

**ML -> SecOps vs SecOps -> ML**

# Наш план



- 1 Что такое MISecOps и как он связан с безопасной разработкой
- 2 Какие «корни» у безопасной разработки моделей
- 3 Куда двигаться?

# КТО Я

Газизова Светлана

Директор по построению процессов DevSecOps  
Positive Technologies

Несколько лет занимаюсь консалтингом в области безопасной разработки. До этого поработала в роли системного аналитика и заместителя ИТ-директора. В информационную безопасность пришла через разработку, поэтому есть понимание, что, как и почему происходит. Опыт в технологиях и консалтинге больше 6 лет. Собирала команды, учила аппсексов, придумывала подходы. Люблю и стараюсь развиваться внутри безопасности приложений во всех направлениях:)



КТО Я

Не МЛ'щик)

как и почему происходит. Опыт в технологиях и консалтинге больше 6 лет. Собирала команды, учила аппсеков, придумывала подходы. Люблю и стараюсь развиваться внутри безопасности приложений во всех направлениях:)

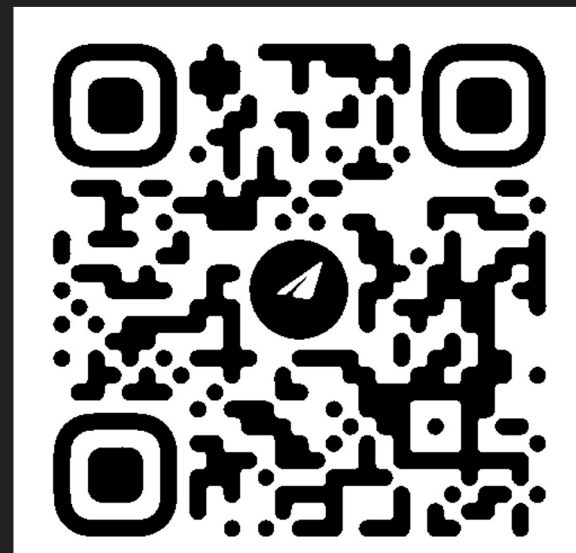
# Я рядом



На связи



Визитка



Telegram-канал  
AppSec Journey

# Что такое MLSecOps?)

# Все, что заканчивается на SecOps



Source Data

Data Engineering/Raw data

Dataprep Development

Training&Testing

Input&output

Policies&Strategy

Ops&monitoring

Secure by Design  
OSA

Secure coding  
SAST

DAST/Fuzz

CSP/CA/SCA

Train Algorithm

Apply Algorithm

# Все, что заканчивается на SecOps



Source Data

Data Engineering/Raw data

Dataprep Development

Training&Testing

Input&output

