

AI Ethics Issues in Real World: Evidence from AI Incident Database

Mengyi Wei

Chair of Cartography and Visual Analytics

Technical University of Munich

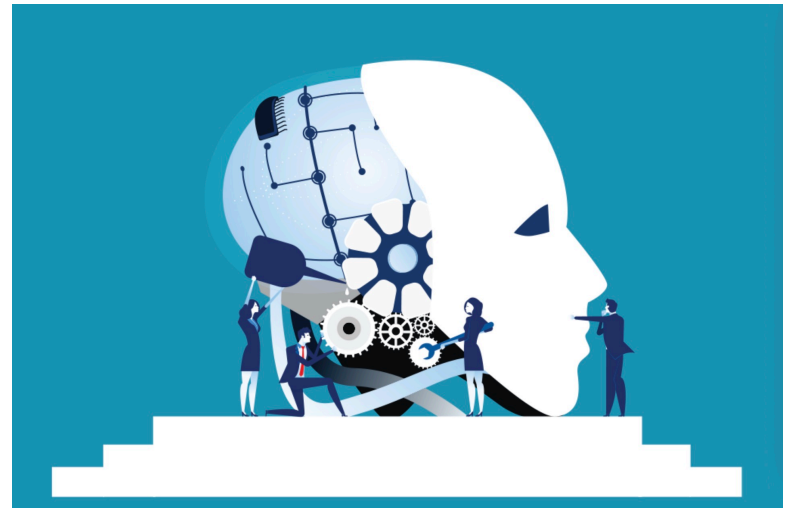
AI Ethics Issues in Real World: Evidence from AI Incident Database

- Introduction
- Method
- Results
- Limitations and Outlook

AI Ethics Issues in Real World: Evidence from AI Incident Database

➤ Introduction

- AI technology has become a mega trend.
- AI guidelines are too theoretical and disjointed from practical problems.
- How AI ethics issues take place in real world?



➤ Method

How to describe AI ethics incidents? ➡ Build a taxonomy of AI ethics incidents in real world

- Data Collection:

150 AI ethics incidents (from: AI Incident Database)

4 attributes: Time, Geographic locations, Application areas, Taxonomy of AI ethics issues

- Content Analysis:

Kappa is computed as:

$$\kappa = \frac{P_A - P_c}{1 - P_c}$$

where:

P_A = proportion of units on which the raters agree

P_c = the proportion of units for which agreement is expected by chance.

| <u>Kappa Statistic</u> | <u>Strength of Agreement</u> |
|------------------------|------------------------------|
| <0.00 | Poor |
| 0.00– 0.20 | Slight |
| 0.21– 0.40 | Fair |
| 0.41– 0.60 | Moderate |
| 0.61– 0.80 | Substantial |
| 0.81– 1.00 | Almost Perfect |

