

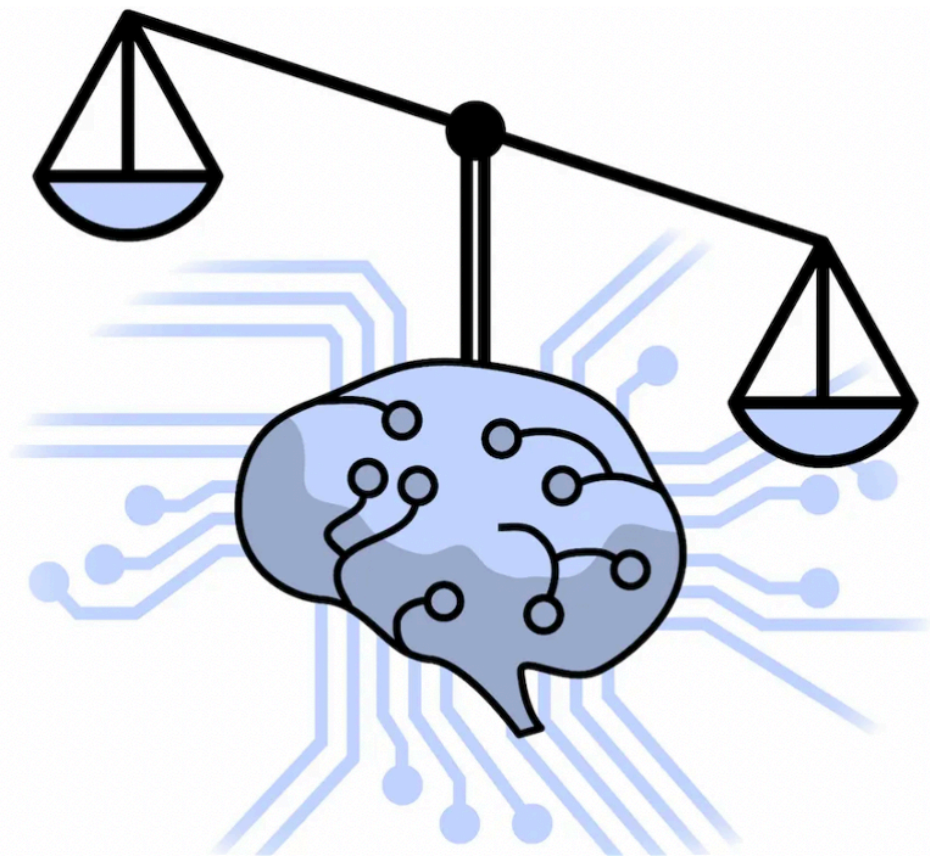
# Applying Algorithmic Accountability Frameworks with Domain-specific Codes of Ethics: A Case Study in Ecosystem Forecasting for Shellfish Toxicity in the Gulf of Maine

