит» и ряд других. В игровой форме их организаторы последовательно склоняли детей к опасным экспериментам над собой, а потом и к суициду. Тогда мы ввели уголовную ответственность для создателей групп смерти как за умышленное убийство. И ряд таких «кураторов» были выявлены и отправлены за решётку.

Комиссия Госдумы по расследованию фактов иностранного вмешательства выявила созданные иноагентами настольные игры, которые искажают историю России. С участием зарубежных организаций был снят и распространялся интерактивный художественный фильм, который учит школьников участию в протестных акциях. Допустить распространения подобного контента было нельзя.

В последние годы Государственная Дума приняла более 30 законов, направленных на противодействие деструктивной работе иноагентов и нежелательных в России организаций. Эту донастройку законодательства мы продолжаем и сегодня. В первом чтении принят пакет законопроектов, которые позволят выдворять из России не только неправительственные зарубежные организации, которые работают на разрушение нашей страны, но и те, которые учреждены иностранными госорганами и также представляют угрозу нашей безопасности и конституционному строю. За сотрудничество и организацию деятельности таких организаций на территории нашей страны предусмотрена административная и уголовная ответственность.

СЕРЫЕ SIM-КАРТЫ УЙДУТ В ПРОШЛОЕ

Мошенники сегодня придумывают всё новые и новые способы обмана граждан и используют самые современные технологии для отъёма денег у потерпевших, но всё же в очень многих случаях мы имеем дело не столько с хакерскими атаками, сколько с обманом и манипуляцией, когда жертвы сами переводят деньги



Встреча В. И. Пискарева с молодёжью в региональном Штабе общественной поддержки в Саранске

преступникам. А помогает уходить от ответственности очень часто использование так называемых серых sim-карт, установить истинных владельцев которых крайне затруднительно.

Госдума недавно приняла в первом чтении внесённый мною совместно с коллегой Антоном Горелкиным законопроект, который поможет укрепить безопасность граждан и существенно сократить возможности для совершения мошеннических действий и других преступлений с использованием мобильной связи. Предложенный нами механизм оформления договоров на оказание услуг связи поможет сделать эту систему более безопасной и прозрачной.

Во-первых, абонентов в обязательном порядке теперь будут оповещать обо всех зарегистрированных на них номерах, и каждый сможет заблокировать договор, который он не заключал. При этом сведения обо всех своих «симках» можно будет проверить в личном кабинете на Госуслугах.

Во-вторых, иностранцы смогут оформить на себя не более 10 номеров и только при личном посещении салона дилера или оператора связи. Для них также станет обязательным прохождение идентификации и (или) аутентификации с использованием биометрии. А код идентификации трубки (IMEI) будет включаться в договор.

Эти и другие предложенные нами меры должны способствовать наведению порядка и улучшить криминальную ситуацию в сфере мобильной связи. Необходимо

исключить пробелы в законах, которые позволяют преступникам уходить от наказания. Ну а ответственность за нарушение вводимых законопроектом требований предлагается внести в Кодекс об административных правонарушениях отдельной законодательной инициативой.

Напомню также, что несколько месяцев назад я вместе с коллегами выступил с инициативой ужесточить наказание для интернет-мошенников путём введения конфискации имущества, полученного в результате совершения преступлений в сфере компьютерной информации. Соответствующие поправки внесены в статью 104.1 Уголовного кодекса и вступили в силу. Эта мера послужит восстановлению справедливости и позволит направить конфискованные средства на возмещение ущерба потерпевшим.

Но, конечно, одними технологическими и законодательными мерами проблему нам не решить. Чтобы успешнее противостоять мошенникам, необходимо усилить разъяснительную работу, которая ведётся в СМИ и банками напрямую, но пока не достигает своих целей. Ведь люди всё равно верят обманщикам, которые часто оказываются тонкими психологами, а это приводит к материальным потерям и личным трагедиям.

КВАРТАЛЬНЫЙ ОТЧЁТ

О НОРМАТИВНОМ РЕГУЛИРОВАНИИ В СФЕРЕ ИБ



Екатерина РАЧКОВА
обозреватель BIS Journal

С 01.04.2024 ВСТУПИЛИ В СИЛУ:

♦ ГОСТ Р 71206–2024 «Защита информации. Разработка безопасного программного обеспечения. Безопасный компилятор языков C/C++. Общие требования». Стандарт направлен на профилактику появления уязвимостей в программах и содержит общие требования к безопасному компилятору программ на языках C и C++, задача которого — не вносить в бинарный код программы ошибки, которых не было в исходном коде программы и которые могут появиться в ходе компиляции, в том числе в ходе выполнения оптимизаций кода программы. Настоящий стандарт определяет требования к динамической компоновке и загрузке программ, выполнение которых необходимо для поддержки ряда возможностей безопасного компилятора, содержит требования к функциям безопасного компилятора и задаёт требования к методике проверки требований к нему.

♦ ГОСТ Р 71207–2024 «Защита информации. Разработка безопасного программного обеспечения. Статический анализ программного обеспечения. Общие требования». Стандарт определяет общие требования к внедрению и выполнению статического анализа ПО, а также исходные данные, необходимые для его выполнения, требования к методам статического анализа, инструментам анализа (статическим анализаторам) и к специалистам, участвующим в анализе, а также методику проверки устанавливаемых требований к инструментам анализа.

♦ ГОСТ Р 71252–2024 «Информационная технология. Криптографическая защита информации. Протокол защищённого обмена для промышленных систем». Настоящий стандарт содержит описание протокола CRISP — CRyptographic Industrial Security Protocol — неинтерактивного протокола защищённой передачи данных, разработанного для применения в промышленных системах. Протокол CRISP может быть использован для обеспечения конфиденциальности и аутентификации сообщений и для защиты от навязывания повторных сообщений. Стандарт введён взамен Р 1323565.1.029–2019.

♦ Предварительный национальный стандарт ПНСТ 905–2023 «Критическая информационная инфраструктура. Доверенные программно-аппаратные комплексы. Термины и определения». Стандарт устанавливает термины и определения основных понятий, относящихся к доверенным программно-аппаратным комплексам, применяемым на объектах критической информационной инфраструктуры. Приведённые определения можно при необходимости изменять, вводя в них производные признаки, раскрывая значения используемых в них терминов, указывая объекты, входящие в объём определяемого понятия. Изменения не должны нарушать объём и содержание понятий, определённых в настоящем стандарте. Данный предварительный национальный стандарт будет действовать в период с 01.04.2024 по 01.04.2027.

♦ Положение Центрального банка Российской Федерации от 17.08.2023 № 821-П «О требованиях к обеспечению защиты информации при осуществлении переводов денежных средств и о порядке осуществления Банком России контроля за соблюдением требований к обеспечению защиты информации при осуществлении переводов денежных средств» (взамен Положения Банка России № 719-П). Новое Положение не несёт практически ничего нового для банков, ранее выполнявших 719-П, зато вводит требования к защите информации для операторов электронной платформы (операторов финансовой платформы/информационной системы, в которой осуществляется выпуск цифровых финансовых активов, операторов обмена цифровых финансовых активов при оказании указанными операторами услуг расчётов по сделкам, совершённым с использованием электронной платформы, пользователям электронной платформы).

ПРЕЗИДЕНТ И ПРАВИТЕЛЬСТВО РФ

11.04.2024 опубликован Указ Президента Российской Федерации от 11.04.2024 № 254 «О внесении изменений в перечень сведений, отнесённых к государственной тайне, утверждённый Указом Президента Российской Федерации от 30.11.1995 № 1203». К числу государственных органов и организаций, наделённых полномочиями по распоряжению сведениями, раскрывающими мероприятия по использованию объектов транспортной инфраструктуры и (или) транспортных средств в Российской Федерации в целях обеспечения безопасности государства, специальные меры по обеспечению безопасности перевозок или сохранности грузов, объёмы воинских перевозок или маршруты транспортировки вооружения, военной техники, отнесено Минцифры Российской Федерации.

07.05.2024 опубликован Указ Президента Российской Федерации «О национальных целях развития Российской Федерации на период до 2030 года и на перспективу до 2036 года» № 309.

Непосредственно отрасли ИТ и информационной безопасности в данном Указе касается:

- ♦ необходимость формирования рынка данных, их активного вовлечения в хозяйственный оборот;

- ♦ автоматизация большей части транзакций в рамках единых отраслевых цифровых платформ и модели управления на основе данных с учётом ускоренного внедрения технологий обработки больших объёмов данных, машинного обучения и искусственного интеллекта в ключевых отраслях экономики и социальной сферы, в том числе здравоохранении и образовании;

- ♦ переход к 2030 году не менее 80% российских организаций ключевых отраслей экономики на использование базового и прикладного российского программного обеспечения в системах, обеспечивающих основные производственные и управленческие процессы, в том числе — для государственных органов, государственных корпораций, государственных компаний и хозяйственных обществ, в уставном капитале которых доля участия Российской Федерации в совокупности превышает 50%, и аффилированных с ними юридических лиц — увеличение к 2030 году до 95% доли использования российского программного обеспечения.

Правительству Российской Федерации в срок до 01.09.2024 предстоит разработать (скорректировать) блок национальных проектов по обеспечению технологического лидерства, привести свои акты в соответствие с настоящим Указом и представить предложения по приведению актов Президента Российской Федерации в соответствие с настоящим Указом.

24.05.2024 опубликовано постановление Правительства Российской Федерации от 23.05.2024 № 639 «Об утверждении Положения о схеме пропуска трафика через технические средства противодействия угрозам устойчивости, безопасности и целостности функционирования на территории РФ информационно-телекоммуникационной сети „Интернет“ и сети связи общего пользования, в том числе пропуска трафика на присоединённую сеть связи оператора связи, оказывающего услуги по предоставлению доступа к информационно-телекоммуникационной сети „Интернет“». Настоящим Положением установлены требования к содержанию схемы пропуска трафика через технические средства противодействия угрозам устойчивости, безопасности и целостности функционирования в России сети „Интернет“ и сети связи общего пользования. Изложен порядок предоставления схемы на согласование, установлены сроки согласования и предоставления ответа заявителю, а также описана процедура и сроки повторного согласования схемы при её изменении. Постановление вступает в силу с 01.09.2024.

14.06.2024 опубликован Указ Президента Российской Федерации от 13.06.2024 № 500 «О внесении изменений в Указ Президента Российской Федерации от 01.05.2022 № 250 „О дополнительных мерах по обеспечению информационной безопасности Российской Федерации“». Настоящий Указ вступает в силу с даты подписания (13.06.2024) и расширяет меры обеспечения информационной безопасности (далее — ИБ) Российской Федерации. С 01.01.2025 органам государственной власти, государственным компаниям, стратегическим предприятиям

и системообразующим организациям будет запрещено не только использовать средства защиты информации из недружественных стран, но и пользоваться их сервисами (работами, услугами) по обеспечению ИБ. Кроме того, Указ устанавливает требования к аккредитованным центрам ГосСОПКА, которые будут привлекаться для реагирования на инциденты в ведомствах и субъектах критической информационной инфраструктуры. Такие центры привлекаются в случае необходимости для предупреждения кибератак и устранения их последствий. Требования к аккредитации таких центров будут разработаны ФСБ России. Также ФСБ России будет осуществлять контроль их деятельности.

ПЕРСОНАЛЬНЫЕ ДАННЫЕ

05.04.2024 опубликовано постановление Правительства Российской Федерации от 01.04.2024 № 408 «О видах биометрических персональных данных, на которые распространяется действие Федерального закона „Об осуществлении идентификации и (или) аутентификации физических лиц с использованием биометрических персональных данных, о внесении изменений в отдельные законодательные акты Российской Федерации и признании утратившими силу отдельных положений законодательных актов Российской Федерации“». Постановление определяет, какие виды биометрических персональных данных подпадают под действие Закона об идентификации и аутентификации физических лиц с использованием биометрических персональных данных: изображение лица человека, полученное с помощью фото- и видеоустройств, и запись голоса человека, полученная с помощью звукозаписывающих устройств. Постановление вступает в силу с 01.09.2024 и действует до 01.09.2027.

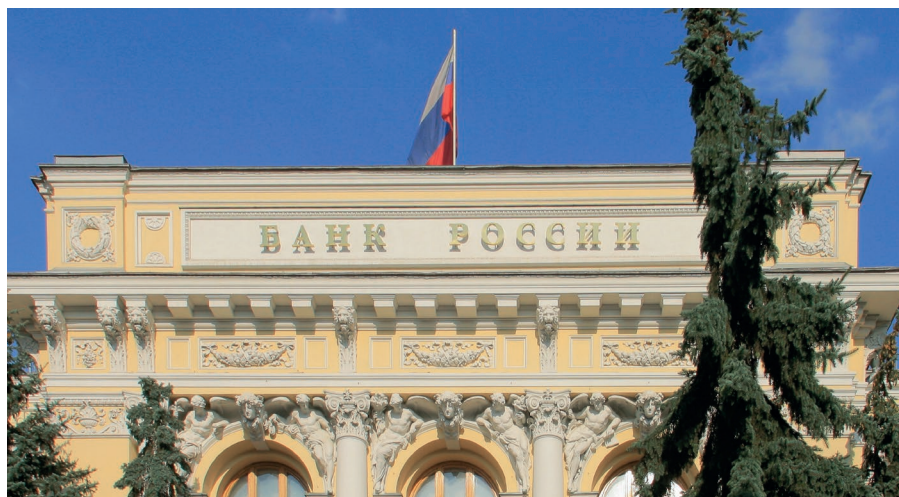
24.05.2024 опубликован проект Указания Центрального Банка Российской Федерации «О требованиях к организации банками процесса сбора и размещения биометрических персональных данных физического лица в Единой биометрической системе и к применению банками Единой биометрической системы посредством их официального сайта в сети „Интернет“, а также их мобильного приложения». Указание будет устанавливать требования к организации банками процесса сбора и размещения биометрических персональных данных в ЕБС и к применению банками ЕБС посредством их официального сайта и мобильного приложения. Сбор комментариев по проекту нормативного акта проводился до 06.06.2024.

ФСТЭК РОССИИ

17.04.2024 опубликован приказ ФСТЭК России от 01.12.2023 № 240 «Об утверждении Порядка проведения сертификации процессов безопасной разработки программного обеспечения средств защиты информации». Данный приказ вступает в силу 01.06.2024.

КИИ

06.05.2024 опубликован методический документ ФСТЭК России от 02.05.2024 «Методика оценки показателя со-



ЦЕНТРАЛЬНЫЙ БАНК РФ

18.04.2024 Центральным банком Российской Федерации опубликовано информационное письмо от 15.04.2024 № ИН-012-56/26 о разъяснении положений части 12 статьи 16 Федерального закона № 572-ФЗ «Об осуществлении идентификации и (или) аутентификации физических лиц с использованием биометрических персональных данных, о внесении изменений в отдельные законодательные акты Российской Федерации и признании утратившими силу отдельных положений законодательных актов Российской Федерации». В частности, уточняется, что использование информационных

систем в целях аутентификации на основе биометрии допускается после предоставления организации, осуществляющей аутентификацию, актов проверки, за которыми рекомендуется обращаться в Центральный банк Российской Федерации с заявлением в форме электронного документа.

05.06.2024 ФСТЭК России опубликовано информационное сообщение от 27.05.2024 № 240/82/1376 «О порядке представления субъектами критической информационной инфраструктуры сведений о результатах присвоения объектам критической информационной инфраструктуры одной из категорий значимости либо об отсутствии необходимости присвоения ему одной из таких категорий». В случае изменения сведений о результатах присвоения объекту критической информационной инфраструктуры (далее — КИИ) одной из категорий значимости либо об отсутствии необходимости присвоения ему одной из таких категорий (далее — сведения) субъект КИИ должен направить во ФСТЭК России новые актуализированные сведения по Форме, утверждённой приказом ФСТЭК России от 22.12.2017 № 236 (далее — Форма). В случае если указанная актуализация не приводит к изменению категории значимости или пересмотру решения об отсутствии необходимости присвоения категорий значимости, актуализированные сведения могут направляться во ФСТЭК России электронным документом, записанным на машинном носителе информации. Указанный машинный носитель информации прикладывается к сопроводительному письму с указанием в нём размера записанных на нём электронных документов. В иных случаях сведения направляются в печатном и электронном виде (в формате.ods и (или).odt) по утверждённой Форме.

Также напоминаем, что с 01.09.2024 вступят в силу требования Постановления Правительства Российской Федерации от 14.11.2023 № 1912, согласно которым субъекты КИИ Российской Федерации должны будут начать переход (который необходимо будет завершить до 01.01.2030) на преимущественное использование на значимых объектах КИИ доверенных программно-аппаратных комплексов (далее — ПАК). Также согласно данному Постановлению не допускается использование субъектами КИИ на принадлежащих им значимых объектах КИИ ПАК, приобретённых с 01.09.2024 и не являющихся доверенными (с некоторыми оговорками и исключениями).

Также 14.05.2024 Банк России опубликовал проект Указания «О внесении изменений в Положение Банка России от 03.08.2023 № 820-П „О платформе цифрового рубля“». Проект указания был разработан в целях уточнения нормативного регулирования обращения цифрового рубля, в том числе в части функционирования платформы цифрового рубля, а также в связи со вступлением в силу с 01.01.2025 отдельных законодательных актов Российской Федерации (в части правовых основ обращения взыскания на цифровые рубли, учитываемые на счёте цифрового рубля должника). Действие проекта предлагалось распространить на кредитные организации, а также на физических и юридических лиц, планирующих совершать операции с цифровыми рублями. До 22.05.2024 проводился сбор комментариев по проекту нормативного акта.

27.05.2024 Банком России также был опубликован обзор отчётности об инцидентах информационной безопасности при переводе денежных средств за I квартал 2024 года.

27.05.2024 Банком России также был опубликован обзор отчётности об инцидентах информационной безопасности при переводе денежных средств за I квартал 2024 года.

ОТРАСЛЕВЫЕ ДОКУМЕНТЫ

06.06.2024 Министерством промышленности и торговли Российской Федерации опубликован приказ от 28.03.2024 № 1310 «Об упорядочении обращения со служебной информацией ограниченного распространения в Министерстве промышленности и торговли Российской Федерации». Приказом определены категории должностных лиц, уполномоченных относить служебную информацию к сведениям ограниченного распространения. Кроме того, установлены порядок передачи служебной информации внешним сторонам, правила снятия пометки «Для служебного пользования» с носителей соответствующей информации, требования к защите служебной информации.

Я КАЛЕНДАРЬ ПЕРЕВЕРНУ –
И СНОВА...

3 СЕНТЯБРЯ

НА ИНФОБЕРЕГ ВСЕМ ПОРА!



Г. СОЧИ,
RADISSON COLLECTION
PARADISE RESORT & SPA

XXIII ВСЕРОССИЙСКИЙ ФОРУМ
ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ.
РЕГУЛИРОВАНИЕ. ТЕХНОЛОГИИ. ПРАКТИКА



ВИПФОРУМ
Организация деловых событий



3-6 сентября
Ждем вас!

ОБЫЧНО НЕ ВИДНЫ...

КОМПЛЕКСНЫЙ ПОДХОД В УПРАВЛЕНИИ ЖИЗНЕННЫМ ЦИКЛОМ ТЕХНОЛОГИЧЕСКИХ СЕРТИФИКАТОВ SSL/TLS И СОПРОВОЖДЕНИИ ИОК RSA



Константин ЛЕОНТЬЕВ
ведущий архитектор и управляющий партнёр
ООО «Клируэй Текнолоджис» (Clearway Integration)

1. ПОЧЕМУ ТЕМА УПРАВЛЕНИЯ ЖИЗНЕННЫМ ЦИКЛОМ ТЕХНОЛОГИЧЕСКИХ СЕРТИФИКАТОВ АКТУАЛЬНА И ВАЖНА?

Когда мы говорим с бизнесом о цифровых сертификатах, на ум сразу

приходят сертификаты разнообразных электронных подписей, пользовательские токены и шифрование пользовательских данных. И все давно уже привыкли к этому пониманию и задачам, которые решает бизнес с использованием сертификатов. Данная статья посвящена менее

известной, но не менее актуальной сегодня теме управления **технологическими сертификатами**, которые пользователям или бизнесу обычно не видны. Исключением являются случаи, когда вовремя не продлён сертификат одного из публичных веб-сайтов, в результате пользователи при входе на сайт видят в браузере яркое и заметное предупреждение о небезопасном веб-сервере (некоторые такие примеры на слуху — Microsoft, Google, Telegram и другие).

Однако руководителям направлений ИТ и ИБ крупных компаний в своей профессиональной жизни приходится сталкиваться с гораздо более серьёзными проблемами, свя-

SSL/TLS-сертификатов с одного года до 6 месяцев, а в будущем даже ещё короче — до 3 месяцев. В то же время аварии и нештатные ситуации, вызванные несвоевременным обновлением даже одного сертификата, могут обходиться бизнесу в десятки и сотни миллионов рублей.

Не стоит на месте и внутрикорпоративная инфраструктура: в ней появляются внутренние, ориентированные на сотрудников, подрядчиков и партнёров корпоративные веб-сервисы и приложения, частные облачные платформы, интеграционные сервисы и шины данных, построенные на веб-технологиях, контейнерных средах Kubernetes/OpenShift и других технологиях, нуждающихся в сертификатах. Активно развиваются внутренние и внешние интеграционные шины и шлюзы API, реализованные на базе семейства протоколов HTTP. Вся эта инфраструктура требует непрерывного выпуска и применения технологических сертификатов всё в большем и большем количестве. В одной из организаций, в которых мы некоторое время назад внедрили решение по автоматизации выпуска технологических сертификатов в середине 2010-х годов количество таких сертификатов составляло всего несколько тысяч в год. Сегодня же в этой организации в год выпускается, обновляется и используется несколько миллионов технологических сертификатов.

Отдельно необходимо отметить, что для инфраструктуры современных корпоративных приложений сертификаты используются не только при шифровании трафика, но, прежде всего, как инструмент идентификации и проверки подлинности клиента и сервера друг для друга (mutual TLS — mTLS), что даёт приложениям более безопасный аналог логины и пароля для авторизации.

В связи с тем, что мы исключили из предмета рассмотрения настоящей статьи все другие типы сертификатов, кроме технологических — следует сделать важное уточнение: далее в тексте будут рассматри-

ваться сертификаты x.509 с алгоритмами шифрования RSA. Такое ограничение понятно и объясняется тем, что большинство современных библиотек кода и информационных систем ещё не в полной мере в рамках семейства протоколов SSL/TLS и IPSec поддерживают сертификаты и криптографию ГОСТ. Безусловно, ГОСТ SSL/TLS получает все более широкое распространение, но чаще пока применяется на каналобразующем оборудовании. Нас же в первую очередь интересует взаимодействие микро-сервисов, приложений и интеграционных шин.

2. КРАТКИЙ ОБЗОР ТРАДИЦИОННЫХ ПОДХОДОВ И ТЕХНОЛОГИЙ

Для общего контекста вспомним, что Инфраструктура Открытых Ключей (ИОК или PKI) — это не только непосредственно Удостоверяющий Центр, а ещё и вся та технологическая инфраструктура сопровождающая УЦ и обеспечивающая взаимодействие с ним. Инструменты и технические решения, которые предлагают нам традиционные Инфраструктуры Открытых Ключей (ИОК, PKI) для решения, описанной в предыдущем разделе задачи, на самом деле весьма ограниченные и, мягко скажем, недостаточно универсальные. Но прежде, чем дать краткий обзор этих средств давайте всё же перечислим сами задачи.

К основным задачам, стоящим перед современной ИОК, предназначенной для выпуска технологических сертификатов, относятся:

1. Формирование подходящей потребителю ключевой пары и её безопасное хранение.
2. Формирование соответствующего потребностям потребителя запроса CSR, сбор сведений о потребителе.
3. Проверка запроса сертификата (CSR) на соответствие ряду требований ИБ.
4. Доставка запроса (CSR) до удостоверяющего центра.
5. Выпуск сертификата непосредственно на УЦ.
6. Доставка подписанного УЦ сертификата до системы-потребителя.

занными с технологическими сертификатами, когда виновником сбоев в работе критически важных бизнес-процессов оказывается просроченный сертификат.

Каких-то пять-десять лет назад, когда использование протоколов SSL/TLS в сети Интернет было ещё рекомендательным, самоподписанные сертификаты — нормой, а очень длительные сроки действия SSL-сертификатов не вызывали ни у кого беспокойства, мы вообще не думали о том, что просроченный сертификат может стать какой-то заметной проблемой.

Сейчас же, под действием растущих угроз и числа уязвимостей, всё стремительно меняется, в том числе ужесточаются общепринятые нормы безопасности в сети Интернет. Некоторые крупные корпорации, влияющие на стандарты Интернет, уже предлагают увеличить длину ключей и сократить сроки действия

7. Установка сертификата в нужное хранилище системы-потребителя.

8. Информирование и конфигурирование системы-потребителя о новом сертификате.

9. Своевременная замена устаревшего сертификата в системе-потребителе на обновлённый с выполнением всех шагов 1–8.

10. Обеспечение необходимого уровня надёжности и производительности ИОК (поддержка миллионов и десятков миллионов сертификатов) для выполнения задач 1–9.

Кроме перечисленных выше задач, есть ещё ряд задач ИОК, о которых часто забывают, но они от того не становятся менее значимыми:

1. Поддержание и распространение актуального списка(ов) доверенных корневых и промежуточных УЦ в рамках всей организации.

2. Поддержание и распространение актуального списка(ов) отозванных сертификатов в рамках всей организации.

3. Логирование не только фактов выпуска, но и фактов доставки и установки готовых сертификатов.

4. Мониторинг доступности ИОК, её отдельных компонентов и главной её функции — выпуска сертификатов.

5. Независимый мониторинг срока истечения выпущенных сертификатов и своевременное уведомление владельцев об этом.

6. Мониторинг и анализ интенсивности выпуска сертификатов различных типов, выявление пиковых нагрузок.

7. Обеспечение функции Личного Кабинета (одного окна) для пользователей сертификатов.

8. Механизмы оперативного отзыва сертификатов и удаления связанного с ними ключевого материала.

9. Учет технологических окон при выполнении обновления сертификатов.

10. Учет требований безопасности по обращению ключевого материала и защите приватных ключей.

11. В случае использования внешних сертификатов (сертификатов, полученных от УЦ за пределами своей организации), отслеживание не только их сроков действия, но и фактов их отзыва.

12. Инвентаризация фактически используемых сертификатов: сетевым сканированием и поиском в хранилищах ОС/Приложений.

13. Балансировка нагрузки и автоматическое переключение между несколькими однотипными УЦ.

14. Сквозной поиск в единой базе всех используемых сертификатов и предоставление необходимых расширенных сведений о сертификатах в другие автоматизированные системы.

15. Резервное копирование и восстановление ИОК и отдельных её компонентов.

16. Контроль за исполнением регулярных и однократных задач по обслуживанию ИОК.

17. Поддержка сложно сегментированных сетевыми экранами сетей.

18. Проверка актуальности версий установленных крипто-провайдеров и крипто-библиотек.

19. Обеспечение жизненного цикла короткоживущих сертификатов (сроком действия до 14 дней) — востребованы в Kubernetes ISTIO Service Mesh, сценариях Network Access Control/Network Access Protection при использовании 802.1x и IPSec, сценариях подтверждения онлайн транзакций и т.п.

20. Поддержка и динамический выбор УЦ от разных производителей.

В крупных организациях все перечисленные действия требуют автоматизации именно из-за большого количества систем и необходимых им сертификатов. При этом крайне желательно, чтобы эта автоматизация была централизованной и управлялась по общим правилам и стандартам под контролем ролевой модели доступа к функциям управления.

Теперь давайте попробуем оценить как традиционные ИОК и связанные с ними технологии позволяют решать перечисленные выше задачи. Начнем с наиболее очевидной и важной части ИОК — с Удостоверяющих Центров. Тут большинство традиционных решений обеспечивает выполнение основных задач: 3 и 5, частично 9 и 10. Из вспомогательных задач ИОК чаще всего оказываются решёнными 2, частично 3, 4, 6, 7, 8, 10 и 14, обыч-

но решена также задача 15. Теперь посмотрим, какие распространенные технологии позволяют расширить функционал традиционных УЦ:

♦ **SCEP** (в реализации Microsoft: сервис NDES) — протокол, описанный в RFC 8894 и обеспечивающий простой механизм запроса и выпуска сертификатов для доверенных сетевых устройств (роутеры, маршрутизаторы, МФУ и т.п.). Также часто этот протокол используется в связке с Mobile Device Management-системами для формирования и распространения сертификатов и списка доверенных УЦ на мобильные устройства пользователей.

♦ **CMP и EST** — группа протоколов (RFC 9480, RFC 5785), разработанных консорциумом вендоров во главе с Verisign, цель которых — заменить SCEP и обеспечить гибкие функции генерации ключевых пар, выпуска и отзыва сертификатов, восстановления и распространения ключей. Поддерживается в основном библиотеками для разработчиков и пакетом OpenSSL v3.

♦ **ACMEv2** — протокол, описанный в RFC 8555, разработанный командой Internet Security Research Group для проекта Let's Encrypt и использующий агента CertBot на Python или библиотеки для множества разнообразных языков программирования. Основан на принципе подтверждения владения DNS-именем доменов и зон с использованием асимметричных ключей в TXT-записях DNS-зон и автоматизирующий выпуск сертификатов SSL/TLS для полного доменного имени сервера или зоны с wildcard-поддоменами.

♦ **Microsoft Enterprise PKI** — семейство проприетарных технологий для ИОК, встроенных в ОС Windows, службу каталогов Microsoft Active Directory и API-вызовов к УЦ Microsoft CA, основанных на стеке протоколов Microsoft RPC. Позволяет автоматизировать процесс распространения доверенных корневых сертификатов (в составе Active Directory GPO), выпуск и обновление клиентских и технологических сертификатов для приложений и серверов, разрабатываемых корпорацией Microsoft.

♦ Системы автоматизации и мониторинга, нацеленные на ИОК, но при этом требующие тщательной настройки и развития.

3. ОБЪЕКТИВНЫЕ ОГРАНИЧЕНИЯ И РИСКИ СУЩЕСТВУЮЩИХ ТЕХНОЛОГИЙ

Если бы эти решения были универсальными, то необходимости в данной статье вероятно бы не было, но перечисленные в предыдущей главе решения не универсальны. Давайте разберём их наиболее существенные ограничения и проблемы.

SCEP — это в первую очередь не сервис или продукт, а протокол, поэтому его должны поддерживать как минимум потребители сертификатов и сам УЦ. Если с точки зрения УЦ проблем с поддержкой этого протокола нет — таких УЦ много, то при пристальном изучении поддержки со стороны клиентов оказывается, что не такое уж и большое разнообразие клиентов поддерживают этот протокол и ещё меньшее количество корректно пользуются всеми его возможностями и делают это безопасно. Например, многие клиенты произведённого в Китае оборудования не поддерживают обновление списка доверенных корневых и выпускающих УЦ по протоколу SCEP. Также очень многие клиенты специфическим образом нацелены именно на проприетарную реализацию SCEP от Microsoft (сервис NDES) в связке с Microsoft Active Directory. Наиболее активно используемыми SCEP устройствами являются те, которые недоступны или ограничены в России в рамках стратегии импортозамещения и действующих санкций. К таким устройствам относится сетевое оборудование Cisco и устройства Apple на базе iOS и её деривативов. Также надо отметить, что протокол SCEP не позволяет понять, успешно ли был выпущен и доставлен сертификат до конечного потребителя, у него нет механизмов, позволяющих контролировать своевременное обновление сертификатов, и нет средств для проксирования или применения в сегментированных корпоративных сетях. Наконец, в протоколе SCEP от-

сутствуют механизмы полноценного контроля доверия клиенту и содержанию его запроса на выпуск сертификата. Если клиент захочет выпустить сертификат с неподходящим для него Common Name или Subject Alternative Name, то только УЦ в режиме ручной проверки запросов и ручного выпуска сертификатов может предотвратить такой сценарий. Поэтому SCEP подходит только для абсолютно доверенных потребителей, самостоятельно проверяющих и формирующих запросы на сертификаты, и потому не позволяет полноценно автоматизировать весь процесс выпуска.

CMP и **EST** — эти протоколы, к сожалению, реализованы в очень ограниченном наборе проприетарных зарубежных продуктов из недружественных стран, например: EJBCA от шведско-американской компании Keyfactor, Nexus Certificate Manager от европейской Nexus Group, некоторые продукты американской компании Entrust, УЦ от финской INSTA и ряд облачных УЦ. Помимо того, что эти продукты и их производители не могут быть использованы в России, в основном их реализация ограничивается серверным функционалом и поддержкой именно протоколов, а потому полноценного контроля за выпускаемыми сертификатами и их применением на конечных ОС потребителей всё равно не обеспечивается, также нет важных механизмов мониторинга сроков действия сертификатов и ключей. В итоге это требует полноценного проекта по разработке и реализации клиентского функционала в применении к конкретным информационным системам и потребителям.

ACMEv2 — этот протокол ориентирован на один-единственный, хотя и важный сценарий выпуска SSL/TLS-сертификатов, а именно на выпуск сертификатов для классических web-серверов и web-сервисов с уникальным доменным именем, которые представлены в Интернет-пространстве мировой инфраструктуры DNS. При этом право получения сертификата с конкретным Subject Name/Subject Alternative Name имеет только такой клиент, который вла-

деет закрытым ключом (паролем), соответствующим публичной TXT-записи в текущем DNS-пространстве. Из этого сразу вытекают и очевидные ограничения и риски протокола: как лучше передать ключ клиенту и хранить его в безопасности, кто и как контролирует и актуализирует TXT-записи в DNS. Как при этом разделить публичную в сети Интернет и внутреннюю корпоративную инфраструктуру DNS и направить запросы к разным серверам ACME. Как бороться со спуфингом и «отравлением» кеша DNS, особенно в корпоративных сетях (по сути, без внедрения стандарта DNSSEC или его альтернативы Secure Dynamic DNS updates в Microsoft Active Directory защититься от получения злоумышленниками «ваших» сертификатов невозможно). ACME — это такой же протокол, и в нем нет специальных механизмов контроля факта установки и гарантии обновления сертификата информационной системы — потребителя.

Microsoft Enterprise PKI — пожалуй, наиболее функционально полное решение из всех описанных выше, однако оно обладает существенным недостатком: оно ориентировано только на стек продуктов корпорации Microsoft. И хотя существуют простые и понятные механизмы интеграции со сторонними решениями (тот же SCEP и поддержка форматов PKCS#), в таких сценариях существенная часть функционала становится недоступной. Например, без работы потребителя сертификатов на ОС Windows и в инфраструктуре Microsoft Active Directory, а также использования Групповых Политик Active Directory (GPO) у вас не получится распространять информацию о доступных УЦ, о корпоративных шаблонах и сертификатах, которые эти УЦ могут по этим шаблонам выпускать, едином корпоративном списке доверенных сертификатов, обеспечить автоматический перевыпуск устаревающих сертификатов (auto enrollment). Все эти механизмы реализованы поверх семейства сетевых протоколов Microsoft RPC, использование которых в сети Интернет, да и в корпоративных сетях, крайне рискованно. Также Microsoft RPC

плохо работает в сегментированных межсетевыми экранами сетях, поскольку устанавливает несколько разнонаправленных TCP-соединений. Надо отметить, что в рамках Microsoft Enterprise PKI нет и средств мониторинга и контроля как самой ИОК, так и подтверждения успешных фактов обновления списка доверенных сертификатов и собственно результатов выпуска сертификатов. Нет и автоматизации установки обновлённых сертификатов в конечную систему потребителя — это тоже делается чаще всего вручную.

При выборе решения по автоматизации управления технологическими сертификатами следует, кроме прочего, сформулировать общие вопросы ко всем технологиям и продуктам:

Все ли ваши корпоративные потребители сертификатов в корпоративной сети нативно поддерживают ту или иную технологию или протокол?

◆ Насколько сложно обеспечить поддержку протокола или технологии вашими потребителями сертификатов?

◆ Есть ли механизмы контроля результатов выпуска и обновления сертификата на целевой системе потребителей сертификатов?

◆ Есть ли механизмы мониторинга и своевременного оповещения об окончании срока действия сертификата для конкретной системы и её владельца?

◆ Есть ли механизмы балансировки и отказоустойчивости у технологии/протокола для повышения надёжности всего процесса автоматического выпуска и обновления сертификатов?

◆ Поддерживается ли балансировка нагрузки на несколько экземпляров однотипных УЦ?

◆ Есть ли механизмы дополнительной проверки запросов на выпуск сертификата, позволяющие понять, какой клиент запрашивает выпуск и насколько ему допустимо выпускать такой сертификат?

◆ В случае сбоя или отказа решения или технологии автоматического выпуска сертификатов, какие временные меры можно предпринять для обеспечения работоспособности систем потребителей?

◆ Поддерживает ли протокол, технология или продукт другие типы криптографических алгоритмов, в частности криптографические алгоритмы семейства ГОСТ?

При рассмотрении всех этих вопросов выясняются некоторые очень специфические моменты, особенно для современных контейнерных сред на базе **Kubernetes** или брокера сообщений **Apache Kafka**. Перечислять все нюансы обеспечения сертификатами SSL/TLS информационных систем и приложения — занятие долгое и неблагодарное для одной статьи, но давайте хотя бы кратко обозначим наиболее очевидные и заметные проблемы прикладных систем — потребителей сертификатов:

◆ Большинство приложений и систем после выпуска или обновления сертификата требуют выполнения специфических команд и действий, которые позволяют им начать использовать новый сертификат. Такие команды обычно выполняются с использованием данных из нового сертификата и должны быть выполнены полностью автоматически;

◆ Доставка, установка и обновление доверенных корневых сертификатов и полной цепочки сертификатов доверенных выпускающих УЦ в соответствующие хранилища;

◆ Обновление сертификатов в Java-приложениях и поддержание актуального списка доверенных УЦ для JVM — сложность состоит в том, что Java-приложения выполняются в JVM и используют специальное хранилище доверенных сертификатов, автоматическое размещение сертификатов в котором требует дополнительных шагов в сценарии обновления;

◆ Корректное автоматическое обновление сертификатов для распределённых кластеров СУБД, таких как etcd, PostgreSQL или брокеров сообщений Apache Kafka — сложность этой задачи состоит в скоординированном обновлении сертификатов и доверия к ним на разных узлах кластеров в такой последовательности, при которой узлы кластера будут сохранять доверие друг к другу. В противном случае кластер начнёт «терять» узлы и кворум, что приведёт к отказу;

◆ Обновление сертификатов для кластеров Kubernetes требует разработки специализированного компонента — оператора, который реализует специальные шаги и поддерживает специфические API-вызовы для выпуска и обновления сертификатов в Ingress и Egress контроллерах;

◆ Обновление сертификатов envoy-proxy в среде ISTIO Service Mesh на кластерах Kubernetes — особая сложность этой задачи состоит не только в том, чтобы заменить встроенный сервис выпуска сертификатов ISTIO Citadel, но и в том, чтобы обеспечить резервирование и высокую производительность самого процесса выпуска сертификатов. Перевыпуск при каждом рестарте pod-ов продиктован тем, что envoy не сохраняет свой закрытый ключ и сертификат. Производительность же нужна вследствие того, что рестарты pod-ов обычно массовые — перезапускаются целые namespaces и даже кластеры в результате чего за короткое время порядка 10–20 секунд стартуют тысячи pod-ов;

◆ Выпуск сертификатов для системных сервисов и приложений Microsoft, таких как контроллеры домена Active Directory, почтовые службы Exchange, сервера IIS, также требует особых специфических подходов по заполнению расширений EKU и правильному указанию SAN в сертификатах.

