Colby Belczyk

Hugging Face link: https://huggingface.co/spaces/crb031/Deon_Ethics

Ethics Checklist from Hugging Face:

1. Data Collection

- Input Variables: Temperature, Humidity, Wind Speed, Rainfall, Fuel Moisture, Vegetation Type, Slope, Region
- Output Variables: Fire Size, Fire Duration, Suppression Cost, Fire Occurrence
- **Are the data sources, such as weather, vegetation, and geographic data, properly licensed and legally available?** Not sure, the origin of the data is not listed. For the data sources to be ethical, the data obtained must be from licensed sources that are legally accessible.
- **Has any sensitive information, such as private property or personal location data, been anonymized?** There is no explicit mention of personal data, and the data provided is vague enough to the point where personal data cannot be identified.
- **Have you obtained consent for data collected from private or proprietary sources, such as satellite imagery or drone footage?** Consent is not stated but consent must be obtained in order to protect and respect privacy laws.

2. Fairness & Justice

- **How will you ensure that the model's predictions are fair and do not disproportionately affect specific regions or communities (e.g., indigenous lands, rural areas)?** The model must be evaluated to ensure it does not disproportionately predict fires in regions that could negatively impact certain communities, such as indigenous lands or areas with fewer resources for fire management.
- **What biases might exist in the historical data (e.g., underreporting of fires in certain regions), and how will you address these to ensure the model does not unfairly target or neglect specific areas?** Historical data may underrepresent fires in rural or less monitored regions. To address this, the model could be trained with augmented or balanced datasets in order to avoid reinforcing these biases.
- **How will you balance fairness in handling both false positives (predicting a fire where there is none) and false negatives (failing to predict a fire)?** False negatives could lead to severe damage, while false positives may cause costly and disruptive evacuations. There should be some way to find the certainty of the prediction and base actions based on this calculated certainty.
- **Have you tested the model across different regions to ensure consistent performance across various forest types (tropical, temperate, etc.)?** The dataset contains various vegetation types and regions. Therefore, the model should be validated across all these conditions.

3. Transparency

- **How will you ensure transparency about the data sources, algorithms, and decision-making process of the model?** Maintain clear documentation of the data sources, algorithms used, and any preprocessing steps. This ensures that stakeholders can trace the model's decision-making process.
- **What information will you make available to government agencies, the public, and environmental organizations?** Information, including model performance and limitations, should be shared with government agencies, environmental organizations, and the public.
- **How will you communicate the model's predictions and limitations to decision-makers so that they understand the risks involved?** Decision-makers must be informed about the model's potential inaccuracies due to the weight of the decisions that are being made.
- **How will you explain false positives and false negatives to the affected communities or stakeholders, especially during critical events like evacuations?** Establish methods to explain false positives and negatives to affected communities to maintain trust, particularly in situations involving high-stakes outcomes. Additionally, reinforce the accuracy of the predictions if they are normally high.

4. Privacy

- **How will you ensure the privacy of individuals whose data might be inadvertently captured (e.g., campers, rural residents) through satellite images, drones, or weather stations?** In cases where the model uses satellite imagery or drone data, ensure that people's identities and movements are not traceable/blurred out.
- **What steps will you take to prevent the misuse of this data, especially in terms of tracking human activities in forest areas without their consent?** Implement strict protocols to prevent tracking or identifying human activities, emphasizing that the data's use is strictly for environmental monitoring.
- **If external data sources, such as drones or surveillance tools, are integrated into the model, how will you balance the need for accurate predictions with protecting individual privacy?** When integrating external data, the privacy of individuals should not be compromised for accuracy. Techniques like data anonymization and secure storage can help manage this balance.

5. Accountability

- **Who will be held accountable if the model incorrectly predicts a wildfire, resulting in unnecessary evacuations or failure to prevent a disaster?** This could involve collaboration between data scientists, fire management authorities, and government bodies.

- **What system will you establish to monitor and adjust the model over time, ensuring it adapts to changing environmental and climate conditions?** Set up a system to continually update and refine the model based on new data and changing wildfire patterns. This helps keep the model relevant as environmental conditions evolve.
- **How will you communicate accountability measures to the public, especially in high-risk areas where wildfire prediction is critical?** Clearly communicate who is responsible for monitoring and updating the model, and outline the steps taken to ensure its reliability.

6. Inclusivity

- **How will you ensure the model includes diverse data from different types of forests (e.g., tropical, temperate) and regions, especially those that may be underrepresented in historical data collection?** The dataset includes various forest types, but it is important to continuously seek data from underrepresented regions to improve model inclusivity.
- **How will you ensure the model accounts for the needs of different communities, including vulnerable populations such as indigenous groups or rural residents who have unique relationships with the land?** Understand and incorporate the unique needs of different communities, especially those vulnerable to wildfires. Engaging local populations can provide insights into region-specific risk factors.
- **If certain regions or communities lack sufficient data (e.g., underreporting, lack of resources), how will you address this to avoid biased predictions?** Address data scarcity in certain regions by collaborating with local agencies to improve reporting or using synthetic data methods to fill gaps, thereby avoiding biased predictions.

7. Sustainability

- **How will the model's predictions affect long-term forestry practices, land management, and firefighting strategies over time?** The model's predictions should promote sustainable forestry practices, such as controlled burns or improved land management, to reduce future wildfire risks.
- **How will you ensure the model remains sustainable, considering the evolving nature of climate change and its effects on wildfire patterns?** Set up a system to continuously update and refine the model based on new data and changing wildfire patterns. This helps keep the model relevant as environmental conditions evolve.
- **What are the broader social and environmental implications if this model becomes widely adopted (e.g., impacts on land use, deforestation policies, wildlife conservation)?** Consider the environmental and social impact of widely adopting this model. This includes potential shifts in land use policies and conservation efforts to maintain ecological balance.