Policies&Strategy

Ops&monitoring

Secure by Design  
OSA

Secure coding  
SAST

DAST/Fuzz

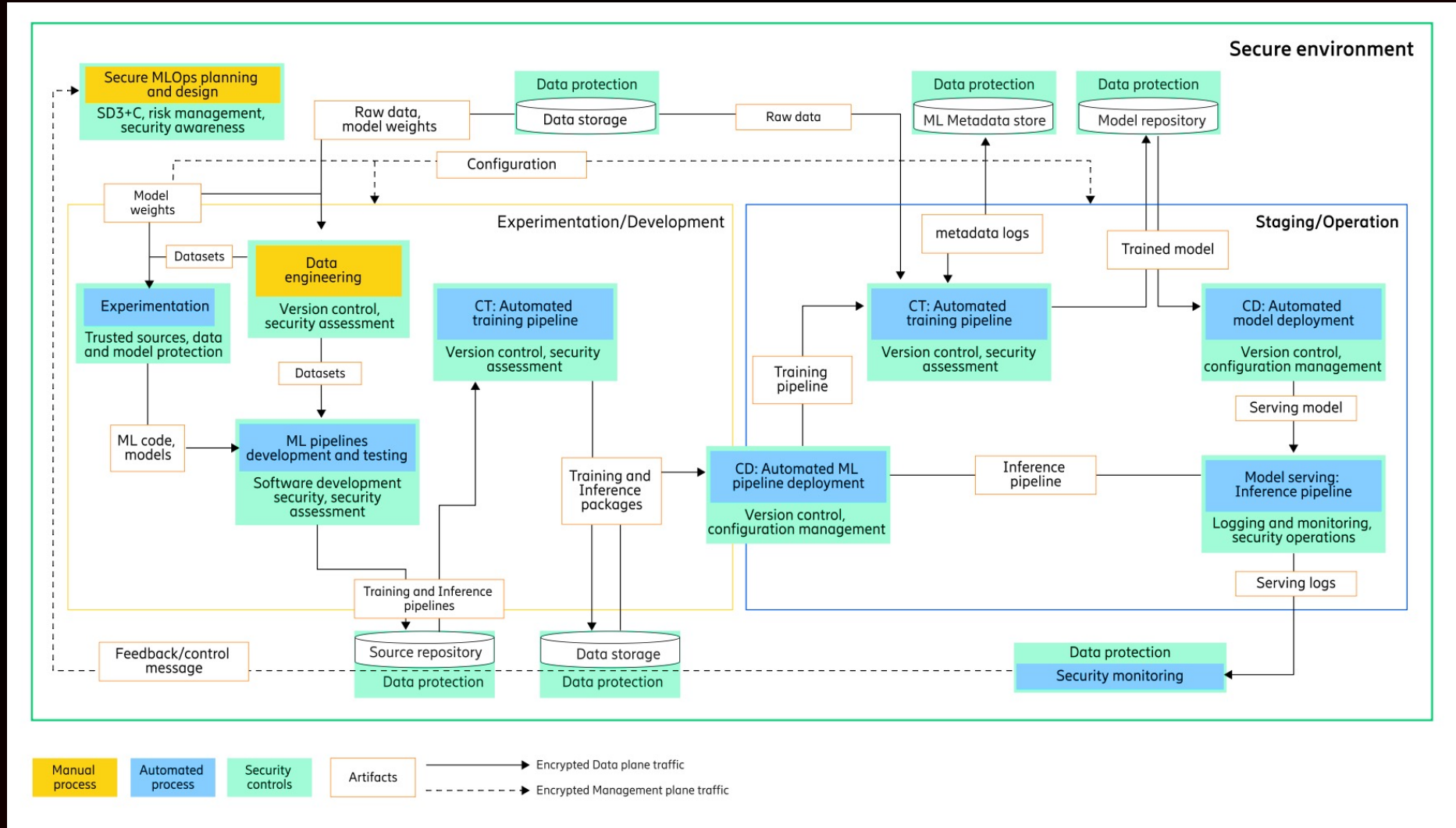
CSP/CA/SCA

Train Algorithm

Apply Algorithm



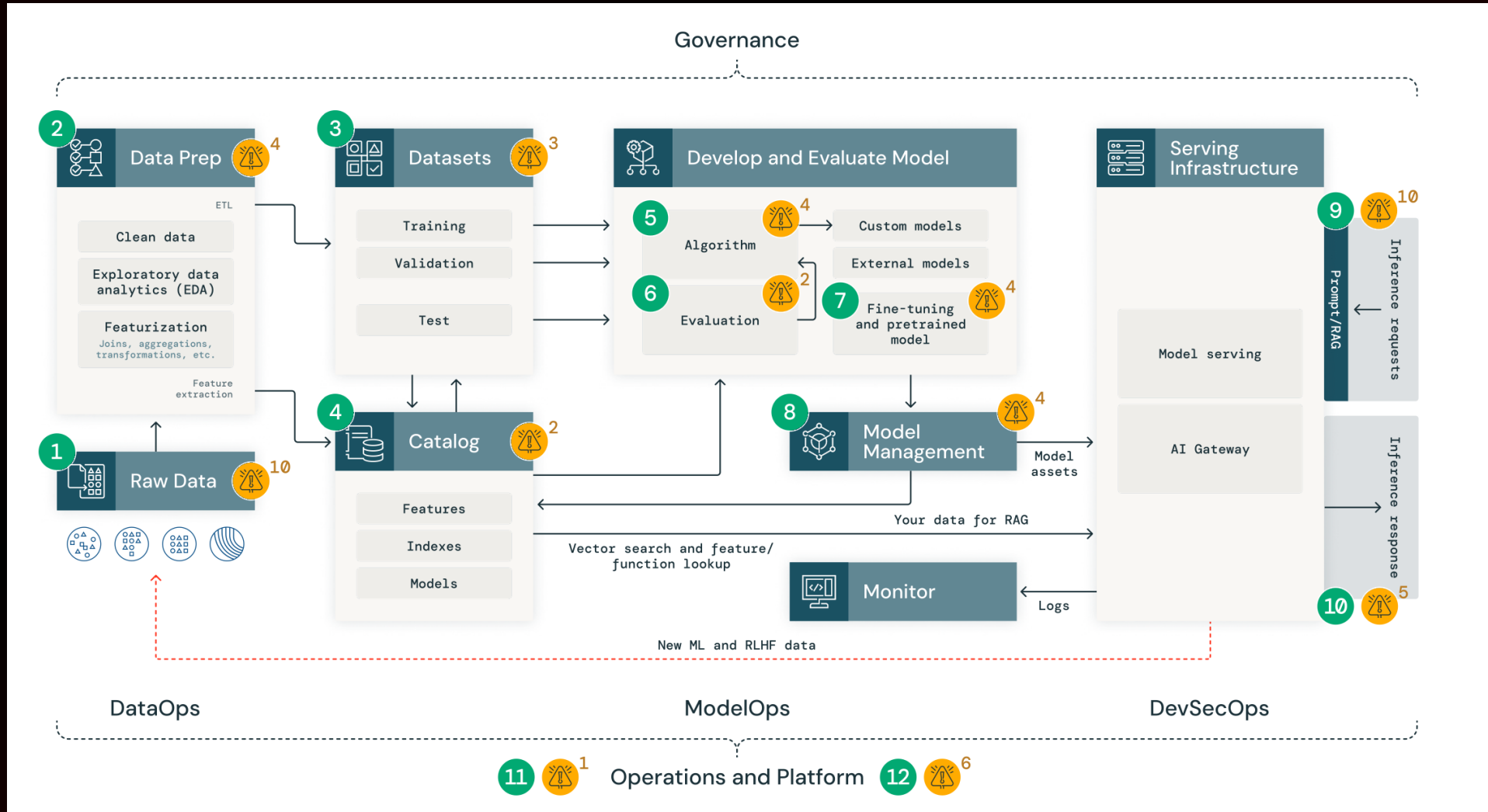
# Архитектура MLOps + Sec



# Что волнует DevSecOps?

Успели! Глянем DASF 

# Схематично



# Работа с данными

Исходные данные

Подготовка  
данных

Создание  
датасетов

Управление

Недостаточный контроль доступа, манипулирование функциями, отсутствие отслеживаемого ЖЦ модели, недостоверность данные: устаревание, некорректность и т.д.

# Работа с моделью

Алгоритм

Оценка

Сборка

Управление  
моделью

Дрейф модели, кража параметров, внедрение вредоносных библиотек, уязвимости Supply Chain и т.д.

Развертывание  
и  
обслуживание  
модели

Inference request

Inference response

Инверсия модели, зацикливание инпута, галлюцинации и т.д.

Ops&Platform

MLOps

MLPlatform

Отсутствие отслеживания инцидентов, отсутствие мониторинга, отсутствие SDLC и т.д.



# А чего по технике?)

## Security for AI Market Map

### Governance

 CRANIUM  credo | ai Arklow  HIDDEN LAYER  PROTECT AI

### Observability

 Humanloop  Helicone CALYPSO AI  Credal.ai **FLOW.**

### Security

#### Model Consumption

#### Detection + Response

 HIDDEN LAYER  Lasso SECURITY  AI Shield Powered by Bosch CALYPSO AI

#### AI Firewall

CALYPSO AI  ROBUST INTELLIGENCE  TROJ.AI  Prompt:  PROTECT AI

#### Continuous Red Teaming

 ADVERSA  ROBUST INTELLIGENCE  LAKERA