ISABELLA GRASSO, ABIGAIL MATTHEWS, DAVID RUSSELL, JEANNA MATTHEWS, NICK RECORD





Machine learning is a rapidly growing field with increasing societal impacts



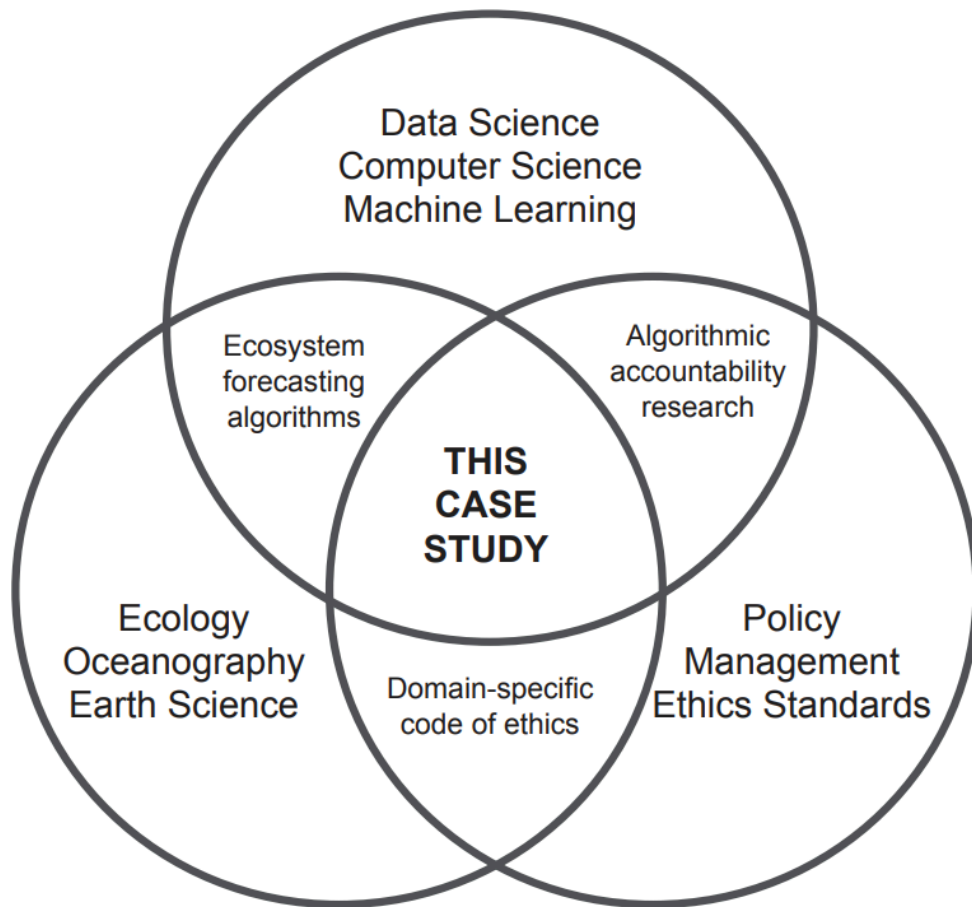
This has led to a mature research field: Fairness, Accountability, and Transparency in Algorithmic Systems

Algorithmic accountability focuses less on the selection of a single ethical standard, but rather on methods for holding a system to an ethical standard determined by domain experts.

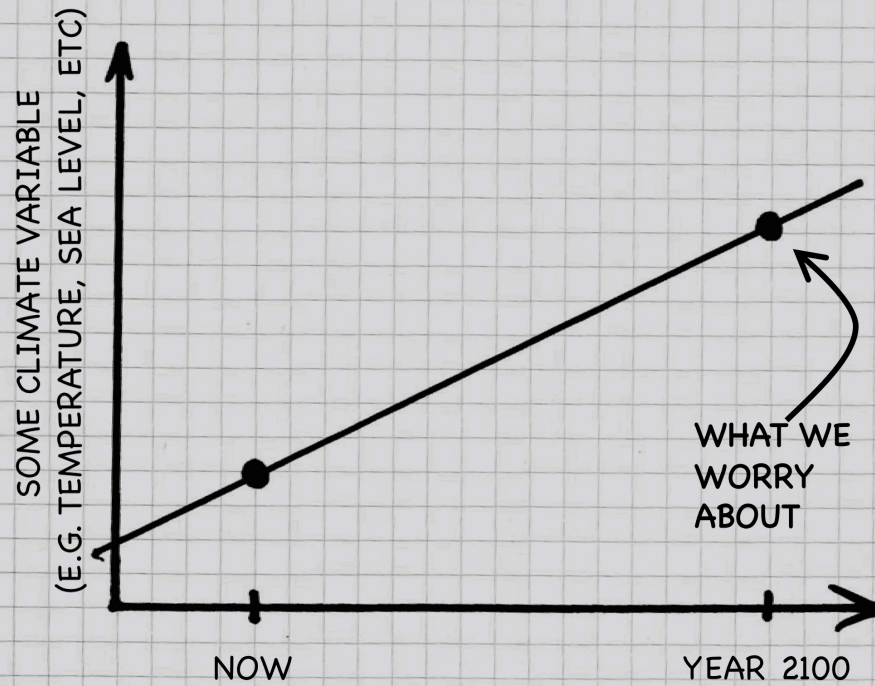
---



However, there are several definitions of fairness, which can be incompatible.



In this study, we explored the relationship between high level algorithmic accountability frameworks and domain-specific codes of ethics through auditing an ecosystem forecasting system.



Machine learning and other powerful forecasting tools are increasingly important in earth and environmental science as a result of climate change.