➤ **Method**

Table 1 Krippendorff's alpha for each variable

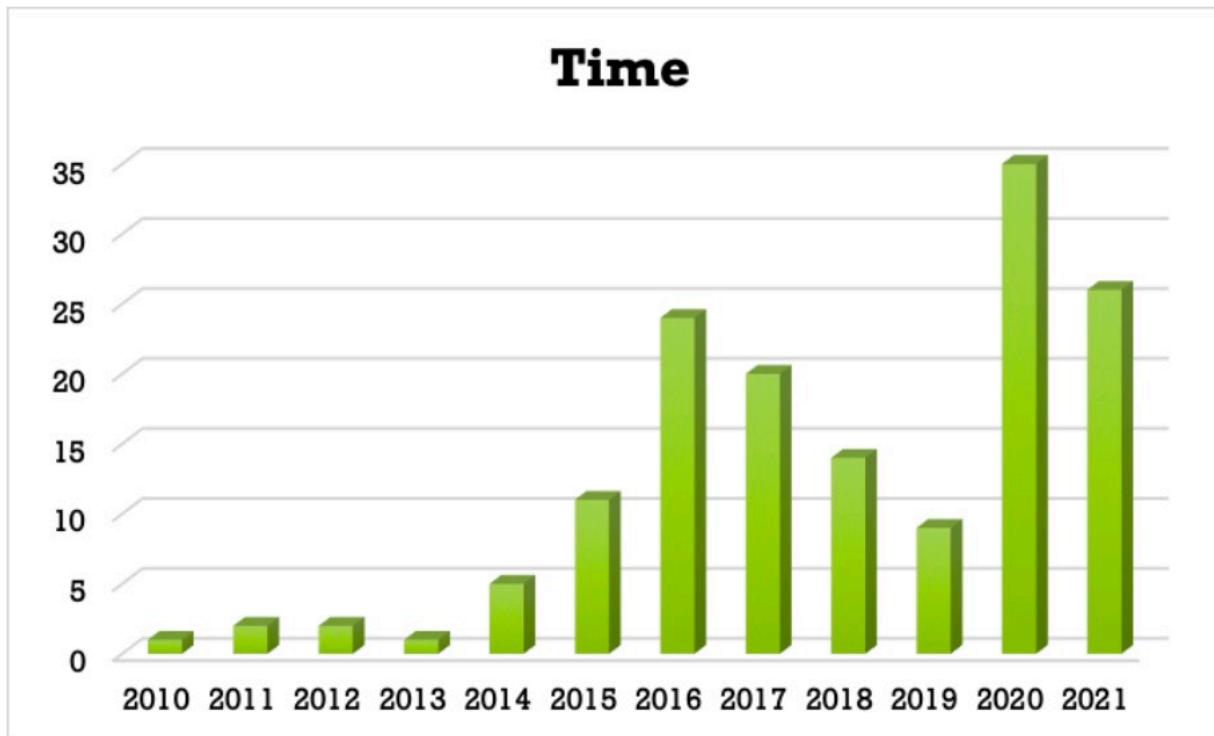
| Content Category | Krippendorff's Alpha |
|------------------------------------|----------------------|
| AI supervision | 0.79 |
| AI recruitment | 0.44 |
| Identity Authentication | 1 |
| Language/vision model | 0.98 |
| Intelligent recommendation | 0.96 |
| Autonomous Driving | 1 |
| Intelligent Service Robots | 1 |
| Smart Healthcare | 1 |
| AI Education | 1 |
| Predictive policing | 1 |
| Smart Home | 1 |
| AI Game | 1 |
| Smart Finance | 1 |
| Privacy | 1 |
| Inappropriate Use(Bad Performance) | 0.90 |
| Unethical Use(illegal Use) | 0.97 |
| Racial Discrimination | 1 |
| Gender Discrimination | 0.98 |
| Unfair Algorithm (Evaluation) | 0.94 |
| Mental Health | 0.86 |
| Physical Safety | 1 |
| Average | 0.94 |

