Lab 4: What dangers does AI pose?

Deadline: Monday 3rd March, 5pm.

In this lab, you will explore some of the current and future threats posed by AI: use this document to jot down some thoughts and quotes as preparation for the discussion we will have in the next supervision.

Introduction

So far, you have learned the basics of Python programming, and are now familiar with some of the most important libraries for linear algebra and data science (namely **numpy**, **matplotlib** and **pandas**). In Lab 3 and 3.5, you were shown how to leverage all this knowledge to build some simple AI models: a **classification** model capable of differentiating between cat and dog images, and a **regression** model capable of predicting future house prices based on the past relationship between house prices and square footage. Whilst these applications only begin to scratch the surface of what we can do with AI and ML, with the knowledge, skills and intuition that you've gained in the process, any future AI project you might want to undertake is within your reach.

But with power, comes **responsibility**. As you probably know, AI is not without its problems. Since AI is fueled by data, if that data reflects human **bias** or contains undesirable correlations, the systems can make decisions that disadvantage certain groups. With the increasing ubiquity of AI systems comes the risk that historically disadvantaged groups will continue to be systematically discriminated against. Beyond that, as we saw in Lab 0, AI is increasingly present in a wide range of industries. While AI efforts to discover new drugs for rare diseases sounds like a good thing, the same technology could be used to develop new chemical weapons. Even if you have the best intentions, it is hard to predict whether the same AI technology could be used maliciously or not.

Research and take notes (~20min)

Below you will find a list of readings to read and videos to watch as preparation for the next supervision which will be seminar-like (discussion-based). After each piece, use the space below to add your initial impressions and maybe some quotes that you can quickly refer to when we discuss these in the next supervision.

- **Article**: How AI Is Using Facial Detection To Spot Rare Diseases In Children [link] **(~7min)**
  Impressions:

- **Video**: "Coded Bias" Trailer [link] – note that the full documentary is available to watch on Netflix. **(~2min)**
  Impressions:

- **Video**: Facial recognition technology will change the way we live [link] **(~6min)**
  Impressions:

[Optional] Further Readings:

- **Article**: Crawford, K. and Paglen, T. (2019). Excavating AI: The Politics of Training Sets for Machine Learning [link]
- **Article**: Hall, L. and Clapton, W. (2021). Programming the machine: gender, race, sexuality, AI, and the construction of credibility and deceit at the border. [link]

Reflect (~20min)

Reflect on what you saw and read (200-400 words). Has your opinion shifted from your initial impressions?