

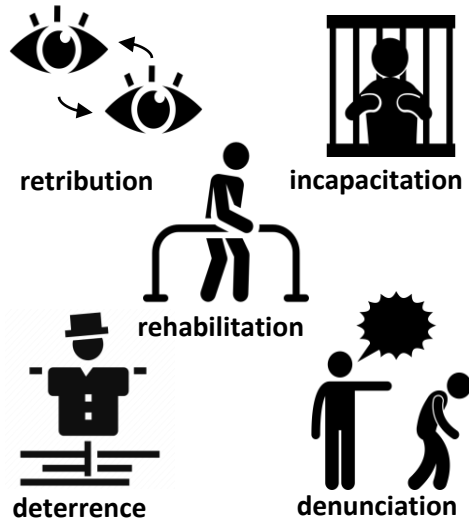
Why does the machine punish?

How does the use of machine learning in criminal sentencing affect the application of the theories of punishment?

1 Background

The theories of punishment

What is our rationale for punishing criminals?



Risk-assessment

- Primary application of machine learning in criminal justice
- Prediction of an offender's chance of recidivism



2 Method

This research is a qualitative literature study. Literature has been retrieved through key-word searches and the snowball method

3 Identified issues

Implicit values to explicit requirements

Not all stakeholders will be familiar with this topic, nor have explicitly thought about them before

The machine as manipulator

The presence of a mechanical tool might unwantedly distort a judge's decision-making process

Incapacitation in the end

Being the only theory truly compatible, the use of risk-assessment ultimately promotes incapacitation

Data matters

To account for rehabilitation, the use of dynamic data is required, even though static data is the more practical choice

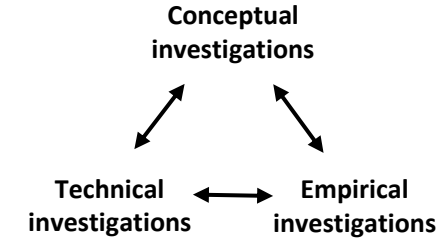
No use for prediction

retribution, deterrence and denunciation are fundamentally incompatible with risk-assessment

4 Proposed solutions

Value-Sensitive Design

- Tripartite framework can accommodate the proposed ways to account for stakeholders' values.
- Iterative structure suggests bridges with established methods like Agile.



Value-aware stakeholder engagement

- Use indirect, in-depth and value-aware questions
- Use value scenarios to observe stakeholders' preferences

Systemic Design

- Consider the placement of the risk-assessment tool within the system
- Engage with lawmakers to adapt the legal system in order to accommodate the risk-assessment tool properly

5 Conclusion

If left unaddressed during development, the use of machine learning based risk-assessment tools will promote incapacitation above all other theories.

Developers should be aware of this effect and can employ various methods to mitigate it.

Matthijs de Lange
m.w.delange@student.tudelft.nl
Supervisor: Evgeni Aizenberg PhD.
Professor: Prof. dr. ir. Inald Lagendijk

