



the performance of the model itself (17). Increasing clinician awareness of AI's biases is critical, but this desire may be paradoxical: Knowing about biases in AI may result in less willingness to use AI-based recommendations for patients that a clinician judges "different" from others. Assuming models are biased in terms of race or ethnicity, for example, could result in clinicians systematically overriding a model's recommendation for that group of patients.

Several strategies exist to identify and address latent biases. One strategy could involve providing clinicians with model-specific, individual-level performance feedback regarding whether they tend to outperform or underperform it, or if they are systematically following or overriding a model only for certain patient groups. Individualized feedback has the potential to improve clinician performance (12). However, a challenge for assessing bias is that clinicians may not see sufficient numbers of patients in different groups to allow rigorous, stratified comparisons.

Patients should be informed about the use of AI in their clinical care as a matter of respect. This includes general messaging about the use of predictive algorithms, chatbots, and other AI-based technologies, and specific notification when new AI-based technologies are used in their individual care. Doing so may improve awareness of AI, motivate conversations with clinicians, and support greater transparency around AI use.

Exactly how much to disclose, and in what format, are unanswered questions that require additional research. There is a need to avoid AI exceptionalism—the idea that AI is riskier or requires greater protection, just because it is AI—and presently patients want to know more, not less (8). That other decisions relying on algorithms, such as clinical risk calculators or computer-aided radiographic or electrocardiogram interpretation, may not be routinely shared with patients is not an argument in favor of secrecy.

Bias has not been a major aspect of drug and device regulations, which focus on overall safety and efficacy. Recent US proposals could extend legal liability to physicians and hospitals, meaning they could be required to provide compensatory damages to patients or be subject to penalties for use of biased clinical algorithms; these could be applied to AI algorithms (13). However, the complexity of AI algorithms and persistent ethical disagreement over when differential performance by race or ethnicity equals true bias complicate liability proposals. Drug and device regulatory agencies might consider making evaluations of bias mandatory for approval (14).

A first step could be requiring evaluations of differential performance and bias under different real-world assumptions in approval processes and other forums, such as in journal reporting of AI research.

In addition, the gaze of AI should be turned on itself. This requires proactive, intentional development of AI tools to identify biases in AI and in its clinical implementation (15). AI may contribute to the emergence of biases, but it also has the potential to detect biases and hence facilitate new ways of overcoming them. Open-source tools, such as AI Fairness 360, FairML, and others, show promise in helping researchers assess fairness in their machine learning data and algorithms. These tools can assess biases in datasets, predictive outputs, and even the different techniques that can be used to mitigate bias according to different metrics of fairness. Their application to health care data and algorithms deserves rigorous scientific examination.

Implementation research is urgently needed to better understand the role of different contextual factors and latent conditions in allowing biases to emerge. Exactly which patients may experience bias under which circumstances requires ongoing rigorous study. In AI, biased data and biased algorithms result in biased outcomes for patients, but so do unbiased data and algorithms when they enter a biased world. All patients deserve to benefit from both fair algorithms and fair implementation. ■

REFERENCES AND NOTES

1. World Health Organization, *Ethics and Governance of Artificial Intelligence for Health* (World Health Organization, 2021).
2. S. Guerrero *et al.*, *Sci. Rep.* **8**, 13978 (2018).
3. Z. Obermeyer, B. Powers, C. Vogeli, S. Mullainathan, *Science* **366**, 447 (2019).
4. S. M. West *et al.*, *Discriminating Systems: Gender, Race and Power in AI* (AI Now Institute, 2019).
5. A. J. London, *Cell Rep. Med.* **3**, 100622 (2022).
6. C. M. Lowe, *Qual. Saf. Health Care* **15** (suppl. 1), i72 (2006).
7. J. Yoo *et al.*, *JMIR Med. Inform.* **8**, e23351 (2020).
8. A. Tyson *et al.*, "60% of Americans Would Be Uncomfortable with Provider Relying on AI in Their Own Health Care" (Pew Research Center, 2023).
9. D. P. Ly, *JAMA Health Forum* **2**, e212333 (2021).
10. W. N. Price 2nd, S. Gerke, I. G. Cohen, *JAMA* **322**, 1765 (2019).
11. A. Bracic, S. L. Callier, W. N. Price 2nd, *Science* **377**, 1158 (2022).
12. N. Ivers *et al.*, *Cochrane Database Syst. Rev.* **6**, CD000259 (2012).
13. K. E. Goodman, D. J. Morgan, D. E. Hoffmann, *JAMA* **329**, 285 (2023).
14. K. Dortche *et al.*, *J. Sci. Policy Gov.* **10**, 38126/JSPG210302 (2023).
15. R. B. Parikh, S. Teeple, A. S. Navathe, *JAMA* **322**, 2377 (2019).

