Algorithmic Portrait

Project Proposal - CART 451
Professor: Sabine Rosenberg
By Audrey Coulombe

Non-technical description of the themes and major elements of the project

As stated by Mimi Onuoha in [The Point of Collection](#), "It's easy to forget that the people collecting a data set, and how they choose to do it, directly determines the data set". Indeed, data is always biased in some ways. The proof of this is that there are big chances that, by changing the way you approach data, you will also change the output. Furthermore, when biased data is used to feed algorithms, it results in biased analysis. Mimi Onuoha also writes about the common practice consisting of discarding certain data because it doesn't fit patterns. This practice adds an additional layer of misinterpretation to these analyses by discrediting marginalized data and amplifying stereotypes and inequities in the process. In [this essay](#), Hito Steyerl uses the term dirty data to refer to this discarded data. She wrote: "Dirty data is simply real data in the sense that it documents the struggle of real people with a bureaucracy that exploits the uneven distribution and implementation of digital technology". She also talks about apophenia and inceptionism to point out that patterns can be seen anywhere, even in random data or noise. Akin to inceptionism, which is a technique for gaining insight into a neural network's "mind" by amplifying patterns it recognizes in an image, *Algorithmic Portrait* aims to create a portrait representing the way people are perceived by algorithms and emphasizing the stereotypes involved in their analysis. In a world of ubiquitous computing where algorithms are widely used to classify people into categories that are assumed to define who they are, paying close attention to the biases and stereotypes they amplify becomes all the more important. *Algorithmic Portrait* is based on this premise.

The project will take the form of an app that asks questions to the user through a questionnaire and collects sensible data on them, such as their political ideologies, the color of their skin, their annual salary and their sexual orientation. Throughout the questionnaire, emphasis will be put on irrelevant justification for collecting these data as an allusion to one of Onuoha's criticisms stating that, even if it is mentioned that the data is collected for this or that *reasonable* purpose, it does not mean that it won't be used for other things later on. Each answer will be collected and stored as a key adjective, which will be used to generate a description of the person. This description will then be fed to a text-to-image artificial intelligence to generate an *algorithmic portrait* of the user. I planned on using Open AI's DALL·E 2 as the image generator. Since social media are known to be a hotspot for the collection of personal data and categorisation of people with algorithms, the app will be mostly aimed at social media users and the generated image will be made so it can easily be posted on such platforms. If possible and if tests show that it improves the overall user experience, the generated portrait will then be adapted with Google's ARCore to be used as an AR lens so that users can use their cellphone to take a selfie with the generated lens. If not, I might also explore how to create the portrait from a picture of the person using the same text-to-image AI.

With this project, I hope to surprise people with unexpected and perhaps obviously biased images. More than that, I hope that these images will stick in people's minds so that they remember that there is no such thing as neutral data or algorithms and that it must be used critically. Tests with text-to-image AI will be needed to see if the generated images convey the intended meaning or if the notion of bias needs to be emphasized. If the biases need to be made clearer, I might as well manually associate stereotypes with each possible answer so that the description fed to the text-to-image AI is based on these stereotypes.

Three other projects that provide motivation and insight

[Spurious Correlations](#) by Tyler Vigen

This project directly relates to the notion of apophenia, which consists of seeing connections within random data. Indeed, *Spurious Correlations* uses algorithms to compare datasets and look for similar patterns. The correlations found are then visually communicated with charts highlighting how patterns can be seen anywhere and how misleading it can be. In fact, this experiment gave way to a lot of surprising and random ressemblance. For example, a correlation was found between the number of people who drowned by falling into a pool and the number of films Nicolas Cage appeared in (see image below). Other than to amuse us and make us think about statistics differently, Vigen's erroneous correlations have little impact on our daily lives. However, it is much more disturbing to think about the algorithms that use similar methods to analyze our behavior and preferences, and thus impact our overall experience of connected interfaces by deciding what we see, who sees our content, the type of ads that target us, etc. Indeed, such algorithms are also based on pattern recognition and, as highlighted in the work of Tyler Vigen, these can lead to misinterpretation. I found this work particularly inspiring for my project because it communicates a concerning problem behind data analysis. It is also relevant to mention that I first explored this project a few years ago, and the fact that I still refer to it today shows how much it has impacted me. Indeed, I've seen many other projects over the past few months that I don't even remember, but I've never forgotten this one. With my project, I would like to attain a similar balance between entertaining and informing people about the biases behind algorithmic analysis so that they remember that there is no such thing as neutral data and that their analysis must be taken with a grain of salt.

