Responsible AI for Conservation

Oliver R. Wearn[1*], Robin Freeman[1] & David M.P. Jacoby[1*]

[1]*Institute of Zoology, Zoological Society of London, Regent's Park, London, NW1 4RY, U.K.*

*correspondence: david.jacoby@ioz.ac.uk; oliver.wearn@ioz.ac.uk
DMPJ – ORCID ID: 0000-0003-2729-3811
ORW – ORCID ID: 0000-0001-8258-3534

Standfirst: AI promises to be an invaluable tool for nature conservation, but its misuse could have severe real-world consequences for people and wildlife. Conservationists discuss how improved metrics and ethical oversight can mitigate these risks.

Machine learning (ML) is revolutionizing efforts to conserve nature. ML algorithms are being applied to predict the extinction risk of thousands of species[1], assess the global footprint of fisheries[2], and identify animals and humans in wildlife sensor data recorded in the field[3]. These efforts have recently been given a huge boost with support from the commercial sector. New initiatives, such as Microsoft's 'AI for Earth'[4] and Google's 'AI for Social Good', are bringing new resources and new ML tools to bear on some of the biggest challenges in conservation. In parallel to this, the open data revolution means that global-scale, conservation-relevant datasets can be fed directly to ML algorithms from open data respositories, such as Google Earth Engine for satellite data[5] or MoveBank for animal tracking data[6]. Added to these will be

‘Wildlife Insights’, a Google-supported platform for hosting and analysing wildlife sensor data which launches this year. With new tools and a proliferation of data comes a bounty of new opportunities, but also new responsibilities.

Potential for AI misuse and misinterpretation

The opaque nature of some ML algorithms means that the potential for unintended consequences may be high and this could have real-world consequences for people and wildlife. Understanding, even in an intuitive sense, how neural networks process a given input can currently be very challenging. This has several ramifications which are not yet fully appreciated in the conservation field. Firstly, it can be difficult to identify the implicit assumptions of an algorithm (for example, how much of the contextual background information it is using when identifying species in images), and therefore the potential risks of using it for this purpose. Secondly, it might be unclear when an algorithm is being asked to make predictions beyond the scope of the training data. Indeed making sure algorithms ‘fail gracefully’ is a major research problem[7]. Thirdly, an algorithm might not be easily interrogated as to why it made a particular decision. Whilst these considerations are well-appreciated within the wider AI community, they have been largely absent from recent discussions around the potential benefits of the technology to conservation.

The use of ML to solve conservation problems without consideration of these factors might have severe negative outcomes. A bias against under-represented classes in a dataset could, for example, mean that a rare species is missed during an Environmental Impact Assessment, leading to the eventual loss of its habitat due to development. Equally, the use of training data with poor coverage of the domain in which predictions are to be made could result in a species being wrongly assessed as extinct on the IUCN Red List, meaning that conservation resources are diverted elsewhere (the 'Romeo Error'[8]). A misclassification error could also wrongly flag local community members as poachers, raising potentially severe legal and safety concerns. Under any of these circumstances the lack of interpretability and accountability we have for an algorithm's decision would be laid bare[9]. As such, perverse outcomes of applying ML in conservation have the potential to waste scare resources, increase the costs of conservation to local communities, and erode trust in science-led approaches to environmental problems.

Better metrics needed

That is not to say that these outcomes are unavoidable. One area of research that conservation might benefit from is the development – by ML researchers and conservationists working together – of better metrics for assessing the usefulness of any given algorithm for actually doing conservation.

Currently, much of the focus is on standard predictive accuracy metrics. Whilst useful for assessing performance in a controlled 'laboratory' setting (using a single or very limited number of datasets, sometimes with pre-treatment of the data such as cropping), accuracy metrics may prove inadequate once the algorithms are released to make automated decisions in the wild. Here, extraneous factors may play a much more important role in the output than anticipated. For example, there has been a recent proliferation of studies presenting deep neural network approaches to classifying imagery taken by autonomous cameras (camera traps) deployed in the field[10,11], with reported accuracies as high as 98%. Perhaps not helped by media reports, these studies can sometimes be seen as a 'silver bullet' to solve some of the major bottlenecks in wildlife monitoring today. However, simple accuracy metrics are unlikely to provide a good indicator of success when an algorithm is transferred to new datasets, for example for a new point in time, a new study site, or on different species. If the model requires re-training, conservationists are unlikely to have the same abundance of training data as the original study. Perhaps more importantly, accuracy metrics may tell us little about how accurately we will in practice be able to monitor the populations of a suite of species.