0.94 > 0.8

Almost Perfect

➤ Results

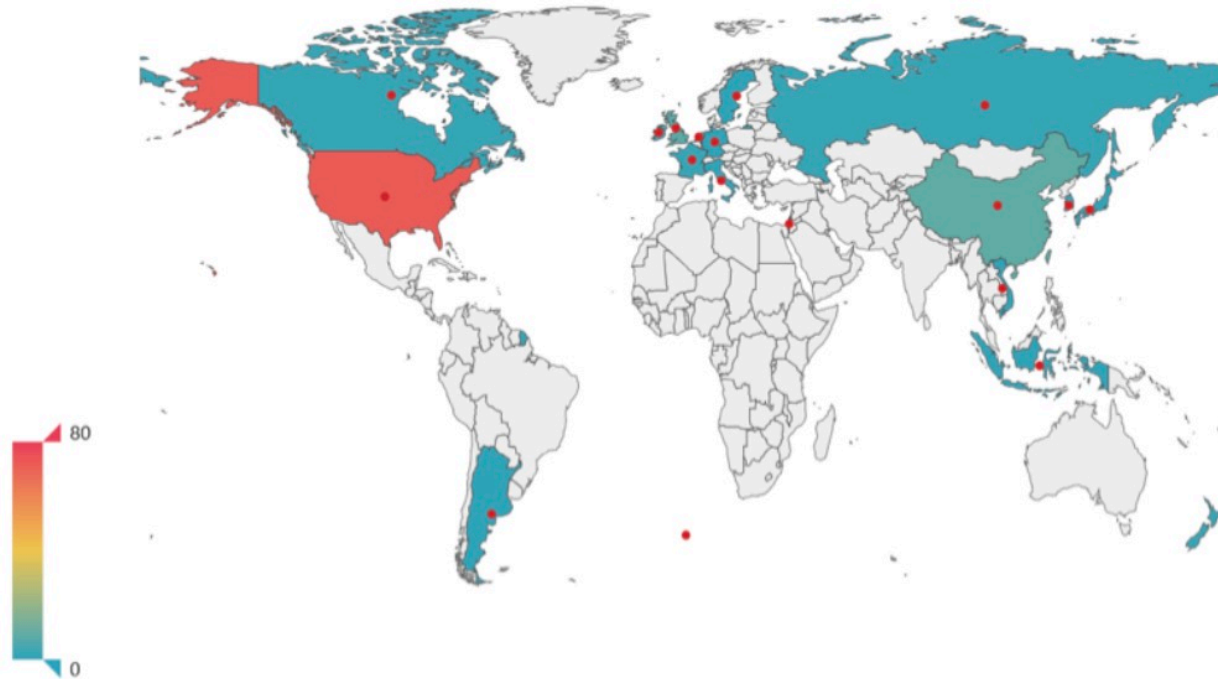
- Temporal evolution of AI ethics incidents



➤ Results

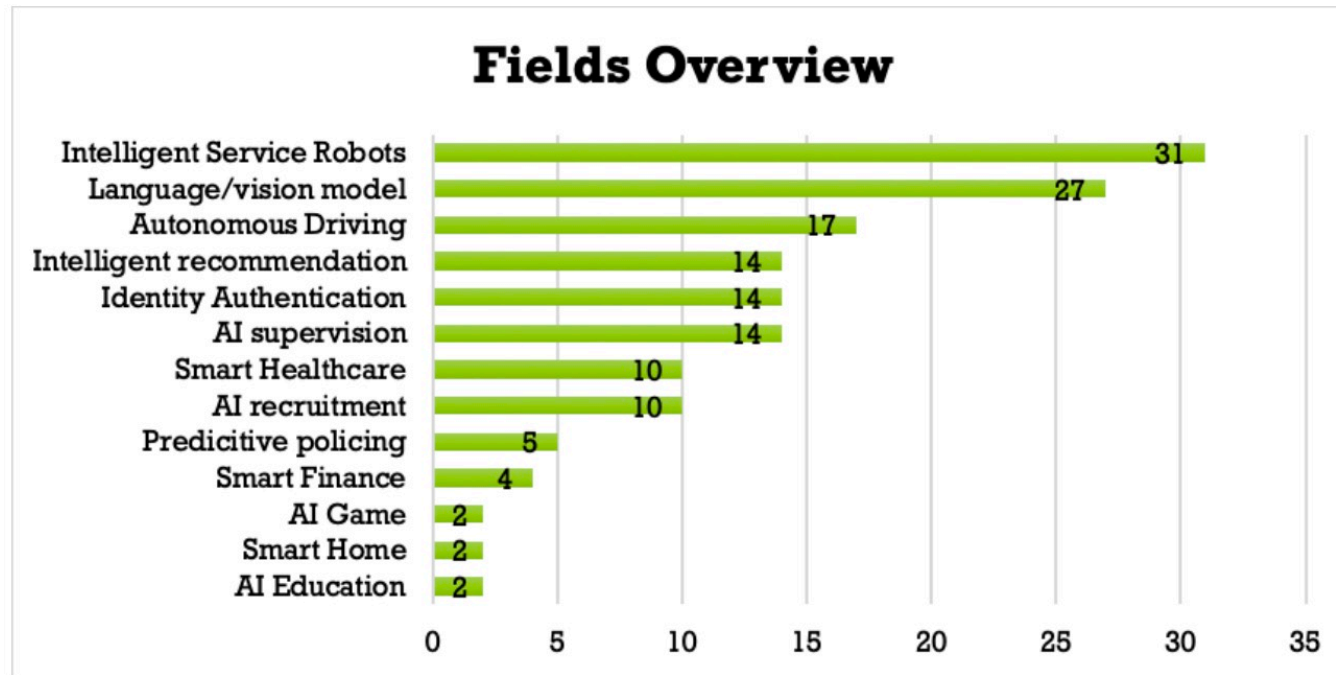
- Geographic distribution of AI ethics incidents

