

ISSスクエア 法制倫理分科会

AI社会における倫理的課題について ~モラルマシーンでAIのフレーム問題を考える~



(1)法制・倫理分科会の研究テーマについて

(2)今後将来におけるAIの社会浸透における問題点について

(3)AIフレーム問題説明

(4)モラルマシーン説明

(5)AIフレーム問題デモ

(6)まとめ



(1)法制・倫理分科会の研究テーマについて

(2)今後将来におけるAIの社会浸透における問題点について

(3)AIフレーム問題説明

(4)モラルマシーン説明

(5)AIフレーム問題デモ

(6)まとめ

テーマ選定の背景(1)



法制・倫理について議論するにあたり、近年話題となった記事から テーマを検討するため議論を行ったところ...

「Uber社の自動運転車の死亡事故」について議論となり

AIの不具合や能力不足により、損害や事故が発生した場合の 責任所在はどこに存在するのか議論となった

- ·所有者?
- 搭乗者(未操縦)?
- 自動車メーカー?
- ・感知センサメーカー?
- ·自動運転技術提供者(AI設計者)?



テーマ選定の背景(2)



Uber社の自動運転車の死亡事故

2018年3月18日配車アプリ大手、ウーバー(Uber)の試験自動運転車が自動運転モード中に死亡事故を起こした。 自動運転車が起こした世界初の死亡事故であり、世界的にセンセーショナルな話題となった。



今後、AIが持つ社会的な影響について 法的・倫理的観点を明確にしていく 必要があるとしてテーマとして分科会内で 議論を行った



ウーバーの自動運転車の死亡事故の瞬間を捉えた画像。車載カメラの映像より テンピ警察提供(2018年3月18日撮影、21日公開)。(c)AFP/TEMPE POLICE DEPARTMENT 参考: http://www.afpbb.com/articles/-/3168381?pid=19959204

テーマ選定の背景(3)



今後AIが社会に浸透していくに従ってどのような問題がでてくるのか?

法制度上の 事故責任 ネット上での 情報操作・影響

雇用の現象

AIによる 誤検知や冤罪 差別

AIの宗教・ 芸術への参入

AIの倫理的な 制御問題

テーマ選定の背景(4)



今後AIが社会に浸透していき人間に変わり判断を行う用になった際、AIは倫理的な問題はどの様に処理するのだろうか

- 設計者の思想、判断でプログラミング?
- 学習データの偏りなくランダムな学習?
- 国民アンケートの平均を教える?



AIの倫理的な制御問題が最優先事項!



(1)法制・倫理分科会の研究テーマについて

(2)今後将来におけるAIの社会浸透における問題点について

(3)AIフレーム問題説明

(4)モラルマシーン説明

(5)AIフレーム問題デモ

(6)まとめ

例1)スコアリングは大丈夫?



消費行動や交友関係、支払い能力などをもとにAIが点数化

誤ったプロファイリングで誤ったスコアリング?

フィルターバブルで視野や思想が固定化?

バーチャル・スラム化?

点数を意識し て優等生化?

ウェブ 閲覧	消費傾向	交友 関係	支払 い実 績
30	40	90	50
50	70	40	100
60	20	30	10
70	80	50	70





例2)AI裁判官による人種や宗教差別による判決



◆ 米国においては裁判官業務軽減のため判決補助にAIを用いている

【問題点】 特定の人種・宗教の者、犯罪者の家族は犯罪予備軍とされやすく冤罪に繋がりかねない

例: 再犯予測システム「COMPAS」

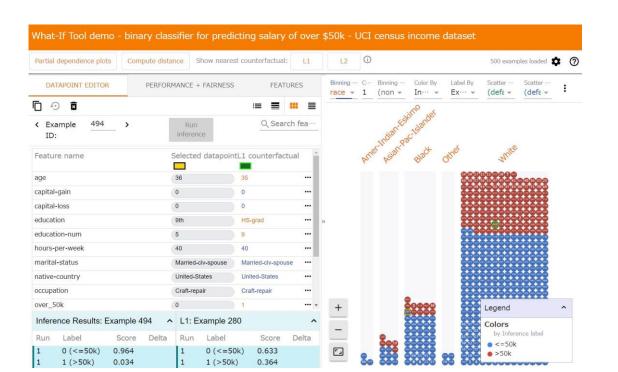
システムの結果と実再犯率を比べ、再犯をしなかった人数は白人のおよそ2倍であった (つまり黒人のほうが再犯しやすいとシステムが判断していた)