ACKNOWLEDGMENTS

The authors thank K. Tang for help with the illustration, M. Akerson for references, and M. Wynia and L. Hunter for comments on earlier drafts. M.D. receives funding from the NIH (grant R01NR019782) and Greenwall Foundation.

PERSPECTIVE

Using machine learning to decode animal communication

New methods promise transformative insights and conservation benefits

By Christian Rutz¹, Michael Bronstein^{2,3}, Aza Raskin⁴, Sonja C. Vernes^{1,5}, Katherine Zacarian⁴, Damián E. Blasi⁶

The past few years have seen a surge of interest in using machine learning (ML) methods for studying the behavior of nonhuman animals (hereafter "animals") (1). A topic that has attracted particular attention is the decoding of animal communication systems using deep learning and other approaches (2). Now is the time to tackle challenges concerning data availability, model validation, and research ethics, and to embrace opportunities for building collaborations across disciplines and initiatives.

Researchers must infer the meaning, or function, of animal signals through observation and experimentation (3). This is a challenging task, not least because animals use a wide range of communication modalities, including visual, acoustic, tactile, chemical, and electrical signals—often in conjunction, and beyond humans' perceptive capabilities. Observational work focuses on recording the signals of interest as well as detailed contextual information, including the identity, state, and behavior of both the senders and receivers of signals, their relationships and past interactions, and relevant environmental conditions. Some signal types may be produced only under certain circumstances, eliciting a specific behavioral response; a classic example is a vervet monkey (*Chlorocebus pygerythrus*) giving an alarm call when it spots a predator, which causes group members to

¹School of Biology, University of St Andrews, St Andrews, Scotland, UK. ²Department of Computer Science, University of Oxford, Oxford, UK. ³Project CETI, New York, NY, USA. ⁴Earth Species Project, Berkeley, CA, USA. ⁵Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands. ⁶Department of Human Evolutionary Biology, Harvard University, Cambridge, MA, USA. Email: christian.rutz@st-andrews.ac.uk



Machine learning is helping to map and understand the vocal repertoire of tool-using Hawaiian crows (*Corvus hawaiiensis*).

seek shelter. Establishing such correlations enables the formulation of hypotheses about signal function that can then be tested experimentally (e.g., with controlled playbacks).

Following this approach, decades of careful research have produced major advances in understanding animal communication (3). But there are considerable challenges, such as avoiding anthropocentric biases in data collection and interpretation, processing ever-increasing volumes of data, charting the full complexity of animals' signaling behavior, and achieving comprehensive functional decoding. ML offers some potential solutions.

Animal signals can be investigated using a rich toolkit of increasingly powerful ML methods, which vary in their modeling objectives, data requirements, and reliance on expert annotation. This includes, among other approaches, supervised learning (e.g., for determining which features accurately predict human-labeled signal types) and unsupervised and self-supervised learning (e.g., for discovering signal repertoires of an individual, group, or population).

Self-supervised deep-learning methods (4) are of interest because they require neither annotated datasets nor predefining features that are potentially relevant for communication. They are also the basis of "foundation models," which are capable of remarkable generalizations across tasks (5). For exam-

ple, large language models that have been trained to predict the next word from a given sequence of words can subsequently be used to carry out much more complex tasks, such as inferring the syntactic classes of, and relations between, linguistic units, or generating realistic text (5).

