

Machine Learning Project: Ethical Assessment

Although collecting and processing weather data may seem like an ethically neutral task, several sources of potential bias could skew our results and even harm the communities ClimateWins is working to serve. Key considerations include:

Personally identifiable information: The current weather data set does not include personal data, nor does it look at geography at a level granular enough to disclose individual locations. However, if used alongside other data, it could reveal sensitive information about the cultural or economic makeup of certain neighborhoods.

Human bias: The data set only includes locations in Europe, yet experts say that underdeveloped countries in the Global South shoulder the greatest burden of [climate disaster](#). Without considering those differences, it would be easy to underestimate the impact on vulnerable communities.

Additionally, because climate change is a politically polarized issue, there may be bias in how the machine learning algorithm is trained. Programmers may assign greater weight to results that align with their preconceptions and dismiss findings that do not. If these biases enter the algorithm, results could be flagged incorrectly.

Data collection bias: In addition to the regional imbalance, there may be bias in the older data due to poorer quality of technology in the 1800s and early 1900s

Latent and selection bias: Without information on how climate change affects other regions, machine learning with this data set would be highly susceptible to these biases. The algorithm could fail to pick up on trends or anomalies because it would be looking for trends like those seen in Europe.