Assignment Discrimination Eddie Conti

The solution suggested by the statistician to solve the issues of Google's image generation can be summarized as it follows: in order to capture the diversity in world's population, when asked to generate an image, to avoid racial biases the algorithm produces the image in accordance with the racial proportion of that prompt.

Even though the proposed solution seems consistent it lacks of solid foundations. First of all, assuming we have the resources to collect statistical data for each prompt, the system would certainly fail. Following the statistician approach: image to ask the AI generator "an image of an asian pope": as the proportion of asian pope is 0% then the algorithm is unable to produce the image, which is very undesirable. Furthermore, there is another very hidden and deep-rooted problem. Suppose that for a given request the percentage for a given ethnicity is y. The value of y is the result of historical data that may not conform to reality. Topics such as inclusiveness, equity are relatively 'modern' and proportions do not take this into account. In simple terms, if we ask the generator to produce an image of an African-American entrepreneur, the proportion calculated with historical data y is much lower than the **current** proportion as current opportunities for people of colour to occupy prominent positions in a company have increased. The problem of using historical demographic data is a problem itself as it encodes prejudices and discrimination, or, in other terms demographics data are "time and bias sensitive".

It is important to underline the origin of the problem. In first place, there is a structural bias which cannot be eradicated due to social disparity and historical discrimination. In addition, there is surely a deployment bias: Jules White, computer science professor at Vanderbilt University believes that you can you these AI tools to produce basically anything that you want if you craft your prompt carefully. If then the model is retrained according to user queries we may induce a feedback bias. A further ethical problem lies in the use of the image: if the generated image implicitly contains judgements and discrimination is then used in the public sphere, we are contributing to the problem itself.

Before providing an alternative approach, I would like to point out that other problems may arise with the generator. The task concerns racial composition, however there are other disparities that should be taken into account such as age, gender or disability. Moreover, even in treating just the problem of racial composition we should be aware of one important aspect: in recent years, the concept of belonging to an ethnic group is constantly evolving: people identify themselves in a very specific way. However, we refer to the concept of group-fairness for groups where there has been systematic discrimination. If this were not the case, then any solution would be fragile.

Taking into account the above anlysis, in my opinion a more robust solution would be the following: in asking for an image regarding people in any aspect, the AI poses a series of questions to the user in order to minimize the possibility unintended discrimination: the algorithm will not produce an output until the space of feasible outputs is small enough. For instance, by default, we can ask specific question such as the color of the skin, the ethnic group, the social context, clothing or accessories and then apply these features to the image. Alternatively, these features can be chosen randomly. Even though it is not possible to remove completely any biases and control the use of the image, which in my opinion should be the responsibility of the user, this would mitigate unwanted prejudices.