Distribution Map of Global AI ethics cases  World Map



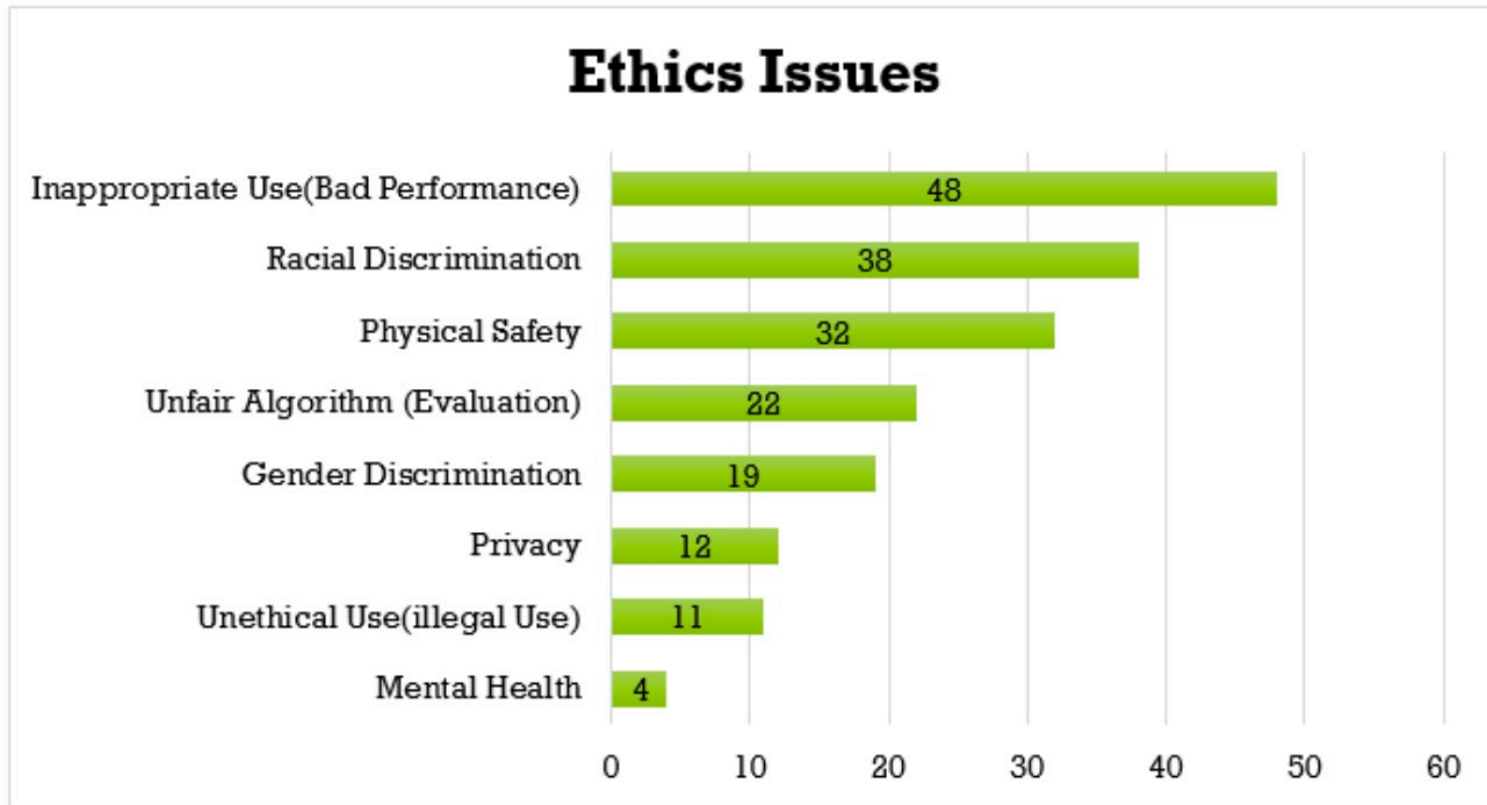
➤ Results

