
AI-

ASTRONOMY'S
-ASSISTED
FUTURE

The next generation of astronomy depends on astronomers and artificial intelligences working together.

BY ASHLEY SPINDLER

Throughout the history of astronomy, the development of new tools and techniques has always aimed at allowing us to do more — observe more, analyze more, explore more. But as the next generation of astronomy becomes the current generation of astronomy, it has become increasingly clear that we might have bitten off more than we can chew.

The problem is data: mountains of it, more data than you can imagine. Our telescopes have become so powerful, our detectors so sophisticated, and our computers so complex that it is simply impossible to analyze all the data they generate and collect.

Not without help, that is. To solve astronomy's big data challenges, teams of researchers around the world are turning to machine learning for answers.

When I started my Ph.D. eight years ago, "next-generation astronomy" referred to the time after the James Webb Space Telescope (JWST) would launch, the Vera C. Rubin Observatory would start taking nightly images of the sky, and the Euclid Space Telescope would begin peering into the depths of the universe. One of those things has already happened, and the other two aren't far off. In that span of time, the number of articles posted to NASA's Astrophysics Data System referencing machine learning has increased sixfold, and something that was once a curiosity is becoming a core part of the astronomer's toolkit.

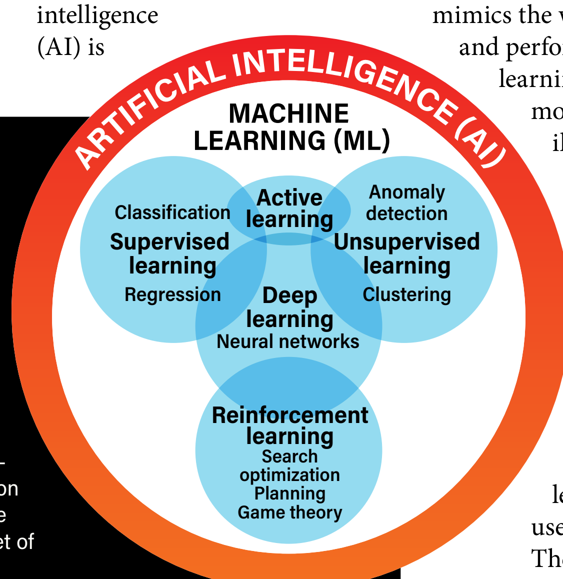
What remains to be seen is just how much machine learning can actually do for astronomers — and perhaps, more importantly, what it can't do.

AI'S APPETITE FOR DATA

Chances are you've heard the terms *artificial intelligence* and *machine learning* thrown around recently, and while they are often used together, they actually refer to different things. Artificial intelligence (AI) is a term used to describe any kind of computational behavior that mimics the way humans think and perform tasks. Machine learning (ML) is a little more specific: It's a family of technologies that learn to make predictions and decisions based on vast quantities of historical data. Crucially, ML creates models that exhibit behavior that is not preprogrammed, but is learned from the data used to train it.

The facial recognition in your smartphone, the spam filter in your emails, and the ability of digital assistants like Siri or Alexa to understand speech are all examples of ML being used in the real world. Many ML technologies are now being used by astronomers to investigate the mysteries of space and time. Astronomy and ML are a match

PREVIOUS PAGES: Artificial intelligence is crucial to helping astronomers make sense of the flood of data that is coming their way from next-generation surveys. *ASTRONOMY*: ROEN KELLY. ILLUSTRATION BELOW: *ASTRONOMY*: ROEN KELLY



AI LINGO

MACHINE LEARNING (ML) is just one technique within the field of artificial intelligence, and within ML, there are many different approaches. For instance, deep learning algorithms process data using neural networks, which are loosely inspired by the human mind. Neural networks can be combined with several other frameworks to improve their performance.

In supervised learning, an algorithm learns by testing itself on a training dataset where the answers — like the type of galaxy classification — are already known. By contrast, in unsupervised learning, an algorithm is left to discover patterns in data on its own, like clusters of stars or anomalies in light curves. Active learning is a form of semi-supervised learning in which a subset of the data is classified by humans to help the algorithm learn.

