Responsible AI Workshop

25-05-2023

Welcome!

Who are we?

Rick van Kersbergen

Master student Applied Artificial Intelligence

graduation internship
Responsible AI Lab

Oscar Oosterling

Master student Applied Artificial Intelligence

graduation internship
Responsible AI Lab

Sophie Horsman

Researcher
Responsible AI Lab

What you’ll be doing today

Schedule

10:00 - 10:30 Introduction / AMANDA presentation

10:30 - 11:00 Data-quality & Privacy

11:00 - 11:10 Break

11:10 - 11:40 Bias & Fairness

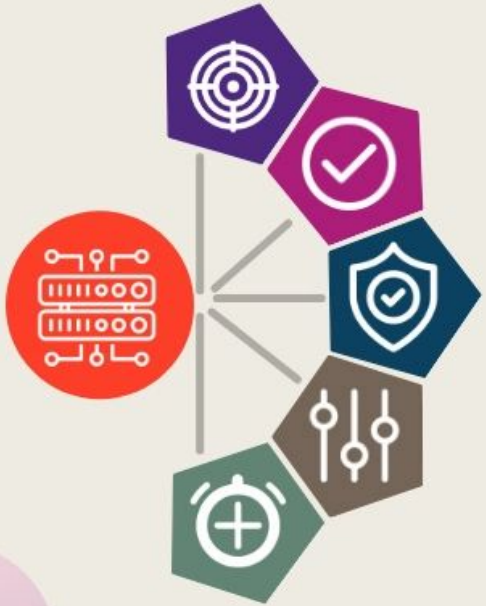
11:40 - 12:00 Ending workshop

AMANDA

How it works in a nutshell

Data-quality and Privacy

Five Dimensions of Data Quality



Accuracy
All details must be correct, error-free, and dependable.

Completeness
The data should be comprehensive, and there should be no missing or obsolete records.

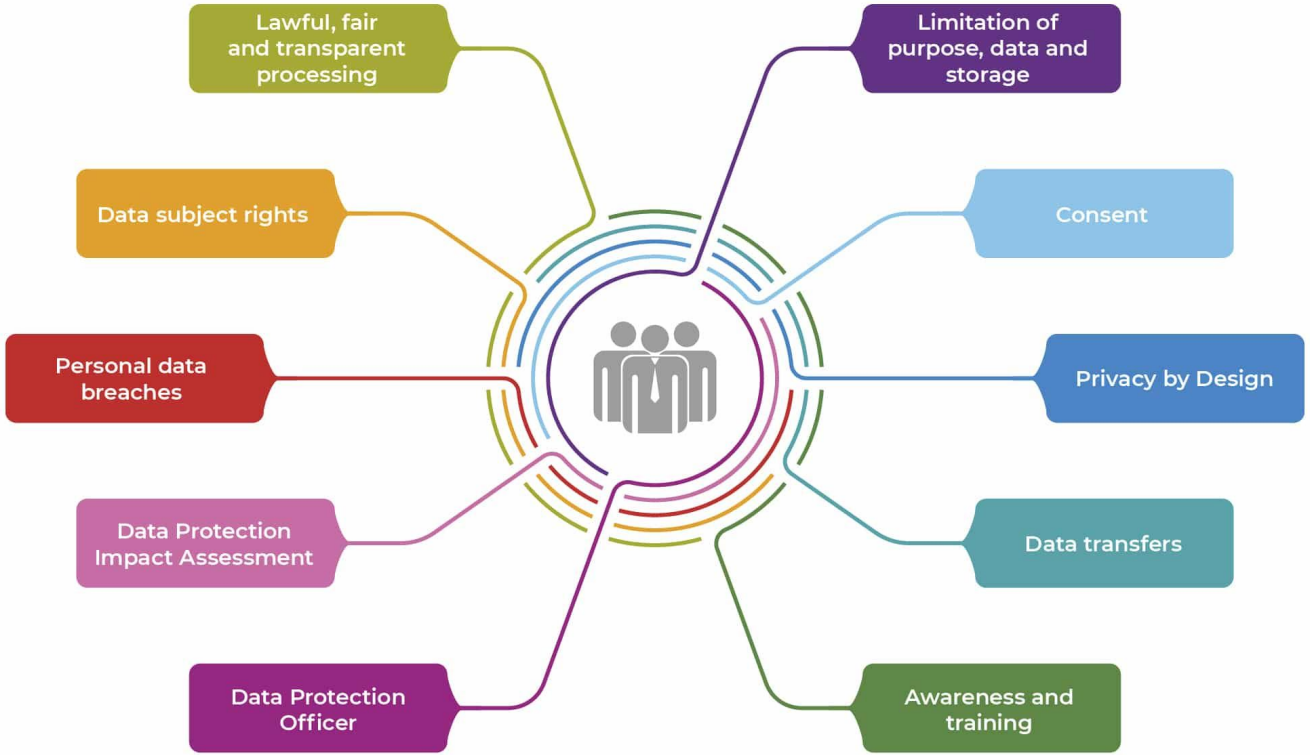
Reliability
Data should be extracted from reliable sources, & hould be consistent across all systems, without any contradictions.

Relevancy
Data must be relevant for the purposes for which it was collected.

Timeliness
Data must be up-to-date and readily available for use, when needed.

What Is Data Quality And Why Is It So Significant? - MOS, 2/11/2021 https://www.managedoutsource.com/blog/what-is-data-quality-and-why-it-so-significant/

Ten key GDPR requirements



With this information…. we had the following questions:

- How was the data for AMANDA gathered?

- And how is decided who belongs in which customer segment?
- How is privacy by design achieved?

