ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ НА СЛУЖБЕ ГРАЖДАН И КАК СРЕДСТВО ФОРМИРОВАНИЯ НОВЫХ КИБЕРУГРОЗ

Кулаева Марина Владимировна
m.kulaeva@it.alania.gov.ru

Намерение государства: превратить Россию в одного из мировых лидеров в области ИИ

В ноябре 2020 г. Владимир Путин повторил, что Россия способна стать одним из глобальных лидеров в области искусственного интеллекта.

Президент РФ поручил правительству оперативно внести проекты законов об экспериментальных правовых режимах для использования технологий искусственного интеллекта. "В наступающее десятилетие нам предстоит провести цифровую трансформацию всей страны, всей России, повсеместно внедрить технологии искусственного интеллекта, анализа больших данных", - сказал Владимир Путин в эфире телеканала "Россия 24".

от 27 декабря 2019 г. № 466-р

г. Владикавказ

Об утверждении концепции развития искусственного интеллекта в Республике Северная Осетия-Алания

В соответствии с Указом Президента Российской Федерации от 10 октября 2019 года № 490 «О развитии искусственного интеллекта в Российской Федерации», Законом Республики Северная Осетия - Алания от 18 сентября 2019 г. № 60-РЗ «О Стратегии социально-экономического развития Республики Северная Осетия - Алания до 2030 года» и в целях формирования системы взглядов на основные направления развития искусственного интеллекта в рамках развития цифровой экономики в Республике Северная Осетия - Алания, а также базовых принципов, основных направлений, целей, задач государственной политики в сфере цифрового развития Республики Северная Осетия - Алания:

1. Утвердить прилагаемую Концепцию развития искусственного интеллекта в Республике Северная Осетия-Алания.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 59276— 2020

Системы искусственного интеллекта

СПОСОБЫ ОБЕСПЕЧЕНИЯ ДОВЕРИЯ

Общие положения

Издание официальное

Принцип № 1 Контролируемость и управляемость систем ИИ

Принцип № 2 Прозрачность и предсказуемость функционирования технологий ИИ

Принцип № 3 Стабильность и надёжность систем ИИ

Принцип № 4 Ответственное и непредвзятое применение ИИ

ИИ в медицине

***Порядок взаимодействия государственных и негосударственных инфосистем в здравоохранении* Постановление Правительства РФ №447 от 12.04.2018**

***Обезличивание данных и обеспечение защиты сведений от несанкционированного использования* Приказ Минздрава России от 14.06.2018 № 341н**

Применение ИИ в транспортном комплексе

Переход от частной разрозненной к комплексной интегрированной автоматизации и цифровизации	Стандартизация информационного обеспечения и унификация информационных ресурсов	Выделение юридически значимых базовых государственных информационных ресурсов с обеспечением их строгой достоверности, соответствия стандартам и с предоставлением доступа к ним с использованием электронных сервисов	Исключение дублирования и параллелизма в функциональных возможностях и информационных ресурсах органов государственного управления транспортным комплексом
Переход от создания множества отдельных новых ИС в интересах решения новых задач цифровизации к созданию новых функциональных средств в рамках существующей единой информационной среды	Обязательность соблюдения единых информационных отраслевых стандартов при создании и развитии ИС	Плановая постепенная трансформация ключевых информационных ресурсов и функциональных возможностей существующих государственных информационных систем в единую информационную среду	Использование при взаимодействии с хозяйствующими субъектами механизма «Единого окна» для унификации взаимодействия и организации доступа к цифровым сервисам и данным органов государственного управления транспортным комплексом

Очевидно, что цифровая трансформация – одно из величайших благ , в том числе, для киберпреступного сообщества, ведь благодаря ей значительно расширяется потенциальная поверхность атаки. При этом к услугам киберпреступников – все те же искусственный интеллект и технологии машинного обучения, которые, как выясняется, можно с той же эффективностью использовать не только для защиты корпоративных сетей, но и для осуществления атак на них.

Смена парадигмы информационной безопасности

- Построение защиты, которую нельзя сломать по своей сути утопично.

- Большинство «интересных» систем либо уже взломаны, либо могут оказаться взломанными, являясь заложниками **«уязвимости нулевого дня».** Времени на устранение проблемы НОЛЬ- программная уязвимость, обнаружена злоумышленниками до того, как о ней узнали производители программы. Не выпущены патчи, а у злоумышленника уже есть метод, используемый для атаки на системы. И далее следует «атака нулевого дня» с использованием «эксплойта нулевого дня», подвид вредоносных программ содержащих исполняемый код, способный воспользоваться одной или несколькими уязвимостями в программном обеспечении на локальном или удаленном компьютере.

