Animal behavior on auto

Researchers can't sit their laboratory mice or fruit flies down and ask them how they're feeling or why they're behaving in a particular way. Instead, humans are left to observe and interpret the various clues their animals provide. Can machines help?

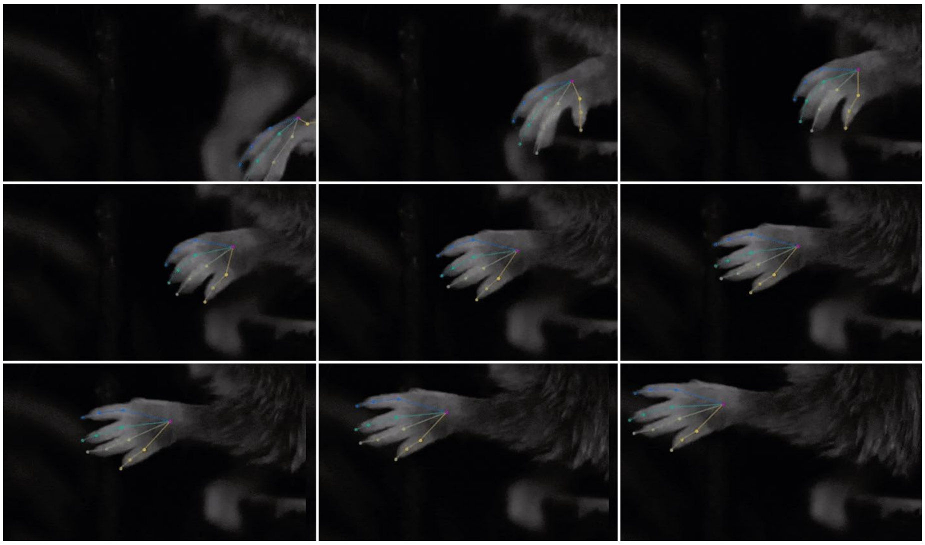
Ellen P. Neff

Charles Darwin once wrote that facial expressions are universal. Pain can be read on a human's face, and on the faces of many animals. That's the basis of the various grimace scales, pain scoring systems first applied to the lab mouse (a white CD1 outbred mouse, to be specific[1]). A mouse in pain will tighten its eyes, scrunch up its nose and cheeks, and flatten its ears and whiskers.

Perhaps researchers (often, undergraduates) make the same face sometimes when they're set to the task of scoring pain or the myriad of other visual tasks needed to classify what an animal is doing. Behavioral phenotyping has long relied heavily on manpower, whether to score a behavior, time it, or just to identify what it is to begin with from ever increasing hours of lab recordings.

Take the ultrasonic vocalizations of rodents, for example. These can be a reflection of an animal's affective state, says Kevin Coffey, a postdoctoral researcher at the University of Washington. Audio can be converted into sonograms—visual representations of frequency by time—from which a human can distinguish different types of ultrasonic calls. But data adds up quickly: one animal might call a few times in a three hour recording; another, thousands. "It was too much to do by hand," he says.

It's a common refrain, but the burden is starting to shift from man to machine. When Annalisa Scimemi, a neuroscientist at SUNY Albany in upstate New York, wanted to break down grooming behavior, she looked to machine learning. Grooming is a natural, stereotyped behavior in rodents that is often used as a measure of compulsive tendencies, but measuring it had been tedious, involving long stretches of a time with a stop watch in hand. Computer vision—that is, a machine's ability to 'see' and identify objects in a picture or video, was improving and computers were doing quite well at distinguishing shapes and colors, Scimemi says. Machine learning algorithms, meanwhile, could roughly follow animals around a cage or arena, but the features she needed to track were indistinguishable from body fur and capable of changing shape, thus tripping up the computers. So the lab added distinguishing features, painting mouse hands (after much training to make the process less stressful to the mice) with fluorescent colors that the computer could more easily see. The simple solution worked to their satisfaction to automatically track grooming behavior—they built a graphical user interface (GUI) for the software they dubbed M-Track[2] and published the open-source code for the tool on GitHub, an immense open-source code repository.



