Every prediction of the future has artificial intelligence. But although computers have beaten humans at chess, Go, <translation>, and more, we still claim that artificial intelligence is something that does not yet exist. Instead, what we have is “machine learning.”

Computers are tools for following rules very quickly. Typically, humans create these rules. Machine learning is where the computers figure out the rules for themselves.

To see how this works, you will first teach a computer how to sort these colored balls, based on the brightness picked up by a camera.

<after hitting some threshold>

This was not too hard, was it? Let’s try something a little more difficult. Now try to teach the computer how to tell whether <a rock is granite. Granite is dark and small.>

<after hitting some threshold>

This is still doable. But real world tasks are far more difficult. Let’s say you’re a real estate company, and you want to find out which houses are worth trying to sell. Can you teach the computer how to figure out which ones are valuable based on the quality of the schools, the number of bedrooms, the number of bathrooms, the availability of public transit, the walkability, the condition, and the price of the neighborhood?

<after enough tries, “give up” button appears>

<automatic training starts>

That was a little too hard, right? Although we know these are all factors, we don’t fully understand how they combine. But computers are very good at doing things fast. This is where machine learning comes in. If we show the computer hundreds of examples, we can adjust the knobs every time it gets one wrong, based on how much that knob contributed to that mistake. This will automatically correct for errors over and over until the predictions are extremely good.

<wait for training to complete, correctness to be shown>

But what exactly did the computer teach itself? Since we didn’t tell it what to do, in order to understand why it does any particular action we have to guess based on what it does to similar inputs. For example, let’s take a look at the houses two different neighborhoods with the same median input.

<button to run through neighborhood 1 and neighborhood 2>

What’s the difference between these two neighborhoods? Neighborhood 1 is historically black and neighborhood 2 is historically white. Did we just create a racist AI?

Apparently, but we didn’t include race anywhere. How did this happen?

The answer is that it’s hidden in our “correct” answers that the computer trained itself from. Because those were based on decisions made by people, who have society’s ingrained biases, the bias is hidden in the machine learning model that the computer created. The algorithm is not neutral.

How do we fix this problem? If you can figure that out, please let us know. But just know that machine learning models are used in situations including determining sentencing of inmates, computing housing prices (as we showed), and determining what ads you’re shown, including political ads. For more information on this, <link>.

This model is much simpler than any used in real life. Generally, hundreds of tiny models like this are combined in ways that let them learn more and more complicated things about the data, which are combined to be able to take things into account in relatively smart ways. To learn more about this technique, called deep neural networks, read <this>.

Machine learning can also be used to create things, such as images (DeepDream, that thing with fake porn, FaceApp, style transfer) or text (Google Translate). To learn about that, read <this>.