25th ISOJ New Google tools for reporting and fact-checking

- **Ashley Edwards,** News Partner Manager, Associations and Local Partners, **Google**
- **Mary Nahorniak,** U.S. Teaching Fellow, **Google News Initiative**

Ashley Edwards [00:00:00] Okay everyone, we're going to get started. Thank you so much for joining us. This lunch session is about new tools for reporting and fact checking. So we're going to go over some new and old tools for you. I'm Ashley Edwards, I'm a news partner manager at Google.

Mary Nahorniak [00:00:23] Hi, I'm Mary Nahorniak, I'm a US teaching fellow with the Google News Initiative.

Ashley Edwards [00:00:28] And our contact information is on that cheat sheet. But we'll have it for you again later.

Mary Nahorniak [00:00:32] We'll bring that back up. Don't worry.

Ashley Edwards [00:00:34] So the goals of the Google News initiative is to advance the practice of quality journalism, strengthen and involve publisher business models, and cultivate a collaborative global news community. So the GNI is in 119 countries. Over our life span, we provided 300 million in global funding. We've worked with over 6000 news publishers and we've trained 2.5 million journalists.

Mary Nahorniak [00:01:05] We're going to add like 40 after this.

Ashley Edwards [00:01:07] Yes.

Mary Nahorniak [00:01:09] So you are part of that number.

Ashley Edwards [00:01:11] All right, so if you didn't get a chance to scan the QR code, it's on the screen again just for another minute. And what we're going to go through today are fact checking tools, pinpoint and data extraction.

Mary Nahorniak [00:01:26] So we will show you a lot of things and don't feel like you have to get all the screenshots, write everything down. It's all in the cheat sheet plus contact information if we forgot something or you have questions. I see phones are down, so I'll move to the next slide which we're going to start with the game.

Ashley Edwards [00:01:40] Yeah. So just to warm everyone up we're going to start with a little fun game of which face is real. Okay. There are a lot of AI generating tools around there, out there right now. And it's very easy to follow things. So please, with a show of hands, which one of these faces is real and which one is AI generated? Is this one real? Who thinks its real? And who thinks this one's real?

Mary Nahorniak [00:02:09] We have more votes for number two.

Ashley Edwards [00:02:11] Okay. So, Mary, do you want to do the reveal? That is the real face. So just some tips. We're going to go through this a little bit more, and this will work into what we're going to train on. But some way to tell a face is AI generated.

Mary Nahorniak [00:02:27] Take a look at the teeth, the ears and sometimes the background will be like weird warping. So like right here, it's kind of covered, but you can think of weird warping by the ears and then like the teeth right here, you can kind of see it.

Ashley Edwards [00:02:44] We'll do a few more. Okay. Who thinks this one is the real face?

Ashley Edwards [00:02:59] Who thinks this one is the real face? More. Less.

Ashley Edwards [00:03:05] Again. This one, is the real one. Take a look at the teeth. Usually is a telltale sign. And you can see here by the collar. It looks a little strange. And then there's something a little off with the stares, sometimes with AI generated images. Okay, we're going to do one more. This one. Okay. Which one? Who votes for this one? As a real face to look around it. And then this one. Right. The reveal. This is the real face. So again, it's the teeth are kind of the giveaway. The teeth kind of look blurred together. A little bit. It looks kind of like a blob. That's a usual telltale sign. Also, if you look here, it looks like a blob. So with a lot of you have questions here. Oh I'm sorry. Okay. We might need this mic then. So this is a helpful tool because with a lot of fake accounts like you'll see fake Facebook accounts, X accounts, things like that. They will use AI generated photos of people. So a good if you're trying to figure out if an account is real or not and you want to start with the photo, maybe reverse image search the photo. And if it's nowhere else on the internet, maybe that's one sign of it being AI generated. But also if it takes you to that website that it's like generates fake faces. And that's kind of like, you know, that it's fake. All right. So we're going to start with something most of you have probably seen. But if you're not reverse image search you can use image files to perform a reverse image search in Google to see where else that photo has been. So you can use an image URL or upload the image directly into Google Images. So for example you can reverse image search this. And see where it comes up. So as you can see, it doesn't really come up anywhere on the internet. It will show you maybe some similar photos with that, it could be a sign that it is reverse AI generated. So here are some new features with reverse image search. It's called this image. So in addition to showing you the image where else it is on the internet, it will tell you information about when the image was first indexed by Google, where it may have first appeared, and where else it appears on the internet. So a common tactic of misinformation is to use photos out of context, for example. So if you reverse image, search an image from a fake news story and it says like this just happened, but it comes up, you know, from a couple years ago, that's a sign that it's manipulated information.