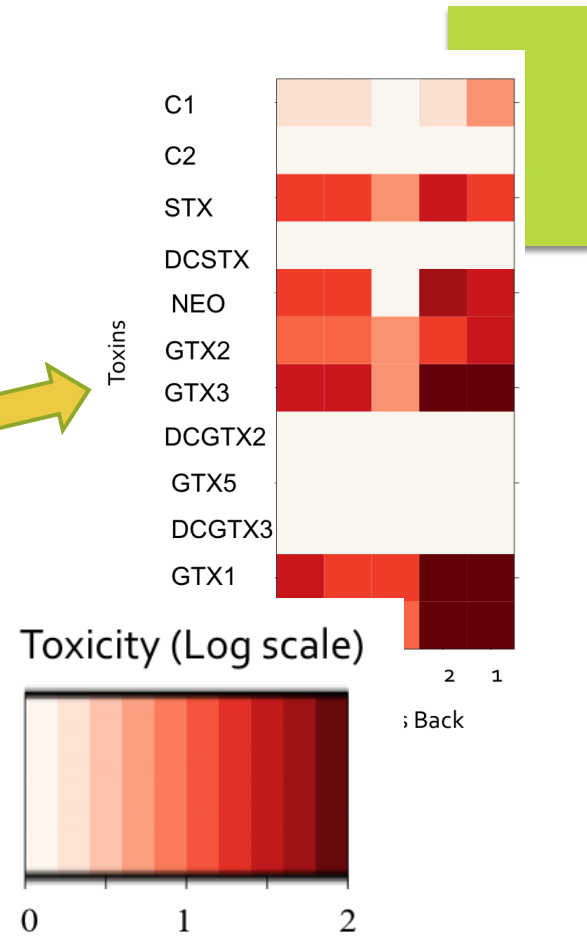
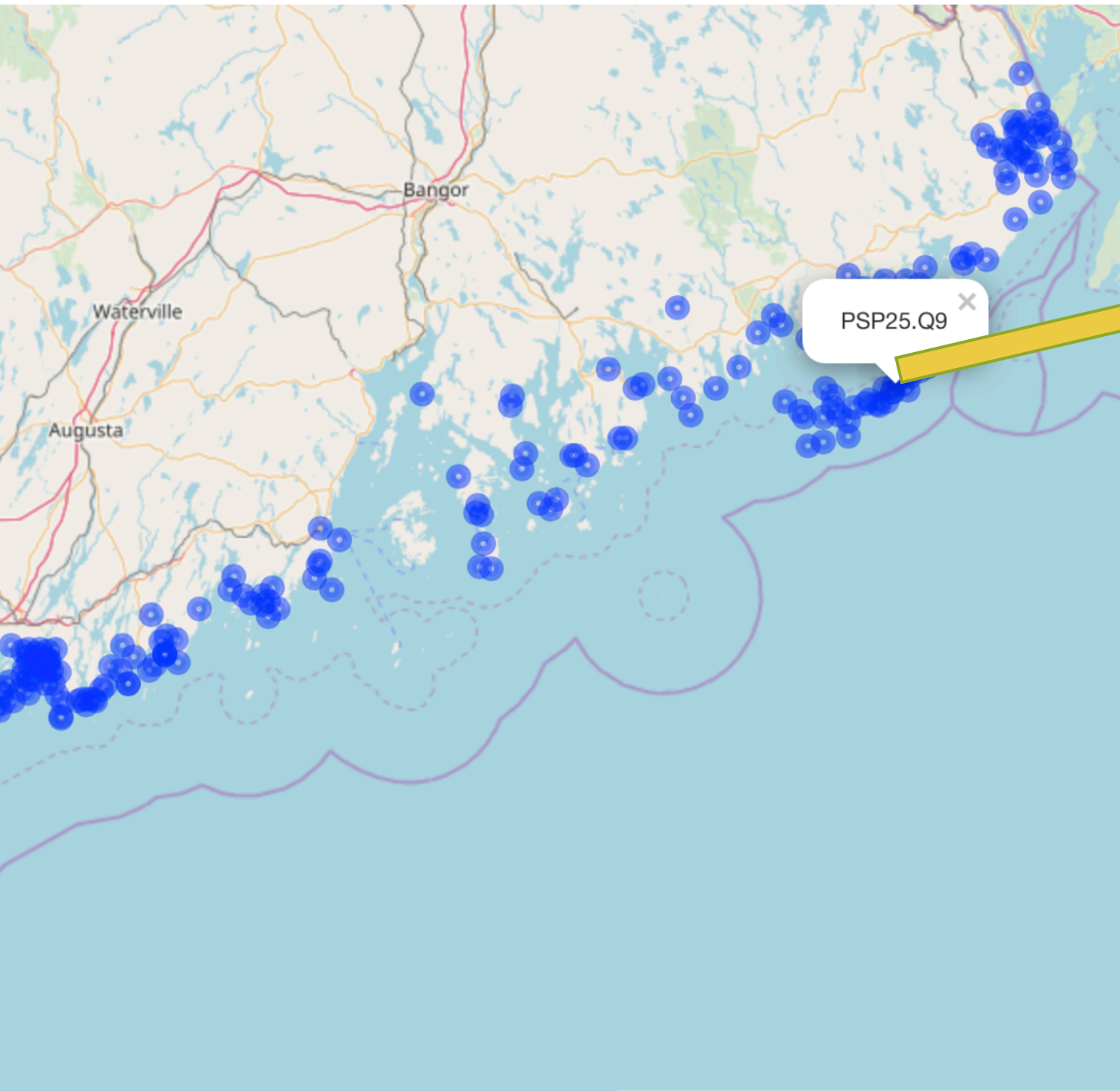