- Следовательно, главная задача любой системы безопасности — максимально быстро обнаружить атаку и атакующего в системе, сократить окно его возможностей настолько, чтобы он не успел нанести непоправимый вред. Речь идет о так называемой **СПОСОБНОСТИ ОБНАРУЖЕНИЯ** (ability to detect), в связи с чем наблюдается рост востребованности средств защиты искусственного интеллекта, позволяющих решать задачи по своевременному выявлению атак и инцидентов. Аналитики IBM заявили о трехкратном росте интереса к технологиям такого типа в течение трех кварталов уходящего года.

Инициативы регуляторов против уязвимостей и недекларированных возможностей

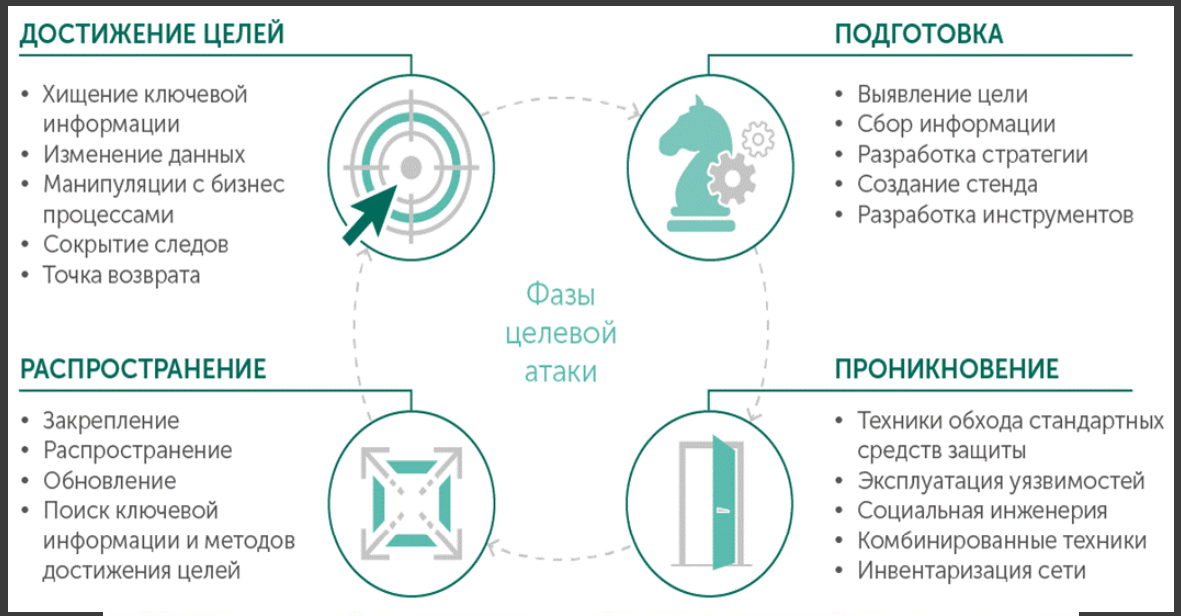
- Совершенствование законодательства в области защиты объектов критической информационной инфраструктуры (КИИ). В нормативных документах по обеспечению безопасности критической информационной инфраструктуры скорректированы неоднозначные термины и формулировки
- Появилось понятие и инструментарий средств ГосСОПКА, сформулированное в приказах ФСБ № 196, 281, 282
- Начала складываться практика привлечения к ответственности по ст. 274 Уголовного кодекса («Нарушение правил эксплуатации средств хранения, обработки или передачи компьютерной информации и информационно-телекоммуникационных сетей»)
- Издан "Методический документ. Профиль защиты прикладного программного обеспечения автоматизированных систем и приложений кредитных организаций и некредитных финансовых организаций" (утв. Банком России 10.07.2020)
- Внесены изменения в Федеральный закон «О связи» и Федеральный закон «Об информации, информационных технологиях и о защите информации», известные как «Суверенный Рунет», гарантирующие бесперебойное функционирование интернета в России. Теперь коммерческие компании (в частности, операторы связи) обязаны проводить регулярные киберучения.
- Advanced Persistent Threat (APT)- таргетированные или целевые кибератаки наиболее развитая и весомая угроза для коммерческих компаний и государственных организаций России

При этом видно, что планомерной государственной программы по повышению осведомленности граждан в вопросах кибербезопасности пока не существует: каждый сам выбирает, как и где ему просвещаться. Поэтому преступники легко могут обманывать доверчивых граждан не только в целях выманивания денег и сведений, но и для пополнения своих рядов малообразованными новичками.