Mary Nahorniak [00:06:20] I'm going to search for an image. Who has had a chance to stop by and see the Ellsworth Kelly Building. That's part of the Museum of Art next door. Anyone? I went

yesterday it is incredibly cool, so I'm just going to search for that. Here it is. It's called Ellsworth Kelly Austin. The name of the work is Austin. I'm going to go to my image results. So this is what it looks like. It's just like 20ft that way. I'm gonna click on any of them. And then from here so everyone can do this. And this is a new public facing, not just for journalists piece of information that gives us some more details about a photo. And so as I click on this three dot menu this works on phones as well and I click about this image. The first thing I see is I can just learn that a version of this is about five years old. So this is really helpful new information that was not previously available on Google. If you wanted to figure out how old something was, you would have to go back in time like chunk by chunk, which you can still do. But now we can just very quickly see. This helps us understand if something is real, or if it was dreamed up yesterday or generated by AI a week ago. We can now understand how old something is. So this is meant to be easy for anyone to find out more about an image, particularly in an election year. Lots of global elections this year, lots of mis and disinformation. So that one is new as of, I think, a couple of months ago, so you may not have seen this one yet.

Ashley Edwards [00:08:00] And on that note. Now this is even newer based on like a couple days ago. You can now see if an image was made specifically with Google AI. So this will now be it'll be indicated prominently when an image was made with a Google AI product. This doesn't work for other AI products. So like there's a lot of other AI generating software out there, but specifically with Google. So for example, this is a photo of, a butterfly that was made with Google AI. So you'll be able to see it right there. Just tap those three little dots. So this is another fun one. An image floating around of dinosaurs on Mars. Maybe a little less believable than the butterflies, but you never know. So as you can see in the image, we just scroll down and right there the credit created with Google. So a lot of images won't be as fantastical as this. Like the butterfly is a little bit more believable, but you'll be able to tell with that. I saw a hand.

Mary Nahorniak [00:09:18] The way that I've had it explained is you can think of it like a watermark. I don't actually know how it works, which I think is what you're asking. Yep, yep. Ooh, that's such a good question. I have to assume, yes, but I don't know that for sure. I would love to experiment with that after we're done here. I think we can figure that out pretty quickly. Cool. Good question. Oh, yes. Right? Yes. It's not a watermark that's pressed onto the photo. It's in the metadata. I think the screenshot is an interesting use case. So it sounds like, Richard, your understanding would also be that it would still show with the credit. Yeah. And another question. So, for folks on Zoom. So we were just talking about how copyright work with an AI generated image? And essentially what we tell folks in trainings, because there are so many tools where you can create images, whether it's Google Earth or any of the AI tools, etc., is just read what it tells you to do, right. And there may be different policies per tool or how you use it as journalists or commercially, a lot of different things. So just take a look at the copyright language. Thanks for the great questions. Are there questions from Zoom? Yes, I will do a better job of that. Sorry hybrid folks.

Ashley Edwards [00:11:25] So another tool that we have is Fact -heck Explorer, which is a searchable database of available fact checks on different things that are going on on the internet. So you can go to Fact-Check Explorer, type in something that you're seeing trending

and see if there's a fact check on it. So for example, one of the first fact checks right here. Eggs, fruits, oats are unhealthy breakfast. So there is some fact. I guess there is maybe something floating around about how this is unhealthy to eat and there is a fact check that was recently uploaded on this. Okay, so very relevant to election season. This is a claim that over 2 million people in Texas, Pennsylvania and Arizona have registered to vote without photo ID. And as you can see from this, these various fact checks, that is false. So especially during election time, you're going to see a lot of claims floating around that you yourself might want to fact check and a good place to start to see if it has already been fact checked is Fact-Check Explorer.