Reaching for it: Deep learning algorithms are expanding what's possible for keeping track of animal behavior. DeepLabCut, pictured in action with a reaching mouse hand, can estimate a variety of features across many different animals without a user having to annotate every frame in a video. Credit: M. Mathis, DeepLabCut

"It was, I think, a very creative idea at the time," says Scimemi. That time was 2016. In just a few short years since, there have been tremendous advances, she says. "Looking forward, it would be useful to have the same tracking accuracy, but without introducing these marker colors… I think that can be done now." Things are moving fast.

"The deep learning field has really pushed what is possible," says Mackenzie Mathis. Mathis, now a fellow at The Rowland Institute at Harvard, wanted to track individual mouse fingers in order to connect the brain to motor outputs. No matter how she tried to add little stickers to her animals' fingers or paint their nails, existing tools couldn't manage the right level of specificity. "At the beginning, there was always something that I had to sacrifice because I couldn't quite measure what I wanted to measure," she says. For a time, she instead turned to building robotic joysticks that the mice could be trained to pull—at least those could provide high enough temporal and spatial precision to allow her to follow where the hand was in space, she says. Nonetheless, she still found herself with terabytes of collected data.

Deep learning, a type of machine learning, brought DeepLabCut[3], neural network-based pose estimation software that's capable of distinguishing mouse digits, and then some. Coffey built a tool called DeepSqueak[4] to pore over all those sonograms, and Alexander Tuttle, a postdoctoral fellow at the University of North Carolina, has been applying deep learning to automate the work of grimace scale scoring[5] If you can train an undergraduate to do it, you can train a computer with enough examples, he says. New applications are emerging regularly.

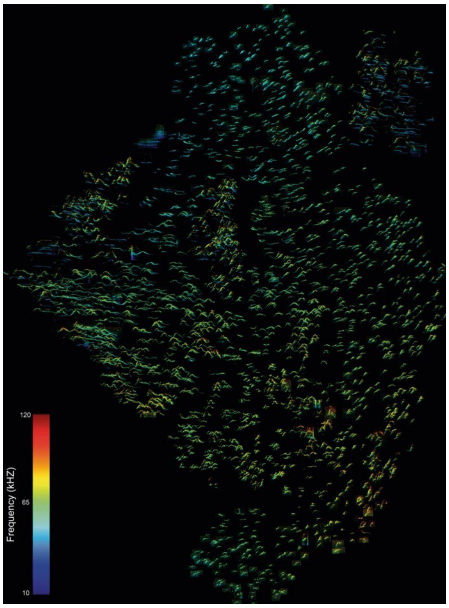
Algorithms are indeed becoming more sophisticated and networks are reaching deeper, letting computer scientists and biologists alike start to tackle their data processing problems in new ways. Even still—no pain, no gain?

Deep machine learning neural convolutional network…what?

The lingo can be tricky. In general terms, 'machine learning' is using statistics to find patterns in big data sets in order to accomplish a specific task, says Kristin Branson, a researcher at the Howard Hughes Medical Institute Janelia Research Campus. In supervised form, it is a process through which a computer algorithm takes input from a user and then implicitly learns criteria it can use to classify new data accodingly. A few years ago, Branson had found herself with flies to follow—thousands of them. Manually annotating videos was going to be too much to tackle, so she and her lab developed a tool they call JAABA, based on supervised machine learning[6] to do the labeling for them. They used to JAABA to annotate 20,000 fly videos and then link the labeled behaviors to responsible neurons in different parts of the brain—a feat that would have been nearly impossible by hand.