Methods that can integrate different data modalities seem particularly promising for facilitating functional decoding because they can provide a fuller account of communication events. ML models have been developed that efficiently learn to link images to words, words to speech, and content across other modality combinations (5), and this approach could be applied productively to animal study systems, for example, by correlating vocalizations with specific behaviors. ML would effectively assist with the challenging task of detecting cross-modal associations (and structure) that can, in turn, inform the design of validation experiments to establish causality (see the figure).

Because many ML methods were originally developed for natural language processing, exciting avenues have started opening up for exploring the much-debated potential similarities between human language and animal communication systems (6). Observations and experimental work suggest that at least some animals, such as southern pied babblers (*Turdoides bicolor*), exhibit some of the

order sensitivity and compositionality that are characteristic of human language (7). ML approaches could leverage large datasets to search for subtlety and complexity that elude traditional methods, potentially expanding the known set of communication features shared across divergent taxa.

There is a growing number of studies that are exploiting the potential of ML for investigating animal communication, including large collaborative initiatives, such as the Earth Species Project (ESP); Communication and Coordination Across Scales (CCAS); Vocal Interactivity in-and-between Humans, Animals and Robots (VIHAR); Interspecies Internet; and Project CETI (Cetacean Translation Initiative), which recently provided a detailed roadmap for ML-assisted work on sperm whale (*Physeter macrocephalus*) communication (2). Although efforts to tackle this grand research challenge are clearly intensifying, the field faces at least two main data-related obstacles: Most methods require vast amounts of data (4), and recordings of a single modality (e.g., vocalizations) are insufficient for functional decoding; additional context is required, including information on the animals' behavior and environment.

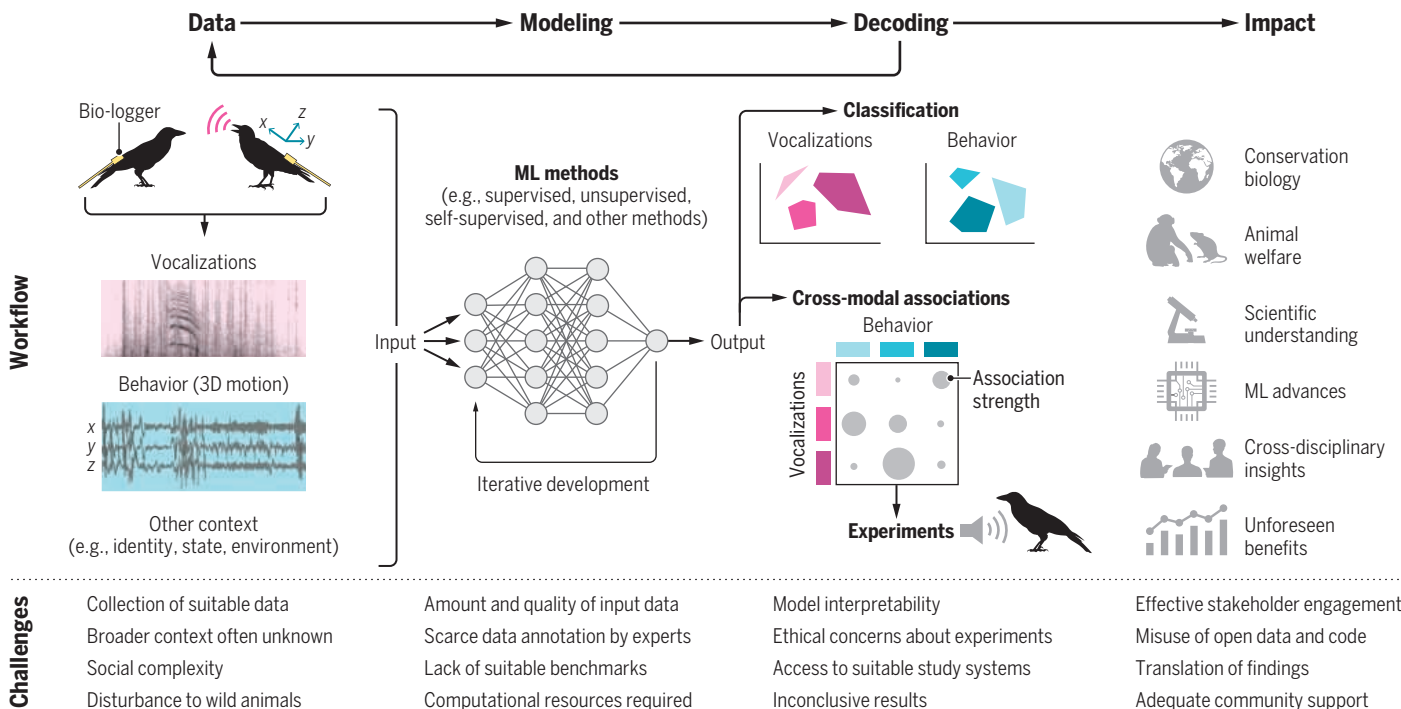
Large volumes of audio and video data are held in community-sourced archives (such as the Macaulay Library or xeno-canto), are being accumulated by passive recording arrays, or can be scraped from the internet. Mining these data sources will provide fascinating glimpses of the richness of animal communication, but on its own, such work is unlikely to achieve breakthroughs in decoding signal function. This is chiefly because robust information on the identities and states of the senders and receivers, and the specific communication context, is usually lacking.

