[Mary] Alright, hi everyone! My name is Mary Ton. I am the Digital Humanities Librarian and I'm delighted that you could join us in-person, online, and in the future.

Today I'm joined by Rebecca Stover and Cadence Cordell as we dive into the world of deepfakes, how computers see and mimic us.

So the first thing that we're going to try is to spot a deepfake. So, before that, we have today's slides. So today's slides are at go.illinois.edu/deepfakes and that's all one word. All lowercase. The link is case sensitive and anytime you see yellow underlined text in the PowerPoint, it means that there's a link there.

So, let's spot a deepfake. What you're about to see is a video that was created in collaboration between a professor of business, Robert Brunner, and the Center for Innovative Teaching and Learning. . . Center for - yes, Center for Innovation in Teaching and Learning, CITL, using a tool called HeyGen.

So what we're going to do is, I'm going to play a short video clip that has the real Robert Brunner and an AI avatar of Robert Brunner. And we're going to have a poll after this to see which one you think is the real one, and which one is the deep fake. So here's the video.

[music from video plays]

[Mary] And for the record, there is no sound, other than some music. So this is a visual test primarily.

[Music continues]

[Mary] So that was the video. Now it's time for the quiz. So, use your cell phones if you're in-person or online. I think the chat or the the quiz should be live. Oops. If it's not working, it should be working now.

And I'm monitoring the results in real time. We're up to 22 votes.

So far we're a little bit divided. I'm gonna give you 10 more seconds.

Actually, now that you have the link, I'm gonna play the video one more time.

Oops, let's see. New technology, who dis?

[music plays]

[Mary] So right now we're sitting at 33 votes. We have 64% of people say that - or 61. The percentage changed on me. So, 33 votes. 34 votes. Okay, okay.

So we have about 60% say that "A" - the person on the right-hand side - is the real Robert Brunner. We have 40% who say "B" is the real Robert Brunner. And I will say this is a more even distribution than what I've seen in previous CITL workshops.

So I'm really - I'm excited that there's some divided on who's the real one.

Okay, let me boop a button real quick. Okay, so our goals for today.

So we have 3 main goals: first, we're going to describe how deepfakes work. So how do computers build models of people, or build and imitate voices? What's happening behind the scenes?

We're also going to consider the current applications and implications, including the legal landscape, which Rebecca will chime in for.

And then we're going to discuss techniques that you can use to spot a deepfake.

So, how computers see and hear us. A lot of the methods that we're going to be talking about resonate with tools that you've already used.

So if you've used a backup camera on a car, if you've searched for text...search for text in a PDF. You've used computer vision in some way. So when we're talking about digital images, we're talking about the basic building blocks, where we always start with the basic building blocks of an image, which is a pixel.

AI tools, generally speaking, break down images into pixels. They then look for patterns in those pixels, and then they use those patterns to generate new arrangements of pixels.

And this is where we come into generative AI tools, aka new images.

Okay. So remember, the computer is looking for patterns between the tiniest blocks of a digital image.

Now, practically speaking, it does this in several steps. So the first thing, it's going to break an image into pixels. It's then going to compare each pixel to the ones surrounding it, and it's going to look for three things.

The first is the red-blue-green color value. Then it's going to look for contrast for outlines, so general shapes. And then it's going to look for contrast for texture.

So is it something relatively solid like Rebecca's cardigan, or is it something that's more patterned like my shirt?

Once it has studied those relationships between pixels, it's going to use those patterns to identify which parts of the image are important, segments, and what features are in those images, and then it's going to classify what those features are.

This brings me to my absolute favorite problem in all of computer vision, known as the Chihuahua or blueberry muffin problem.

Computers can have a difficult time distinguishing between photos like these for three reasons. They have similar red-green-blue color values in that you have a light tan and a dark blueberry color. They have similar outlines in that there are a lot of round shapes, and they have a similar texture: predominantly light colors with some dark circles in between.