Ashley Edwards [00:12:37] Question? Yes. So for those on zoom, the question was how do we scrape the fact checks? We actually have like a markup tool, that developers on the back end can add to. It's a little technical, but we'll give you the information that can be used to indicate to Google that your content is a fact check or contains fact checks. So after this we can give you the information on that as well. And we're going to show on the screen where it is.

Ashley Edwards [00:13:16] Yeah. Of course. Yes, so for those on Zoom, she asks if there is a verification process on the fact checks that are in our repository. Yes, we work with several big fact checking organizations that accredit different fact checkers. And that's how we do that. So we can give you the information, like if you want this on your website, we'll give you that information on how to do that. And another good point is that you have to be in Google News in order to be able to use this and come up. So we're not just like adding fact checks to like any random scammer on the internet.

Ashley Edwards [00:14:12] Yes. Okay, so another new part of a Fact-Check Explorer is that you can now search Fact-Check Explorer for images. This is available to everyone now. As well of last week that you can search this much like a reverse image search, but you can search an image, right? In Fact-Check Explorer. So let's do an example. This is one of my favorite, every time there is a storm or a hurricane in the US. This picture starts trending of a shark swimming on a flooded highway. This may seem funny, but like a lot of people fall for this. My husband fell for this like a couple of months ago where he sent me. This photo was like, did you see what happened in Florida? I'm like, yeah, I did. I also saw this photo for like the past ten years. So now you can search an image right in Fact-Check Explorer. So as you can see, you upload that and you can see that this comes up, there's a lot of fact checks from September 2022. And if we do an image context on this, we can see this photo has been floating around for a very, very long time. So the image context part is not available. It's a closed beta that you can all sign up for. We'll give you the information. So this part is not available for everyone just yet. But like I said you can see an image. The image context shows that this was first indexed by Google in 2011. So even though it's popping up every year, and it may seem like we all know this is fake, but a lot of people do not. A lot of people, when they see it, it's the first time they're seeing it, so this is a good way to fact check this sort of content. So we're going to show where you can sign up for this beta. And for those on Zoom we can send this information as well. Yeah, it's all in the cheat sheet to apply for the beta to see this. Great and Mary is going to take over for Pinpoint now.

Mary Nahorniak [00:16:28] All right. As we switch gears to a whole other tool and person, just check in for any questions. So the question from Zoom is, can we use the links to fact check videos that circulate on social media? No. It needs to be an image upload. What you could do is use a tool like engVid, or just take a screenshot as you play a video, like we kind of saw in the previous session and you can fact check that, right? So you can upload your screengrab. And in fact, the shark photo that Ashley showed is a screenshot. So a little bit back to kind of the watermarking AI question. That's one that I just grabbed a screenshot myself off of a tweet. One of the more recent times and it went viral, possibly when her husband fell for it and we were talking about it. So a video can't be uploaded? No, but a screengrab of it can and then it will. It's a good first stop, because if there is a fact check that's been done by a reputable news organization, it should come up. If it's too early, if something is just breaking and you're not sure yet, and it may not pull up a fact check that doesn't tell you anything yet. It's just that something is brand new. It's a really nice, first stop shop. Then you move on to reverse image search and other tools like I mentioned EngVid. That's a really good one. We're not demoing that today, but big fans. Yes. So of the PowerPoint. We're not going to share the slides but we do have the cheat sheet, so at the end of the session I'll make sure to bring that QR code back up. So folks on Zoom can point their phones at their computer screens as well.

Ashley Edwards [00:18:04] And then oh my god, sorry. With the cheat sheet. You'll see all these other tools, links to the Google News Initiative training center where you can take some more lessons on these tools. But also we do training sessions for the newsroom. So if your colleagues and your newsroom want a deep dive into any of these tools or want to go over this again, we can set that up so our contact information is there as well.