The premises of machine learning go back decades—the term was coined in 1959. Deep learning approaches have arrived in the past few years. Deep learning is a type of machine learning that goes a bit, well, deeper, taking higher dimensional data and passing it through a series of layers that processes the information and determines an output. Users designate training data and set weights that then guide the algorithm to the desired outcome. Deep learning algorithms are often based on artificial neural networks that parse connections in the data back and forth, akin to neurons in the brain "We're still very much short of what the brain can do, of course," says Mackenzie Mathis. "But in that sense, it's very much inspired by that."

Not unlike the workings of the brain, neural networks are a little bit of a black box still, says Francisco Romero Ferrero, a PhD student in Gonzalo de Polavieja's lab at the Champalimaud Foundation in Lisbon who helped build the idtracker.ai tool[7]. The underlying architectures of different neural networks can vary, but many deep learning algorithms used with images or video fall into the category of convolutional neural networks, an approach that takes advantage of the structure present in natural images, he says. A convolutional algorithm takes an input from one layer, performs some calculations, and then passes that information on to another level. The idtracker.ai tool for example, designed to keep track of multiple animals in a video, uses two convolutional networks: one sorts out which animal is which when they touch or cross while a second keeps track of individual identities.



Sorting out songs: Rodent ultrasonic vocalizations occur beyond human hearing, but tools like DeepSqueak can up the throughput of sorting through recordings. Credit: K. Coffey, DeepSqueak

From man to mouse, by way of machine

Applying such computer concepts in the first place might seem daunting, but researchers need not start from scratch. "There is so much energy put into these networks," says Coffey. "We just applied it." Prompted by his principle investigator John Neumaier's insistence that the lab automate the process of picking calls out from noise in their recordings, their lab technician Russell Marx started exploring existing options. They eventually chose a convolutional neural network called Faster R-CNN, which was designed for self-driving cars. In the automated automotive world, cars need to process what's in a large scene that's constantly changing; they do so through something called 'region proposal,' Coffey says. In DeepSqueak, the network is looking for the peaks that correspond to a vocalization on a sonogram rather than say, the sidewalk or another car. That's an advantage over previous software to define calls, which were often based on template matching and could be easily disrupted by background noise, he says. A separate classification network can then take on the work of separating the calls into different categories.

Tuttle, a behavioral neuroscientist by training, shares a similar story. With undergraduate Mark Molinaro, Tuttle dove in to GitHub until they found a deep learning network they could cobble together into what they needed to start automating grimace scale scoring. In the first iteration of the automatic Mouse Grimace Score (aMGS) tool, published just last year, they used Google's Inception V3 model to sort white mice into 'pain' and 'no pain' categories. Now working with collaborators who come from a more computer science-focused background, they've switched to the YOLO network—'You Only Look Once.' ("Computer scientists definitely have a sense of humor when they name their software," says Tuttle.) He presented preliminary data at the American Pain Society Meeting in April that suggests the revised YOLO-based network is up to about 80% accuracy for identifying pain in both white and (much more commonly used) black mice; with additional training and tweaking, they hope to match the 90% accuracy of their first paper by the summer. YOLO is also offering more nuance, sorting faces into no/low, moderate, and high pain read outs, he says.

DeepLabCut takes its inspiration from a human pose estimation tool called DeeperCut and is built on a ResNet (short for 'residual neural network'). Its developers also took advantage of a concept called transfer learning, pre-training the DeepLabCut neural network with a trove of existing images from the ImageNet database. "We take a network that's already learned another task," says Mackenzie Mathis "We're just refining it and saying, 'Okay, you already know what lots of things look like, but we just want you to find all the fingers." DeepLabCut can reach human labeling accuracy with only about 200 training images.