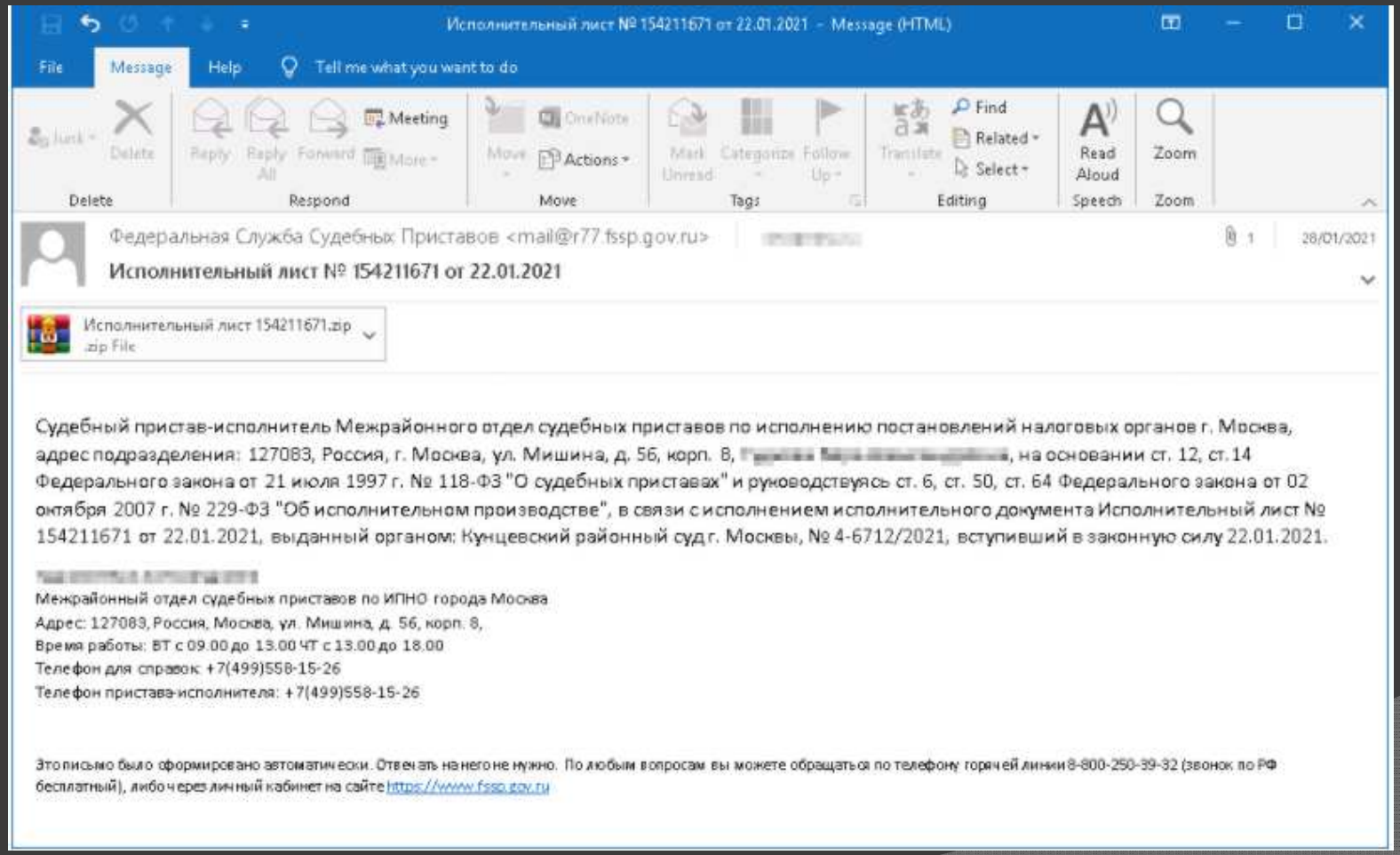
Стать киберпреступником не сложно(

Пример сценария реализации угрозы безопасности информации

Фазы целевой атаки



Криптография для упаковки вредоноса с целью сокрытия от средств защиты при сигнатурном анализе



Банк данных угроз безопасности информации https://bdu.fstec.ru/

УБИ.222: Угроза подмены модели машинного обучения — Вид

Описание угрозы	Угроза заключается в возможности подмены нарушителем модели машинного обучения, используемой в информационной (автоматизированной) системе, реализующей технологии искусственного интеллекта. Данная угроза обусловлена слабостями разграничения доступа в информационных (автоматизированных) системах, использующих машинное обучение. Реализация данной угрозы возможна при наличии у нарушителя непосредственного доступа к модели машинного обучения
Источники угрозы	Внутренний нарушитель с высоким потенциалом
Объект воздействия	Программное обеспечение (программы), использующее машинное обучение; модели машинного обучения
Последствия реализации угрозы	Нарушение конфиденциальности Нарушение целостности