If you've seen "The Mitchells vs. the Machines", there's a version of this problem when the - this is not a spoiler, by the way- The machines have difficulty classifying the family dog as "dog, pig, dog, pig, loaf of bread".

So if you see the movie, keep an eye out for the Chihuahua or Blueberry Muffin problem.

Now, we can help computers better understand the patterns that they're seeing in images, and you've participated in this process.

So if you've ever had to identify all the stop lights in an image, you've helped the computer identify important features, or important segments rather, that's what you're seeing on the left.

And then you've helped the computer identify important features on the right. So saying this particular set of pixels, that's an A, that one's an R, that one's a C, that one's an H.

This is a process that is called "supervised learning", and it means that humans are supervising the ways that computers are interpreting information and providing feedback.

But a computer doesn't always need human supervision in order to learn how to do certain kinds of tasks better.

One of the most popular techniques for doing this is generative adversarial networks. And this is a form of unsupervised learning. It's most often used for image classification tools. And it's helpful to think about the algorithm splitting itself into three different people.

The first person is a real artist. In this case, it's Da Vinci, who's creating real works of art.

The second person is the counterfeiter or the generator, the Moriarty in this example, who is trying to create - or, uses what they've learned about the patterns and images and using that information to generate new images.

The program then feeds both the real images and the computer-generated images to a discriminator, who's looking to see if it can correctly identify which ones are real and which ones have been generated by the counterfeiter.

What's going to happen is with every success, the discriminator says: Aha, I've learned how to recognize those counterfeit images better. I've learned something and I'm going to incorporate that when I identify tools, or when I look at future pictures.

But the counterfeiter is learning as well. And so the counterfeiter says: hmm. The detector was able, or the discriminator was able to see this was not accurate. I need to modify it to do something different, until the counterfeiter gets so good at counterfeiting images, that the discriminator has difficulty distinguishing between what is real and what is fake.

Unsupervised learning is also the reason why we are no longer using things like AI detection tools in our classrooms. So because language models have been trained to beat any kind of AI detection, we found that AI detection tools just simply don't work.

So, yeah, this is this is part of the problem of deepfakes, in that it's really, really, really difficult to automate the process of identifying deepfakes because they were trained to beat other machines doing the same thing.

Okay, so...

After a computer has learned what patterns make a cat, for example. It's going to use that information to generate new images. And this process is something that we tend to describe as diffusion.

It's one of the most popular models, or popular techniques, things that are running tools like Midjourney, for example.

So what it does is as part of the learning process, it takes an image, like this image of a cat, and it introduces more and more and more static, more and more and more noise, until it can't quite see the cat anymore. And what it's trying to do is trying to figure out the minimum number of pixels that it takes in order to recognize a cat.

This information is really helpful because then it starts with, oh, I can take that, that minimum version of a cat, and then reverse engineer, add more and more and more pixels, until I get something that's like the cat on the left-hand side of the screen.

So, part of the generative process is taking those patterns and adding more and more pixels to that pattern in order to get you an image.

Now remember noise, because we're gonna come back to it in a moment where we talk about tools that you can use to protect your work, and how these tools are using things like random pixels to help protect your images and your voice.

So just because it recognizes patterns doesn't mean that the computer understands anatomy. So we have examples of AI-generated hands. They're really creepy, so I won't stay on this slide for long.

But what you're seeing is an example of a computer understanding what hands generally speaking should look like, but it doesn't understand that humans tend to have five fingers, and that they move and interact with each other in certain ways.

So hands are actually becoming one of the key ways - Yeah. Yeah. (inaudible - need to fix audio)

Okay, you that you need to speak up when that happens, because I can't see you. Alright, let's pause... Okay, I'm gonna keep talking and J.P. is gonna fly into action.

[muffled audio fixing sounds]

I tried turning it on and off again. Does this help the issue? I'm going to talk yet again or I'm gonna go back to our slide on diffusion so that we don't have to look at hands... for this entire process. I'm going to also try switching to one of the table mics.

