

Visualizing the Carbon Intensity of Machine Learning Inference for Image Analysis on TensorFlow Hub

TAEWON YOO¹, HYUNMIN LEE², SEUNGYOUNG OH², HYOSUN KWON³, HYUNGGU JUNG²

¹KYUNG HEE UNIVERSITY, ²UNIVERSITY OF SEOUL, ³KOOKMIN UNIVERSITY



Paper



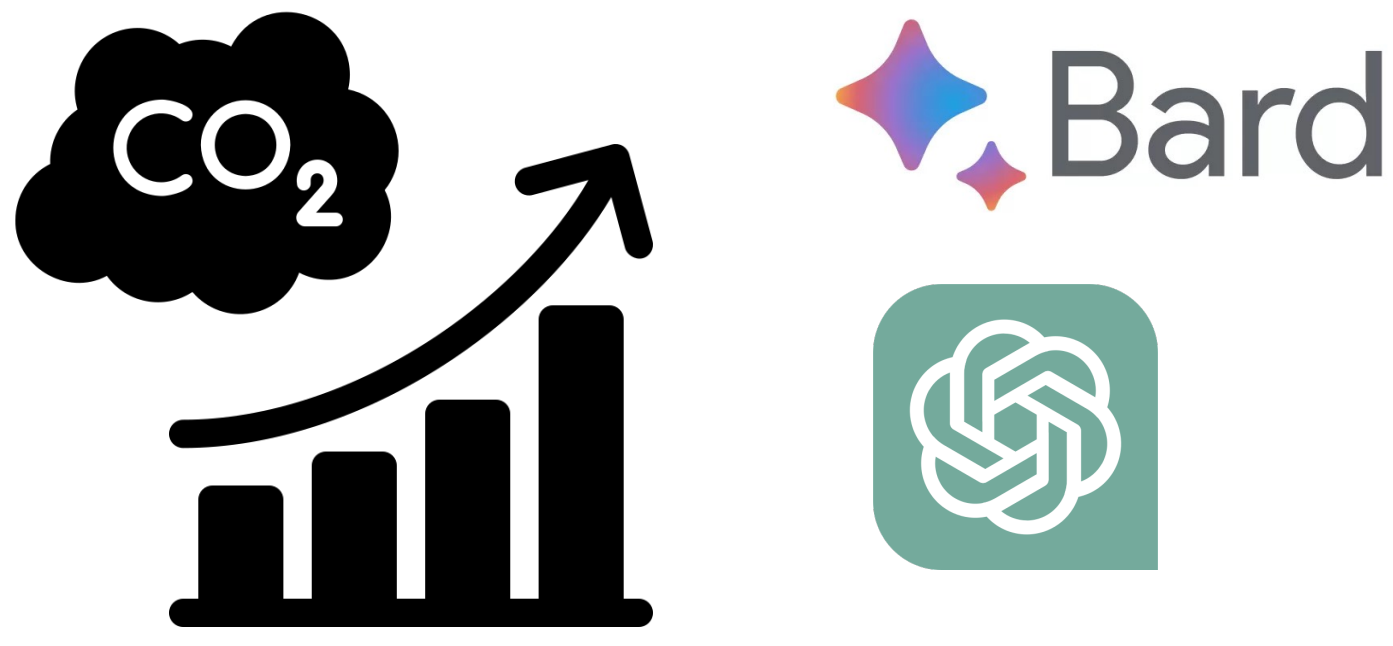
HCAIL Website



MOTIVATIONS

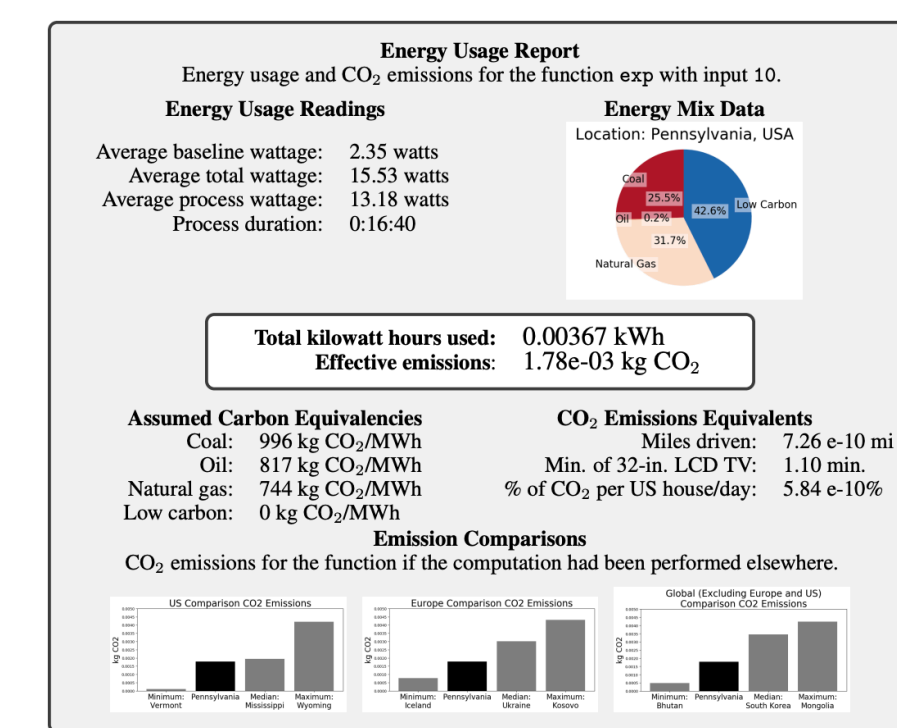
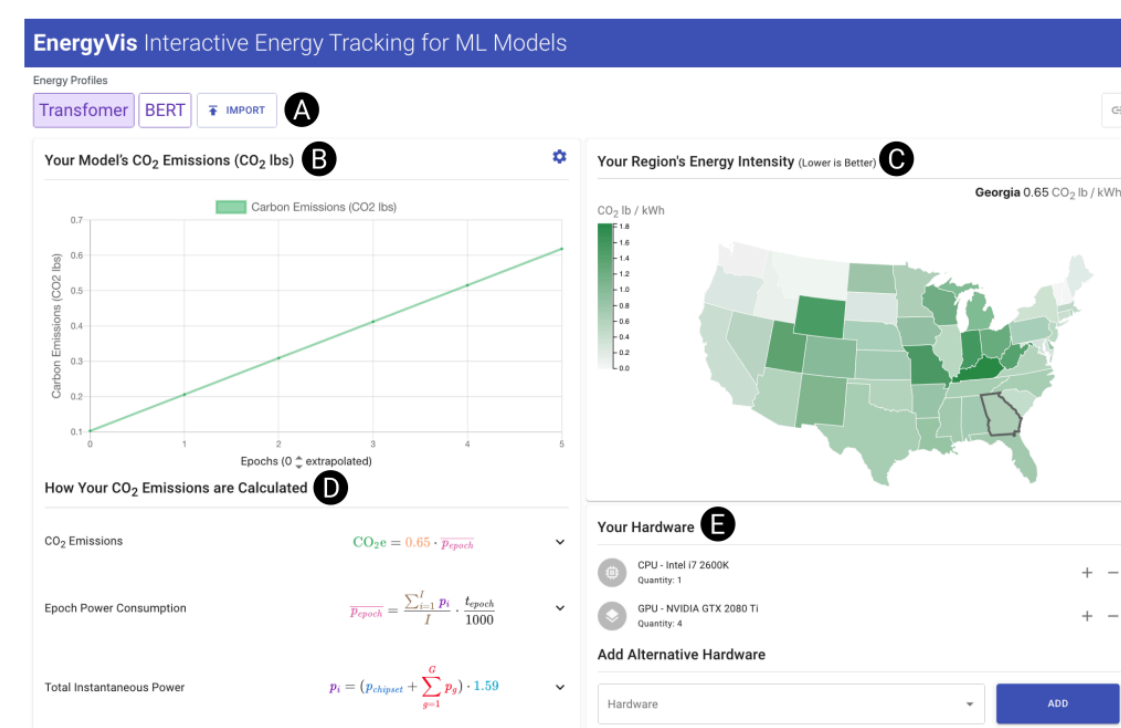
Contact: **Taewon Yoo** (twyoo@khu.ac.kr) **Hyunggu Jung** (hjung@uos.ac.kr)