4. ДОСТУПНОСТЬ РАЗЛИЧНЫХ ТЕХНИЧЕСКИХ РЕШЕНИЙ В УСЛОВИЯХ САНКЦИЙ И ИМПОРТОЗАМЕЩЕНИЯ

Наверное, этот раздел будет самым коротким в статье — все известные автору имевшиеся продукты и решения по автоматическому управлению технологическими сертификатами в настоящий момент либо подлежат импортозамещению, либо вообще недоступны для использования в России. Из чего следует очевидный вывод: такое решение должно быть создано в России и интегрироваться с российскими ОС, СУБД, системами резервного копирования, виртуализации и ПО с открытым исходным кодом. Критически важно в переход-

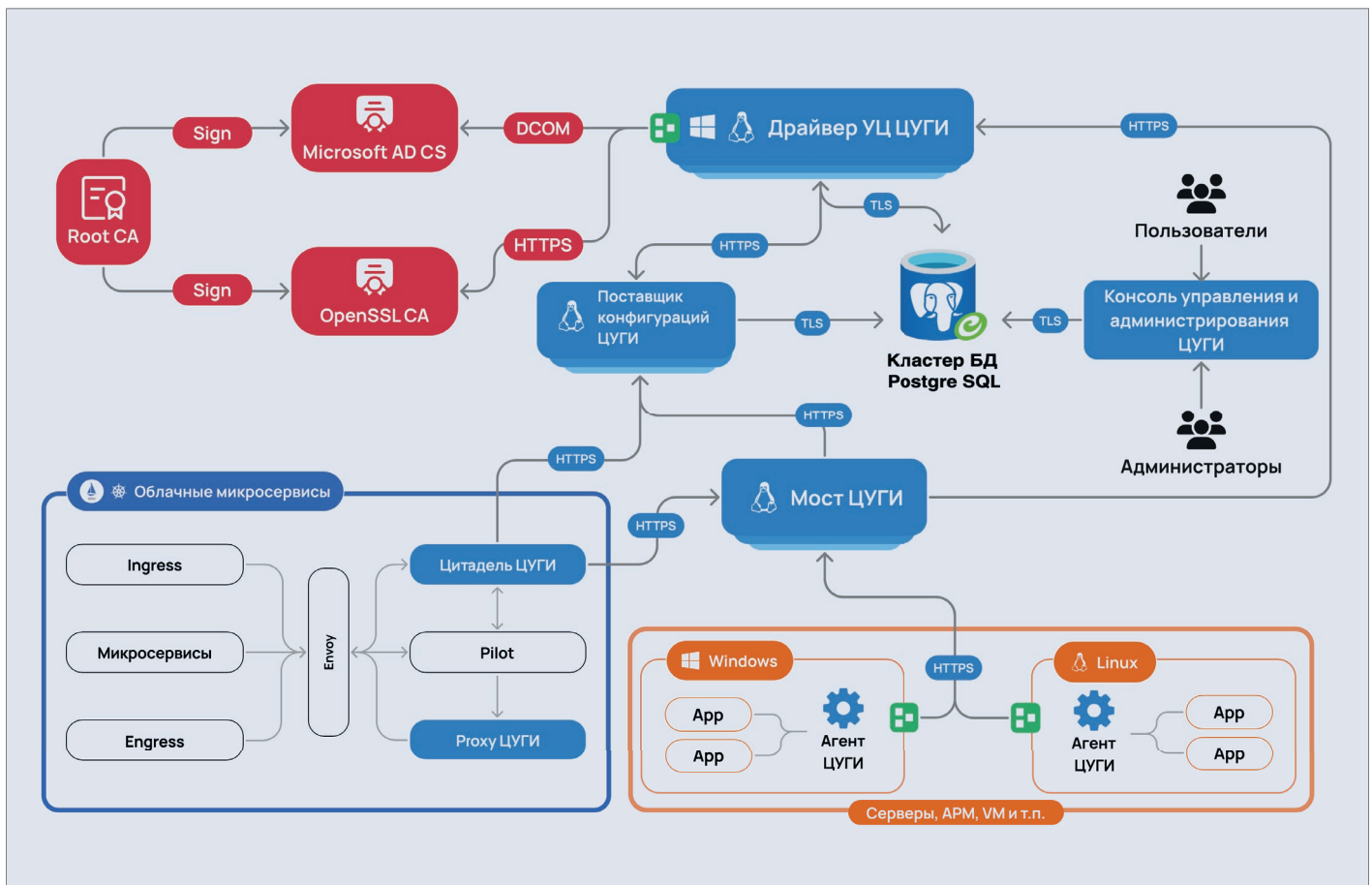


Рисунок 1. Этот компонент обеспечивает автоматизацию выпуска, доставки и установки сертификатов

ный период, пока программные продукты западных вендоров, таких как Microsoft, Cisco, Oracle и прочих, будут сосуществовать в организациях с новыми импортозамещенными продуктами и технологиями, обеспечить работоспособность и совместимость и тех, и других.

5. ОПЫТ ПО КОМПЛЕКСНОМУ ИМПОРТОЗАМЕЩЕНИЮ ИОК

В завершение статьи перейдем к краткому обзору нашего опыта внедрения системы ЦУГИ (Централизованное Управление Гетерогенными Инфраструктурами — в реестре отечественного ПО № 13033), которая в итоге позволяет решить все из перечисленных выше задач технологической ИОК и обеспечить безболезненный переход от инфраструктуры Microsoft Active Directory Certificate Services на технологические УЦ RSA на базе OpenSSL или УЦ российских производителей.

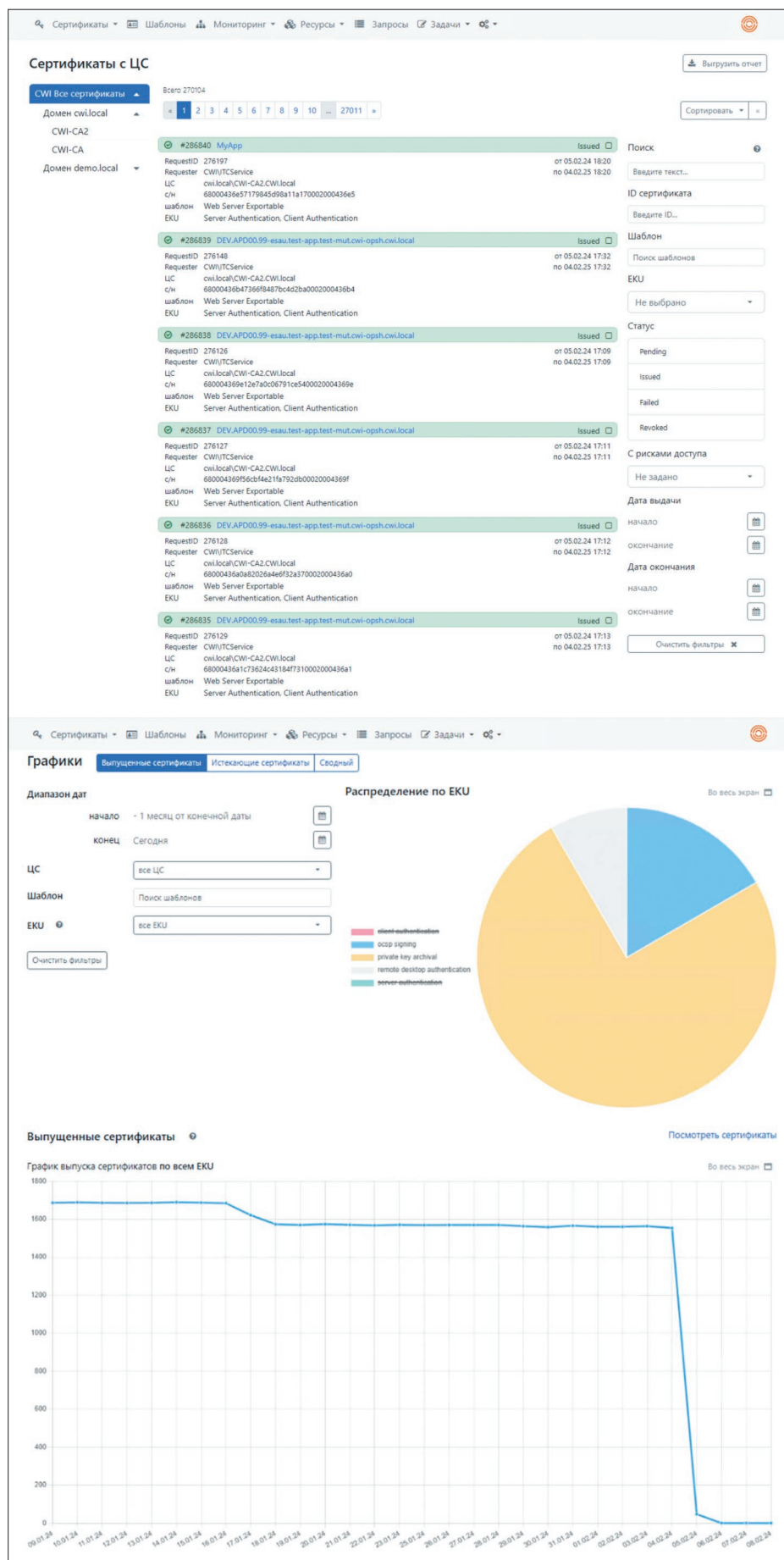
5.1 Импортозамещение серверов УЦ RSA Microsoft CA

В этом вопросе выбрано самое очевидное решение — использование встроенного в большинство ОС семейства Linux пакета OpenSSL который при определенных настройках и программной обвязке позволяет реализовать почти все основные функции технологического УЦ. Такое решение дополнено модулем, поддерживающим протокол SCEP и автоматизирующим процесс формирования и публикации списка отзыва сертификатов (CRL, Delta CRL) и точек доверия (AIA). Как показывают испытания, решение достаточно производительное и при этом не требующее заметных вложений. Однако следует отметить, как будет видно дальше, подобный УЦ с лёгкостью может быть заменён на другие промышленные решения без необходимости какой-либо перенастройки ИОК и информационных систем — потребителей сертификатов.

5.2 Автоматизация выпуска технологических сертификатов для АРМ, серверов и контейнерных сред

Этот компонент — Единая Система Автоматического Управления Сертификатами (ЕСАУС), пожалуй, наиболее важный во всей архитектуре. Именно он обеспечивает автоматизацию выпуска, доставки и установки сертификатов. Кратко представим из чего такая подсистема состоит (см. схему):

- ◆ кроссплатформенный Агент, поддерживающий ОС семейства Windows, Linux, MacOS, IBM AIX и др.;
- ◆ ЦУГИ Citadel, ЦУГИ операторы Kubernetes — компоненты кластеров Kubernetes, поддерживающие процесс выпуска сертификатов для служб и компонентов Kubernetes;
- ◆ мосты ЦУГИ — сервера, обеспечивающие балансировку нагрузки, отказоустойчивость и управление очередями сообщений, а также связность топологии в сегментированных сетях;
- ◆ драйвер УЦ — компонент системы, реализующий набор вендор-не-



зависимых стандартных вызовов для выпуска и отзыва сертификатов при подключении к УЦ различных производителей, а также балансировку нагрузки между экземплярами УЦ;

♦ сервер управления политиками и конфигурациями системы с пользовательской и административной консолью.

Именно агенты и операторы реализуют все те разнообразные сценарии и шаги по формированию запросов на сертификаты (CSR) и следующие после выпуска и доставки сертификата шаги по его корректной установке и конфигурированию приложения-потребителя. Отметим, что агенты при своей первичной установке требуют авторизации, а при запросе сертификата собирают расширенную информацию об окружении и процессах, которые запрашивают сертификат. Именно эта информация и используется потом на Мостах ЦУГИ для проверки запроса на соответствие корпоративным политикам выпуска сертификатов. Только после успешной проверки запрос (CSR) направляется через кластер драйверов УЦ для выпуска на один из множества УЦ. При этом запросы могут как балансироваться, так и для конкретных политик направляться всегда на один УЦ, если этого требуют особенности приложения потребителя. Все происходящие процессы конечно же журналируются, а сама система имеет два независимых плеча, работающих в режиме Active/Active, в географически разнесённых ЦОД.

5.3 Мониторинг Инфраструктуры Открытых Ключей RSA и состояния сертификатов

Эта сопутствующая для ИОК задача для достижения описанных выше целей имеет критически важное значение. Фактически при массовом исполь-

← **Рисунок 2.** Кроме архитектурных решений, необходимо реализовать автоматический структурированный мониторинг как самих УЦ RSA, так и всех вспомогательных компонентов ИОК. Для этой цели реализована специализированная система мониторинга, в которую встроена индивидуальная «модель здоровья» ИОК

зовании сертификатов и выпуске их десятками тысяч в день нам необходимо обеспечить очень высокую доступность и надёжность всех компонентов системы. Кроме архитектурных решений, необходимо реализовать автоматический структурированный мониторинг как самих УЦ RSA, так и всех вспомогательных компонентов ИОК. Для этой цели реализована специализированная Система Мониторинга ИОК (СМИОК), в которую встроена индивидуальная «модель здоровья» ИОК, включающая в себя (рис. 2):

- ◆ проверку состояния процессов УЦ в ОС сервера;
- ◆ снятие счётчиков производительности УЦ;
- ◆ анализ журналов и логов УЦ;
- ◆ проверку доступности и содержания точек распространения CDP/AIA;
- ◆ анализ состояния и работоспособности HSM (при наличии);
- ◆ постоянный пробный выпуск тестового сертификата;
- ◆ анализ шаблонов сертификатов в Microsoft Active Directory;
- ◆ анализ выпущенных сертификатов на типовые риски и требования безопасности;
- ◆ поиск и фильтрацию единой БД всех выпущенных в организации сертификатов.

Также, помимо классических задач мониторинга, система обеспечивает дополнительные функции мониторинга и сопровождения ИОК, а именно (рис. 3):

- ◆ оповещения при отзыве или достижении предельных сроков жизни сертификатов;
- ◆ мониторинг срока действия выпущенных на внешних УЦ сертификатов;
- ◆ мониторинг списка отзыва для сертификатов, полученных на внешних УЦ;
- ◆ централизованное снятие резервных копий конфигурации и БД УЦ;
- ◆ сканирование сетевых сегментов для поиска портов, поддерживающих SSL/TLS-подключение и фиксацию фактически установленных сертификатов;
- ◆ учёт и ведение разовых и регулярных задач дежурной смены ИОК;

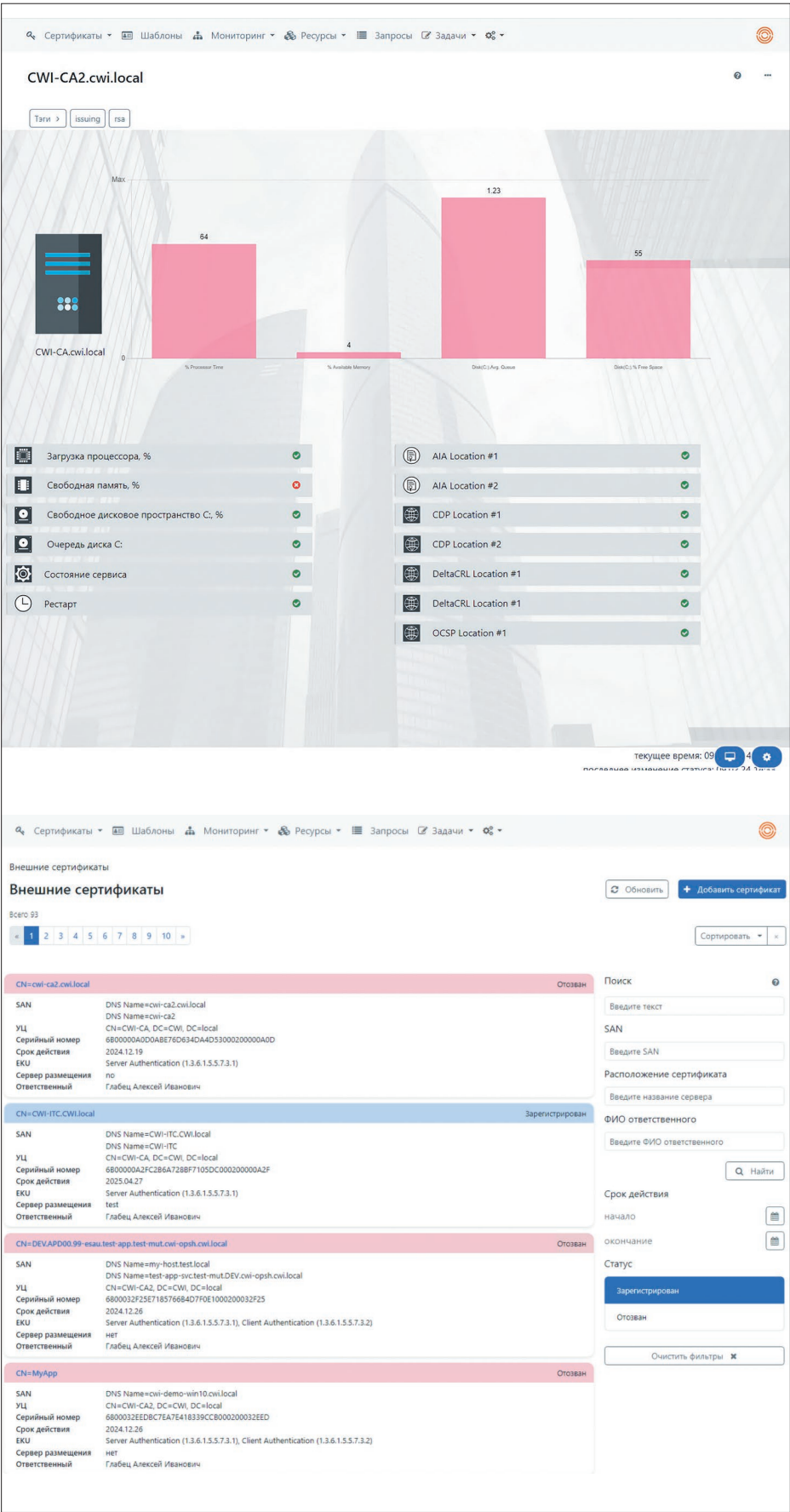


Рисунок 3. Помимо классических задач мониторинга, система обеспечивает дополнительные функции мониторинга и сопровождения ИОК

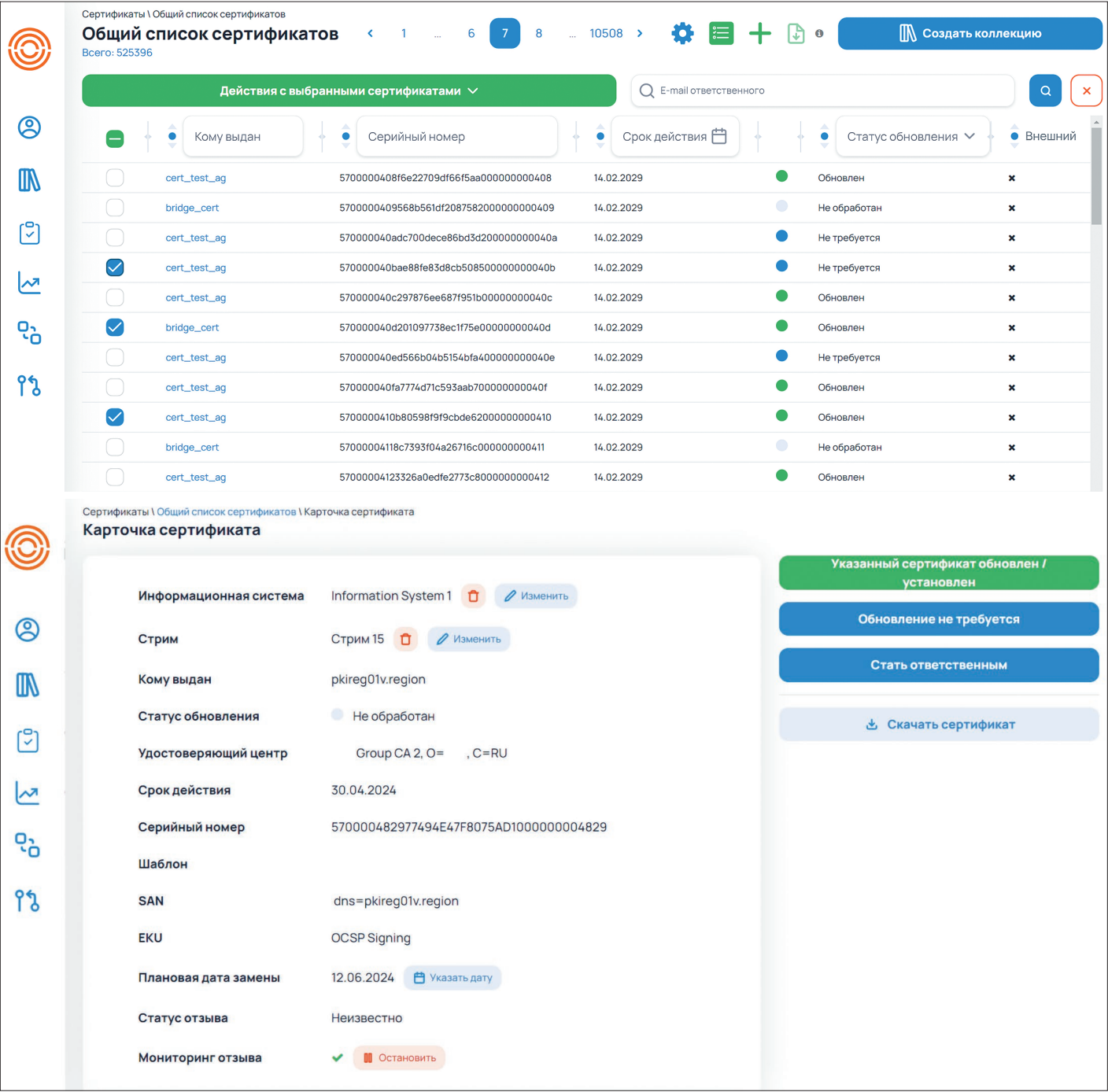


Рисунок 4. Для пользователей и администраторов информационных систем, которые являются потребителями технологических сертификатов, личный кабинет — это единое окно для решения всех основных задач

♦ инвентаризация версий установленных криптобиблиотек и провайдеров.

5.4 Самообслуживание потребителей сертификатов через Личный Кабинет

Этот компонент архитектуры Личный Кабинет Пользователя Сертификатами (ЛКПС) предоставляет интерфейс к Единому Справочнику Сертификатов (ЕСС), используемых в организации,

и отвечает за работу пользователей и администраторов информационных систем, которые являются потребителями технологических сертификатов. Для таких пользователей Личный Кабинет является единым окном для решения всех своих основных задач:

- ♦ указание контактов и способов оповещения пользователя;
- ♦ поиск и сортировка по единому справочнику всех сертификатов организации;

- ♦ создание коллекций сертификатов для удобства управления сертификатами;
- ♦ назначение себя или своих подчинённых ответственными за конкретные сертификаты;
- ♦ просмотр карточки сертификата, где все поля разобраны и проведён анализ рисков;
- ♦ просмотр в карточке сертификата расширенной информации о сертификате, его статусе и местах ис-

пользования, поступившей из других систем: системы мониторинга, системы автоматического выпуска, учёта СКЗИ, CMDB, Active Directory и т.д.;

- ♦ ручной запрос на отзыв или перевыпуск своего сертификата;
- ♦ отметки о статусе выполнения задач установки и обновления сертификатов в своих прикладных системах;
- ♦ связывание сертификатов с данными корпоративной CMDB и кодами информационных систем (место установки/применения сертификата).

5.5 Интеграция технических решений в бизнес-процессы

Рассмотрим совместимость архитектуры ЦУГИ с наиболее актуальными отечественными решениями. В первую очередь опишем возможности интеграции с УЦ, выпускающими RSA сертификаты российского производства:

- ♦ КриптоПро Кластер УЦ – интеграция реализована для выпуска RSA сертификатов.
 - ♦ SafeTech CA – драйвер на финальной стадии тестирования, релиз запланирован в 3 квартале 2024 года.
 - ♦ Aladdin eCA – драйвер в разработке, релиз ожидается в 2024 году.
- Кратко обозначим и другие средства интеграции:
- ♦ Универсальное API для работы с внешними системами – в завершающей стадии разработки.
 - ♦ Системы мониторинга на базе Prometheus – реализованы разнообразные метрики для мониторинга.
 - ♦ Ряд распространённых корпоративных бизнес-приложений – интеграция в разработке.

6. ВМЕСТО ЗАКЛЮЧЕНИЯ

Описанные в статье подходы и программные продукты не просто теория, а зарекомендовавшие себя в реальной эксплуатации (некоторые компоненты успешно работают уже более 2 лет) системы и сервисы построенные на платформе ЦУГИ производства ООО «Клируэй Текнолоджис» (торговая марка Clearway Integration).

Одним из успешно использующих их заказчиков является Банк ВТБ. В частности, на момент написания

статьи в Банке ВТБ из намеченных к концу 2024 года 700 информационных систем и приложений уже подключены и обсуживаются в режиме полностью автоматического выпуска сертификатов более 150. При этом разработана методическая и техническая документация, необходимая для ускорения и упрощения процесса подключения информационных систем потребителей сертификатов к ЦУГИ.

Планы развития ЦУГИ на ближайшую перспективу включают (рис. 5):

- ♦ Создание функционала для более гибкого управления расписанием обновлений и технологическими окнами обновления сертификатов;
- ♦ Разработку мобильного приложения Личного Кабинета Пользователя для управления Сертификатами;
- ♦ Проектирование и реализация модуля ЦУГИ для учёта СКЗИ и/или

интеграция с существующими рыночными решениями;

- ♦ Создание и развитие публичной библиотеки сценариев выпуска сертификатов для конкретных приложений и систем;
 - ♦ Разработку новых драйверов УЦ для УЦ различных производителей, ещё не интегрированных с системой;
 - ♦ Реализацию протоколов CMP, EST и ACMEv2 (SCEP уже реализован) в Агенте и в Драйвере УЦ для расширения сценариев интеграции;
 - ♦ Интеграцию с доверенными хранилищами секретов (аналоги HashiCorp Vault) для расширения сценариев обращения с закрытыми ключами, включая безопасный Key Recovery;
- Сертификацию отдельных модулей и решений на базе ПО ЦУГИ во ФСТЭК и ФСБ.

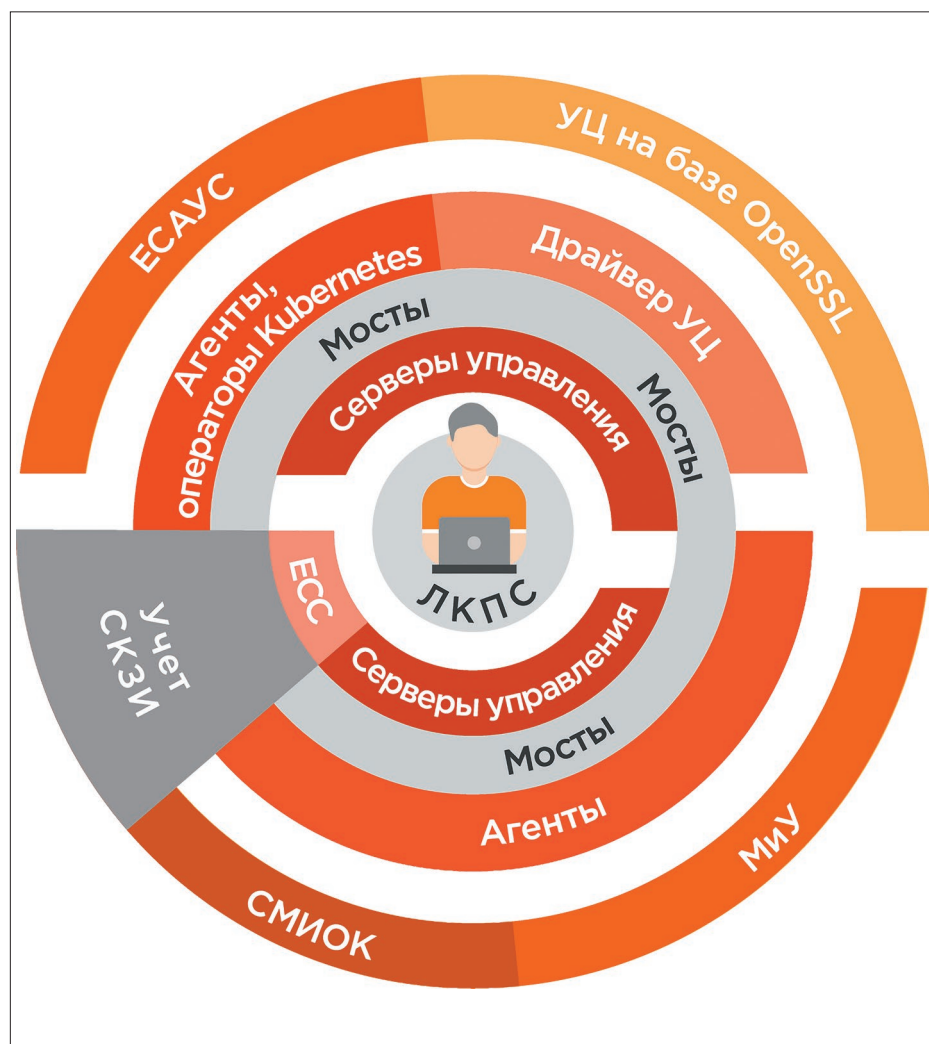


Рисунок 5. Планы развития ЦУГИ на ближайшую перспективу

НАШ ОТВЕТ MICROSOFT

SAFETECH ЗАПУСТИЛ ОТЕЧЕСТВЕННЫЙ СА,
ПРИЗВАННЫЙ ЗАМЕНИТЬ ИНОСТРАННОЕ РЕШЕНИЕ



Денис КАЛЕМБЕРГ
генеральный директор SafeTech



Александр САНИН
генеральный директор SafeTech Lab

В данной статье мы расскажем о комплексном решении SafeTech CA, разработанном компанией SafeTech для выпуска сертификатов стандарта RSA (в том числе технологических), которое обладает не только всеми свойствами Microsoft CA, но и широким дополнительным функционалом.

НА 100% РОССИЙСКИЙ ПРОДУКТ

Microsoft Certificate Authority (Microsoft CA, центр сертификации Microsoft) для компаний всего мира де-факто является «сервисом по умолчанию», так как он бесплатен и встроен в Windows Server. Российские организации не стали исключением, и долгие годы Microsoft CA для отечественного бизнеса был одним из ключевых компонентов в работе с инфраструктурой открытых ключей. С его помощью выпускали, приостанавливали и отзывали сертификаты, используемые внутри компании для различных технологических нужд. Все привыкли, что Microsoft CA «просто есть» и что он решает свои задачи. При этом к удобству и интерфейсам сервиса вопросов всегда было много, но, когда продукт бесплатен и поставляется в комплекте с серверной ОС, их обычно не задают.

Однако после ухода Microsoft из России и появления рисков ограничения доступности его сервисов у организаций, заботящихся о непрерывности функционирования

ИТ-инфраструктуры, появилась необходимость в импортонезависимом решении. Как минимум для того, чтобы иметь что-то в резерве на случай негативного сценария в виде внезапной недоступности Microsoft CA. Для компаний, относящихся к субъектам критической инфраструктуры (КИИ), отечественный СА не только должна осмотрительность, но и требование законодательства с четко установленными сроками. А именно: с 2025 года субъекты КИИ должны перейти с иностранного софта на отечественный. Такое требование содержится в Указе Президента России.

На сегодняшний день на российском рынке есть несколько отечественных СА. Одно из решений — наше, то есть разработанное компанией SafeTech, резидентом ИТ-кластера фонда «Сколково». SafeTech CA — не адаптация open-сурсного решения, а полностью собственная разработка. Это на 100 % российский продукт, выпущенный компанией, которая более 13 лет создаёт решения в области информационной безопасности.

НЕ КОНКУРЕНТ

Отметим, мы не позиционируем SafeTech CA как конкурента, например, удостоверяющим центром от «КриптоПро» или «ИнфоТекСа». Наша цель — именно замена сервиса выпуска технологических сертификатов, используемых повсеместно. А это совершенно иной рынок и иная целевая аудитория. Нашим решением

даже может больше заинтересовать эксплуатационное подразделение ИТ, чем департамент информационной безопасности. Безусловно, оба понимают, о чём речь, но тут есть явная и чёткая грань: наше решение обслуживает технологические сертификаты, а в крупных компаниях для тех или иных нужд таких сертификатов могут быть десятки или даже сотни тысяч, и все они выпускаются по простым и понятным шаблонам и протоколам, и это весьма далеко от ГОСТ криптографии и процессов, принятых там.

ДРУГИМ ПУТЁМ

Порой ИТ-компании, выпускающие своё решение вместо продукта ушедшего вендора, стремятся достичь максимальной идентичности как по интерфейсу, так и по функционалу. Мы пошли другим путём. Microsoft CA, хотя и является привычным и бесплатным, но весьма скуден по набору функций. На наш взгляд, в 2024 году просто странно иметь административный интерфейс, который не менялся уже четверть века. При создании нового продукта копировать морально и технически устаревшие особенности привычного Microsoft CA не было смысла, и потому мы в своём решении сделали упор на гораздо большую функциональность, современные интерфейсы, гибкость, масштабируемость и кросс-платформенность!

Именно поэтому мы в рамках разработки взяли кроссплатформен-

ный стек — Java. С одной стороны, это позволило решению разворачиваться и работать как в Microsoft, так и в Linux-инфраструктурах, а с другой стороны — обеспечить очень высокую производительность. Первые проведённые тесты и замеры продемонстрировали скорость выпуска сертификатов на уровне 6–8 тысяч в минуту в базовой конфигурации серверов. Этого показателя хватит с большим запасом подавляющему числу клиентов. Решение поддерживает развёртывание в любом формате:

- ◆ классическая виртуализация (VmWare, KVM, HyperV и т.д.);
- ◆ контейнерная виртуализация (Docker, OpenShift/K8S);
- ◆ установка на физические сервера.

SafeTech CA способно закрыть все базовые задачи, которые решал Microsoft CA. В текущей версии 1.0 мы реализовали протокол Microsoft WS-Trust X.509v3 Token Enrollment Extensions (MS-WSTEP). Кроме того, решение поддерживает интеграцию с MS Active Directory. В совокупности это делает возможной замену Microsoft CA и обеспечение auto enrollment-сервиса на полностью отечественном ПО.

Кроме того, миграция на SafeTech CA с Microsoft CA — бесшовная: SafeTech CA можно развернуть в существующей инфраструктуре параллельно с Microsoft CA и начать выпускать на нём новые сертификаты какого-то определённого типа. Затем постепенно расширять типы выпускаемых сертификатов. И через какое-то время вывести сервис Microsoft CA из эксплуатации.

АРХИТЕКТУРА

Архитектура SafeTech CA строится на микросервисной модели, что, с одной стороны, обеспечивает лёгкую масштабируемость и отказоустойчивость, а с другой — позволяет довольно гибко управлять роадмэпом (дорожной картой развития) продукта. С момента релиза SafeTech CA прошло меньше 2 месяцев, и всё это время мы находились в непрерывном общении как с потенциальными заказчиками, так и с рядом интеграторов. Собрав первые отзывы, мы существенно изменили сроки

выпуска функциональных модулей, передвинув в начало наиболее востребованные и убрав в конец экзотические. В настоящее время в продукте, помимо ядра, включающего в себя основные сервисы (CA core, Gateway, Discovery service, CRL/AIA services, OCSP), а также стандартного интеграционного REST API, мы реализовали модуль MS Enrollment, который обеспечивает поддержку управления сертификатами для пользователей MS Active Directory и компьютеров MS Windows.

Следующий на очереди модуль — SCEP. Он обеспечивает поддержку управления сертификатами для сетевых устройств и ПО (Cisco, MS Intune и т.д.). Пожелание по скорейшей реализации данного протокола мы услышали от 100 % компаний, с которыми общались. Мы считаем хорошей практикой, когда потенциальные заказчики и эксперты в предметной области могут влиять на роадмэп продукта, что в нашем случае и произошло. Это позволит решению развиваться быстро, эффективно и в правильном направлении.

В дальнейшем планируем реализовать модули ACME (обеспечивают поддержку открытого интернет-стандарта ACME (Kubernetes, Openshift, Let's Encrypt) и CMP (обеспечивает поддержку протокола CMP — базового стандарта для API OpenSSL, Bouncy Castle, Nexus и т.д.). И последним — модуль EST (обеспечивает поддержку управления сертификатами для MDM-систем и мобильных устройств).

НАВСТРЕЧУ ПАРТНЁРУ

Кроме того, мы уже работаем над специализированной версией SafeTech CA, использующей сертифицированное криптографическое ядро от нашего партнёра — компании «КриптоПро». Это позволит, с одной стороны, заместить импортные HSM на хорошо зарекомендовавший себя «КриптоПро HSM», не меняя при этом используемые криптографические алгоритмы. С другой стороны, с запуском этой версии ГОСТ криптоалгоритмы в решении будут поддерживаться автоматически. На сегодняшний день мы сосредоточились

на RSA. Но важно отметить, что для SafeTech CA не имеет значения, какой криптографией пользоваться. Это всего лишь вопрос подключения/отключения криптоядра, на остальную функциональность продукта выбор криптографии не влияет. Но первая версия SafeTech CA была запущена именно на RSA-криптографии по той лишь причине, что инфраструктура в широком смысле «не умеет в ГОСТ», как бы нам ни хотелось все выпускаемые сертификаты делать по ГОСТу...

Все нововведения в функциональности автоматически будут доступны и в специальной версии SafeTech CA с сервисами от «КриптоПро». Это опять же о преимуществах микросервисной модели, которая позволяет нам не вести две независимые ветки разработки, а делать это в одном едином ключе.

МЫ ВЫБРАЛИ ИМПОРТОНЕЗАВИСИМОСТЬ

Вместо заключения хотел бы ещё раз подчеркнуть: мы не пошли по пути традиционного импортозамещения и не стали делать свой продукт как кальку с Microsoft CA. Мы выбрали импортонезависимость, сделав наше решение более удобным и функциональным.

И потому SafeTech CA обладает всеми атрибутами современного, хорошего продукта:

- ◆ адекватным, современным и удобным UI (интерфейсом) администрирования;
- ◆ максимально широким по функциональности интеграционным REST API;
- ◆ возможностями интеграции с системами сбора и корреляции событий информационной безопасности (SIEM);
- ◆ управлением доступом в рамках современного OAuth OIDC с возможностью использования внешних систем безопасного доступа и каталогов субъектов.

Кроме того, ключевым преимуществом SafeTech CA является поддержка всех распространённых протоколов: MS Enrollment, SCEP, ACME, CMP, EST, HSM, OCSP — для решения любых задач в рамках современного центра сертификации.

КОРПОРАТИВНЫЕ ЦЕНТРЫ СЕРТИФИКАЦИИ И РОССИЙСКАЯ КРИПТОГРАФИЯ



**Станислав
СМЫШЛЯЕВ**
заместитель
генерального
директора
компании
«КриптоПро»,
д. ф.-м. н.

У ряда иностранных решений в области информационных технологий до 2022 года отсутствовали отечественные аналоги, и вопрос их импортозамещения встал особенно остро после ухода соответствующих компаний с российского рынка. В области информационной безопасности это относится, в первую очередь, к тем типам решений, для которых не требуются сертификаты соответствия требованиям ФСБ России. При этом зачастую области применения решений, требующих сертификации, и решений, которые ее не требуют, крайне близки. В таких случаях импортозамещение за короткий срок становится решаемой задачей: возможно частичное переиспользование существующих наработок.

Именно так получилось с платежными HSM после того, как летом 2022 года французский производитель Thales заявил об уходе из России. HSM-модули этого производителя стояли в подавляющем большинстве банков, работающих с пластиковыми картами. При этом возникшая проблема была быстро решена, поскольку разработка отечественных платежных HSM началась раньше, с 2016 года, а в июне 2023 года она

была завершена: был сертифицирован программно-аппаратный криптографический модуль КриптоПро HSM версии 2.0 R3 с платежным функционалом, позволявшим полностью заменить зарубежные аналоги. Создание этого решения стало возможным, в частности, благодаря многолетнему опыту разработки HSM для защиты ключей электронной подписи, ключей удостоверяющих центров, обеспечения криптографической защиты электронных сообщений — то есть для решения тех задач, где необходимо наличие сертификата соответствия требованиям ФСБ России.

Одна из важнейших задач в настоящее время — создание отечественного аналога корпоративного центра сертификации Microsoft (Microsoft CA), который мог бы обеспечивать, в частности, совместимость с ним по интерфейсам. Отсутствие ранее созданных российских аналогов данного продукта связано с тем, что применение подобных корпоративных центров сертификации зачастую осуществляется в случаях, не требующих соответствия решений требованиям ФСБ России и потребности в российских продуктах данного типа ранее не было.

Задача по созданию российского аналога сейчас решается, в частности, коллегами из компании SafeTech в рамках развития их продукта SafeTech CA. При этом, с учетом специфики российского рынка, для ряда корпоративных клиентов важна возможность работы продукта на базе сертифицированных криптографических ядер и российских криптографических алгоритмов, в том числе, с возможностью хранения ключей в защищенных программно-аппаратных модулях. Решить данную задачу

возможно с помощью существующих сертифицированных программных и программно-аппаратных СКЗИ, таких как КриптоПро CSP версии 5.0 R3 и КриптоПро HSM версии 2.0 R3.

Безопасность любой информационной системы, доверие в которой основывается на инфраструктуре ключей электронной подписи и сертификатов открытых ключей, критически зависит от уровня защищенности корневых ключей. В корпоративных центрах сертификации, в случаях, не требующих применения сертифицированных средств УЦ (таких как наш продукт КриптоПро УЦ версии 2.0), ключи в любом случае должны быть защищены должным образом. Этого можно достичь с применением программно-аппаратных криптографических модулей, а в тех случаях, когда требуется сугубо программная реализация, с помощью сертифицированных программных СКЗИ, обеспечивающих максимально возможный уровень безопасности в условиях ограниченного использования аппаратных компонент. Поддержка таких решений производства нашей компании, имеющих сертификаты соответствия требованиям ФСБ России по классам КС1-КС3 и КВ/КВ2, которая реализуется в корпоративных центрах сертификации, позволяет существенно повысить безопасность их применения у конечных заказчиков. Безопасная реализация криптографического ядра, постоянное их совершенствование с учетом обновлений операционных систем и обнаруживаемых в них уязвимостей, высочайшая производительность как программных, так и программно-аппаратных средств — всё это даёт возможность быть уверенным в успешном решении данной задачи.