Conservation has been here before

As well as better metrics, we need better ethical oversight of the use of AI in conservation. We have been here before: a promising new tool is developed, followed by a period of mass uptake amongst conservationists. We then enter a period of critical appraisal, eventually resulting in the well-considered and effective use of the approach. A good example of this is Population Viability Analysis (PVA), a widely-used tool to predict the risk of a species going extinct in the future. PVA first emerged in the 1980s and then saw a surge in use during the 1990s, especially after software became available offering a 'canned' approach. Towards the end of that decade, various researchers began to critique the use of PVAs in conservation, expressing the view that it could act as a 'loaded gun' in the wrong hands, rather than an aid to conservation[12]. Best-practice guidelines were eventually promoted[13,14], nearly two decades after the tool first emerged.

The AI community as a whole is already grappling with the concepts of 'fair AI' and goal alignment – central tenets of the Beneficial AI movement – and there is much that conservationists could learn from[15,16]. There is also an emerging consensus within the broader AI community on what responsible and ethical guidelines for AI development look like (e.g. the Asilomar AI Principles or the Biosphere Code Manifesto). Such guidelines for conservation could be designed to steer algorithm development in the right direction for humanity and wildlife in ways that are collaborative, maximally beneficial, liberating and yet robust to misuse and corruption (for example by those involved in the illegal wildlife trade). As conservationists, we are already familiar with ethical oversight of our practices, in particular with respect to the care and husbandry of animals in research (for example, animals kept in captivity for the purposes of ex-situ conservation, or animals captured and released for the purposes of research). Robust ethical review processes already exist in many research departments and ethics statements on the use of animals are often provided, or required, in published research. A pragmatic approach may therefore be to encourage the inclusion of a 'Responsible AI' statement, which outlines the ethical review process, provides responsible guidance on the limits to an algorithm, and gives a description of the training data involved. This would not only promote greater transparency but would also ensure that researchers are able to demonstrate that they have considered both the generalities and the limitations of their method.

Given the potentially severe social and environmental costs of AI misuse and misinterpretation in conservation, we ask whether we might avoid the pitfalls of the past by building, from the outset, the technical and ethical capacity to harness these new tools responsibly. With this in mind, we have outlined two potential goals for the conservation and AI communities to tackle in the immediate term: the development of metrics to better allow conservationists to assess the usefulness of an algorithm, and the formulation of ethical guidelines for the responsible use of AI in conservation. Importantly, these metrics and guidelines will need to exist in the *application* domain, not just within the machine intelligence field. Critical to this will be the input of the AI community. Now is the time to bring together conservationists, AI experts and industry, to ensure maximum benefit with minimum harm comes from the application of AI to protect the Earth's most threatened species and habitats.

Competing Interests: The authors declare no competing financial and non-financial interest.

References

1. Darrah, S. E., Bland, L. M., Bachman, S. P., Clubbe, C. P. & Trias-Blasi, A. *Divers. Distrib.* **23,** 435–447 (2017).
2. Kroodsma, D. A. *et al. Science (80-.).* **908,** 904–908 (2018).
3. Mac Aodha, O. *et al*. *PLoS Comput. Biol.* **14,** 1–19 (2018).
4. Joppa, L. N. *Nature* **552,** 325–328 (2017).

