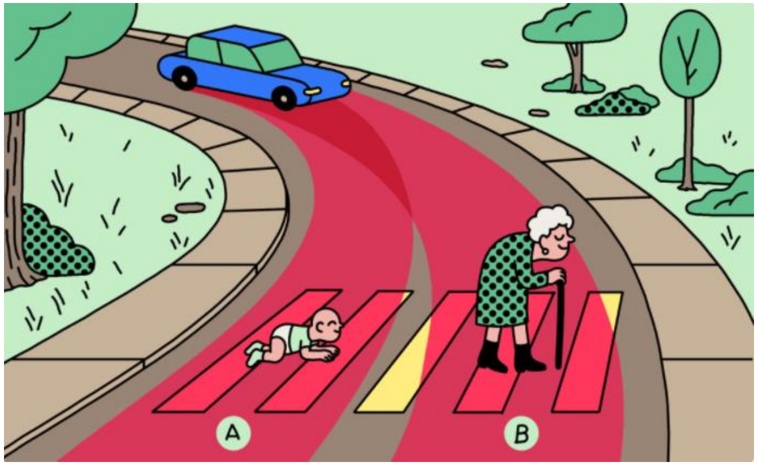
AI Ethics Issues in Real World: Evidence from AI Incident Database

Mengyi Wei

Chair of Cartography and Visual Analytics

Technical University of Munich

mengyi.wei@tum.de



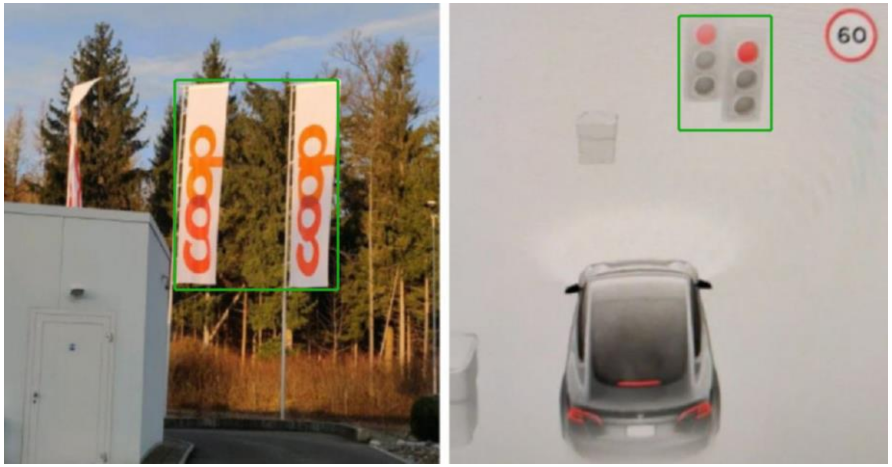
AI Ethics Issues in Real World: Evidence from AI Incident Database

- 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 and how repetitive AI failures can be mitigated?