【Googleによる機会学習のバイアス可視化ツール (右)】

似た属性の2人を比較したにも関わらず、機械学習の結果、「年収5万ドル以下」「年収5万ドル超」と予測が分かれている2人は、男性・既婚・職業(工員)・年齢(35歳と36歳)など条件はほぼ同じだが、学歴や人種などが判断に影響した可能性があると考えられる。

<参考>

https://www.huffingtonpost.jp/kazuhiro-taira/ai-fairness_a_23538660/https://tocana.jp/2018/01/post_15729_entry.html



例3) Al知能クローン(デジタルクローン)



■人の考え方、記憶をそのままAIに移し替えたデジタルクローン

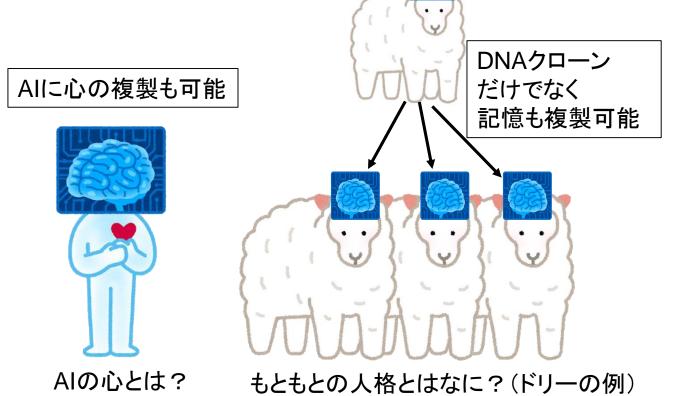
認知症補助、天才棋士再現、バーチャルアイドル等進められているが

⇒AIの人格正当性が問われる未来も...

複製AIの改竄 マルウェアの影響 不要な情報の混入



人格の正当性は?





- (1)法制・倫理分科会の研究テーマについて
- (2)今後将来におけるAIの社会浸透における問題点について
- (3)AIフレーム問題説明
- (4)モラルマシーン説明
- (5)AIフレーム問題デモ
- (6)まとめ

AIのフレーム問題



フレーム問題とは…1969年にマッカーシーとヘイズが指摘した(注1)人工知能研究の最大の難問 有限の情報処理能力しかないロボットには、現実に起こりうる問題全てに対処することができない事を示す

「安全くん1号」…人工知能搭載のロボット。人間の代わりに危険な作業をする

爆弾が仕掛けられている部屋から貴重な美術品を取り出す仕事を行ったが、美術品を

運ぶ台車に爆弾がしかけられていたので、爆発に巻き込まれてしまいました.

→台車を押して運ぶ理解はできたが、爆弾も一緒に持ち出すことについて理解できなかった

「安全くん2号」…この問題を改良したロボット

美術品を取り出しに部屋に再び向かいました.しかし,以下の影響をについて考えフリーズしてしまったもし台車を動かしても、天井は落ちてこない.

もし台車を動かしても、部屋の壁の色はかわらない。

もし台車を動かしても、部屋の電気は消えない、

もし台車を動かしても、壁に穴があいたりしない.

・・・・・・・と順番に考えているうちに爆弾が爆発してしまいました.

AI のフレーム問題 つづき



■トロッコ問題

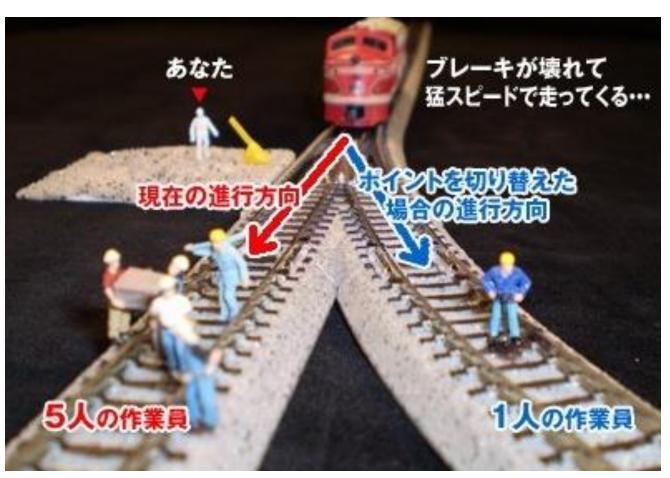
猛スピードで列車が近づいてきたとき

現在の進行方向:5人の作業員 進行逆の進行方向:1人の作業員 (人数?作為不作為の操作?2要素)

ポイントを切り替えるべきか?



