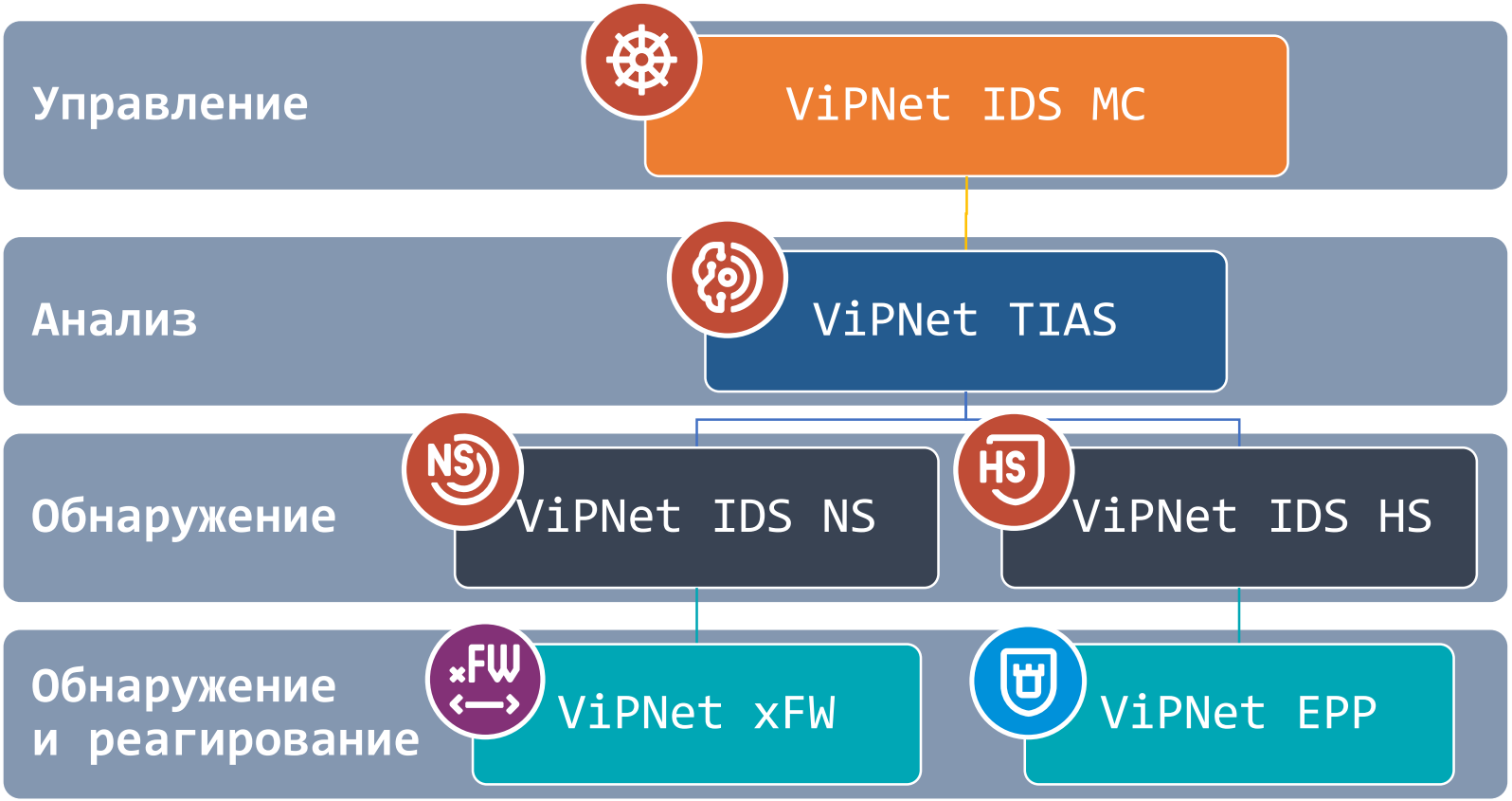
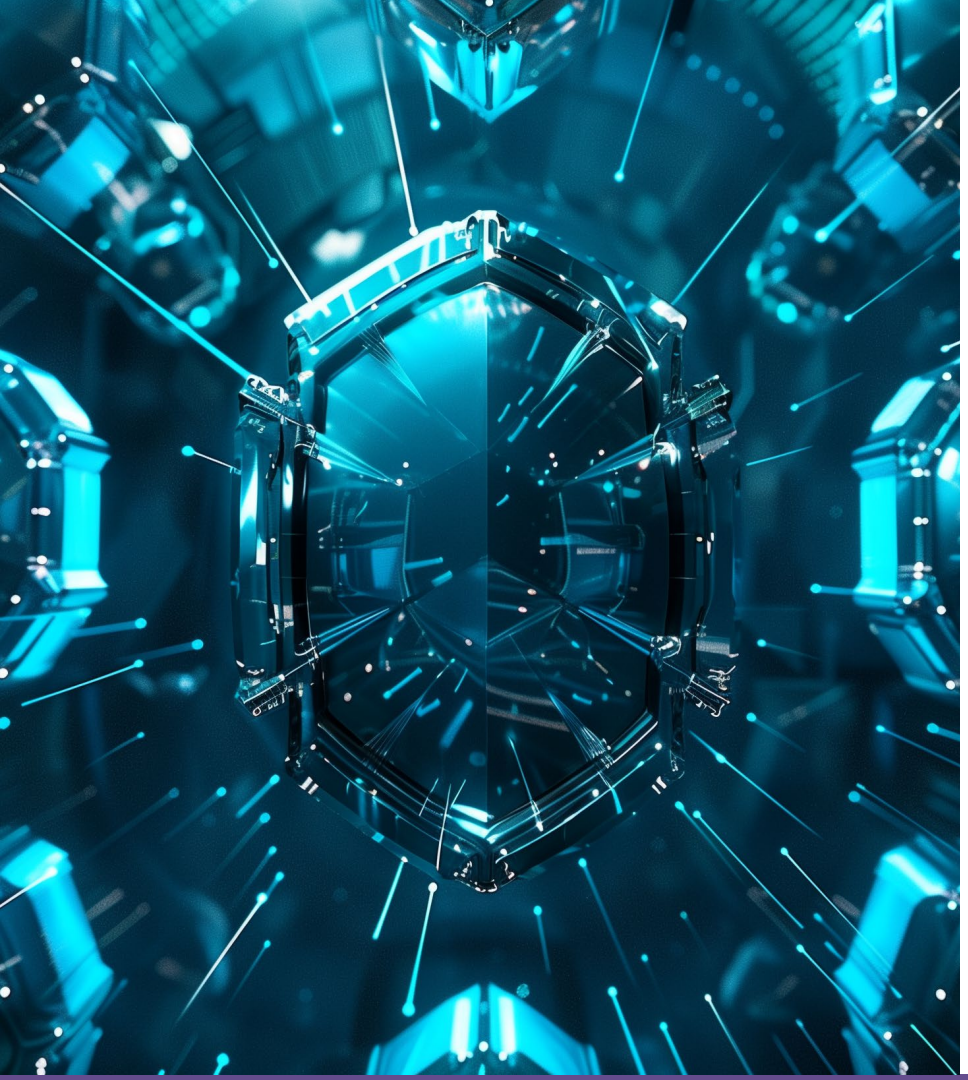
Использование машинного обучения для выявления компьютерных атак