Growth of Carbon Emissions Generated by Large ML Models



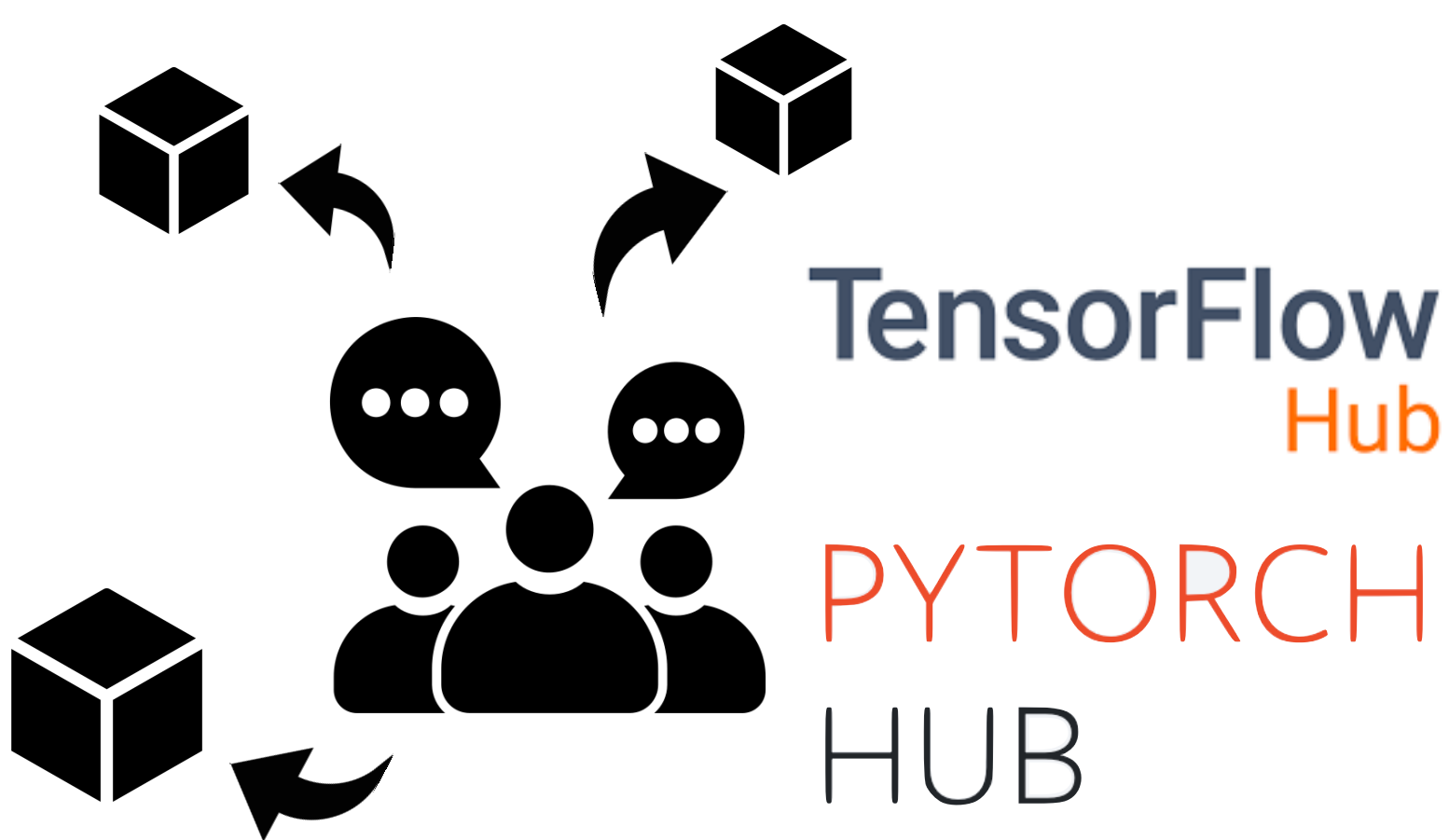
Large ML models such as Bard and ChatGPT are producing significant carbon emissions during the model development phases.