Machine learning and other powerful forecasting tools are increasingly important in earth and environmental science as a result of climate change





# The ecosystem forecast



## Model Cards for Model Reporting

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru  
mmitchellai, simonewu, andrewzaldivar, parkerbarnes, lucyvasserman, benhutch, espitzer, tgebru}@google.com  
deborah.raji@mail.utoronto.ca

**ABSTRACT**  
Machine learning models are increasingly used to perform critical tasks in areas such as law enforcement, medicine, education, and employment. In order to clarify the intended use cases of machine learning models and minimize their usage in contexts where they are not well suited, we recommend that released models be accompanied by documentation detailing their perfor-

**KEYWORDS**  
datasheets, model cards, documentation, disaggregated evaluation, fairness evaluation, ML model evaluation, ethical considerations

**ACM Reference Format:**  
Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru. 2019. Model Cards for Model Reporting. In *FAT\* '19: Conference on Fairness, Accountability, and Transparency*.

## Datasheets for Datasets

TIMNIT GEBRU, Google  
JAMIE MORGENSTERN, Georgia Institute of Technology  
BRIANA VECCHIONE, Cornell University  
JENNIFER WORTMAN VAUGHAN, Microsoft Research  
HANNA WALLACH, Microsoft Research  
HAL DAUMÉ III, Microsoft Research; University of Maryland  
KATE CRAWFORD, Microsoft Research; AI Now Institute

The machine learning community currently has no standardized process for documenting datasets, which can lead to severe consequences in high-stakes domains. To address this gap, we propose *datasheets for datasets*. In the electronics industry, every component, no matter how simple or complex, is accompanied with a datasheet that describes its operating characteristics, test results, recommended uses, and other information. By analogy, we propose that every dataset be accompanied with a datasheet that documents its motivation, composition, collection process, recommended uses,

We stepped through two algorithmic accountability frameworks developed by Gebru et al. to audit the foreco

**Contribution to the Symposium: 'The Effects of Climate Change on the World's Oceans'  
Food for Thought**

**Ethical considerations and unanticipated consequences  
associated with ecological forecasting for marine resources**

Alistair J. Hobday<sup>1\*</sup>, Jason R. Hartog<sup>1</sup>, John P. Manderson<sup>2</sup>, Katherine E. Mills<sup>3</sup>, Matthew J. Oliver<sup>4</sup>,  
Andrew J. Pershing<sup>3</sup>, and Samantha Siedlecki<sup>5</sup>

<sup>1</sup>CSIRO Oceans and Atmosphere, Hobart, Tasmania 7000, Australia

<sup>2</sup>NOAA/NEFSC Oceans & Climate Branch, Highland, NJ, USA

<sup>3</sup>Gulf of Maine Research Institute, Portland, ME, USA

<sup>4</sup>University of Delaware, Lewes, DE 19958, USA

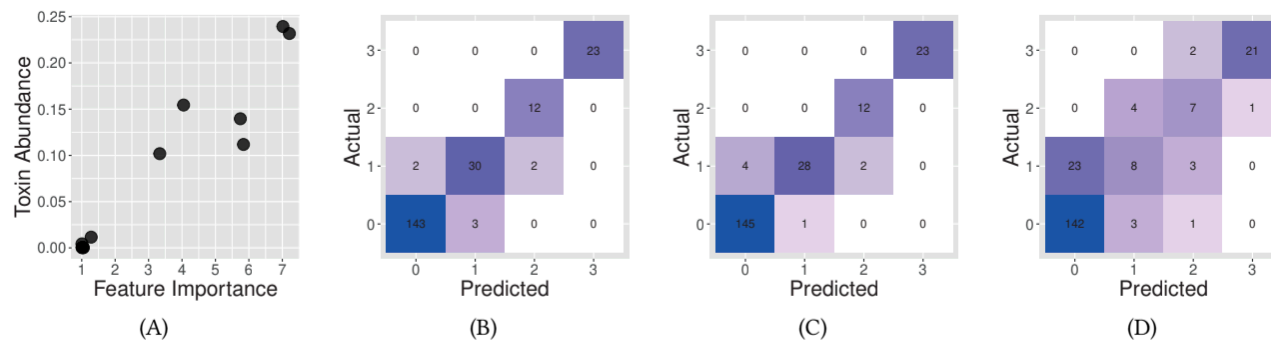
<sup>5</sup>University of Connecticut, Groton, CT 06340, USA