Under supervision

Learning algorithms don't run entirely on autopilot—training is an important aspect. "In supervised machine learning, you have a very specific task that you're trying to accomplish," says Branson. The user knows what the output should be and can, in theory, label all their data accordingly; to be a help, the computer must learn to do the same in a predictable way. But, "it's a basic assumption of machine learning that the distribution of your training data is exactly the same as the distribution of your test data. And if that's not the case, then machine learning could fail completely," says Branson. If an algorithm encounters a totally novel scene for example, it can struggle with how to label it. Training can be even more important with more complex networks. That said, "one thing that's remarkable about deep learning is it has a huge number of parameters, but it doesn't over-fit as much as we think it should," she says. "A current field of interest in machine learning is understanding why deep networks, in particular deep convolutional networks, work as well as they do."

To get an algorithm up to speed, users first define classifiers that describe a particular aspect of their data. They then manually label a set of training images. The learning algorithm then searches for a classifier function that can predict these labels from the input training image. If it makes the correct predictions, according to whatever accuracy threshold the investigator thinks is appropriate for their particular question, the tool can be ground truth-ed against a novel set of investigator-labeled images that it has not seen before. In JAABA for example, the user annotates what the animal is doing, such as a fly walking or jumping, and then the tool implicitly learns criteria that distinguish those behaviors. Users can always go back and add some more training rounds or, in the case of neural networks, continue to tweak the weights of the network before setting the tool to the task of labeling new data.

How much training is needed depends on the algorithm. Deep learning approaches intended for visual data of humans for example usually need very large training sets because humans look quite different from one another, says Talmo Pereira, a PhD student at Princeton University who has been working on the LEAP pose tracking tool[8] with Mala Murthy and Joshua Shaevitz' labs. Most lab animals are much more homogenous in both appearance and behavior, meaning a neural network can rise to par with a smaller training set. LEAP's algorithms were built from scratch but were inspired by 'stacked hourglass' networks, which are used in techniques to create heat maps from user inputs that can then predict where a particular feature is in each frame as its processed through the network. For best performance, some manual preprocessing is involved—the animals should be more or less centered within the video, Pereira says, but their network can be trained with as few as 100 user-labelled frames.

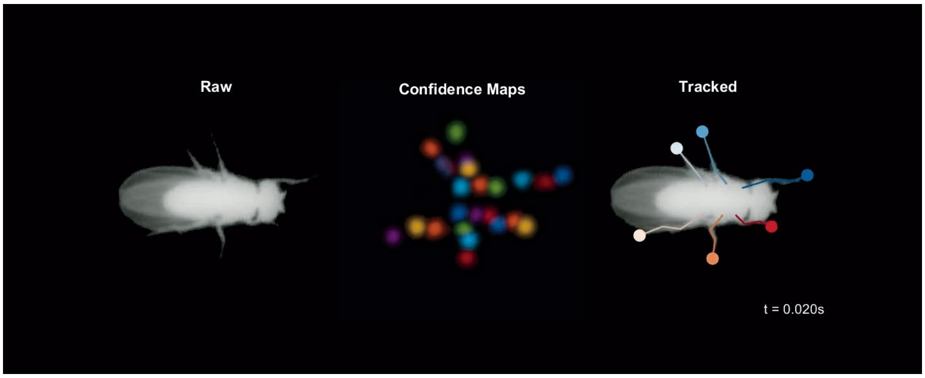
Up and running

Behind the scenes, there can be a lot going on—the current iteration of DeepLabCut, for example, consists of about 16,000 lines of code, says co-developer Alexander Mathis (he too found himself frustrated with existing tracking capabilities, wanting to follow just a mouse's nose along an odor trail). A user doesn't necessarily need to directly interact with all that though. Many of the developers of machine learning based-tools for the animal lab have had user-friendliness in mind and have spent considerable time on point-and-click graphical user interfaces (GUIs) that can make the underlying algorithms accessible to the less computer-savvy.

Building a good GUI can be challenging though in and of itself. "That was really the bigger thing than getting the neural networks to work," says Coffey. The first release of DeepSqueak worked beautifully in their hands, he says, but issues arose with other user's systems, their audio files, and the way that they record their animals. "The hardest part was definitely making it a…turnkey solution for anybody. It was easy to make it work for us," he says. Alexander Mathis recalls the time spent just making sure DeepLabCut would function across different operating systems. "That is actually quite a headache, to be frank," he says. "You just need a lot a testing, and you need to do the same kind of testing in all the different platforms."