Mary Nahorniak [00:18:33] Anything else from Zoom? Okay. Thumbs up. Who has heard of Pinpoint? Sweet, who has used Pinpoint. Sweet. Okay, that's great, 100%. If you've heard of it, you've used it. I love that. That's already great. So Pinpoint is a tool that the Google News Initiative worked with within Google to develop for only two groups of people, which are journalists and academic researchers. Those are the only groups of people that can access it right now. It's a free tool. It does a bunch of different things. I'm going to show you most of those. It is rapidly iterating. So every couple of weeks I get information about a new feature. I'm going to show you two brand new ones, including a generative AI feature. And what it's meant to do is help you make sense of large document collections. So whether that's at the reporting level, maybe you got, public records, tranches been delivered to you or you've been collecting something over time. This is a great way to help you get your arms around it on day one. And potentially, like, if you build something, let's say you cover city council. You're just constantly putting in the meeting agendas. You're downloading their Zooms off of YouTube, you're getting any other things that they are putting out, and you're just running that collection over time so that when news pops up, spot news, you can search through your collection for when's the last time this developer was referenced, or this the teacher salary strike, right? Whatever that might be. So you can kind of build things that make sense over time. So I'll give some more use cases as we go along. But just kind of the basics help you make sense of large document collections. And when I say large, I mean large, potentially up to, 200,000 documents in a single collection. When you get a Pinpoint account, which you all can and recommend that you do, you get two

gigs of storage if you use it for important work and you run out of storage, please write to the Pinpoint team whose information is on the cheat sheet or to us, and we will work to get you more space. That does not detract from the space that's part of your Drive or your Gmail. So if, like me, you're constantly at like 91%, it's totally separate. So don't have to worry about that. Okay that was introducing Pinpoint. So there's are all of the different ways that we have talked to journalists about how they've used it. And we're continually hearing of new use cases too. So if you have used it or you intend to and you're using it for something really interesting, we would love to hear from you. We would love any kind of feedback as well, so I can't use it until it does this. I love it because it does that. Whatever. We're always looking for new ways of making it come alive. Since we have 5 or 6 hands in the room of folks who have used it, does anybody want to just chime in with how they have? And I'll repeat it back on the mic. So hey, look at that. A traveling mic here is a fella in the back.

Volunteer [00:21:21] We've used it to look through reports from, facilities that treatment, kids in like residential mental health facilities and the incidents reports that they have filed. So we put them in a collection so that we can find different parts of that and use it with campaign finance reports, including handwritten reports. So there's different.

Mary Nahorniak [00:21:48] Perfect. Okay. So I think the folks on Zoom heard that. Anybody else want to weigh in? Those are awesome examples. Oh, we got another one.

Volunteer [00:21:58] We've been using it for migrant farmworker housing inspection reports, which are a mess and vary by state, by state. So it's been pretty handy, but it's a challenge.

Mary Nahorniak [00:22:09] Yeah. Wonderful. Okay. Great example. So I know I mean, a number of organizations will look at the same thing in many places. Right. So whether that's what you're doing or not looking at. Housing inspections or campaign finance forms across all 50 states. You can build 50 collections. It does search through handwriting, so a lot of those things are handwritten police reports, restaurant inspections, you name it. If it can be read, it can be found, and you can even have something handwritten and copy it and paste it into another file and it will become copy and pasteable typed out text without you having to retype it. So there's a lot of time saved, which I think is one of the ways to help think about how to integrate Pinpoint into your work is letting a tool like this do some of the heavy lifting things like, transcription. Let's we'll talk about things that take a lot of time. Data entry, extracting data off of files that you can then use your limited time and energy, because we only have so much every day to do the things that only you can do, which is developing news stories, interviewing folks, new sources, writing your wonderful stories, whatever that might be. So I'm going to walk through the slides just again to help you kind of get our arms around it together. And then we'll just do a live demo and I'll show you what a bunch of that stuff looks like. So searching through documents. So that's one of the first ways you've uploaded a big collection. It's going to quickly break out for you. Let me see if I have that. It's going to break out names, locations and organizations. Anything that Pinpoint and Google understands to be one of those. It's going to pull it out into a sidebar so you can very quickly see, oh, this person gets talked a lot about in these 900 documents that I have and haven't had a chance to read through yet, and maybe that's where

