Towards Ethical Deployment of AI for Conservation Systems

Christine Kaeser-Chen
christinech@google.com
Google Research

Tanya Birch
tanyak@google.com
Google Inc.
Wildlife Insights

Katherine Chou
kic@google.com
Google Research
Google Health
Lewa Wildlife Conservancy

Tomer Gadot
tomerg@google.com
Google Inc.

Hartwig Adam
hadam@google.com
Google Research

Serge Belongie
sjb344@cornell.edu
Google Research
Cornell Tech

Tim Robertson
trobertson@gbif.org
Global Biodiversity Information Facility

Eric Fegraus
efegraus@conservation.org
Conservation International
Wildlife Insights

Dan Morris
dan@microsoft.com
Microsoft AI for Earth

ABSTRACT

The ability to collect, aggregate, and process "big data", particularly with artificial intelligence (AI) tools, has the potential to facilitate breakthrough research in conservation. As institutions develop and deploy such systems at scale, it is critical that they be designed with ethics, fairness, and transparency best practices in mind. In this short paper, we outline potential ethical concerns around AI for conservation as a starting point for further work to advance.

CCS CONCEPTS

• **Social and professional topics** → **Codes of ethics**; • **Computing methodologies** → **Artificial intelligence**; **Machine learning**.

KEYWORDS

artificial intelligence, conservation, AI ethics

1 INTRODUCTION

Machine Learning (ML) has shown great promise in conservation applications alongside traditional approaches. Citizen science organizations such as iNaturalist have used Artificial Intelligence (AI) for species classification, helping contribute research-grade observations to the scientific community. Remote sensing scientists have been applying AI for petabyte-scale geospatial analysis in forest monitoring, and conservation area planning [6]. Further research results include identifying animals using wildlife camera traps and microphone arrays [1][7]. When thoughtfully designed and appropriately deployed, AI has the potential to significantly scale the impact of current conservation efforts.

Recent advances in cloud and ML infrastructure have accelerated the development and deployment of AI solutions. Large volumes of high-fidelity data and easy access to expert and non-expert human labelers significantly reduces the time needed to prepare a training dataset for ML. Cloud computing platforms enable developers not only to store a large amount of data, but also train and serve state-of-the-art ML models. Global aggregators such as Global Biodiversity Information Facility (GBIF) federate numerous datasets, and help researchers share their datasets with broader audiences.

However, the same infrastructural maturity that accelerates AI solutions to conservation challenges has also made it easier to scale other tasks that have recently received increased attention for perpetuating existing societal disparities. For example, face recognition has been shown to bias against darker-skinned individuals when it has not been trained on a sufficiently diverse dataset [2].

Similar ethical issues now also face the AI for conservation community. Building upon previous discussions [14], we highlight four areas of ethical concerns central to the field, and invite community members to brainstorm mitigation and prevention methods.

2 ETHICAL AI FOR CONSERVATION

2.1 Explainability

AI-powered analytics is already a form of scientific evidence for downstream research such as population estimation [12] and natural resource monitoring. If users of ML-derived data are unaware of a model's biases, they may entrust the system with too much influence and credibility. As these research results may impact policy in certain applications, the evidence provided by AI systems needs to be explainable to stakeholders.

Explainable models benefit AI developers as well: certain applications in conservation, such as species identification, require substantial human expertise. Explainable AI systems help human experts examine root causes of possible erroneous AI predictions, enabling them to intervene and improve the system.

The ML community has started developing quantitative methods on explainability for both datasets and models [8][9][11][13]. But explainability cannot always be achieved post-hoc: we may need a development workflow where explainability is a consistent priority. Instead of predicting class probability, for example, models can be trained to predict an explanation of class membership, which can be verified by human experts. For applications where understanding the *causal mechanism* is important (e.g., predicting the impact of human activities on migration patterns), causal models may provide additional explainability and control to scientists [4].

2.2 Data Rights and the Fair Use of Data

The public availability of datasets and models built for conservation requires a cautious analysis of the potential benefits to the AI and biological science communities (such as on lila.science), along with the potential consequences of malevolent access. In some instances, releasing models or datasets that facilitate the identification of at-risk species may create threats to wildlife.

Transparency to data and model providers as to how their assets may be used is key to gaining trust in AI safety. In other domains, an institutional review board (IRB) often assists with assessing these tradeoffs; perhaps an analogous process may be instituted around conservation data and models. Considerations for such a board would include potential misuse, along with safeguards that may minimize risk. A conservation IRB might also make recommendations to limit distribution only to vetted organizations.

Discretion must be exercised when releasing location data for threatened species. For some applications, data revealing precise locations must be obfuscated within a reasonable range, determined on a species-by-species basis (e.g., species with limited range or abundance might need larger buffers than migratory species) and paying attention to any mosaic effect that might inadvertently reveal location [3]. In other cases, no location data for at-risk species should be made public if it puts the species further at risk.

Standard practices and services to enable easy citation tracking, such as with DOI, should be provided. Fully qualified data citations, combined with unique identifiers, give data providers attribution and encourage data sharing.

2.3 Unfair Bias and Misuse

The AI for conservation community is well-intentioned, but misuse can still happen. **Unfair bias in model design** can occur when an individual labeler's interpretation instead of an adjudicated truth is used as ground truth, resulting in bias towards well known or traditional classes. **Unfair bias in training data** due to insufficient data, poor data quality, or even systematically missing data for certain scenarios, can result in minority bias, where groups are insufficiently represented to learn the correct statistical pattern.