- How will customers be made aware?

Now it’s up to you!

Discuss ~ 20 min and write down questions you come up with

Break

10 minutes

Bias and Fairness

Before we can look at **Fairness**…

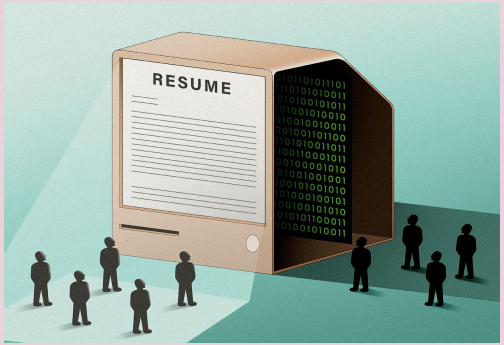
… we have to understand **Bias**

“The action of supporting or opposing a particular person or thing in an unfair way, because of allowing personal opinions to influence your judgment”

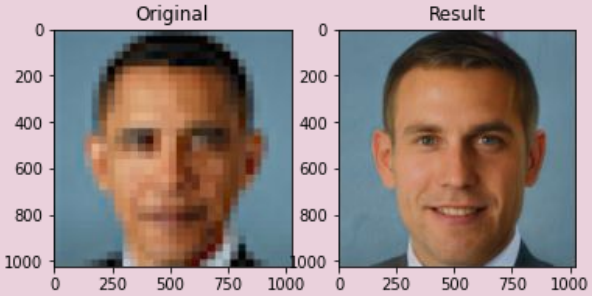
~ Cambridge English Dictionary

…Otherwise, defined as *prejudice*

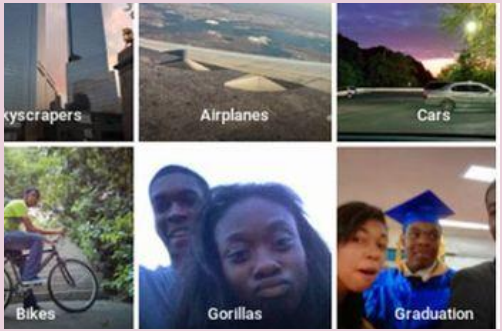
Bias can lead to harms!



Allocation harms



Quality-of-service harms



Representation harms

These **harms** reveal **bias** against or towards…

Sensitive features

Features that contain **personal**, **racial** or **social-economic** indicators that may be used for discrimination

- Race
- Colour
- Sex including gender, pregnancy, sexual orientation, and gender identity
- Religion or creed
- National origin or ancestry
- Citizenship
- Age
- Pregnancy
- Familial status
- Physical or mental disability status
- Veteran status
- Genetic information

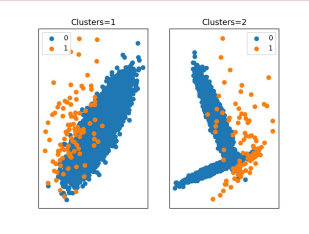
Agarwal, S. and Mishra, S. (2022) Responsible AI: Implementing ethical and unbiased algorithms. Cham: Springer International Publishing AG.

These **harms** caused by **bias** for or against **sensitive features** lead to…

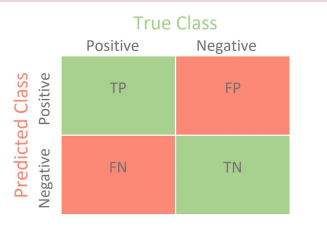
Fairness issues

Mistreatment of people caused by **harms**, based on **biased** opinions

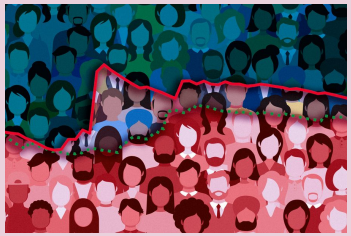
These harms can be caused by…



Unequal Distribution of data



Using the wrong metric



Using sensitive and proxy variables

And data-quality problems!

Agarwal, S. and Mishra, S. (2022) Responsible AI: Implementing ethical and unbiased algorithms.Cham: Springer International Publishing AG.

J. Brownlee, 'Why Is Imbalanced Classification Difficult?', MachineLearningMastery.com, Feb. 16, 2020. https://machinelearningmastery.com/imbalanced-classification-is-hard/ (accessed May 23, 2023).

'Confusion Matrix for Your Multi-Class Machine Learning Model | by Joydwip Mohajon | Towards Data Science'. https://towardsdatascience.com/confusion-matrix-for-your-multi-class-machine-learning-model-ff9aa3bf7826 (accessed May 23, 2023).

'A technique to improve both fairness and accuracy in artificial intelligence', MIT News | Massachusetts Institute of Technology, Jul. 20, 2022. https://news.mit.edu/2022/fairness-accuracy-ai-models-0720 (accessed May 23, 2023).

Questions we had…

How is decided which features are necessary for the prediction of the model?

Is the model even suitable as a solution for the problem you're trying to tackle?

Could the outcomes of AMANDA lead to uneven benefits for certain groups of people?

Now it’s up to you!

Discuss ~ 20 min and write down questions you come up with

Why should KPN care?

Responsible AI is not just the right thing to do…

…it’s also a quality assurance and image boost

(And in two years you’ll have to anyway)

Action points

Communicate! Be transparent and explainable!

Know what you intend to use and do

Know how you intend to use and do

Know why you intend to use and do

Keep asking yourself question like you've learned today

Noticing any issues in your project? Discuss with your team/supervisor

And above all, ask yourself…

Is it ethical to solve this problem using AI to begin with?

Further Questions?