In reinforcement learning, an algorithm learns behaviors by receiving rewards when it performs as desired. This allows it to steadily improve at tasks like driving a car within the boundaries of a traffic lane, playing the board game Go, or writing human-like sentences as a chatbot. In astronomy, reinforcement learning has been used in telescope operations to find the most efficient observing schedules and improve the performance of adaptive optics systems.

Neural networks can even be pitted against one another with one algorithm acting as a sparring partner, attempting to find flaws in the main algorithm's output. This setup is called a generative adversarial network and can produce realistic-looking astroimages — or augment existing images with realistic details. —*Mark Zastrow*

The Vera C. Rubin Observatory's survey of the southern sky will produce about 20 terabytes of raw data every night. NOIRLAB/NSF/AURA

made in the heavens because the one thing astronomers have too much of — data — is the thing that ML models can't get enough of.

We're familiar with gigabytes and terabytes of storage, but data at that scale is old news in astronomy. These days, we're interested in petabytes: 1,000 TB, or 1 million GB. It would take just 10 PB of storage to hold every single feature-length movie ever made, at 4K resolution; it would take over 100 years to watch them all.

The Vera C. Rubin Observatory, currently under construction in Chile, will be tasked with mapping the entire night sky in unprecedented detail, every single night. Over 10 years, Vera Rubin will produce about 60 PB of raw data, studying everything from asteroids in our solar system to galaxies in the distant universe.

No human being could ever hope to analyze all that data, and that's from just one of the next-generation observatories being built. So the race is on among astronomers in every field to find new ways to leverage the power of AI.

TAMING THE ZOO

One such astronomer is Mike Walmsley, a postdoctoral researcher at the University of Manchester in the U.K. and one of the leading AI researchers in astronomy. Walmsley is the brains behind integrating ML systems into Galaxy Zoo, a citizen science project that has classified the morphologies of over a million galaxies. In its original form, Galaxy Zoo's citizen scientists tackled the mammoth task of visually inspecting galaxies from the Sloan Digital Sky Survey, which were so numerous that professional astronomers could not complete the task alone. But the increased scale of modern sky surveys has outpaced the rate at which Galaxy Zoo volunteers can classify galaxies. The project's latest dataset includes 8.7 million galaxies. "It would have taken about 200 years for Galaxy Zoo volunteers to measure these alone," says Walmsley.

To solve this massive data problem, Walmsley enlisted machine learning to pick up the slack. He developed Zoobot, an AI model that is about as accurate at classifying galaxies as asking 15 people.

"Adding AI is like giving volunteers power tools," he says. "Where one person alone might classify a few hundred galaxies, our AI can learn from them and together classify millions more."

What's unique about Walmsley's work is that ML hasn't been used

Machine learning (ML) is a family of technologies that learn to make predictions and decisions based on vast quantities of historical data.

The Square Kilometre Array is envisioned to eventually consist of some 2,000 radio dishes across Africa (pictured here as a photoillustration) and up to 1 million antennas in Australia. Even in its smaller initial form (which is currently under construction) of roughly one-tenth that size, it will generate up to 1 terabyte of data per second. SPDO/TDP/DRAO/SWINBURNE ASTRONOMY PRODUCTIONS/CC BY 3.0

to entirely replace the role of professional and amateur astronomers, but rather to work together with human classifiers. The AI model used by Galaxy Zoo utilizes a concept called active learning, where the model is able to send images that it isn't certain about back to the citizen scientists to provide more information about what kind of galaxy is being inspected. By using this method, Walmsley and the Galaxy Zoo team were able to dramatically reduce the time it takes to classify hundreds of thousands of galaxies.

In fact, keeping human intelligence in the loop is important for the future of astronomy research. "Unlike our volunteers, Zoobot only works well when classifying galaxies similar to those it has seen before," says Walmsley. "And it lacks that uniquely human skill of noticing when something looks a little bit strange."