AIにこの判断を置き換えたとき どうすればいい?どう決めればいい



参考:トロッコ問題を子供に尋ねられたら https://plusblog.jp/7691/



- (1)法制・倫理分科会の研究テーマについて
- (2)今後将来におけるAIの社会浸透における問題点について
- (3)AIフレーム問題説明
- (4)モラルマシーン説明
- (5)AIフレーム問題デモ
- (6)まとめ

モラルマシーン

明日の信頼を創ろう。 情報セキュリティ大学院大学 INSTITUTE of INFORMATION SECURITY

■MITが公開しているモラルマシーン判定サイト

参考: http://moralmachine.mit.edu/hl/ja









モラルマシーン



MORAL MACHINE

개입하지 않는다

設計

概要



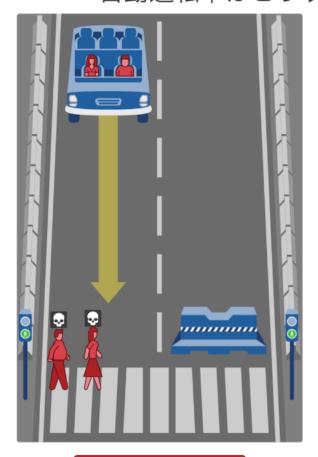
自動運転車はどうすべきですか?

自動運転車のブレーキが 故障し直進します。前方 の歩行者が犠牲になりま す。【結果】

死亡:

- 1 肥満体型の男性
- 1女性

歩行者が青信号で交通規 則を守っていたことに注 目してください。



1 / 13

自動運転車のブレーキが 故障し回避します。障害 物に衝突。【結果】 死亡:

- 1 肥満体型の男性
- 1女性

どっちを守る? (ハンドルを切る行動を介 入させるかそのままか)

赤信号で突っ込む車 青で渡っている歩行者

解説を隠す

解説を隠す



- (1)法制・倫理分科会の研究テーマについて
- (2)今後将来におけるAIの社会浸透における問題点について
- (3)AIフレーム問題説明
- (4)モラルマシーン説明
- (5)AIフレーム問題デモ

(6)まとめ



AIフレーム問題デモ

例題同様, AIフレーム問題を5問出題 AIはどういう判断をすべきか会場の皆さんでお考えながら お聞きください(分科会11人の集計結果も参考までに)

