

Альянс
в сфере
искусственного
интеллекта



**Проблема вагонетки:
беспилотные автомобили
(предварительное решение)**

Проблема вагонетки: беспилотные автомобили

Проблемы:

Основная задача алгоритма - не допустить столкновения в принципе:

- Нет разницы в штрафе для алгоритма за столкновение с живым/неживым объектом – алгоритм минимизирует риск столкновения;
- Разница в штрафе есть – алгоритм минимизирует ущерб живому объекту.

НО:

Но нужно ли различать живые объекты между собой? Такую систему ИИ нужно специально дорабатывать.

Решения:

1. Ориентироваться на то, что максимальная ценность - **человеческая жизнь**
2. Разрабатывать алгоритм таким образом, чтобы **минимизировать риски для человеческой жизни**, вызванные действием устройства под управлением алгоритма
3. При этом **не категоризировать** людей ни по каким признакам

1.

Допустимо ли в целом закладывать в систему ИИ этический выбор о причинении вреда разным людям или группе людей?

2.

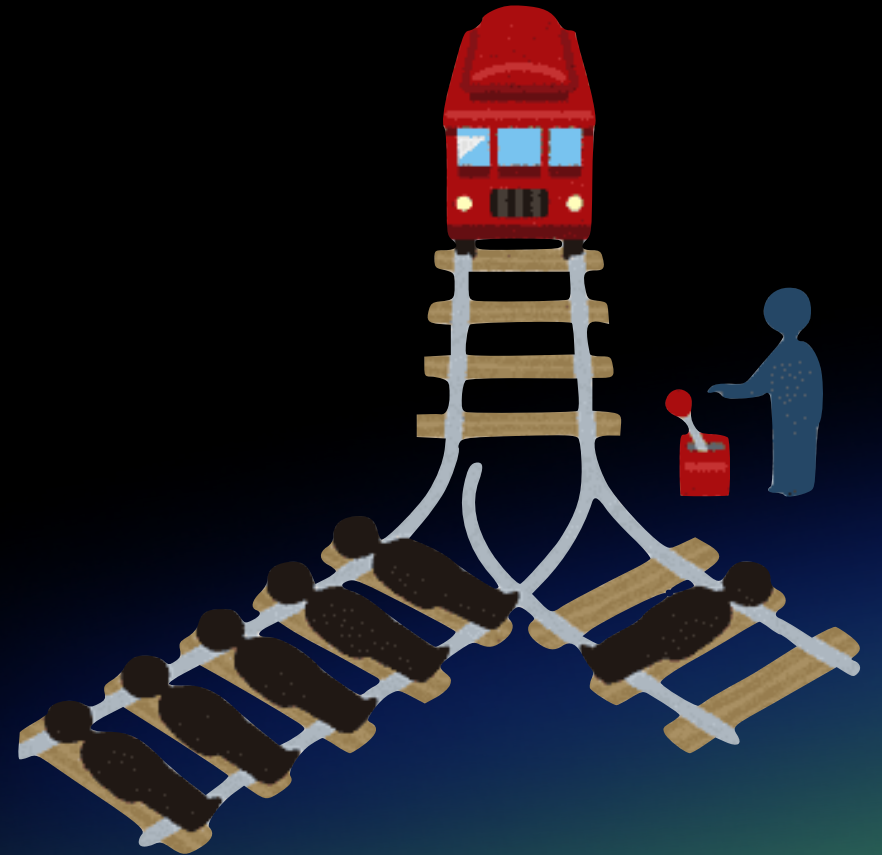
Каким должен быть профессиональный и моральный выбор разработчика?

Рекомендация

При необходимости решения аварийной ситуации СИИ всегда должен в максимальном приоритете **избегать опасности нанесения вреда человеку**, какие бы потери не последовали

При невозможности избежать человеческих жертв алгоритм должен стараться **минимизировать их число и значимость ущерба**

Не учитывать приоритеты при выборе между ущербом различным людям



Итоги голосования подписантов Кодекса

62%

ЗА принятие Рекомендации в исходном виде

16%

ЗА, с учетом замечаний

19%

ПРОТИВ принятия Рекомендации в исходном виде и **ЗА** дальнейшую проработку

3%

ВОЗДЕРЖАЛИСЬ



По итогам голосования решено:

Доработать предварительную Рекомендацию, с проведением дополнительного исследования