СТРОГАЯ АУТЕНТИФИКАЦИЯ — КЛЮЧ К ДОВЕРЕННОЙ КОРПОРАТИВНОЙ ИНФРАСТРУКТУРЕ



Сергей ХАЛЯПИН
директор
департамента
внедрения
и пресейла
компании
«Аладдин»

Если провести аналогию ИТ-инфраструктуры компании с домом, для создания безопасного пространства необходимо заменить замки и ключи, чтобы быть уверенным, что доступ в помещение разрешён только его жителям. В организациях, особенно в тех, которые относятся к КИИ или обрабатывают информацию с высоким уровнем значимости, необходимо развернуть целую систему управления «ключами» и их носителями.

Решить задачи для создания доверенного пространства внутри и одновременно защиты от внешних угроз позволяет функция строгой аутентификации на основе инфраструктуры открытых ключей (PKI), которая ранее строилась с использованием продуктов компании Microsoft. Сейчас необходимо обеспечить бесшовный переход с PKI, построенной на Windows, на отечественные Linux-продукты.

Ключевым элементом системы является корпоративный центр сертификации. Его основными задачами являются создание и функционирование PKI в рамках компании, обслуживание всех объектов и компонентов ИТ-систем ключами (цифровыми сертификатами). Aladdin Enterprise CA обеспечивает

совместимость с Microsoft CA, позволяет плавно мигрировать в новую ИТ-инфраструктуру на базе российских ОС. В этом процессе важную роль играет совместная работа с технологическими партнёрами:

- ◆ производителями ОС и служб каталогов (ГК «Астра», «РЕД СОФТ», «Базальт СПО»);
- ◆ разработчиками СКЗИ («КриптоПро»);
- ◆ партнёром в области систем управления и мониторинга технологическими сертификатами и ИОК (Clearway Integration).

Важным требованием для построения строгой аутентификации является наличие нескольких факторов (знание, владение и/или биометрия) и использование инфраструктуры открытых ключей.

Выполнение этого требования обеспечивается линейкой токенов и смарт-карт JaCarta.

Для организаций, которым нужна юридическая неотказуемость действий пользователя, рекомендуется использовать биотокен JaCarta. Этот продукт устраняет вероятность входа в систему злоумышленника с чужим токеном. На устройстве хранится отпечаток пальца конкретного пользователя, только он может активировать bio-токен для дальнейших операций.

Одновременно с импортозамещением инфраструктурных решений, таких как служба каталогов или центр сертификации, происходит перевод рабочих мест пользователей на ОС Linux. Пользователи, привыкшие, что в ОС Windows многофакторная аутентификация и PKI в рамках компании поддерживаются из «коробки», сталкиваются с тем, что в ОС Linux такая под-

держка отсутствует. Для реализации функционала Windows Smart Card Logon компания «Аладдин» предлагает продукт Aladdin SecurLogon для полноценной поддержки PKI и строгой многофакторной аутентификации. Организации, ещё не развернувшие у себя PKI, с помощью Aladdin SecurLogon могут обеспечить усиленную аутентификацию как на отдельно стоящих рабочих местах, так и в одноранговых сетях.

В средних и крупных организациях процесс управления жизненным циклом сертификатов и их носителей должен быть автоматизирован. Компания «Аладдин» предлагает для решения этой задачи JaCarta Management Server (далее — JMS). С его помощью администратор обеспечивает учёт и управление токенами, смарт-картами, программными токенами, сертификатами и другими объектами PKI. JMS предоставляет пользователям возможность самообслуживания через выделенный портал, а специалисты по информационной безопасности получают инструмент автоматизации применения политик ИБ и учёта выдачи токенов и сертификатов с формированием необходимой отчётности.

Все продукты, перечисленные в этой статье, зарегистрированы в российском реестре программного обеспечения Минцифры РФ, имеют необходимые сертификаты ФСТЭК или находятся в стадии сертификации. Продукты «Аладдин», собранные в единую инфраструктуру, обеспечивают безопасную доверенную среду, в основе которой лежит строгая аутентификация пользователей, оборудования и сервисов.

СТАНДАРТ БЕЗОПАСНОСТИ ФАПИ.СЕК

ДЕТАЛИ И ГЛАВНОЕ



Юрий КАРДЮКОВ
руководитель
направления Identity
Platform и Open API
eKassir

Рынок финансовых технологий движется в сторону поэтапного перехода к свободному информационному обмену на основе Открытых API. В конце 2023 года Ассоциация ФинТех, оператор среды открытого банкинга, опубликовала исследование¹ ряда реальных бизнес-кейсов, в которых были применены открытые программные интерфейсы. Двусторонняя передача данных между банками, оформление цифровой ипотеки, быстрое согласование медицинских услуг — все эти задачи успешно решаются «в один клик» при помощи Открытых API.

Подключаясь к доверенной среде Открытых API, участники финансового рынка внедряют в свои бизнес-процессы все преимущества информационного обмена нового вида. Например, банки могут делегировать нагрузку по разработке цифровых клиентских услуг финтехам, которые, получая доступ к данным, создают эффективные решения и удовлетворяют спрос пользователей на удобные приложения и сервисы.

Регулировать новую форму информационного обмена будет Банк России. Сейчас финансовый регулятор разрабатывает единые правила передачи данных через Открытые API, определяет модели, формат и структуру финансовых сообщений, а также уста-

навливает стандарты и способы обеспечения безопасности информационного обмена. Рассмотрению важнейшего из них посвящена эта статья.

ЗНАКОМЬТЕСЬ, ФАПИ.СЕК!

Концепция Открытых API опирается на применение двух компонентов:

♦ **Прикладные стандарты** — такие как стандарты «Инициирования перевода денежных средств клиента»² и «Получения информации о счёте клиента»³. С последующим внедрением технологии открытого информационного обмена на российский рынок количество таких стандартов будет возрастать.

♦ **Протоколы безопасности финансовых операций** — **ФАПИ.ПАОК** и **ФАПИ.СЕК**.

Летом 2021 года Банк России утвердил и ввёл в оборот один из ключевых стандартов Открытых API — СТО БР ФАПИ.СЕК-1.6-2020⁴ «Безопасность финансовых (банковских) операций. Прикладные программные интерфейсы обеспечения безопасности финансовых сервисов на основе протокола OpenID» (далее — ФАПИ.СЕК).

Стандарт ФАПИ.СЕК описывает обязательные требования по использованию криптографических механизмов и взаимной аутентификации на уровне транспортного слоя (mTLS). Он задаёт алгоритмы подписи и шифрования сообщений в процессе информационного обмена и унифицирует требования для систем аутентификации клиентов в финансовых приложениях и для подтверждения операций.

Прикладные стандарты определяют бизнес-составляющую Открытых API, и каждый из них должен учитывать реализацию всех положений протоколов безопасности.

² <https://www.cbr.ru/StaticHtml/File/59420/standard 3.pdf>

³ <https://www.cbr.ru/StaticHtml/File/59420/standard 2.pdf>

⁴ <https://www.cbr.ru/StaticHtml/File/59420/Standart 1.6-2020.pdf>

¹ <https://www.fintechru.org/api/download/?id=6149&fid=3268>

В отличие от СТО БР ФАПИ.ПАОК-1.0.2021⁵, который ещё не реализован участниками и не представлен на сертификационном стенде Ассоциации ФинТех, ФАПИ.СЕК получил распространение в России и уже применяется в пилотных проектах по открытому банкингу. Разобраться с механизмами организации безопасного обмена данными, регламентированными этим стандартом, может его ретроспектива.

ЯВНОЕ СОГЛАСИЕ ПОЛЬЗОВАТЕЛЯ

От других моделей информационного обмена Открытые API отличает обязательность явного согласия пользователя на передачу и хранение данных. Без осведомлённого участия владельца бизнес-информации запустить обмен невозможно. Согласие бывает двух видов.

♦ **Долгосрочное согласие** применяется в Открытых API, которые не изменяют состояние ресурсов пользователя. Например, оно используется для получения доступа к счёту пользователя. Долгосрочное согласие позволяет поставщику финансовых сервисов получать у банка информацию от имени пользователя на длительный срок. Благодаря этому финтех может работать с данными владельца, даже когда того нет в сети.

♦ **Краткосрочное согласие** используется только один раз. Согласие такого рода невозможно вызвать повторно, и действует оно в течение ограниченного времени. Применяется в Открытых API, которые вносят изменения в состояние ресурсов пользователя, — например, для проведения платежа.

Рассмотрим ситуацию, когда существует некий кешбэк-сервис, предоставляющий клиенту финансовое приложение для возврата денежных средств. Сам сервис выступает в роли потребителя Открытых API — а как в этом случае проявляются обязательства, накладываемые на участников информационного обмена стандартом ФАПИ.СЕК?

Пользователь скачивает приложение, регистрируется в нём и выдаёт от своего имени (как клиент банка) долгосрочное согласие кешбэк-сервису, которое тот использует для доступа к клиентскому счёту. Теперь сервис может взаимодействовать с транзакциями по счёту на протяжении всего периода действия явного согласия. При этом информация передаётся напрямую, от банка к сервису, что значительно снижает возможность её компрометации или перехвата злоумышленниками.

⁵ Стандарт обеспечения безопасности финансовых сервисов при инициации OpenID Connect клиентом потока аутентификации по отдельному каналу.

ИНФОРМАЦИОННЫЙ ПОТОК ФАПИ.СЕК

Стандарт СТО БР ФАПИ.СЕК-1.6-2020 разработан на основе технологии защищённой авторизации OAuth 2.0 и профилирующего её протокола OpenID Connect. OAuth 2.0⁶ — протокол авторизации, который позволяет одному сервису (приложению) получить доступ к данным другого приложения (сервера ресурсов). В этом случае сервис может авторизоваться от имени владельца ресурса, но при этом аутентичные учётные данные ему не передаются. OpenID Connect (OIDC)⁷ — расширение протокола OAuth 2.0, отвечающее за аутентификацию владельца ресурса и передачу стороннему сервису точной информации о нём.

СТО БР ФАПИ.СЕК-1.6-2020 не новый стандарт, он базируется на OAuth 2.0, OIDC и Financial-Grade API. Чтобы проследить «эволюционный» путь стандарта Банка России, необходимо рассмотреть преимущества и недостатки его предшественников (рис. 1).

Для этого вернёмся к примеру с приложением кешбэк-сервиса и опишем пути передачи информации, обеспечивающие взаимодействие в рамках стандарта ФАПИ.СЕК:

1. Пользователь регистрируется в приложении, затем изъявляет желание передать кешбэк-сервису доступ к своим банковским счётам и для этой цели выбирает нужный.

2. Чтобы подтвердить подлинность обращения, кешбэк-сервис передаёт аутентификационный запрос на сервер авторизации, после чего перенаправляет пользователя в выбранный им банк.

3. Пользователь видит знакомый экран входа в банковское приложение и авторизуется по своим реквизитам — логину, паролю и/или второму фактору аутентификации, доступа к которым кешбэк-сервис не имеет. Далее пользователь получает запрос на авторизацию согласия кешбэк-сервиса для временного доступа к счёту и, если всё в порядке, авторизует его.

4. Сервер авторизации выпускает и передаёт токен доступа кешбэк-сервису, который, вызвав прикладные Открытые API, при каждом следующем обращении этот токен предъявляет.

ПРОТОКОЛ АВТОРИЗАЦИИ OAUTH 2.0

Приведённому выше примеру с кешбэк-сервисом подходит аналогия с подписантом договора, который оформляет юридическую

⁶ <https://datatracker.ietf.org/doc/html/rfc6749>

⁷ https://openid.net/specs/openid-connect-core-1_0.html

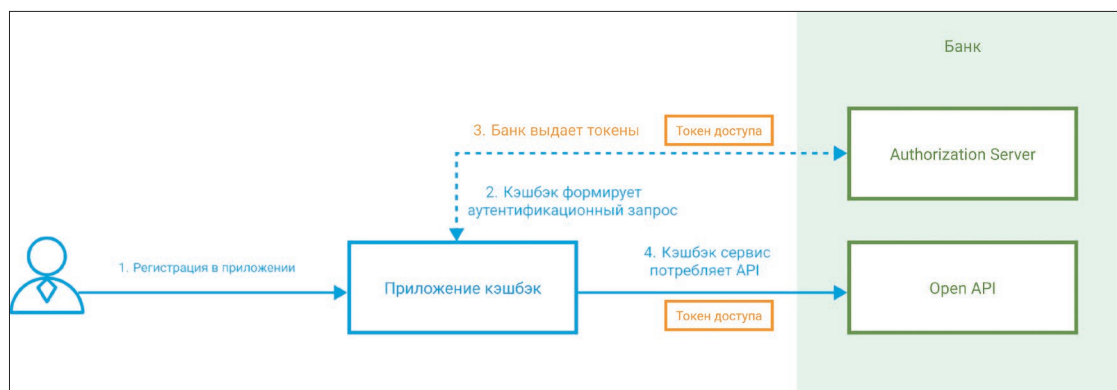


Рисунок 1. Информационный поток стандарта ФАПИ.ОСЕК

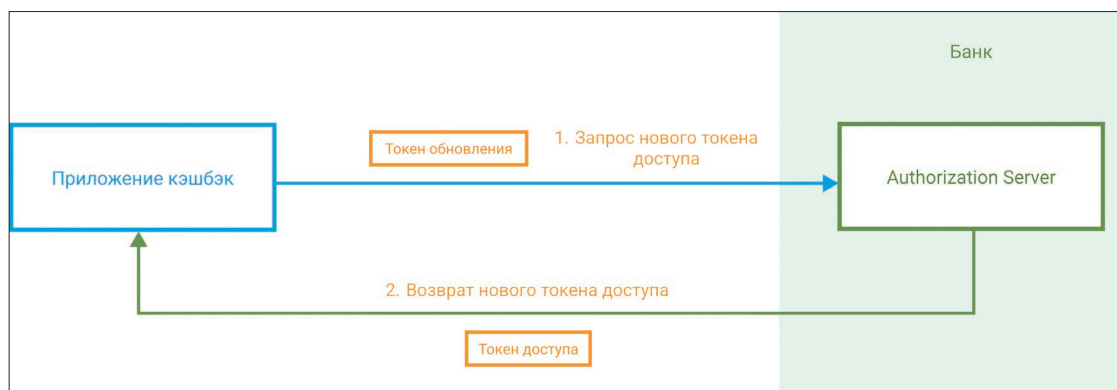


Рисунок 2. Принцип работы токенов доступа и обновления

доверенность с правом действовать от своего имени контрагенту. Когда пользователь цифровой финансовой услуги собирается делегировать поставщику третьей стороны такое право, то сделать это безопасно он может при помощи протокола OAuth 2.0 (рис. 2).

Для этого протокола авторизации важны два объекта — токен доступа (access token⁸) и токен обновления (refresh token⁹). Первый характеризует крайне ограниченное время жизни. Обычно он действует не более часа, так что для выдачи долгосрочного согласия одного токена доступа недостаточно. Проблему решает токен обновления: используя его, кэшбэк-сервис может запрашивать новые токены доступа на протяжении всего времени обращения к ресурсам пользователя.

Тем не менее полностью решить задачу, возложенную на стандарт Открытых API, протокол OAuth 2.0 не может: разработчиками он позиционируется как фреймворк, который нуждается в дальнейшем профилировании.

ПРОТОКОЛ АУТЕНТИФИКАЦИИ OPENID CONNECT

Протокол OpenID Connect представляет собой расширительный слой OAuth 2.0, допол-

няющий его идентификационным токеном (ID token), через который банк передаёт информацию о пользователе третьей стороне (кэшбэк-сервису из примера выше). В ID token можно «защитить» подтверждающую личность пользователя информацию — полные имя и фамилию или адрес электронной почты. Когда кэшбэк-сервис отправит запрос на сервер авторизации в следующий раз, в ответ он получит оба токена (идентификации и доступа) и со стопроцентной точностью удостоверит подлинность своего обращения.

В отличие от OAuth 2.0, протокол OpenID Connect поддерживает защиту запроса аутентификации, поэтому злоумышленникам гораздо сложнее подменить его содержимое. OpenID Connect содержит набор параметров, дополнительно увеличивающих информационную безопасность. Такой механизм гарантирует целостность аутентификации сервиса, то есть его неотрекаемость, которая достигается при помощи цифровой подписи.

МЕЖДУНАРОДНЫЙ СТАНДАРТ FINANCIAL-GRADE API

Протокол OpenID Connect носит общий характер, благодаря чему его можно адаптировать под различные требования к информа-

⁸ <https://oauth0.com/docs/secure/tokens/access-tokens>.

⁹ <https://oauth.net/2/refresh-tokens/>.

ционной безопасности. В частности, именно на основе OIDC был разработан международный стандарт для финансовых операций в банковской отрасли – Financial-Grade API (FAPI)¹⁰.

Financial-Grade API определяет два профиля:

- ♦ **базовый (Baseline)** применяется к Открытым интерфейсам, которые не изменяют состояние (доступ к пользовательскому счёту);

- ♦ **продвинутый (Advanced)** повышает требования к защите данных относительно базового и применяется к Открытым интерфейсам, которые изменяют состояние (проведение платёжных операций).

Базовый профиль не допускает аутентификацию между кешбэк-сервисом и банком с применением стандартных реквизитов, таких как логин и пароль. Он предполагает использование методов защиты с применением криптографии.

Для продвинутого профиля необходим request object (объект запроса) – JWT (JSON Web Token), включающий необходимые для запроса авторизации OAuth 2.0 параметры: client_id, response_type и прочие. Без этого невозможно обеспечить целостность и конфиденциальность параметров запроса авторизации. Если request object отсутствует, то любые обращения должны быть немедленно отклонены.

Продвинутый профиль вводит обязательное условие взаимной аутентификации на транспортном уровне (mutual TLS) между кешбэк-сервисом и банком. Чтобы оно было выполнено, не только банк, но и сам сервис должны предъявить к проверке цифровой сертификат X.509 (рис. 3).

Даже если токен доступа будет перехвачен при передаче, воспользоваться им злоумышленники не смогут, так как в него зашит отпечаток транспортного сертификата кешбэк-сервиса, который, в свою очередь, надёжно защищён.

ОТЛИЧИЯ РОССИЙСКОГО СТАНДАРТА

Спецификации некоммерческой организации OpenID Foundation, которая определяет технологии OpenID и порядок модели Открытых API со структурированными данными, предполагают, что Financial-Grade API могут использоваться на международном уровне в качестве «коробочного стандарта».

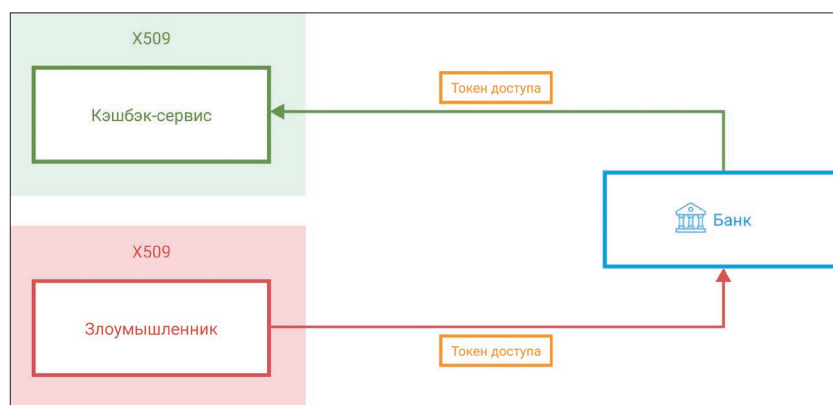


Рисунок 3. Защита аутентификации в профиле стандарта FAPI

То есть каждая страна вправе адаптировать FAPI под свои реалии.

Стандарт ФАПИ.СЕК, выпущенный Банком России, базируется на основных положениях FAPI, но с включением заметных дополнений, которые не наследуются от оригинального протокола и связаны с необходимостью соответствовать нормативным и законодательным актам Российской Федерации.

Ключевым отличием для России является исключительное применение отечественной криптографии согласно технической спецификации ТК26 «Использование российских криптографических алгоритмов в протоколах OpenID Connect»¹¹ (JWT, JWS, JWE, JWK на отечественных стандартах алгоритмов криптографической защиты).

Работа над внедрением на российский рынок Открытых API продолжается, и применение стандартов вида ФАПИ в качестве базисных должно существенно её облегчить. Как одна из стран, успешно развивающих технологии открытого банковского дела, Россия будет использовать эти стандарты в дальнейшем, хотя специфика национального финансового рынка, безусловно, скорректирует их профили.

Кроме нашей страны, открытый банкинг на основе ФАПИ сегодня развивают семь государств: США, Великобритания, Канада, Австралия, Бразилия, Нигерия и Новая Зеландия. Можно с уверенностью говорить, что большая часть рынка окончательно определилась и выбрала стандарты ФАПИ в качестве функциональных «строительных блоков» для решения разнообразных проблем, вызванных запуском сред открытого информационного обмена, где прозрачность и согласие пользователя стоят во главе угла.

¹⁰ <https://curity.io/resources/learn/what-is-financial-grade/>

¹¹ https://www.ruscrypto.ru/resource/archive/rc2022/files/07_gruntovitch.pdf

С ЛИЦЕВОГО СЧЕТА ПОХИЩЕНО...

СУДЕБНЫЕ ЭКСПЕРТИЗЫ В СФЕРЕ ДБО И ПРОТИВОДЕЙСТВИЕ МОШЕННИЧЕСТВУ



Владимир АВДЕЕВ
ведущий эксперт
технического
департамента
RTM Group

Появление методических рекомендаций № 3-МР от 28.02.2024, в которых ЦБ предписывает кредитным организациям принимать дополнительные меры для защиты своих клиентов, спровоцировало рост количества технических судебных экспертиз. Потребность в них вызвана частыми спорами между банками и клиентами о природе совершения некоторых операций (в том числе без согласия клиентов). Рассказываем об особенностях проведения экспертиз в сфере дистанционного банковского обслуживания (ДБО) и их значимости в контексте противодействия мошенничеству.

В ЧЁМ СУТЬ ЦБ РФ 3-МР?

Методические рекомендации ЦБ РФ 3-МР направлены на усиление информационной работы кредитных организаций с клиентами с тем, чтобы сократить число несанкционированных операций. Они включают в себя меры по улучшению ИБ, по аутентификации пользователей, по контролю за финансовыми операциями и обнаружению подозрительных транзакций. Также они обязывают кредитные организации предупреждать клиентов (через материалы, проведение мероприятий) о том, что недопустимо сообщать посторонним лицам конфиденциальные данные.

Рекомендации вступают в силу с 25 июля 2024 года и отменяют предыдущие рекомендации Банка России от 19 февраля 2021 года. Что же нового привнесено по сравнению с 2021 годом? Принципиальные изменения заключаются в следующем:

- ◆ расширенный объём рекомендаций;
- ◆ проведение обучающих мероприятий;
- ◆ внедрение двухфакторной аутентификации;
- ◆ дополнительные рекомендации по противодействию вовлечения в дропперство.

ПРЕДМЕТ ИССЛЕДОВАНИЯ В ЭКСПЕРТИЗАХ ДБО

К сожалению, несмотря на усилия кредитных организаций по информированию клиентов и принятию мер обеспечения безопасности, они не всегда способны в полном объёме предотвратить случаи хищения денежных средств.

Предположим, инцидент произошёл и деньги были похищены со счёта клиента, который считает, что антифрод-система банка не справилась со своей задачей. Специалисты кредитной организации, имеющие сильных юристов и техническую поддержку, легко докажут, что пострадавший не выполнил условия договора, а инструменты защиты сработали безупречно.

Нередко в спорных случаях клиенты могут обратиться в независимые организации для проведения исследований правильности функционирования системы ДБО. Назначая экспертизу, суд может указать её конкретный вид — компьютерная или компьютерно-техническая. В ходе исследования нужно будет получить ответы на следующие вопросы:

1. Обладают ли спорные операции признаками перевода денежных средств без согласия клиента?
2. Выполнялся ли в исследуемый период вход в личный кабинет банка с личного устройства клиента?
3. Направлялись ли клиенту push или SMS-уведомления следующего содержания <...>?
4. Имеются ли на устройстве клиента следы функционирования вредоносного программного обеспечения?

ОБЪЕКТЫ ЭКСПЕРТИЗЫ ДБО

Для обеспечения всестороннего исследования судьи нередко предоставляют полный доступ к материалам дела, что позволяет эксперту

ознакомиться со всеми доказательствами. И наиболее ценными объектами являются:

- ♦ документы, подтверждающие договорные отношения между клиентом и банком (договор, соглашение об обслуживании клиентов и др.);
- ♦ журналы событий с серверов банковской системы; чаще всего объектом исследования является выгрузка логов за исследуемый период, например журнал входа в личный кабинет клиента банка и история работы клиента в системе ДБО;
- ♦ журналы событий с серверов мобильного оператора (в случае если клиент выбрал оповещения посредством SMS);
- ♦ личное устройство клиента (ПК, мобильный телефон и т.д.);
- ♦ версия мобильного приложения банка за исследуемый период (APK / IPA файл) может предоставляться, если проводится анализ уязвимостей.

ЭТАПЫ ПРОВЕДЕНИЯ ЭКСПЕРТИЗЫ

1. Предварительное ознакомление с материалами дела. Изучение нормативной и документальной базы, позиции сторон, свидетельских показаний и т.д.

2. Исследование клиент-банковского взаимодействия:

- ♦ изучение журналов банковской системы, связанных с аутентификацией клиента, а также со спорными операциями, помогающими понять их последовательность и характер;
- ♦ анализ устройства клиента, включая выявление возможных уязвимостей, вредоносного ПО и т.д.

3. Исследование журналов мобильного оператора позволяет выявить информацию о передаче данных, совершении звонков или отправке SMS, потенциально связанных со спорными операциями.

4. Синтез. На основе предоставленных материалов и проведённых исследований эксперт восстанавливает хронологию событий.

5. Формулирование выводов и подготовка экспертного заключения. Выдается заключение эксперта с выводами по поставленным вопросам.

ПЕРСПЕКТИВЫ ПРИ РАЗРЕШЕНИИ СПОРОВ

Хотя статистика свидетельствует о том, что зачастую суд признаёт правоту банков (примерно в 62 % случаев клиентам отказывается в удовлетворении иска в полном объёме), вышеперечисленные примеры иллюстрируют, что для определения причин произошедшего инцидента, а также установления истины в подобных спорах требуется как техническое, так и нормативное исследование. Именно они нередко служат важным доказательством при принятии решения судом и влияют на исход дела.

Учитывая, что в 2024 году Центробанк расширил объём рекомендаций по усилению информационной работы кредитных организаций с пользователями, есть надежда на сокращение числа операций, производимых без согласия клиентов. И тогда, как следствие, снизится и количество судебных споров, связанных с неправомерными переводами денежных средств и заключением кредитных договоров.

Приведём примеры исследований случаев использования различных тактик злоумышленников.

Социальная инженерия. С лицевого счёта клиента банка похищено свыше 500 тыс. руб. Перед экспертом стояли задачи выяснить: производился ли вход в личный кабинет банка со смартфона клиента (Samsung A50), а также обладают ли спорные операции признаками перевода денежных средств без согласия клиента. Были выявлены несоответствия спорных операций признакам, утверждённым приказом Банка России от 27.09.2018 № ОД-2525, а именно нетипичное для клиента время входа в приложение (ночное), место и устройство (iPhone XR). Несмотря на это, система пропустила эти спорные операции. Однако представители банка ссылались на имеющийся в материалах дела опрос клиента, в котором тот сообщил о подозрительном звонке со стороны сотрудника «отдела безопасности» банка.

Фишинг. В другом случае на электронную почту клиента поступило письмо, содержащее во вложении вредоносный код, загружающий на компьютер жертвы «стилер», который крадёт сохранённые в браузере пароли и данные банковских карт. После чего со счёта жертвы была списана значительная сумма денежных средств.

В ходе досудебного расследования инцидента была установлена вся цепочка атаки и даны рекомендации, прежде всего по установке на компьютер средств защиты информации. В результате осознавший свою невнимательность клиент передумал подавать иск в сторону банка.

Внедрение ВПО. В данном случае предметом спора стало платёжное поручение, направленное с корпоративного ноутбука бухгалтера организации клиенту системно значимого банка. С целью выявления признаков нарушения условий договора по эксплуатации системы ДБО на исследование было передано само устройство с токеном электронной подписи, заявление о предоставлении комплексного банковского обслуживания, выписка по операциям на счёте клиента. Эксперты посредством методов цифровой криминалистики установили, что пострадавший нарушил меры безопасности: ключ электронной подписи был постоянно вставлен в ноутбук, антивирусные базы устарели и не обновлялись, была активирована нелегальная операционная система. Кроме того, на данном ноутбуке были выявлены следы функционирования программного обеспечения для удалённого администрирования, а также найден «троян-банкер», обладающий функционалом перехвата учётных данных, относящихся к банковским системам.

НОВЫЕ ВОЗМОЖНОСТИ — НОВЫЕ УЯЗВИМОСТИ

КИБЕРБЕЗОПАСНОСТЬ ЗАГОВОРИЛА БОЛЬШИМИ ЯЗЫКОВЫМИ МОДЕЛЯМИ

Нейросетевые технологии становятся неотъемлемым инструментом в борьбе с киберугрозами. Человеческий ресурс уже давно не в силах угнаться за развивающимся информационным пространством, поэтому отличать плохое от хорошего «в цифре» поручили большим языковым моделям — LLM (Large Language Models).

Вместе с экспертом компании «АйТи Бастион» разберёмся, в чём плюсы LLM-моделей, как они используются в сфере информационной безопасности и как их самих защищать от киберугроз.

#1

Что же такое большие языковые модели? LLM представляет собой класс алгоритмов машинного обучения, основанный на моделях естественного языка. Эти модели, созданные на базе огромных объёмов текстовых данных, способны анализировать, понимать и генерировать тексты с удивительной точностью и чёткостью. Их применение в области информационной безопасности открывает новые перспективы для обнаружения и предотвращения киберугроз.

Одной из ключевых областей применения LLM в сфере ИБ является анализ сетевых данных и обнаружение аномального поведения пользователей. Нейронные сети на основе моделей LLM могут анализировать большие объёмы логов и сетевых трафиков, идентифицируя паттерны, которые способны указывать на потенциальные угрозы безопасности.

Это позволяет оперативно реагировать на атаки, минимизируя потенциальные ущербы и исключая возможные потери.

Также преимущества использования LLM включают в себя их способность адаптироваться к новым угрозам без необходимости ручной переконфигурации или переобучения модели. Благодаря своей способности к обучению на больших объёмах данных модели LLM могут непрерывно улучшать свою производительность и эффективность в обнаружении киберрисков.

Более того, LLM не ограничиваются только анализом текстовой информации. Они также могут использоваться для анализа метаданных, аудио- и видеофайлов. Это делает их универсальным инструментом для обнаружения угроз безопасности различных типов.

#2

Разберёмся на конкретных угрозах, когда именно используются LLM в сфере ИБ.

1. Обнаружение мошенничества. LLM могут быть обучены распознавать тексты фишинговых сообщений и мошеннических писем. Благодаря способности к глубокому анализу языковых структур эти модели могут выявлять подозрительные и аномальные паттерны в текстах, которые часто ускользают от внимания человека.

2. Анализ инцидентов. LLM способны обрабатывать огромные объёмы данных, включая отчёты о безопасности и логи инцидентов. Используя эти данные, модели могут

помочь специалистам по безопасности в выявлении и классификации угроз, а также в предоставлении рекомендаций по реагированию на инциденты.

3. Создание тестовой среды. С помощью LLM можно создавать реалистичные сценарии кибератак для тренировки специалистов по информационной безопасности. Модели могут генерировать симулированные диалоги и электронные письма, которые имитируют поведение злоумышленников, обучая команды быстро и эффективно реагировать на угрозы.

4. Поддержка принятия решений. LLM могут анализировать текущие тенденции в сфере кибербезопасности и предоставлять обоснованные прогнозы относительно будущих угроз. Это позволяет руководителям и специалистам по безопасности принимать взвешенные решения, основанные на данных и вероятностных моделях.

5. Автоматизация рутинных задач. LLM могут автоматизировать множество рутинных задач, таких как классификация инцидентов, анализ логов и даже написание отчётов о безопасности. Это освобождает время специалистов для решения более сложных и творческих задач.

Таким образом, применение LLM в сфере информационной безопасности открывает новые горизонты для защиты цифровых активов. Их способность к обработке и анализу естественного языка делает их ценным инструментом в арсенале любого специалиста по кибербезопасности. По мере развития этих технологий

мы можем ожидать ещё большего усиления их роли в обеспечении информационной безопасности.

#3

Однако, несмотря на все свои преимущества, применение LLM в сфере ИБ также сталкивается с рядом вызовов. Поэтому необходимо **обеспечивать и безопасность самих моделей LLM**. При больших плюсах для бизнеса их использование связано с рядом киберрисков. Рассмотрим ключевые аспекты, которые нужно учитывать при таком подходе.

Важно учитывать: когда данные отправляются в облако для обработки, всегда есть риск их утечки или несанкционированного доступа к ним. Облачные провайдеры могут хранить и обрабатывать информацию в разных условиях, например, исходя из своей территориальной принадлежности, что усложняет контроль за их работой. А ещё эти процессы могут относиться к разным юрис-

дикциям, что также может привести к возможным вопросам в разрезе законодательных и нормативных требований.

Есть риск, что злоумышленники могут попытаться внедрить вредоносные данные непосредственно в сам набор для обучения модели. Это может привести к созданию модели с уязвимостями и ошибками.

Не стоит забывать про DoS-атаки, которые могут сделать облачные сервисы недоступными, «выключить» их из процесса, что может нарушить работу моделей и доступ к данным.

Кроме того, преступники могут попытаться извлечь информацию о LLM, чтобы воссоздать, к примеру, её копию или цифрового двойника или найти её уязвимости для дальнейших целенаправленных атак.

И конечно, использование LLM в принципе может привести к появлению определённых этических вопросов, таких как предвзятость самой модели и корректность её работы.

Поэтому информационная безопасность LLM-моделей требует комплексного подхода. Необходимо учитывать защиту данных на всех этапах жизненного цикла модели: от сбора данных и обучения — до её развёртывания и эксплуатации в рамках конкретной инфраструктуры. Потому что использование LLM без киберзащиты может привести не только к искажению и некорректности получаемых с её помощью результатов, но и к снижению их эффективности.

В целом применение LLM в сфере информационной безопасности открывает новые возможности для борьбы с киберугрозами и защиты цифровых активов. Эти передовые технологии становятся неотъемлемой частью современных стратегий безопасности и позволяют бизнесу эффективно реагировать на постоянно меняющуюся киберугрозную среду.

BIS-КОММЕНТАРИЙ

А КАК НА ПРАКТИКЕ?



Алексей ИВАНОВ
заместитель
технического
директора
по вопросам
инноваций
компании
«АйТи Бастион»

Наша команда специализируется на использовании передовых технологий для обеспечения ИБ. Мы применяем рекуррентные нейронные сети для анализа данных, поступающих с группы шлюзовых устройств внутри сетей заказчиков. Эти данные состоят из логов, которые содержат данные о действиях пользователей — от обычных операций до потенциально подозрительных действий.

Чтобы точно определить нормальное поведение пользователей, мы начинаем с выделения эталонной группы, которую называем «корректными» пользователями. Они характеризуются типичными паттернами действий и поведения в сети. Мы используем данные об их активностях для обучения наших нейронных сетей. Это делается, чтобы обеспечить моделям понимание того, что считается нормальным поведением.

Однако мир киберугроз постоянно меняется, злоумышленники развивают и адаптируют свои методы. Поэтому полагаться только на эталоны нельзя. Вместо этого в режиме реального времени мы анализируем поступающие логи от всех пользователей, не принадлежащих к эталонной группе. Мы используем наши обученные модели, чтобы оценить, насколько их действия коррелируют с поведением эталонных пользователей.

Например, если в системе происходит необычное перемещение файлов или подозрительная активность в попытках несанкционированного доступа к системным ресурсам, наши алгоритмы могут автоматически выделить эти действия и «поднять тревогу». Мы предоставляем детальные отчёты о таких событиях офицерам безопасности и руководителям ИБ-служб заказчиков, которые могут принять соответствующие меры для защиты сети, обеспечивая мгновенное обнаружение аномального поведения и реакцию на потенциальные угрозы безопасности сети.

Такой подход позволяет нам не только обнаруживать и предотвращать потенциальные угрозы, но и постоянно совершенствовать нашу систему безопасности, делая её более чуткой к изменениям в поведении пользователей и адаптивной к новым вызовам.

ОТ АНАЛИЗА КОНТЕНТА К АНАЛИЗУ КОНТЕКСТА

НОВАЯ ПАРАДИГМА РАЗВИТИЯ DLP-СИСТЕМ

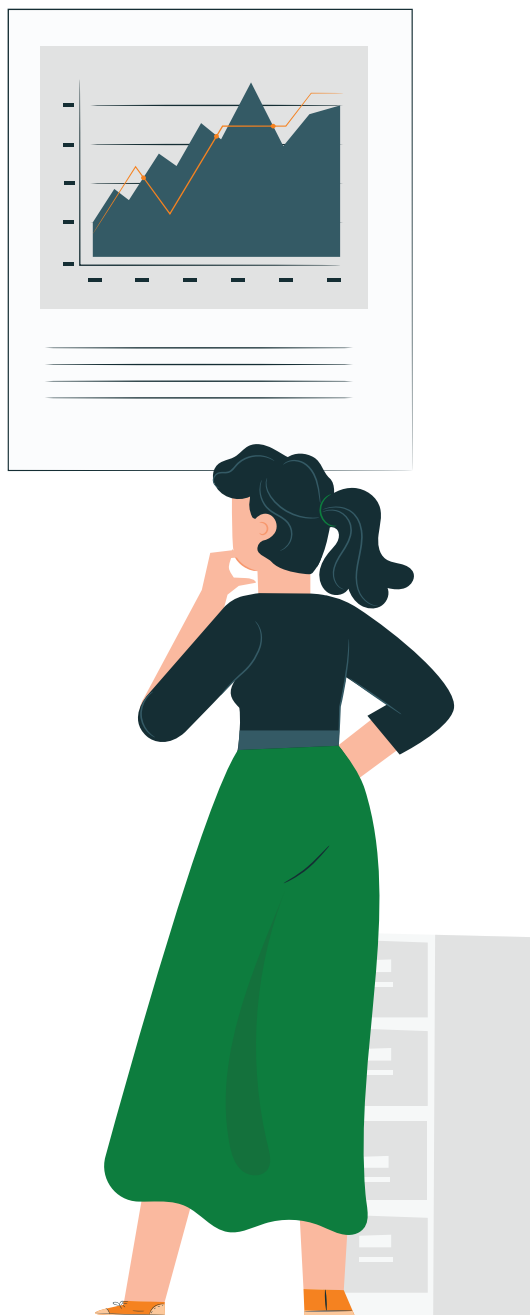


**Александр
КЛЕВЦОВ**
руководитель
по развитию
InfoWatch
Traffic Monitor

На сегодняшний день DLP-системы научились качественно и точно выделять в информационных потоках компании конфиденциальную ин-

формацию. Это позволяют делать технологии контентного анализа: они анализируют содержимое писем, сообщений в мессенджерах, файлов в облачных хранилищах, на съёмных носителях и так далее. Пользователи DLP-системы настраивают технологии анализа таким образом, чтобы они отражали информационные активы компании, которые следует защищать. Это могут быть и корпоративные решения, и созданные с нуля. Например, настройки словарей позволяют определять тематику любого документа вне зависимо-

сти от структуры, цифровые отпечатки выявляют фрагменты, редакции или черновики конкретных документов, регулярные выражения детектируют реквизиты и атрибуты, в том числе персональные данные. В последние годы стали появляться технологии, анализирующие изображения, чертежи, записи голоса. Словом, современная DLP-система способна распознать практически любой конфиденциальный документ. Но достаточно ли этого для своевременного предотвращения утечек?



Привет, как обещал, скидываю
контакт репетитора Иван
89721234567

20:50 ✓

Рисунок 1. Такое сообщение может быть утечкой, а может быть и правда передачей контактов репетитора

хороший актив и широкую возможность кросс-продаж. Клиент в определённый момент оформил ипотеку, а спустя полгода рефинансировался в другом банке, который предложил ему более низкую ставку по кредиту. Похожая история произошла ещё с несколькими крупными клиентами, и эта цепочка заинтересовала службу экономической безопасности. Как выяснилось, один из сотрудников банка систематически передавал контактные данные VIP-клиентов менеджеру другого банка и получал за это вознаграждение.

На что могла бы отреагировать DLP-система в данном случае, чтобы выявить утечку своевременно? Например, на номер телефона? Но если в DLP-системе будет создана политика, реагирующая на номер телефона как на инцидент, то любая подпись в письме, любой легитимный обмен контактами будут вызывать ложноположительное срабатывание (ЛПС). И таким образом сигналы DLP-системы фактически не будут нести никакой ценности, потому что их будет бесконечное множество (рис. 1).

Предположим, что офицер безопасности решил просматривать вручную все сообщения, содержащие номер телефона. Но даже если представить в теории, что штата специалистов ИБ хватит на ручной разбор всего «белого шума», такой подход тоже не даст результата: чтобы отличить номер телефона репетитора, которым решили обменяться сотрудники, от контакта клиента, офицеру безопасности нужно знать наизусть всю контактную информацию всех клиентов. Разумеется, это нерабочий вариант.