High-quality datasets are available for some taxa, enabling swift progress with core model-development objectives. But it is clear that community mobilization and appropriate resourcing are required to ensure that species experts are fully involved in the annotation and interpretation of existing recordings and can lead targeted efforts to collect new data at scale, in both the laboratory and field. For wild animals, a range of methods can be used to collect suitable datasets, including observation of focal subjects, autonomous cameras and audio recorders, drones and robots, and animal wearables (bio-loggers). Some bio-logging devices can collect audio and body-motion data simultaneously for the same individual, providing valuable input for multimodal ML models (see the figure).

The journey could be the biggest reward. Training the lens of ML on a broad range of taxa will likely uncover surprising degrees

Using multimodal data and experiments to understand animal signals

Machine-learning (ML) methods can be used to integrate information on sender, receiver, and communication context, revealing patterns that may inform hypotheses about signal function and, in turn, the design of controlled experiments. ML-assisted research on animal communication will likely generate important benefits, such as improving animal conservation and welfare, but is not without its challenges; addressing ethical concerns is a top priority.



of previously hidden complexity in animals' communicative behavior. Many of the species that appear to use only a handful of basic call types may turn out to possess rich vocal repertoires, and those that are renowned for their sophisticated communication may be shown to be more impressive still. There are early indications of the discovery potential of ML, as highlighted by a recent study that explored individual and group differences in the vocal behavior of zebra finches (*Taeniopygia guttata*) (8).

The ability of ML to produce systematic inventories of vocal (or other signaling) output across a diverse range of taxa will enable unprecedented comparative analyses, helping researchers to pinpoint the evolutionary drivers, genomic signatures, life-history correlates, and cognitive and sensory foundations of different communication systems. At the same time, longitudinal recordings for individual subjects could reveal how communication skills arise and mature (9).

But perhaps most importantly, advances in this field could boost animal conservation and welfare. For example, in critically endangered species, such as the Hawaiian crow (*Corvus hawaiiensis*), comparisons with historical baseline data could generate a detailed record of how population bottlenecks have altered vocal repertoires, potentially leading to impoverished commu-

nicative capabilities (10); lost calls of high fitness relevance, such as those involved in foraging, courtship, or antipredator behavior, could then conceivably be reintroduced. Furthermore, there is increasing recognition that socially transmitted information may affect population viability (11), as illustrated by foraging specializations in killer whales (*Orcinus orca*) (12). Where vocal dialects can be established as "cultural markers," ML approaches would enable automated mapping of social population structure and identification of animal groups at risk of losing critical knowledge.

