

# Machine Learning

Dr Changjiang He, Dr Kuo-Ming Chao  
Computer Science | School of Art  
University of Roehampton

# Lesson 11.2

## How to Avoid Data Bias

- The prevention of data bias in machine learning projects is an ongoing process.
- Though it is sometimes difficult to know when your machine learning algorithm, data or model is biased, there are a number of steps you can take to help prevent bias or catch it early.

- To the best of your ability, research your users in advance. Be aware of your general use-cases and potential outliers.



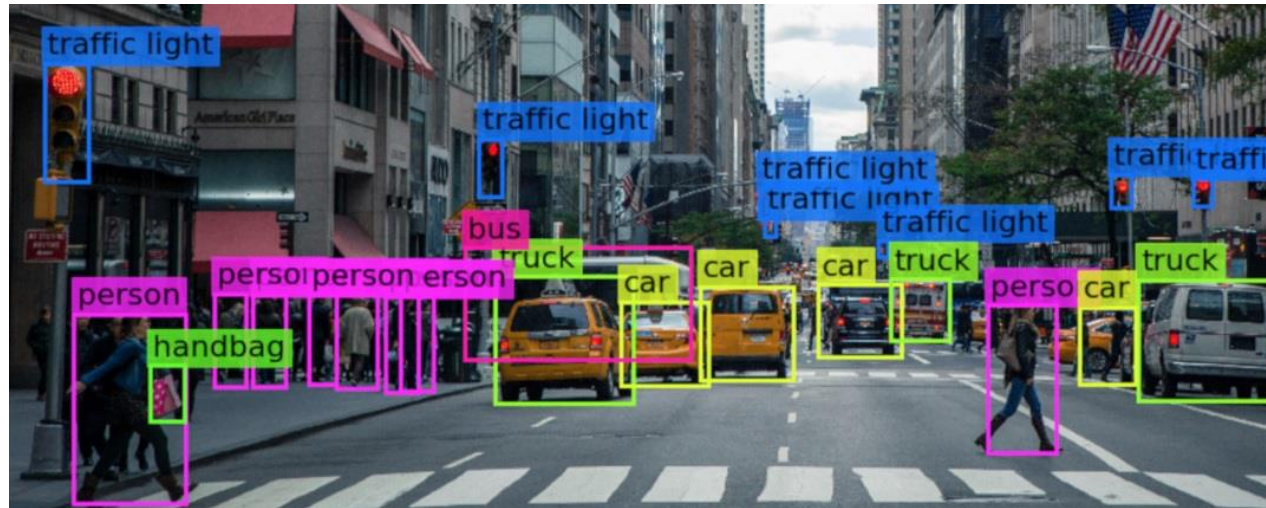
- Enlist the help of someone with domain expertise to review your collected and/or annotated data. Someone from outside of your team may see biases that your team has overlooked.

- Ensure your team of data scientists and data labellers is diverse.



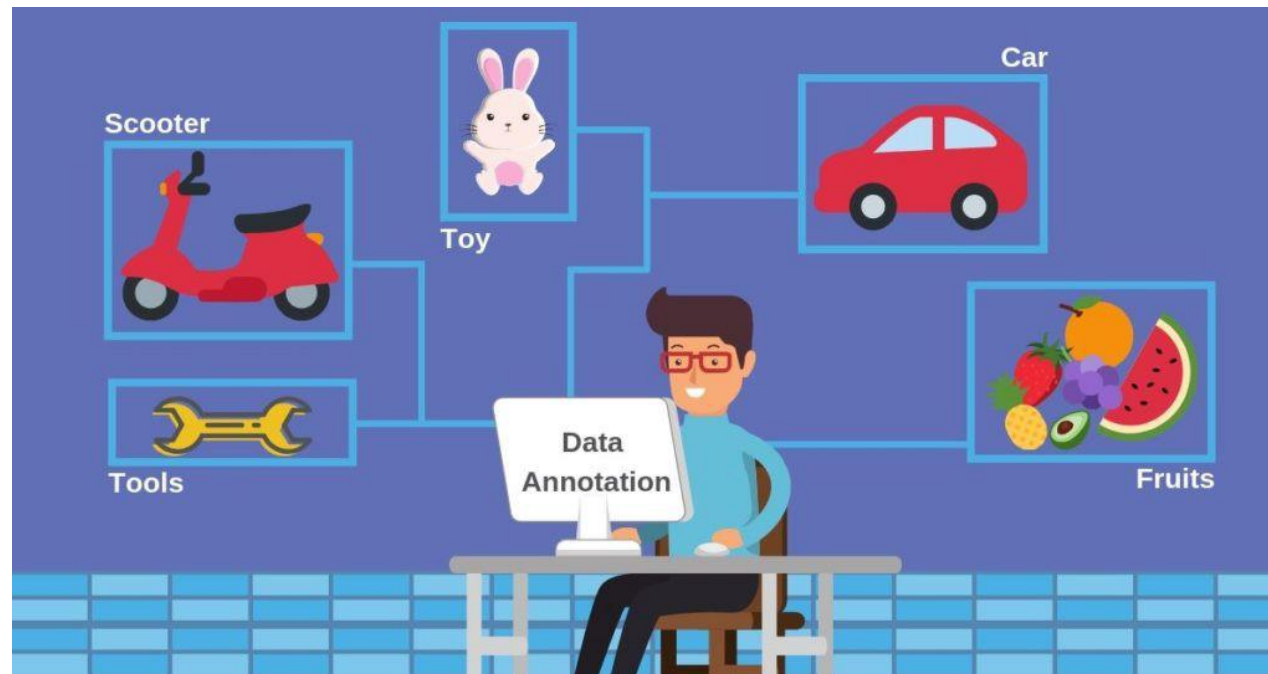
- Where possible, combine inputs from multiple sources to ensure data diversity.