И в данном случае получается, что и человек (офицер безопасности),

и машина (DLP-система) могут определить, что сообщение содержит номер телефона, но не могут определить ценность этого номера. А она может быть очень велика: в случае с утечкой контакта VIP-клиента под видом контакта репетитора это несколько сотен миллионов рублей. И здесь анализа контента, то есть анализа непосредственно содержимого, с определением категории и структуры данных, недостаточно. Для эффективной защиты данных здесь не хватает понимания того, как они задействованы в бизнес-процессах компании, какую роль в них играют.

В результате мы сталкиваемся с ситуацией, когда специалисты ИБ понимают, что подобные «мелкие», с точки зрения объёма данных, утечки возможны, но закрыть этот риск невозможно. DLP может детектировать крупные фрагменты конфиденциальной информации — документы и сообщения достаточно объёмные, чтобы их можно было отнести к какой-либо категории, выгрузки из бизнес-систем с большим количеством строк и столбцов, но для детектирования таких маленьких, но опасных утечек нужен более серьёзный фильтр.

Рассмотрим второй пример. Письмо VIP-клиенту с индивидуальным предложением имеет следующий фрагмент: «Уважаемый Иван Сергеевич, хотим вам предложить выгодные условия по депозитам». Письмо отправляется на адрес s.i.petrov@gmail.com. Возможно ли определить, что сотрудник клиентского отдела случайно или намеренно не отправил письмо по похожему адресу — например, s.petrov@gmail.com? Для этого нужно на первом шаге определить конкретную пор-

КОНТЕНТ УЖЕ НЕ «НАШЕ ВСЁ»

Рассмотрим первый пример, где инструментов контентного анализа недостаточно для эффективной защиты данных. Сотрудник отправил сообщение в мессенджере Telegram: «Привет, как обещал, скидываю тебе контакт репетитора Иван 89012345678». На первый взгляд, безобидное сообщение, но в действительности это реальный кейс утечки в финансовой организации. Произошла следующая ситуация: банк работал с крупным VIP-клиентом с большими остатками на счетах, получая таким образом

Привет, напиши про ремонт
в Камышине ооо "зерно и солнце"

20:52 ✓

Медная труба 6,35 x 0,76мм 15м
3150

20:57 ✓

Рисунок 2. Оба примера — намеренная передача конфиденциальной информации. В первом случае — передача третьим лицам подряда на ремонт сельхозтехники, во втором случае — утечка закупочных цен в ретейле. Оба кейса встречались в практике ГК InfoWatch

цию персональных данных, а на втором — отправляются ли они владельцу персданных, а не третьему лицу. Для DLP-системы классического дизайна это нерешаемая задача: чтобы связать конкретные персданные с конкретным получателем, пришлось бы создавать отдельное правило в политиках безопасности для каждого клиента с указанием легитимного адреса электронной почты. 100 000 клиентов — 100 000 правил. И здесь мы вновь видим, что недостаточно ни инструментов контентного анализа, ни инструментов политик.

Отдельная проблема заключается в том, что существует конфиденциальная информация, которую сложно классифицировать: Ф. И. О., адреса, названия контрагентов, продуктов, товарных позиций, любые числовые показатели — например, закупочные цены. Такие реквизиты и атрибуты не имеют строгой структуры и тем более контрольных сумм и проверочных алгоритмов. И если офицер безопасности решит создать регулярное выражение, которое будет детектировать, например, единичные нелегитимные отправки товарной позиции + цены, то мы получим огромное количество ЛПС. А те реквизиты, которые поддаются классификации, — номер телефона или email — возвращают нас к проблеме: недостаточно детектировать отправку «какого-то

телефона», нужно определить в потоке номеров конфиденциальные данные клиентов, передача которых является утечкой.

Здесь же стоит добавить, что, кроме сложной классификации, есть и ещё одна особенность — огромный объём защищаемой информации, которая регулярно обновляется и изменяется. Отслеживать вручную все изменения и отражать в политиках безопасности невозможно (рис. 2).

ОТ КОНТЕНТА К КОНТЕКСТУ

Важнейший вывод, к которому мы приходим, — контентного анализа для противодействия утечкам и другим нарушениям, связанным с информацией, недостаточно.

Резюмируем причины:

- ♦ часть конфиденциальной информации плохо поддаётся классификации, попытки её выявлять будут приводить к большому количеству ЛПС;
- ♦ часто недостаточно выявить только категорию информации — нужно принять решение по конкретному значению, например, один номер телефона — утечка, другой — всего лишь подпись в письме;
- ♦ огромные объёмы отдельных категорий конфиденциальных данных: клиентские базы, базы сотрудников, прайс-листы и т.д. Эти данные меняются ежедневно, отразить их в политиках невозможно;

♦ политики безопасности часто имеют достаточно общие правила и описывают маршруты распространения данных на уровне масштаба подразделений, отделов, и никто не описывает связку, например, ПДн и легитимного email-адреса для данного субъекта ПДн.

Поэтому мы создали в DLP-системе InfoWatch Traffic Monitor инструмент, который при принятии решений учитывает контекст — ценность той или иной порции данных, владельцев и статусы документов, роли контрагентов и другие признаки, исходя из бизнес-процессов. Как это возможно и откуда взять этот самый контекст? Этот инструмент «забирает» контекст оттуда, где он отражён в корпоративных информационных системах (КИС) — ABS, ERP, CRM и других, в зависимости от специфики бизнеса.

Первым кейсом внедрения этого инструмента контекстного анализа стал банк, работающий с юридическими лицами. Бизнес поставил перед нами несколько задач:

♦ данные партнёров и клиентов могут отправляться только на те email-адреса, которые закреплены в качестве контактов контрагентов; если в письме есть посторонние адреса, то письмо должно быть заблокировано;

♦ важно определять статус документов: если он уже публичный, то можно отправлять за периметр компании, если документ пока в разработке, то нужно блокировать отправку;

♦ необходимо блокировать письма, содержащие коммерческую тайну, если с контрагентом на текущий момент не подписан NDA.

С помощью API InfoWatch Traffic Monitor был интегрирован с ABS, CRM, СЭД банка и импортировал реквизиты клиентов и партнёров: легитимные контакты для отправки документов, номера счетов, названия компаний, адреса, название продуктов, статус NDA и т.д. По сути, была реализована автоматическая политика с тысячами правил, которая принимает решение, исходя из контекста каждого контрагента и документа. За счёт синхронизации с КИС

правила этой автоматической политики поддерживаются в актуальном состоянии. Таким образом, защищается всегда актуальная версия данных. В этом кейсе мы и получили тот самый переход от контентного анализа к контекстному.

Другое внедрение нового инструмента InfoWatch Traffic Monitor мы реализовали в производственной компании, которая стремилась защитить данные своих сотрудников: — полисы ДМС, трудовые договоры, расчётные листы по заработной плате и другие. Бизнесу было важно, чтобы в массовых почтовых рассылках документы получали только сотрудники, которым они были адресованы, а любые ошибки и попытки отправить документ постороннему лицу DLP-система должна блокировать. При этом сам сотрудник может распоряжаться своими документами как захочет — например, пересылать их на личный почтовый адрес или членам семьи. Интеграция с ERP-системой позволила реализовать следующую логику: если в письме определяются данные конкретного сотрудника, то это письмо можно пропустить только в двух случаях — либо отправитель — сотрудник отдела кадров, а фактический получатель совпадает с email сотрудника, либо фактический отправитель совпадает с email сотрудника. В остальных случаях письмо блокируется. Количество сотрудников во всех подразделениях компании более 20 000, и с помощью инструмента контекстного анализа InfoWatch Traffic Monitor проверяют на легитимность несколько десятков тысяч писем ежемесячно. Чего только стоят два зарплатных листа в месяц для каждого сотрудника.

НОВАЯ ПАРАДИГМА РАЗВИТИЯ РЫНКА DLP-СИСТЕМ

Стоит подчеркнуть, что оба описанных внедрения выдвигали требования не просто по выявлению нарушений, а именно по предотвращению утечек — то есть блокировке писем. А это значит, что мы не могли себе позволить никаких отложенных обработок, решение должно приниматься незамедлительно: не полу-

Мы не могли себе позволить никаких отложенных обработок, решение должно приниматься незамедлительно: не получится перехватить письмо, разобрать его, обратиться к ABS, получить ответ и только после этого принять решение

чится перехватить письмо, разобрать его, обратиться к ABS, получить ответ и только после этого принять решение. Это привело бы к накоплению очередей обработки, учитывая интенсивность переписки в enterprise-сегменте. Производительности, которая позволила DLP-системе работать в режиме блокировки, удалось добиться благодаря тому, что данные из корпоративных систем превращаются в высокопроизводительный индекс, например, сравнение сообщения с клиентской базой в 10 млн записей занимает 0,1 секунды. Важно, что индексируются конкретные значения — названия компаний, телефоны, адреса, названия продуктов и другие. И проблема неклассифицированных данных и большого количества ЛПС становится неактуальна, потому что каждое сообщение сравнивается с массивом конкретных значений, даже если это массив в несколько миллионов значений. Алгоритмы индексации и поиска — собственная и запатентованная разработка ГК InfoWatch.

Второе ключевое качество инструмента контекстного анализа InfoWatch Traffic Monitor — решение может учитывать нюансы бизнес-процессов компании. Как мы видим из примеров выше, на уровне контекста можно защищать данные партнёров, клиентов, сотрудников и связанных с ними информационных активов. Для этого в инструменте предусмотрена технология гибкого описания правил. Такой подход позволяет реализовать проект, в котором, например, будут защищаться данные поставщиков и связанных с ними товарных позиций, чтобы такой ключевой параметр, как закупочные цены, не попал

в руки третьим лицам. Технология по конфигурированию и принятию решений получила отдельный патент в июне 2024 года.

Третья ключевая особенность нового инструмента — открытое API, которое позволяет интегрироваться с любой корпоративной информационной системой. Банк, в котором мы впервые внедряли решение, используя SDK, самостоятельно разработал интеграцию DLP-системы с ABS, CRM и СЭД, в случае с производственной компанией помогли наши инженеры.

В заключение отмечу, что технология, созданная InfoWatch, уже прошла череду пилотов и внедрений у заказчиков. Мы получили самые позитивные отклики и помогли клиентам решить задачи разного масштаба и сложности, от уже известной проблемы с зарплатными листками, о которой упоминалось выше, до принципиально новых задач, таких как, например, защита каждой порции персональных данных с учётом их владельцев. На этапе проведения пилотов мы помогаем выявить эти проблемы, и это полностью win-win-процесс: компании-заказчики выводят защиту данных на более высокий уровень, мы как производители расширяем потенциал технологии и наращаем экспертный опыт, а рынок DLP-систем получает принципиально новый подход, который в перспективе может стать новой парадигмой его развития.

ОНИ СРЕДИ НАС

ПОЧЕМУ ИНСАЙДЕРСКИЕ АТАКИ ТАК ОПАСНЫ



Анатолий ПЕСКОВСКИЙ
эксперт
департамента
противодействия
киберугрозам
АО НИП
«Информзащита»

Информационная безопасность — сложная система, которая должна обеспечивать защиту данных, учитывая возможность атаки с разных сторон. Однако компании сегодня зачастую смещают фокус на борьбу с внешними угрозами в лице киберпреступников, упуская из виду собственных сотрудников с доступом к корпоративным сетям и данным. Они также могут стать внутренними нарушителями и своими действиями или бездействием нанести вред организации.

При проведении аудита почти у половины компаний обнаруживаются критические недостатки в процессе управления доступом к конфиденциальным данным. «Информзащите» в 70 % проектов по внутреннему пен-тесту удается добраться до чувствительной информации под учетными записями рядовых сотрудников, что свидетельствует о плохой защищенности.

ЯЗЫКОМ ФАКТОВ

Статистика говорит, что такое распределение внимания компаний негативно сказывается на уровне информационной безопасности. Каждый год количество связанных с инсайдерскими атаками инцидентов увеличивается на 25 %, при этом реальная цифра может быть еще выше, так как не все подобные кейсы регистрируются. В 2023 году в России около 28 % всех случаев нарушения информационной безопасности были связаны с инсайдерами, 62 % организаций столкнулись с такими вызовами. Среди основных факторов, способствующих этому росту, выделяются недостаточная обученность сотрудников (55 % от всех случаев), использование в профессиональной деятельности личных устройств (50 %) и удаленная работа (45 %). Все это подогревается и стремительно растущим спросом на инсайдерскую информацию: в даркнете он ежегодно увеличивается на 50 %.

Чаще всего инсайдерские атаки приводят к двум видам инцидентов: утечке данных (40 %) и несанкционированному доступу к информационным системам (35 %). Например, только в 2023 году из российских финансовых организаций утекло 170,3 млн записей персональных данных клиентов, что больше в 3,2 раза, чем в 2022 году, и в 57 (!) раз

больше, чем в 2021 году. При этом основными типами угроз являются злоупотребление привилегиями сотрудниками (45 %), ошибки и небрежность работников (35 %) и кража данных (20 %).

Среди наиболее подверженных таким проблемам отраслей находятся финансовый сектор, государственные структуры, промышленность и IT-компании. Финансовый ущерб от одного случая инсайдерской атаки в среднем обходится пострадавшей организации в 10 млн рублей, не говоря о часто сопровождающих репутационных потерях.

ЧЕТЫРЕ ТИПА

Все вышеперечисленные данные свидетельствуют, что отношение бизнеса к инсайдерским атакам должно быть пересмотрено. Для этого важно понимать, какие виды внутренних нарушителей есть, кто из сотрудников наиболее подвержен тому, чтобы стать угрозой для информационной безопасности, а также основные причины таких атак.

Существуют четыре основных типа внутренних нарушителей. Во-первых, наиболее очевидные в этом случае, **злонамеренные инсайдеры**. Эти сотрудники специально своими действиями наносят вред организации, чаще всего из-за собственных финансовых выгод, желания мести, идеологических убеждений и так далее. Далее стоит сказать о **неосторожных инсайдерах** — тех, кто становится вредителями по невнимательности, из-за недостатка знаний и несоблюдения установленных правил работы с конфиденциальной информацией. Еще один тип нарушителей — **компрометированные инсайдеры**. К такой категории относятся работники, чьи данные были украдены и теперь используются злоумышленниками для доступа к корпоративной информации. Четвертая

Чаще всего инсайдерские атаки приводят к утечке данных (40 %) и несанкционированному доступу (35 %). В 2023 году из российских финансовых организаций утекло 170,3 млн записей персональных данных клиентов, что больше в 3,2 раза, чем в 2022 году, и в 57 (!) раз больше, чем в 2021 году

группа вредителей — **контрагенты и посетители**, которые могут стать ситуативными инсайдерами, получив физический доступ к защищенной информации в офисе компании.

КТО ОНИ?

Какие сотрудники чаще всего становятся «внутренними врагами»? В первую очередь стоит говорить о тех, кто имеет доступ к защищаемой информации и при соответствующей мотивации или по неосторожности может нанести вред компании. Работники, не имеющие доступа к конфиденциальной информации, но допущенные в помещения, где она хранится, а также обслуживающие информационные системы сотрудники и доверенные лица организации также являются потенциальной угрозой. Студенты-практиканты и стажеры в силу своей неопытности и незнания правил работы с внутренней информацией компании тоже могут нанести ей вред.

Проблемой для информационной безопасности бизнеса могут стать и любые посторонние люди, например, курьеры или гости, получившие доступ в помещения компании. Именно они относятся к четвертой категории инсайдеров. «Информзащита» в рамках работ по физическому пентесту часто моделирует действия злоумышленников, которые посещают офис заказчика. К примеру, имитируется ситуация с проходом в контур заказчика без бейджика или предварительно скопировав бейдж неосторожного сотрудника в метро или другом месте. Другая проверка состоит в разбрасывании потенциально вредоносных носителей в рабочих пространствах с целью выяснить, кто из сотрудников решит вставить какой-то из них в свой компьютер. К сожалению, хакерам сегодня не нужно прилагать усилия для поиска внешних возможностей взломать систему, достаточно найти ангажированного сотрудника, который из корыстных помыслов или из-за недалекости сойдет на обходную информацию.

СЕМЬ ПРИЧИН

Можно выделить семь основных причин возникновения инсайдерских атак:

1. Недостаточный контроль доступа. Иногда компании предоставляют слишком широкие привилегии сотрудникам, в том числе временным, и подрядчикам. Это, очевидно, приводит к повышению риска несанкционированного доступа к критически важным системам и данным, что может привести к утечке. Например, в декабре 2022 года стало известно о признании бывшего работника Twitter виновным в осуществлении шпионажа за пользователями социальной сети. Работая в компании, он передавал Саудовской Аравии личные данные пользователей;

2. Отсутствие систем мониторинга и аудита. Организации даже при наличии таких инструментов могут использовать их недостаточно эффективно. В результате такого отношения невозможно своевременно выявлять подозрительную активность и предотвращать инсайдерские атаки;

3. Недостаточное обучение и осведомленность сотрудников. Из-за отсутствия регулярного обучения по вопросам ИБ для сотрудников возникает их небрежное отношение, например, открытие фишинговых писем или другие действия, которые приводят в случайной передаче конфиденциальной информации;

4. Отсутствие строгих политик безопасности. Когда их нет или же они прописаны не слишком четко и строго, повышается риск утечки данных и других связанных с инсайдерскими атаками инцидентов;

5. Недостаточное внимание к обеспечению физической безопасности. При условии отсутствия эффективных мер контроля и мониторинга для любых посторонних людей, получающих доступ в офисные помещения, растет риск несанкционированного доступа к конфиденциальной информации и злоупотребления со стороны внешних лиц;

6. Неправильная сегментация сети. В случае наличия такой ошибки в построении системы информаци-

онной безопасности увеличивается масштаб потенциально нанесенного вреда в результате инсайдерской атаки, так как взлом одного компьютера ведет к доступу ко всей корпоративной сети.

7. Отсутствие многофакторной аутентификации (MFA). Неиспользование такого механизма ставит под угрозу конфиденциальные данные, так как скомпрометированная учетная запись сотрудника будет давать злоумышленнику доступ ко всей системе.

КОМПЛЕКСНЫЙ ПОДХОД

Применение комплексного подхода к безопасности поможет значительно снизить риск инцидентов, вызванных внутренними нарушителями. Среди инструментов можно отметить разработку политик безопасности, в том числе на основе требований и рекомендаций ФСТЭК и ФСБ, инцидент-менеджмент, для конкретного механизма реагирования на кейсы с инсайдерскими атаками, обучение и проверку сотрудников, а также регулярные аудиты и проверки на соответствие требованиям законодательства и внутренним политикам компании.

Противодействие внутренним угрозам было и остаётся одной из наиболее трудоёмких задач специалиста по безопасности, пренебрежение которой делает вероятность утечки информации на порядок более осязаемой. Эффективная защита от внутренних нарушителей — это непрерывный процесс, он требует комплексного подхода, включающего использование разнообразных технических инструментов и методов управления, а также обучение и повышение осведомленности сотрудников. Современные технологии позволяют значительно снизить риски, связанные с внутренними угрозами, и обеспечить более высокий уровень защиты информации в организации.

II ФОРУМ «ТЕХНОЛОГИИ ДОВЕРЕННОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»

ДОВЕРИЕ ДОЛЖНО БАЗИРОВАТЬСЯ НА НАУКЕ

Марина БРОДСКАЯ
обозреватель BIS Journal

В Москве прошёл II форум «Технологии доверенного искусственного интеллекта», объединивший представителей профильных государственных ведомств, исследователей в области технологий ИИ, криптографии и информационной безопасности, а также специалистов компаний, реализующих проекты в области искусственного интеллекта.

Открылся Форум традиционной ключевой дискуссией «Доверенный искусственный интеллект». В её работе приняли участие представители государственных ведомств, регуляторов и исследовательских центров. Модератор дискуссии Игорь Качалин (АНО «НТИ ЦК») отметил основные тенденции развития ИИ и задачи, стоящие перед разработчиками. Он напомнил, что Национальная стратегия развития ИИ на период до 2030 года была разработана в 2019 году и обновлена в 2024 году на фоне геополитических событий, нехватки числительных мощностей, дефицита кадров, развития отечественных решений и изменений в области ИИ.

МИНЦИФРЫ РФ

Об актуальности ИИ рассказал Александр Шойтов (Минцифры РФ).

Он оценил искусственный интеллект как одну из самых перспективных технологий современности. «По различным оценкам, ИИ в сравнении с другими технологиями считается лидером в плане перспектив развития. И весь мир движется в этом направлении», — отметил замминистра. Он напомнил о трёх мировых подходах к регулированию ИИ — европейском, американском и китайском, а также о документах, принятых в России: Кодексе этики ИИ, Национальной стратегии развития ИИ, Комплексе мероприятий в рамках нацпроекта «Экономика данных» и Национальных целей развития. По словам спикера, во всех документах указывается, что ИИ должен быть не только эффективным, но и безопасным, доверенным.

На понятии «доверия», которое сегодня носит верхнеуровневый характер в ИИ и подразумевает требования разработчиков, потребителей и регуляторов, А. Шойтов остановился подробнее. Он напомнил, что общие вопросы «доверия» прописаны в документах ФСТЭК России. Активно обсуждаются понятия надёжности, функциональной корректности, технологической независимости, ИБ и устойчивости. «Они до конца не увязаны», — рассказал замминистра.

Он отметил риски атак на ИИ и рассказал об организациях, с которыми Минцифры работает по вопросам доверенного ИИ. Академия криптографии выпускает теорети-

ческие исследования в области ИБ, доверия и обезличивания. Институт системного программирования РАН (ИСП РАН) занимается доверенным ПО, доверенным ядром и создаёт модели ИИ с элементами доверия.

Также замминистра рассказал, что 21 мая был создан Консорциум исследований в области доверенного ИИ. Аналогичная структура практически в это же время создана и в США. Консорциум будет предлагать защищённые решения. В краткосрочной перспективе среди его задач — тестирование уже существующих систем.

Замминистра напомнил, что в России принят риск-ориентированный подход к развитию и регулированию ИИ, который предполагает определение рисков ИБ и их снятие для самой сложной модели системы. Таковой А. Шойтов считает модель ядерного реактора.

ДОВЕРЕННОГО ИИ НЕТ!

Тему регуляторики ИИ продолжил замдиректора ФСТЭК России Виталий Лютиков. Он напомнил, что на I форуме высказал тезис о том, что доверенного ИИ в РФ нет и в ближайшие годы не появится. Прошёл год — ситуация не изменилась. «Нет доверенного ИИ! Я исхожу в этом утверждении из понятия, которое даёт обновлённая Национальная стратегия развития ИИ», — подчеркнул В. Лютиков. Документ определяет доверенный ИИ как технологическое решение, соответствующее стандартам безопасно-



сти, разработанное с учётом объективности, непредвзятости, этичности и исключения применения для нанесения угрозы гражданам и государству. «Пока по этим составляющим мы не разработаем требования, а главное, не сформируем методики подтверждения требований, говорить о доверенном ИИ преждевременно, — заявил спикер. — Если опираться на Стратегию в понимании того, что такое ИИ, то он состоит из информационно-коммуникационной инфраструктуры, ПО и сервисов обработки данных и принятия решения. То есть, чтобы ответить на вопрос доверия ИИ, нам необходимо к этим трём сущностям предъявить стандарты безопасности. Пока мы этого не сделаем, говорить о том, что у кого-то доверенные технологии ИИ, а у кого-то недоверенные, бессмысленно». Также нельзя использовать те наработки ИИ, которые есть сегодня, в промышленно-сти, считает спикер.

ФСТЭК России прорабатывает вопрос установки набора требований к государственным системам при использовании ИИ с учётом особенностей моделей применения. Второй вопрос — разработка методического документа, рекомендаций для разработчиков и операторов систем ИИ

для формирования концептуальных подходов к вопросам специфичных угроз для ИИ. «Эти документы будут открыты и публичны для обсуждения в сообществе», — отметил спикер.

Проблема регулирования ИИ является мультиминистерской, считает В. Лютиков. «Одно ведомство не сможет решить эту задачу, потому что, даже исходя из понятий, приведённых в Стратегии, стандарты безопасности должны охватывать аспекты информационно-коммуникационной инфраструктуры, ПО и главное — процесс обработки данных и принятия решений», — отметил спикер, обратив внимание на важность третьей сущности — процесса обработки и принятия решений, которая «является главной, и там появляется очень много отраслевой специфики». Исходя из понятия доверенного ИИ, объективности и этичности его применения и исключения создания угроз, отраслевую специфику необходимо учитывать с точки зрения регулирования.

НЕЧЁТКАЯ ЛОГИКА

«Атомный реактор, который подключён к интернету и управляется ИИ, — это пока недостижимая история. Надо ли это делать — предмет дискуссий. Но определённые сферы

применения ИИ точно есть», — считает Евгений Абакумов («Росатом»).

Он затронул вопрос доверия в применяемых системах ИИ на объектах КИИ «Росатома» и остановился на датасетах, которые нужно использовать для обучения, а также на проблемах технологий обработки информации и дефицита вычислительных средств, которые позволяют использовать эти технологии. В России таких вычислительных архитектур немного, отметил спикер, и они не являются лидерами в своих сегментах.

Вторая проблема — нечёткая логика, которую не хочется использовать в системах КИИ, подчеркнул спикер. Он сообщил, что рассматривает технологии ИИ в логике систем поддержки принятия решений, когда ИИ является подсказчиком. На данный момент этому подсказчику доверять трудно.

Представитель «Росатома» поддержал позицию В. Лютикова, что на сегодня в промышленных системах нет доверенного промышленно-эксплуатируемого ИИ. При этом есть много сфер применения, имеющих кросс-отраслевую экспертизу, считает спикер. Он призвал участников Форума идти по пути развития прикладных информационных систем и вычислительных архитектур.



НА МИРОВОЙ АРЕНЕ

Представитель МИД России Артур Люкманов продолжил тему ИИ в мире. Однако он выделил европейскую, американскую и азиатскую модели развития ИИ и отдельно китайскую, охватывающую, в отличие от других, все сферы общества и государства. Чиновник отметил отсут-

ствие четкого представления, что такое ИИ, разное качество экспертизы в разных странах и шум вокруг темы ИИ. Он негативно оценил попытку компаний из США за столбить внешние рынки, заложив в основу регулирования доступа к системам ИИ права человека и вопросы демократии.

Российская Федерация активно работает через структуры ООН и Международный союз электросвязи (МСЭ), на площадке которого предлагает рассматривать и продвигать вопросы ИИ строго функционально.

ДОВЕРИЕ БАЗИРУЕТСЯ НА НАУКЕ

О достижениях и наработках отечественных ученых в области ИИ рассказал Арутюн Аветисян (ИСП РАН), завершивший дискуссию. Он отметил многолетний вклад Академии наук и российских ученых в развитие научного подхода к построению технологий ИИ, а также призвал к диалогу ИТ-отрасли с гуманитариями по вопросам доверия к ИИ. «Должно быть четкое понимание, как вы контролируете процесс разработки», на который накладывается отраслевая специфика и этические вопросы, считает ученый. Всё это впоследствии преломляется и превращается в конкретные этапы жизненного цикла разработки.

А КАК НА ПРАКТИКЕ?

ОБ ЭТОМ РАССКАЗАЛИ НА КОНФЕРЕНЦИИ
«ИНТЕЛЛЕКТ НОВОГО ПОКОЛЕНИЯ»

О применении ИИ на практике говорили участники мероприятия «Интеллект нового поколения», организованного газетой «Ведомости» и компанией FESCO.

ЧЕТЫРЕ НАПРАВЛЕНИЯ

Так, компания FESCO выделяет четыре основных направления внедрения ИИ: производство, маркетинг и коммерция, внутренние коммуникации, образование и работа с персоналом. Работы по внедрению ИИ в производство находятся в начальной стадии, однако уже можно говорить об успешной аналитике данных, оптимизации бизнес-процессов и сокращении производственных затрат. Например, пилотный проект по оптимизации склада контейнеров вы-

явил места, где внедрение ИИ помогает снизить стоимость операций, повысить эффективность управления контейнерным парком и сократить число перемещений контейнеров в терминалах.

В ИНТЕРНЕТ-ТОРГОВЛЕ И МАРКЕТИНГЕ

Основная задача инструментов на базе генеративного ИИ в e-commerce и маркетинге — экономия времени, считают в компании «Сбермаркетинг». Компания реализовала три проекта, направленные на генерацию текстов для контекстных объявлений, графики и транскрибирования аудио- и видеоматериалов с кратким изложением содержания. Экономия времени составляет до 70% при работе с текстовыми объявлениями.

Использование технологий ИИ для позиционирования продуктов перед выводом на рынок дает решение за 5 часов. Людям на это требуется несколько дней мозговых штурмов.

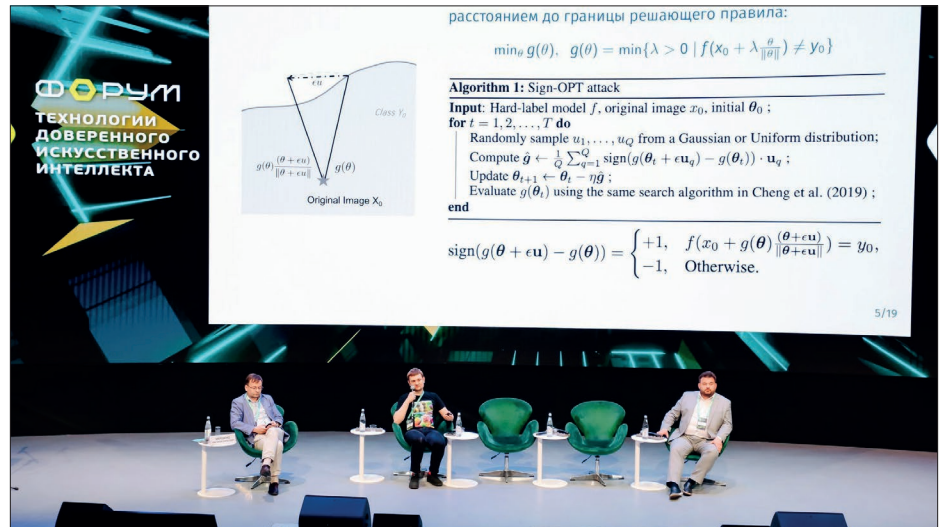
Также ИИ внедрен в сервисы продвижения платных объявлений, оценки адекватной цены товара по фото и данным компании «Авито». Учитываются различные критерии, проводится анализ сигналов и оценка, какие объявления привлекут внимание клиентов с учетом поисковых запросов.

В АПТЕКЕ

С помощью чат-бота продвигают продукты в аптечном маркетплейсе «Ютека». Компания использует генеративный ИИ для подготовки

Спикер напомнил историю развития вопроса и задачи, которые уже решены при участии сотрудников ИСП РАН и его партнёров из НКЦКИ, «Лаборатории Касперского», «Иннополиса», МИФИ, МФТИ, СПбГУ и Сколково. Также остановился на задачах, которые решаются в рамках второй волны искусственного интеллекта. Специалисты начинают приземлять базовые технологии и вести отраслевые исследования в области медицины, логистики, промышленного интернета вещей. При этом спикер отметил заметное отставание страны в микроэлектронике и призвал стремиться к тому, чтобы создавать отечественную базу и аппаратуру для внедрения новых моделей. А.Аветисян считает, что, несмотря на все трудности, необходимо продолжать системно развивать данное направление.

Учёный отметил, что разработки мирового уровня доверенного ИИ и сопутствующих технологий ведутся в России на базе лучших научных



и математических школ, к ним должны подключаться технологические партнёры и бизнес. Только совместным трудом отрасль может решить задачи и правильно выстроить процесс, считает А.Аветисян.

В ходе дальнейших мероприятий форума его участники рассмотрели научно-технологические осно-

вы доверенного ИИ, практические вопросы безопасности данных и систем ИИ, угрозы и атаки на системы искусственного интеллекта и методы их защиты. Многие доклады стали продолжением работ, о которых говорилось на I форуме «Технологии доверенного искусственного интеллекта».

текстов объявлений и полных отчетов для фармкомпаний-рекламодателей. Еще один инструмент, разрабатываемый с партнерами, сканер кожи по фото для оценки заболеваний и выдачи рекомендаций, а также проведение анализа возраста кожи, которым активно пользуются косметические бренды. Скрипты и кейсы в этом случае прописываются под конкретный заказчика и его продуктовую линейку.

В КОЛЛ-ЦЕНТРЕ

«Самолет финтех» внедряет ИИ на первой линии колл-центра. Это дает снижение стоимости звонка клиента на 40%, сокращение времени ожидания на 5 минут.

НА СКЛАДЕ

Интересный кейс использования ИИ при автоматизации рутинных складских операций представила компания «Автомакон». Одновременная работа беспилотных систем и людей потребовала переработки AGV-роботов для перемещения нестан-

дартных поддонов, их неравномерной загрузки и упаковки грузов. Итогом стала реализация проекта на складе, где ежедневно перемещается порядка 70% грузов напольного хранения и работают 700 человек. Проект уже привлек внимание заказчиков, увидевших в нем перспективы для развития.

АНТИФРОД

Компании также активно используют ИИ для борьбы с фродом. Тут применяется классическое машинное обучение, необходимы сложная структура нейросетей и моделей обучения, учет объективных сигналов по взаимодействию с реальными пользователями.

БЕЗ ЧЕЛОВЕКА НИКУДА

Несмотря на активное применение ИИ, во всех сервисах по-прежнему работают люди, которых невозможно заменить, отметили все эксперты. Только сочетание креативности человека и возможностей ИИ в логике систем поддержки принятия реше-

ний, анализа больших данных дает отличные результаты.

О ГАЛЛЮЦИНАЦИЯХ

Говоря о качестве получаемых результатов, все эксперты обратили внимание на такие проблемы ИИ как галлюцинации — попытки выдавать желаемый результат за действительность. Для борьбы с этим явлением обращаются к разработчикам, поскольку качество результатов от действий ИИ зависит от исходных обучающих датасетов. Проводится дообучение ИИ под конкретные задачи, создаются экспертные системы. Сети тестируются, а персонал учит правильно составлять промпты.

КРИТЕРИЙ ИСТИНЫ — ПРАКТИКА!

Критерии успешности решений с ИИ — внедрение, считают все эксперты. Однако для перехода от пилота к реализации компаниям необходимы серьезные вложения во всех направлениях и достаточный уровень зрелости.

POSITIVE HACK DAYS 2024

230 ТЫСЯЧ ГОСТЕЙ И ОНЛАЙН-ЗРИТЕЛЕЙ

Марина БРОДСКАЯ
обозреватель BIS Journal

Киберфестиваль Positive Hack Days 2024 стал самым масштабным за всю историю. За четыре дня его посетили более 100 тыс. гостей, а за онлайн-трансляциями основных сессий следили более 130 тыс. зрителей.

Огромные пространства «Лужников» вместили всех желающих. Организаторы учли опыт прошлого года, когда на набережной Парка культуры и отдыха им. Горького было не протолкнуться, а шатры, где выступали спикеры, проходили соревнования и мастер-классы, явно не были рассчитаны на такой наплыв зрителей. Главной площадкой PHDays-2024 стала большая спортивная арена «Лужников», а мероприятия, активности и концерты проходили по всей территории спортивного комплекса.

ГОСУДАРСТВО

О своём видении безопасного будущего спорили участники дискуссии

«Безопасное кибербудущее Человека — Государства — Мира. 100 лет тому вперёд». Модератор сессии, сооснователь компании Positive Technologies и проекта «Сайберус» Юрий Максимов предложил оттолкнуться от настоящего и спросил участников их мнение о происходящем.

Глава Минцифры России Максуд Шадаев отметил, что сегодня ведомство развивает различные государственные цифровые сервисы и их безопасность. «Вопросы кибербезопасности также вошли в нашу жизнь: чем больше сервисов и технологий, тем выше риск атак. Мы обязаны решить эту проблему, чтобы не тормозить дальнейшую цифровизацию».

О неизбежности перехода в цифровой мир и актуализации проблем кибербезопасности рассказал основатель АФК «Система» Владимир Евтушенков. Он отметил, что его мнение об отсталости отечественных технологий за последнее время изменилось. Спикер «уверен, что мы не только не отстали, но и можем быть лидерами».

Активный промоутер российского кибербезопасности за пределами страны, как его представил Ю. Максимов, акционер «Абрау-Дюрсо» Борис Титов срав-

нил современных хакеров с бандитами времён 1990-х и отметил, что в условиях новых рисков возникла необходимость построения защиты страны в цифровом пространстве.

Мир находится в уникальной ситуации, считает генеральный директор Positive Technologies Денис Баранов. «Правительства уже осознают всю мощь кибероружия, но ещё плохо понимают все детали», — подчеркнул он. По мнению спикера, главная задача его поколения и тех, кто младше, — «довести состояние ИТ-систем до такого уровня, чтобы можно было предотвратить глобальный ущерб». В этой области он видит вызов, достойный лучших инженеров.

Обеспечение цифровой безопасности считает критически важной задачей для всех отраслей председатель Совета федеральной территории «Сириус», руководитель фонда «Талант и успех» Елена Шмелёва. Она уверена, что подготовку специалистов надо вести «со школьной скамьи, а для этого необходимо, чтобы преподаватели уже сейчас имели базовые компетенции, а в основе образовательных программ лежал опыт реальных специалистов». Значимым конкурентным преимуществом отечественной школы спикер считает традиции в области преподавания математики и естественных наук, которые необходимо сохранять и развивать.

Оценивая будущее, М. Шадаев отметил, что вопрос зрелости технологий и их восприятия будет определяться по тому, насколько быстро начнётся внедрение технологий беспилотного вождения. Уже через пять-семь лет беспилотные автомобили появятся на дорогах, считает министр, приводя в пример активное внедрение в жизнь беспилотных робокурьеров «Яндекса». «Я как человек, у которого нет прав, просто жду этого мо-



Фото: Positive Hack Days 2024



Фото: Positive Hack Days 2024

мента. Думаю, что это совсем близко, не за горами. Я не испытываю страха, уже несколько раз ездил на беспилотной машине, это очень круто», — заявил М. Шадаев.

В. Евтушенков полагает, что в скором будущем кибербезопасность станет неотъемлемой частью всех сфер. Он также призвал развивать перспективные технологии, например квантовые вычисления или фотонику, и порадовался, что принятый ранее подход «купим всё» перестал работать в 2022 г. Ситуация в сфере инноваций меняется очень быстро, считает глава АФК «Система».

По мнению Д. Баранова, стоит ожидать прорыва в области создания высокоэффективных систем, которые смогут предотвратить любой неприемлемый ущерб. Однако главной задачей сферы ИБ является борьба со страхами, которые стоят на пути прогресса, считает глава РТ. Б. Титов назвал недостаток вычислительных и энергетических мощностей сдерживающим фактором кибербезопасности. Однако, по его мнению, в ближайшие 20 лет появятся новые источники энергии и высокопроизводительные вычислители на новых принципах.

По мнению участников дискуссии, искусственный интеллект и системы машинного обучения приведут к исчезновению ряда профессий и появлению новых бизнес-моделей. Однако в целом необходима защита приватной стороны жизни человека и его мыслей от тотального проникновения ИИ.

РЕГУЛЯТОРИКА

Дискуссия о том, что такое OSINT (англ. Open-Source Intelligence, метод сбора информации из открытых источников), была проведена по инициативе государственных органов с целью выяснения позиции отрасли, рассказал модератор сессии Алексей Лукацкий (Positive Technologies) по итогам обсуждения. А в ходе дискуссии каждый из участников обсуждения по-своему обозначил и понимание темы, и отношение к её регулированию.

Андрей Масалович («Инфорус») отметил, что OSINT находится в серой зоне, где каждое действие незаконно и его можно оспаривать. Организационно он разделил OSINT на три уровня: нижнюю, типовую контрразведку в целях экономической безопасности, контроль конку-

рентов и секторов рынка на среднем, и стратегический уровень, на котором ведётся захват рынков.

Дмитрий Борошук (BeholderIsHere Consulting) вспомнил про массовые утечки или непродуманное раскрытие информации пользователей о себе в соцсетях, которые дают огромный слой сведений о гражданине и его интересах. Ещё один аспект — IoT, который слабо защищён в настоящее время.

Разную мотивацию компаний, которые начали заниматься OSINT, подчеркнул Владимир Бенгин («Солар»). Это и экономическая безопасность, и киберразведка самих ибэшников — «чисто техническая история», в ней нет юридических проблем, если есть договоры и получено согласие заказчиков. В. Бенгин напомнил ещё один раздел OSINT — защиту бренда компании и её партнёров от внешнего воздействия хакеров, которые от имени «чужого бренда» ведут атаку через фишинговые сайты или иначе воздействуют на клиентов организации, нанося как минимум ущерб её репутации. При этом у самой компании, от имени которой действуют злоумышленники, проблем с безопасностью нет.

Важным источником информации для OSINT являются люди, невольно раскрывающие важные сведения о себе или компаниях в соцсетях и мессенджерах. Утечки ПДн — проблема, которая находится в рамках закона, если исследования проводятся в рамках подписанного согласия. Но это серая зона, поскольку чаще всего такого согласия на исследование жизни человека нет, считает В. Бенгин. Компания может быть защищена со всех сторон, однако нельзя исключать человеческий фактор и вероятность воздействия на сотрудников, которые могут стать объектами шантажа со стороны злоумышленников, напомнил Вадим Соловьёв (Positive Technologies).

