Autonomous Bus Ethical Decision-Making for Moral Dilemmas

Background

Transport for London Crime and Anti-Social Behaviour Report:

During April to September 2022, there were 17923 crimes reported on London public transportation, which has 8% increase comparing with pre-pandemic period, 45% of these crimes resulted in injuries.

National Standard for Driving Buses and Coaches by DVSA :

Human bus drivers are required to follow safety management protocols during emergencies. Bus drivers are allowed to decide a place that is not part of the fixed route that would be suitable to stop, to ensure safety of persons.

Autonomous Buses:

Carry a higher responsibility for the safety of passengers and other road users, requiring more sophisticated route planning and decision-making algorithms to navigate through traffic and respond to unexpected situations.

Author:
Zijie Huang

E-mail:
Z.huang@bristol.ac.uk

Challenges

Incommensurable Moral Values:

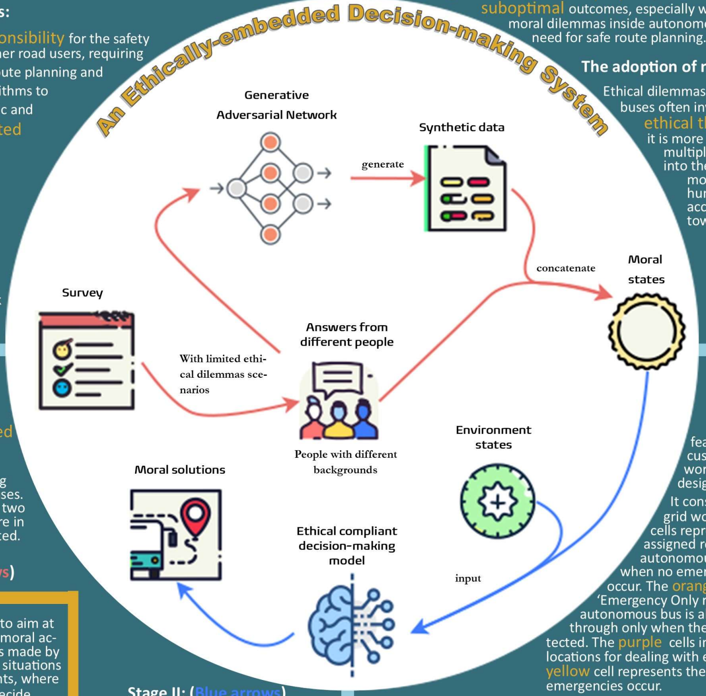
There are various metrics available to assess certain objects, such as the Cambridge Crime Harm Index (CCHI) for measuring the severity of crime harm to victims and Injury Severity Score (ISS) for assessing trauma severity. However, in ethics, there can be incommensurability in values when they do not share the same scale of measurement. For example, values from the CCHI and ISS cannot be easily compared and contrasted. Thus, harms caused by a crime incident inside a bus may not be comparable to the one caused by an injury.

Miserable policies generated from the agent:

The linear scalarization of reward functions in multi-objective decision-making models may oversimplify the problem and result in suboptimal outcomes, especially when faced with moral dilemmas inside autonomous buses and the need for safe route planning.

The adoption of moral principles:

Ethical dilemmas inside autonomous buses often involve multiple ethical theories. Therefore, it is more reasonable to adopt multiple moral principles into the decision-making model to ensure that human values are accurately represented towards a safe route planning.


An Ethically-embedded Decision-making System
Generative Adversarial Network
generate
Synthetic data
concatenate
Moral states
Survey
With limited ethical dilemmas scenarios
Answers from different people
People with different backgrounds
Environment states
input
Ethical compliant decision-making model
Moral solutions

Methods

We proposed an ethically-embedded decision-making system for handling moral dilemmas arising inside autonomous buses. The system comprises two main stage as the figure in the centre demonstrated.

Stage I: (Red arrows)

A survey is designed to aim at collecting data on the moral acceptability of decisions made by autonomous buses, in situations of unavoidable incidents, where participants must decide which incident takes precedence over the other. The survey presents the participant with a hypothetical moral dilemma, whereby there is simultaneously a criminal and medical incident, of which both have a similar severity.

To gather more moral dilemma scenarios and the corresponding participants' answers for training a general ethical decision-making model, the Generative Adversarial Networks (GANs) is utilized to generate synthetic data from the original collected data. Lastly, the synthetic data and original collected data are concatenated together and serve as moral states to the decision-making model.

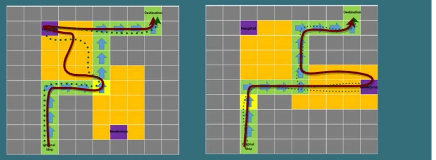
Stage II: (Blue arrows)

The human-value aligned data from Stage I and other environment states from different dilemma scenarios are considered as inputs to the ethical decision-making model. An Ethical compliant multi-objectives Thresholded Lexicographic Deep Q-learning (e-TLDQ) is proposed to ensure finding ethical-optimal policies, which lead to the autonomous bus generating moral solutions for route planning under various moral dilemma scenarios.

Results

An interactive playground that features a customizable grid world environment is designed.

It consists of a 10 X 10 grid world. The green cells represent the 'Pre-assigned route' that the autonomous bus should follow when no emergency situations occur. The orange cells are the 'Emergency Only route' that the autonomous bus is allowed to drive through only when the emergency is detected. The purple cells indicate the allocated locations for dealing with emergencies. The yellow cell represents the location where emergencies occur.



The above figures showcase two examples of route planning solutions in different environments (scenario A and B). The burgundy-colored line represents the majority of solutions provided by the participants, while the dark green dotted line represents the primary solution generated by the proposed method. Both the participants and the agent's solution opted for the shortest route to address the emergency while reaching the destination. It is noteworthy that the route matching rate between user study and our proposed method is approximately 76.8%.