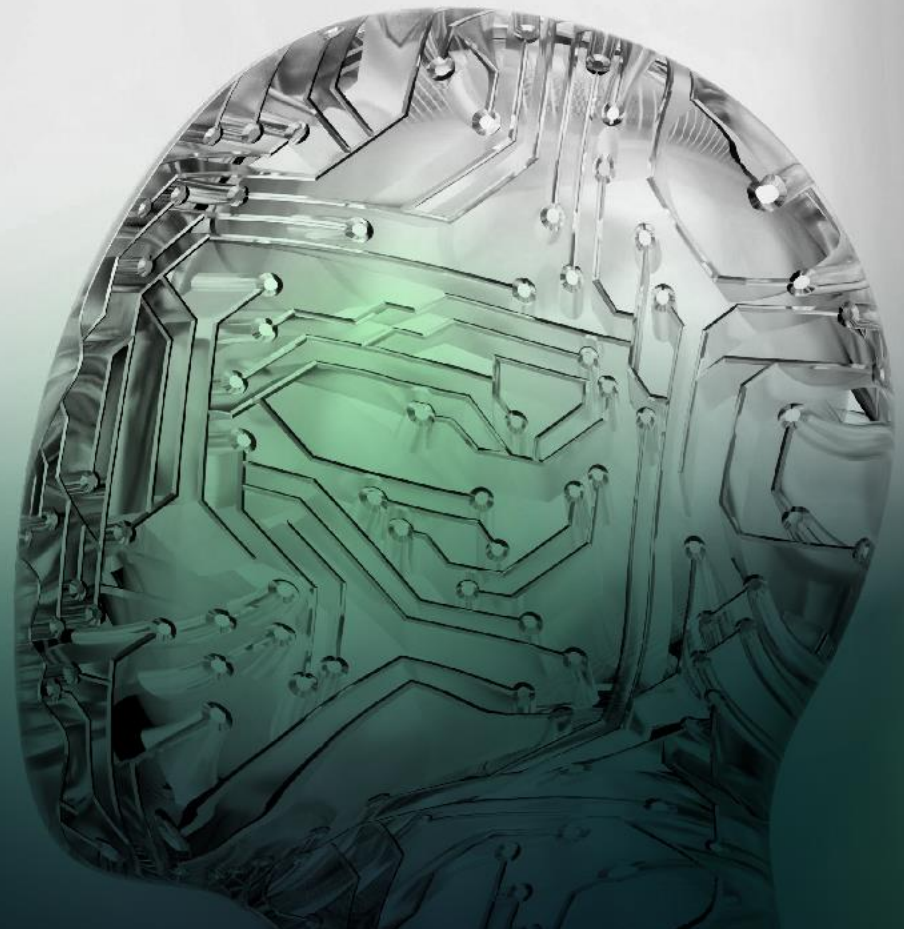
КОНСОРЦИУМ
ИССЛЕДОВАНИЙ
БЕЗОПАСНОСТИ
ТЕХНОЛОГИЙ
ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА



Тестирование технологий искусственного интеллекта

Ганелин П.В., советник по стратегии АНО «НТЦ ЦК»

РАБОЧАЯ ГРУППА №2

РАБОЧАЯ ГРУППА №2

ОБЩАЯ ИНФОРМАЦИЯ

Лидер группы:

Петр Ганелин, НТЦ ЦК

Активные участники группы:

- ✓ НТЦ ЦК
- ✓ Академия Криптографии
- ✓ ИСП РАН
- ✓ Газинформсервис
- ✓ SwardFish Security
- ✓ РТУ МИРЭА
- ✓ Русское биометрическое общество

Ты записался в группу тестирования технологий ИИ?

Образ результата: проведение тестирования технологических решений ИИ и их элементов для включения в **реестр доверенных решений ИИ (РГЗ).**

РАБОЧАЯ ГРУППА №2

ЗАДАЧИ

***Тестирование различных технологических элементов систем ИИ*:**

- ✓ **разработка наборов метрик, систем мониторинга и анализа логов**
- ✓ разработка комплексных методик тестирования и испытаний в соответствии **с требованиями (РГ1)**
- ✓ создание методики аудита на протяжении **жизненного цикла (РГ4)** технологического решения ИИ
- ✓ тестирование дата сетов (размеченные, неразмеченные, создание эталонов обучающих сетов)
- ✓ тестирование обучающих фреймворков
- ✓ тестирование моделей ИИ и используемых в них алгоритмов

Методики по видам и типам тестирования:

- ✓ на соответствие **применимой регуляторике (РГ1)** (по ИБ, отраслевой регуляторике)
- ✓ **функциональное тестирование** (тестирование на соответствие заявленным характеристикам)
- ✓ информационная безопасность (специфичные **угрозы технологий ИИ (РГ1)**)
- ✓ методы тестирования

Обеспечение процесса тестирования:

- ✓ разработка и создание испытательных стендов
- ✓ разработка специального программного обеспечения для проведения испытаний технологий ИИ в различных сферах
- ✓ автоматизации процессов испытаний и оценки результатов

Существующие наработки

НТЦ ЦК

Разработана **Методика** оценки качества систем искусственного интеллекта, предназначенных для защиты от компьютерных атак на Web приложения.