По мнению Виталия Лютикова (ФСТЭК России), тактика разведки активно используется злоумышленниками и необходимо использовать такие же методы защитникам, чтобы закрыть имеющиеся уязвимости. Другой вопрос, насколько эти действия соответствуют закону. Если специалист по ИБ становится обладателем чувствительной информации в результате нормализации, обработки и анализа данных, он должен выполнить требования, которые налагает законодательство.

По мнению представителя регулятора, должна появиться предметная область OSINT. Методика становится новой и растущей областью ИБ, дополняется инструментами ИИ и LLM, поэтому необходим системный подход к решению вопросов OSINT, повышение грамотности людей, расширение правоприменения не только для тех, кто собирает данные, но и для тех, кто допускает утечки.

Отвечая на вопрос о необходимости лицензирования деятельности, замглавы ФСТЭК России отметил, что в ближайших планах его нет.

ВСЁ О CISO

Традиционный кадровый вопрос рассматривался на многих площадках PHDays 2024 и в разрезе дефицита специалистов, и в рамках взаимодействия компаний и вузов. Отдельная секция была посвящена работе руководителей по информационной безопасности: карьерному росту, роли

CISO в организации и его жизненному циклу в компании.

По данным Минцифры, среднее время поиска руководителя по ИБ занимает полгода, а средний возраст CISO вырос за последние годы с 41 до 45 лет. Компании не спешат нанимать молодых руководителей, отметил Владислав Иванов (H&N), который поделился собственным опытом восхождения по карьерной лестнице безопасника и управления отделом ИБ компании, взаимодействия с руководством, коллегами и партнёрами.

Роль и цели главного безопасника в организации обсудили участники другой сессии. Они рассмотрели две противоположные точки зрения на результативность работы CISO: тех, кто считает основной мерой успеха рост бизнес-показателей компании, и тех, кто полагает, что основная задача CISO — классический подход к обеспечению ИБ, выполнение требований регуляторов и снижение числа инцидентов. Несмотря на разные точки зрения, участники сессии отметили, что CISO должен уметь говорить с топ-менеджментом, предлагать ему разные инструменты и подходы. Опыт — лучшая школа для подготовки CISO, отметили эксперты. Под конкретные задачи бизнеса с учётом специфики компании нужен свой руководитель по ИБ. И главные потери от инцидентов — не финансовые, а репутационные, которые пока невозможно посчитать, полагает эксперт Алексей Волков.

Тема развития и карьерного роста специалиста по ИБ обсуждалась отдельно. Важными качествами соискателей на должность CISO эксперты считают хорошие навыки коммуникации, мотивация к самообразованию, аналитическое мышление и умение быстро принимать решение, считает Лев Шумский (Yandex Cloud).

Этапам жизненного цикла руководителя ИБ в организации — от желания сменить работу и начала переговоров о приёме в другую компанию до ухода по различным обстоятельствам — был посвящён доклад А. Волкова. Он остановился на задачах и проблемах CISO: проблемах адаптации на новом месте, получении результатов, психологических аспектах, во-

просах самореализации и выгорания. Он также дал советы участникам мероприятия, как правильно построить свою стратегию работы, чтобы будущее перестало быть неопределённым.

О том, есть ли жизнь после инцидента, какова в ней роль CISO, как предотвратить негативное влияние киберугроз на компанию и бизнес-показатели, поговорил с участниками сессии Владимир Ташкеев (Infosecurity).

STANDOFF 13

Основной этап кибербитвы, в которой белые хакеры представляли два виртуальных государства F и S, проходил в стороне от основных мероприятий. За три дня участники реализовали 415 критических событий, в том числе 77 уникальных, и обнаружили 617 уязвимостей. Успешные атаки были проведены во всех сегментах государств. Больше всего пострадала финансовая отрасль, где были реализованы 23 критических события, меньше всего досталось атомной промышленности (два уникальных критических события). По итогам основного этапа битвы среди всех активных команд с небольшим перевесом победило Государство F.

Команды защитников предотвратили 115 критических событий в атомной отрасли и энергетике, расследовали 182 успешные атаки и обнаружили 851 инцидент. В заданиях для кибербитвы были учтены 76 % техник матрицы MITRE ATT&CK, из которых атакующие использовали 67 %, а защитники выявили 66 %.

В финал — плей-офф и плей-ин — попали четыре лучшие команды из каждого государства. В ходе плей-офф ни одной команде не удалось реализовать недопустимые для Positive Technologies события, победитель определился по итогам плей-ин. Standoff 13 выиграла команда DreamTeam, она и получила главный приз — 800 тыс. рублей. Следом расположились команды Cult, DeteAct x SPbCTF и GISCYBERTEAM.

В процессе плей-ин атакующим также не удалось реализовать недопустимые для Innostage события. Специальным призом за хорошую попытку атаки была награждена команда Wetoх.

«SOC С МАШИНКАМИ В ПЕСОЧНИЦЕ И БЕЗ»

250 ТЫС. РУБЛЕЙ — ЗА ИМЯ ДЛЯ SOC
«ГАЗИНФОРМСЕРВИС»!

Анна КАТАНКИНА
корреспондент BIS Journal

Компания «Газинформсервис» объявила о выходе на рынок коммерческих центров мониторинга и реагирования. Об этом на пресс-конференции «SOC с машинками в песочнице и без» сообщил Николай Нашивочников, заместитель генерального директора — технический директор «Газинформсервиса».

Николай Нашивочников рассказал, что идея создать коммерческий SOC давно витала в стенах компании, которая имеет двадцатилетний опыт работы на рынке информационной безопасности. События 2022 г. и запрос со стороны заказчиков в условиях дефицита кадров и финансовых издержек развеяли сомнения, стало очевидно, что услуга внешнего SOC будет востребована на рынке.

Работы по созданию Центра мониторинга начались в мае прошлого года. Вопреки первоначальному плану уложиться в полгода они продлились год. При разработке Центра компания использовала свой многолетний опыт, большой массив накопленного экспертного опыта, разработки центра компетенций и RnD-команды. Сегодня «Газинформсервис» готов представить потребителям SOC, который оказывает пять услуг: по инвентаризации активов, управлению уязвимостями, мониторингу и реагированию на инциденты. Работа ведётся с использованием продуктов компании Ankey SIEM NG, и Ankey ASAP — платформы аналитики данных для обнаружения атак и инсайдеров, так

же применяются решения класса IRP/SOAR, UEBA, SIEM, PAM, Asset Management, Vulnerability Management, Hardening and Compliance, EDR. В перспективе или на стадии глубокой проработки находятся решения класса DevSecOps и MLSecOps, PSV, BAS, TI, тестирование на проникновение силами собственной и сторонних Red Team. Николай Нашивочников подробно рассказал об услугах SOC и перспективах его развития.

«Повышение эффективности и скорости реагирования поставлено во главу угла», — сказал спикер. В процессе выбора решений для SOC компания отказалась от стандартных коробочных SIEM, тестировала и заменила несколько продуктов, в том числе дорабатывала и донастраивала собственный SIEM. «Этот путь мы прошли, и он оказался для нас запоминающимся», — отметил технический директор «Газинформсервиса» и допустил, что все решения, на текущий момент включённые в экосистему SOC, могут быть видоизменены или вовсе заменены в будущем. Компания делает ставку на надёжность и масштабируемость, которая заложена как парадигма в основу центра мониторинга, и на поведенческую аналитику, предоставляемую платформой Ankey ASAP.

«Мы прекрасно понимаем, что запуск Центра мониторинга — это только начало пути, и видим несколько дорожных карт развития услуги. Одна из этих дорожных карт ограничена текущим годом. Мы продолжим развивать вторую и третью линии SOC, обязательно проведём несколько тестирований с точки зрения защищённости инфраструктуры Центра мониторинга силами команды Red Team GIS», — рассказал Н. Нашивочников о ближайших планах компании.

В феврале 2024 г. «Газинформсервис» создал Аналитический центр кибербезопасности (АЦКБ), который недавно выпустил первый отчёт. «Это важный шаг в истории компании: благодаря SOC появились свои датасеты, и теперь лаборатории «Газинформсервиса» есть с чем работать в настоящем и будущем», — отметил Н. Нашивочников. В состав АЦКБ входят три лаборатории «Газинформсервиса»: лаборатория технологий искусственного интеллекта, лаборатория исследований кибербезопасности, развивающая GIS-киберполигон и пополняющая платформу знаний в области ИБ с упором на расследование киберинцидентов, и лаборатория стратегического развития продуктов безопасности. Последняя обеспечивает внедрение инновационных решений на базе технологий ИИ в продукты компании.

«Киберполигон был создан для работы с доверенными датасетами, эмуляцией действий в инфраструктуре, оттачиванием детектирования и т.п. В конце года на базе киберполигона компания планирует проводить учения. Я думаю, что киберполигон частично будет открыт к концу года», — считает Н. Нашивочников.

Рассказывая о дорожной карте развития SOC, представитель компании отметил, что в августе текущего года планируется расширение команды центра за счёт аналитиков L2 и аналитиков угроз и начала регулярного тестирования SOC силами собственной Red Team. В начале осени, согласно дорожной карте 2024 г., будет введена услуга «Киберразведка», которая позволит как определять ландшафт угроз, так и проводить ретроспективный анализ. Также в планах «Газинформсервиса» переложить задачи обслуживания инфраструктуры Центра мониторинга и реа-



гирования на плечи команды SOC и подключить системы к киберполигону АЦКБ. В октябре компания GIS намерена расширить команду аналитиков L3 за счёт специалистов по форензике, углублённому расследованию и устранению последствий инцидентов. Тогда же инфраструктура SOC будет протестирована сторонней Red Team.

Форензика — одно из активно развивающихся направлений компании, отметил техдиректор. Собрана команда, которая прошла обучение, в этом «Газинформсервис» активно помогают партнёры. Новые возможности и новые инструменты уже позволили выявить и расследовать реальный инцидент в инфраструктуре самой компании.

В полную силу SOC заработает в конце текущего года, когда будут запущены услуги третьего уровня, включая киберразведку и Threat Intelligence. Сейчас компания активно подключает фиды и тренируется, отметил спикер.

Рассказывая о том, каким образом внешний SOC от «Газинформсервис» поможет снять часть нагрузки с ИБ-подразделений заказчиков, Н.Нашивочников подчеркнул, что при подключении услуги компания проводит аудит инфраструктуры

предприятия на соответствие базовому уровню кибербезопасности. В дальнейшем ведётся подключение инструментария как на стороне потребителя, так и со стороны исполнителя. В процессе мониторинга и реагирования, который осуществляется в режиме 24/7, снимается нагрузка с сотрудников заказчика, частично разрешаются кадровые вопросы.

Говоря о перспективах спроса на услуги стороннего SOC и преимущества перед конкурентами, Н.Нашивочников отметил, что «Газинформсервис» — один из ведущих разработчиков продуктов ИБ и интеграторов в России, и он делает ставку на собственную качественную экспертизу и на автоматизацию процессов, активное использование методов машинного обучения и искусственного интеллекта, которые кратно повышают эффективность реагирования и скорость детектирования. SOC от «Газинформсервис» — это уникальный сервис, адаптированный под потребности заказчика с учётом актуальных векторов атак. Предлагаемые мультивендорные комплексные решения с применением собственного ПО от «Газинформсервиса» обеспечивают гибкость и масштабируемость оказываемых услуг и не потребуют

от заказчика дополнительных расходов на закупку программного и аппаратного обеспечения.

Оценивая перспективы внедрения SOC на промышленных предприятиях и объектах критической информационной инфраструктуры, представитель «Газинформсервиса» отметил повышение спроса на такие услуги в сегментах ИТ-инфраструктуры. Однако в случае с КИИ стороны ограничены действием Федеральным законом № 187-ФЗ «О безопасности критической информационной инфраструктуры РФ» и подзаконными актами. «Газинформсервис» имеет лицензию ФСТЭК России на оказание услуг по мониторингу.

«Сегодня мы имеем огромную экспертизу, которой готовы делиться в рамках услуг для одного из основных заказчиков», — отметил спикер.

И главное! Компания объявила конкурс на выбор собственного названия для SOC: www.namesoc.ru. Сбор заявок проводится до 10 июля 2024 г., до 10 августа будут отобраны 20 лучших предложений, а 10 сентября — назван победитель. Он получит денежный приз в размере 250 тыс. рублей.



UserGate V

«МЫ ПОДНЯЛИ ПЛАНКУ АМБИЦИЙ»

Анна КАТАНКИНА
корреспондент BIS Journal

Компания UserGate, ведущий российский разработчик решений в области информационной безопасности, провела пятую ежегодную конференцию, на которой представила новую продуктовую линейку и поделилась планами развития на фоне изменения ландшафта киберугроз. В состав продуктового портфеля UserGate теперь входят не только решения, обеспечивающие безопасность сети и конечных устройств, средства обнаружения вторжений, но и услуги аудита и консалтинга, анализа киберинцидентов, а также центр мониторинга информационной безопасности (SOC), заявили представители компании.

«МЫ НЕ ПРОСТО МЕЧТАЕМ»

«Прошедший год был интересным и сложным. Мы подняли планку амбиций, — отметил Михаил Пеньковский,

вице-президент по продажам и маркетингу UserGate. — Мы запускаем новые продукты, новые услуги, консалтинг, образовательные курсы — мы видим себя провайдером безопасности и защиты цифровой природы. Здесь наши амбиции становятся безграничными, и мы не просто можем мечтать об этом, мы уже это делаем».

М. Пеньковский выделил три фундаментальные направления RND, которыми занимается компания: работа над продуктами в текущих итерациях на ближайший год, команда, которая ведёт исправление ошибок, и перспективные разработки. Также работают мобильные команды agile, которые позволяют гибко масштабировать готовые решения. Ведётся адаптация продуктов UserGate для работы в облачной инфраструктуре, идёт активное взаимодействие с ведущими облачными провайдерами. Предусматривается адаптация инструментов UserGate для работы в частных облаках, которые активно развивают заказчики. «Это огромная точка роста», — подчеркнул спикер, отметив высокий уровень российской научной и инженерной школы.

ЛАБОРАТОРИЯ ОСОБОГО НАЗНАЧЕНИЯ

В рамках реализации стратегии долгосрочного развития в прошлом году создана Технологическая лаборатория UserGate. Она занимается только инновационными технологиями. Специалисты лаборатории работают в сотрудничестве с крупными заказчиками и ведут изыскания и разработки на перспективу трёх-пяти лет. Подразделение объединило людей с особым складом ума, которым нравится работа по разработке новых технологий и функционала, отметил М. Пеньковский.

«Лаборатория полностью оторвана от технологического процесса, который есть в компании. Мы инвестируем в проекты, которых сегодня ещё нет в России и которые появятся в перспективе трёх-пяти лет. Возможно, в следующем году мы покажем что-то в виде прототипов», — заявил директор-сооснователь UserGate Александр Левченко.

В ПЕРСПЕКТИВЕ — ПРОЦЕССОРЫ

На конференции А. Левченко традиционно представил новые программ-

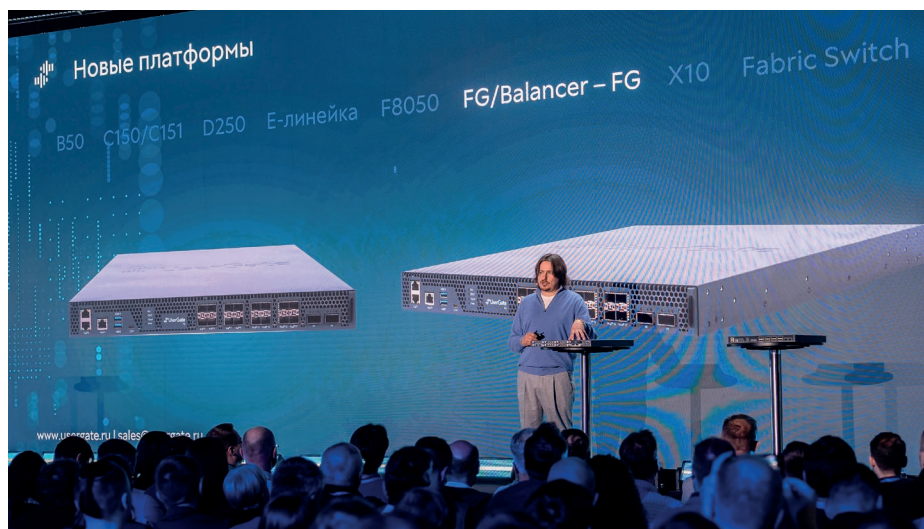
но-аппаратные платформы, которые были созданы в течение года и охватывают все сегменты бизнеса — от МСП и брэнч-офисов до крупных компаний, дата-центров и облаков, образуя экосистему решений. Программно-аппаратные комплексы разработаны компанией UserGate и реализованы новым отечественным партнёром. По словам А. Левченко, новые ПАК затмят решения старого поколения. Он подчеркнул, что все разработки аппаратных платформ — проектирование устройств, промышленный дизайн — ведутся конструкторским бюро компании. Сегодня UserGate всё ещё зависит от других поставщиков микроэлектроники, но, возможно, когда-нибудь станет производить и процессоры. «Главная задача сегодня — обеспечить доверие к продукции, которая разработана и собрана в России. Мы со своей стороны полностью контролируем все технологии и компоненты, которые входят в состав наших решений», — отметил директор-сооснователь UserGate.

О том, как разрабатывался промышленный дизайн новых платформ с учётом эргономики, функционала, экономики и экологии, рассказала участницам конференции промышленный дизайнер Виктория Хромова.

NGFW ЗРЕЕТ НЕРАВНОМЕРНО

Выделяя значимые для компании события последних двух лет и востребованность у потребителей отечественных межсетевых экранов уровня сети (NGFW), Иван Чернов, менеджер по развитию UserGate, отметил, что не все заказчики учли возможность ухода западных вендоров при построении риск-моделей. Поэтому в критический момент на российском рынке возникла паника и массовый спрос на NGFW. Сейчас межсетевые экраны, которые массово стали разрабатываться отечественными вендорами, не самое слабое место в кибербезопасности, считает И. Чернов. До определённого момента сегмент оборудования для ИБ развивался последовательно, и невозможно одновременно перейти на отечественные решения.

«Западные решения есть, сохраняется серая зона поставки. Задача



Александр Левченко, директор-сооснователь UserGate

ответственного вендора — создать дорожную карту и синхронизироваться с заказчиками», — подчеркнул И. Чернов, отметив, что работы по импортозамещению много и часто связаны с развитием ИТ.

Сегодня UserGate NGFW — один из самых функциональных в стране, отметил И. Чернов. Компания представляет самую широкую инсталляционную базу и потому является лидером по продажам по итогам 2022–2023 годов, значительно оторвавшись от конкурентов. При этом компания активно собирает обратную связь и пополняет бэклог разработки.

И. Чернов подчеркнул, что функционал NGFW зреет неравномерно и заказчики оказывают влияние на этот процесс. Представленная осенью 2023 г. новая версия на базе ОС UGOS версий 7.0 и 7.1 работает стабильнее и качественнее во многом благодаря общению с потребителями и налаженной обратной связи. «Сегмент СМП менее требовательный к функционалу, в сегменте Enterprise за два года в компании реализовали более пяти тысяч проектов, объём работы колоссальный и помогает развиваться», — подчеркнул И. Чернов. Большой пул заказчиков, на которых тестируются пилотные версии, позволяет повысить уровень релизов. Поэтому устойчивый рост компании в перспективе 2024–2025 гг. и планируемый рост выручки — совместная работа ком-

пании и заказчиков. По всем портфелям продуктов будут размещать новые решения и диверсифицировать источники. Планируется, что российский рынок ИБ к 2026 г. вырастет вдвое, при этом более трети его объёма будет приходиться на ИБ-услуги.

В 2024 г. UserGate делает ставку на стабильность и производительность своих межсетевых экранов. NGFW 7.1 — самый стабильный релиз, утверждает Александр Кистанов, директор по продукту UserGate. По его словам, компания тратит 70 % рабочего времени на покрытие технического долга: закрывает баги, проводит тестирование решений, добавляет новые фичи и улучшения. Работа ускорена за счёт расширения команды разработчиков.

КОНСАЛТИНГ ВЕРНЁТСЯ СТОРИЦЕЙ

Дмитрий Кузеванов, CISO, руководитель центра мониторинга и реагирования компании, напомнил участникам конференции, что количество инцидентов безопасности в стране и мире растёт, увеличивается количество критических уязвимостей, в том числе в иностранных продуктах, которые остались без поддержки со стороны западных вендоров и без возможности обновления программного обеспечения. На фоне массовой цифровизации бизнеса и расширения использования ИТ-инструментов в бизнес-процессах российский рынок испытывает де-



Михаил Пеньковский, вице-президент по продажам и маркетингу UserGate

фицит квалифицированных кадров. Особенно это заметно в сфере МСП, где конкурентоспособность зависит от возможности использовать ИТ-инструменты, но вопросам информационной безопасности уделяется мало внимания. Квалификация ИТ-специалистов не успевает за сменой парадигмы в освоении новых инструментов, возникает пробел в уровне осведомлённости. Вопросы безопасности малого бизнеса необходимо рассматривать не только с точки зрения кибербезопасности, но и с позиции социальной ответственности, это будет зрелым подходом предпринимателей к вопросам ИБ, считают в UserGate. Компания предлагает системный подход и готова предоставить консалтинг и обучение своим партнёрам и заказчикам. «UserGate не планирует держать внутри свои достижения, а готов передавать весь свой накопленный опыт клиентам и партнёрам», — подчеркнул спикер. Компания ставит цель через три года получать 20 % выручки за счёт услуг.

SOC-AS-A-SERVICE

Компания UserGate значительно расширила деятельность центра мониторинга и реагирования (SOC), включив в его состав новые направления и компетенции: полноценные услуги аудита и консалтинга в области ИБ, экспертное содействие в проведении анализа киберинцидентов. Уровень компетенции специалистов SOC, который построен на разработ-

ках компании, позволяет вывести его на уровень коммерческой эксплуатации и предоставлять полноценный сервис по облачной модели SOC-as-a-Service.

Подробнее о SOC-as-a-Service рассказал лидер направления Дмитрий Шулинин, имеющий большой опыт построения центров мониторинга и реагирования. Спикер отметил, что SOC UserGate уже готов принимать клиентов, которые осознали необходимость такого центра, но пока не имеют возможности построения собственного inhouse SOC. Им предлагают услуги аналитики и сбора телеметрии, мониторинга и реагирования на инциденты в режиме 24/7 365 дней в году. Такая услуга позволит заказчикам получить доступ к лучшим практикам ИБ и снизить нагрузку на внутренние команды компаний. Такое решение дешевле построения собственного SOC, отметил докладчик. Также он рассказал о перспективах развития направления до конца 2025 г.

СУВЕРЕННОСТЬ И НЕЗАВИСИМОСТЬ

Руководство UserGate не прогнозирует стагнацию бизнеса в 2024–2025 годах в рамках реализации Указов президента России по импортозамещению и переходу российских компаний на отечественные решения к 1 января 2025 г. Процесс миграции на российские решения не так прост, как кажется со стороны, в силу сложности инфраструктур, которые нуж-

но защищать. Многие иностранные решения закупались на длительную перспективу, и перенос таких решений на отечественные платформы — процесс длительный и дорогостоящий. Поэтому UserGate ожидает дальнейшего устойчивого роста числа проектов и объёма доходов. Это заслуга не только сотрудников компании, но и заказчиков, их запросов.

«Суверенность и независимость — императив существования бизнеса, — считает М. Пеньковский. — Отечественные вендоры уже начали производить свои решения. Производителям надо выйти из зоны комфорта и объединиться, поскольку все работают над одной задачей. Необходима дорожная карта и обсуждение заинтересованных сторон, как плавно и без напряжённости перейти на отечественные продукты». Он отметил, что в прошлом году в некоторых сегментах бизнеса, включая большие проекты, рост зрелости решений выроскратно.

«По сути, мы соревнуемся с силой, которая в 50 раз нас превосходит, — считает М. Пеньковский, поскольку западный рынок по объёму в 50 раз превышает российский. — Мы выросликратно не только в денежном выражении, но и в плане опыта, в том числе негативного. Важно, что заказчики проявили терпение и понимание».

ВЫХОД ЗА РУБЕЖ

Отвечая на вопрос о перспективах развития бизнеса и возможностях выхода на иностранные рынки, М. Пеньковский рассказал о готовящемся запуске полноценного подразделения в Минске (Республика Беларусь), которое будет включать не только офис продаж, но и развитие центра разработки и компетенций. Есть планы по выходу на другие дружественные рынки. За пределами этой зоны необходимо учитывать геополитические и культурные аспекты, требования местных регуляторов, особенности каналов сбыта, считает спикер. Для того чтобы сделать это эффективно, необходима стратегия, а также содействие государства в части лоббирования интересов российских компаний и нивелирования методов борьбы конкурентов вне рамок технологического соперничества.

ПРОСТРАНСТВО ДОВЕРИЯ: ПРОМЫШЛЕННАЯ ЭКСПЛУАТАЦИЯ

О ДОКАЗУЕМОСТИ
ЭКВИВАЛЕНТНОСТИ
(ПО СУЩЕСТВУ)
УРОВНЕЙ НАДЁЖНОСТИ
ЭЛЕКТРОННЫХ
ПОДПИСЕЙ ПРИ ИХ
ТРАНСГРАНИЧНОМ
ВЗАИМНОМ ПРИЗНАНИИ



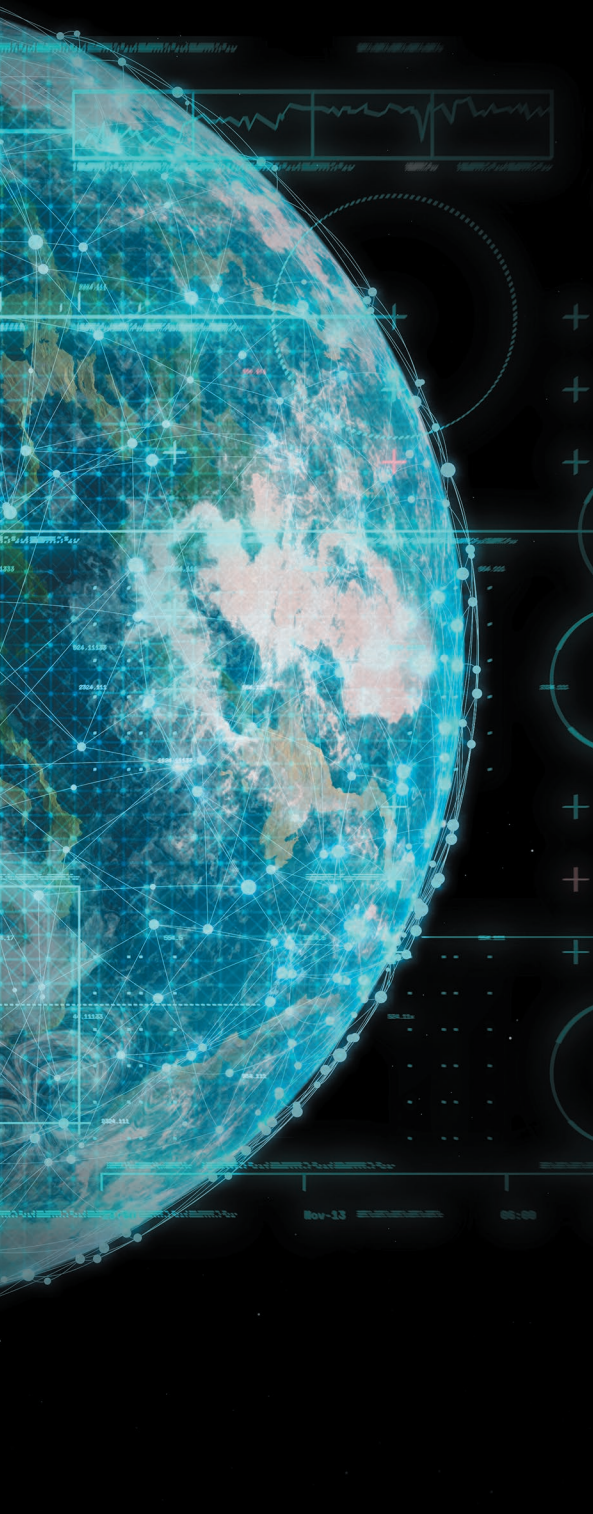
Сергей КИРЮШКИН

советник генерального директора — начальник удостоверяющего центра ООО «Газинформсервис», эксперт Ассоциации «РОСЭУ», эксперт сети экспертов ООН по безбумажной торговле и транспорту UNNEXT, заместитель председателя программного комитета международной конференции PKI-Форум, к. т. н.

Ключевым вопросом трансграничного взаимного признания электронной подписи является доказуемость эквивалентности (по существу) уровней надёжности электронных подписей, применяемых по обе стороны границы в рамках трансграничного электронного документооборота (ЭДО).

Компания «Газинформсервис», сотрудничая с ФНС России и ФТС России, выполняя требования ФСБ России и Минцифры России, являясь аккредитованным удостоверяющим центром и разработчиком линейки PKI-продуктов Litoria, совместно с отечественными операторами ЭДО и партнёрами из других стран занимается вопросом признания

иностранной электронной подписи с середины 2000-х годов. В конце 2023 года на двух направлениях (с Беларусью и Казахстаном) начата промышленная эксплуатация трансграничного пространства доверия в сегменте В 2В на основании соглашения между участниками электронного взаимодействия (63-ФЗ, статья 7, часть 4). Ещё по восьми направлениям



развёртывание трансграничного пространства доверия (ТПД) в сегменте В 2В находится на предварительных фазах (табл. 1).

На ряде международных площадок, в том числе ЭСКАТО ООН, СЕФАКТ ООН, ЕЭК, АТЭС, ОЧЭС и др., обсуждаются подходы к оцениванию уровней надёжности электронных подписей для их взаимного признания.

Фаза развития ТПД между РФ и	Описание фазы	Условная нумерация
Беларусью	Промышленная эксплуатация	1
Казахстаном	Промышленная эксплуатация	1
Китаем	Предпромышленная эксплуатация	2
Арменией	Пилотный проект	3
ЕС	Пилотные проекты	3
Азербайджаном	Тестирование	4
Республикой Корея	Тестирование	4
Узбекистаном	Подготовка к тестированию	5
Кыргызстаном	Переговоры	6
Индией	Переговоры	6

Таблица 1. ТПД уровня В 2В для российского бизнеса на июнь 2024

Визуально это можно представить в виде диаграммы Венна «зона ответственности двух команд» (рис. 1). Область взаимного доверия ЭП разных стран находится в зоне пересечения этих объёмных фигур, представленных на данной визуальной модели. Как видно, надёжность каждой подписи зависит от национального законодательства, особенностей национальной инфраструктуры открытых ключей и стойкости используемой криптографии. Сложности данной задачи добавляет то, что эти области со временем меняются (принимаются новые национальные стандарты, меняется национальное законодательство, меняется национальный домен доверия удостоверяющих центров, например, при аккредитации новых УЦ или формировании новых требований к квалифицированным сертификатам). Потому что каждая страна может модернизировать и менять криптографию, зако-

нодательство и организацию работы УЦ по своему усмотрению.

Доказуемость эквивалентности (по существу) уровней надёжности электронных подписей разных стран является краеугольным камнем международных договоров (соглашений) о взаимном признании электронной подписи, на каком бы уровне они ни заключались — между государствами (G2G) или между хозяйствующими субъектами (B 2B 2G). Общей отправной точкой для создания национальных систем электронной подписи является Типовой закон об электронных подписях UNCITRAL (Model Law on Electronic Signatures, MLES), принятый ООН в 2001 году.

Статья 12 MLES определяет принципы признания юридической силы иностранной электронной подписи. Таких принципов два:

1) юридическая сила электронной подписи не зависит от страны, в которой был подписан электронный

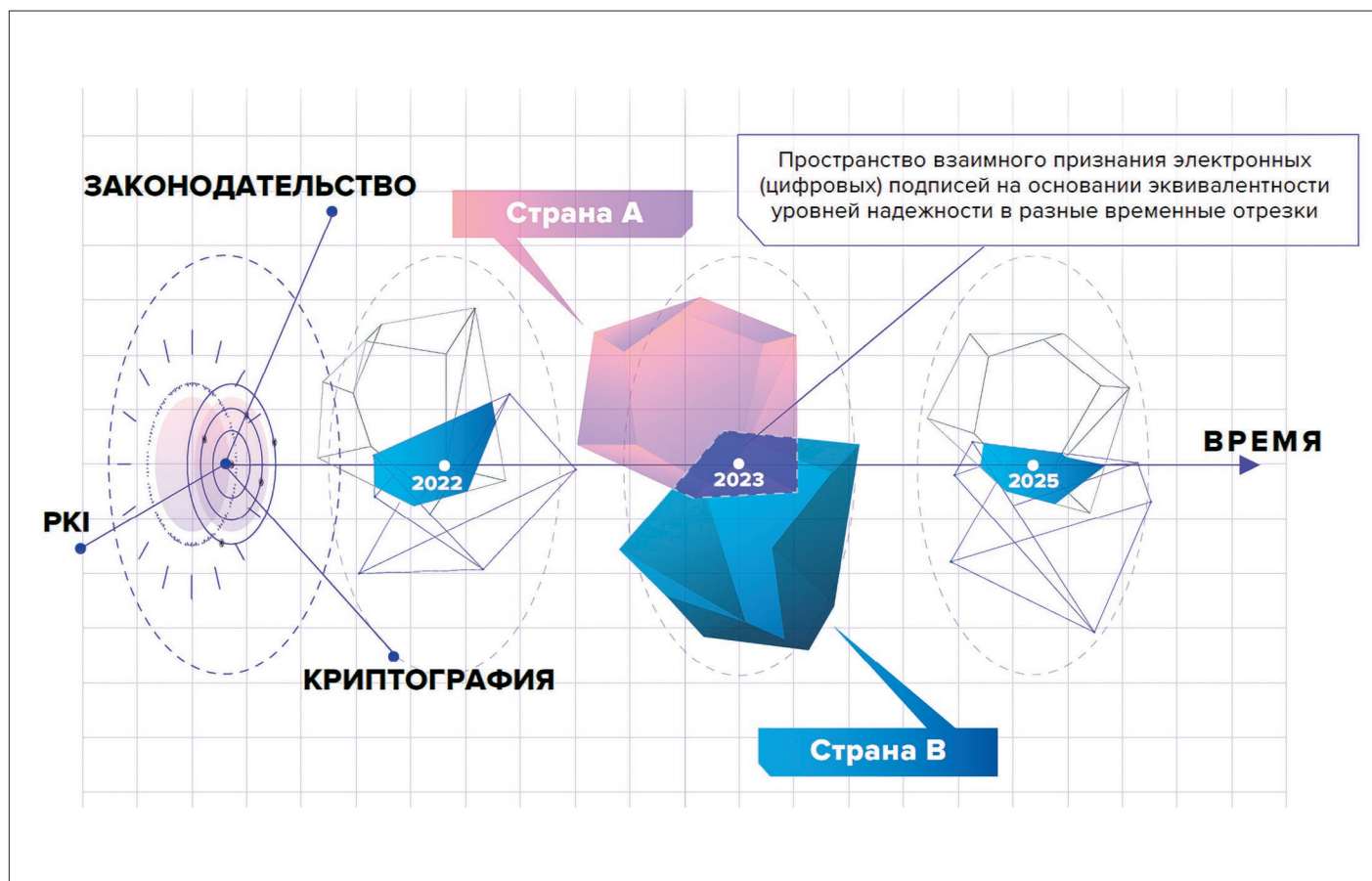


Рисунок 1. Модель области эквивалентности уровня надёжности электронных подписей разных стран

документ и где выдан сертификат электронной подписи;

2) условием признания юридической силы иностранной электронной подписи является эквивалентность (по существу) уровней надёжности электронных подписей стороны (страны) подписанта и доверяющей стороны (страны).

В руководстве по принятию MLES (п. 158) определено, что основные критерии, по которым следует оценивать уровень надёжности иностранной электронной подписи, изложены в статьях 6, 9 и 10 типового закона. Эти статьи посвящены:

Ст. 6 — тому, как может быть соблюдено требование в отношении наличия электронной подписи;

Ст. 9 — тому, что как должен организовывать свою работу удостоверяющий центр;

Ст. 10 — как можно при этом (ст. 6 и 9) оценить надёжность национальной инфраструктуры ЭП.

Ни одно из положений статей 6 и 9 не должно вызывать затруднений при потенциальном аудите по этим

канонам иностранных поставщиков сертификационных услуг (удостоверяющих центров) со стороны российской профильной экспертной группы. Так же понятна и может быть оценена российская организационная и технологическая инфраструктура ЭП со стороны наших иностранных партнёров. Можно быть уверенными, что она этим требованиям соответствует. Но вот в статье 10 на этот счёт есть пробел, который не позволяет на сегодня утверждать о полной готовности российской PKI к такой оценке. Об этом пробеле в регулировании отечественной PKI говорит п. е) статьи 10 MLES:

е) регулярность и объём аудита, проводимого независимым органом

Это очень важный фактор. Независимая система аудита национальной инфраструктуры открытых ключей является основой доказательства уровня надёжности иностранной электронной подписи, всех проце-

дур, решений, технологий и гарантий, связанных с ней.

Евразийский экономический союз реализовал трансграничную среду доверия электронных подписей в сегменте G2G. А поскольку это G2G, то системы независимых аудиторов там пока нет. Страны гарантируют свои уровни надёжности ЭП другим странам за счёт своего авторитета.

Другой известный пример — глобальная система доверия WebTrust для центров сертификации. Независимые аудиторы в ней оценивают соблюдение этих требований поставщиками услуг доверия (в том числе — центрами сертификации) по всему миру, кроме пространства СНГ, где нет ни одного аккредитованного независимого аудитора по системе WebTrust.

Отличным примером реализации данного подхода является трансграничная среда доверия иностранных электронных подписей Европейского Союза, основанная на общих требованиях европейского регламента eIDAS. Независимые аудиторы — представи-

тели аудиторских компаний, аккредитованных по EidAS, оценивают соблюдение этих требований.

В настоящее время идёт работа по выработке подхода к оцениванию эквивалентности (по существу) уровней надёжности электронных подписей в рамках рамочного соглашения об упрощении процедур безбумажной торговли на площадке ЭСКАТО ООН (СРТА), которое РФ ратифицировала в 2023 году. Кроме того, первые подходы к организации взаимных международных ау-

дитов национальных PKI сделаны в Статье 5 «Ознакомительные визиты» «Соглашения между Правительством Республики Беларусь и Правительством Российской Федерации о порядке признания электронной подписи (электронной цифровой подписи) в электронном документе при трансграничном электронном взаимодействии», подписанном 15 апреля 2024 года. Есть шанс и смысл использовать эти лучшие практики и научиться оценивать уровень надёжности иностран-

ной электронной подписи на основе этого канонического подхода, по сути, описанного в типовом законе об электронных подписях, на основе которого созданы национальные законы об электронной подписи подавляющего большинства стран мира (табл. 2). Это открывает широкие возможности масштабирования трансграничного пространства доверия в рамках электронного документооборота В2В (и не только) со всеми экономическими партнёрами во всём мире.

1. Австралия	19. Грузия	37. Макао	55. Тайвань
2. Австрия	20. Дания	38. Малайзия	56. Туркменистан
3. Азербайджан	21. Израиль	39. Мальта	57. Турция
4. Алжир	22. Индия	40. Мексика	58. Узбекистан
5. Аргентина	23. Индонезия	41. Молдавия	59. Украина
6. Армения	24. Иран	42. Нидерланды	60. Филиппины
7. Бангладеш	25. Ирландия	43. Норвегия	61. Финляндия
8. Беларусь	26. Испания	44. Перу	62. Франция
9. Бельгия	27. Италия	45. Польша	63. Хорватия
10. Бермудские острова	28. Казахстан	46. Португалия	64. Чехия
11. Болгария	29. Кипр	47. Россия	65. Чили
12. Бразилия	30. Киргизия	48. Румыния	66. Швейцария
13. Великобритания	31. КНР	49. Сингапур	67. Швеция
14. Венгрия	32. Колумбия	50. Словакия	68. Эстония
15. Вьетнам	33. Корея	51. Словения	69. ЮАР
16. Германия	34. Латвия	52. США	70. Япония
17. Греция	35. Литва	53. Таиланд	
18. Гонконг	36. Люксембург	54. Таджикистан	

Таблица 2. Страны, в которых законодательство об ЭП основано на MLES¹

¹ По данным актуальных исследований компании «Газинформсервис», а также с использованием материалов «Общее руководство по законодательству в области ЭП: краткое изложение законодательства и исполнимости по странам»/© Adobe Systems Incorporated 2016, «Глобальный индекс кибербезопасности и профили по киберблагополучию. Отчёт. Исследовательская группа ABI Research по заказу группы кибербезопасности МСЭ». Апрель 2015 г.

ПРОЗРАЧНОСТЬ ОБРАЗОВ ВЛЕЧЁТ...

