SIT799 Human Aligned Artificial Intelligence

Distinction Task 9.1: AI Safety

Overview

In Week 9, our focus shifted to a crucial aspect of artificial intelligence: safety. Within the AI domain, safeguarding against potential risks takes on utmost importance, especially when the technology is deployed in critical applications, such as autonomous driving vehicles. A prime illustration of this is the recognition of traffic signs, where the ramifications of erroneous predictions can be far-reaching, possibly leading to tragic consequences. Given the profound implications involved, this assignment delves deeply into the pivotal subject of AI safety, with a specific emphasis on the context of image classification for traffic signals.

Through an examination of safety within the realm of traffic sign classification, our objective in this assignment is to unveil and mitigate potential risks inherent in such systems. This endeavor is driven by the pursuit of a more profound comprehension of how AI algorithms can inadvertently introduce risks and uncertainties.

To complete this assignment, you need to refer back to Week 9 lecture material.

Requirements

It is required to use Python 3.11. Please install the packages provided in the **requriments.txt** file as follows:

```
> virtualenv python3.11_task_9.1D
> source python3.11_task_9.1D /bin/activate
> pip install -r requirements.txt
```

Submission Details

Convert the Jupyter Notebook to a **PDF** and submit that document. You may have to install pandoc to convert a Jupyter Notebook to a PDF document:

http://pandoc.org/installing.html

Instructions

Download the Jupyter Notebook from the task resources and complete all the tasks.