Self-driving car mistakes red letters on flag for red traffic lights

➢ Method

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

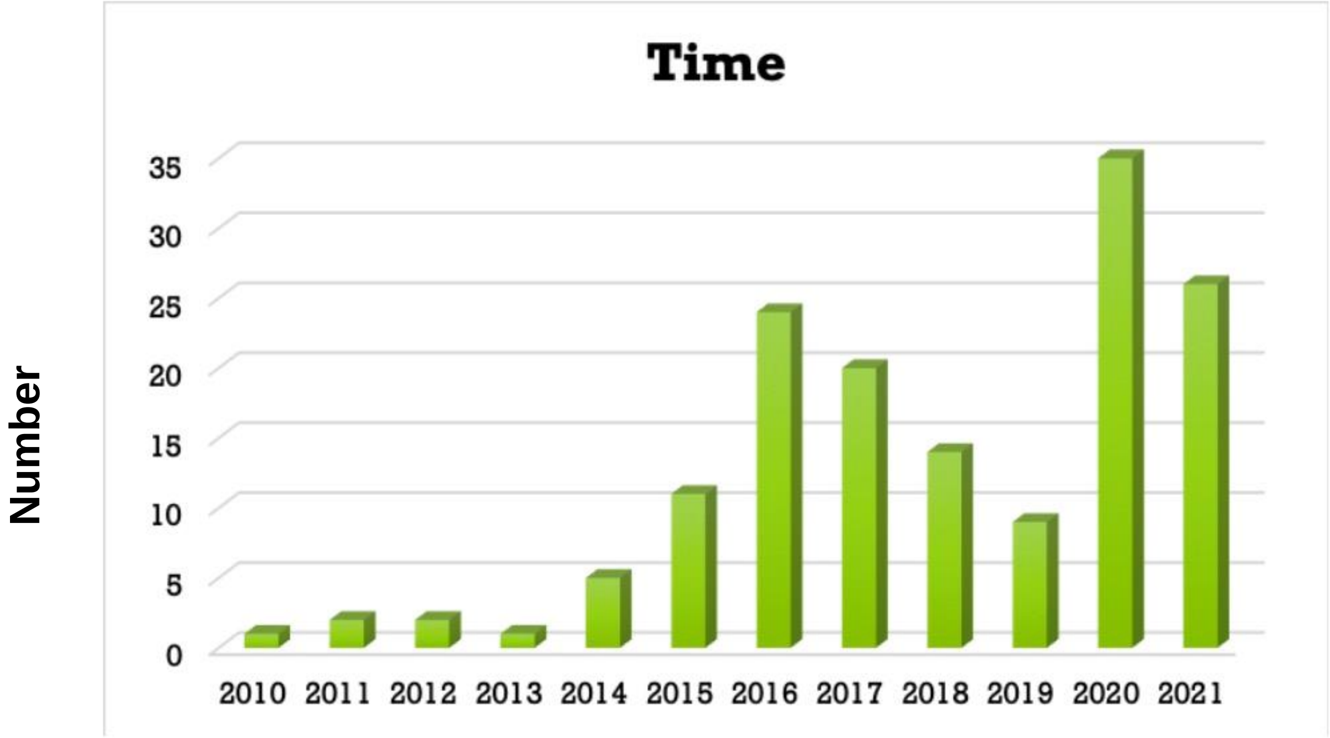
- ## Data Collection:

150 AI ethics incidents from AI Incident Database (https://incidentdatabase.ai/)

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

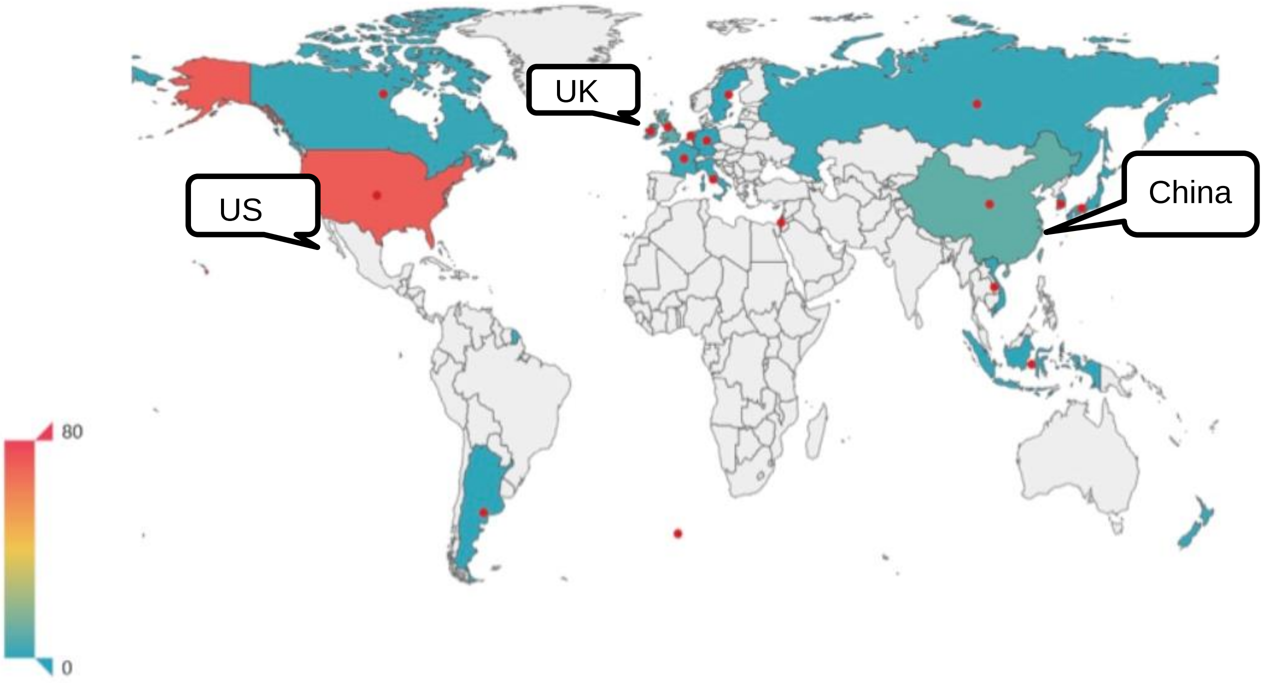
No	Title	Time	Location	Application areas	AI ethics issue
46	Robot passport checker rejects Asian man's photo for having his eyes closed	2016.12.07	New Zealand	Identity Authentication	Racial Discrimination