Hi, party people online, can you still hear me? I'm not talking about pancakes, cats. Let's give the dog some love. Okay, fabulous. Then we'll switch to this mic.

Don't forget to verbally say, yeah, we have some problems. What point did the audio cut out? Oh, muffled? Yeah. Okay. Alright.

If there was a part that you want me to come back to in the Q&A, I'm more than happy to do that. And thank you for the heads up. So, eldritch abomination that is hands. Ugh.

But there are ways for a computer to understand three-dimensional objects. So taking pictures around an object, a computer can use the same information about red-blue-green color values, about color, contrast, and texture in order to take a series of two-dimensional photos and stitch them into 3 dimensions.

So this is a prelude for the ways that they were - how visual special effects can recreate human faces. And there was a significant breakthrough for the movie, "The Curious Case of Benjamin Button" in 2009, where the VFX team used Paul Ekman's facial action coding system, which is a series of about 70 facial expressions, to map Brad Pitt's face, base the digital models of Brad Pitt's face off of that, and then digitally took Brad's Pitt's digital head and placed it onto actors' bodies.

It's a really fascinating process, and I would very much recommend Ed Ulbrich's TED Talk that goes into more detail about how this was created.

So on the top part of the image, you see that they had to paint Brad Pitt in fluorescent paint in order to use this technology in 2009, and on the bottom you're seeing him act through scenes after they've created the digital model, to give them a better sense of how the skin on the face should move.

This was a radical break from traditional motion capture techniques, and I apologize that it's a little difficult to see with the lighting in this room.

But on the left, you see the new face mapping technique that the visual special effects team for Benjamin Button developed. And on the right, you see a more traditional version of motion capture, where there are dots painted on someone's face.

And what those dots and motion capture help the computer do is translate those points into polygons to make that three-dimensional model.

And the more polygons you have in that digital 3D model, the more detail that you can, and so they were able to achieve far more polygons with the scanning, or the method of creating the map of Brad Pitt's skin, than they were using other traditional motion capture methods.

So this - you have definitely seen instances where they have replaced an actor's, or replaced someone's face, or overlaid someone else's face on top of - or, person one's face on top of person two's body.

So this was 2009. More recently, Ryan Reynolds in "Free Guy" got his face painted on this incredibly muscular wrestler.

So you see the top photo is him actually filming his performance, and you'll see that there are photos in 360 in the round, so that they can capture all the nuance in his face.

Technology has evolved since Benjamin Button in that we're no longer painting actors with fluorescent paint.

I'm sure they're very glad of that. But I'm also an "Indiana Jones" stan, and so they've used this technology too to de-age actors, as they did in the last Indiana Jones movie, or virtually resurrect actors. So this is the technology behind Carrie Fisher's appearance in "Rogue One".

But you might be asking, we've been talking a lot about images. What about sound? Sound, surprisingly, is an image recognition problem too.

So, but before we get into talking about how it is an image recognition problem, we need to talk about two important pieces of sound.

So the first thing we need to do is talk about frequency. So frequency refers to how high or how low in pitch something is.

So something with a relatively high pitch. I don't wanna quite go...[unintelligible high pitch sounds]

Where as something that is like Johnny Cash range has lower frequency and lower pitch.

The other thing that image, or that's that's important for deepfakes of voice, is amplitude.

One way to think of amplitude is thinking about it in terms of how loud something is or how intense the sound is. So something with a relatively small amplitude is going to be very quiet. But it might, so this voice has the same frequency as this voice, but a different amplitude.

Something that is very loud, which I won't model for those of you who are wearing headphones. Something that is very loud will have a very large amplitude.

So this is where it gets interesting, in how sound is actually an image recognition problem.

What happens is, the first step to processing sound or studying patterns in sound, is to transform something that's from sound input into visual input.

So what you see here is a spectrogram of the phrase "nineteenth century". And what's happening is we have frequency or pitch on the left, or going up and down the XY axis. So things that are at the top of this graph are going to be very high in pitch, and things that are towards the bottom are going to be very low in their pitch.