Previous Studies to Support ML Developers to View the Carbon Emissions Generated during the Model Development Phases



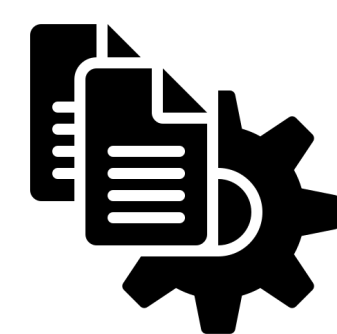
Previous studies created tools for ML developers to view the carbon intensity of ML models during the training process. For example, Shaikh et al. developed EnergyVis, an interactive tool for ML developers to track the energy consumption of ML models and explore alternative deployment locations.

ML Developers Sharing Models on Online Communities

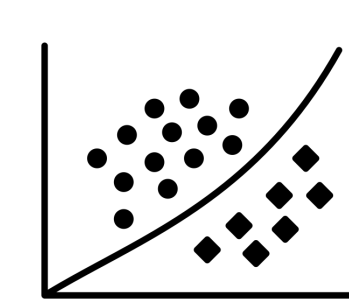
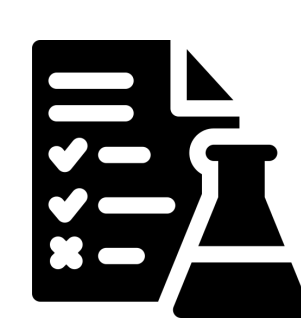


ML developers are sharing pre-trained models on online communities such as TensorFlow Hub.

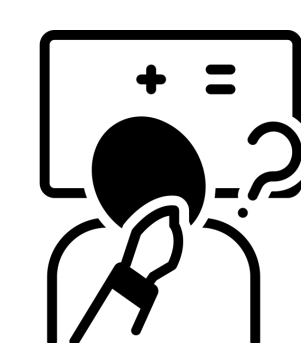
ML Model Development Phases



(1) Data Processing (2) Experimentation



(3) Training



(4) Inference

ML development cycle has four phases: data processing, experimentation, training and inference.

Gap

	Data processing	Experimentation	Training	Inference
EnergyVis	X	X	O	X
Eco2AI	X	X	O	X
CarbonTracker	X	X	O	X
CodeCarbon	X	X	O	X
Experiment-Impact-Tracker	X	X	O	X
Green Algorithms	X	X	O	X
ML CO2 Impact	X	X	O	X

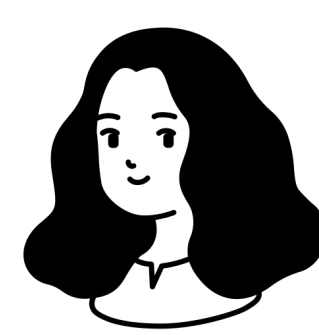
No studies proposed visualization tools enabling ML developers to explore the carbon intensity of models in online communities during the **inference phase**.

How can we develop a visualization tool that supports ML developers in exploring the carbon intensity of ML models in TensorFlow Hub (TF Hub) during the inference phase?