Светлана Старовойт
Руководитель продуктового направления

Решение ViPNet TDR





Назначение решения

Назначение решения Threat Detection and Response – выявление угроз информационной безопасности, связанных с компьютерными атаками, и предоставление специалисту по ИБ информации, позволяющей адекватно и своевременно среагировать на такую угрозу

Предпосылки: увеличение объемов данных

Сетевые и узловые сенсоры решения TDR регистрируют события ИБ в том объёме, обработка которого неавтоматизированными средствами нецелесообразна

Разнообразие событий ИБ и объёмы их регистрации продолжают расти

Предпосылки: несовершенство метаправил

Осуществление корреляции событий в условиях высоко изменчивой среды* с помощью детерминированных алгоритмов посредством ручного составления и поддержки правил становится все более затруднительным

* Меняется не только набор правил, на основе которых регистрируются «сырые» события ИБ, но и характерные детали осуществления угроз ИБ и ландшафт инфраструктуры

1.5 Сила метаправил и недостатки ML

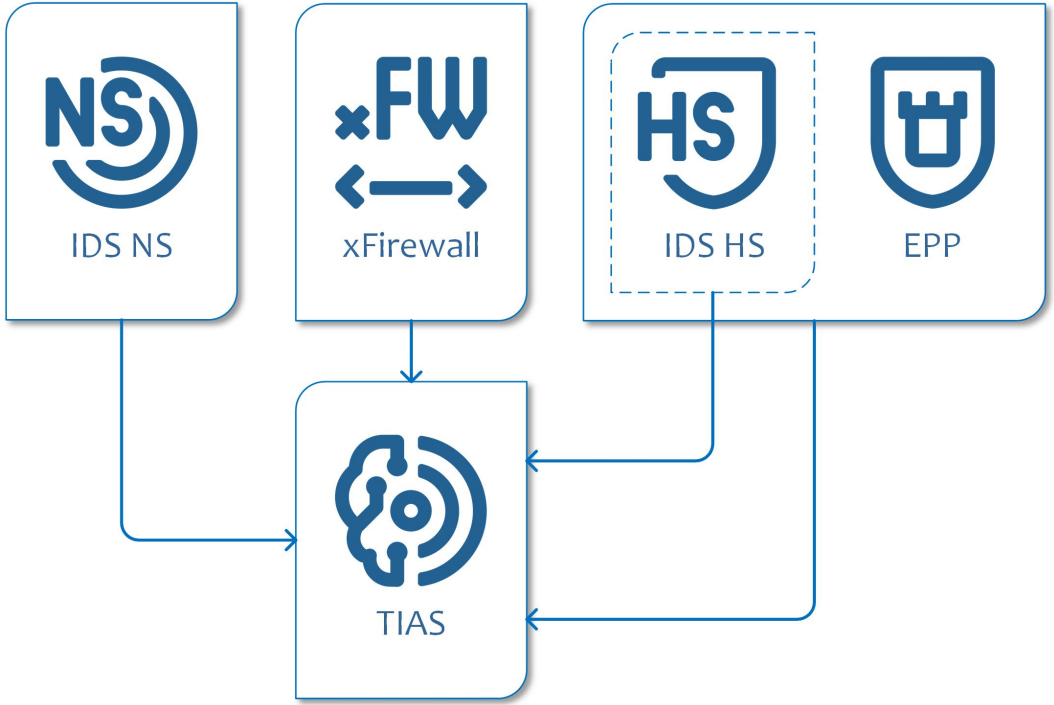
ТехноФест 2024 infotecs

Критерий

- Интерпретируемость результата
- Привлечение эксперта ИБ для поддержки инструмента в актуальном состоянии
- Гибкость инструмента при появлении новых событий

MetaRules	Machine Learning
Высокая	Низкая
Необходимо	Не требуется
Отсутствует	Высокая

