Ethical Considerations in Computer Science

Required Reading - Week 1

Table of contents

Computer science is a field that involves creating and using computer programs and systems to solve problems and improve human life. Computer science has many applications in various domains, such as health care, education, business, and entertainment. However, computer science also poses some ethical challenges that need to be addressed by researchers, developers, and users of these technologies.

1 Ethical Issues in Computer Science

One of the ethical issues in computer science is the quality and fairness of the data used to create and evaluate computer programs and systems. Data is the fuel for computer science, and the quality and fairness of the data can affect the performance and outcomes of the programs and systems. For example, if the data is incomplete, inaccurate, outdated, or biased, the programs and systems may produce erroneous or discriminatory results.

1.1 80 Million Tiny Images Dataset

A recent case that illustrates this issue is the 80 Million Tiny Images dataset by MIT (Torralba, Fergus, and Freeman 2008). This dataset was intended for training machine learning systems to recognize objects in images. It contained 79,302,017 32×32 pixel color images, scaled down from images extracted from the World Wide Web in 2008 using automated web search queries on a set of 75,062 non-abstract nouns derived from WordNet. The words in the search terms were then used as labels for the images.

However, in 2020, a paper by researchers Abeba Birhane and Vinay Prabhu found that some of the labeling of several publicly available image datasets, including 80 Million Tiny Images, was causing models trained on them to exhibit racial and sexual bias (Birhane and Prabhu 2021). For example, they found that some of the labels used for images of people were derogatory, offensive, or harmful. They also found that some of the labels used for images of objects were inappropriate or inaccurate.

The researchers argued that these labels were not only unethical and harmful to human dignity and rights but also detrimental to the scientific validity and reliability of the computer programs and systems created with them. They pointed out that these labels were not representative of the real-world diversity and complexity of objects and people and that they could introduce biases and errors into the programs and systems that could affect their performance and outcomes in various domains.

The researchers also criticized the lack of transparency and accountability in the creation and use of these datasets. They noted that there was no clear information about how the images were collected, how the labels were generated, how the quality and fairness of the data were ensured, and how the data was used by other researchers or developers. They also noted that there was no mechanism for users to report or correct any problems or issues with the data.

As a result of this paper, MIT decided to retire the 80 Million Tiny Images dataset from use by its creators and other researchers. They apologized for any harm caused by the dataset and asked other researchers not to use it for further research and to delete their copies of the dataset. They also stated that they would review their data collection practices and policies to prevent similar issues in the future.

1.2 Conclusion

This case shows that computer science is not a neutral or objective field but rather a social and ethical one. It shows that computer science can have positive or negative impacts on society depending on how it is designed, developed, and used. It also shows that computer science requires careful consideration of ethical principles and values such as respect for human dignity and rights, fairness and justice, transparency and accountability, and social responsibility.

As first-year students who are interested in computer science solutions, you should be aware of these ethical issues in computer science broadly. You should think about how you can address them in your own projects. You should also be curious about how computer science is used in various domains and contexts while being critical about how it affects different groups of people. You should also be willing to learn from others who have different perspectives on computer science solutions.

Computer science is a powerful field that can offer many benefits to society. However, it also comes with many challenges and risks that need to be addressed ethically.

Birhane, Abeba, and Vinay Uday Prabhu. 2021. "Large Image Datasets: A Pyrrhic Win for Computer Vision?" In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1536–46. IEEE.

Torralba, Antonio, Rob Fergus, and William T Freeman. 2008. "80 Million Tiny Images: A Large Data Set for Nonparametric Object and Scene Recognition." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (11): 1958–70.