PLANET HUNTERS

There are many ways to look for the signals of exoplanets, but the most prolific methods with current technology involve studying the variation of a star's brightness over time. If a star's light curve shows a characteristic dimming, it could be a sign of a planet transiting in front of the host star. Conversely, a phenomenon called gravitational microlensing can cause a large spike in a star's brightness, caused by the exoplanet's gravity acting as a lens that magnifies a more distant star along the line of sight. Detecting these dips and spikes means sifting through thousands, or even millions, of light curves studiously collected by space telescopes like NASA's Kepler telescope and the Transiting Exoplanet Survey Satellite.

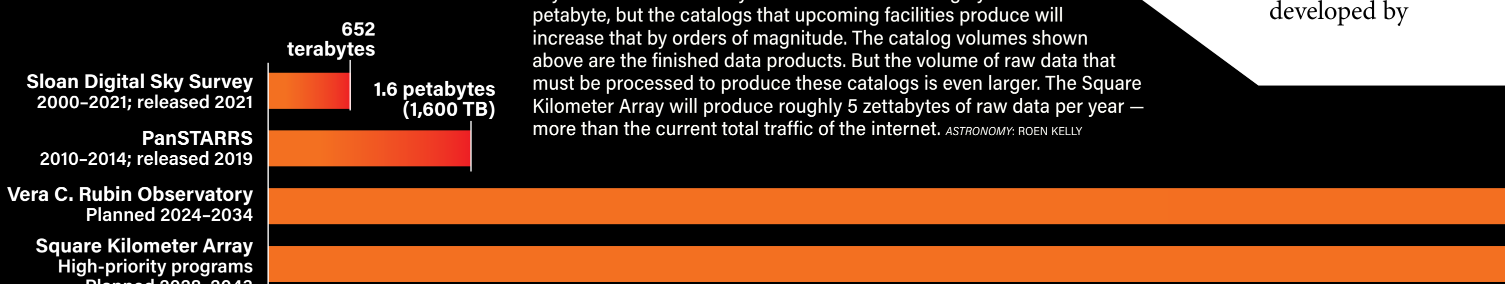
Using the huge libraries of observed light curves, astronomers have been able to develop ML-based models that can outperform humans in identifying possible exoplanets. But AI can do much more, such as help us identify which planets might be habitable. With next-generation observatories, such as the Nancy Grace Roman Telescope and JWST, astronomers hope to use ML algorithms to detect water, ice, and snow on rocky planets.

AI can even reveal new fundamental insights into mathematics and astronomy. In a paper published last May in *Nature Astronomy*, a team of researchers reported that ML algorithms helped them discover a more elegant understanding of exoplanet microlensing, unifying multiple interpretations of how an exoplanet's configuration with its host star can vary. The report came just months after researchers at DeepMind in the U.K. reported in *Nature* new AI-aided fundamental insights into mathematics.

GALACTIC FORGERIES

While many ML models are trained to distinguish between different types of data, others are intended to produce new data. These generative models are a subset of AI techniques that create artificial data products, such as images, based on some underlying understanding of the data used to train it. The series of DALL-E models developed by

DATA DELUGE



The amount of data from astronomical surveys is set to skyrocket. Current surveys are on the scale of roughly a petabyte, but the catalogs that upcoming facilities produce will increase that by orders of magnitude. The catalog volumes shown above are the finished data products. But the volume of raw data that must be processed to produce these catalogs is even larger. The Square Kilometer Array will produce roughly 5 zettabytes of raw data per year — more than the current total traffic of the internet. *ASTRONOMY*: ROEN KELLY

M. J. SMITH ET AL. (U. HERTFORDSHIRE), DOI:10.1093/MNRAS/STAC130

Generative algorithms are so good that even professional astronomers can struggle to distinguish between the real and the fake. This mosaic features dozens of synthetically generated images of objects in the night sky — and just one real image. **Can you spot it?** *The answer is on the next page.*

the research company OpenAI have pushed this concept into the public eye. These models generate an image from any written prompt and have set the internet alight with their uncanny ability to construct images of, for instance, Garfield inserted into episodes of *Seinfeld*.

You might think that astronomers would be wary of any kind of fake imagery, but in recent years, researchers have turned to generative models in order to create galactic forgeries. A paper published last January in *Monthly Notices of the Royal Astronomical Society* describes using the method to produce incredibly detailed images of fake galaxies, which can be used to test predictions from enormous simulations of the universe. They can also help develop and refine the data-processing pipelines for next-generation surveys.