You also may have noticed that there are different colors in this graph, some parts of the graph are dark and some parts of the graph are light.

This refers to the intensity. So things that are towards the bottom of the graph, especially towards the left-hand side of the graph, are very intense and very low frequency.

And then from here it becomes a matter of recognizing the patterns in pixels.

So here we see an example of the soundscape of Mount Rainier. Rainier? [not sure of pronunciation] I always have been told by those who live in Washington State, it's Mount Rain-ier. Because it's always raining there.

But what you see are little boxes around distinct sounds. So things like bird calls or insects tend to be very high in frequency. Whereas jets are much, have much lower frequency and they're kind of rumbly.

So you can use the patterns and pixels in a spectrogram, not just to identify sounds, but then to recreate those sounds through voice generation, and using AI to generate sounds.

Okay, so use cases. There are quite a few really impressive use cases.

One of them is to use AI to generate the lips and the voice of the actor and to translate that into multiple languages. So here's an example. I will say, this is from a rated R movie, but the saucier words have been bleeped out.

[Video clip] Now we're stuck on this stupid tower in the middle of nowhere, and I don't blame you, and now we're stuck on this stupid-

Stuck on this stupid freaking tower in the middle of freaking nowhere. And it's all my fault.

[Japanese dubbing, followed by Spanish dubbing]

[Mary] And so what you're hearing is, not just have they mapped the lips onto the actress' face, but they've also dubbed the actress in her voice using pre-trained models that can speak in these languages.

And I, one of my favorite examples of AI avatars is actually using these to speak multiple languages.

One of my favorite moments from the faculty retreat was seeing Michel, the director of CITL, speak in Chinese and Portuguese, which are languages that he doesn't speak in real life, but it was really cool to see his AI avatar speaking them fluently.

So, HeyGen is a tool that you can use. You can use either pre-trained AI avatars, or you can create your own avatars, as we saw in the Robert Brunner example, but it also has additional features like text to speech, etc, etc.

It allows you to speak foreign languages, but it also can be a tool for maintaining identity-or, or, maintaining your anonymity if you prefer to use an AI avatar instead of your real image and real voice.

Now, of course, there is a significant sinister side to deepfakes. We have already seen examples of deep-faked images created as part of the election cycle, and also deep-faked videos of President Zelenskyy from the Ukraine as propaganda.

And it's going to be an important information literacy - an important issue in the election, when these fake images are mistaken as real images.

This is perhaps one of the worst statistics that I have to present.

When we're talking about lip syncing and we're talking about special effects, I get really excited. And I think that they're really cool, especially in helping promote communication. This is less than 2% of use cases.

It's estimated that 98% of use cases is to deepfake pornographic material of women without their consent. And we currently don't have any legislation to prevent this.

The current legal landscape is really tricky, and there are reasons why it's really difficult to regulate this particular issue without causing significant ramifications for other types of data mining and text mining and the kinds of potentially good research that we can do using similar techniques.

So for talking about the legal landscape, I am going to turn things over to Rebecca.

[Rebecca] Thank you. Can everyone hear me?

So as Mary mentioned, the legal landscape is really complicated and still very much in process. A lot of determinations on copyright and AI are just still under consideration in the courts right now.

But for now, what we know is that generally speaking, training AI on copyrighted material we think is going to be considered fair use.

So for copyrighted material. Part of what copyright is, is that, you know, if I take an article that the New York Times published, I can't publish it myself in my own name because that would be taking something that they made.

However, there is an exception in copyright called fair use, which allows for a limited use of copyrighted material without permission from the copyright holder for purposes like criticism, comment, news reporting, teaching, scholarship, or research.

Courts have found that text and data mining are quintessential fair use cases.

So for example, a large scale text mining such as in HathiTrust is a fair use case, and this was a legal case that was determined.

A transformative use is an aspect of fair use, which is one that alters the original work with a new expression or meaning or message.

