Lab 5: Discover your bias

Odette Scharenborg and Ruben Wiersma

Machine Learning 2019/2020

- **WHAT** This nonmandatory lab consists of several insight exercises and questions.

- **WHY** The exercises are meant to help you understand implicit bias and why it is important.

- **HOW** Follow the exercises in this notebook either on your own and discuss them with a friend. Use Mattermost to discuss questions with your peers.

Note: this lab is supposed to take a maximum of 2 hours. The other 2 hours are reserved for the voluntary mid-term exam.

1 Part 1

You're most likely not aware of your own implicit biases, since implicit biases tend to go unnoticed. To find out about your implicit biases, you are going to carry out the Harvard implicit bias test, a well known and widely used test for implicit bias.

$\rightarrow$ **Exercise 1.1** Perform the Harvard implicit bias test.[1]

$\rightarrow$ **Exercise 1.2** Discuss your experiences with a lab partner:

- Which biases did you expect?

- Did those biases show up in the test?

- What was a surprise to you?

- Were there any differences in implicit biases between you and your lab partner?

  - Can you explain where they come from? Think about cultural, language, educational, social, gender-based differences between you and your lab partner.

---

[1]If you cannot open the test by clicking on the name, copy and paste the url in your browser: `https://implicit.harvard.edu/implicit/education.html`

2 Part 2

In the lectures, we discussed several examples where implicit biases caused a machine learning algorithm to exhibit unwanted behaviour. One of these cases was Tay, the Twitter chatbot developed by Microsoft.

→ **Exercise 2.1** Read the story about Tay.[2]

→ **Exercise 2.2** Discuss the Tay case with a lab partner:

- What caused the issues with Tay?

- How could they be solved?

- What other problems, besides biased training data, could have influenced Tay's results?

---

[2]`https://chatbotslife.com/the-accountability-of-ai-case-study-microsofts-tay-experiment-ad577015181f`