NEXT-GENERATION ASTRONOMERS

In the summer of 2022, Walmsley, a few other astronomers, and I organized a special session on AI in astronomy at the Royal Astronomical Society's National Astronomy Meeting, held at the University of Warwick. Among all the amazing science being presented, perhaps the most exciting thing for me wasn't the research itself, but the people doing it.

OUR FAKE COSMOS: *ANSWER*

NASA, ESA, J. HESTER AND A. LOLL (ARIZONA STATE UNIVERSITY)

This image of the Crab Nebula (M1) is the only real image in the mosaic on page 41.

Overwhelmingly, ML research in astrophysics and astronomy is being driven by early-career researchers, particularly doctoral students, who are bringing new, unique perspectives to the field.

For instance, Emily Hunt is a Ph.D. student at the University of Heidelberg in Germany and works with data from the European Space Agency's Gaia satellite. Gaia observes the stars in our own galaxy and beyond, and its catalog contains precise positions for over 1 billion stars. With data of this scale, using AI isn't just a choice for astronomers, it's a necessity.

"Searching through Gaia data by hand to look for open clusters would be like looking for thousands of needles in a galaxy-sized haystack," says Hunt. "Put simply, our science is not only greatly improved with ML, it would be pretty much impossible without ML."

Unlike a lot of ML research, Hunt's work doesn't rely on deep neural networks, the workhorses of AI whose function is inspired by the human mind.

Instead, Hunt has explored the effectiveness of using different kinds of clustering models. As the name suggests, this family of algorithms identifies groups of nearby points in a dataset — for example, clusters of stars in a catalog. According to Hunt, with this method it "takes seconds to find a cluster that a human might need hours to find." Using ML, Hunt is hoping to publish the largest-ever catalog of open star clusters — just one example of how this next generation of astronomers will revolutionize the field.

SEARCHING FOR SERENDIPITY

AI is also primed to make discoveries that we cannot predict. There's a long history of discoveries in astronomy that happened because someone was in the right place at the right time. Uranus was discovered by chance when

Galaxy Zoo's AI model was able to classify the galaxies at right as ellipticals. The galaxies at left are examples that the model was unsure of how to classify. In active learning, such images are sent back to humans, who can provide the model with additional feedback and training. MIKE WALMSLEY

William Herschel was scanning the night sky for faint stars. Vesto Slipher measured the speed of spiral arms in what he thought were protoplanetary disks, eventually leading to the discovery of the expanding universe. And Jocelyn Bell Burnell's famous detection of pulsars happened while she was analyzing measurements of quasars.

But the next generation of astronomy, with its petabytes of raw data, poses a significant barrier to the possibility of serendipitous discovery. With so much data on hand — and limited resources to follow up every potential discovery — how might we find the weird and unexplained phenomena that we don't even know we're looking for?

AI could hold the answer with a field of techniques called anomaly detection. These algorithms are specifically trained to sift through mountains of images, light curves, and spectra, looking for samples that don't look like anything we've seen before. One example could be so-called jackpot gravitational lenses, a rare alignment of galaxies resulting in two or more magnified images around a single nearby galaxy.

Perhaps soon, an AI could join the ranks of astronomy's greatest discoverers.

THE FUTURE OF ASTRONOMY AND AI

You are probably wondering at this point: Will AI put astronomers out of a job? Probably not, though there's no doubt that the way we do our jobs has already changed, as AI and ML are quickly becoming core tools for astronomers.

But astronomy also has a lot to offer researchers working on the cutting edge of machine intelligence. From studying the evolution of galaxies, to hunting for alien worlds, and even tracing the origins of martian meteorites, astronomy offers vast quantities of research data from a variety of sources that most fields can only dream about.

This is one of the most exciting parts of this intersection of fields — both AI researchers and astronomers can push each other forward. And in the next generation of astronomy, with its petabytes of raw data from facilities like the Vera Rubin and JWST, we can't possibly imagine what these algorithms might find.