ML could also be used to identify animal signals that are associated with stress, discomfort, pain, and evasion, or with positive states, such as arousal and playfulness. This could provide momentum for improving the living conditions of livestock and other captive animals and may even enable the assaying of wild populations to measure the impact of anthropogenic stressors. Ecological "soundscape" analyses are, at present, largely focused on species detection, but it should be possible to listen in on animals' welfare at the landscape level (13). This idea could be developed further by looking beyond communication, for example, by developing ML tools that can examine satellite-recorded animal movement tracks for signatures of disease, distress, or human avoidance.

Despite manifold potential benefits, ML-assisted research on animal communication raises major ethical questions, such as under what circumstances it is acceptable to conduct playback experiments with wild animals. Advanced chatbots may enable researchers to initiate communication with animals before signal function is fully understood, potentially causing unintended harm. For example, broadcasting vocalizations to wild humpback whales (*Megaptera novaeangliae*) could inadvertently trigger changes in singing behavior on an ocean-basin scale. These issues must be tackled head-on and not as an afterthought. Cross-stakeholder consultation is urgently needed to develop best-practice guidelines and appropriate legislative frameworks (14).

Other challenges and opportunities lie ahead. For example, it is important to coordinate research efforts across existing initiatives and to enhance engagement of experts on animal communication, tracking, conservation, and welfare. Despite rapid technological advances, progress in this field will continue to depend on careful consideration of each study species' biology, a detailed knowledge of communication contexts, and controlled behavioral experiments (3). This expertise is essential for informing and validating ML analyses and intensifying data interpretation and collection efforts. Professional societies

and networks could help coordinate inclusive community-driven collaborations.

Workflows should be developed using study systems where data collection and experimental validation are relatively straightforward. In captive settings, researchers can ensure excellent experimental control as well as the highest ethical and welfare standards; good models include rodents, bats, and birds. Such work can be complemented with analyses of extensive field datasets that are already available for some species. Once methods have been established, they can be cautiously applied to the considerably more challenging problem of studying difficult-to-observe wild animals.

Present developments in ML are exceptionally fast paced. Beyond the use of deep-learning methods, there is scope for trialing other ML frameworks, such as reinforcement learning and meta-learning (i.e., learning from the output of other ML models). As models are developed, formal “benchmarking” will be key to improving the reliability and efficiency of analysis pipelines (15), although safeguards must be put in place to prevent the misuse of open resources, such as attempts to disturb, kill, or weaponize animals.

ML holds the potential to generate transformative advances in our understanding of animal communication systems, uncovering unimagined degrees of richness and sophistication. But it is essential that future advances are used to benefit the animals being studied. ■

REFERENCES AND NOTES

1. D. Tuia et al., *Nat. Commun.* **13**, 792 (2022).
2. J. Andreas et al., *iScience* **25**, 104393 (2022).
3. J. W. Bradbury, S. L. Vehrenkamp, *Principles of Animal Communication* (Sinauer, 2011).
4. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning* (MIT Press, 2016).
5. R. Bommasani et al., arXiv:2108.07258 (2022).
6. A. R. Fishbein et al., *Philos. Trans. R. Soc. London Ser. B* **375**, 20190042 (2020).
7. S. Engesser, S. W. Townsend, *WIREs Cogn. Sci.* **10**, e1493 (2019).
8. J. Goffinet et al., *eLife* **10**, e67855 (2021).
9. B. C. Roy et al., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 12663 (2015).
10. A. M. Tanimoto et al., *Anim. Behav.* **123**, 427 (2017).
11. P. Brakes et al., *Science* **363**, 1032 (2019).
12. A. D. Foote et al., *Nat. Commun.* **7**, 11693 (2016).
13. B. Williams et al., *Ecol. Indic.* **140**, 108986 (2022).
14. O. R. Wear, R. Freeman, D. M. P. Jacoby, *Nat. Mach. Intell.* **1**, 72 (2019).
15. M. Hagiwara et al., in *2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2023), pp. 1–5.

