

## How you might feel when AI can help

Hi. Let's get started. So part of using machine learning is recognizing when it makes sense to use machine learning in your project. I had a little bit of a reading about that in the reading list. And I'd like to elaborate on that right now just to walk you through it, so you know when to recognize those instances when machine learning might be able to help you.

The feelings you feel when AI might help you are probably not the feelings you're having of joy and happiness. It's usually when you're faced with maybe a big document dump or a whole pile full of data, and you just don't think you can get through it all. Specifically, it may be those moments when you feel like you will never be able to sort all of these documents, or images, or we'll never be able to watch all of this video. Maybe you get some dash cam video. And it's thousands of hours, and you can't watch it all. Or we'll never be able to read all these documents. Maybe you're looking for the names in a document or all the companies in a document or a document dump. And you just don't have time to go look for every single company in there. Or those situations where you want to find more like something. You have something that you're looking for, and you want to find more like that in your data, or your images, or your videos.

OK. Well, let's start out with: We'll never be able to sort all these documents or images. Here's an example. We have these maps at courts where we have flight paths of helicopters. The one on the left is not as interesting as the one on the right. Right? Because it's circling, and we get these images all day for days and days on end. And we wanted to see if a computer could detect the interesting pattern, which is the circular pattern. Right? Because the circular pattern shows something is up. In this case, it was a big fire. So if we could detect that, maybe that can be an alert that something's going on.

It turns out this is doable, and you don't need a whole lot of images to do it. In fact, with just about 300 images, I was able to train a model to do this. And here's how. First I sorted these images into a couple of different folders. Zero for not circling, and one for circling. And then I took a model that already existed. This is a model that's been trained on 14 million images, already. And I just additionally trained my almost 300 images, and I was able to blend that through a principle called transfer learning. And in the result, is a model which I like to call an artisanal model, which is just for me. It's just to solve my particular problem. This is the kind of thing that you can do, too. And we're actually going to walk through exactly this example later on in this class.

So in the end, you can take a pile of images, have this artisanal model, and sort them either as they come in or if they're just a huge folder full of them on your harddrive. In a similar way, we can take tweets and sort them into something like whether or not they're fact checkable or not fact checkable. You know, there are thousands and thousands of tweets out there, and we're actually going to look at a particular set of these that were around the Texas legislature. That's the TXLedge hashtag, and we we're going to actually make this model later, too. And you can see how you can have it sort documents in this case, very little documents, into two different kinds of categories. Categories that you would think in a way only a human could detect. Right? Whether or not something is fact checkable. Turns out we can teach a computer to do that, too.

Then we have this situation where we'll never be able to watch all of this video. Hours and hours of video footage, you know, how might you do that? Well, it turns out it's not that dissimilar from what we just saw because videos are really just a collection of images, so if we split videos into images, we can then build artisanal models to look at all those images. We're actually going to make something where we can detect bicycles inside a video. You might not think that's pretty newsworthy, and you're probably right. But maybe there are other things you could imagine where you're trying to find a particular scene or situation in a video, and you want to train a computer to find it so that you don't have to look through thousands of hours of video in order to find those particular instances.

Another situation is you'll never be able to read all these documents, so then you have the situation where you have a pile of documents. So if you know the person you're looking for, that's not too hard, right? You can search for, or control F, or do a find, for a particular person's name. But what if you don't know all the people who might be mentioned in that pile? What you can do is you can use machine learning to say, "Hey, what are all the names in this pile? Who are all the people in this pile?" And then it will tell you, "Oh, Andrew Cuomo is in there, and Bill de Blasio is in there, too, but so is Gwyneth Paltrow."

And finally, the situation where you want to find more like this. This is a little bit more complicated. But you can imagine in the same way, if you wanted to find tax returns, you could have a little search engine on your documents and look for the words "tax return," and you'll get all the tax returns. But what if they don't say tax return? What if they just look like a tax return? It's actually possible to make a model that can look for things that look like tax returns, even if they don't exactly say "tax return" on it.

This particular example is a little bit more sophisticated, and I'll show you where to find it. So once you're done with this class, this is the kind of thing you might be able to explore. Finally, I want you to think of machine learning as basically 20 enthusiastic interns. They're here to help you, and they want to do their best with what you've told them. And they're going to be able to do it in less time than you would be able to do it all by yourself. But remember, these interns, these machine learning models don't have any journalism experience. In fact, they have no real life experience at all. All they know is what we're going to teach them, or what they've been taught by others. As such, they are never perfect.

We have to make sure that just because we're using a computer and it seems to be smart, you have to remember that it's not perfect. It won't find some of the things you're looking for, and it'll find things that you're not looking for. Always do your journalistic due diligence to check other sources, and inspect the information, and do your own reporting.

Next up, we'll take a look at a free, powerful computer that you can use to do your machine learning on. It's called Google Collab Notebooks, and we'll learn all about them and how to use them.