You can use a copyrighted work in a way that it was not originally intended, such as in AI or like AI art, etc.

However, the kind of pre-existing case that exists is in one where physical books were ingested into HathiTrust with no, like, restrictions or licenses that they had, and no AI was trained on this, like, body.

So, is it enforceable scraping the web using training AI on these, like, basically on the internet. We're still not entirely sure. So is training AI fair use? It is up to the courts.

There is currently filed on, I believe, December 27th, 2023, so a few months ago, a court case of New York Times v. Microsoft et al., so including Microsoft Copilot, and also Open AI, which makes ChatGPT.

And the New York Times is saying that, you can't use our immense body of work to train it because it's unlawful, it limits the Times' ability to provide that service and it violates the Times' copyright.

This case has not been determined yet, so it'll be really interesting to see what the courts end up saying about it.

So could we use this video that we're doing right now to train AI? Many copyright experts think the answer is yes, but there also might be other measures preventing you from using the text.

So one of these examples is Terms of Service or licensing agreement.

So, DeviantArt, which is a website where artists can post their art, has within their, kind of, terms of service, you can market as "no AI", which means that it will not be used, AI, it's like prevented from being used.

However, for TikTok - we look at TikTok's Terms of Services.

So you or the owner of your user content still own the copyright, in user content sent to us, but by submitting user content via the services-TikTok - you hereby grant us an unconditional, irrevocable, non-exclusive, royalty-free, fully transferable, perpetual worldwide license to use, modify, adapt, reproduce, make derivative works of, publish and/or transmit and/or distribute and/or to authorize other users of the services and other third parties to view, access, use, download, modify, adapt, reproduce, make derivative works of, publish, and/or transmit your User Content in any form on any platform, either known or hereinafter invented. So, perpetual. Anytime in the future.

[Attendee] Essentially... [some chatter to turn on mic] You own it, but you don't own it. Essentially, you own it, but now everyone owns it. But they own it through TikTok. So TikTok can sell your stuff to other people. Why do people like TikTok?

[Mary] Pretty pictures?

[Rebecca] Dancing challenges? So. This is all kind of within the Terms of Service languages, so the kind of stuff that will... A lot of people will just click through because it is buried in a really big wall of text.

And you don't really think about that when you're just like clicking through your service agreement.

And that is one of the really complex and really concerning things about these big Terms of Services. So we have an example. From TikTok, specifically. Slightly different issue. So in, yeah.

[Video Clip] Anywhere you hear a voice and don't see a face. It's kind of what I do.

[voiceover] And now it seems she does a lot of this as well. [AI voice] Hi TikTok. My name is Misty.

[some muffled voices] [Rebecca] Oh no, that was, that was, that was it. I just did the small clip.

So as we see, in the first part of that video was a lady called Bev Standing, who is a voice actress from Canada and records a lot of commercials. etc.

She made recordings, that she created for the Chinese Institute of Acoustics Research, for a text to speech tool.

They, she found out via, I believe, a friend or her child, that her voice was being used on TikTok without her consent and without compensation.

So in TikTok's text to speech application, her voice was being used as we hear in the video. She did not know. TikTok did not tell her, and she was not being compensated for her work.

She sued for damage to her brand. And it was settled out of court. So - And TikTok is using someone else's voice now. I believe it is an actual voice actress that they're paying, but, point being is that her voice was able to be kind of just taken and used based on these other recordings that she had done without her permission or knowledge.

And this is a really big issue right now for voice actors and actresses. That their voices can be transformed and used without their knowledge and then they're out of jobs and they're not being told.

And this actually came up in the SAG-AFTRA agreement. And right now their agreement says that it requires an AI firm to get consent from actors before it uses their voices based on their likenesses and also gives voice actors the ability to die, their voice being used in perpetuity without their consent.

However, the language is still complicated and it's not completely solved yet. So a lot of this is going to be wait and see and come to next year's Savvy researcher, or the year after.

[Mary] So switching to if your voice can be used to train an AI model. Out of [muffled] transformative use of your image.