- Create a gold standard for your data labelling. A gold standard is a set of data that reflects the ideal labelled data for your task. It enables you to measure your team's annotations for accuracy.



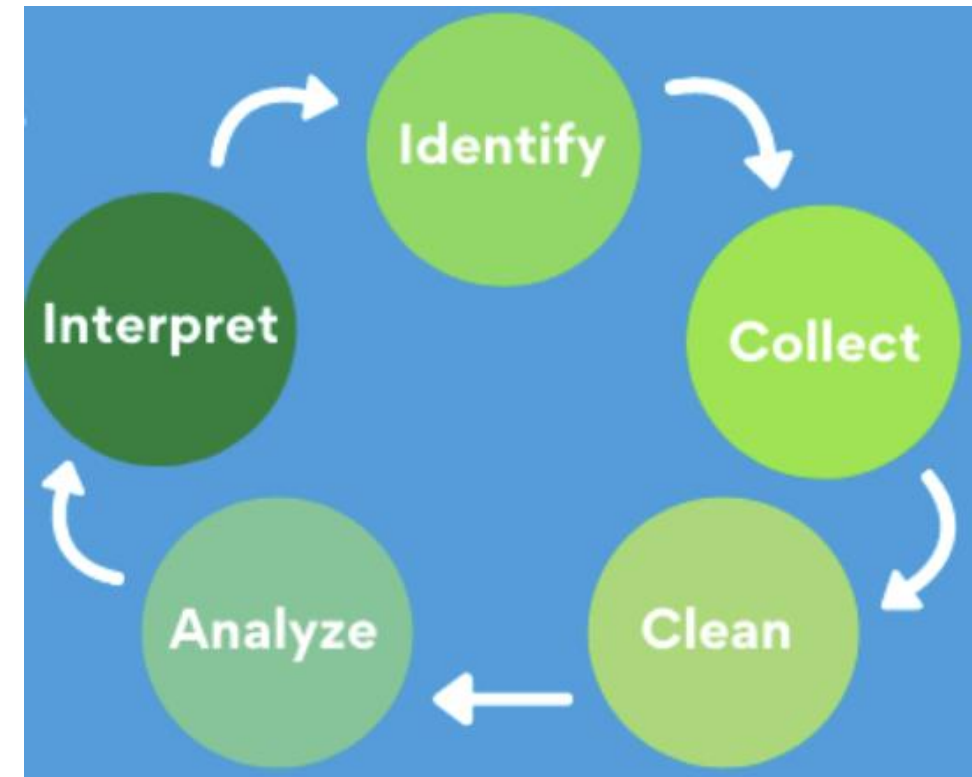
- Make clear guidelines for data labelling expectations so data labellers are consistent.

- Use multi-pass annotation for any project where data accuracy may be prone to bias. Examples of this include sentiment analysis, content moderation, and intent recognition.





- Analyse your data regularly. Keep track of errors and problem areas so you can respond to and resolve them quickly.
- Carefully analyse data points before making the decision to delete or keep them.
- Make bias testing a part of your development cycle.





# MSc project intensive

All your project preparation sessions will take place over 2 days.  
Put them in your calendar now!

- **Monday 5 June**
- **Tuesday 6 June**

These are on-campus sessions. By the end of the two days you will have completed your project initiation form (topic expression of interest) and project proposal forms (aims and objectives), and have been allocated a supervisory team.

Your final submission date will be **6 September**.