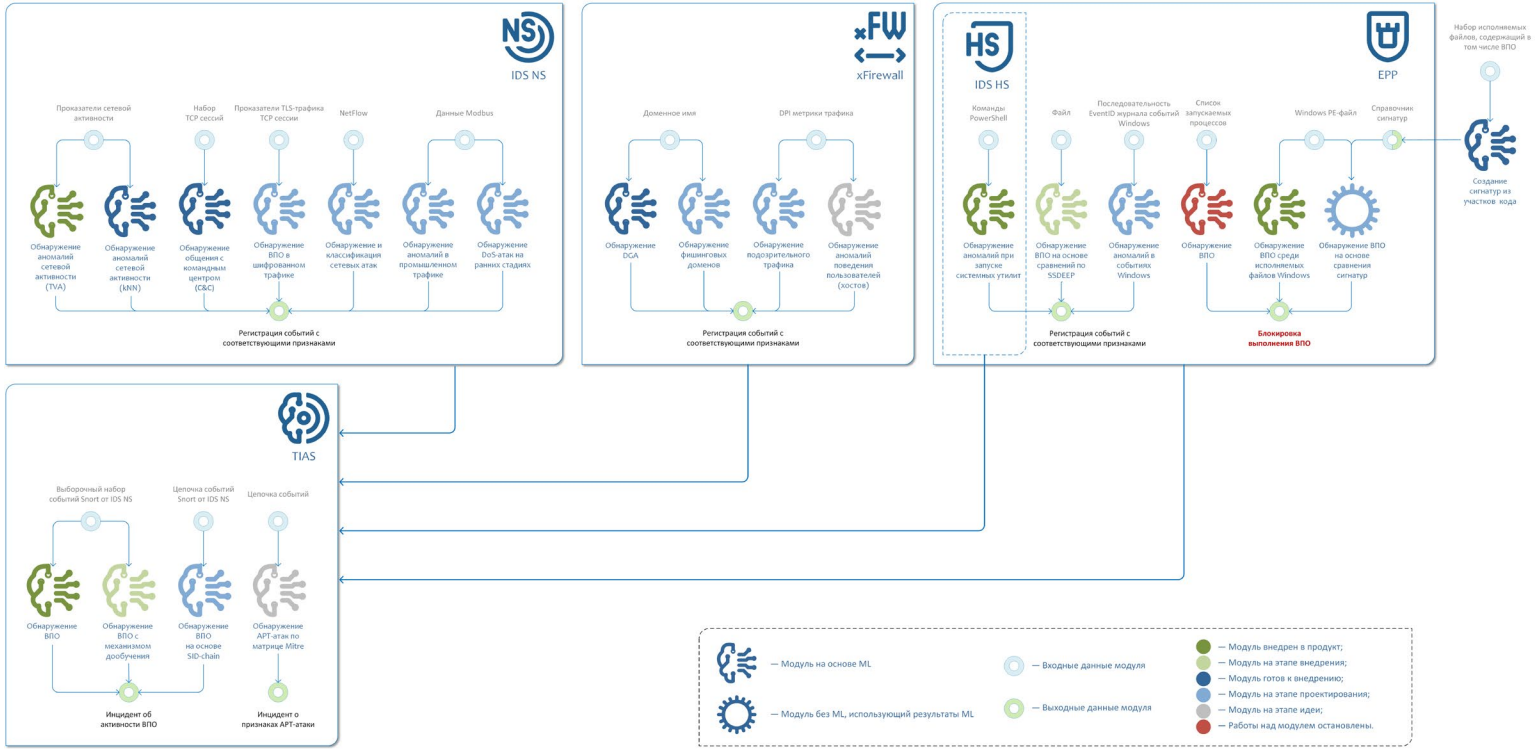
Сфера внимания ML на данном этапе



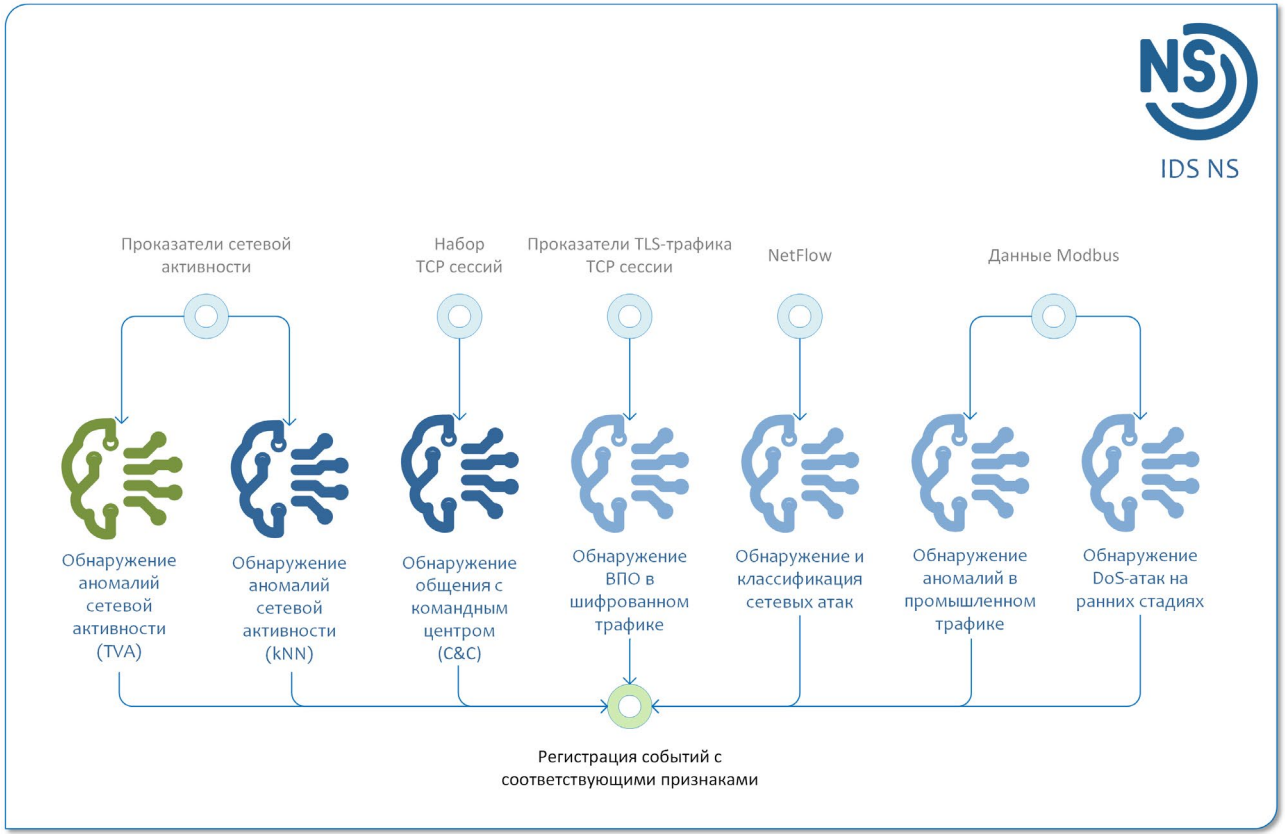
ТЕХНО 2024 infotecs ФЕСТ

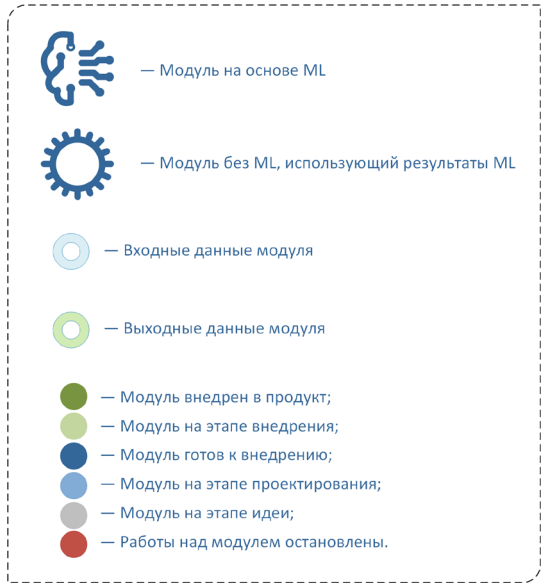