ACKNOWLEDGMENTS

The authors thank V. Baglione, D. Canestrari, M. Cusimano, F. Effenberger, D. Gruber, M. Hagiwara, B. Hoffman, M. Johnson, S. Keen, I. Kuksa, J. Lawton, J. Y. Liu, M. Miron, R. Patchett, L. Rendell, R. Seyfarth, P. Shaw, and additional colleagues from ESP and Project CETI for discussion, comments, and support. Conflicts of interest and funding sources are listed in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.adg7314

10.1126/science.adg7314

POLICY FORUM

Do we want less automation?

AI may provide a path to decrease inequality

By **Ajay Agrawal, Joshua S. Gans, Avi Goldfarb**

Impressive achievements made through artificial intelligence (AI) innovations in automating the tasks required in many jobs have reinforced concerns about labor market disruption and increased income inequality. This has motivated calls for change in the direction of AI innovation from being guided by task automation to instead focusing on labor augmentation (1). But task automation and labor augmentation are not polar opposites. Instead, automation of some tasks can lead to augmentation of labor elsewhere. Furthermore, AI automation may provide a path to reversing the trend of increasing income inequality by enabling disproportionate productivity improvements for lower-wage workers, allowing them to perform at levels that would previously require years of education and experience.

People have worried about automation—using machines to do the work that humans do—for centuries. Over that time, automation has proceeded apace, with entire sectors such as agriculture and manufacturing going from majority to minority shares of employment in many countries. Despite these sweeping changes, continual productivity improvements have not brought about technological unemployment. There has always been more for people to do, even if the fruits of economic growth have in recent times not lifted the incomes of all (2). Nonetheless, with the accelerated pace of innovations in AI technologies that specifically target the automation of cognitive rather than physical labor, many economists have become concerned that this time will be different, leading to substantial disruption and increased inequality but with little benefit to productivity and standard of living (3, 4).

The economic history of the internet and computing over the past 50 years suggests that worries about inequality are not unfounded. These were what economists call skill-biased technologies (5), placing increasing demands on the skill, education, and know-how of the workforce. As computers and the internet diffused in the

decades after 1980, demand for skills grew faster than supply, and inequality increased because of the disproportionate increase in wages for high-skilled workers.

More recently, consider all of the new products based on generative AI that automate the tasks of reading, writing, editing, summarizing, composing music, creating images, synthesizing speech, translating languages, programming computers, and producing videos. In each case, they promise time-saving productivity boosts by substituting capital (hardware or software) for the labor time that would otherwise be devoted to such tasks. If a machine can do these tasks, what will become of the people who did them previously? What if, instead of seeking to automate existing tasks, the mindset of the innovators was to provide tools that make existing workers more productive in their current jobs?

AUTOMATION AND AUGMENTATION

Automation and augmentation need not be opposites; the economic definitions of each are not presented as such. Economists studying automation consider a job to be a collection of tasks [as developed in (4, 6) and applied in (7–9)]. Automation occurs when a machine (capital) is substituted for labor performance of one or more tasks in a workflow. This will typically increase productivity, which determines long-term economic growth and the average standard of living. With technological innovation having long been recognized as a key source of productivity growth, there is close to a consensus among leading economists that AI-driven technological change is likely to increase productivity, economic growth, and the average standard of living (10). However, if workers are replaced by machines, then this improved productivity can come at a substantial cost to displaced workers.

If technology creates new tasks, then displaced workers can find new things to do (4). Augmentation can therefore be seen as the development of new tasks (3), particularly if those tasks complement existing human labor (1). There is a challenge, however, in identifying such tasks and determining which inventions are likely to lead to their development. The difficulty is that one person's automation is another's augmentation. Automating one task may create even more new tasks.

Rotman School of Management, University of Toronto, Toronto, ON, Canada. Email: avi.goldfarb@rotman.utoronto.ca



Using machine learning to decode animal communication

Christian Rutz, Michael Bronstein, Aza Raskin, Sonja C. Vernes, Katherine Zacarian, and Damin E. Blasi

Science, **381** (6654), .

DOI: 10.1126/science.adg7314

View the article online

<https://www.science.org/doi/10.1126/science.adg7314>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

Science (ISSN 1095-9203) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.
Copyright © 2023 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works