5. Gorelick, N. *et al. Remote Sens. Environ.* **202,** 18–27 (2017).

6. Kranstauber, B. *et al. Environ. Model. Softw.* **26,** 834–835 (2011).

7. Amodei, D. *et al.* arXiv preprint arXiv:1606.06565 (2016).

8. Collar, N. J. *Oryx* **32,** 239–243 (1998).

9. Doshi-Velez, F. & Kim, B. arXiv preprint arXiv:1702.08608 (2017).

10. Tabak, M. A. *et al. Methods Ecol. Evol.* (2018).

11. Norouzzadeh, M. S. *et al. Proc. Natl. Acad. Sci.* 1–10 (2018).

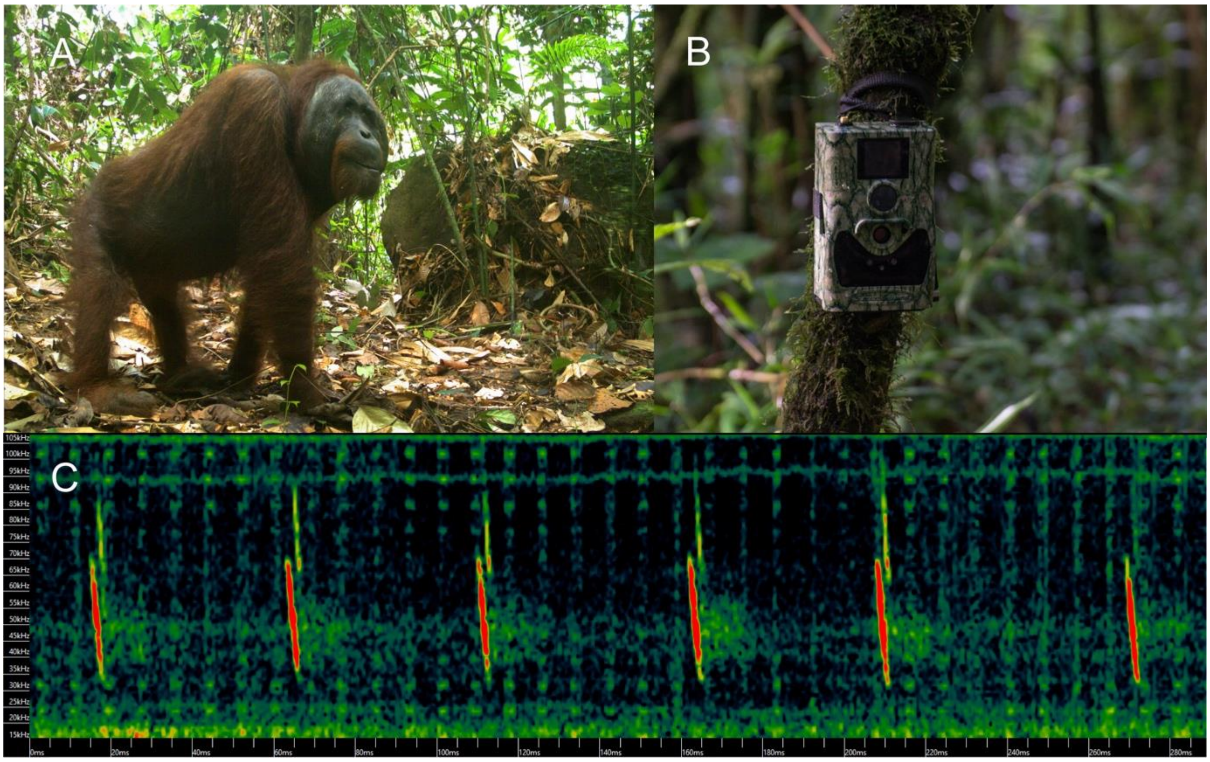
12. Burgman, M. & Possingham, H. P. in *Genetics, Demography and Viability of Fragmented Populations* (eds. Young, A. G. & Clarke, G. M.) 97–112 (Cambridge University Press, 2000).

13. Reed, J. M. *et al. Conserv. Biol.* **16,** 7–19 (2002).

14. Ralls, K., Beissinger, S. R. & Cochrane, J. F. in *Population Viability Analysis* (eds. Beissinger, S. R. & McCullough, D. R.) 521–550 (University of Chicago Press, 2002).

15. Crawford, K. & Calo, R. T. *Nature* **538,** 311–313 (2016).

16. Zou, J. & Schiebinger, L. *Nature* **559,** 324–326 (2018).



Machine learning algorithms on the front line of conservation. ML methods are applied to identify wildlife or people using sensors deployed in the field. An image of a Critically Endangered Bornean orangutan (*Pongo pygmaeus)*, **A**, captured using a camera trap, **B**. A threatened bat species (*Natalus primus*) detected on a sonogram from *in situ* acoustic monitoring, **C**.