On the other hand, if model deployment context differs significantly from the training context, **overexposure to risk during deployment** can cause unintended consequences. One possible mitigation is to ensure proper and fair data and model use through access controls and review boards as described before. Identifying protected groups, having red teams, and setting up periodic model performance reviews that evaluate for equal allocation, accuracy, and outcomes prior to deployment can also help address misuse.

Institutions publishing datasets and models to advance conservation research should consider including *datasheets* or *release notes* with their datasets and models [5][10], documenting current limitations and targeted application scenarios of the published resources. As a community, we have the obligation to make our dataset and models more diverse and representative over time.

2.4 Implications for Human Privacy and Safety

As the AI community strives to contribute to conservation goals, the intention of the AI application, whether for measuring change in habitat, population health, or preventing illegal action, must be made transparent to prevent scope creep. Additional caution must be exercised when AI systems impact human privacy and safety.

In some cases, human information may be incidentally recorded by the system, and further work on protecting personally identifiable information (PII) is needed. In other cases, human information is intentional recorded and critical to the application. When collecting and making use of such data, the AI for conservation community must follow internationally accepted laws and norms with respect to surveillance and human rights. For example, humans in camera trap images can help monitor illegal activity but also present an issue for privacy. Irrespective of intent, sharing data that contains human information must comply with existing PII-related policies. Opportunities exist for building AIs to pre-process data, and redact or obfuscate all PII from the media prior to publishing.

Systematic biases against human communities could be amplified and reinforced through AI. An example is applying AI for optimizing patrol against illegal wildlife capturing activities. In a recent crowdsourced surveillance project on poaching, volunteers from across the world used an online platform to annotate potential human intruders based on remote sensing camera footage. If AI were trained on this data, it could cause unfair bias against communities which volunteers perceive to be more threatening, and lead to disparate patrolling.

Transparency in AI systems can mitigate this issue. Organizations adopting AI for assessing human impacts should be transparent regarding their technical partners, and engage with domain experts and local communities when developing the application. Vendors supplying such AI technology should support examination and verification on the makeup of their training data and model.

3 RECOMMENDATIONS AND NEXT STEPS

In this short paper, we outline several ethics concerns around AI for conservation. To facilitate further research and discussions, we recommend the following next steps:

- Make explainability a requirement for AI systems used in conservation research and policy making.
- Establish community-wide data use terms and citation best practices. High-value datasets and models should be handled with additional guidelines provided by conservation IRBs.
- Published datasets and models should include additional release notes, documenting known biases in the datasets, fairness metrics, and targeted deployment scenarios.
- When human information is recorded, the data collection procedure, sharing mechanism, and targeted use must be reviewed by conservation IRBs prior to execution.
- Conservation policy-makers and institutions should be transparent regarding the supply chain for AI-powered tools. Datasets and technology of public interest should be made available for examination and verification.
- We should continue engaging in community discussions on ethical AI, and collaborate with stakeholders to establish a more complete set of guidelines for future development.

ACKNOWLEDGMENTS

We thank our collaborators Sayali Kulkarni, Kiat Chuan Tan, Yulong Liu, and Chenyang Zhang for their contribution to the discussion.

REFERENCES

[1] Sara Beery, Grant Van Horn, and Pietro Perona. 2018. Recognition in Terra Incognita. In *Proceedings of the European Conference on Computer Vision*. Munich, Germany.

[2] Joy Buolamwini and Timnit Gebru. 2018. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*. PMLR. http://proceedings.mlr.press/v81/buolamwini18a.html

[3] A.D. Chapman and Oliver Grafton. 2008. Guide to Best Practices for Generalising Sensitive Species Occurrence Data. (2008). https://www.gbif.org/document/80512

[4] Adnan Darwiche. 2018. Human-level Intelligence or Animal-like Abilities? *Commun. ACM* (2018). https://doi.org/10.1145/3271625

[5] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna M. Wallach, Hal Daumé III, and Kate Crawford. 2018. Datasheets for Datasets. *CoRR* (2018). http://arxiv.org/abs/1803.09010

[6] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment* (2017). https://doi.org/10.1016/j.rse.2017.06.031

[7] Matt Harvey. 2018. Acoustic Detection of Humpback Whales Using a Convolutional Neural Network. Retrieved May 2, 2019 from https://ai.googleblog.com/2018/10/acoustic-detection-of-humpback-whales.html

[8] Pang Wei Koh and Percy Liang. 2017. Understanding Black-box Predictions via Influence Functions. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*.

[9] Gilles Louppe. 2014. *Understanding Random Forests: From Theory to Practice*. Ph.D. Dissertation. https://doi.org/10.13140/2.1.1570.5928

[10] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2018. Model Cards for Model Reporting. *CoRR* (2018). http://arxiv.org/abs/1810.03993

[11] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. https://doi.org/10.1145/2939672.2939778

[12] Daniel Rubenstein, Jason Parham, Charles Stewart, Tanya Berger-Wolf, Jason Holmberg, Jon Crall, Belinda Low Mackey, Sheila Funnel, Kasmira Cockerill, Zeke Davidson, Lizbeth Mate, Cosmas Nzomo, Rosemary Warungu, Dino Martins, Vincent Ontita, Joy Omulupi, Jennifer Weston, George Anyona1, Geoffrey Chege1, David Kimiti, Kaia Tombak, Andrew Gersick, and Nancy Rubenstein. 2018. *"The State of Kenya's Grevy's Zebras and Reticulated Giraffes: Results of the Great Grevy's Rally 2018"*. Technical Report.

[13] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. 2017. Axiomatic Attribution for Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*.

[14] Oliver R. Wearn, Robin Freeman, and David M. P. Jacoby. 2019. Responsible AI for conservation. *Nature Machine Intelligence* (2019). https://doi.org/10.1038/s42256-019-0022-7