ML-модули

3.1 ML-модули в составе продуктов

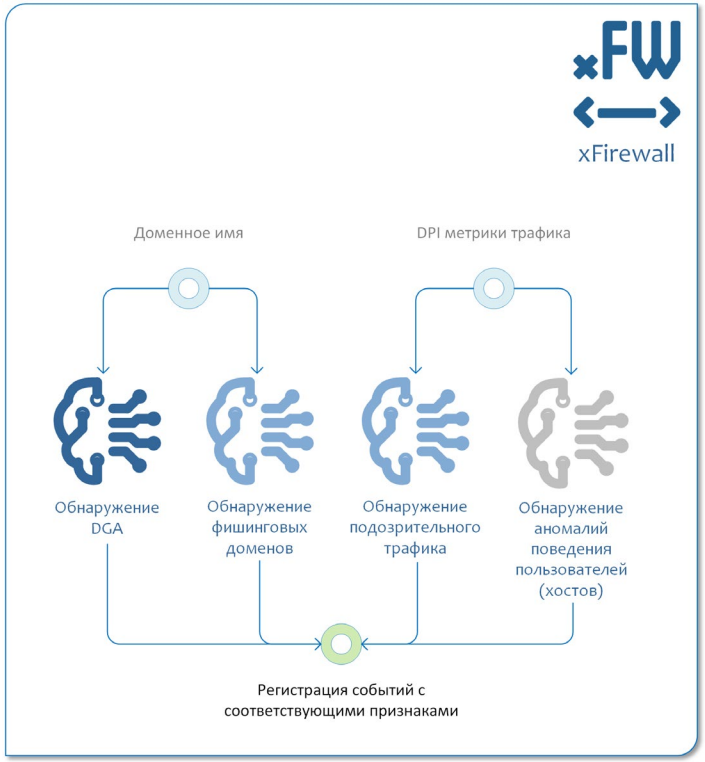


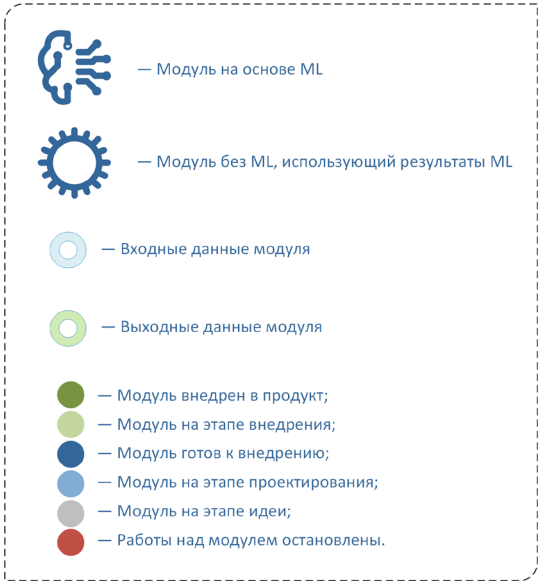
ML-модули для IDS NS



