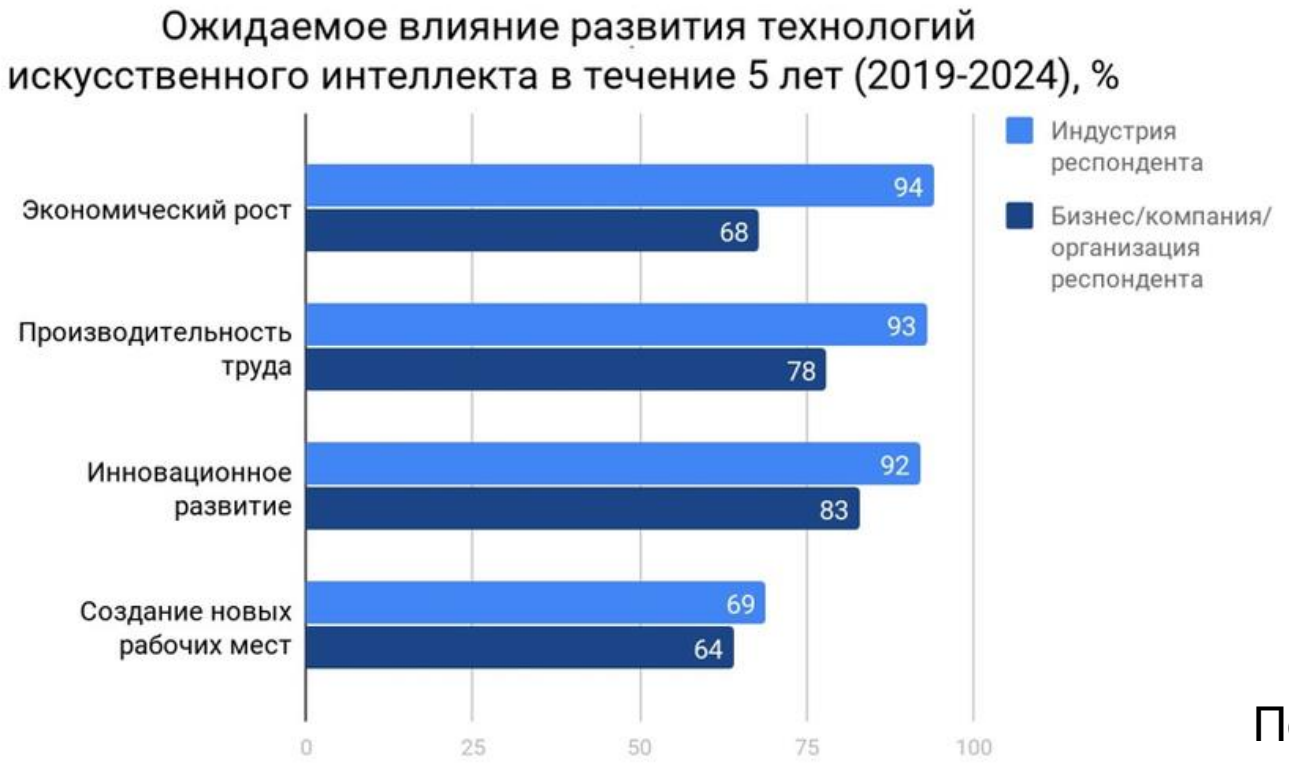
БЕЗОПАСНОСТЬ
информационных технологий
ФОРУМ - САНКТ-ПЕТЕРБУРГ

Подход к доказательству безопасности систем ИИ

к.т.н., доцент Попов
Илья Юрьевич

Статистика развития ИИ



По данным Tadviser

УПБ для систем ИИ различных областях применения

УПБ 1. Отказ системы приводит к ущербу продукции/оборудования.

Интенсивность отказов: 1 отказ на 100 т. часов.

УПБ 2. Отказ системы приводит к травматизмам персонала.

Интенсивность отказов: 1 отказ на 1 млн. часов.

УПБ 3. Отказ системы приводит к гибели персонала

Интенсивность отказов: 1 отказ в 10 млн. часов.

УПБ 4. Отказ системы приводит к техногенной катастрофе.

Интенсивность отказов: 1 отказ на 100 млн. часов.

