Ethics Case Study

Judge Fatigue vs Machine Learning System

Acknowledgement

This sample case study is derived from actual case that is documented in the 2019 CSIRO Discussion Paper “Artificial Intelligence: Australia’s Ethics Framework”.

Context

An analysis of judicial parole hearings in a particular jurisdiction revealed there was a distinct pattern correlating severity of sentences with certain times of the day (i.e. more lenient early in the morning or straight after lunch, more severe late in the morning or late in the afternoon).

Dilemma

A machine learning (ML) system was trained to analyse all the factors that a judge would consider along with the judgement in 1,000 cases; it was configured to compensate for those cases determined late in the morning or late in the afternoon, which had exhibited greater severity of sentence. The resulting ML system produced much less biased outcomes (sentences), eliminating the “judge fatigue” pattern.

Unfortunately, because it was a deep learning system, its decision-making logic was opaque, and therefore it was not possible for the system to provide evidence as to why certain decisions were proposed.

How would you help the officers in this jurisdiction to decide whether to implement this system as a replacement for human judges?

Options

The officers of jurisdiction have following options for adopting this system:

1. No, do not implement until its explicability is improved.
2. No, but it can be used to advise judges in a testing environment to ascertain that no biases exist within the model and train the system for better results.
3. Yes, but not until it has been trained on many more cases.
4. Yes, but a judge should be able to overrule it.
5. Yes, but the appellant can have the right to appeal to a human judge.
6. Yes, its accuracy is clearly much better than human judges.
7. Something else, which is…

Considerations

To uphold ACS values, officers need to be mindful of the following from the Code of Professional Ethics:

2.2 h Develop systems which are robust, secure, and user-friendly: System algorithm needs to be explained before it can be made ready for implementation. Outliers or mistakes in the input dataset can significantly affect the deep learning process.

2.3.1 c Be impartial and fair and do not discriminate unfairly against people in interpersonal interactions or in the design and function of systems: Deep learning systems can learn and improve over time based on user behaviour, so the system design needs to consider this factor and train them on large quantity of high-quality datasets before implementing.

2.3.1 e: Support and contribute to a healthy workplace, that is respectful and supportive of others: AI and Machine learning are not intended to replace human judges but to augment their capabilities.

2.3.2 c Ensure that the public interest is defended: We do not have enough information about the cases to ensure the judgement is correct (unless they were all the same e.g. all for speeding fines etc. Otherwise, it is comparing apples with oranges if the cases are different).

Ethical Decision

Using this system without human judges can lead to incorrect decisions as the system algorithm is not explicable. so here our best option is #1.

Option #2 can also be considered valid but only if system is used in a controlled environment under human judges' supervision to notice patterns, understand the system's reasoning and ensure accountability.

Further attention

The officers in jurisdiction should not consider replacing human judges completely with an automated system as human reasoning, creativity and emotional intelligence cannot easily be replicated by automated systems. AI based systems can aid and revolutionise the decision-making process by assisting to inform, support and advise people involved in the justice system. We recommend that the jurisdiction goes through a workplace health and safety assessment to address the fatigue issues and incorporate appropriate measures to address the risk factors.