О СИСТЕМЕ МОНИТОРИНГА БЕЗОПАСНОСТИ КОНТЕЙНЕРОВ В RUNTIME



Алина КНЯЗЕВА
руководитель
отдела
информационной
безопасности
Дзена



Татьяна ХУРТИНА
программист
группы
внутренней
автоматизации
ИБ VK

Сейчас большинство разработчиков приложений используют контейнеры для развёртывания инфраструктуры, выбирая этот способ по ряду факторов — портативность, возможность масштабирования, отказоустойчивость, гибкость в процессах доработки и эффективность. Контейнеризация уже прочно вошла в процесс DevOps, но известно, что образ любого контейнера может содержать ошибки и уязвимости, которые могут привести к угрозам работы всего сервиса. Поэтому крайне важно вовремя отследить, какие образы используют сервисы компании и насколько они безопасны.

КОГДА В ПРОЦЕСС DEVOPS НЕОБХОДИМО ВКЛЮЧИТЬ SECURITY?

При разработке ПО приложений чаще всего используются сторонние библиотеки (зависимости), но ключевой вопрос безопасности — насколько им можно доверять. В распоряжении разработчиков огромное количество различных ресурсов, но не все они лишены уязвимостей. Например, в популярных Java-библиотеках или криптографических пакетах до сих пор часто встречаются:

- ♦ Уязвимость Log4j (CVE-2021-44228) сохраняет в логах строку, которая мо-

жет загрузить и запустить из системы вредоносный скрипт злоумышленника, что приведёт к полному захвату приложения и сервера.

- ♦ Уязвимость Spring Framework (CVE-2021-22698) даёт возможность дистанционного запуска вредоносного кода (RCE).

- ♦ Уязвимость OpenSSL Heartbleed (CVE-2014-0160) даёт возможность перехватывать информацию, защищённую шифрованием SSL/TLS. Злоумышленники могут украсть секреты, данные пользователей, информацию из электронной почты и документов, не оставив никаких следов.

Кроме того, активно развивается тенденция общего технического усложнения всех процессов — с появлением новых прогрессивных инструментов оценки защищённости уязвимости также становятся более изощрёнными.

Известно, что вредоносные зависимости наследуются из разных источников, и если не знать, где и какие библиотеки задействованы, то неясно, куда обратиться для устранения проблем. К тому же может возникнуть ситуация, что закрытие уязвимости не так критически и важно, например, если она находится в изолированной части системы.

Всё это подводит к логической необходимости включения системы сквозного мониторинга для по-

иска критически важных уязвимостей и обеспечения комплексной безопасности уже функционирующего приложения. Внедрив грамотно выстроенный процесс, компания может уверенно контролировать состояние безопасности инфраструктуры, обеспечивать защиту от угроз и быстрее реагировать на инциденты.

Но возникает вопрос, как правильно организовать и технически реализовать мониторинг контейнеров в Runtime, особенно если это сложная инфраструктура с большим количеством компонентов? В этой статье на примере опыта Дзена расскажем, как была реализована эта задача с помощью доработанных инструментов Open source.

КОНТЕКСТ

У Дзена нестандартная инфраструктура, развёрнутая во внутреннем облаке с иерархической системой организации структур и сервисов. Дзен использует десятки тысяч контейнеров, каждый из которых содержит тысячи уникальных образов с сотнями различных зависимостей. Для комплексного решения вопросов безопасности в Дзене реализована функция архитекторов ИБ — они проводят анализ продукта и формируют задание по безопасности, консультируют команды разработчиков и в случае публикации уязвимости помогают её быстро обнаружить и предложить решения по исправлению.

Поэтому перед нашей командой автоматизации ИБ стояла задача обеспечить их необходимыми инструментами контроля. Нам надо было создать удобные технические решения, которые закрыли бы вопросы по инвентаризации сервисов — сколь-

ко и каких сервисов запущено, какие они имеют компоненты и лицензии компонентов, и по работе с уязвимостями — быстрому поиску, приоритизации и устранению. Мы сразу решили, что будем создавать собственное решение на базе Open source.

Для понимания, как данное решение было встроено в классическую модель организации инструментов DevSecOps, уточним, что с помощью данной разработки мы работаем сразу на четырёх этапах процесса — Release, Deploy, Operate и Monitor (рис. 1). На этапе Release происходит анализ контейнеров и пакетов, на этапе Deploy — отслеживание зависимостей, а в Operate — анализ уже запущенных контейнеров, что позволяет реализовать механизм оперативного обнаружения уязвимых зависимостей и контейнеров. И всё это, по сути, и складывается в систему Monitor.

НАША РЕАЛИЗАЦИЯ НА OPEN SOURCE

Сначала надо было составить представление, из каких компонентов состоит сервис, и для этого нам потребовалось создать Software Bill of Materials (SBOM) — структурированный список компонентов программного обеспечения, используемых в контейнерах, и мы выбрали формат CycloneDX. В качестве инструмента для создания SBOM было выбрано решение Trivy, которое позволяет выявить известные уязвимости в компонентах для последующего анализа и реагирования на потенциальные угрозы.

Чтобы предоставить аналитикам ИБ удобный инструмент для сканирования на уязвимости и управление рисками, сначала мы выбрали платформу для анализа компонентов Dependency-Track. Почему выбор пал на это решение?

Dependency-Track — это интеллектуальная платформа, которая позволяет выявлять и снижать риски в цепочке поставок программного обеспечения. Она применяет уникальный и очень выгодный подход, используя возможности и спецификации программного обеспечения (SBOM). Кроме того, платфор-

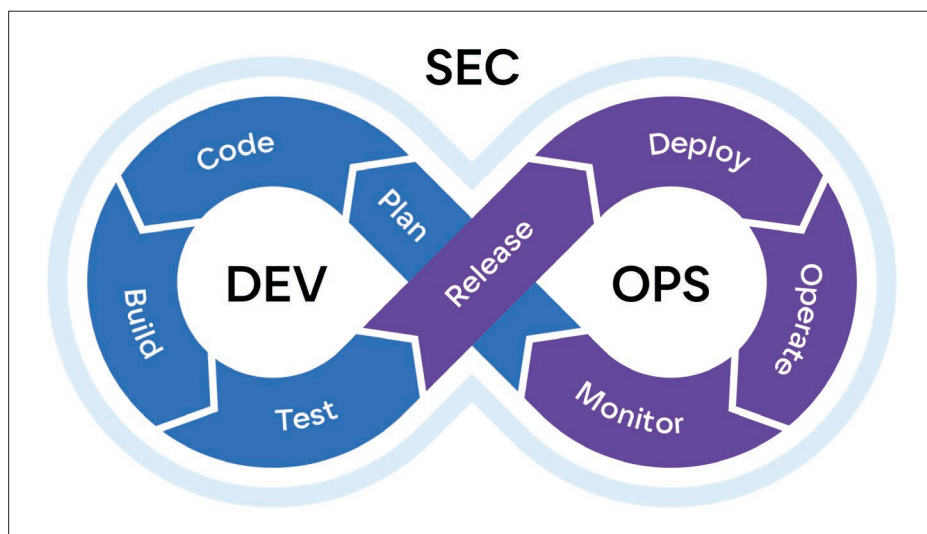


Рисунок 1. Работаем сразу на четырёх этапах процесса — Release, Deploy, Operate и Monitor

ма имеет встроенные инструменты для выявления уязвимых компонентов — National Vulnerability Database (NVD), Sonatype OSS Index, GitHub Advisories, Snyk, OSV, VulnDB (Risk Based Security).

Однако в процессе тестирования оказалось, что Trivy сканирует собственные SBOM на уязвимости лучше, чем это делают сторонние решения, в том числе Dependency-Track. Поэтому мы кастомизировали данные решения: сделали ещё один плагин для выявления уязвимых компонентов для Dependency-Track на базе Trivy — Trivy-REST. Этот инструмент предоставляет REST API для взаимодействия с Trivy, обеспечивая возможность автоматизированного сканирования компонентов и получения результатов анализа.

Одновременно с этим был также создан собственный сканер, который включает в себя планировщик для выполнения следующих задач:

- ◆ автоматическое создание SBOM новых Docker-образов;
- ◆ автоматическая отправка SBOM всех Docker-образов в Dependency-Track.

В итоге у нас получилась следующая архитектура решения — рис. 2.

МЕХАНИКА ПРОЦЕССА

1. Сканер периодически запрашивает список запущенных в облаке образов, обеспечивая актуальность данных мониторинга.

2. Сканер получает информацию по каждому образу из registry. В качестве уникального идентификатора образов служит хеш-сумма (Digest)

3. Сканер посредством Trivy создаёт SBOM в формате CycloneDX и сохраняет его в базу данных.

4. Также сканер посредством другого планировщика создаёт проект в Dependency-Track и отправляет SBOM. Отправляются как новые SBOM, так и уже существующие. Поскольку база уязвимостей периодически обновляется, следует пересканировать уже существующие образы.

5. Dependency-Track, используя Trivy-REST, анализирует компоненты.

Последовательность взаимодействия внутри системы получилась такая — рис. 3.

КАКОЙ ЭФФЕКТ?

Благодаря собранному решению удалось добиться следующих результатов.

◆ **Актуальность данных.** Регулярное сканирование контейнеров и создание SBOM позволяет всегда получать самую последнюю информацию о составе контейнеров и их уязвимостях.

◆ **Автоматизация процесса.** Использование планировщиков обеспечивает непрерывный мониторинг безопасности, устраняя необходимость вручную запускать процессы сканирования и анализа.

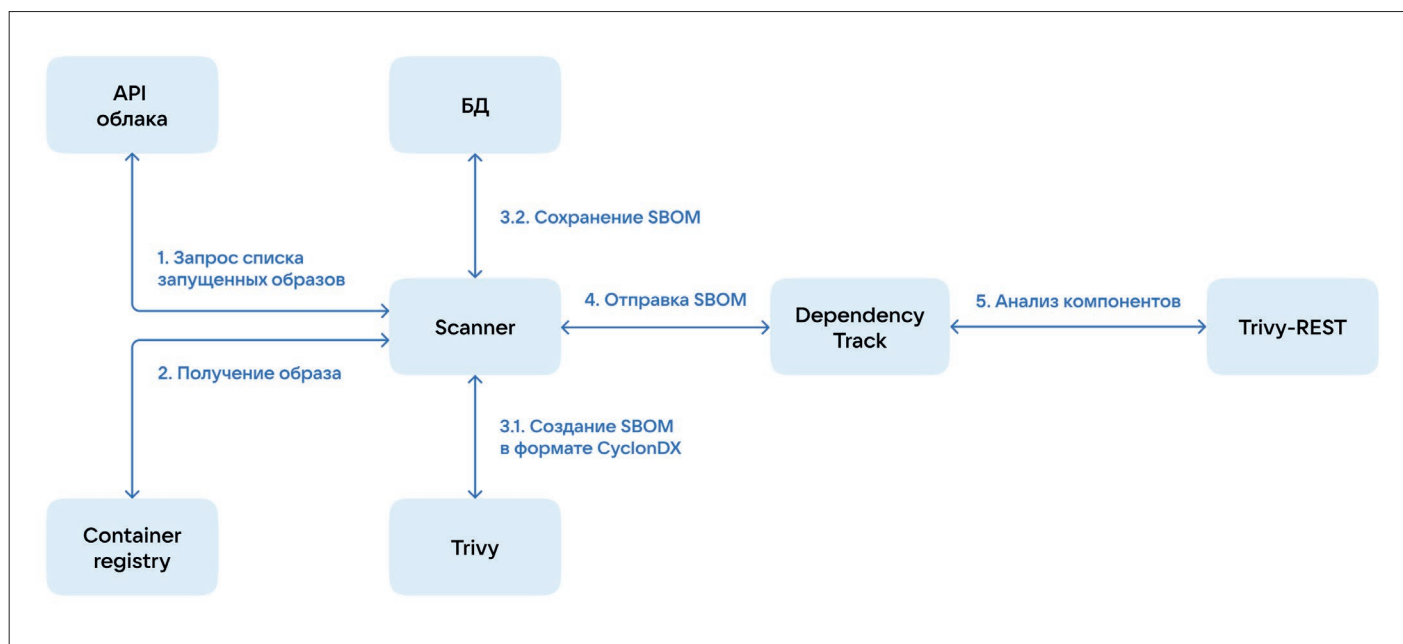


Рисунок 2. Архитектура решения

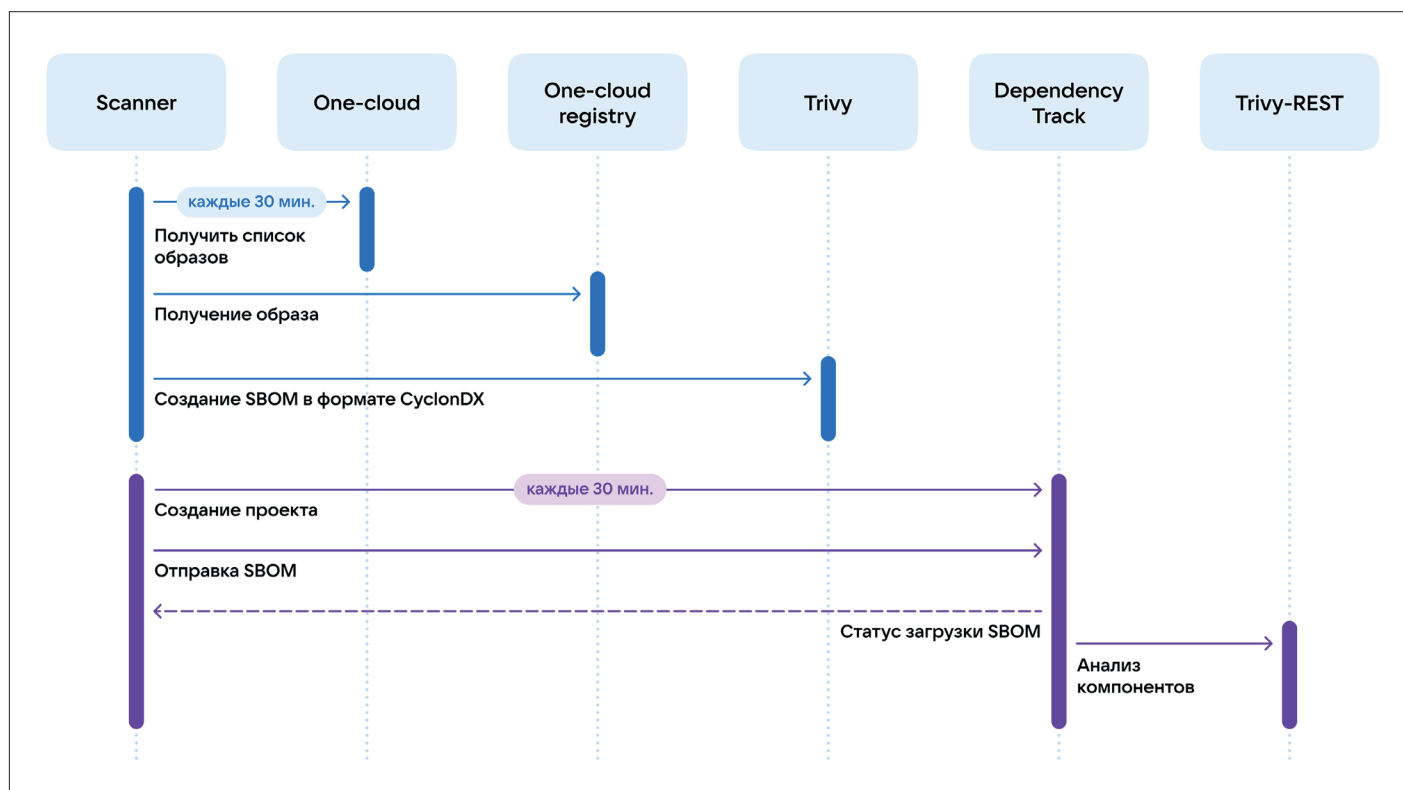


Рисунок 3. Последовательность взаимодействия внутри системы

♦ **Наглядность.** Dependency-Track предоставляет удобный интерфейс для анализа безопасности контейнеров с централизованным доступом к информации об уязвимостях. Это облегчает принятие решений о приоритетности той или иной уязвимости в составе приложения и позволяет оперативно реагировать на критически важные.

♦ **Универсальность.** Так как SBOM достаточно собрать один раз, то облако довольно быстро проверяется на уязвимости при добавлении новых образов.

Отметим, что данная система мониторинга безопасности контейнеров в Runtime может быть легко и в любой момент интегрирована

в процессы обеспечения безопасности. Даже если приложение уже функционирует, подобная сквозная автоматизированная система позволит увеличить эффективность взаимодействия между командами ИБ и разработки, особенно в части приоритизации задач по устранению уязвимостей.

ИСКАЖЕНИЕ РЕАЛЬНОСТИ

СПАСЁТ ЛИ ФАКТЧЕКИНГ ОТ ЛОЖНЫХ ВОСПОМИНАНИЙ?



Андрей ЖАРКЕВИЧ
ведущий рубрики,
эксперт Start X

Нейросети на базе больших языковых моделей (Large Language Model, LLM) переживают бурный рост. Бизнес в восторге от потенциала экономии с помощью LLM. Клиники используют чат-ботов для быстрой диагностики заболеваний по набору симптомов. Отделы маркетинга и редакции сокращают авторов, поскольку уверены, что тексты для них быстрее и дешевле напишет нейросеть. И если раньше вопросы было принято задавать «Гуглу», то сейчас в моде ChatGPT или его «братья по разуму».

Масштабы доверия к ответам нейросетей поражают воображение и вызывают обоснованный скепсис у знающих людей. Причина в том, что нейросети не только успешно проходят тесты на человечность, но и имеют склонность к вполне человеческому заблуждению — конфабуляции, или ложным воспоминаниям.

В материале поговорим о том, почему LLM придумывают несуществующие факты, обозначим связанные с этим риски и выясним, существуют ли методы для эффективного решения этой проблемы.

КАК РАБОТАЮТ LLM

Если не углубляться в детали, в основу работы LLM положена статистика. В процессе обучения нейросети собираются статистические данные о сочетаниях слов в реальных текстах. Это требует огромных массивов данных для обучения, в ходе которого собираются не только возможные варианты взаимного расположения слов, но и другие характеристики текстов, определяющие стиль, эмоции, настроение и многое другое. Например, разработчики сберовского GigaChat использовали для обучения своей модели 7,50 петабайта исходных данных. По объёму это в 2,5 раза больше, чем вся мировая литература.

В результате такого обучения нейросеть получает возможность строить фразы на базе заданного вопроса или продолжать предложение. Например, если мы попросим LLM продолжить предложение «Основные принципы информационной безопасности — это», то получим примерно такой ответ:

«Основные принципы информационной безопасности — это конфиденциальность, целостность и доступность информации».

Нейросеть проанализировала данные обучения и выбрала сочетания слов, которые с максимальной вероятностью присутствуют в предложениях, начинающихся с заданного текста.

ЛОЖНЫЕ ВОСПОМИНАНИЯ

Собранные во время обучения закономерности и взаимосвязи в огромных объёмах текстовых данных позволяют языковым моделям генерировать текст, похожий на человеческий. Но это также означает, что иногда они могут выдавать результаты, которые статистически вероятны, но фактически неверны. Такое явление называют галлюцинацией, или конфабуляцией LLM.

Суть его в том, что нейросеть генерирует текст, который выглядит связным и правдоподобным, но содержит фактические неточности, несоответствия или даже полностью сфабрикованную информацию, например ссылки на несуществующие интернет-ресурсы, публикации в выдуманных СМИ или события, которые никогда не происходили.

Вот как выглядит галлюцинация у ChatGPT (рис. 1). У модели спросили, когда в 1980-х в Аргентине проходил чемпионат мира по горным лыжам FIS. В вопросе содержится очевидная ошибка, поскольку Аргентина никогда не принимала чемпионат мира по горным лыжам. Несмотря на это, модель уверенно отвечает: «Чемпионат мира по горным лыжам FIS проходил в Аргентине в 1985 году».

Когда задают дополнительный вопрос о том, кто выиграл конкретные соревнования на этом вымышленном чемпионате, модель придумывает дополнительные детали, заявляя: «Скоростной спуск среди женщин выиграла Микела Фигини из Швейцарии, а победителем в скоростном спуске среди мужчин стал швейцарец Пирмин Цурбригген».

Эти спортсмены в самом деле были лучшими в скоростном спуске на чемпионате

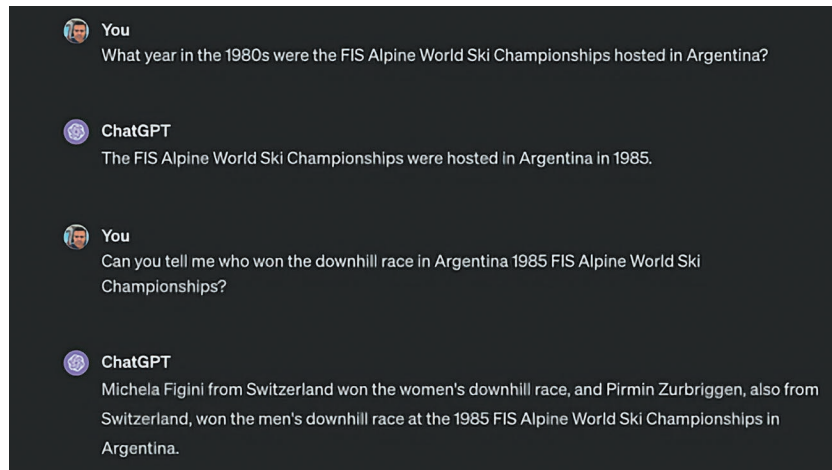


Рисунок 1. Так выглядит галлюцинация у ChatGPT

мира по горным лыжам в 1985 году. Только проходил он в Италии, а не в Аргентине, как написал ChatGPT.

БРЕД ИЛИ КРЕАТИВ?

В общем случае модели не отличают правду от вымысла. У них есть сложная структура вероятностей и связей, на базе которых они генерируют ответы на вопросы. Готовность моделей подробно рассказать о воображаемом событии без каких-либо признаков неопределённости называют галлюцинацией, но проблема в том, как использовать созданный моделью ответ. Ведь, если нам нужны не факты, а придуманная история, галлюцинацию можно рассматривать как продукт творческой деятельности.

Креатив предполагает создание новых идей, историй или концепций, которые не обязательно должны быть фактически точными. Например, LLM может создать вымышленную историю о разумном метеорите, который путешествует по просторам вселенной в поисках разума. Однако, если бы та же модель написала новость о том, что учёные NASA внедрили в пролетающий неподалёку от земли метеорит компьютер с ChatGPT и отправили его искать разумную жизнь, это был бы яркий пример галлюцинации. Причина в том, что выдуманная информация представлена как фактическая.

Распознавание и смягчение галлюцинаций имеет решающее значение для обеспечения надёжности и достоверности контента, создаваемого LLM. Критически важно это там, где фактическая точность имеет первостепенное значение, например, когда модели используются для постановки диагнозов и назначения лечения или когда чат-бот самостоятельно обрабатывает обращения клиентов на сайте.

К ЧЕМУ ПРИВОДИТ ДОВЕРЧИВОСТЬ

Фантазии LLM не настолько безобидны, как может показаться на первый взгляд. Вот лишь несколько из недавних случаев, когда доверчивость собеседников больших языковых моделей создала проблемы.

В 2023 году в Южном округе Нью-Йорка разразился скандал, причиной которого стал иск, поданный одним из адвокатов. В документах по делу юрист процитировал шесть несуществующих судебных решений с цитатами и внутренними ссылками, которых не существовало в природе.

Источником фальшивки стал ChatGPT, к которому обратился адвокат, решивший оптимизировать работу с помощью современных технологий. Он сделал это впервые и не знал, что ответы нейросетей нужно досконально проверять.

Американский радиоведущий из Джорджии Марк Уолтерс был крайне удивлён, когда обнаружил, что ChatGPT распространял о нём ложную информацию, обвиняя его в растрате денег.

Из-за галлюцинации чат-бота на сайте Air Canada в феврале 2024 года компании пришлось выплатить пассажиру компенсацию в несколько сот долларов США. Виртуальный ассистент пообещал ему, что он может получить скидку в связи с тяжёлой утратой в течение 90 дней после покупки билета по обычной цене, в то время как в тарифах авиакомпании подобного пункта не было.

Галлюцинации нейросетей могут использовать в своих целях злоумышленники. Например, эксперты описали метод атаки через цепочки поставок, получивший название «галлюцинация пакетов ИИ».

Атака использует галлюцинации ChatGPT и других LLM для поиска ссылок на несуществующие СМИ, блоги и статистику, а также на вымышленные исправления для CVE и отсутствующие в природе библиотеки кода для Python, Node.js и других языков программирования.

Злоумышленник начинает с формулировки вопроса, в котором просит ChatGPT подобрать пакет, решающий некую проблему. ChatGPT выдаёт названия библиотек, часть из которых могут быть вымышленными. Ситуация становится опасной, когда ChatGPT рекомендует пакеты, не опубликованные в репозиториях npmjs, PyPi или аналогичных.

Получив рекомендацию с отсутствующим в реестрах пакетом, злоумышленник может создать вредоносный пакет с таким именем и опубликовать его. Когда другой пользова-

тель задаст аналогичный вопрос, он может получить от ChatGPT рекомендацию использовать вредоносный пакет, причём в этот раз он уже будет доступен для загрузки из официального репозитория.

ФАКТЧЕКИНГ ДЛЯ LLM

Было бы странно, если бы для защиты от недостоверных ответов LLM не предложили организовать фактчекинг. Ведь, если языковые модели претендуют на роль технологии, которая продвинет человечество в будущее, нужен способ смягчить проблемы галлюцинаций на этом пути.

Одной из таких технологий стала расширенная поисковая генерация (Retrieval Augmented Generation, RAG). Её суть в том, что LLM, перед тем как выдать ответ на запрос пользователя, добавляет к данным обучения дополнительную информацию из внешних источников, например из поискового запроса или из выдачи внутренней системы хранения документов. Другими словами, в контекст запроса к нейросети добавляется уточняющая информация, благодаря которой пользователь получает более полный, точный и актуальный ответ.

Например, если пользователь спросит у LLM стоимость акций Сбербанка, то без RAG дать полезный ответ не получится, поскольку в обучающем массиве данных вряд ли найдётся курс SBER за сегодня. Но если сделать запрос в «Яндексе» типа «биржевой курс SBER сегодня» и добавить его к запросу, то LLM сгенерирует актуальный и достаточно точный ответ.

Каждый ответ, который выдаёт LLM, может быть добавлен во внешние данные, используемые в RAG. Теоретически это может повысить точность ответа.

Важно, что модели, использующие RAG, могут ссылаться на источник своих утверждений. В результате, если LLM выдаёт неверный ответ и он идентифицирован, источник этой неверной информации может быть точно определён и удалён или исправлен.

ChatGPT от OpenAI использует RAG, когда выполняет веб-поиск, связанный с вопросом пользователя, чтобы выдать более актуальную информацию и ссылку на источник, который пользователь может проверить. Аналогичным образом действует модель Gemini от Google.

Процесс внедрения RAG в большие языковые модели вполне закономерный этап развития. Правда, оказалось, что разные LLM работают по-разному при использовании RAG, лучше или хуже справляясь с обработкой информации, которую RAG отправляет

обратно в LLM. Более того, обработка полученной от RAG информации может не улучшить точность выдачи, а ухудшить её и даже вызывать новые галлюцинации.

С такой ситуацией столкнулись исследователи из Университета Мэриленда, обнаружившие, что GPT-3.5 может давать сбои даже при получении уточняющих данных от RAG. Это происходило в тех случаях, когда предоставленный контекст выходил за рамки данных обучения модели. При этом LLM генерировал правдоподобные галлюцинации, интерполируя фактическое содержание.

Исследование Пекинского университета выявило, что «внедрение поиска неизбежно увеличивает сложность системы и количество гиперпараметров для настройки».

А ещё RAG может создать проблему, значительно расширив контекстное окно — общее количество символов или слов, которые LLM должна обрабатывать.

Исследователи из WeChat предложили решить проблему с обработкой моделями полученных от RAG данных с помощью специального «метода обучения уточнению информации» под названием INFO-RAG. Его суть в том, что чтобы использовать для обучения LLM данные, полученные заранее с помощью RAG. Новый набор данных отбирается из авторитетных источников, разбивается на части, и модель обучается прогнозировать последнюю часть предложения, полученного из RAG.

ДОВЕРЯТЬ, ПРОВЕРЯТЬ, ОБУЧАТЬ

Галлюцинации в больших языковых моделях становятся серьёзным стоп-фактором на пути массового внедрения этих технологий в широкий спектр приложений. Поскольку LLM стремительно захватывают сферы деятельности и умы жителей Земли, жизненно важно устранить риски и последствия, связанные с галлюцинациями, от распространения дезинформации до подрыва доверия к контенту, создаваемому ИИ.

RAG и другие обособленные технологии проверки корректности выдачи LLM решают проблему лишь частично. Более перспективным представляется разработка нового поколения нейросетей, в архитектуру которых на этапе проектирования встроены механизмы защиты от галлюцинаций и другие средства обеспечения корректности выдач.

Только таким образом можно обеспечить уверенность в том, что нейросети будут внедряться и использоваться с пользой для общества, станут надёжными помощниками для бизнеса и обычных людей.



ДЕТАЛИ И ГЛАВНОЕ

О КАДРОВЫМ ПОТЕНЦИАЛЕ В СФЕРЕ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ



Игорь КАЧАЛИН
генеральный директор
АНО «НТЦ ЦК»

Университет «Сириус» провёл конференцию «Кадровый потенциал в сфере информационной безопасности», участники которой обсудили вопросы подготовки квалифицированных специалистов и проблемы российского образования. Мероприятие прошло при поддержке Минцифры России, ФУМО, ФСБ и ФСТЭК России, Академии криптографии РФ, компаний — лидеров от-

расли кибербезопасности. **BIS Journal** попросил поделиться своими впечатлениями и мыслями по итогам конференции генерального директора АНО «НТЦ ЦК» Игоря Качалина.

НЕСКОЛЬКО СЛОВ В ПРЕДИСЛОВИЕ

Проводя эту конференцию, Университет «Сириус» дал возможность обменяться мнениями как работодателям, так и тем, кто занимается подготовкой специалистов разного профиля и разного уровня. Подобные мероприятия позволяют выявить и донести до вузов те проблемы, которые есть у работодателей. Понять, какие сложности и проблемы есть у тех, кто занимается подготовкой кадров, найти выход из ситуации и осознать, что это не тупик.

С точки зрения подготовки кадров нужно сделать так, чтобы работа

по специальности в области ИБ была привлекательной для молодых людей, доставляла им удовольствие, была нужной для государства — в этом её ценность. Участие в этой конференции АНО «НТЦ ЦК» позволило нам в какой-то мере донести до сообщества потребности в исследователях и научных кадрах.

Наша молодая организация (АНО «НТЦ ЦК» работает чуть больше года — ред.) работает больше в исследовательской сфере. Основываясь на своём опыте, я полагаю, что студенты, получающие высшее образование, в первую очередь должны иметь исследовательские навыки, а не только профессионально знать языки программирования или иностранные языки, уметь общаться и программировать. Важно уметь искать и добывать новые знания, реализовывать их в востребованных на рынке продуктах.

Какими компетенциями должны обладать специалисты-исследователи, которые хотят работать в НТИ ЦК? На самом деле требования заложены в самом названии организации — цифровая криптография. С одной стороны, есть базовое понятие — криптография. С другой — слово «цифровая» говорит о новой эпохе в развитии России — цифровизации.

Найти место криптографии в новых условиях, когда технологии развиваются очень быстро, достаточно сложно. Ведь криптография — наука неторопливая, она требует глубоких исследований. Наша главная задача — добиться, чтобы отечественные криптографические механизмы широко использовались в ИТ, внедряемых в системы государственного управления. Отсюда возникают требования к специалистам.

Эти люди должны иметь подготовку в области информационных технологий, поскольку никакие новые технологии без ИТ не существуют. Это должны быть специалисты, знающие, что такое криптография, её слабые и сильные стороны. Неправильное применение хороших алгоритмов может приводить к серьёзному упущению в области безопасности.

Требования высокие, они вряд ли применимы к выпускникам, пришедшим непосредственно со студенческой скамьи. Но мы готовы брать и молодых специалистов, доводить их квалификацию до нужных нам требований. В тех проектах, которыми мы сейчас занимаемся, есть место исследователям, программистам, математикам и физикам.

В рамках государственных проектов, которые реализует АНО НТИ ЦК, в прошлом году мы привлекали коллективы студентов и преподавателей к научным исследованиям. Это был успешный опыт, результаты аналитических работ показали, что направление востребовано и имеет высокий потенциал. В текущем году мы развиваем это направление, будем привлекать студенческие группы к проведению научно-исследовательских работ.

Если же говорить о передаче опыта между поколениями, мне кажется, нельзя подходить менторски к взаи-



модействию с молодёжью. Она в чём-то умнее нас, иначе не было бы прогресса, это диктует жизнь. Надо учиться слушать и слышать молодёжь, её потребности и навыки. Находить применение устремлениям молодых, направлять их в нужное русло, на решение задач, которые стоят перед организацией. В этом и должно состоять наставничество, без насаждения своих взглядов.

ЕСТЬ ЛИ НАУКА В ВУЗАХ?

На этой конференции мне довелось стать модератором круглого стола «Наука в ИБ — потребности и возможности университетов, компаний и государства». Интересная получилась дискуссия. Началась она с вопроса: «Так есть ли наука в вузах?»

Коллеги не согласились с моим мнением, что в сборниках материалов к научным конференциям последних лет нет работ по ИБ, хотя практически все программы подготовки по направлению 10.00.00 «Информационная безопасность» открыты для доступа. Они отметили, что ряд направлений ИБ или не попадают в категорию инфобеза, например квантовые технологии, или кодируются и публикуются в закрытом виде, такие как криптография.

Получилось, что вопрос не в отсутствии работ, а в методике их классификации и специфике самой науки, связанной с ИБ.

Область инфобеза стоит на стыке математики, информационных и информационно-коммуникационных технологий, физики, экономики и юриспруденции. И порой даже возникает вопрос, что шире — ИБ или ИТ. С другой стороны, для любого объекта остро стоит вопрос защиты информации. Но в целом ИБ — узкая тема.

Тематика доверия, защищённости, надёжной эксплуатации — это те фундаментальные направления, которые успешно разрабатываются в вузах и используются на практике. Ещё больше точек соприкосновения в прикладных направлениях. В гражданской критической информационной инфраструктуре предстоит много сделать и категоризировать. В качестве примера успешного научного сотрудничества эксперты привели опыт взаимодействия МИРЭА и Госкорпорации «Росатом».

Машинное обучение, искусственный интеллект — серьёзная тема для фундаментальной и прикладной науки. И здесь стоит вспомнить вклад в математику российских учёных. В этом году исследователи



из Массачусетского технологического института (США) представили новую архитектуру нейронной сети, получившую название Kolmogorov-Arnold Networks (KAN). Она построена на основании трудов Андрея Колмогорова и Владимира Арнольда. Архитектура обещает произвести революцию в понимании и создании нейронных сетей, по крайней мере, так говорят её создатели.

Искусственный интеллект построен на математическом аппарате, который можно применять там, где он применим, и решать новыми средствами новые задачи. Мы выглядим не хуже западных коллег в области науки и обеспечения ИБ новыми разработками, в том числе фундаментальными результатами в математике. У нас есть свои результаты, которыми можно гордиться. Так, есть опыт создания АСУ ТП плавучих АЭС малой мощности. Здесь важен набор средств и методов защиты для надёжной и функционально устойчивой защиты работы станций. На основе математического аппарата учёными из вузов создаются динамически меняющиеся системы ИБ таких станций. Эта задача позволяет создать новый комплекс мето-

дов и средств для проактивной защиты объектов КИИ.

То есть участники дискуссии сошлись во мнении, что наука в университетах есть и будет, без этого не может существовать вуз, а обучение по ИБ невозможно без научного сопровождения.

О ФИНАНСИРОВАНИИ

Для развития науки и проведения научно-исследовательских и опытно-конструкторских работ (НИОКР) нужно устойчивое финансирование. Регулярных заказов на научные исследования можно ожидать со стороны госсектора и его силового блока, где планирование расходов ведётся на перспективу 3–5 лет. В госзаказе прописывается научная составляющая, которая является обоснованием дальнейшей деятельности.

И если госсектор планирует расходы на науку, то бизнес идёт на них при необходимости планировать фундаментальные разработки на длительную перспективу. Иной подход скорее исключение, чем правило. Долгосрочное планирование научных исследований есть у госкорпорации «Росатом», которая финансирует научные исследования в Московском

и Ивановском энергетических университетах. Ивановские учёные приняли активное участие в разработке проекта АСУ Курской АЭС-2 нового типа, которая будет запущена в конце года.

Огромный пласт возможностей сотрудничества участники конференции видят в гражданском секторе, где ведутся НИОКР, а вузы выступают как соисполнители. Этот пласт разработок планируется на годы вперёд и хорошо коррелируется с кадровым обеспечением.

Есть западный опыт частных компаний, которые заказывают исследовательские работы у университетов. Но российский бизнес часто живёт в парадигме конца 1990-х и не представляет текущее состояние научной базы вузов. Участники круглого стола заметили, что необходимо активизировать работу в этом направлении и через 3–4 года вузы получат волну заказов НИОКР.

Ещё один аспект исследований — отрицательный результат. Госзаказ не допускает такого итога работ. Но научные исследования — вещь непредсказуемая и не всегда согласуется с техзаданием заказчика. В таком случае стоит рассматривать иные формы взаимодействия сторон и финансирования работ. Такие подходы позволят заказчикам и исследователям рисковать, активизировать как минимум прикладные НИОКР.

НА ДОЛГОСРОЧНУЮ ПЕРСПЕКТИВУ

Опыт взаимодействия коммерческих компаний и учёных из вузов и академических институтов — отдельный вопрос. Совместную работу организовать сложно, поскольку наука рассматривает фундаментальный подход, а производство идёт по реализации, которая обеспечивает достаточный уровень ИБ и далека от науки. Интересен опыт компании «ИнфоТекС», которая создала исследовательское подразделение, ставшее интерфейсом между разработчиками и учёными, что позволяет учитывать требования бизнеса к разработке.

Компания разработала стратегию работы с вузами, которая включает



краткосрочные ознакомительные программы и долгосрочное сотрудничество, работу по нескольким трекам. Многое зависит от открытости вузов, их готовности сотрудничать, а не только желания получить финансирование.

ОСТРЫЕ ВОПРОСЫ

Острый вопрос – уровень подготовки преподавательского состава. Сложилось поколение учёных и преподавателей, которые никогда не выходили из стен вуза. Они выросли от студентов до профессоров, но не имеют опыта практической работы. Возникает разрыв НИОКР и производства. Возможно, стоит вывести аспирантуру из системы высшего образования и принимать на учёбу только людей, имеющих опыт практической работы, тогда не будет возникать противоречий между наукой и практикой.

Если вендор открывает в университете свои учебные подразделения под НИР, формирует региональные кластеры разработки, от этого выигрывают все стороны. Вместе с учебной программой начинаются

научные исследования, публикации и внедрения с партнёром, отметили участники дискуссии. Но здесь нужно встречное движение со стороны вузов и координация работ. Это огромная долгосрочная история, которая упрощает работу в прикладной тематике, но требует терпения со стороны участников.

Показательна инициатива по центрам ИИ, когда промышленные партнёры подключаются к финансированию. Силами ИСП РАН при участии АНО НТЦ ЦК, частных компаний и регуляторов ведутся фундаментальные исследования прикладных доверенных систем ИИ и больших языковых моделей. Однако многим вузам не интересен индивидуальный подход, который даёт развитие кадрового потенциала, но не идёт в зачёт организации и не позволяет финансировать административный персонал.

ПОДВОДЯ ИТОГИ

Завершая дискуссию, я попросил коллег выделить главное, что должно помочь и заказчику, и университетам результативно развивать

вузовскую науку. Получилось следующее:

- ◆ нужна история успеха; надо идти к большим проектам маленькими шагами;
- ◆ открытость и партнёрство с обеих сторон, долгосрочное планирование и сотрудничество;
- ◆ партнёры не должны предъявлять друг другу завышенных требований и ожиданий;
- ◆ желательна готовая модель, которую небольшие вузы могли бы внедрить;
- ◆ промышленный партнёр должен идти в вуз, учить и финансировать учёных, развивать науку – выращивать команду. В университете всегда есть люди, готовые заниматься практическими работами;
- ◆ стороны должны ставить амбициозные стратегические задачи на 5–10 лет, которые потребуют научных исследований.
- ◆ вузы должны ориентироваться на мировой уровень, и тогда взаимодействие сторон будет эффективным.

Записала
Марина Бродская

А ЕСТЬ ЛИ У ВАС ПЛАН, МИСТЕР ФИКС?

НА НЕИСПОЛНЕНИЕ ДИРЕКТИВ ПРЕДПОЛАГАЕТСЯ ОТКАЗ ОТ ПРИНЦИПА ДОБРОВОЛЬНОСТИ

Анна КАТАНКИНА
корреспондент BIS Journal

Регуляторы ведущих стран мира ужесточают требования к критически важной инфраструктуре и вводят обязательные требования по срокам информирования об инцидентах, планам реагирования и нормам отчётности. На неисполнение директив предполагается отказ от принципа добровольности.