my story is. Maybe it isn't right, but it's something to understand. We talked about handwriting, so I want to talk about how I end up saying documents a lot, right? It feels like it kind of loses its meaning. Of course, I think of PDFs and Word docs. I think it's important to hover here, because I want to talk about all the other things that it can take in and ingest. So PDFs, PowerPoints, if you cover an organization or something like the military, big companies, they love to share information on PowerPoints. It can read all of that. It can read text within a photo. So if I took a picture of this room and looked for text, I could read maybe like a sticker on your laptop or the name on the water bottle. All readable. This morning I took a picture of the name tags that hadn't yet been collected, and I haven't put it in Pinpoint yet, but I could quickly see, oh, was so-and-so here yet? Had they picked up their name tag by 9 a.m.? And then transcription. So it supports audio and video files, meaning upload an audio or video file of any of these types, which is pretty much all of them, and it's going to immediately start transcribing it for you. You don't have to tell it to do that. It understands that's what you want it to do. And it can do that in multiple languages. So I'll show you where that lives as well. This is one of the newer features. This is now available to anyone with a Pinpoint account, which is extracting data off of PDFs that are structured the same. So all campaign finance forms or any of the forms we just heard from the folks in the audience, 9 90s for nonprofits, you name it. If you have a bunch of them, you're going to annotate one, and it's going to pull all of that information off of all of them simultaneously. I don't like to say that technology is magic, but that one kind of feels like it. And if you've ever spent days or even weeks doing that manually by data entry, you don't have to anymore, so that is available for everyone. This is what that looks like. Again, I'll do it live, but just to show you some more interesting. So it's pulling out key value pairs. This thing goes with that thing, A and B. Whole tables. And again, you don't have to do this. Page by page or PDF by PDF. You do it on one as long as the others in your collection are structured, the same is going to do it at the same time. And then you can export that as a CSV. Then you can really start working with it, right? You can start analyzing it, looking for what's interesting, build a visualization. Whatever you then would do. So that's what the extraction looks like. Cool. It's also important to talk about privacy and security, and some of you might be bubbling up those questions in your minds right away. I see some head nods. So, everything that you were to upload into Pinpoint is private to you and you alone, unless you choose to share it. And you can share it two ways you can share it. I think of it as sideways, but like you would share a Google Doc with colleagues with a person or two, maybe with your editor, say this is where this is queued, whatever that looks like. Or you can now share it publicly and you can make things available to your audience that you've collected as a public service or something that they may want to look through on their own. So some of what I heard here might be things that are interesting for people to look through. What's this facility like, this assisted living facility, this child care facility. What incident reports that they had, restaurant inspections. Inspections in my neighborhood. So we'll look at where those live too. So again, private to you and you alone, unless you choose to share it. What that means also is that Google is not looking at them. We're not training on them. They're not showing up in search results. They're not any part of LMM research. They're just yours. If there is a file that is so close to the vest that you wouldn't email it or you wouldn't put it in Drive, you probably wouldn't put it in Pinpoint either. So just making smart decisions about what that is. But it's really important to emphasize again, they're just yours. And we want this to be a tool that's useful. And so if it wasn't secure and it wasn't private,

it wouldn't be useful. Okay, so that's collaboration you talked about generative AI. That's that brand new feature. So I'll show you this at the end, but this is what it looks like. So you can now ask questions of a document and you can, as of last week, ask questions of an entire collection. So let's say you've got 30,000 documents. What do you want to know about them? Right. Why did you collect them in the first place? You can ask what are key themes. Give me an executive summary. You can search through for a how many times this person appears or this particular phrase that's used by whomever. Lots of interesting things to do. The other way that this is distinguished, similar to our other AI products, is it will take you to where it found that answer in the documents set. So you don't have to take the AI's word for it. You get to click on the source link and it'll take you there. So all make sense. Question. Yes. So the question was, does this work for transcripts of audio and video? And the answer is yes. Cool. I want to show you the tool. We'll come back to the case studies. Any other questions at the moment before we just kind of get right in there. Oh, thank you.

Volunteer [00:29:01] Thank you. I just have two, but you can. For time, you can choose if you want to answer one. Okay. I'm curious if there's an API that you can hook into with this.

Mary Nahorniak [00:29:10] There is not.

Volunteer [00:29:10] There's not, okay. And then you said that it can handle documents if they have the same structure. Is there some flexibility if you have a bunch of documents, but perhaps the table structure is different because of change over time? Yeah.