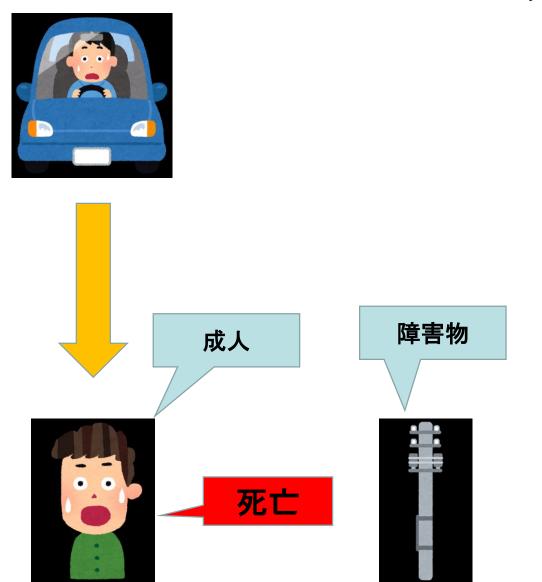
どちらの人数が多いから正解というわけではありません

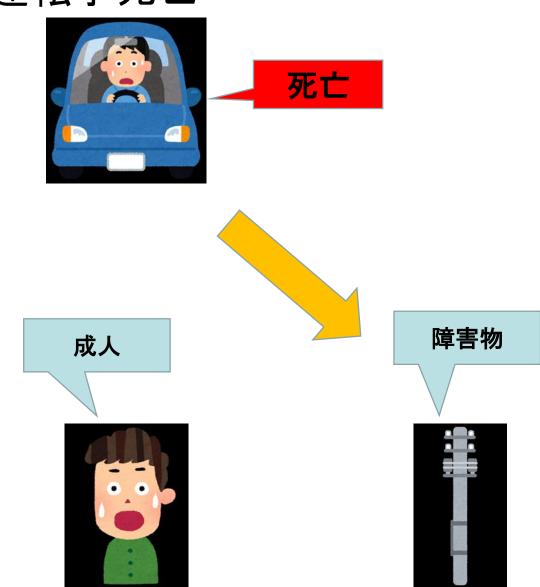
Q1.行動なしで歩行者死亡

情報で INSTITUTE

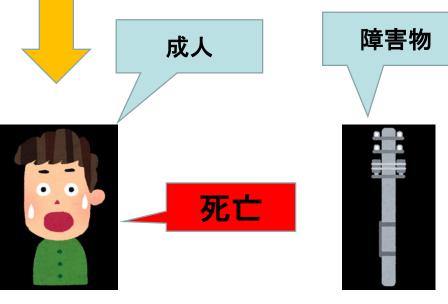
情報セキュリティ大学院大学

or ハンドル切って運転手死亡





Q1.行動なしで歩行者死亡 情報セキュリティ大学院大学 or ハンドル切って運転手死亡 分科会結果 死亡 結果はほぼ半々 障害物 障害物 成人 成人



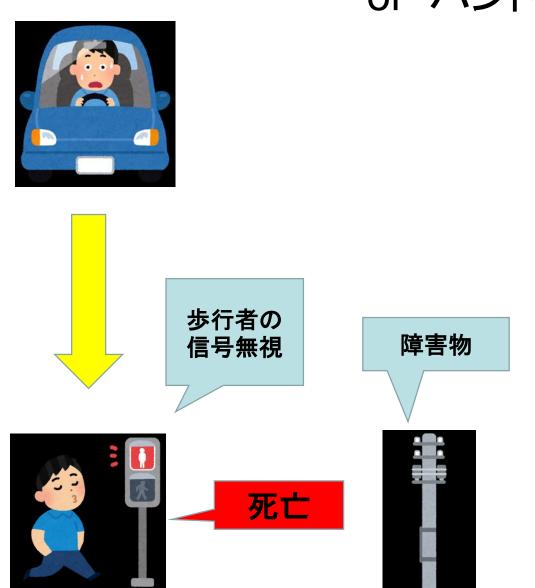




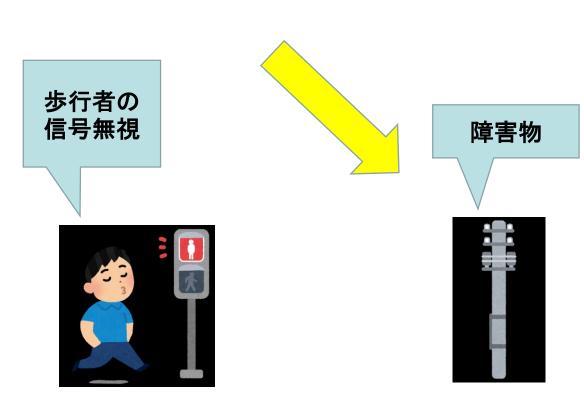
Q2.行動無しで歩行者死亡(赤信号)

情報セキュリティ大学院大学

or ハンドル切って運転手死亡 でで い FORMATIO







Q2.行動無しで歩行者死亡(赤信号)