The tools can be run from most any computer, but speed can become an issue without a machine that has graphic processing unit (GPU), the technological advance that has been a big contributor to the deep learning field, says Pereira. GPUs are a type of hardware that consists of multiple processing cores; individually, each core isn't as fast as that of a central processing unit, he says, but because power can be combined across the cores, the speed at which a network performs its calculations increases. "You could do it with basically any computer but without a GPU, it'll be very slow," says Coffey. GPUs vary in quality (and expense), but are becoming more common in research facilities and even laptops in recent years. "We haven't found anybody that…didn't have at least one machine in the lab that has some GPU in it," says Coffey. For those still without, moving the software to the cloud, such as through Amazon Web Servers or Google Cloud Services, is an option several tool developers are considering for their users.

For users that want to tinker themselves, the code underlying the various tools is often open online. "I think deep learning did so well because of the openness of the computing community," says Alexander Mathis. "People share their codes all the time." GUIs can help eliminate (or at least minimize) the learning curve, but those with more computer expertise can modify the code to suit their own needs and also help troubleshoot issues to improve the tools further.



Following flies: An illustration of how LEAP figures out limb locations in a frame. Credit: T. Peirera, LEAP

Algorithms in action

Romero Ferrero and his colleagues in de Polavieja's lab have been working on the problem of tracking the individual identities of multiple animals in a group. One individual is easy to follow around at this point, but when multiple animals touch or occlude one another, computers can get mixed up. An early version (early being 2014) of the idTracker software used 'shallow' machine algorithms to identify the visual footprint of an individual animal against a consistent background and then keep track of it as it moved amongst other conspecifics. It worked with about 5 to 10 animals at a time. But animal behavior can change as the size of the group grows—deep learning let the lab increase the scale. The deep learning iteration, idtracker.ai, can keep track of up to 100 individuals (they stopped at 100, Romero Ferrero says, because for fishes, the tanks were getting small; for flies, it was a challenge just to get that many into an arena in the first place). Users input the video and parameters to separate the animals from the background and the algorithm automatically selects the training data and identifies the animals through the video. Much information can be gained by studying the trajectories of animals in a group—the de Polavieja lab, for example, studies concepts like aggression and transfer learning, in which animals with experience at performing a task are mixed with novices.

But (most) animals are much more than a center of mass facing a particular direction (the basis by which many tracking algorithms follow an animal's location over time)—there are limbs and fingers, wings and tails that the animals use to communicate and interact with their environment that can also reveal interesting information. For Pereira, the idea of applying deep learning to his research was to 'take a modern spin on the study of animal behavior.' More specifically, to connect the brain to its outputs via natural behaviors rather than artificial assays, such as T-mazes or lever pressing, that were designed to control for variability but that don't necessarily reflect an animals' natural behavior and instincts. Ten years ago, 'state of the art' for measuring natural behavior was hand scoring, he says, an approach that goes back decades but that can be somewhat arbitrary in nature and quickly become complicated when the 'behavior' in question isn't clear cut. The more accurately and objectively that you can measure a behavior, the better your understanding can be of neural activity and the computations the brain itself is performing, he says. In particular, they are interested in fly courtship and will be using LEAP to help tease apart the cues in fly wings during the ritual, with the intent to then manipulate the brain and see how behavior changes as a result.

