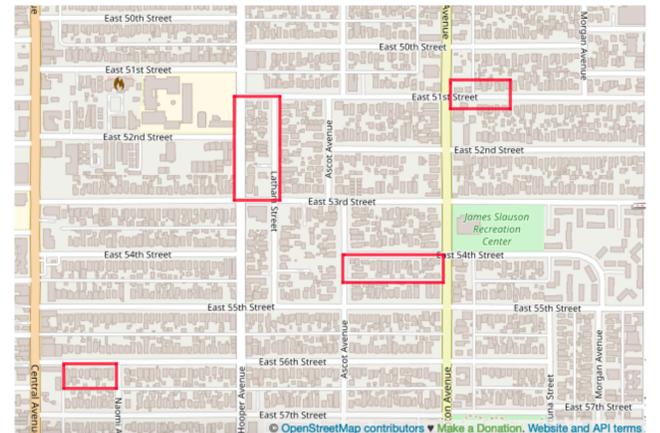


# The Danger Behind Biased Data in Predictive Policing

“Stopping crime before it starts.” [1] What might sound like a quote from 2002’s movie “Minority Report” is already a reality. The Los Angeles Police Department is one of the very first institutions worldwide that use PredPol, a commercial system for predictive policing. It aims to predict potential criminal activity by using machine-learning algorithms.

What can have a real benefit for society, however, also has a downside. Civil rights advocates are raising privacy issues as they are being monitored. Additionally, the proprietary system is not transparent. Which algorithms calculate how high the risk potential of individuals is? Which factors influence the decision? Could the decision be based on biases that lead to discrimination?



PredPol is trying to predict areas where crimes are likely to be committed. (Own example image)

## Goal

This poster proposes to conduct a research project that reverse-engineers PredPol to find out how their machine-learning algorithms work. Furthermore, it aims to find out if biases influence its predictions what would make it unethical to use it.

## Problem

Although predictive policing appears to have a positive effect on reducing criminality [2], evidence suggests that biased data can have a significant impact on who gets controlled by the police. Therefore it is suspected to even increase these biases. Civil rights advocates worry that predictive policing leads to profiling individuals before they even commit a crime. As the used machine-learning algorithms learn to make predictions by analyzing patterns and then search for similar patterns in new data input, biases in the training data might be a problem. This behavior eventually can lead to the discrimination of minorities and communities [3].

## Solution

It is crucial to make predictive policing transparent and assure that the data used is free from biases. Therefore, an interdisciplinary team is needed, as technical, economic, sociological, and legal aspects must be covered. The following solution is proposed:



### Reverse-Engineering PredPol

- Understanding the way PredPol is trying to predict who will commit a crime and where a crime will happen by analyzing its machine-learning model.
- Finding out which characteristics influence the predictions in data sets.



### Analyzing Training Data

- Analyzing the data that e.g., the Los Angeles Police Department uses to train PredPol’s algorithms.



### Identifying Biased Data

- Conducting a study with domain experts that check the data set for biases
- Comparing the prediction of the model to the actual crime history of individuals that were likely to commit crimes according to PredPol’s forecast.



### Data Cleaning

- Cooperating with the authorities to clean data set from biases to avoid discrimination.

**Tobias Wuttke**

Student IT-Systems Engineering (Master)  
Hasso Plattner Institute, Potsdam, Germany

E-Mail: tobias.wuttke@student.hpi.de  
Poster for Lecture Series in Practical Data Engineering  
Professor Dr. Tilmann Rabl, Professor Dr. Felix Naumann

### References

- [1] Joel Rubin (21 August 2010). "Stopping crime before it starts". The Los Angeles Times. Retrieved 19 December 2013.
- [2] Mohler, G., Short, M. B., Malinowski, S., Johnson, M., Tita, G. E., Bertozzi, A. L., and Brantingham, P. J. (2015), "Randomized Controlled Field Trials of Predictive Policing," Journal of the American Statistical Association, 110, 1399–1411.
- [3] P. Jeffrey Brantingham, Matthew Valasik & George O. Mohler (2018) Does Predictive Policing Lead to Biased Arrests? Results From a Randomized Controlled Trial, Statistics and Public Policy, 5:1, 1-6.

### Connection to the Lecture

Professor Dr. Gjergji Kasneci’s talk “Applications of AI in the Credit Information Business” discussed the possible use of AI for individual SCHUFA scoring that is comparable to predicting the likelihood individuals will commit crimes.  
Professor Dr. Birgit Beck’s talk “Ethical Data Engineering” covered the ethical use of data engineering methods and techniques that inspired to discuss the ethical use of predictive policing.  
Professor Dr. Ziawasch Abedjan’s talk “Data Cleaning” covers the structural analysis and cleaning of data that also is applicable in the proposed solution of this poster.