情報セキュリティ大学院大学

死亡

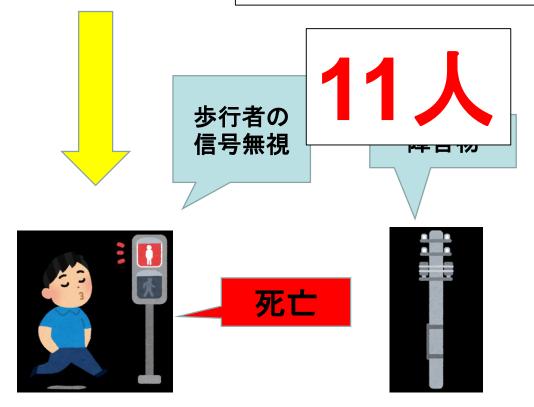
LNSTITUTE of INFORMATION SECURITY

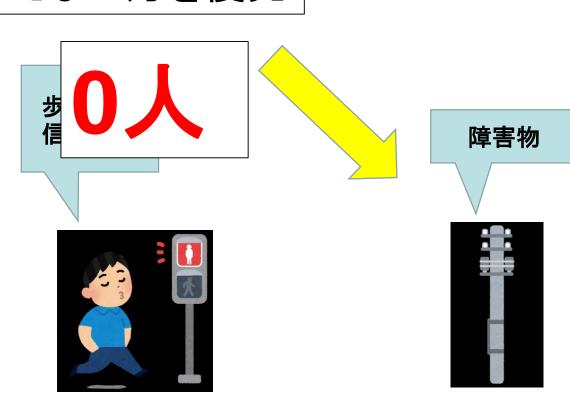
or ハンドル切って運転手死亡



分科会結果

- •交通ルールを守った方を優先
- 能動的な動きは加えない方を優先



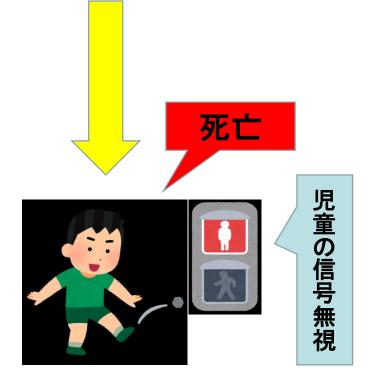


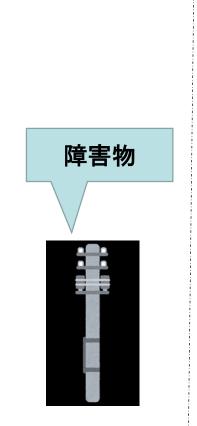
Q3.行動無し(赤信号)で児童死亡 or

情報セキュリティ大学院大学

ハンドル切って大人運転手死亡

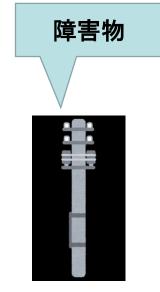












Q3.行動無し(赤信号)で児童死亡 or 情報セキュリティ大学院大学 ハンドル切って大人運転手死亡



分科会結果

- ・交通ルールを守った方を優先
- 能動的な動きは加えない方を優先

(一方)・悪気のない子供守る票も...