\*Corresponding author: tel: +61 3 62 32 5310; e-mail: [alistair.hobday@csiro.au](mailto:alistair.hobday@csiro.au).

Hobday, A. J., Hartog, J. R., Manderson, J. P., Mills, K. E., Oliver, M. J., Pershing, A. J., and Siedlecki, S. Ethical considerations and unanticipated consequences associated with ecological forecasting for marine resources. – ICES Journal of Marine Science, doi:10.1093/icesjms/fsy210.

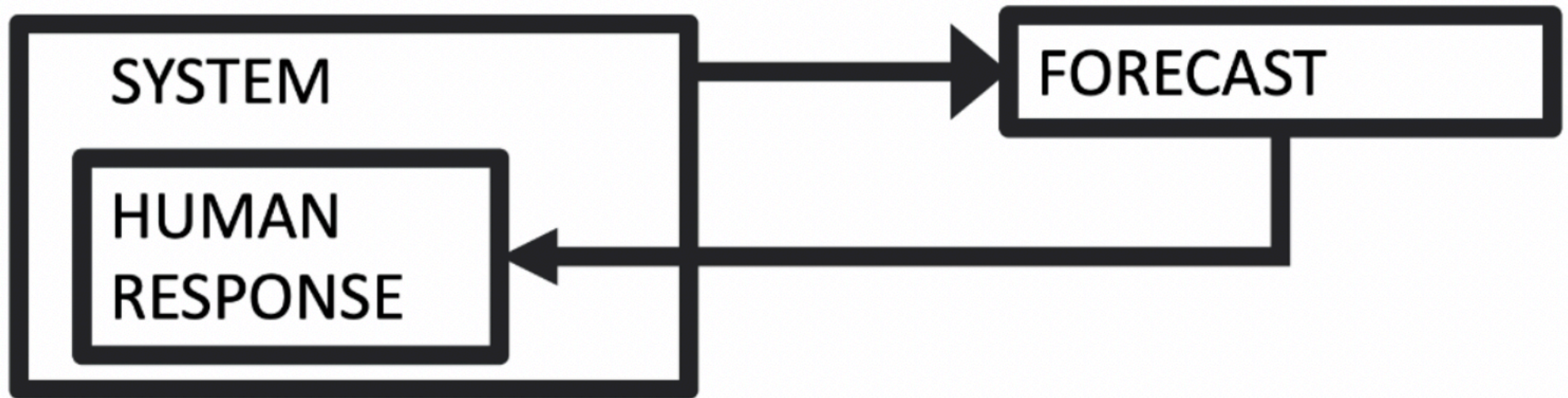
We also audited the system using a domain-specific  
code of ethics developed by Hobday et al.

# changes made to model as a result



**Figure 3: (A) Feature importance of random forest model using the twelve toxins to predict the total toxicity vs relative toxin abundance within samples. (B) Forecast results using all toxins. (C) Forecast results using the eight most important toxins. (D) Forecast results using six least important toxins, for comparison.**

# Changes made to model as a result



# Conclusions

- ▶ Operating under the standards of algorithmic accountability, ecosystem forecasters take ownership not just for the predictive power of their models, but also the human and environmental consequences.
- ▶ An algorithmic accountability framework compliments domain specific codes of ethics by incorporating the domain expertise of machine learning researchers into auditing a system.
- ▶ Ethical standards developed by Hobday et. al overlap with, but are not the same as algorithmic accountability principles developed by Gebru et al.

# Broader impacts

- ▶ The task of building domain specific ethical codes into algorithmic accountability frameworks applies more generally: essentially anywhere that algorithms are increasingly replacing, or supporting, human decision making
- ▶ Algorithmic accountability frameworks when adapted with domain-specific codes of ethics are advantageous both in terms of utility and fairness
- ▶ This methodology offers a key way to answer calls to uphold fairness and human values in each domain in which AI algorithms are used rather than search for one universal definition of fairness



Questions?