- Temporal evolution of AI ethics incidents



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

- Geographic distribution of AI ethics incidents



Cumulative number of AI ethics incidents from 2010 to 2021

➢ Method

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

- ## Content Analysis:

Krippendorff’ s alpha is computed as:

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

where:

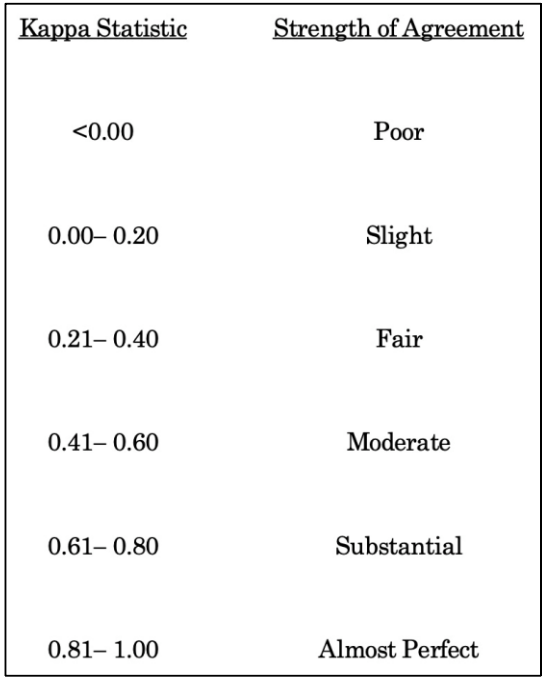
P_A = proportion of units on which the raters agree

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

➢ Method

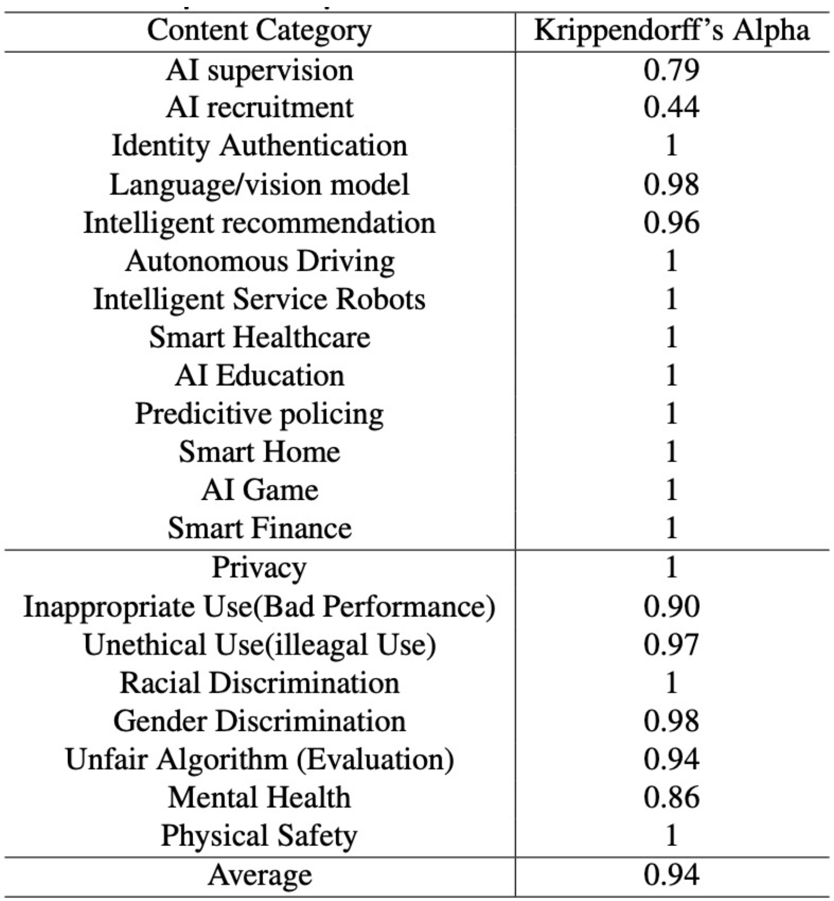
0.94 > 0.8

Almost Perfect



Kappa Statistic	Strength of Agreement
<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

