

# Balance in Data

---

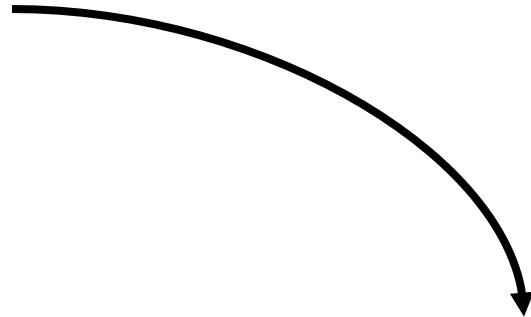


# Who are we?

---



Andrea Walker



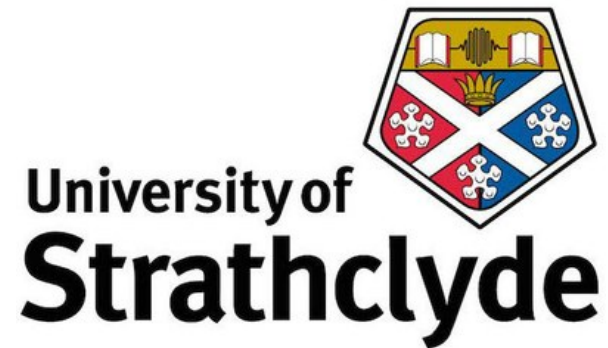
University  
of Glasgow



# Who are we?

---

## Suzie Beith



# Gender Balance At University

---



Data Analytics



Data Science



Web  
Development

# CYBG Data Analytics and Data Science

---

**Machine Learning Club** - a new exciting club set up by Finlay Macrae. Open to everyone with an interest in data, analytics, machine learning and artificial intelligence...

The club is currently filled with enthusiastic colleagues who are eager to learn about the field, it would be great to see some of you there, especially more women at our **Lunch & Learn** sessions!

The **CYBG Big Data** team have been using models

- ❖ to detect fraudulent transactions,
- ❖ inform customers what their discretionary spending is and
- ❖ to identify vulnerable customers who may require help to deal with their finances.

# Why Is Gender Equality Important In Data Analytics and Data Science?

---

What features are useful here?

**What does this result mean  
for the data owners?**

*What mode is the most appropriate for the data and the question?*

How does it impact the  
environment the data was  
taken from?

What metric is the most important here?

## **To answer these questions Diversity is Key!**

# Why Representative Data Matters

---

Algorithms are highly dependent on the data that is put into them.

A model which has been trained on images of brown dogs and white cats, may classify a white dog as a cat, because it is unaware that white dogs exist.

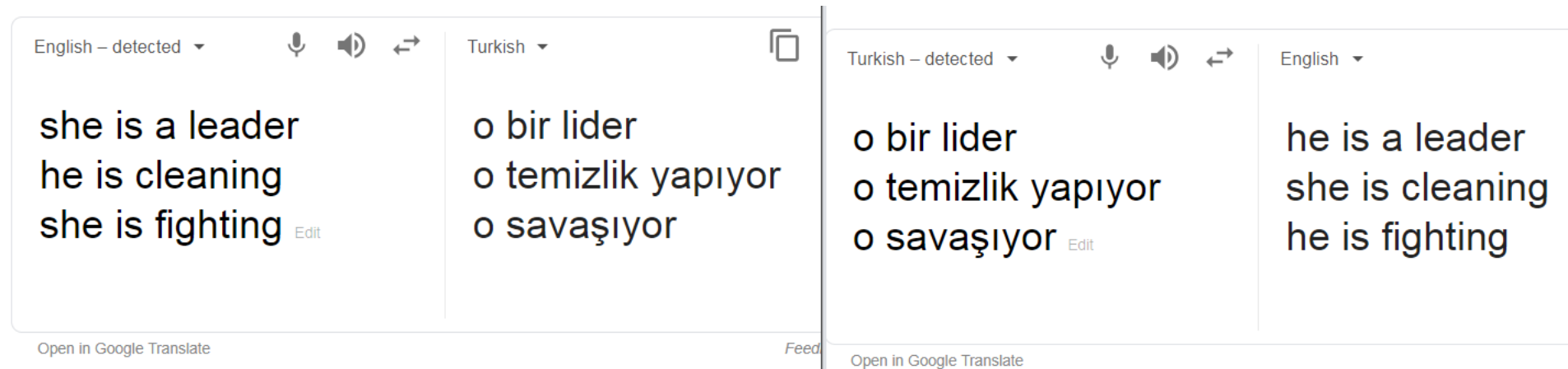


If we accept the results of algorithms, without the proper testing and explanations, how can't we be sure that the race, sex or social-economic background of the data subjects aren't being used inappropriately.

# Why Representative Data Matters



In 2017, media reported that Google Translate was translating from Turkish, a gender-neutral language, to English with signs of sexism.\*



\* Is artificial intelligence sexist?, Jodie Wallis, The Globe And Mail



# Why Representative Data Matters

---



**Last year, an algorithm that was being tested as a recruitment tool by Amazon, was found to be sexist and had to be scrapped, according to a Reuters report.\***

The artificial intelligence system was trained on data submitted by applicants over a 10-year period, much of which came from men, it claimed.

\* <https://www.bbc.co.uk/news/technology-45809919>

# Why Representative Data Matters

---

Recognising gender using images of faces is a commonly researched area of machine learning. You may have witnessed it in action on your iPhone or on Facebook.\*



\* <http://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212>

# Key issues for low representation

---

❖ There is a **gap in the data collected** for women and girls, especially in developing countries\*.

In 2014, Data2X identified key gaps in gender data based on need, population coverage, & policy relevance.\*\*

How can we track progress, recognise sources of inequality and create policies and change to help, if we don't have all the information?

\*<https://www.gatesnotes.com/2019-Annual-Letter>

\*\*<https://www.data2x.org/what-is-gender-data/>

# How can improving gender balance in the field of data help?

---

More recognition of the lack of data about women and girls.

More women being inspired to use data to fight inequality across all fields.

More young girls inspired to take up data science, to use data to help improve issues they have experience of.



**More women in data**

**More supporters for the campaigns fighting for closing the gender data gap.**

Data being used to improve women-centric experiences which may not be considered by those with different experiences.

# What Can We Do?

---

- ❖ Publicise the issue and work to inform women that this is an exciting field they can get involved in.
- ❖ Join the ML Club at work to learn more about data and the ways it is being used in the bank, and in the wider world.

# What are your ideas?

---

- ❖ Is there anything you think analytics and data science could do for your department at CYBG?
- ❖ What data from your day to work do you think could provide interesting insight?
- ❖ Is there an issue in your local area, that insight from data could help? If so, are you aware of what data is available?

# Balance in Data

---

Thank you for listening 😊