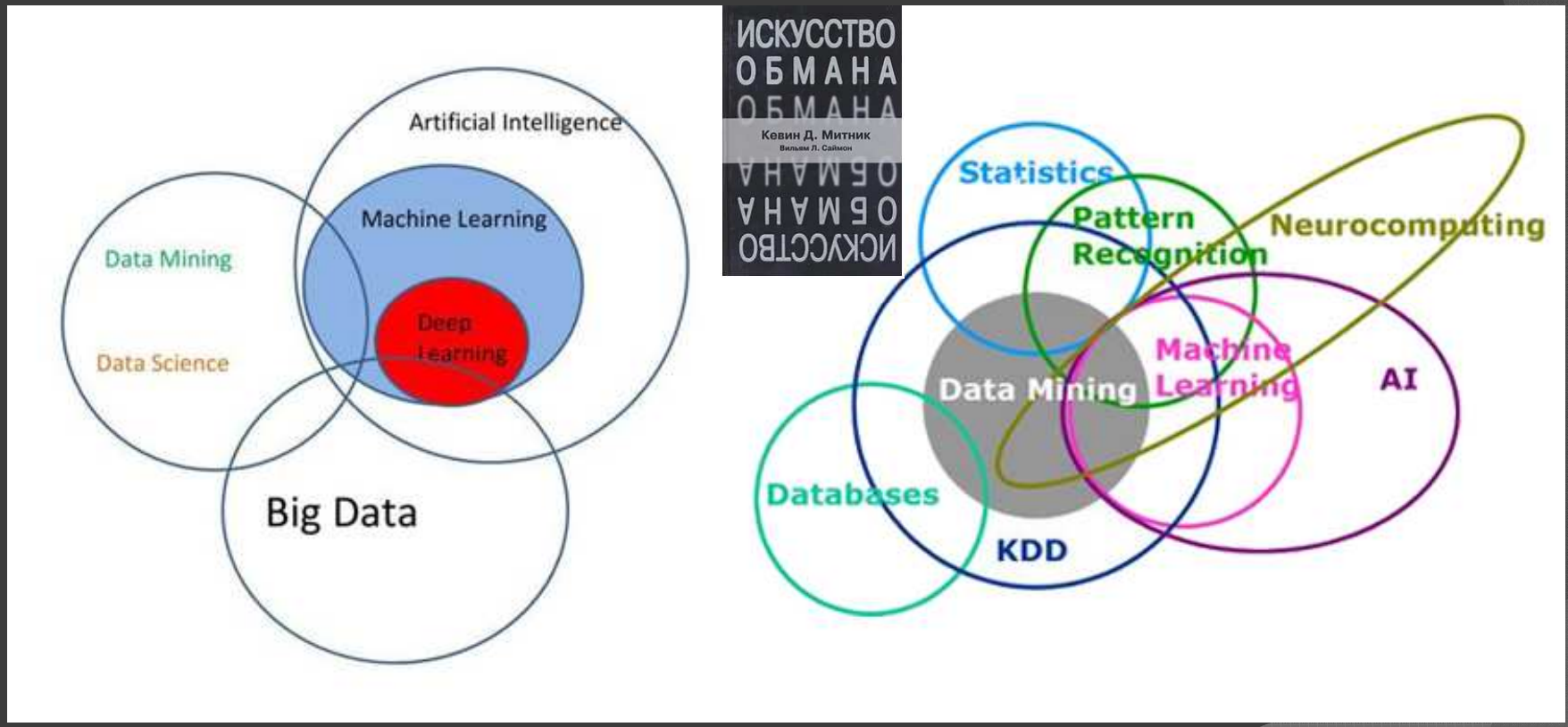
УБИ.221: Угроза модификации модели машинного обучения путем искажения («отравления») обучающих данных — Вид

Описание угрозы	Угроза заключается в возможности модификации (искажения) модели машинного обучения, используемой в информационной (автоматизированной) системе, реализующей технологии искусственного интеллекта. Данная угроза обусловлена: - недостатками реализации процесса машинного обучения; - недостатками устройства алгоритмов машинного обучения. Реализация данной угрозы возможна при наличии у нарушителя возможности воздействовать на процесс машинного обучения
Источники угрозы	Внутренний нарушитель со средним потенциалом Внешний нарушитель с высоким потенциалом
Объект воздействия	Программное обеспечение (программы), использующее машинное обучение; модели машинного обучения; обучающие данные машинного обучения
Последствия реализации угрозы	Нарушение целостности