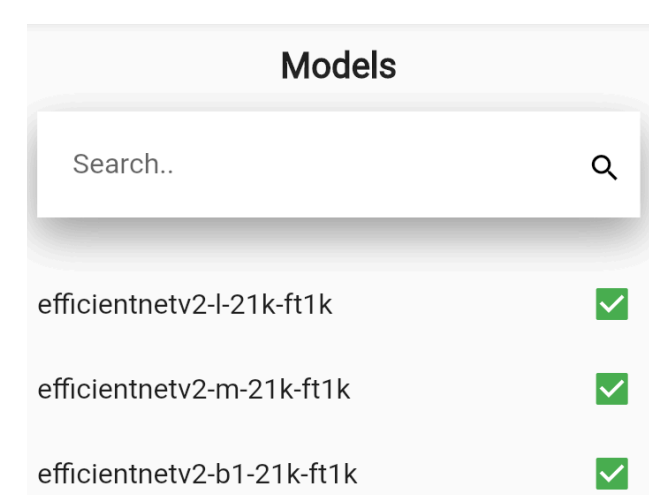
MIEV Tool Design



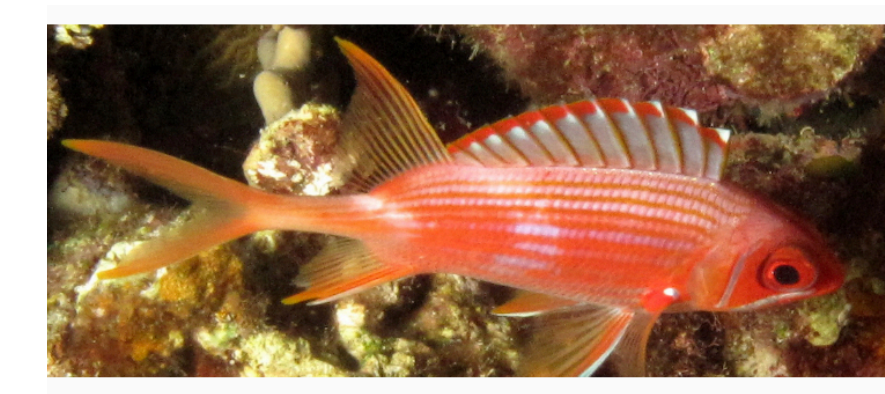
USAGE SCENARIOS



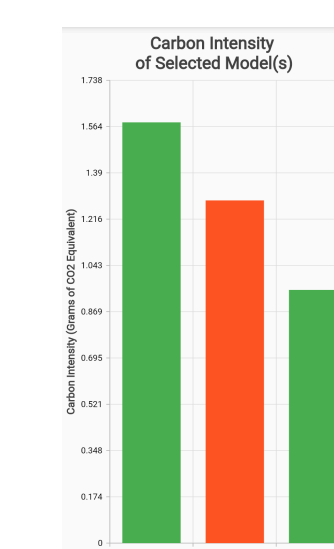
- Olivia, a graduate student needs to deploy an image classification model for her research project.
- She aims to consider both the classification accuracy and carbon intensity generated during the inference phase.



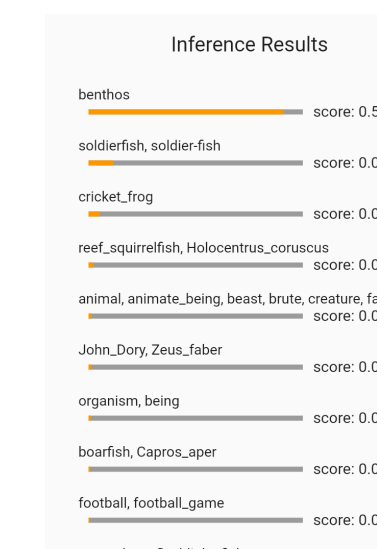
- Olivia selects the three models published on TF Hub as candidates.



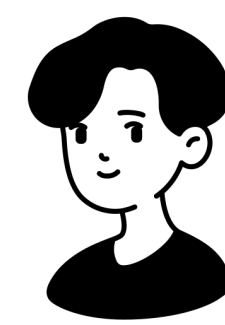
- Olivia uploads an image to test the classification result, and clicks the "Create Chart" button.



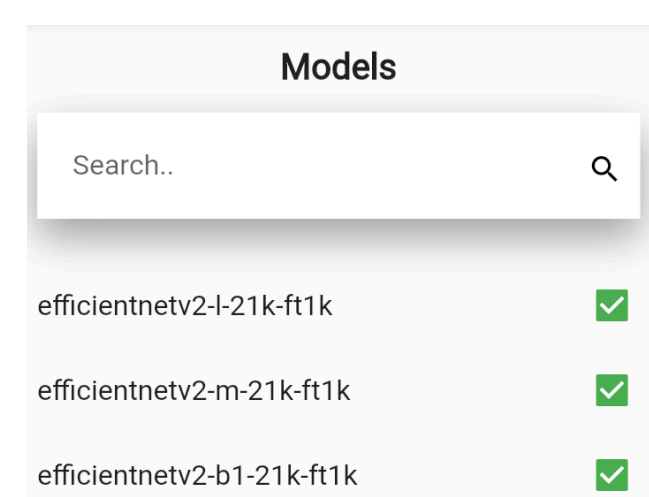
- Olivia compares the carbon intensity generated when the three models classify the image.
- She found that the first model has the highest carbon intensity during the inference phase.