- Application areas of AI ethics incidents



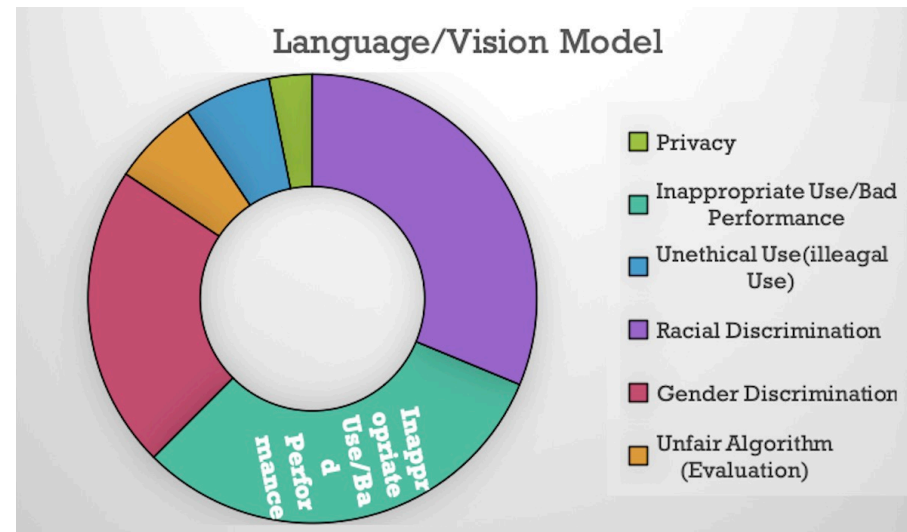
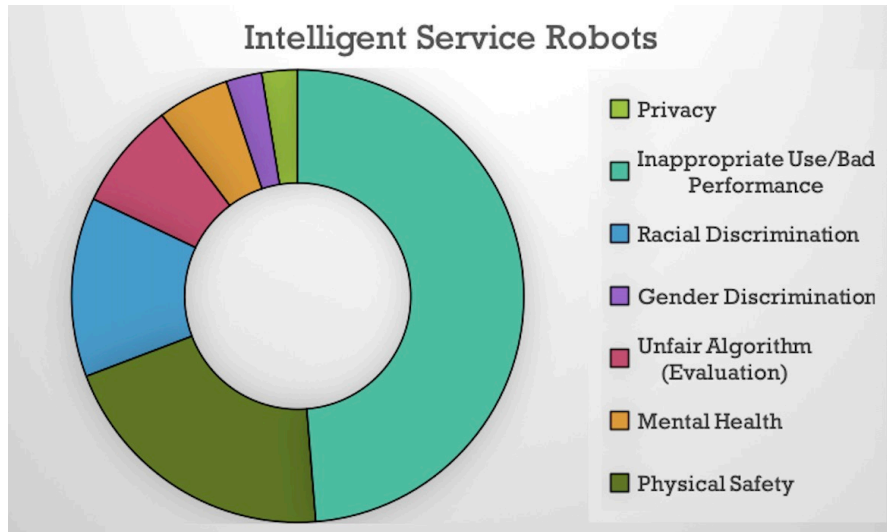
➤ Results