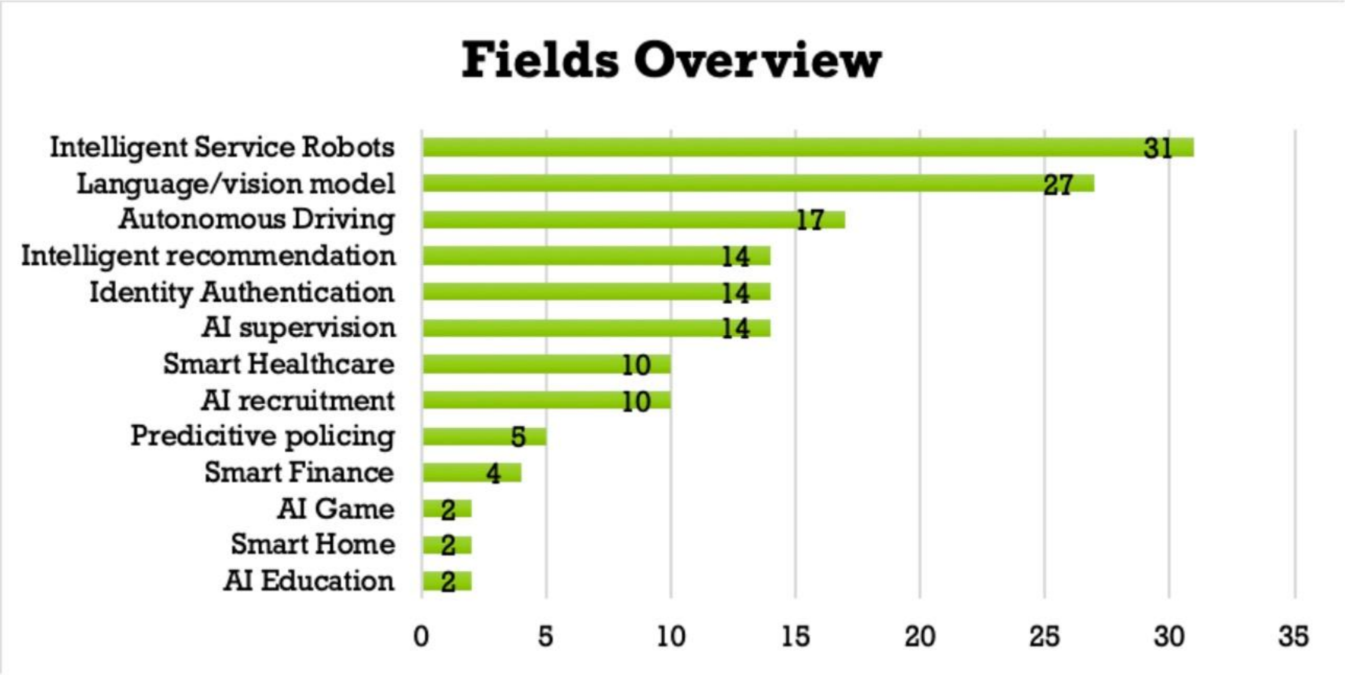
Table 1 Krippendorff' s alpha for each variable



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
Predicitive policing	1
Smart Home	1
AI Game	1
Smart Finance	1
Privacy	1
Inappropriate Use(Bad Performance)	0.90
Unethical Use(illeagal 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