УБИ.220: Угроза нарушения функционирования («обхода») средств, реализующих технологии искусственного интеллекта — Вид

Описание угрозы	Угроза заключается в возможности нарушения функционирования («обхода») нарушителем средств, реализующих технологии искусственного интеллекта. Данная угроза обусловлена следующими причинами: - отсутствие в обучающей выборке необходимых данных; - наличием недостатков модели машинного обучения;
Источники угрозы	Внутренний нарушитель со средним потенциалом Внешний нарушитель с высоким потенциалом
Объект воздействия	Программное обеспечение (программы), реализующие технологии искусственного интеллекта

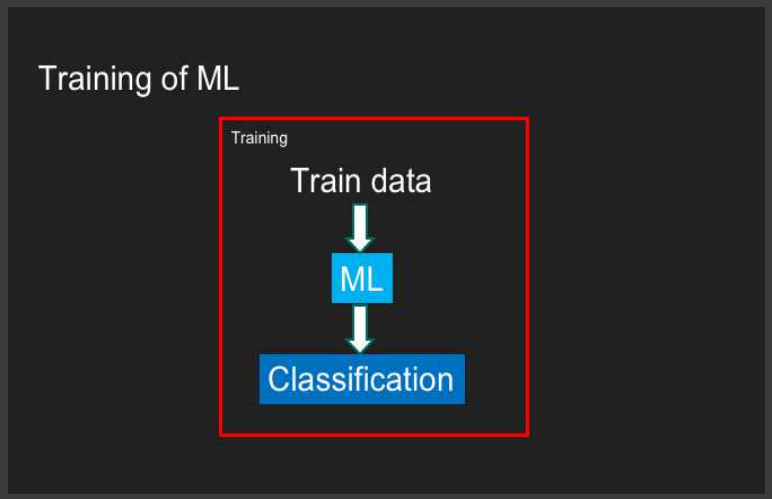
Искусственный интеллект в кибербезопасности



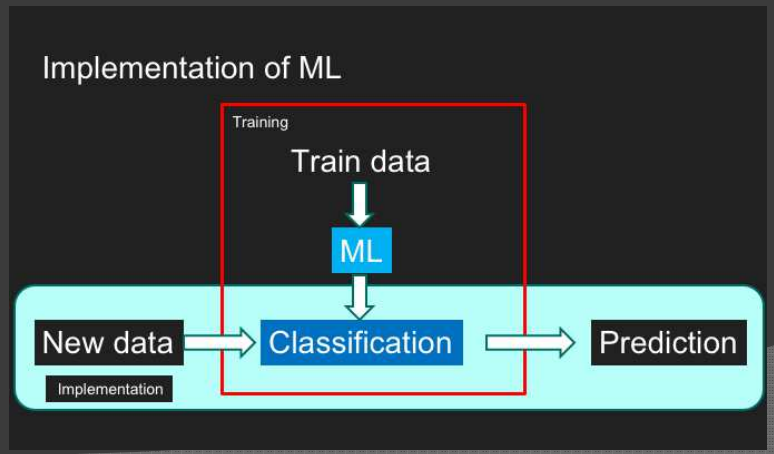
The scheme of the classical approach

The scheme of ML work

Машинное обу-чение состоит из двух процессов. Первый- это тренировка, когда человек берет данные, обучает модель и в итоге получает некий классификатор.



Второй процесс- это уже использование ML, когда обученный классификатор внедряется в систему, а затем на вход системы подают новые данные, которые классификатор не видел. В результате мы получаем предсказания от классификатора.