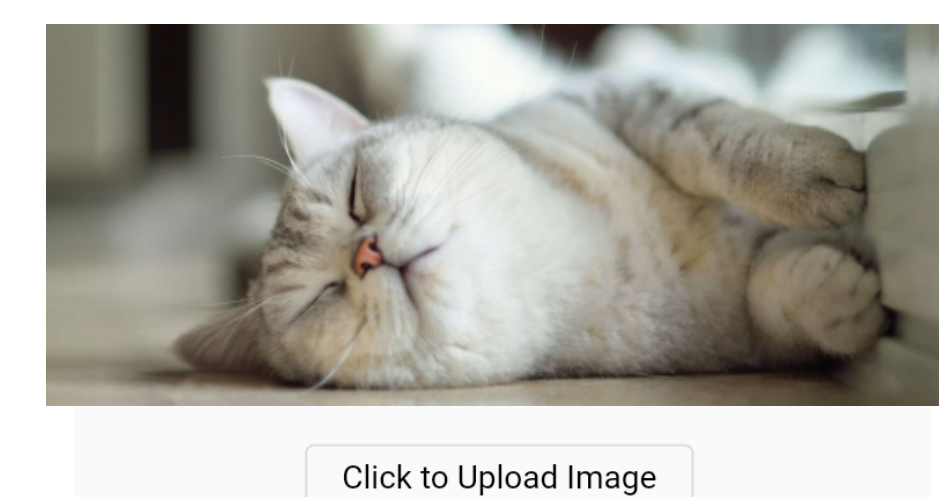
- Olivia also compares the classification performance.
- She found that the second model has higher classification accuracy than other models.



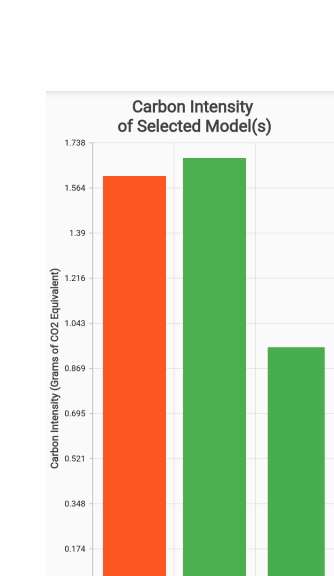
- Bruno is an ML developer sharing image classification models on TF Hub.
- Due to government regulations, Bruno's company is trying to reduce the amount of carbon intensity generated by the ML models.



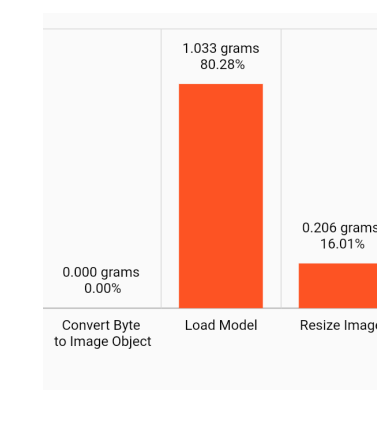
- Bruno selects three models that he uploaded on TF Hub.



- Bruno uploads an example image for classification result, and clicks the "Create Chart" button.



- Bruno found that the first model had the highest carbon intensity during the inference phase.
- Then he clicks the bar to find out which step has a high carbon intensity.



- Bruno found that the carbon intensity was higher than other models when the model received the resized image and return the inference result.

POTENTIAL USE CASES

- Providing ML developers with alternative deployment options regarding carbon intensity can potentially motivate them to consider the environmental impact of ML models.
- Fostering collaborations and providing ML developers with tools for considering environmental impacts can empower developers to collectively work towards more sustainable AI development.

CONTRIBUTIONS

- We developed MIEV that demonstrates exploring the carbon intensity for ML models on TF Hub during the inference phase.
- We presented use cases of ML developers considering the carbon intensity in the model inference phase.
- We provided design insights toward collaborative work among ML developers to lead sustainable AI development.

ACKNOWLEDGEMENTS

- We appreciate HCAIL members for their constructive feedback on our initial manuscript.
- This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2022R1C1C1010883)

FUTURE WORK

- Extend the functionality of MIEV beyond the image domain, such as text, video, and audio domains.
- Conduct user study to improve tool's usability and efficacy, and to assess the tool's impact on users' understanding of the carbon intensity that ML models produce during the inference phase.