- Taxonomy of AI ethics issues



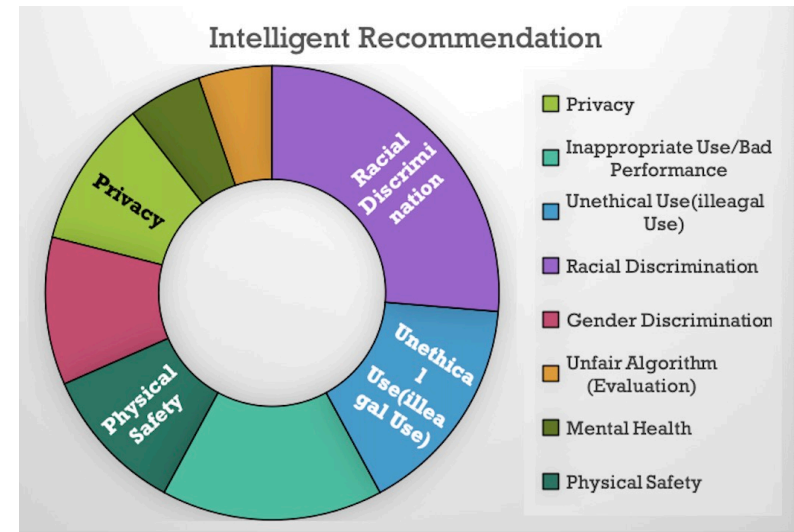
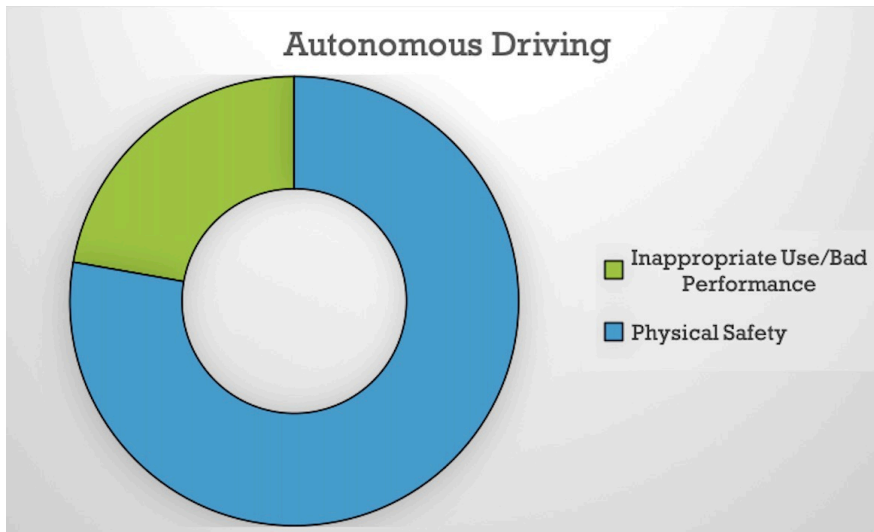
➤ Results

- Distribution of AI ethics issues in different fields



➤ Results

- Distribution of AI ethics issues in different fields



➤ **Limitations and Outlook**

- **Limitations:**

- The size and variety of data are limited.
- Only manually analyze the AI incident database

- **Outlook:**

- Expand the number of AI ethics incident
- Build NLP models to analyze topics and sentiments
- More work to refine the theoretical and operable parts of the guidelines

Thank you !