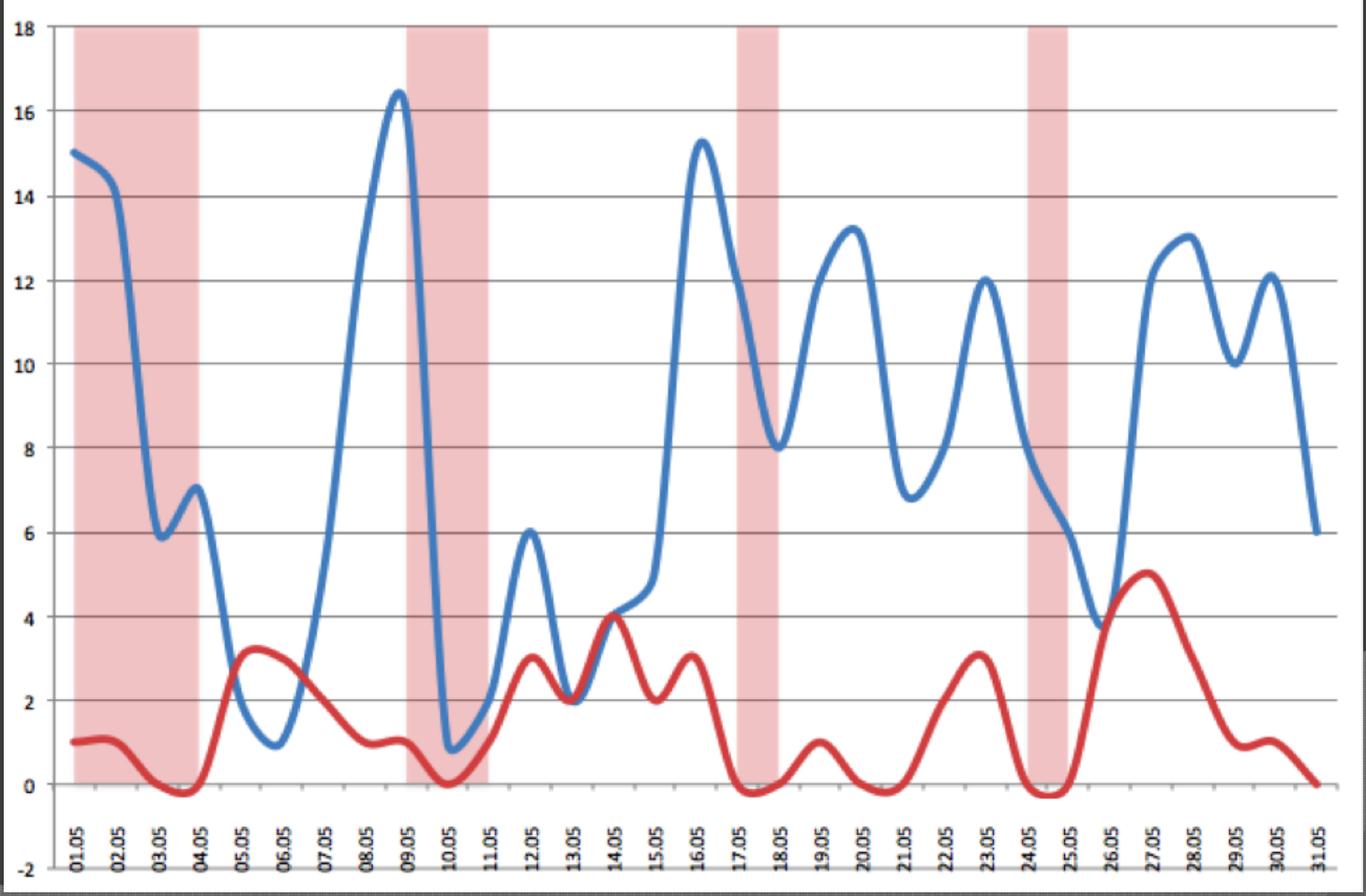
Взяты четыре паттерна поведения человека при чтении электронной почты, анализ которых поможет определить его действия:

В какое время суток человек пользуется почтой: утром, днем, вечером.

Сколько устройств использует: телефон, компьютер или сразу несколько устройств одновременно.

В каких локациях человек находится, когда пользует-ся почтой.