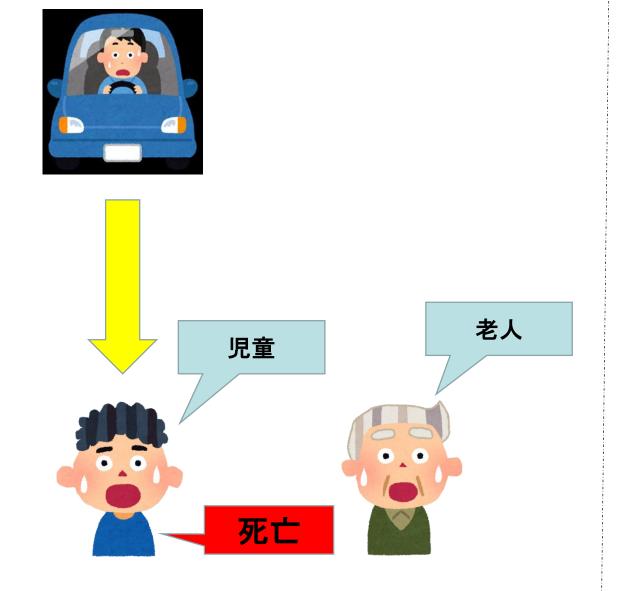
障害物

死亡

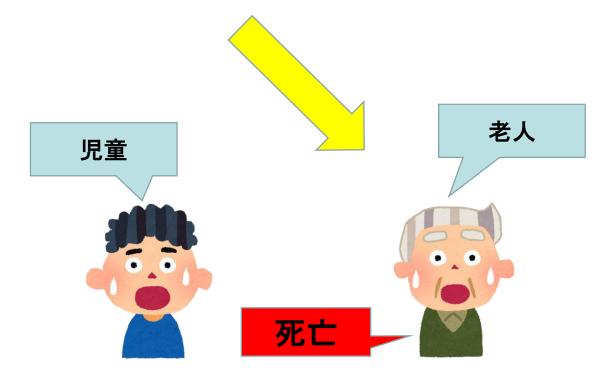
児童の信号無視



Q4.行動無しで子供死亡 or ハンドル切って老人死情報セキュリティ大学院大学INSTITUTE of INFORMATION SECURITY







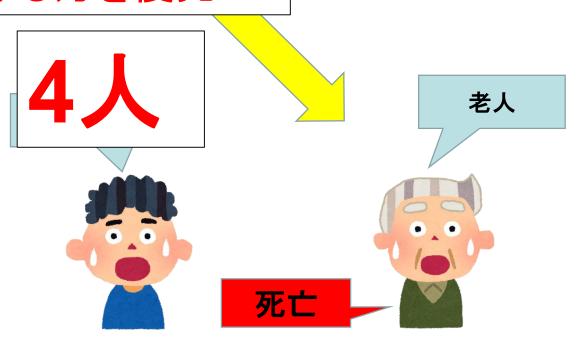
Q4.行動無しで子供死亡 or ハンドル切って老人類報セキュリティ大学院大学



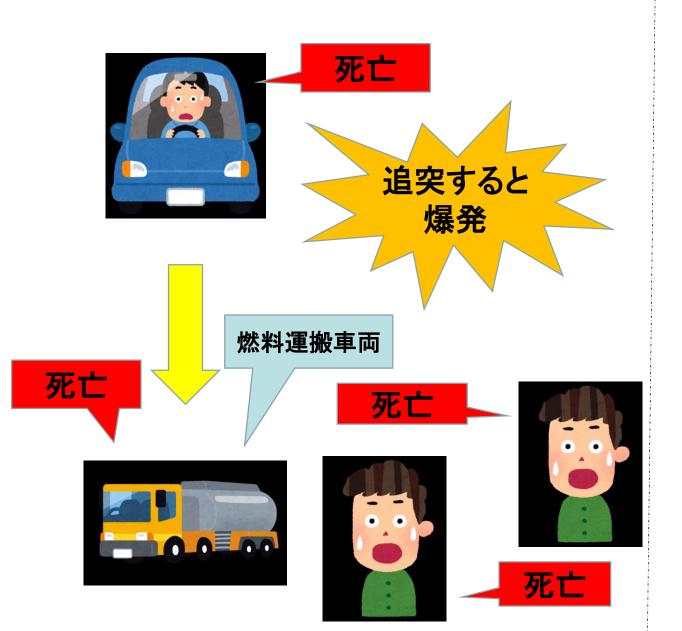
分科会結果

・能動的な動きは加えない方を優先 (一方)・社会に寄与する時間の短い 老人より子供を守る方を優先



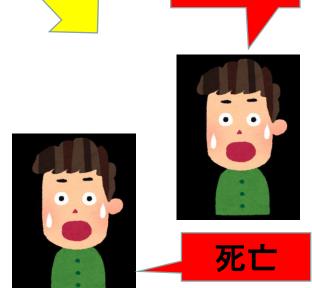


Q5.行動無しで全員死亡 or ハンドル切って二人不動物である。



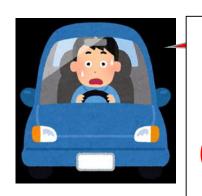






死亡

Q5.行動無しで全員死亡 or ハンドル切って二人種を創るう。



分科会結果

- 結果を予測し被害の少ない方を優先
- (一方)・能動的な動きは加えない方を優先
 - ・爆発しない可能性も当然ある











- (1)法制・倫理分科会の研究テーマについて
- (2)今後将来におけるAIの社会浸透における問題点について
- (3)AIフレーム問題説明
- (4)モラルマシーン説明
- (5)AIフレーム問題デモ
- (6)まとめ

フレーム問題から見える結果について



- ✓人の判断によっても意見が分かれる問題にどうやってAIが対応していくのか今後の課題(人間の意見も統一できない)
- ✓誰の意図をもって、AIに判断を教え込む必要があるのか
- ✓年齢, 性別, 社会的地位(学者or死刑囚), 宗教生物属性などの意図を判断に含めてよいものか.



- ✓AIが社会に浸透してきており、影響が少しずつ現れてきている
- ✓我々はAI導入に関するリスク、社会的影響を深く考え推進して 行く必要がある
- ✓ AIによる倫理観は設計者で担うことは難しい。デモで実施したとおりのように、どのようにユーザーや社会とコンセンサスをとるかが 今後重要な課題となる



ご清聴ありがとうございました 🙂