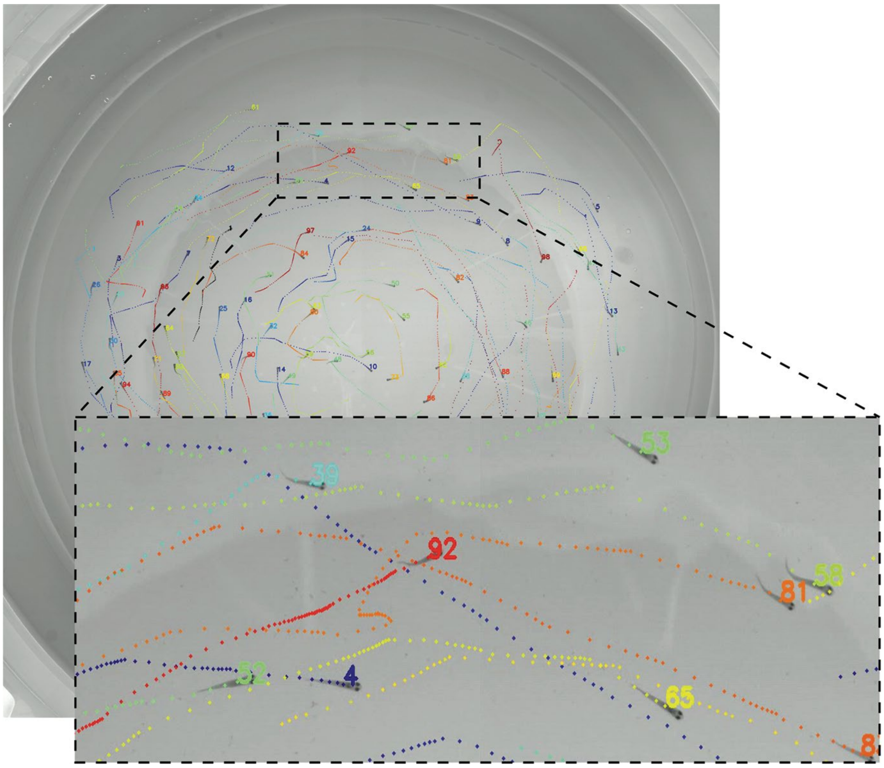
"I'm really happy to see is that there are a lot more neuroscientists caring about behavior," says Branson. In the past, behavior was often condensed down into one number, but people are realizing that's an oversimplification, she says. "It's really nice to see…all of these tools being developed, and people using them."

DeepSqueak and the aMGS too will be put to use for scientific questions in their developers' labs—the former, as a means to measure neurological conditions related to the serotonin system as well as drugs of abuse; the latter, for understanding the mechanisms of pain and how to treat it. Both might also have welfare applications. Coffey suggest DeepSqueak might be helpful for monitoring distress in rodent colonies—humans can't actually hear USVs, but machines can. Managing and treating pain in experimentally manipulated animals is a priority across disciplines.

Emerging tools need not only apply to new studies. Prior recordings can often be re-analyzed, as could old data languishing on forgotten hard drives. The Mathis' mention collaborators at a nonhuman primate lab who in the past would spend months training their animals to wear stickers on their arms; they are now returning to that data to follow the limb in its entirety, not just a few select points, such as the wrist. "It's a beautiful re-use of data," says Mackenzie Mathis. Nor will some of these tools just to the 'traditional' lab species, either. Coffey has seen researchers apply DeepSqueak to animals like lemurs and dolphins. The pose estimators abound with examples applied to bees, giraffes, cheetahs, horses… "There's no stipulations on what type of animals it can track," Mackenzie Mathis says of DeepLabCut. "As long as you can see it, you can track it."

Where to next?

Machine learning has arrived in the animal lab, and it's changing the way researchers handle the drudgery of data processing. For example, a six-hour cocaine self-administration session that would 60 hours of manual processing can be now completed in about 15 minutes, says Coffey. But as demands on manual labor diminish, new questions are arising about all that neatly labeled and categorized data. "What do you do with it?" asks Coffey. "How do you relate it back to behavior?" Not that that's necessarily a bad problem to have. "For me, there's always this hope: maybe, one day, one can really spend 80% of the time doing science, and not 95% of the time on extracting the data that one needs from an experiment," says Alexander Mathis. The time spent building the tools (and helping new users learn them) has been time well spent, but the developers all seem ready to turn their algorithms to their own research questions in earnest.