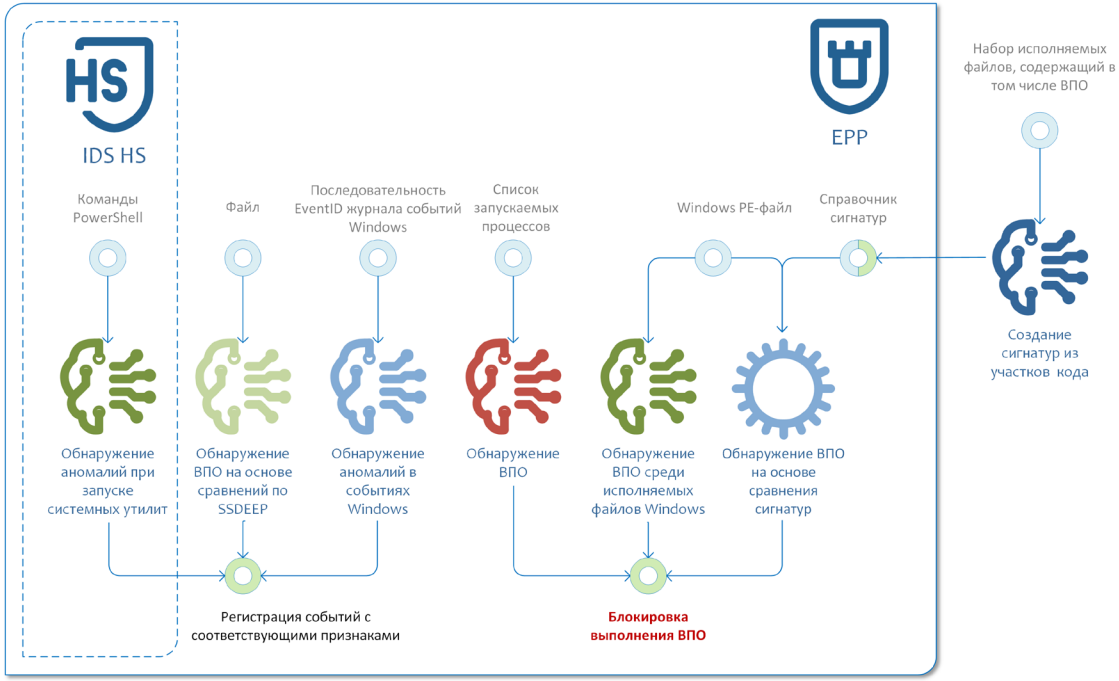


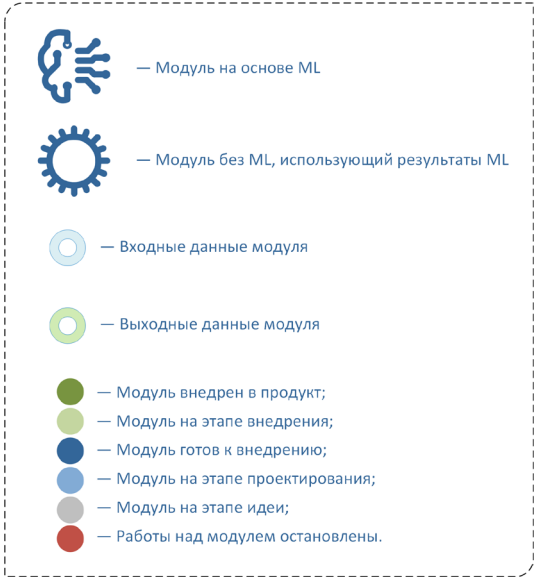
ML-модули для xFirewall



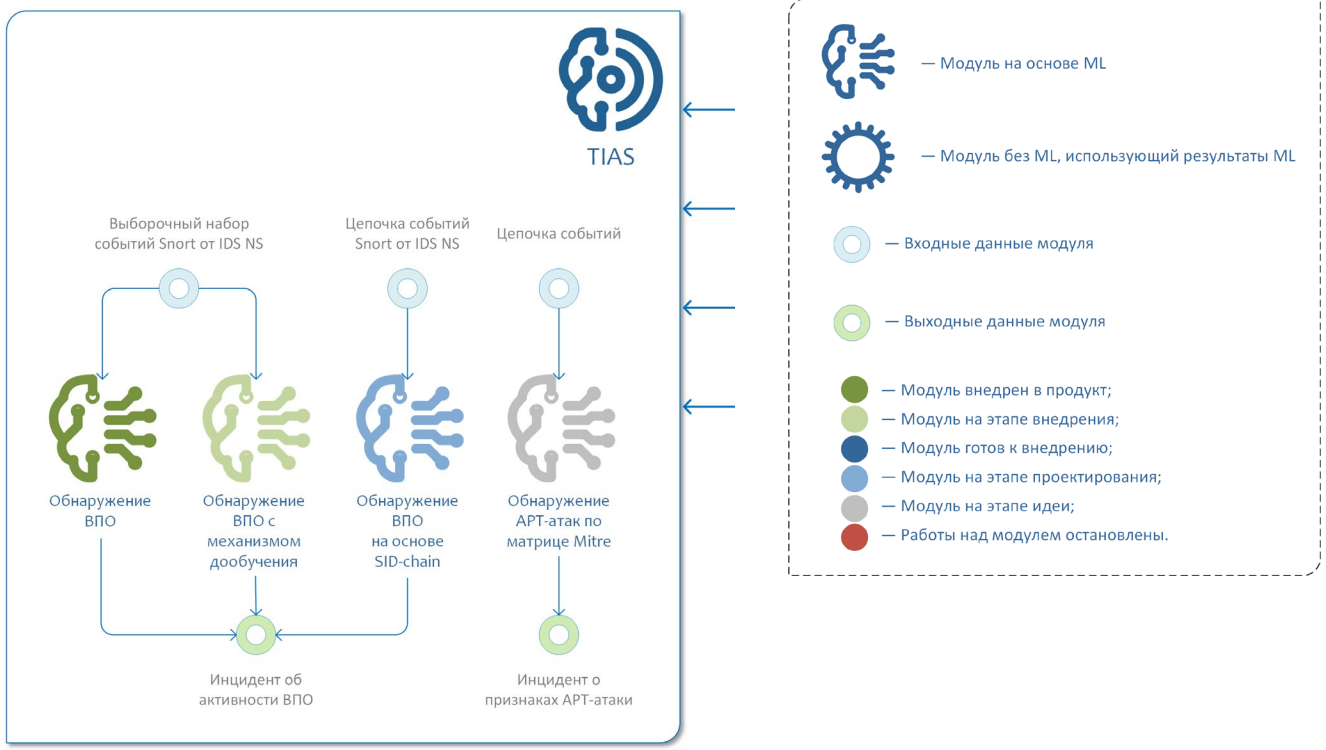