#### Data Leak Protection

 Nightfall  PRIVATE AI  LAKERA

#### Vulnerability Scanning + Monitoring


 PROTECT AI  HIDDEN LAYER  ROBUST INTELLIGENCE  Giskard TROJ.AI  AI Shield Powered by Bosch

#### Model Building + Serving

#### PII Identification/Redaction

 PRIVATE AI  gretel  Kobalt Labs skyflow

#### Synthetic Data

 TONIC  gretel  hazy MOSTLY AI

#### Federated Learning

 MITHRIL SECURITY  DynamoFL  Devron  RHINO HEALTH  OWKIN  nimbleedge  FEDML

А если честно?

# Берем это:



## Open Source инструменты

- **ModelScan** - защита от атак на сериализацию ML-моделей.
- **NB Defense** - Безопасные Jupyter Notebooks.
- **Garak** - сканер уязвимостей LLM.
- **Adversarial Robustness Toolbox** - Библиотека методов защиты моделей машинного обучения от adversarial-атак.
- **MLSploit** - MLsploit - облачный фреймворк для проведения интерактивных экспериментов с исследованиями в области адверсивного машинного обучения.
- **TensorFlow Privacy** - Библиотека алгоритмов и инструментов машинного обучения с защитой конфиденциальности.
- **Foolbox** - инструментарий на языке Python для создания и оценки adversarial-атак и защитных средств.
- **Advertorch** - Инструментарий на Python для исследования стойкости атак.

# «Золотая» Орда



Open-Source вики по безопасной  
разработке. Наполнена по MLSecOps

# Выводы?



Ну, мы видим зарождение нового витка безопасности. Пока что мы в начале пути, и конца-края этому пути не видно.

Давайте вместе сделаем что-то крутое!

(с) Джейсон Стетхем