Mary Nahorniak [00:29:21] So, if you're doing the exporting data feature and if your documents don't exactly match, it probably just won't export the ones that don't match the one that you annotate. Exactly. So you group them into a collection of but these are all the same. And that might be 30 and it might be 300. It just won't extract if the data on that one doesn't match. If data extraction is what you do and you run into some. We've heard interesting things about when stuff's redacted or what if the handwriting goes down the side of the page, please reach out. We are always looking for ways of making sure that the tool can deal with documents in the real world. Thank you for those questions.

Volunteer [00:30:05] You might get to this, but one of the challenges that we've had is, Pinpoint like individual documents to look at as opposed to like, the same report, but in the same PDF, like multiple reports in the same PDF. And so I'll just throw that out there. You might get to that or not yet.

Mary Nahorniak [00:30:24] I actually have not heard that one. That is great for me to understand. So. The PDF is many pages and it's got the same thing multiple times in there. Yeah, I would think, okay, the way to do that is probably to break apart the PDF. That's a great, that's a great use case. All right. So let's take a look in here. So again it's free. It's just for journalists and researchers. If you would like to get access to that generative AI that is a closed beta. The link is right here on the home page. This home page is just redesigned last week. So you can just click apply here. That's also in your cheat sheet. That generative AI is only

available in these countries. I just want to read them out. The US, Argentina, Brazil ,Chile, india and Mexico. So we hope that that will spread globally more soon. But those are those are where that exists right now. Pinpoint is available globally. Cool, all right. I'm just going to click in to get started. I'm going to actually log in. So the first thing I want to show you is the explore page. So even if you're not a person that's in going to upload documents or that's not the work that you do, you can still make really good use of this explore page where the public collections live. So if and if you were to publish something publicly to your audience, this is where it would go. So all kinds of interesting organizations here, you will probably recognize a lot of them. I usually just kind of scroll through it. We don't have to look at them all, but CNN Brazil, Sea Pear Chile, La Nacion, New York Times, The Washington Post, the Associated Press, all kinds of good stuff. I'm going to go to the AP because we all know and love the AP, and this is what they have made public for anyone to use. And we can take a look at it as well. So lots of newsy stuff. Federal court documents related to obscene things from the Lewiston, Maine shooting. Three sets of three, collections about the Lahaina wildfires last year. All stuff that the AP has either done public records requests or they've collected through other means and again, made public as a public service. So these are all there for anyone to use. And I have some case studies of folks just simply finding interesting stories in here. This can also be a way of getting your own collection started. Right. So maybe you're reporting on one of these things. You've got your own seven documents and you're really looking for some of the ones the AP has already found. You can go download them. I'm just going to open one of these. This is what a collection looks like. And here's this document. So I could just go in and I can download this original file. Lots of interesting things that can be done with the collections here. This is what the extraction sidebar looks like. So this is what happens anytime you put even just one document into a collection. It's going to immediately pull out all of the names. Anything that Pinpoint understands to be the same person, if it's a public figure, it will roll up together. We can't teach you that about private citizens, but it does understand that JFK, for example, is referenced as John F Kennedy, JFK, President Kennedy, all the same guy. And I have a JFK document collection, so I bring that one up to organizations. So this can be really interesting. Again, the city council example some developer, right. Some the teachers union, whatever, all the times that gets mentioned. So if I click on FEMA right here, it's going to pull up every document where FEMA is referenced. And then locations as well. So of course the locations here in Hawaii. So this is kind of what a collection looks like. Again, this is one that the AP created for us. They're also date extraction. This is relatively new as well. Anytime we release a new feature by the way for the pinpoint users that already have accounts, new features apply going forward. So if you go I don't have the dates on my collections. It's probably because you built that collection before these were extracted. But this is where that lives now. So I can go in here and I can say what documents are available from before the wildfires, which were in early August. In. The hardest part of the demo is getting the dates right, which I've already done wrong. So now what I have is any document that's from January to August of last year, August 1st, and that references FEMA. So I've started to just kind of really narrow in on what's here. What is it that I'm looking for, right? so that's what those public collections look like. And again, if you were to publish publicly, this is where they would show up. If you think that you might do that. A pro tip would be work on, so what we have to do on the back end is you will essentially say, I work for this organization, and then the team will then go make sure that that's true, so that nobody is publishing collections on