Разработаны **Предложения** в проект требований по обеспечению информационной безопасности информационных систем, реализующих технологии искусственного интеллекта.

ИСП РАН

Разработана **Методика** и **ПО** тестирования фреймворков машинного обучения.

GIS ГАЗИНФОРМ СЕРВИС

Разработана **Матрица угроз** безопасности при разработке ПО с применением ML-моделей с описанием 61 угрозы и 141 доступного решения. Разрабатывается **ПО системы тестирования** моделей ИИ.

SWORDFISH SECURITY

Разработана **Методология** тестирования безопасности и доверенности AI-моделей. Разрабатывается **ПО системы тестирования** моделей ИИ.

МИРЭА Лаборатория Доверенного ИИ

Разрабатывается **ПО Автоматизированного тестирования** наборов данных.

РБО

На примере биометрических технологий разработаны **Стандартизированные подходы** по проведению независимых доверенных испытаний.

Тестирование технологических элементов ИИ

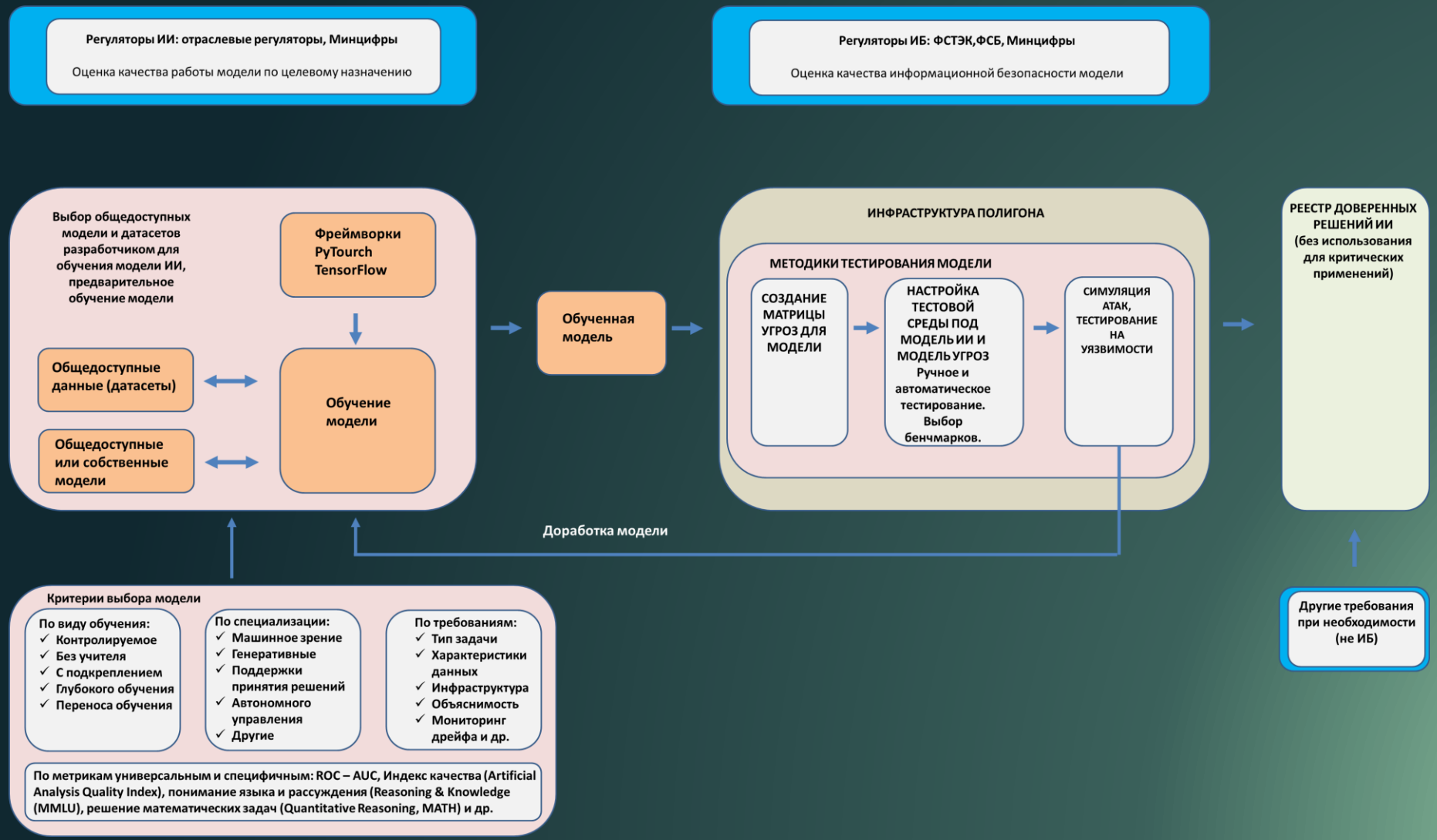
Объекты тестирования:
- Датасеты (обучающий, эталонный)
- Обученные модели
- Вспомогательное ПО (фреймворки МО и другое специализированное ПО)