SEC МОДЕРНИЗИРУЕТ СТАРЫЕ ЗАКОНЫ

Комиссия по ценным бумагам и биржам США (SEC) объявила о новых правилах, требующих от ряда финансовых учреждений иметь чёткие планы действий на случай утечки информации о клиентах.

Правила, принятые в качестве поправки к предыдущим требованиям от 2000 г., распространяются на брокеров-дилеров, финансовые порталы (напр., Kickstarter или GoFundMe), инвестиционные компании, зарегистрированных инвестиционных консультантов и трансфер-агентов. Они обязаны «разрабатывать, внедрять и поддерживать письменные политики и процедуры» для обнаружения и устранения нарушений, связанных с информацией клиентов.

Поправки также вводят правила, обязывающие компании разработать процедуры уведомления клиентов, к конфиденциальной информации которых произошёл несанкционированный доступ или произошла её утечка.

Как отметил глава SEC Гэри Генслер в своём заявлении, эти поправки необходимы, поскольку за более чем два десятилетия с момента вступления в силу исходного постановления «характер, масштаб и последствия утечек данных существенно изменились».

Основная идея требований заключается в том, что, если произошло нарушение, орга-

низация должна уведомить об этом, считает Генслер. Этот шаг важен для инвесторов.

Организации должны уведомить жертв как можно скорее и не позднее, чем через 30 дней после того, как стало известно об инциденте, связанном с утечкой информации о клиентах. Уведомление должно содержать подробную информацию об инциденте, утечке данных и о том, что пострадавшие могут сделать для своей защиты.

Поправка вступит в силу через два месяца после публикации правила в Федеральном реестре. Крупным компаниям даётся 18 месяцев на выполнение требований закона, более мелким — два года. SEC не уточнила, каким образом она делит предприятия по размеру.

Новые поправки вступают в силу на фоне постепенного перехода компаний к новым правилам SEC по отчётности об инцидентах. Требования обязывают публичные компании уведомлять агентство о «существенных» инцидентах. Несколько крупных компаний, в том числе Microsoft, Hewlett Packard, Frontier и др., уже были вынуждены подать документы формы 8-K об инцидентах в области кибербезопасности.

Попытки отменить эти требования SEC предпринимают члены палаты представителей, которые считают, что Комиссия плохо подготовлена к решению проблем, связанных с кибербезопасностью, и что отчёты об инцидентах подвергают пострадавшие компании новым атакам. В ответ Белый дом заявил, что наложит вето на любую законодательную попытку отменить правило SEC.

Эксперты по кибербезопасности приветствуют действия SEC. Они отмечают, что многолетняя практика добровольности соблюдения правил ИБ способствовала вялому отношению многих компаний к вопросам кибербезопасности, в том числе и к сообщениям о нарушениях и утечках, и потому компании не готовы соблюдать новые требования.

ЕС ГОТОВИТСЯ К НОВОЙ ЭРЕ NIS 2

18 октября 2024 г. в ЕС вступит в силу Директива NIS 2 (Network and Information

Systems Directive 2) — закон, который придёт на смену устаревшей Директиве NIS 2016 г. Он направлен на повышение устойчивости критически важных секторов по всему Евросоюзу.

Директива уделяет значительное внимание персональной ответственности заинтересованных сторон в управлении рисками кибербезопасности и вводит финансовые санкции за несоблюдение требований закона. Широкая сфера применения, строгие санкции и ключевая роль руководства — опорные точки, которые обеспечат прочную основу для построения киберзащиты организаций на фоне развивающихся рисков безопасности, считают авторы закона. Организации, работающие на территории ЕС или поставляющие продукцию в страны Евросоюза, в рамках NIS 2 должны реализовать 10 важнейших мер кибербезопасности, чтобы соответствовать требованиям законодательства.

NIS 2 расширяет сферу действия, он охватывает такие сектора, как цепочки поставок, производство продуктов питания и государственное управление. Закон будет введён во всех странах — членах ЕС и вводит правила ограничения размера. Все средние и крупные предприятия, работающие в критически важных секторах, подпадают под действие директивы. Эти меры направлены на усиление кибербезопасности в отраслях, имеющих важное значение для экономики.

Основная цель NIS 2 — повысить киберустойчивость организаций по всему ЕС путём укрепления активной позиции в области ИБ и обеспечения тесного сотрудничества между организациями и их поставщиками. NIS 2 также предполагает развитие сотрудничества между секторами, обмен информацией об угрозах и способах борьбы с ними.

В рамках NIS 2 организации должны соблюдать строгие требования безопасности. Они включают надёжные методы управления рисками, планирование непрерывности бизнеса, корпоративную подотчётность и обязательства по отчётности.

Директива определяет конкретные обязательства и сроки предоставления отчётности. Установлены предельные сроки уведомления регуляторов об инцидентах: 24 часа для первичного извещения о происшествии, 72 часа для уточнения информации и один месяц — на подготовку полного отчёта. Компании должны обеспечить оценку серьёзности ущерба и информировать

все заинтересованные стороны об их роли в процессе уведомления.

NIS 2 возлагает ответственность на корпоративное руководство за надзор, одобрение и прохождение обучения мерам кибербезопасности организации. Руководство компаний несёт ответственность за контроль и утверждение мер кибербезопасности, а также за обеспечение соблюдения требований NIS 2. Невыполнение этих обязанностей, неспособность устранить киберриски может привести к штрафам для руководства, включая потенциальную ответственность и временное отстранение от руководящих должностей.

Невыполнение требований, изложенных в NIS 2, может повлечь за собой значительные штрафы и репутационный ущерб. 10 млн евро или 2 % от общего годового оборота — максимальный штраф, который может быть наложен на предприятие за несоблюдение директивы.

БЕЛЫЙ ДОМ УКРЕПЛЯЕТ РОЛЬ CISA

В 2013 г. администрация Б. Обамы выпустила «Директиву президентской политики по безопасности и устойчивости критической инфраструктуры», которая стала предшественником Закона об Агентстве по кибербезопасности и инфраструктурной безопасности (CISA) 2018 г. Этот новаторский для того времени документ подчёркивал важность растущего риска кибератак на критически важную инфраструктуру. Однако со временем программы кибербезопасности нуждаются в переоценке. В апреле этого года президент Д. Байден подписал новую директиву, которая усилила роль CISA в защите КИИ от киберугроз.

Меморандум о национальной безопасности (NSM) давно ожидался в отрасли ИБ. Он закрепляет роль CISA в обеспечении национальной безопасности, но не учитывает изменения в критической инфраструктуре, которые произошли за последнее десятилетие, пишут эксперты. В исходном документе 2013 г. были определены 16 отраслей, которые работают напрямую с агентствами по управлению отраслевыми рисками (SRMA) и департаментами правительства. К этим отраслям относятся химическая промышленность, критически важные производства и службы экстренной помощи, вопросы продовольствия и сельского хозяйства, финансовые услуги.

В NSM 16 отраслей остались без изменений, в закон не были включены космос

и биоэкономика — две критически важные отрасли, рекомендованные для включения CISA.

Космос играет огромную роль в сфере телекоммуникаций, интернет-услуг, спутников и GPS. Но чиновники заявили, что космическая инфраструктура широко сегментирована и является частью других отраслевых агентств. Биоэкономика подпадает под другой указ 2022 г., который способствует пониманию технологий, формирующих отрасли.

Роль CISA в новом законе прописана более чётко. NSM считает CISA «национальным координатором по безопасности и устойчивости» критической инфраструктуры страны и отраслевых агентств. Агентство официально реализует предусмотренные законом полномочия, чтобы возглавить национальные усилия, направленные на управление и снижение рисков для кибер- и физической инфраструктуры, работает на межведомственном уровне и далее поддерживает реализацию ролей и обязанностей SRMA.

В рамках новой роли CISA будет отвечать за оценку прогресса в улучшении приоритетов безопасности и устойчивости в отраслевых агентствах, а также за выявление угроз и рекомендации мер по улучшению кибербезопасности. CISA будет поддерживать обмен важной информацией по безопасности на уровне правительства.

КИБЕРНАЛОГ ПРИЗНАЛИ НЕСВОЕВРЕМЕННЫМ

Правительство Нигерии приостановило введение сбора на кибербезопасность при внутренних денежных переводах. Новым налогом власти планировали ограничить использование криптовалют, которые, как считают в правительстве, виновны в ослаблении национальной валюты страны.

6 мая Центральный банк страны распространил циркуляр для финансовых учреждений, включая коммерческие, негосударственные, беспроцентные банки и банки, предоставляющие платёжные услуги, которым объявил о введении сбора в размере 0,5 % на электронные транзакции в целях финансирования кибербезопасности. Регулятор поручил начать сбор налога в течение двух недель.

Гражданское общество выступило против нововведения. 9 мая группа Нигерийского экономического саммита (NESG) выразила обеспокоенность по поводу сроков введения сбора. В своём заявлении группа призвала фе-

деральное правительство пересмотреть налог из-за опасений по поводу множественного налогообложения и инфляционного давления, отягощающего нигерийцев. Эксперты подчеркнули, что в условиях растущей инфляции, финансовой изоляции, роста стоимости жизни и курсов валют сроки введения сбора на кибербезопасность являются неподходящими.

14 мая правительство официально сообщило о приостановке реформы налогообложения, об этом заявил министр информации Мохаммед Идрис.

Первоначально сбор на кибербезопасность в Нигерии был предложен в законе 2015 г. Средства планировалось направить в Национальный фонд кибербезопасности и использовать для укрепления потенциала ИБ Нигерии. Ожидалось, что этот сбор будет приносить около 3 трлн найр (\$ 1,9 млрд ежегодно, однако правительство не представило никакого финансового расчёта. Реализация этой нормы в мае 2024 г. в итоге признана несвоевременной.

Нигерия — одна из трёх крупнейших экономик Африки. Сегодня страна переживает серьёзный экономический кризис. Годовая инфляция превышает 30 %, падают иностранные инвестиции, растут расходы на жизнь. Средний гражданин Нигерии с трудом может позволить себе купить основные продукты питания.

На этом фоне страна является центром киберпреступности, особенно мошенничества с использованием социальной инженерии и программ-вымогателей. Ухудшение экономической ситуации может привести к увеличению киберрисков для граждан и бизнеса, отмечают наблюдатели. Выходом из ситуации могут стать в первую очередь экономические реформы.

ЖЕНСКИЕ ОРГАНИЗАЦИИ ТРЕБУЮТ КИБЕРЗАЩИТЫ

Системные проблемы могут поставить под угрозу безопасность женщин при внедрении искусственного интеллекта (ИИ), а гендерные предубеждения в широко используемых системах ИИ создают серьёзное препятствие для позитивного использования ИИ в контексте мира и безопасности в странах Юго-Восточной Азии, говорится в исследованиях, проведённых структурой «ООН-женщины» и Университетским институтом ООН в Макао (UNU Macau).

Женские правозащитные организации региона подвергаются высокому риску столкновения с киберугрозами. И хотя они осознают

эти риски, такие организации не способны подготовиться к ним или отразить кибератаки. Это один из ключевых выводов исследования «ИИ, женщины, мир и кибербезопасность в странах Юго-Восточной Азии», проведённого при поддержке правительства Австралии в рамках Программы сотрудничества в области кибербезопасности и критических технологий (СТССР) Министерства иностранных дел и торговли (DFAT) и правительства Республики Корея в рамках инициативы «ООН-женщины».

По прогнозам, к 2030 году ИИ добавит 1 трлн долларов США к ВВП стран Юго-Восточной Азии. Понимание влияния новых технологий на повестку дня женских организаций имеет решающее значение для оказания поддержки странам региона в регулировании технологий и снижении их рисков.

Исследователи рассматривают возможные риски, уделяя особое внимание четырём типам гендерных предубеждений в ИИ: дискриминации, стереотипам, изоляции и отсутствию безопасности. Их необходимо решить прежде, чем регион сможет в полной мере извлечь выгоду из новых технологических разработок, считают эксперты ООН. В исследовании рассматривается взаимосвязь между ИИ и женскими кампаниями в соответствии с тремя типами применения ИИ: мирный, нейтральный и конфликтный.

В отчёте отмечается, что во всех категориях существуют благоприятные и неблагоприятные последствия ИИ для гендерно-чувствительного мира и участия женщин в миротворческих усилиях. И хотя использование ИИ в мирных целях имеет множество преимуществ, в том числе для повышения инклюзивности и эффективности предотвращения конфликтов, отслеживания доказательств нарушений прав человека, он используется неравномерно между полами. Широко распространённые гендерные предубеждения снижают вероятность того, что женщины получают выгоду от применения новых технологий. Отдельно подчёркиваются риски, связанные с использованием технологий ИИ в военных целях. Для улучшения динамики использования ИИ в регионе необходимо снижение рисков, связанных с системами ИИ для продвижения повестки дня правозащитниц, особенно в социальных сетях, с помощью чат-ботов и мобильных приложений, и содействие развитию инструментов ИИ, специально созданных для поддержки гендерно-чувствительного

мира в соответствии с пожеланиями женского правозащитного движения.

В докладе «Угрозы кибербезопасности, уязвимости и устойчивости среди женщин-правозащитников и гражданского общества в Юго-Восточной Азии» исследуются риски и уязвимости кибербезопасности с целью повышения киберустойчивости, а также прав человека и цифровых прав женщин в странах региона. Несмотря на растущую осведомлённость о рисках, с которыми сталкиваются в виртуальном пространстве женщины и девочки, эксперты оценивают как слабое понимание специалистами влияния гендера на кибербезопасность, процессы и практики, используемые для защиты цифровых систем и сетей от киберрисков, и наносимый этими рисками вред.

Исследование фокусируется на кибербезопасности, ориентированной на человека, и уделяет особое внимание человеческому фактору, а не техническим навыкам, отмечают эксперты. Они полагают, что киберугрозы носят гендерный характер, в результате чего правозащитницы подвергаются особым нападениям из-за направленности их работы и, вероятно, атакуются женоненавистниками с целью сексуального притеснения.

Результаты исследований подчёркивают, что цифровые технологии занимают центральное место в работе женских правозащитных объединений, которые более чувствительны к восприятию киберугроз и рисков по сравнению с CSO, которые не занимаются гендерными вопросами и правами женщин. Наибольшие различия в уровне угроз между группами касались онлайн-преследований, троллинга и доксинга (распространение в интернете чувствительной информации о персоне без разрешения).

Эксперты рекомендуют содействовать инклюзивным и совместным подходам к разработке требований и политик кибербезопасности, средств предотвращения угроз и реагирования на кибератаки для правозащитных объединений с учётом опыта женских правозащитных движений. Особое внимание следует уделять организациям, подвергающимся риску и действующим в политически нестабильных, конфликтных и кризисных ситуациях.

Презентация исследований состоялась во время молодёжной конференции Структуры «ООН-женщины» «Ген-Форум 2024: молодые лидеры для женщин, мира и безопасности в Азиатско-Тихоокеанском регионе», которая прошла в Таиланде.

ОДКБ. РЕСПУБЛИКА ТАДЖИКИСТАН

ОБЗОР ЗАКОНОДАТЕЛЬСТВА В ОБЛАСТИ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

**Александр
ВИНОГРАДОВ**
ФГУП «ЦНИИХМ»

Сергей МЕНЬШИКОВ
основатель Belarusian
InfoSecurity Community

Андрей СИТНОВ
независимый
эксперт

Валерий СОКОЛЕНКО
заместитель начальника
службы информационной
безопасности ICBC в г. Алматы

Следующий обзор законодательных документов в области информационной безопасности (ИБ) продолжает государство, также член Организации Договора о коллективной безопасности (ОДКБ), Республика Таджикистан (далее – РТ). Государственная политика в этой республике в сфере информационной и кибербезопасности регулируется следующими нормативными актами и развивается по следующим направлениям:

БАЗОВЫЕ АКТЫ

Базовыми актами, закрепляющими общие положения о доступе к информации, конфиденциальности и защите информации, являются следующие документы:

Закон РТ от 2 декабря 2002 года № 71 «О защите информации». Закон устанавливает основополагающие принципы обеспечения защиты информации и регулирования правовых отношений, возникающих в этой области, определяет общие требования к защите информации в РТ, функции Государственного органа по сертификации средств защиты информации. В качестве объекта защиты Закон устанавливает документированную информацию, по отношению к которой установлены определённые правила и ограничения её использования законодательством РТ, владельцем или собственником такой информации.

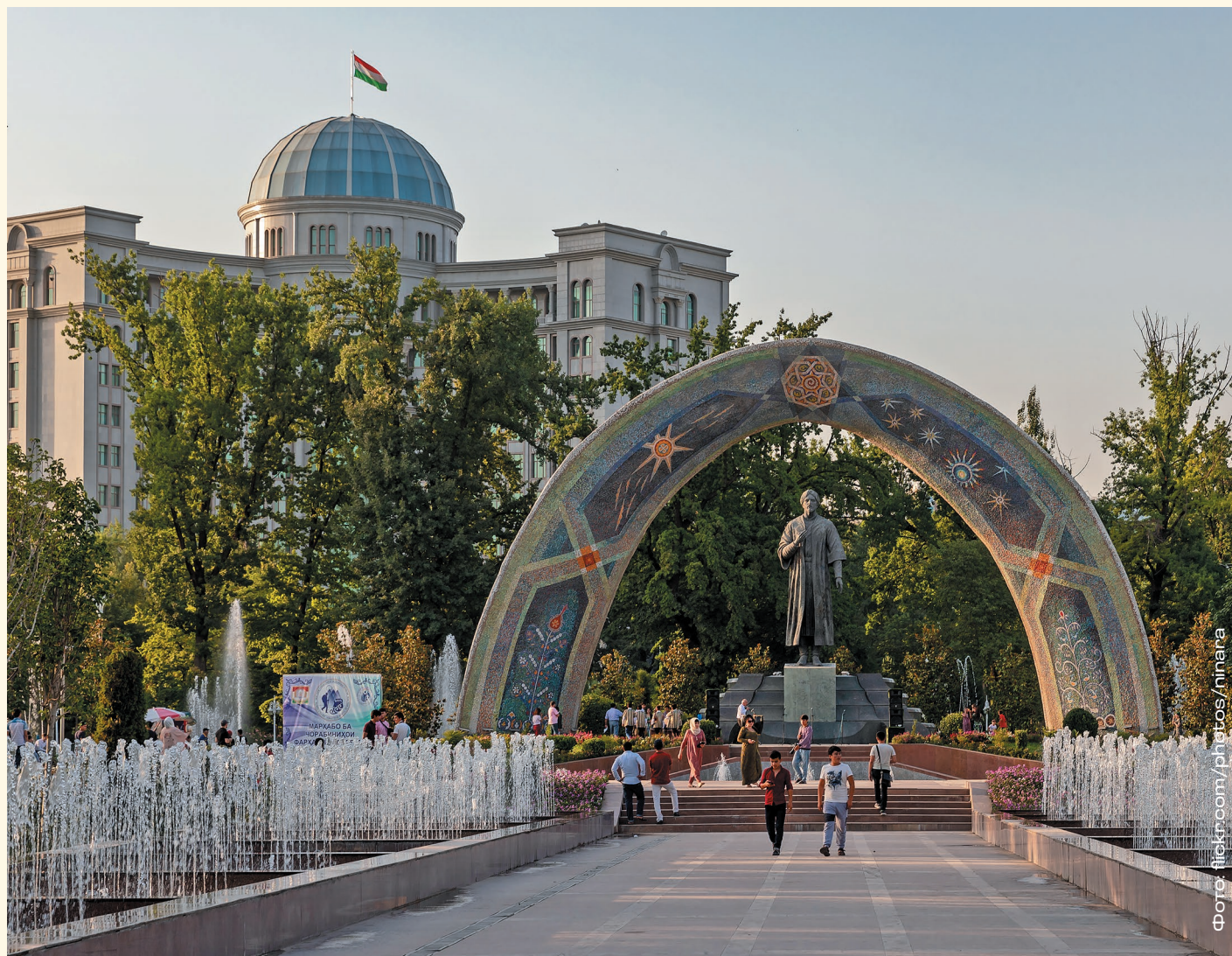
Закон РТ в редакции от 03.07.2012 № 848 «Об информации». Данный Закон определяет правовые основы обеспечения безопасности человека и гражданина, общества и го-

сударства, регулирует систему обеспечения безопасности, её функции, цели и основные направления, устанавливает полномочия Президента Республики Таджикистан, органов государственной власти, органов самоуправления посёлков и сёл и организаций, независимо от их организационно-правовой формы в области обеспечения безопасности.

Закон классифицирует информацию в зависимости от порядка её предоставления или распространения на:

- ♦ информацию, которая в соответствии с законами подлежит предоставлению или распространению;
- ♦ информацию, распространение которой в РТ ограничивается или запрещается.

Данный закон также определяет в качестве отдельной категории информацию о личности (персональные данные). При этом указывается, что основными данными о личности являются национальность, образование, семейное положение, материальное положение, религиозная принадлежность, состояние здоровья, а также адрес и место рождения. Источниками документированной информации о личности являются выданные на её имя документы, подписанные ею документы, а также сведения о личности, собранные государственными и местными органами власти. Запрещается сбор, хранение, использование и распространение информации о частной жизни, а равно информации, разглашающей личную и семейную тайну, тайну телефонных переговоров, почтовых, телеграфных и иных сообщений личности без её на то согласия, кроме случаев, предусмотренных законодательством РТ.



ОБЩИЕ СВЕДЕНИЯ

Республика Таджикистан — государство в Центральной Азии, расположено в предгорьях Памира, не имеет выхода к морю. По площади — наименьшее государство Средней Азии. Граничит с Узбекистаном на западе и северо-западе, на севере — с Кыргызстаном, на востоке — с Китаем, на юге — с Афганистаном.

Столица — Душанбе. Государственный язык — таджикский; русский язык является языком межнационального общения. Таджикистан — единственное персоязычное государство в бывшей советской Средней Азии.

Таджикистан является членом ООН, ШОС, СНГ, ЕврАзЭС, Организации Исламского Сотрудничества (ОИК), МВФ, ОБСЕ, ВТО.

Органы, утверждающие и согласовывающие законы и подзаконные акты в области ИБ: Президент РТ, Правительство РТ, Маджлиси милли Маджлиси Оли¹.

Все рассмотренные в нашей публикации документы имеют общереспубликанское значение и применяются для организации и обеспечения информационной и кибербезопасности.

¹ Национальный совет Высшего собрания Республики Таджикистан — верхняя палата Маджлиси Оли (Высшего собрания) Республики Таджикистан.

КОНЦЕПТУАЛЬНЫЕ ДОКУМЕНТЫ

Государственная политика РТ в сфере информационной и кибербезопасности регулируется следующими актами:

Указ Президента РТ от 7 ноября 2003 г. № 1175 «О концепции информа-

ционной безопасности Республики Таджикистан».

Концепция информационной безопасности Республики Таджикистан представляет собой совокупность официальных взглядов на цели, задачи, принципы и основные направления обеспечения ин-

формационной безопасности Республики Таджикистан.

Концепция служит основой для:

- ♦ формирования государственной политики в области обеспечения национальной информационной безопасности Республики Таджикистан;
- ♦ подготовки предложений по совершенствованию правового, методического, научно-технического и организационного обеспечения национальной информационной безопасности Республики Таджикистан;
- ♦ разработки целевых программ обеспечения информационной безопасности Республики Таджикистан.

Постановление Правительства РТ от 30 июня 2004 года № 290 «Программа обеспечения информационной безопасности Республики Таджикистан». Программа обеспечения информационной безопасности Республики Таджикистан (далее — Программа) разработана в соответствии с Указом Президента РТ от 7 ноября 2003 года № 1175 «О Концепции информационной безопасности Республики Таджикистан».

Программа направлена на реализацию и поддержку основ государственной политики по обеспечению информационной безопасности, как главная составная часть национальной безопасности.

Общегосударственный уровень важности проблемы обеспечения информационной безопасности требует концентрации усилий научного и технического потенциала органов государственной власти, предприятий, учреждений и организаций Республики Таджикистан, что определяет государственный характер Программы.

АДМИНИСТРАТИВНОЕ РЕГУЛИРОВАНИЕ

Административное регулирование деятельности в сфере сертификации и технической защиты информации, в том числе связанной с оборотом криптографических средств регламентируется следующими актами:

Закон РТ от 3 июля 2012 года № 839 «О криптографии». Закон предусматривает упорядочение общественных отношений в сфере образования и использования криптографии в процессе обмена и сохранения информации путём коммуникационно-автоматизированных информационных каналов.

Закон РТ от 15 марта 2023 года № 1965 «Об электронном документе и электронной подписи». Данный Закон уста-

навливает правовые и организационные основы, порядок оборота и использования электронных документов, а также создания и применения электронной подписи в процессе формирования и использования электронных документов. Закон определяет три вида электронных подписей, которые могут использоваться при оформлении документов и сделок: простая электронная подпись, усиленная электронная подпись и защищённая электронная подпись. Защищённая электронная подпись используется только при подписании защищённых электронных документов. Защищённые электронные подписи создаются удостоверяющими центрами на основании заключённого договора с участником информационной системы.

Постановление Правительства РТ от 2 апреля 2015 года № 190 «Положение о требованиях, условиях и правилах защиты информации от иностранной технической разведки». Положение подготовлено с целью реализации статьи 14 Закона РТ «О защите информации» и пункта 24 Перечня мероприятий Государственной программы развития и внедрения информационно-коммуникационных технологий в РТ на 2014–2017 годы, утверждённой постановлением Правительства РТ от 3 июля 2014 года № 428.

Данное Положение определяет основные каналы утечки информации, содержащей государственные секреты и другие охраняемые законом виды информации в военной, экономической, политической, научной и технической сферах РТ (далее — защищаемая информация) к иностранной технической разведке, регулирует требования, условия и правила защиты информации от иностранной технической разведки.

Постановление Правительства РТ от 1 октября 2004 года № 404 «Положение о сертификации средств защиты информации по требованиям безопасности информации, аттестации объектов информатизации, порядка их государственной регистрации». Положение разработано на основании Закона РТ «О государственной тайне», Закона РТ «О защите информации», Закона РТ «О сертификации продукции и услуг» и регулирует взаимоотношения в данной области.

ТАЙНЫ

Тайны, определённые законодательством РТ:

Закон РТ от 18 июня 2008 года № 403 «О коммерческой тайне». Закон регули-

рует отношения, связанные с отнесением информации к коммерческой тайне, передачей такой информации, охраной её конфиденциальности в целях обеспечения баланса интересов обладателей информации, составляющей коммерческую тайну, и других участников регулируемых отношений, в том числе государства на рынке продукции, труда и услуг. Определение понятия коммерческой тайны полностью соответствует определению аналогичного закона Российской Федерации.

Закон РТ от 26 июля 2014 года № 1095 «О государственных секретах». Данный Закон определяет правовые основы и систему защиты государственных секретов в интересах обеспечения национальной безопасности РТ, регулирует общественные отношения, связанные с отнесением сведений к государственным секретам, их засекречиванием, распоряжением, защитой и рассекречиванием.

Закон РТ от 28 июня 2011 года № 721 «О Безопасности». Настоящий Закон устанавливает общие правовые нормы получения, использования, распространения и хранения информации, закрепляет право субъекта информационных отношений на информацию во всех сферах общественной и государственной жизни Республики Таджикистан, а также систему информации, её источники, определяет статус участников информационных отношений, регулирует доступ к информации и обеспечивает её защиту, защищает личность и общество от ложной информации.

Закон РТ от 3 августа 2018 года № 1537 «О защите персональных данных». Закон определяет правовые и организационные основы деятельности, связанной со сбором, обработкой и защитой персональных данных. В отличие от российского закона «О персональных данных», в соответствии с которым к ПД относят широкий круг информации — любая информация, относящаяся к прямо или косвенно определённой или определяемому физическому лицу, в Законе РТ к персональным данным относят сведения о фактах, событиях и обстоятельствах жизни субъекта персональных данных, позволяющие идентифицировать его личность. То есть упор делается на возможность идентифицировать личность по персональным данным.

Закон РТ от 19 мая 2009 года за № 524 «О банковской деятельности». Регламентирует понятие и защиту банковской тайны, под которой понимаются сведения о дея-

тельности и финансовом положении клиента, которые стали известны кредитной организации при обслуживании и отношении с клиентом или с третьим лицом, разглашение которых может причинить клиенту материальный или моральный ущерб. В частности, к банковской тайне относятся сведения о наличии банковских счетов, их владельцев, о наличии в них денежных средств и об операциях, проводимых клиентами по этим счетам, о денежных переводах физических лиц без открытия счёта и об их отправителях (получателях), о денежных средствах и других ценностях клиента, которые хранятся в кредитной организации.

Постановление Правления Национального банка Таджикистана от 19 ноября 2021 года № 167 «Инструкция № 246 О требованиях по использованию информационно-коммуникационных технологий и обеспечению информационной безопасности при организации деятельности бюро кредитных историй». Инструкция разработана в соответствии с частью 2 статьи 7 Закона РТ «О кредитных историях» и определяет требования по использованию информационно-коммуникационных технологий, а также обеспечению информационной безопасности при организации деятельности Бюро кредитных историй.

КИБЕРПРЕСТУПНОСТЬ

Правовое регулирование в сфере борьбы с киберпреступностью:

Уголовный кодекс РТ от 21 мая 1998 года № 574

Содержатся следующие составы преступлений, которые можно отнести к киберпреступности:

- ♦ неправомерный доступ к информации, хранящейся в компьютерной системе, сети или на машинных носителях, сопровождающийся нарушением системы защиты, те же деяния, повлёкшие по неосторожности изменение, уничтожение либо блокирование информации, а равно вывод из строя компьютерного оборудования либо значительный ущерб;

- ♦ изменение информации, хранящейся в компьютерной системе, сети или на машинных носителях, а равно внесение в них заведомо ложной информации, причинившее значительный ущерб или создавшее угрозу его причинения, то же деяние, сопряжённое с неправомерным доступом к компьютерной системе или сети;

- ♦ уничтожение, блокирование либо приведение в непригодное состояние компьютерной информации или программы, вывод из строя компьютерного оборудования, а равно разрушение компьютерной системы, сети или машинного носителя;

- ♦ незаконное копирование или иное неправомерное завладение информацией, хранящейся в компьютерной системе, сети или на машинных носителях, а равно перехват информации, передаваемой с использованием компьютерной связи;

- ♦ принуждение к передаче информации, хранящейся в компьютерной системе, сети или на машинных носителях, под угрозой оглашения позорящих сведений о лице или его близких, предания гласности сведений о таких обстоятельствах, которые потерпевший желает сохранить в тайне, а равно под угрозой применения насилия над лицом или его близкими либо под угрозой уничтожения или повреждения имущества лица, его близких и других лиц, в ведении или под охраной которых находится эта информация;

- ♦ изготовление с целью сбыта, а равно сбыт специальных программных или аппаратных средств для получения неправомерного доступа к защищённой компьютерной системе или сети;

- ♦ разработка компьютерных программ или внесение изменений в существующие программы с целью несанкционированного уничтожения, блокирования, модификации или копирования информации, хранящейся в компьютерной системе, сети или на машинных носителях, а также разработка специальных вирусных программ, заведомое их использование или распространение носителей с такими программами;

- ♦ нарушение правил эксплуатации компьютерной системы или сети лицом, имеющим доступ к этой системе или сети, если это повлекло по неосторожности уничтожение, блокирование, модификацию компьютерной информации, нарушение работы компьютерного оборудования или причинение иного значительного ущерба.

СТАНДАРТЫ

Национальные стандарты РТ:

- ♦ СТ РТ 1070–2005 «Средства вычислительной техники. Защита от несанкционированного доступа к информации. Общие технические требования»;

- ♦ СТ РТ ИСО 21188–2012 «Инфраструктура открытых ключей для сферы финансовых

услуг. Практические приёмы и структура политики»;

- ♦ СТ РТ ISO/IEC 20000–1–2012 «Информационные технологии. Менеджмент услуг. Часть 1. Технические требования»;

- ♦ СТ РТ ИСО/МЭК 27001–2013 «Информационная технология. Методы и средства обеспечения безопасности. Системы менеджмента информационной безопасности. Требования»;

- ♦ СТ РТ 1050–2002 «Информационная технология. Защита информации от утечки за счёт побочных электромагнитных излучений при её обработке средствами вычислительной техники. Общие технические требования и методы испытаний»;

- ♦ СТ РТ 1049–2002 «Защита информации. Оценка качества программных средств. Общие положения»;

- ♦ СТ РТ 1046–2002 «Защита информации. Автоматизированные системы управления. Виды испытаний»;

- ♦ СТ РТ 1047–2002 «Защита информации. Автоматизированные системы управления. Общие требования»;

- ♦ СТ РТ 1045–2002 «Защита информации. Защита от несанкционированного доступа к информации. Показатели защищённости от несанкционированного доступа к информации»;

- ♦ СТ РТ 1051–2002 «Защита информации. Информационная технология. Методы испытаний изделий электронной техники»;

- ♦ СТ РТ 1074–2005 «Защита информации. Испытания программных средств на наличие компьютерных вирусов. Типовое руководство»;

- ♦ СТ РТ 1048–2002 «Защита информации. Нормы эффективности защиты средств передачи, обработки, хранения и отображения секретной информации»;

- ♦ СТ РТ 1072–2005 «Защита информации. Объект информатизации. Факторы, воздействующие на информацию. Общие положения»;

- ♦ СТ РТ 1069–2005 «Защита информации. Основные термины и определения».

Надеемся, что этот обзор поможет вам не «заблудиться» в Законодательстве РТ. Мы постарались систематизировать и тем самым облегчить путь изучения и применения на практике нормативных документов в области ИБ одного из наших соседей, члена ОДКБ – Республики Таджикистан.

«МЫ РОЖДЕНЫ ДЛЯ ВДОХНОВЕНИЯ!»

225 ЛЕТ СО ДНЯ РОЖДЕНИЯ А. С. ПУШКИНА



Борис СТОЛПАКОВ
историк криптографии

Шестого июня 2024 г. исполняется 225 лет со дня рождения А. С. Пушкина. Одновременно шестое июня — это и День русского языка, учреждённый Организацией Объединённых Наций в 2010 г. Ведь русский язык — один из шести официальных языков ООН и ЮНЕСКО. Время для праздника выбрано вполне закономерно. Именно Пушкин давно и по праву признан создателем русского литературного языка.

«ОНА ПО-РУССКИ ПЛОХО ЗНАЛА...»

Вероятно, оценить по достоинству вклад Пушкина в развитие русского литературного языка возможно, лишь представив себе, каким было читательское сообщество и его литературный язык в пушкинское время, то есть в 20–30-е годы XIX века.

Как известно, в ту пору многие представители образованного слоя, прежде всего дворяне, и говорили, и читали, и писали на французском. Добавим, что в российских городах начиная с конца XVIII столетия проживало немало французов-эмигрантов.

Сам Александр Сергеевич напоминает в «Евгении Онегине»:

*«...Ещё предвижу затрудненья: родной земли спасая честь,
Я должен буду, без сомненья, письмо Татьяны перевести.
Она по-русски плохо знала, журналов наших не читала
И выражалась с трудом на языке своём родном...».*

«ЖАЛОБЫ БЕСПОЛЕЗНЫ»

Даже Н. М. Карамзин заявлял: «Жалобы бесполезны. Связь между умами древних и новейших россиян прервалась навеки. Мы не хотим подражать иноземцам, но пишем, как они пишут, ибо живём, как они живут...»

Министр народного просвещения А. С. Шишков пытался сопротивляться: «Где чужой язык употребляется предпочтительнее своего, где чужие

книги читаются более, нежели свои, там при безмолвии словесности всё вянет и не процветает... Доколе мы не возлюбим языка своего, обычаев своих, до тех пор во многих наших науках и художествах будем мы далеко позади других. Надобно жить своим умом, а не чужим».

«ВСЛУШИВАЙТЕСЬ В ПРОСТОНАРОДНОЕ НАРЕЧИЕ, МОЛОДЫЕ ПИСАТЕЛИ»

Вот в таком окружении А. С. Пушкин вступил на литературное поприще, вполне закономерно представ перед читателями как западник. Но уже в 20-е годы он предложил и убедительно продемонстрировал иной путь: «У нас ещё нет ни словесности, ни книг, все наши знания, все наши понятия с младенчества почерпнули мы в книгах иностранных, мы привыкли мыслить на чужом языке; <...> учёность, политика и философия ещё по-русски не изъяснялись, <...> леность наша охотнее выражается на языке чужом.

Всё должно творить в этой России и в этом русском языке. Живой народный язык, сберёгший в жизненной свежести дух, который придаёт языку стойкость, силу, ясность, целостность и красоту, должен послужить источником и сокровищницей для развития образованной русской речи.

Вслушивайтесь в простонародное наречие, молодые писатели, — вы в нём можете научиться многому, чего не найдёте в наших журналах...»

«АЙ ДА ПУШКИН!»

Удивительно, как быстро и надёжно и сам Пушкин, и его последователи сумели преобразить и развить литературную русскую речь! Но для этого они и сами должны были преобразиться. Так, духовно созревший Пушкин отвратился от философов — «просветителей»: «Гуманизм сделал французов язычниками». Собственные духовные перемены стали неожиданными для Александра Сергеевича. «Ай да Пушкин, ай да сукин сын!» — записал он в дневнике после завершения «Бориса Годунова» в конце 1825 г. Вот его мысли конца 20-х — начала 30-х годов: «Не должно торопить времени, и без того уже довольно деятельного. Лучшие и прочнейшие изменения суть те, которые происходят от улучшения нравов, без насильственных потрясений политических, страшных для человечества». Те же мысли Пушкин вложил в уста Петра Андреевича Гринёва в «Капитанской дочке». Болью за духовное состо-



АЛЕКСАНДРЪ ПУШКИНЪ

Гравированный портрет А. Пушкина. Работа Е. И. Гейтмана, 1822 г., по акварельному портрету, сделанному учителем рисования на уроке в Лицее

яние общества пронизано замечание Александра Сергеевича: «Ныне нет в Москве мнения народно-го: ныне бедствия или слава отечества не отзываются в этом сердце. Грустно было слышать толки московского общества во время последнего польского возмущения. Гадко было видеть бездушного читателя французских газет, улыбающегося при вести о наших неудачах».

МЫ ПОВТОРЯЕМ ТЕПЕРЬ ЕЩЁ БЕССМЫСЛЕННО СЛОВО «ПРОСВЕЩЕНИЕ»

Пушкину вторил Гоголь: «Мы повторяем теперь ещё бессмысленно слово «просвещение». Даже и не задумывались над тем, откуда пришло это слово и что оно

значит. Слова этого нет ни на каком языке, оно только у нас. Просветить не значит научить, или наставить, или образовать, или даже осветить, но всего насквозь высветлить человека во всех его силах, а не в одном уме, пронести всю природу его сквозь какой-то очистительный огонь. Слово это взято из нашей церкви».

Н. В. Гоголь уже в 1832 г. сказал об Александре Сергеевиче: «В нём, как будто в лексиконе, заключилось всё богатство, сила и гибкость нашего языка. Он более всех, он далее раздвинул ему границы и более показал всё его пространство. Пушкин есть явление чрезвычайное и, может быть, единственное явление русского духа». «Нужно было, чтобы выболтали мы на чужеземных наречьях всю дрянь, какая ни пристала к нам вместе с чужеземным образованием, чтобы все те неясные звуки, неточные названия вещей, — дети мыслей невыяснившихся и сбивчивых, которые потемняют языки, — не посмели бы помрачить младенческой ясности нашего языка и возвратились бы мы к нему, уже готовые мыслить и жить своим умом, а не чужеземным. Пройдёт эта речь уже насквозь всю душу и, ударивши по всем струнам, какие ни есть в русском человеке, вызовет нам нашу Россию, — нашу русскую Россию, не ту, которую показывают нам грубо какие-нибудь квасные патриоты, и не ту, которую вызывают к нам из-за моря очужеземившиеся русские, но ту, которую извлечёт она из нас же и покажет таким образом, что все до единого, каких бы ни были они различных мыслей, образов воспитания и мнений, скажут в один го-

лос: «Это наша Россия; нам в ней приятно и тепло, и мы теперь действительно у себя дома, под своей родной крышей, а не на чужбине».

«ЧЕРЕЗ НЕГО УМНЕЕТ ВСЁ, ЧТО МОЖЕТ ПОУМНЕТЬ...»

Драматург А. Н. Островский сказал о Пушкине: «Через него умнеет всё, что может поумнеть... Он дал всякой оригинальности смелость, дал смелость русскому писателю быть русским... раскрыл русскую душу... Следующее поколение, воспитанное исключительно Пушкиным, когда сознательно оглянулось назад, увидало, что предшественники его и многие его современники для них уже даже не прошедшее, а далеко давнопрошедшее».

И. С. Тургенев в знаменитой речи на открытии памятника А. С. Пушкину в Москве в 1880 г. подчеркнул, что поэту «одному пришлось исполнить две работы, в других странах разделённые целым столетием и более, а именно: установить язык и создать литературу». «Он дал окончательную обработку нашему языку, который теперь по своему богатству, силе, логике и красоте формы признаётся даже иностранными филологами едва ли не первым после древнегреческого».

ОТКРЫТИЕ ПУШКИНА

Иван Сергеевич Шмелёв, которому было семь лет в 1880 г., написал о своём открытии Пушкина: «В книжечке Ступина увидал я красивый кораблик с парусами и прочитал по складам стишки:

Ветер по морю гуляет и кораблик подгоняет.