What can you do to protect your face, your voice, your image in this digital age?

So there are a couple of different tools that we recommend.

This is also an evolving field, so I'm not entirely sure what the best option is in terms of video yet. This is one of the things that we're investigating over the summer.

But for images, there's two really fabulous tools already available created by an artist collective through UChicago.

The first is Glaze. So remember what we were talking about noise and stable diffusion and how stable diffusion, when you're generating images, requires you to add more and more and more pixels and uses noise to better understand the patterns in those pixels?

What Glaze does is it introduces pixels randomly into an image that aren't perceivable by most humans, but does make a difference to the ways that AI understands the pattern's specific image.

So you see the top row of images are original art pieces.

You'll see without Glaze, that second middle row, that they are fairly close in composition and style, but the bottom row shows how much difficulty a computer has mimicking this artist's style once she's added Glaze to the image.

Now, they've also created Nightshade, which goes one step further and poisons the model, tricking it into thinking that the image is completely different.

So it adds noise that tricks the computer into thinking that an image of a dog is actually a cat. And it completely borks the results that it gets after that point.

There's also a free Python script called Antifake that introduces noise into audio files, but the challenge is that noise and audio files is more perceptible to humans than random pixels in an image so it does diminish the audio quality.

Now, how to spot a deepfake? You're probably going to see quite a few pop up on social media, in the news.

Katy Perry's mom quite recently and quite famously got tricked into thinking that Katy Perry was actually at the Met Gala.

So what can you do to help spot a deepfake?

One of the best ways is the SIFT method. So the first is to stop before doing anything else with the image. Re-sharing it, reposting it, etc, etc. Two, investigate the source. Is this a reliable news source? Is this coming from a library or an image repository or something like Reuters or AP that serve a lot of different

news agencies? Is this coming from Reddit? A site you've never heard of. So thinking about where this image is or where this thing is coming from.

Finding better coverage. Is there other examples of this particular recording or image or video. Or is this a fairly unique thing?

And then, tracing claims to their original context, does the person - or, does the content make sense given this person's political background or their area of expertise?

One of the tools that you can use to help you find especially images is to do a reverse image search. So these are things like TinEye or Google Lens that will help you uncover the original context of an image.

So things like, hey, there's this really cool photograph of a Civil War soldier riding a dinosaur and it only has just appeared within the past month.

If you use TinEye and Google Lens and trace it back to a library, well, maybe they're in the process of digitizing their collection and maybe there is such a thing as a Civil War soldier riding a dinosaur. (quietly) I doubt it.

But we go against things that we know about the context. For the sake of example, if we can trace it back to a reputable source, it's most likely real.

If it's on a sub thread of Reddit, I would be a little bit more suspicious.

There are some other things that we can look for, especially in video content.

And I want you to be thinking about these when we look at the video of Robert Brunner again in a moment.

But there are some visual cues that we can use to see if this is something real or something that has been deepfaked.

So gestures, if someone is very animated and tends to talk with their hands, that is usually a indication that it is - it was more likely to be real, because it's very difficult to generate someone who has a lot of hand gestures and moves frequently.

It's much easier to deep fake someone that's standing still.

So if you look at the deepfaked video of President Zelenskyy, he's at a podium. He's not moving his hands and he has a very limited range of movement and turning his head.

Hands, we saw. We saw what terrible things...what terrible things generative AI does to hands.

And this isn't a perfect method. Because you can have an actor who's making gestures and then replace their face with someone else's.

But if we're looking at completely AI generated avatars like the one that HeyJen can do, you'll notice that you don't often see their hands.

Backgrounds. So it's easier to manipulate a video of someone if the background is solid, as opposed to has a lot of texture, has a lot of things going on behind them. So looking at the background, if it's a fairly consistent or solid background, that's a little sus.

And then camera movement. Again, it's really, really hard, unless you have a multi-million dollar budget like the Indiana Jones movies, to track someone's movement through space, because anything that you do with the face also has to mimic the lighting conditions of the space that someone is moving through.