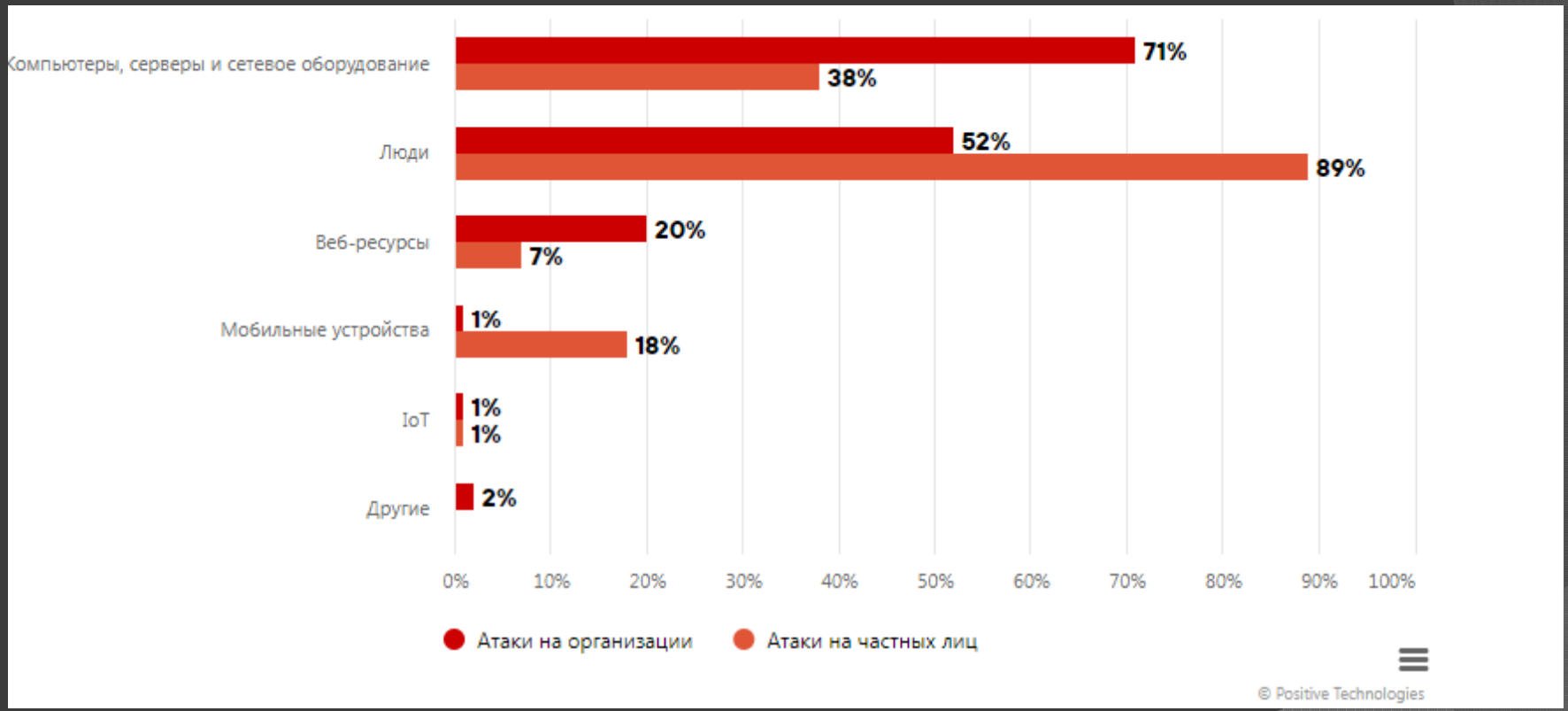
В каком порядке человек проверяет письма: сверху вниз или снизу вверх. Мы можем определить это по тому, как он отвечает или удаляет из ящика рассылки и прочий мусор.