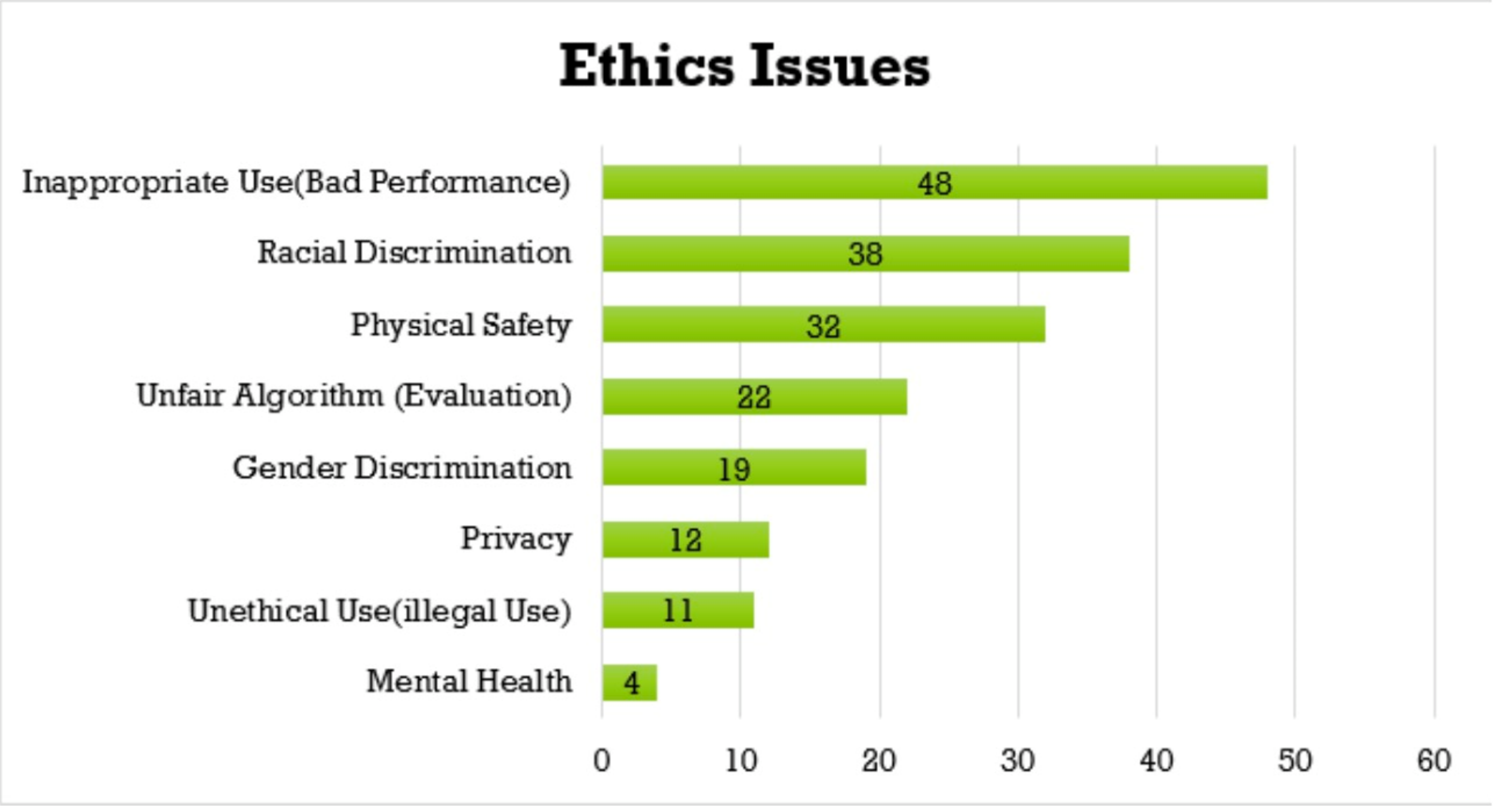
➢ Results

- Application areas of AI ethics incidents



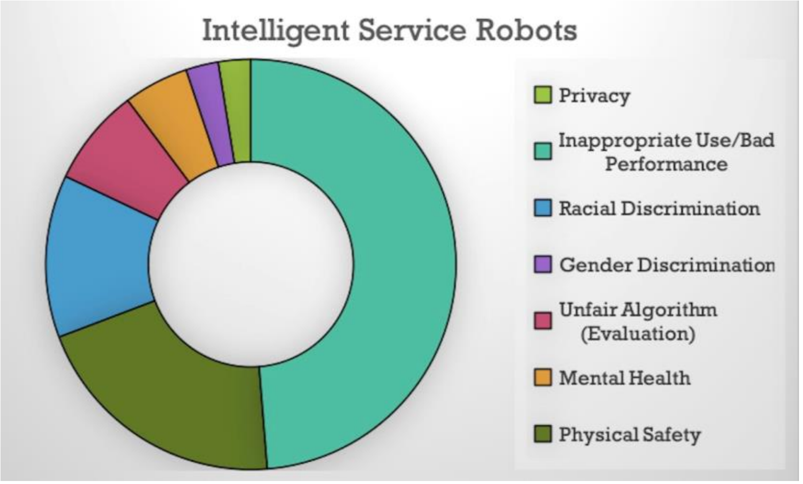
➢ Results

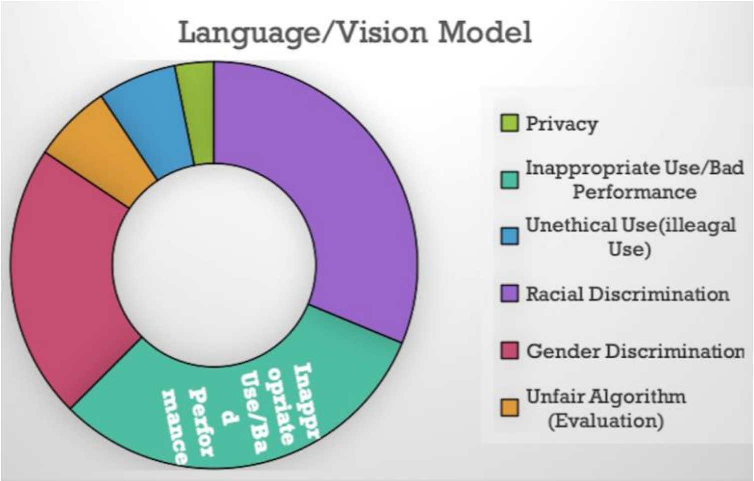
- Taxonomy of AI ethics issues

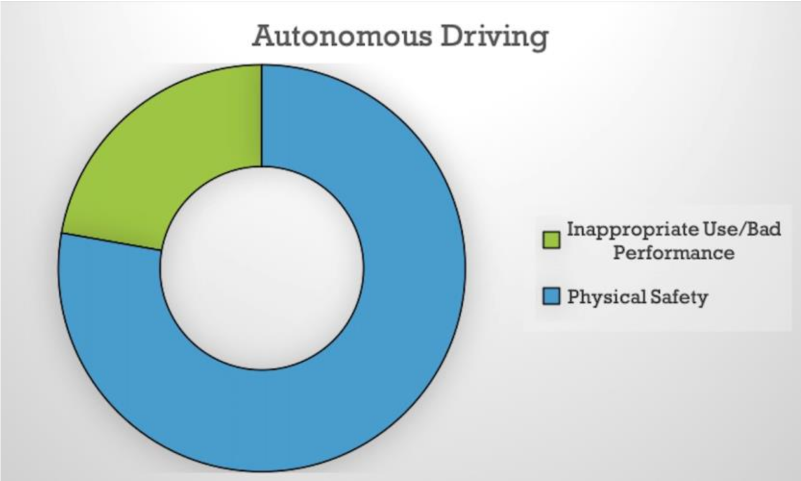


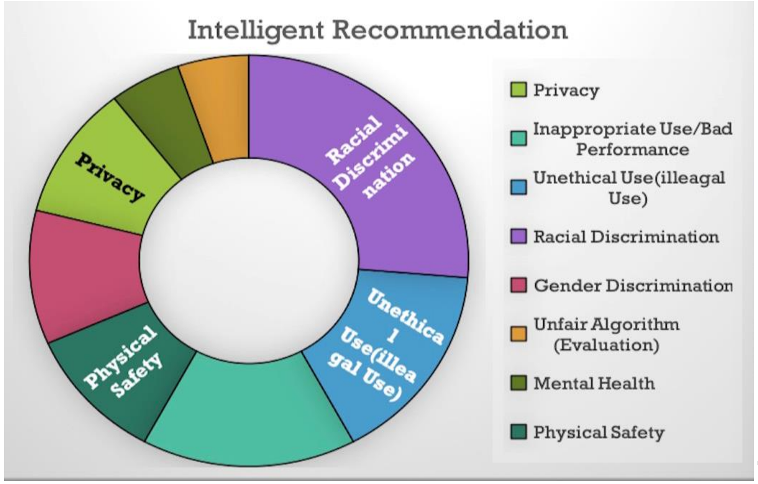
➢ Results

- 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, without applying NLP models to analyze topics and sentiments.

- 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

Assist principle makers in formulating more practical AI guidelines.

Thank you !

Mengyi Wei

Chair of Cartography and Visual Analytics

Technical University of Munich

mengyi.wei@tum.de