There are also sound cues as well. So you'll notice that the video that I showed you does not have sound.

That's because it's easier to pick up, or at least for me, it's easier to pick up on AI generated voices than it is on AI generated images.

So a lack of sound, because it's harder to fake the human voice, is a indication that something might require a little bit more research.

Voice tone. So someone who tends to talk in more, or this can be really hard for people who tend to talk in more of a monotone.

So if I'm talking like this, my voice tone doesn't really modify until maybe I get to the end of the sentence.

That is easier to mimic than say, someone who has more of a radio announcer's voice, where my voice is going up and down in pitch and I'm adding emphasis.

I'm adding different kinds of speaking. Like, if I wanted to do a little bit of vocal fry, I could.

So looking at the things like the tone of voice, the melody of the voice, if you will.

The more melodic someone's voice is, the more up and down in pitch, the more animated, the more varied, the more likely it is to be real.

And then of course there's content cues. Is that something that this person would do?

If I showed you a video of Robert Brunner singing along to "Redbones' ooga-chaka, ooga-ooga-ooga-chaka". That doesn't make sense in a professional context.

Is it something that we could do with the video material that we currently have? Probably. Does the university have policies to prevent us from doing this? No. Is this a problem? Yes.

But thinking about, is this natural for this person to be talking about this content, or does this make sense in the context that I'm viewing it, is also a helpful tip-on.

So get your camera, your phone cameras ready. We're gonna look at the video again, and then I'll flash a QR code and you can complete the survey, but keep these things in mind as you're looking at the video.

And remember, it doesn't have any sound except for music.

[music]

[Mary] Okay, quiz time. So, I'll flash the QR code one more time.

It always makes me feel like a celebrity when people are looking at their QR codes. For those of you who are online, there's just, I'm looking out at a group of people who have their phones up and it's really delightful.

So, we're gonna go back to the video, let you see it one more time, and then we'll review the results.

Let's see, there we go.

[music]

So I'll give you 10 more seconds, because we only have 25 votes at the moment and we had 36 votes, and for the record I can't see what you voted from - in the previous poll or in the this.

All the polls are anonymous. So don't feel shy. I'm not grading you. I'm not quizzing you It's just helpful to see what what you think it.

So we have 30, 37 votes, 38 votes. Yay! Have even more people weighing in.

So I'm gonna give you 5 more seconds. I will say it was really difficult for me at first to distinguish who is real and who is fake.

And the detail that I was, like, really drawn to initially is the U of I pin in the image on the right - or the left hand, sorry.

Robert Brunner "A" has a U of I pin on his lapel, and that's something that's unique, that's something that's context appropriate.

That's something that's distinguished. That's something that I wouldn't necessarily expect to see on a AI generated avatar.

However, as the video goes on, you notice - and in fact, we've actually switched.

So when we first showed this video, most of you thought that "A" was the accurate, or was the real Robert Brunner.

But now we have about 28% people saying that "A" is the real, and 72% of people saying "B" is the real and the answer is "B", and it's the hands that were the dead giveaway for me, especially, so when we look at this video again, you'll see that Robert Brunner "B" is a little bit more animated than Robert Brunner "A".

So you'll also see a "B"s thumbs at one point, and you'll notice that "A" keeps his hands outside of the frames. So, aha, similar gestures. But we actually see fingers in "B".

Even though he very cleverly took off the U of I pin. It's a really, really good and a really, really convincing deepfake.

Or AI avatar, I should say, because this was created with Robert Brunner's consent and blessing and his input.

So, I do want to make that very clear. And when we tend to talk about deepfakes, we're talking about unauthorized...

Things that are created without consent. Okay, yes. Mike.

[Attendee] Can I add about what was telling about that? [Mary] Yes!

[Attendee] Sure, so the one on the left, like the contrast is just too perfect, whereas the one on the right has like elements that would show you that it was made with a camera that's shooting in like, a raw format, so it's all gray. So you see how like the highlights are really... Like the one on "A" looks too idealized, whereas "B" has artifacts of humans using technology.