ML-модули для EndPoint Protection





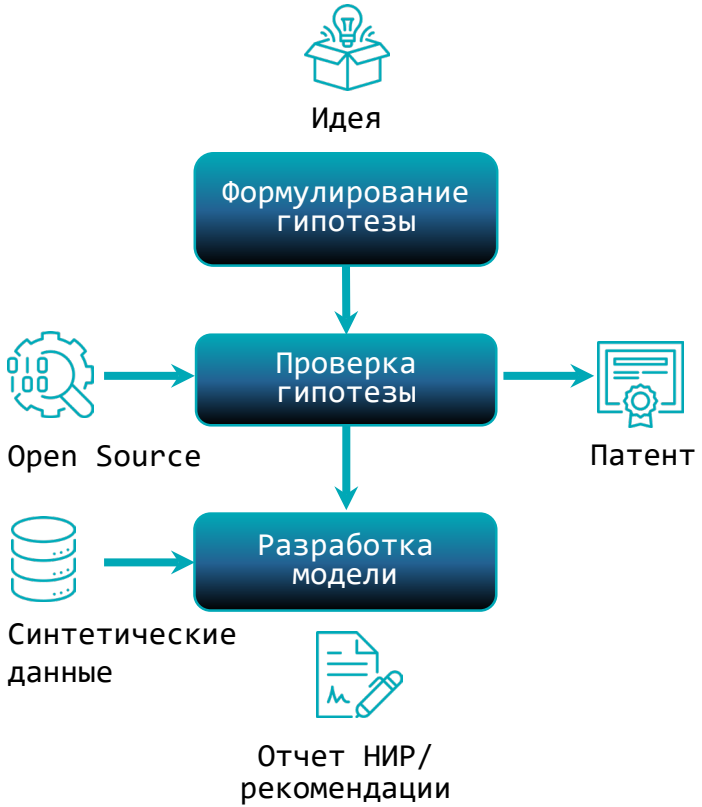
ML-модули для TIAS



ТЕХНО 2024 infotecs Фест

Процесс разработки

1 Проверка гипотезы