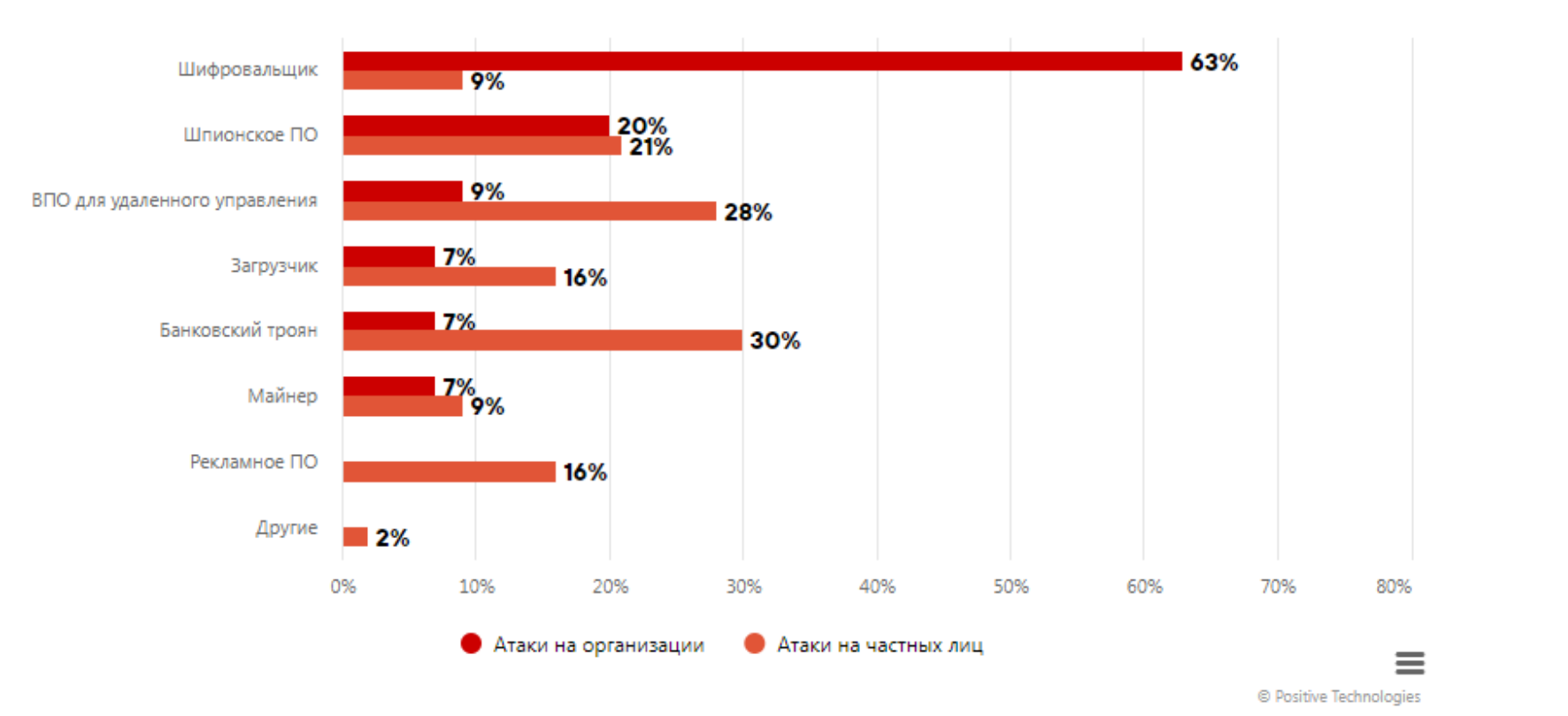
Всего за 500$

- слежка за действиями пользователей;
- запуск файлов и выполнение команд;
- запись снимков экрана;
- включение веб-камеры и микрофона;
- сканирование локальной сети;
- загрузка файлов из интернета.

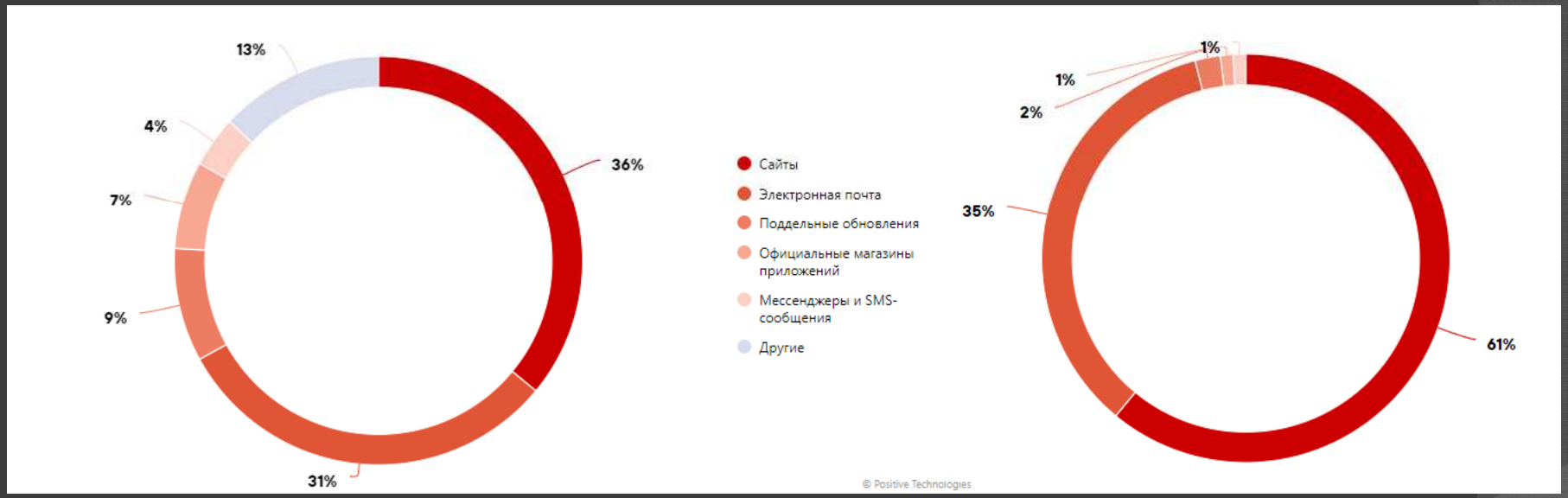
Объекты атак, II квартал 2021 г.



Типы вредоносного ПО, II квартал 2021 г.



Способы распространения вредоносного ПО, II квартал 2021 г.



Методы социальной инженерии применялись в 56% инцидентов...