Он бежит себе в волнах, на раздутых парусах.

Под ними я прочитал: «Пу-шкин». Потом «Пушкин» связался во мне с зимой, со снегом на нашем дворе, с сарайчиком, где стояли мои лубяные салазки.

...Я уже учился в гимназии, в четвёртом классе. Приходил к нам раз в месяц «наводить бухгалтерию» старший участковый паспортист. Его всегда угощали перед работой — больше стараться будет. Как-то пришёл он после Рождества и объявил нам: На подписку собираю, по знакомству, лавошникам даже рекомендую... Всего Пушкина за полтора целковых, цельную кучу книжек, дешевле пареной репы! По случаю смерти Пушкина. Кануло пятьдесят лет, и теперь все могут печатать. Не поскучите для потомства, в память знаменитого человека! Вот, подпишитесь на листочке, напишите-ка две строчки...

И вот, как-то весной, почтальон приносит тючок в рогожке. А когда его распорол, оттуда посыпались пухлыми кирпичиками чудесно пахнущие свежей краской томики «Полного собрания сочинений А. С. Пушкина». И тогда я открыл его, от детского портрета — крутолобого кудрявого мальчугана — до Памятника ему, открыл до конца, всего.

И всего его прочитал и перечитал. Встретил и «про кораблик», и «про зиму», пережил снова первую с ним детскую свою встречу, с незабываемой свежестью и радостями неосознанных и загадочных ощущений, необъяснимых и до сегодня».

БЛАГОДАРНАЯ ПАМЯТЬ

Сергей Вавилов, 20-летний студент-физик, будущий президент Академии наук СССР, записал в своём дневнике:

«4/17 марта 1911... Я знаю, что всю жизнь буду я под действием двух полюсов «зеркального» существования: наукой и Пушкиным».

В 1937 г. Пушкинские памятные торжества прошли в 231 городе 42 стран, в СССР они стали крупным государственным мероприятием.

СЛУЧАЙ — МОЩНОЕ, МГНОВЕННОЕ ОРУДИЕ ПРОВИДЕНИЯ...

Вскоре после гибели Пушкина его друг А. И. Тургенев вспоминал: «Я находил в нём сокровища таланта, наблюдений и начитанности о России, особенно о Петре и Екатерине, редкие, единственные. Никто так хорошо не судил русскую новейшую историю: он созревал для неё и знал и отыскал в известность многое, чего другие не заметили... Ему оставалось дополнить и передать бумаге свои сведения».

Любителям точных и гуманитарных наук можно предложить для размышлений пушкинский набросок рецензии 1830 г. на издание «Истории русского народа»:

«Россия никогда ничего не имела общего с остальной Европою; история её требует другой мысли, другой формулы. Не говорите: иначе нельзя было быть. Коли было бы это правда, то историк был бы астроном, и события жизни человечества были бы предсказаны в календарях, как и затмения солнечные. Но Провидение не алгебра. Ум человеческий, по простонародному выражению, не пророк, а угадчик, он видит общий ход вещей и может выводить из одного глубокие предположения, часто оправданные временем, но невозможно ему предвидеть случая — мощного, мгновенного орудия Провидения...»

ОН ДАЛ НАМ ВОЗМОЖНОСТЬ, И ОСНОВАНИЕ, И ПРАВО...

Философ И. А. Ильин: «Кто хочет понять Пушкина и его восхождение к вере и мудрости, должен всегда помнить, что он всю жизнь прожил в той непосредственной, прозрачной и нежно-чувствующей детскости, из которой молится, поёт, плачет и пляшет русский народ...»

Он дал нам возможность, и основание, и право верить в призвание и в творческую силу нашей родины, благословлять её на всех её путях и прозревать её светлое будущее, — какие бы ещё страдания, лишения или унижения ни выпали на долю русского народа».

Русский язык за 200 лет, с 30-х годов XIX века, не претерпел каких-либо существенных изменений. Язык, на котором говорил и писал Пушкин, по сей день понятен и общеупотребим на обширной территории нашего Отечества и за её пределами. Хочется верить, что и завтра, как и сегодня, грамотная русская речь будет свободно звучать и печататься на языке, близком к пушкинскому, напоминая нам об этом великом Божиим даром.

«БЕЛОРУССКИЙ МОТИВ В ПЕСНЕ ВЕРЕСКОВ, В ПЕСНЕ РАКИТ...»

СОВЕТСКАЯ ПАРТИЯ В ОПЕРАЦИИ «БОДИГАРД»



Борис СТОЛПАКОВ
историк криптографии

Лето 1944 г. стало последним военным летом в Европе, его события вселили и укрепили общую надежду на близкое завершение ужасов войны. Наступление войск антигитлеровской коалиции началось 6 июня операцией «Оверлорд» – высадкой союзников в Нормандии. Тогда же Сталин направил Черчиллю послание, где информировал британского премьер-министра: «Ваше сообщение об успехе начала операции „Оверлорд“ получило. Оно радует всех нас и обнадеживает относительно дальнейших успехов. Летнее наступление советских войск, организованное согласно уговору на Тегеранской конференции, начнется к середине июня на одном из важных участков фронта <...>. Обязуюсь своевременно информировать Вас о ходе наступательных операций».

ЗАМЫСЕЛ «МИСТИФИКАЦИИ И ОБМАНА ПРОТИВНИКА»

Далее постараемся раскрыть смысл одной фразы из решений Тегеранской конференции: «Соответствующими штабами должен быть согласован план мистификации и обмана противника». По существу, здесь упомянут план защиты информации о точном месте и времени открытия второ-

го фронта. Ценнейшей информации, для доступа к которой противник ни перед чем бы не остановился. Как же защитить столь важную и роковую для врага информацию? Невозможно скрыть перемещение и накопление сотен тысяч военнотехнических, а также десятков тысяч единиц боевой техники. Значит, усилия надо направить на убедительное внушение разведке противника ложных представлений о намерениях англо-американских войск, т.е. на создание эффективной дезинформации. И этот замысел удался. Разведку немцев дезинформировали мероприятия, проведенные весной 1944 г. в соответствии с планом специальной операции «Бодигард» (англ. – телохранитель).

«ТЕЛОХРАНИТЕЛЬ» ДЕСАНТА В НОРМАНДИИ

Проект плана операции «Бодигард» был разработан в Лондоне, одобрен Черчиллем и после этого в марте 1944 г. предложен руководству ГШ Красной Армии для согласования и разработки дополнительных мероприятий, касающихся советско-германского фронта. Надо сказать, что союзники прислали в Москву объяснительную записку к плану ещё в феврале. Британский вариант плана был многоаспектным, он «предусматривал проведение одновременно до 35 дезинформационных операций оперативно-тактического уровня». Были продуманы методики навязывания ложной информации, использования двойных агентов и т.д. Они должны были обеспечить «достижение трёх важных целей».

1. Заставить германское высшее командование поверить в то, что высадка англо-американских войск на побережье Нормандии является лишь начальной стадией вторжения и проводится незначительным количеством союзных войск.

2. Убедить противника в том, что основное вторжение англо-американскими войсками бу-

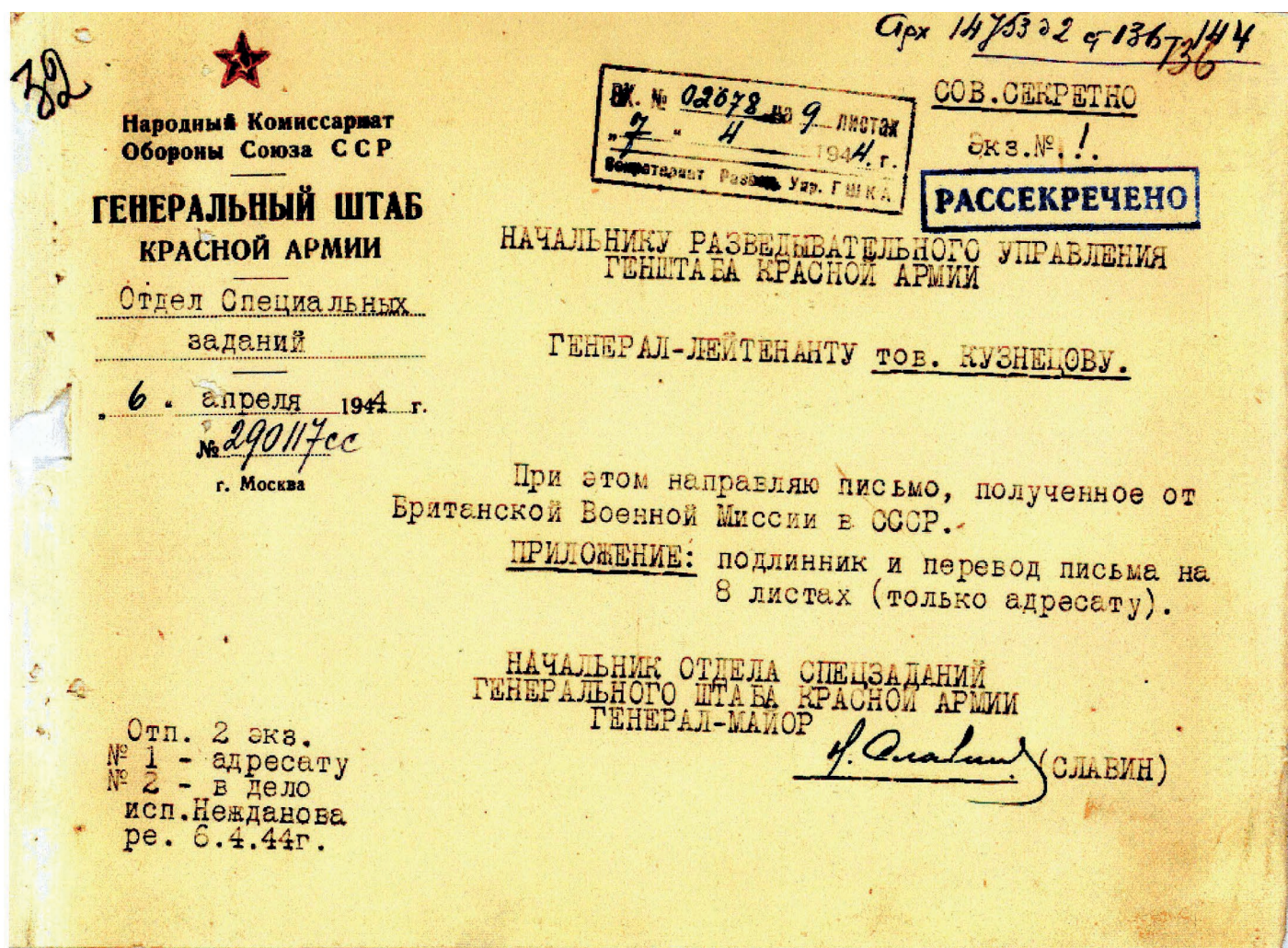


Иллюстрация 1. Сопроводительное письмо Н. В. Славина к письму британской военной миссии от 06.04.1944. С сайта mil.ru

дет проводиться в районе пролива Па-де-Кале или в Норвегии, а сосредоточение союзных войск в Южной Англии является лишь отвлекающим манёвром.

3. Уверить противника, что летом 1944 г. союзники намерены нанести решительный удар на балканском направлении».

«НАПАДЕНИЕ» КАК СРЕДСТВО ЗАЩИТЫ СВОИХ СЕКРЕТОВ

Разработка мероприятий советской части плана «Бодигард» была поручена разведывательным управлениям Генштаба КА и ВМФ.

С целью ускорения согласований планов к весне 1944 г. Великобритания и СССР организовали канал связи: Британский комитет начальников штабов — Британская военная миссия в Москве — г.-м. Н. В. Славин — начальник РУ ГШ КА Ф. Ф. Кузнецов — Генштаб КА (илл. 1).

Согласно замыслу генштабистов, «основные усилия по дезинформации германского командования предстояло предпринять на двух направлениях.

1. Демонстрировать подготовку вторжения совместно с британскими войсками в Норвегию.

2. Обозначить реальную подготовку к широкому наступлению Красной Армии на Балканах».

Окончательный вариант плана «Бодигард» для советско-германского фронта был утверждён Сталиным 17 апреля 1944 г. под кодовым названием «Нападение». Первая роль в нём была отведена Карельскому фронту и Северному флоту. Карельский фронт занимал особое положение среди советских фронтов. Протянувшийся на 1300 км от Баренцева моря до Ладожского озера, он единственный из фронтов не имел непрерывной линии соприкосновения с противником. Его штаб располагался в Беломорске.

Командующим Карельским фронтом и Северным флотом были сообщены непосредственно дополнительные указания, в которых, в частности, говорилось: «В соответствии с пунктом 12 (в) плана „Бодигард“ необходимо заставить противника поверить в предполагаемое „наступление“ Красной Армии на Северную Финляндию или Норвегию во взаимодействии с англо-американскими войсками, группирующимися с такими же задачами в Шотландии, Северной Ирландии и Исландии».

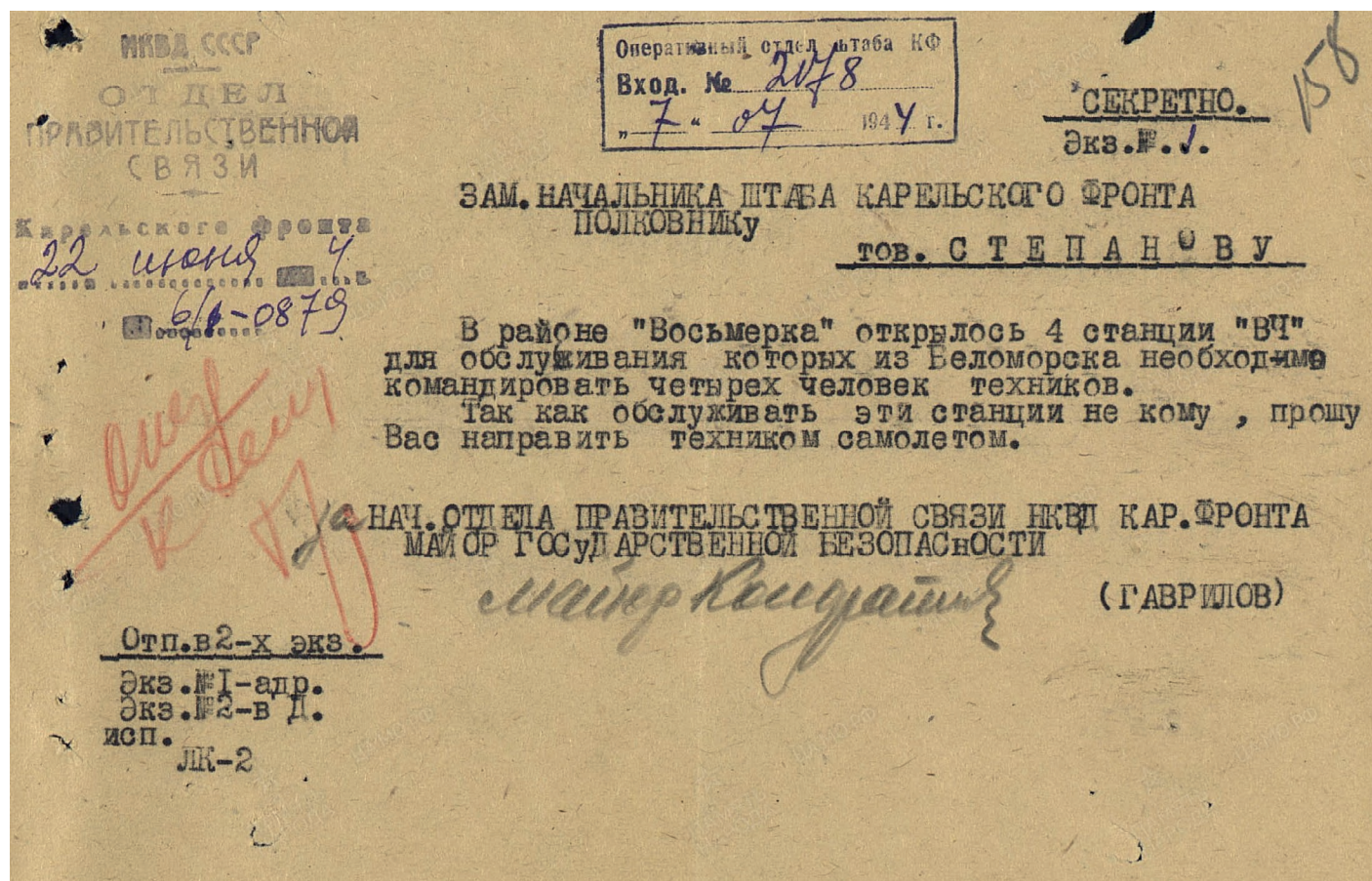


Иллюстрация 2. Донесение о нуждах станций ВЧ-связи

При этом, информируя командование Карельского фронта и СФ, начальник РУ Генштаба писал: «Во взаимодействии с англо-американцами ГШ Красной Армии намечает в качестве объектов нападения Петсамо, Киркенес и Берлевог. Окончательная подготовка к нападению на указанные объекты 1 мая 1944 года». Всё справедливо, так оно и случилось в дальнейшем. Уже 26–28 апреля 1944 г. авиация Северного флота бомбила Берлевог, расположенный в глубине Норвегии, стремясь подавить батареи противника. Но всё же основные удары были нанесены в другие сроки и другими силами.

Использовали для дезинформации и финский центр радиоразведки, расположенный в г. Сортавала. Он активно работал против советских войск и в глазах немцев заслужил авторитет. Немедленно началась реализация предусмотренных планом многочисленных мероприятий. Самыми простыми были, конечно, сосредоточение войск и морских судов, а также усиление радиообмена между радиостанциями флотских и армейских частей. Более значимым стало наступление Красной Армии на Карельском перешейке, а затем в Карелии, начавшееся 9 июня мощной артподготовкой. Об открытии новых станций ВЧ-связи в период наступления на карельском участке фронта свидетельствует рас-

секреченный ныне документ, приведённый на илл. 2.

КРАХ ГРУППЫ АРМИЙ «ЦЕНТР»

Плоды дезинформационной операции «Бодигард» очень пригодились и в главном летнем наступлении Красной Армии 1944 года – в операции «Багратион» против группы армий «Центр». Подтверждающее свидетельство невольно предоставил активный участник тех боёв генерал К. Типпельскирх, командовавший 4-й полевой немецкой армией, входившей в эту армейскую группу. Крупный военный специалист, он уже в первые послевоенные годы подготовил капитальную «Историю Второй мировой войны». Главу, посвящённую событиям лета 1944 г. на просторах Белоруссии, К. Типпельскирх по своим живым впечатлениям озаглавил «Крах немецкой группы армий «Центр».

КРАСНОРЕЧИВОЕ СВИДЕТЕЛЬСТВО

Привожу выдержки из этой главы:

«В середине мая войскам Восточного фронта было вменено в обязанность „при любых обстоятельствах оборонять и удерживать занимаемые ими рубежи“. К этому времени не было ещё никаких данных, которые позволили бы предугадать направление или направления, несомненно, готовившегося летнего наступления русских. <...>

Группам армий „Центр“ и „Север“ предсказывали „спокойное лето“. Относительно того, что первый удар противника последует севернее или южнее Карпат — скорее всего, севернее, — мнение было единодушным. Помимо этого, немецкое командование справедливо предполагало, что русские не начнут наступления до тех пор, пока на Западе не появится сильный второй фронт, который наверняка лишит немцев возможности перебрасывать силы с Запада на Восток».

«На фронте группы армий „Центр“ намерения противника стали выясняться примерно к 10 июня. Именно здесь, где немецкое командование меньше всего ожидало наступления, стали появляться, очевидно, признаки крупных приготовлений русских. Радиоразведка сообщала о новых армиях; авиация отмечала усиление железнодорожных перевозок и интенсивное движение на шоссейных дорогах. Дивизионы артиллерийской и инженерной разведки установили, что на ряде участков фронта начали пристрелку крупные силы перебросенной сюда русской артиллерии. Пленные сообщали о появлении в тылу противника „ударных частей“. <...> Прошло ещё несколько дней, и для командования группы армий „Центр“ стало совершенно очевидным, что противник развёртывает на этом фронте крупные силы. Кроме того, стали отчётливо вырисовываться направления предстоящих ударов на Бобруйск, Могилёв, Оршу и Витебск. Полученная в результате сопоставления самых разнообразных наблюдений картина приготовлений противника была настолько определённой и ясной, что для предположения о возможности имитаций и ввода в заблуждение совершенно не оставалось места.

14 июня у начальника Генерального штаба сухопутных сил состоялось совещание с участием всех начальников штабов армий и групп армий. В то время как начальники штабов группы армий „Север“ и обеих южных групп единодушно сообщали о том, что на их фронте нет никаких признаков подготовки ожидавшегося в скором времени наступления русских, начальники штабов армий группы „Центр“ столь же единодушно указывали на уже почти завершённое развёртывание крупных русских сил перед фронтом их армий. Однако в Генеральном штабе сухопутных сил у Гитлера настолько глубоко укоренилось — чему в немалой степени содействовала категорическая точка зрения Моделя, возглавлявшего фронт в Галиции, — предвзятое мнение о наибольшей вероятности русского наступления на фронте группы армий „Северная Украина“, что отказаться от него они уже не могли. Разумеется, развёртывание русских сил перед фронтом группы армий „Центр“ нельзя было отрицать, однако ему в оценке русских планов приписывалась лишь подчинённая роль. <...> Доминирующей оставалась точка зрения, пред-

полагавшая нанесение русскими основного удара на фронте группы армий „Северная Украина“. Вследствие этого там была сосредоточена большая часть танковых дивизий и все были уверены, что именно там наконец вновь удастся противопоставить „удар удару“».

Но ответить ударом на удар в другом месте фронта у этой группировки не получилось.

НАКАНУНЕ НАСТУПЛЕНИЯ

Подготовка Белорусской наступательной операции проходила в обстановке строгой секретности.

Главная ударная сила наступления — 1-й Белорусский фронт генерала К. К. Рокоссовского — получила в своё распоряжение три радиодивизиона и 1-й отдельный радиополк ОсНаз Ставки Верховного Главнокомандования. Годы тяжёлого труда и учёбы радиоразведчиков принесли свои плоды. «К началу операции „Багратион“ радиоразведкой была выявлена дислокация штабов всех занимавших оборону немецких армий и входивших в их состав корпусов и дивизий, а также достаточно точно установлены полосы их обороны и стыки между ними».

Сведения, добытые радиоразведкой накануне наступления, и оперативность доведения их до командования позволили своевременно избегать ударов немецкой авиации по местам дислокации советских войск. Так, в начале июня в полосе предстоящего наступления 1-го Белорусского фронта «радиоразведка установила, что немецкие самолёты-разведчики обнаружили на аэродромах бобруйского направления до 600 советских самолётов и зафиксировали много следов гусениц танков вблизи фронта. Эти данные были немедленно доложены командиром 545-го радиодивизиона командующему фронтом Рокоссовскому. Реакция командующего последовала незамедлительно: самолёты были сосредоточены по запасным аэродромам; места дислокации танков замаскированы, следы их передвижения, которое осуществлялось по ночам, уничтожены».

Перед началом наступления фронты перешли в режим радиомолчания, распоряжения передавались лично. Партизаны получили указания — дезорганизовать работу тыла противника, особенно связь и военные перевозки, первый подрыв рельсов провести в ночь на 20 июня.

УДАР НЕПРЕОДОЛИМОЙ СИЛЫ

23 июня 1944 года на врага обрушился удар непреодолимой силы. Хотя наступление началось почти от стен Смоленска, уже 3 июля был освобождён Минск, а к концу июля советские войска достигли предместий Варшавы. Правда, высокой ценой.

Все дни и ночи не смолкал эфир. В наступлении участвовали войска пяти фронтов — трёх

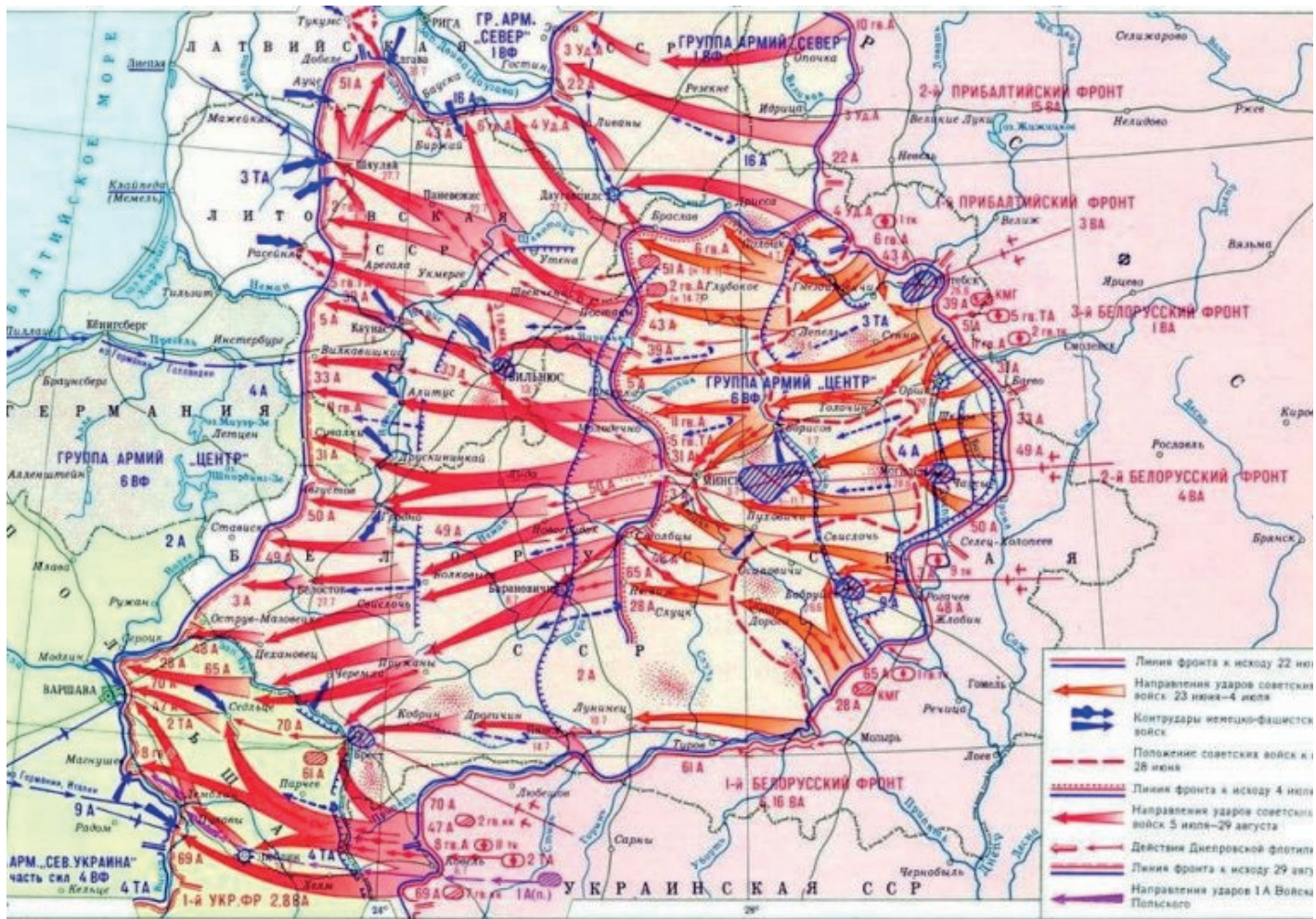


Иллюстрация 3. Белорусская наступательная операция 23.06–29.08 1944 г.

Белорусских и двух Прибалтийских (илл. 3). Обеспечение их взаимодействия и контроль расположения вражеских частей нельзя было наладить без участия радиоразведки. «Радиоразведчики всех фронтов напряжённо следили за радиосетями противника. Все изменения в дислокации войск противника вскрывались и докладывались командующим фронтами».

Особая роль была определена командованием отдельному радиополку ОсНаз. Ещё в начале наступления «1-й отдельный радиополк Ставки Верховного Главнокомандования, наблюдая за радиосвязью немецкого Генерального штаба, своевременно выявил перегруппировку войск противника и переброску 2-й армии с правого фланга группы армий „Центр“ на участок прорыва немецкого фронта, образовавшийся между 4-й и 9-й армиями». Эти сведения очень помогли успеху наступления.

ОСНОВНЫЕ ЗАДАЧИ — «ЗАБИВКА» И ИМИТАЦИЯ

131-й радиодивизион в Белорусской операции создавал помехи радиосвязи противника из районов северо-западнее Орши. Специалисты ди-

визиона применили новый тактический приём: вместо излучения непрерывных сигналов или передачи цифровых групп стали имитировать работу радиостанций противника. Этим они загружали радиолинии и вводили немцев в заблуждение.

Одной из основных задач стала постановка радиопомех окружённым частям гитлеровцев. В операции «Багратион» дивизион ставил помехи соединениям 9-й немецкой армии в районе Бобруйска, готовившимся, как стало известно из перехваченных радиogramм, к прорыву из окружения. Позже, совместно со 130-м дивизионом радиопомех, нарушал связь штабов группировки немцев, окружённой восточнее Минска. А ещё позже, в Висло-Одерской операции, дивизионы «забивали» радиостанции окружённой группировки фашистов в Познани.

ПРОСЬБА О ПРОДОЛЖЕНИИ ВЗАИМОДЕЙСТВИЯ

Быстрое изменение фронтовой обстановки летом 1944 г. активизировало работу канала сообщений, связанных с «Бодигардом». Британское послание от 25 июля было особенно важным (илл. 4).

В РУ Генштаба Красной Армии в начале августа был подготовлен проект официального ответа, вот его текст:

«Главе британской военной миссии в Москве генерал-лейтенанту Барроуз, через посредство полковника Тернера.

В письме от 25 июля 1944 года на моё имя Вы, по поручению начальников объединённого англо-американского штаба, запрашиваете, согласен ли советский Генеральный штаб с политикой сговывания возможно больших сил противника в Скандинавии, Италии и на Балканах на протяжении всего 1944 года, и в случае положительного ответа на этот вопрос со стороны советского Генерального штаба, просите продолжения взаимодействия и, особенно, в поддержании угрозы в направлении Болгарии в соответствии с планом „Бодигард“.

В связи с этим считаю необходимым довести до Вашего сведения, что советский Генеральный штаб не только согласен с политикой сговывания возможно больших сил врага на указанных направлениях, но и, как Вам хорошо известно, в течение всего летнего периода 1944 года практически успешно её осуществляет. <...>. Таким образом, в проводимых советским правительством мероприятиях и боевых действиях Красной Армии за летний период 1944 года соответственно продолжают осуществляться меры, предусмотренные планом „Бодигард“».

ВЗАИМОДЕЙСТВИЕ ПРОДОЛЖАЕТСЯ

Участие Генерального штаба и его Разведывательного управления в операции «Бодигард» на скандинавском и балканском направлениях продолжалось до конца 1944 г. Взаимодействие союзных штабов просматривается по происшедшим тогда событиям.

Как известно, в сентябре 1944 г. Финляндия и Болгария признали своё поражение и выступили против Германии. Армии 3-го Украинского фронта смогли пройти без помех через территорию Болгарии вглубь Балкан.

В октябре началось освобождение Заполярья. Уже 18 октября советские войска вошли на территорию Северной Норвегии во время Петсамо-Киркенесской операции. Немцы вскоре отступили за линию своих горных укреплений, куда принудительно эвакуировали и около 2/3 немногочисленных местных жителей, уничтожив их жилища и постройки. К концу октября активные боевые действия прекратились, но немецкое командование держало в Норвегии значительные силы вплоть до момента капитуляции Германии. По соглашению с норвежским правительством в изгнании, размещавшимся в Лондоне, советское военное управление Северной Норвегией продолжалось до сентября 1945 г.

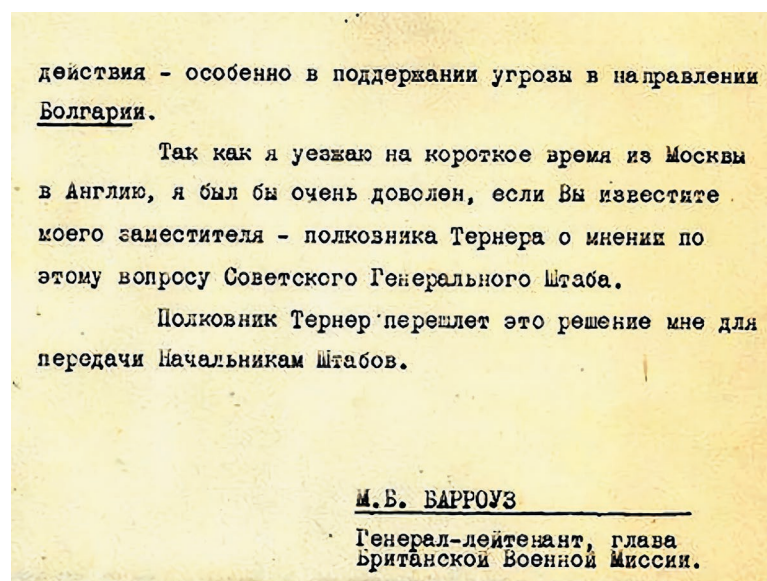
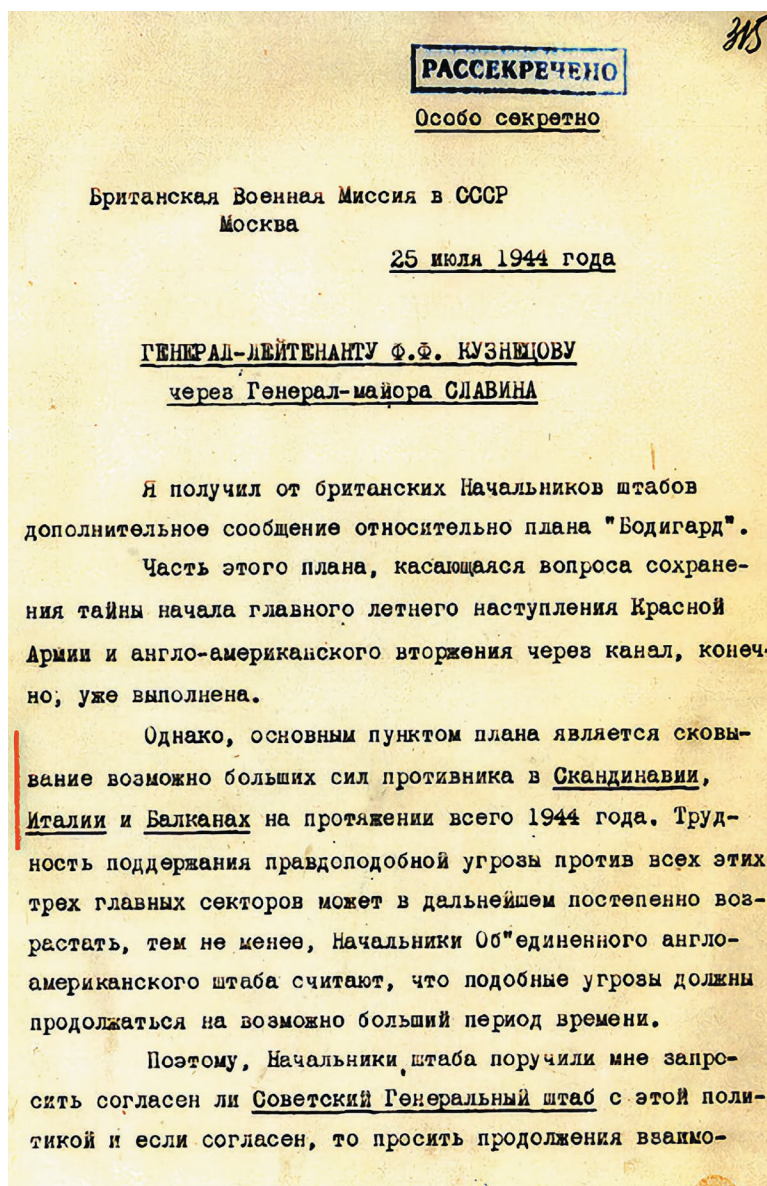


Иллюстрация 4. Лицевая и оборотная сторона письма М. Б. Барроуза, с сайта mil.ru

МАЛ ХОСТОВИК, ДА ДОРОГ

ЗАМЕТКИ БЕЗОПАСНИКА

Александр ИГОНИН
эксперт BIS Journal

Я часто пишу о том, как важно не только развернуть в периметре средства защиты информации, но и надлежащим образом их обслуживать и поддерживать в работоспособном состоянии. Сегодня хочется поговорить о локальных агентах СЗИ, устанавливаемых на хосты корпоративной сети. Довольно очевидно, что в идеальной ситуации эти агенты должны быть установлены везде, где это возможно.

Степень покрытия хостов корпоративной сети локальными агентами — простая и относительно легко измеримая метрика состояния защищённости инфраструктуры. Здесь есть свои особенности: например, нельзя просто взять список всех хостов и сверить его со списком, выгруженным из консоли СЗИ, потому что не все типы ОС могут этим СЗИ поддерживаться, а ряд хостов может быть исключён из покрытия в связи с недостатком ресурсов или особенностями корпоративной политики. Но даже с учётом этого вопрос «На каком проценте хостов, пригодных к установке локальных агентов, эти агенты реально установлены?» звучит куда проще, чем, например, «Насколько полно наш SOC покрывает атаки по классификации MITRE?» или «Какой процент существующих рисков безопасности отражён в матрице рисков?»

Из моего личного опыта, на практике посчитанная метрика покрытия обычно выглядят неприглядно и грустно, тоскливо болтаясь где-то в районе 50–60 % (а в тяжёлых случаях — и 30 %). Насколько при этом падает приносимая соответствующим СЗИ польза — вопрос дискуссионный, но я бы грубо оценил процент падения эффективности равным проценту «недопокрытых» хостов. При этом никто не отменял закон подлости: атака на компанию, как назло, пройдёт именно через них.

Вот пример из реальной практики: организация заключила контракт с MSSP на мониторинг событий безопасности. Всего на серверы и рабочие станции было установлено порядка двух тысяч сенсоров, степень покрытия Windows-инфраструктуры агентами при этом составила солидные 70 %. Через некоторое время бдительный аналитик SOC заме-

тил отсутствие сенсора на одном из контроллеров домена (на все остальные сенсоры были установлены). Отбившийся от стада контроллер был «возвращён в стойло», и буквально через час сработало правило SIEM, обнаружившее подозрительный процесс, выполнявший дампы базы данных Active Directory. Дальше события развивались вполне предсказуемо: эскалация, запуск процедуры реагирования на инцидент, изучение под микроскопом всех активностей на хосте, его очистка и так далее. Однако из-за отсутствующего долгое время агента атака была обнаружена с лагом в несколько месяцев, что сильно затруднило определение вероятного пути компрометации и её масштабов.

Другой пример — хосты с устаревшей и снятой с поддержки ОС, а также высоконагруженные серверы. Как показывает практика, даже в крупных и известных компаниях нередко ситуация, когда критичные для бизнеса сервисы разворачивались настолько давно, что сейчас уже никто не понимает, ни как это работает, ни как это поддерживать, ни что делать, если всё вдруг «упадёт». Вот и крутится себе какое-нибудь веб-приложение под Windows 2003, на которую не то что антивирус, но даже и обновления безопасности администраторы ставить отказываются, мотивируя это тем, что «если сломается, уже не починим». Для атакующих такие системы — настоящий рай: взломать легко, заметить никто не сможет, а проблемы со стабильностью злоумышленников, разумеется, не волнуют.

Ну и последняя типичная ситуация — хосты (как правило, серверы), которые более не используются, но ещё не выведены из эксплуатации и работают. Это тоже лакомый кусочек для хакеров. Но к теме теневой ИТ-инфраструктуры вернёмся как-нибудь в другой раз.

Так что, приобретая очередное чудо-СЗИ, неплохо бы сразу внедрить и метрику по контролю степени его покрытия. А чтобы эта метрика выглядела не слишком печально, придётся организовать адекватную поддержку — об этом мы говорили в апрельской заметке прошлого года. И пусть кому-то эта метрика покажется скучной и незамысловатой, но многие самые важные вещи в нашей индустрии часто выглядят именно так.



ФОРУМ

РОССИЯ 20²⁴

17 19

сентября



XXII

**международная
конференция**
по проблематике
инфраструктуры
открытых ключей
и электронной
подписи

г. Санкт-Петербург
Отель «Коринтия»



pki-forum.ru

DLP-СИСТЕМА НОВОГО ПОКОЛЕНИЯ

x3

раза быстрее —
от первого подозрения
до принятия решения



Центр
расследований
InfoWatch

события,
персоны,
файлы,
аналитика,
риски

автоматизированная настройка



infowatch.ru

ИИ наводит порядок
в информационных
активах



учится находить новые
активы в трафике —
машинное обучение

надёжный контроль



точно отличает
конфиденциальную
информацию



контролирует все
необходимые каналы
передачи данных

проактивное реагирование



помогает расследовать —
пути информации
и хронология событий



позволяет предотвращать —
поведенческая аналитика
и аудит хранения



даёт возможность делать выводы
на будущее — аудит мест хранения,
поиск ложных срабатываний
в статистике