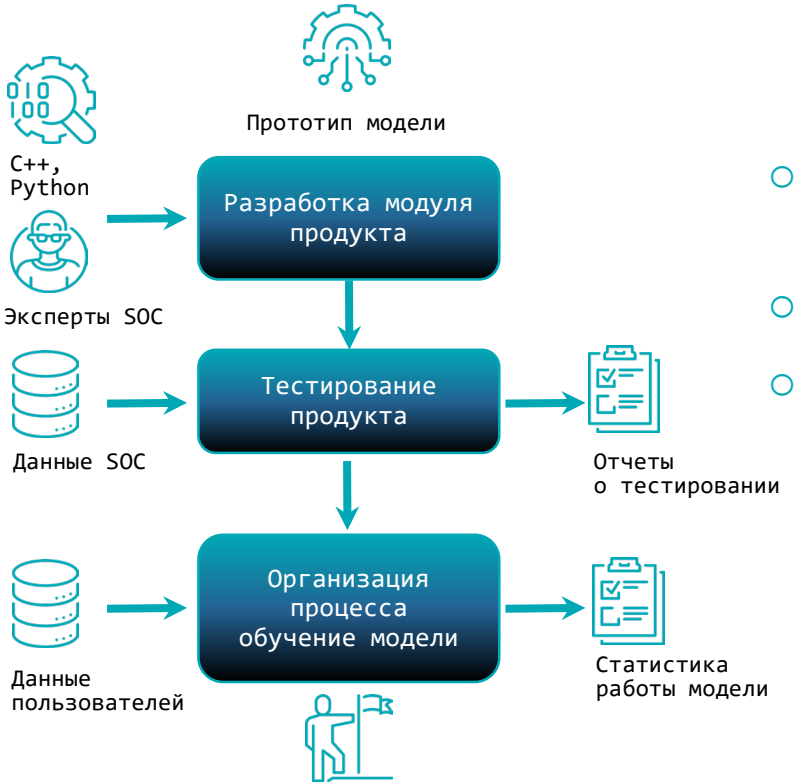


- Привлечение студентов и учёных
- 2 патента в области ML
- Использование готовых инструментов и синтетических данных для ускорения процесса