[Mary] So, yeah, so lighting conditions, there's a higher contrast-

[Attendee] There was also in his - on his armpit, or in that like gap.

You'll notice a bunch of green screen artifacts, where like, this was shot on a green screen- [overlapping voices]

And so, where - which is obvious on "A", or on "B", but not on "A". [overlapping voices again]

Thus showing that this is more of a generative content. I'm looking at some of like the image quality and video quality of artifacts as tells. The subject matter is part of it, but it's the way the subject matter is presented to the viewer.

[Attendee] That "B" has more. Of what you would expect from a man-made thing. [Mary] Right, right.

Mary] So when we're talking about artifacts in this case, we're talking about the little slivers of the green screen that you can see sometimes, when if you use like the "Blur my background"-

[Attendee] It's where the computer is trying to determine whether it's subject or background, and it's getting half pixels, like half translucent pixels or partially opaque pixels.

So you're getting, that's what I meant by green screen artifacts, where you're seeing the algorithm failing. If you play it, you can see it, underneath his arm. On the right, all the way to the far right corner, lower right corner.

[Mary] Yeah. So it's similar to, you know when you're watching the news, and the weatherman gestures to the map on the wall, and like, there are moments where you see the map through the weatherman, or the weatherwoman.

That's the effect that we're talking about here. So we're looking at the lower righthand part of the screen?

[Attendee] There. See, they're kind of moving around and bouncing in that shadow?

[Mary] Your eyes are better than mine - oh, wait, yeah I do! Like, it's flickering right there. Ohhh.

[Attendee] That's the green screen failing. The algorithm that's trying to differentiate subject matter from green screen is failing right there, and that's what it looks like.

[Mary] Very cool. So really quickly, because I know that we're reaching the top of the hour, some resources.

So library resources, we have, of course we have plenty of workshops, especially during the semester. A little quiet over the summer, but be sure to keep your eye out for the Savvy Researcher calendar.

We are posting all of our recordings of the workshops that we've been giving on AI, and of the arts and humanities in our "Digital Humanities at Illinois" Media Space channel.

And of course, the library also has a generative AI libguide with advice, resources for if you are considering using AI in your research and in your creative practice.

I also want to give a shout out to some of the spaces that you can play with AI.

So there's SCIM Lab, which is in the lower level of the Grainger Engineering Library. They have motion capture and video generation techniques, or capabilities with TensorFlow, and the best way to find a time to explore that space is to request a consultation with Jake Metz through the link that's on the SCIM Lab page.

The hours are a little bit, it might be closed over the summer, but if you request a consultation, you should get a response.

There's also of course, CITL's Innovation Studio that has HayGen, but also some other popular AI generation tools like the journey. So you can visit their space. I forgot to double check, what their hours are over the summer, but I suspect that there might be some CITL folks online.

So if those hours aren't correct for the summer, please send a message to Cadence and we'll make sure that's related in the Q&A.

And finally, I am delighted to help answer questions. So my expertise is in artificial intelligence in the humanities.

I also do some work around 3D modeling cultural heritage objects, and I will quite selfishly invite you to next week's workshop on "Digitizing in 3 Dimensions".

So if you're curious to see how we can use computer vision and this kind of 360 degree photography to create digital models and then to take those digital models in 3D print them, please join us next week.

For those of you joining us in the future, there will be a recording of that workshop on the DH Media Space Channel.

There's also Sarah Benson, our an amazing copyright librarian who can handle further questions about the legal status of AI generated images and deepfakes, and then Jake Metz, who is our point person for questions about AI tools, including those that you can use to protect yourself.

And I realize I have gotten his email wrong, with apologies to Jake.

If you use the consultation request form through the SCIM Lab, that's the best way to get in touch with him.

So with that, I thank you so much for joining us today or in the future. I'm going to turn off the recording and open the floor for questions.