Тестирование по направлению:
- Функциональность
- Надежность
- Защищенность

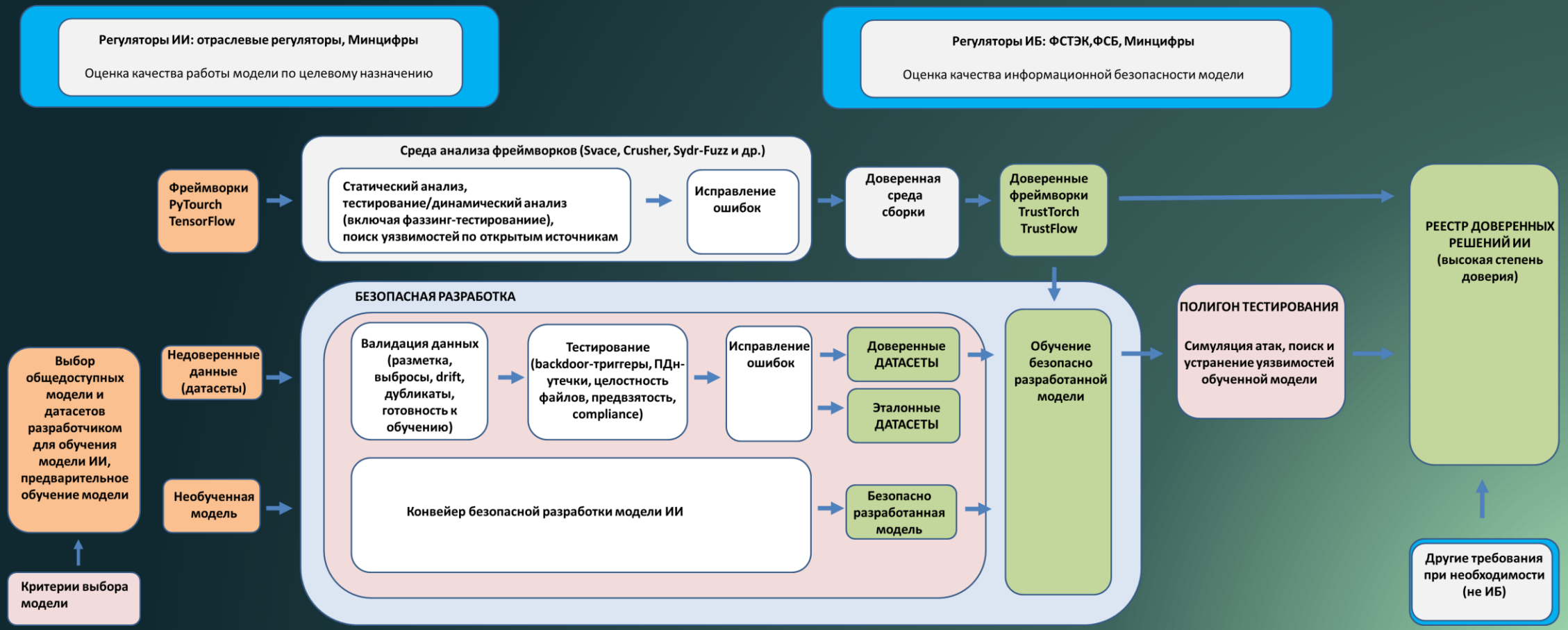
Тестирование на этапах жизненного цикла:
- Статическое
- Динамическое
- Фаззинг

Тестирование по виду проведения:
- Ручное
- Автоматизированное

Тестирование технологий ИИ (без использования в КИИ)



Безопасная разработка и тестирование технологий ИИ для КИИ (высокая степень доверия)



Что дальше?

Тестирование пилотных регионов

Разработка наборов метрик тестирования