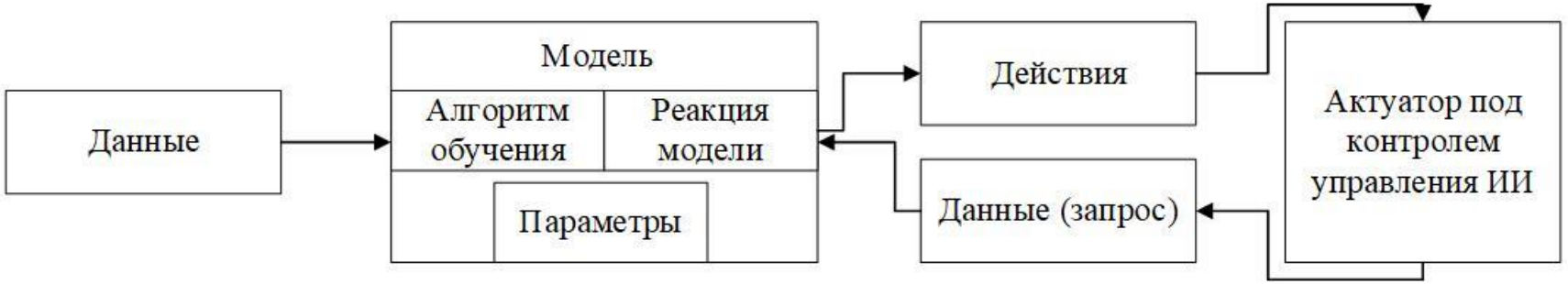
Название системы	Возможный УПБ
Система обработки данных	Уровень полноты безопасности присваивается в зависимости от того, какой конечный результат необходим в процессе обработки данных
Вспомогательные системы помощи	УПБ 0. Оператор всегда может отключить данный тип системы
Распознавание речи	Уровень полноты безопасности присваивается в зависимости от того, какие конечные функции задействованы
Система распознавания образов	Уровень полноты безопасности присваивается в зависимости от того, какие именно функции системы задействованы
Беспилотные транспортные средства	УПБ 1-4. БТС всегда имеет уровень УПБ, уровень зависит от территории эксплуатации.

Верификация и валидация

- ✓ Верификация - подтвердим соответствие продукта требованиям разработчиков (установленные требования выполнены).

- ✓ Валидация - подтвердим соответствие продукта конкретным требованиям использования (потребности пользователя).

Типовая модель системы ИИ



Данные, как основа система ИИ

1.Данные должны быть репрезентативны.

Обучение должно проходить в **типичной** среде для этого типа системы и окружения. Среда должна быть такая, чтобы воздействия были **типичными** для этого типа использования, включая все изменения в среде.

На данном шаге необходимо ввести ограничения среды, связанное с безопасностью работы системы.

2. Возможность дообучения системы после ввода в эксплуатацию.

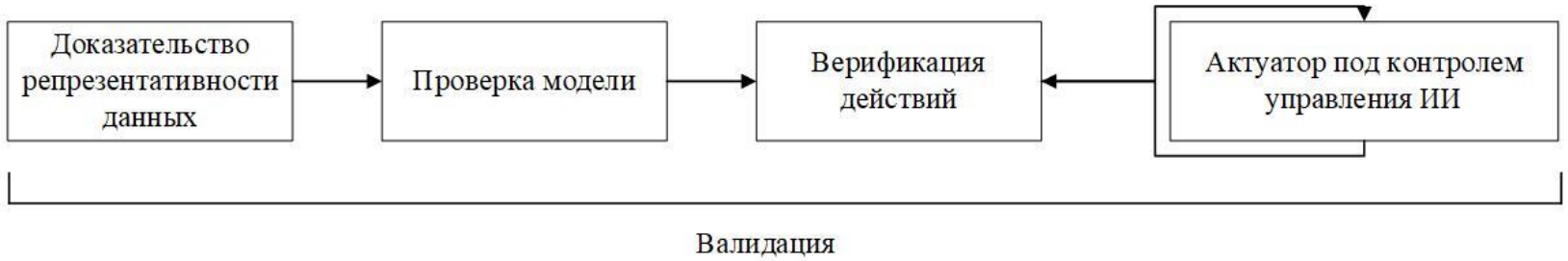
Модель ИИ

Модель системы ИИ - это её ядро. Модель должна быть правильно обучена, иначе на сколько бы не были репрезентативные данные, модель системы никогда не будет правильной. Необходимо точно удостовериться, что используемые данные репрезентативны для реальных данных.

На данном этапе необходимо использовать средства для обнаружения выбросов.

Необходимо использовать меру качества соответствия. Причем в некоторых случаях данная мера рассчитывается подгонкой, зависит от функции потерь.

Процесс валидации системы ИИ



Вывод

1.Система ИИ не имеет уровня полноты безопасности, так как её поведение не имеет критического значения.

2.Система ИИ поддерживается достаточно простой системой управления, которой необходимый высокий уровень полноты безопасности. Основной задачей такой системы должна являться проверка всех опасных решений по более простым алгоритмам и подавлять опасные реакции. То есть, данная система имеет отношение к безопасности и берет на себя полную ответственность за безопасность.

Спасибо

Попов Илья Юрьевич

ilyapopov27@gmail.com

+79216454390