Who's who? Deep learning takes animal tracking to more animals. Idtracker.ai, pictured with zebrafish, can follow up to 100 animals at a time without losing track of identity.
Credit: F. Romero Ferrero, idtracker.ai

As researchers work through algorithms and build the GUIs on top of them, the rapid pace of machine learning highlights the importance of interdisciplinary learning and lab work. “I think the emergence of these fields should probably ring the bell that the boundaries between disciplines are much more fluid than probably they were 20 years ago,” says Annalisa Scimemi at SUNY Albany. “It’s a good time to be creative, but also it’s a good time to break boundaries between disciplines.” It’s important that biologists and computer scientists come together, she says; the former might not know how to code, while the latter might lack the biological context to develop the right tools for the biological question at hand. Branson makes sure to have both in her lab, such as her JAABA co-developer Alice Robie. “I’m a computer scientist, who knows a little bit about biology. Alice is a biologist who knows about computer science, and so we can talk to each other. But I think we also help each other translate the two fields that we are specialized in,” says Branson. That can make it easier to make the most appropriate tweaks to the data collection in order to make the computer analysis easier on the tail end, says Robie.

There are a lot of labs working on machine learning applications for the animal lab—this article provides just a few examples—and many are endeavoring towards similar goals. “All of these researchers right now are building sort of their own neural networks, but we also might be tracking really similar behaviors,” says Mackenzie Mathis. Given how open the community can be, she suggests that continuing to share networks in the future could help standardize efforts across labs, an important element to improving reproducibility of results. “I think that’ll be a really useful thing, not only for…speed of training, but for consistency,” she says. “People will be really analyzing data in a similar way.”

With all the effort out there, especially as it becomes easier and easier for biologists to apply computer concepts to their work, a word of caution is merited as interested users look for the ‘best’ tools. “We’re still in the Wild West,” says Tuttle. “I think that we have to be very careful in what we use and sort of decide on the most robust models.” The tools are built by humans and like anything built by humans, there is the potential for bias, he says. One lab’s code to approach a particular problem will differ from another’s and it will take time and lots of testing from multiple labs to sort through any hidden flaws in the underlying algorithms. Bias, however, may be just as present across the different humans who have long borne the brunt of manual phenotyping in the past. “If we can get a computer to be as accurate as a human being in doing what an undergraduate can, then it’s going to be good for us,” Tuttle says.

The larger field of machine learning is pacing ever onward. “Deep learning is really changing the way that people will do science,” says Mackenzie Mathis. “But it’s also hard to predict because it moves so fast.” ❐

Ellen P. Neff
Lab Animal, New York, NY, USA.
e-mail: ellen.neff@us.nature.com

Published online: 13 May 2019
https://doi.org/10.1038/s41684-019-0312-z

References

1. Langford, D. J. et al. *Nat Methods* **7**, 447–449 (2010).
2. Reeves, S. L., Fleming, K. E., Zhang, Z. & Scimemi, A. *PLoS Comput Biol* **12**, e1005115 (2016).
3. Mathis, A. et al. *Nat Neurosci* **21**, 1281–1289 (2018).
4. Tuttle, A. H. et al. *Mol Pain* **14**, 1744806918763658 (2018).
5. Coffey, K. R., Marx, R. G. & Neumaier, J. F. *Neuropsychopharmacology* **44**, 859–868 (2019).
6. Kabra, M., Robie, A. A., Rivera-Alba, M., Branson, S. & Branson, K. *Nat Methods* **10**, 64–67 (2013).
7. Romero-Ferrero, F., Bergomi, M. G., Hinz, R. C., Heras, F. J. H. & de Polavieja, G. G. *Nat Methods* **16**, 179–182 (2019).
8. Pereira, T. D. et al. *Nat Methods* **16**, 117–125 (2019).