Ashley Spindler *(she/they) is a lecturer in astronomy and data science at the University of Hertfordshire in the U.K. She spends her time studying the evolution of barred galaxies with the help of AI.*

HOW A NEURAL NETWORK WORKS

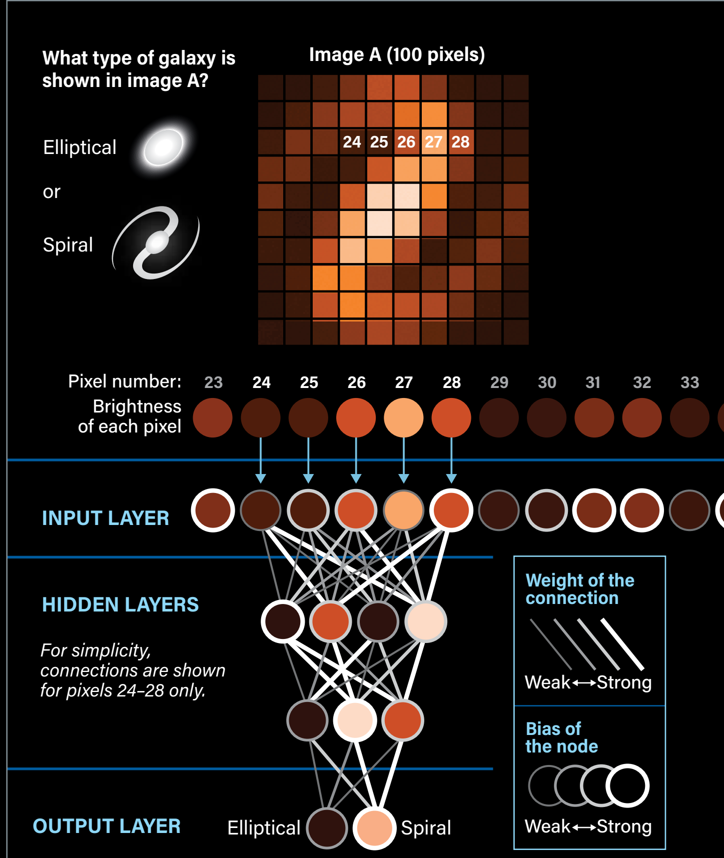
A NEURAL NETWORK is a sequence of computations arranged like a network of neurons, where values are stored and manipulated as they propagate through the network.

For example, take a simple neural network designed to classify a galaxy in a 8x8-pixel monochrome image as either spiral or elliptical. The brightness value of each pixel is fed to a node in the input layer, 91 in all in our case. If that value exceeds a certain threshold, that neuron "fires" and feeds its value to neurons in the next layer, the first of multiple so-called hidden layers.

Each neuron in that layer performs a computation on the values it is fed: First, the values are multiplied by a number specific to the connection from which they came, called a weight. Then they are summed and added to another number specific to that node, called a bias. Weights effectively act as a measure of the strength of that connection in the network, and the biases indicate how sensitive the neuron is to firing.

The strength of a neuron's signal is determined by the weights, biases, and the signal received — modified by a mathematical function called an activation function — which is then sent to the neurons in the next hidden layer. This process repeats, triggering patterns of neurons, until the values arrive at the final layer, the output layer. The output neurons are like options on a multiple-choice question: one for an elliptical galaxy and one for a spiral galaxy. The neuron with the highest value is the network's choice for that image.

In the beginning, the weights and biases for each connection and neuron are set to random values, and the network's choice is no better than a random guess. To train the network, the actual galaxy type — elliptical or spiral, as determined by a human — is propagated backward through the network, and the weights and biases are adjusted to improve the algorithm's performance. This process can be repeated thousands or millions of times. In more complex neural networks, additional mathematical operations can be performed at the nodes of each layer; this may allow the network to learn to detect edges or textures in the image. The result is a network of nodes with weights and biases tuned to act on fresh input data and make the decision it was intended to make. — *Mark Zastrow*



The strengths (or weights) of the connections between neurons and the sensitivity (bias) of the neurons themselves are depicted here with varying line widths. Note that only some connections and neurons are shown, for legibility. *ASTRONOMY*: ROEN KELLY