



# Этические и правовые вопросы Искусственного Интеллекта



Андрей Кулешов,  
[a.kuleshov@phystech.ai](mailto:a.kuleshov@phystech.ai)



# Дискламация

- Представлены собственные взгляды автора
- Ошибки собственные ошибки автора
- Использованы материалы
  - CFA Institute
  - “Superintelligence” by Nick Bostrom, 2014
  - “The Ethics of Artificial Intelligence”, Machine Intelligence Research Institute
  - New England Complex Systems Institute
  - Рабочие документы ISO/SC42



# Центр НТИ ИИ

- Создан на базе МФТИ
- 7 приоритетных направлений
- Партнерство с крупными предприятиями, инновационными компаниями, ВУЗами
- Поддержка создания спин-офф компаний
- Подготовка специалистов ИИ с использованием «Систем Физтеха»



# Практические решения

Контекст

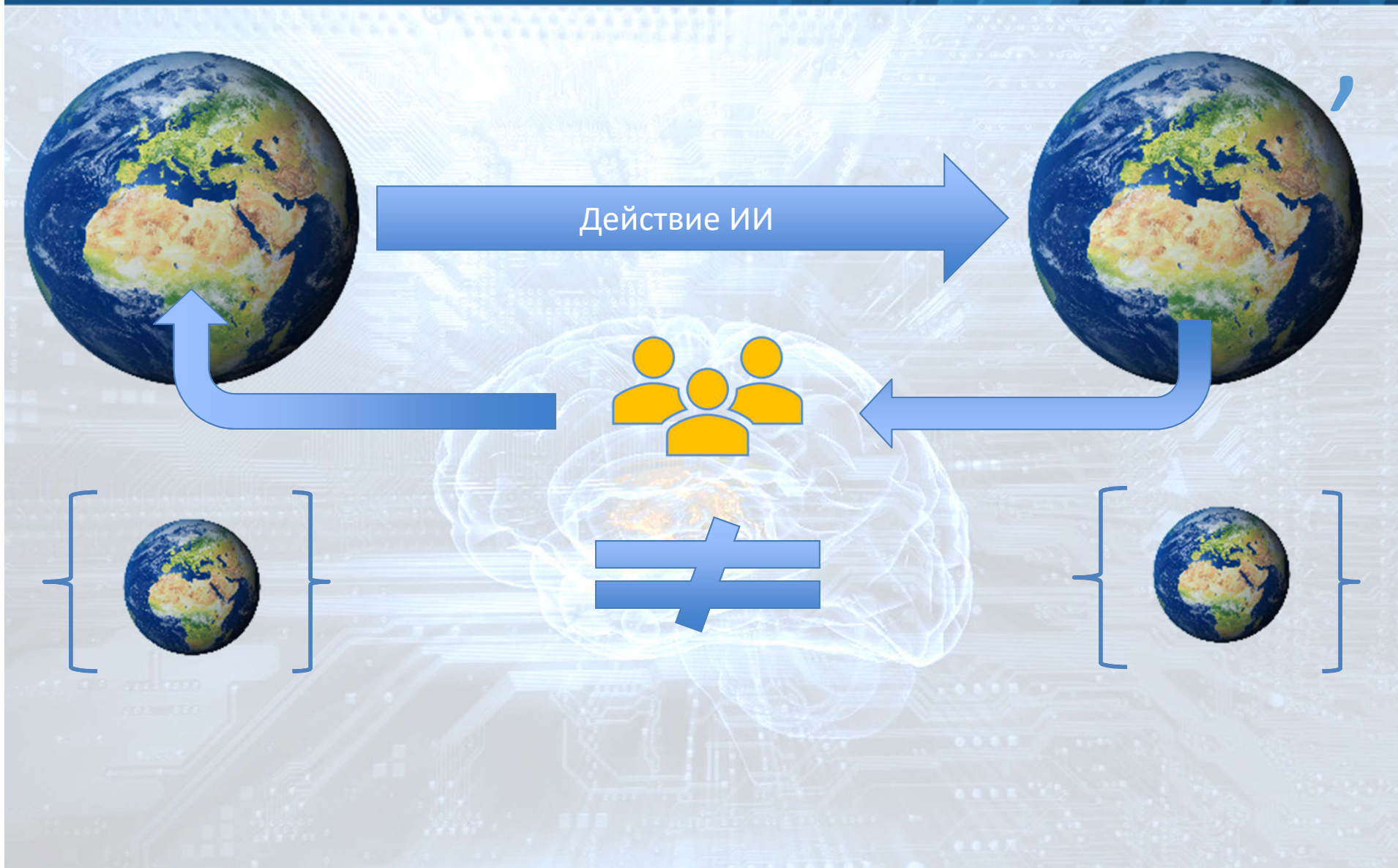
Общие  
принципы

Практические  
методы  
реализации

Стандартизация  
и Правовое  
оформление

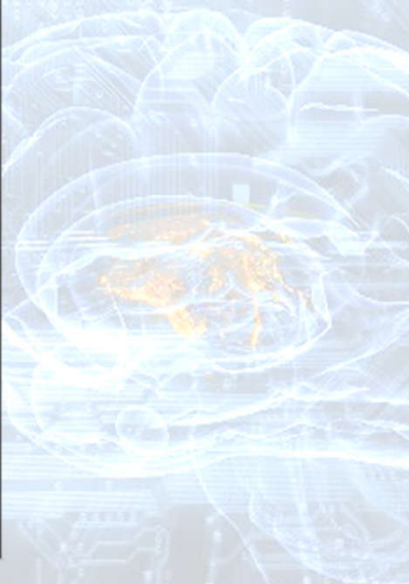


# Место этики в применении ИИ





# О принципах





Пример 1

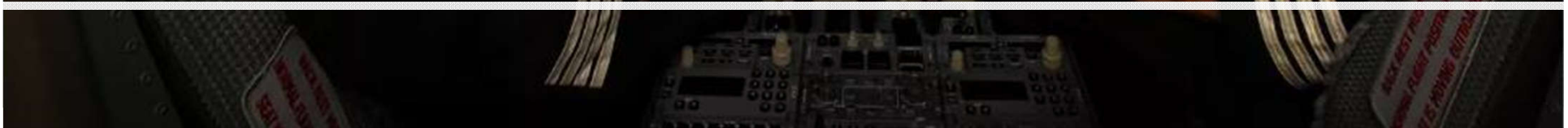


Пример 2





Пример 3





# Стандартизация и регулирование

- Функциональное соответствие
- Доверенность
- Безопасность
- Точность
- Робастность
- Прозрачность
- Предсказуемость
- Непредвзятость
- ...





# Стандартизация и регулирование

- Множество контекстов в которых допускается использование системы ИИ
- Тестовые контексты, и недопустимые состояния
- Критерии полезности системы и метрика позволяющая сравнивать полезность разных состояний



# Формирование метрик

- Явное представление
- Эволюционный выбор
- Стимулированное обучение
- Накопление значимости
- Мотивационные «подпорки»
- Узнавание значимости
- Модулированная эмуляция
- Институциональное проектирование