Source: https://tylervigen.com/spurious-correlations
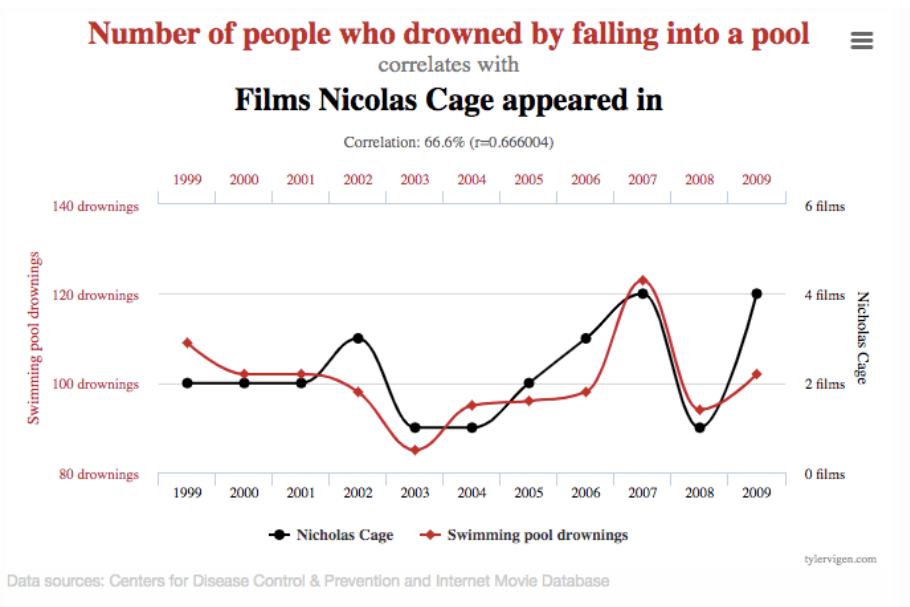
Classification.01 (2017) by Mimi Onuoha

This artwork is about the algorithmic categorisation of people. It critiques the way our society blindly perceives statistics as undeniable facts, without even questioning the processes involved. Classification.01 takes the form of neon tubes in the shape of curly brackets, placed on a wall. When two people stand in front of them, cameras combined with various algorithmic processes are used to analyze their appearance and determine whether or not they are similar. The neon tubes light up when they are classified as "similar." As often, the process behind these algorithmic classifications is opaque: viewers have no idea why they were categorized in this way or what data it was based on. Similarly, I want to reflect the opacity of the analysis made by the AI DALL-E 2. People will see certain details in their portrait and not understand where they come from. Like Classification.01, I also want to raise questions about why a certain categorization was made, as well as get people to think about how algorithmic analyses amplify the cognitive biases of humanity.

Source: https://mimionuoha.com/classification01

Movie *Coded Bias*, directed by Shalini Kantayya and based on Joy Buolamwini's discovery

This film exposes the inability of facial recognition algorithms to detect dark-skinned people and accurately classify women. It highlights how the data that feeds these algorithms is as biased as humans. This is because the most privileged people, such as white men, produce the most data and are therefore overrepresented in that data. When algorithms are fed with this biased data, it results in biased algorithms that discriminate against underrepresented groups, which are often people of color and women. The film also shows how these discriminations can contribute to the amplification of inequalities and stereotypes, such as when police use biased algorithms to identify potentially dangerous people, which too often results in racial profiling. The denunciations made in this film are the basis of my criticism of algorithmic biases. Indeed, my wish is that the portraits resulting from my application shock and show clear traces of these biases.

Source: https://www.codedbias.com/