2 Прототипирование

- Аналитики данных и специалисты по ML
- Автоматизированные процессы работы с данными
- Рабочие прототипы для встраивания

3 Внедрение в продукт



- Разработчики продуктовых команд и эксперты SOC
- Только реальные данные
- Сбора обратной связи о работе модели

4 Результаты

- 50 проверенных гипотез
- 30 разработанных прототипов
- 5 внедренных моделей

ТЕХНО 2024 infotecs ФЕСТ

Проблемы использования методов машинного обучения

Безопасность

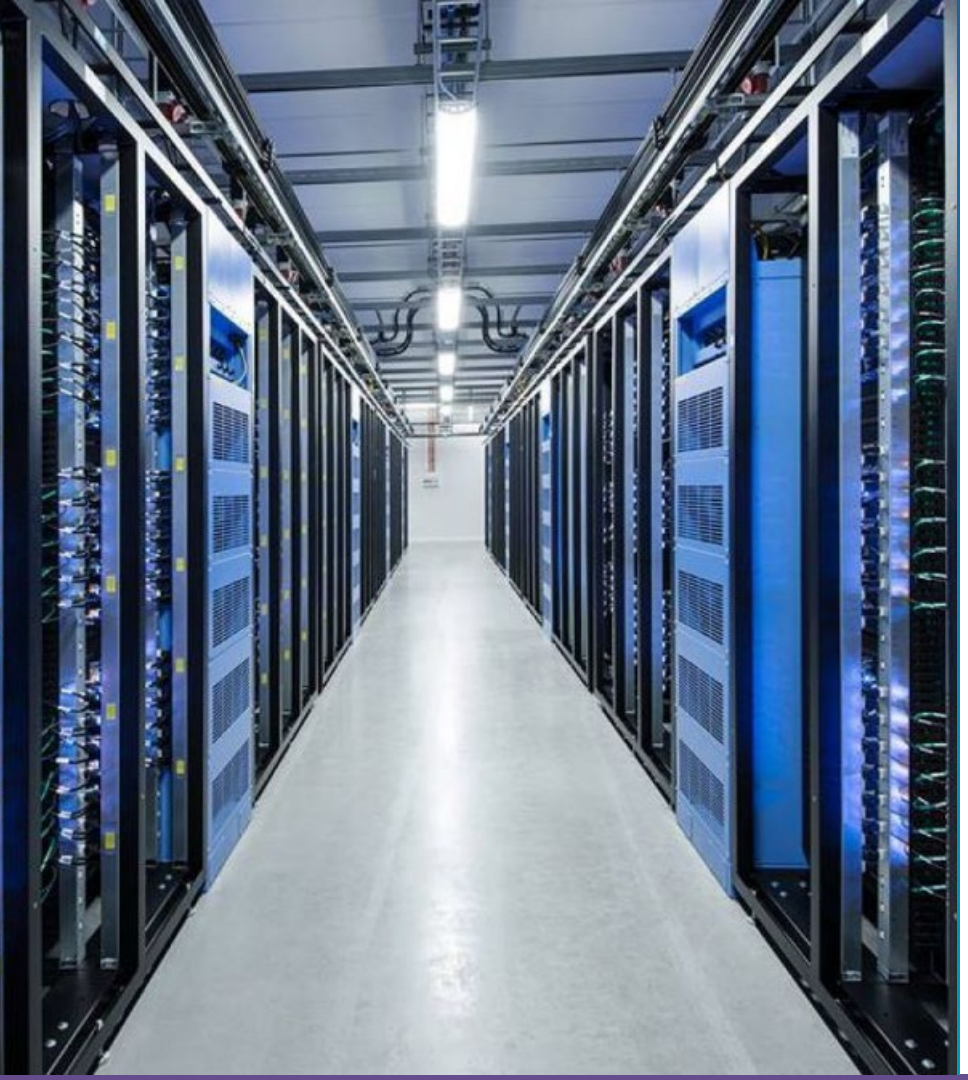
- Атаки на Open Source библиотеки и на данные
- SDL или MLOps
- Передача данных третьим лицам
- Дополнительное регулирование

Доверие к работе модели

- Интерпретируемость результата работы модели
- Дрейф данных
- Доверие при передаче данных

Данные для обучения и проверки

- Синтетические данные
- Открытые источники
- Данные SOC
- Данные пользователей
- Разметка экспертами или пользователями



Ресурсоемкость

Где работает/обучается модель?

- на конечном узле
- в ЦОДе
- в облаке

> Если мы все еще не собираем, коррелируем и обогащаем данные, то мы отстаем от современных вызовов и потребностей

AI

ТЕХНО 2024 infotecs Фест

Светлана Старовойт
Руководитель продуктового направления

Подписывайтесь на наши соцсети