The Many Faces of Bias

One of the promises of machine learning is that it can potentially offer a superior alternative to human decision-making. Algorithms offer objective interpretations of data and can be trained and developed to produce consistent results, unlike human decision-making which can oftentimes be subjective and inconsistent. Thus, an advantage of machine learning is that it isn't susceptible to 'human error' or mistakes. Even more impressive is the potential for machine learning to predict outcomes with a level of accuracy that surpasses that of the world's leading experts.

While these are exciting possibilities, one of the major challenges of machine learning is the problem of bias. The presence of bias can be the downfall of an otherwise perfect model, resulting in skewed outcomes and low accuracy. This can be a frustrating experience for those who dedicate their time and energy into creating a model, only to discover that it doesn't work very well. Even more concerning, however, is the possibility that if bias finds a foothold in machine learning, the model may effectively function to automate bias and inequality and deploy it at scale. When all goes well, machine learning can be consistently accurate. But the presence of bias can cause machine learning to be consistently inaccurate in such a way that leads to discriminatory or unfair outcomes.

Recent discussions of bias have primarily focused on the use of machine learning in high-stakes contexts where inaccurate outputs can have major consequences. For instance, when algorithms are used to inform decisions regarding bail in the criminal justice system, the granting of loans in financial institutions, or the hiring of job applicants, the presence of bias can result in some individuals or groups being systematically disadvantaged in a way that seriously impacts their life prospects. As one might expect, biased algorithms in high stakes contexts have made national news headlines, as demonstrated by [Amazon's hiring algorithm](#) that revealed gender bias and the [COMPAS recidivism tool](#) that revealed racial bias.

While the negative impact of bias is most apparent in high stakes contexts, bias remains a serious challenge for any application of machine learning—including tinyML. This may be somewhat surprising when one considers the kinds of tasks with respect to inference that tinyML is meant to tackle with the use of microcontrollers. For instance, if bias leads a keyword spotting or activity detection application to be inaccurate, how problematic could this be? Even though tinyML will not be responsible for complex inference tasks in high-stakes contexts, the unique features of tinyML raise a different kind of concern.

TinyML is at the cutting-edge in terms of low latency, low power, and small size. This is significant because it enables tinyML to be pervasive and find applications across a wide range of contexts. The number of IoT devices has already grown considerably in recent years, and with the emergence of tinyML this number is expected to surge upwards. It's precisely because tinyML is uniquely capable of being everywhere, surrounding us, that the problem of bias remains a major concern. If one were to consider the impact of a single device that is inaccurate

due to bias, it may not seem especially problematic. But when one considers the likelihood that tinyML will soon become a feature of a wide variety of our daily interactions with machine learning, it's easy to see how the presence of bias could amount to a more serious concern. For example, if an individual fails to trigger the wake word of a personal voice assistant, this may seem like a minor inconvenience. But if this individual also fails to trigger the wake word for their light switch, thermostat, coffee maker and so forth, then this individual would seem to be systematically disadvantaged insofar as they cannot enjoy the benefits of machine learning applications.

In order for tinyML to make our lives easier and more accessible than ever before, designers will need to remain vigilant in rooting out bias from their applications. So how does bias find its foothold in machine learning algorithms? One reason this can happen is because humans are still responsible for creating machine learning applications, so whatever biases are present in the data scientists may be carried over into the project. The most common reason machine learning becomes biased, however, is owing to the underlying data.

For instance, a dataset may be biased if it is incomplete. If the dataset doesn't fully represent the diversity of end users, then this can lead the model to be less accurate for certain individuals or groups. Researchers have shown that voice recognition tools struggle to identify African American Vernacular English, leading widely popular personal voice assistants to work less well for black individuals. Similarly, research shows that voice recognition struggles to identify non-native English speakers or those with speech impairments. After recognizing the importance of making products accessible to everyone, Google launched "Project Euphonia" in 2019 with the goal of collecting more data from individuals with speech impairments or heavy accents to fill the gaps in current voice datasets.

However, tackling the problem of bias isn't always as simple as collecting more data from underrepresented users. Rather, bias has many faces and can present at any stage of the data lifecycle, including creation, collection, sampling and processing. A general piece of advice is to pay close attention to the data that is being used to develop an algorithm. It is worth reflecting on questions like the following: From whom is the data being collected and is it representative of the end-user population? What device is being used to collect the data? What decisions have been made about how to process and label the data?

Carefully analyzing the data is a huge step towards mitigating bias, as identifying the exact form of bias is necessary for making a determination about how to address it. For instance, if the dataset is incomplete, then the solution may be collecting additional data that better represents the end-user population. Alternatively, if the data is faulty due to labeling errors then the solution may be conducting a review of the data to correct these errors. In the next video, you will learn to identify some common forms of bias. In addition, you will learn about some potential solutions that are being developed on the industry level, as well as what you can do as an individual designer to tackle the problem of bias in machine learning.