behalf of an organization they don't work for. And so that takes some time to just verify that you match with this, right? So if you want to publish something publicly, don't do it on the day that you roll out your investigation, start working on that like two weeks ahead of time. I want to also show you before I hop over to my workspace where all those language settings live. So they're up here in the settings toolbar and they're global. And I bring that up because there is a way around it. So the first question is entity display language, which means my documents are in this language. My documents are mostly in English. Audio file, spoken language. My audio and video files are in this language. So if you switch between languages, maybe you interview somebody in English. You want that transcribed. The next one you're going to upload is in Spanish, change the language here before you do the upload or the language is different. Does that make sense? So you can't say put these documents in this language, in these, in this other. But the use case of I'm working in multiple languages, you just change it before you make the next upload. These are where and these are the languages by the way. So they're about 12 to 15. It's not all the world languages. It's only the ones that we know. The transcriptions will be accurate enough to be useful. And we don't release new languages until we know that that's true. We did recently release Russian and Ukrainian for news reasons, so you'll see those in here. There are more languages in entity display because that's easier to do than an accurate transcription of audio. Make sense? Cool. All right. And then here in my workspace. This is where you build collections. And I'm going to do the data extraction in just a moment here. But, just call out. So you can see the little lock makes it clear this is mine. And if I've shared it or if someone's shared it with me, I get that little symbol. So just to really make clear what the statuses of any collection. And you can see here that I have some that my colleagues have shared with me. I'm going to actually go into this one. So this is a collection of NASA documents, largely related to the Apollo missions. Again, this collection was created a while ago, so we don't see that date extraction here. I could re-upload these, and I would have that. We get people that are named a lot: astronauts, presidents, organizations like NASA and Capcom, locations like Earth, which is always referenced in NASA transcripts. Just always love that one. And I want to show you some of those special search features, so there's a division within NASA, The Spacecraft Tracking and Data Acquisition Network. And I've searched for Stephen and it's called this photo. So the pinpoint highlighter is the green. And I think of it a little bit like the tool is being like, if you're looking for this you probably want to look at this photo, right? I've got a photo in here of that, was anybody at Nycar? Cool, the Pinpoint user. Wonderful. So, I'm going to go find it. But in the hotel in Baltimore where Nycar was, the wallpaper was principal newspapers, so I took a photo of that. Search for the word Baltimore in this collection, and it's bringing up my photo that, again, I just toss it into pinpoint. This is a picture of wallpaper, and yet I'm able to annotate it. I've searched for the word Baltimore. It's found it, this highlighter. I applied it on a different day and I can copy this whole thing and I can open up a new document. And it's all pasteable. So just handwriting text within photos making it really, really usable that it's copy and paste able. Y'all with me on that. Cool. Does anybody like I know what I need that for? All right. I'm going to go back to my NASA collection. Sorry, I just like the wallpaper thing. I'm going to show you the handwriting. I have heard from folks anecdotally that even if they aren't able to read the handwriting, if they are copying and pasting it, that Google's guess is pretty good. So if it's maybe if it's, I always say if it's legible, it can be read. Maybe if it's not legible, it can be read. I haven't tested this yet, but keep that in mind.

Ashley Edwards [00:39:18] I've tested it. I uploaded, I was doing some family research, and I was uploading, like, old voter rolls. And it did read cursive handwriting from like 1950 something. So not saying it's going to read everything, but pretty well. And these were like scanned paper documents with cursive writing on them and it understood them.

Ashley Edwards [00:39:39] That's amazing. So this is what this, this one example looks like. Pinpoint has found the word Loomis, which I searched for, and I found the word state, which the handwriting is pretty good, but it's also sideways and it's in cursive, and yet it's still readable. So I couldn't do what I need to do with this. And again, we think about things that are handwritten, whether it's archival material as actually as give an example, maybe your own notes, margin notes, maybe you walk past somebody at a hearing and you take a little photo of their notebook. It's, it can be found. Any questions on what we're seeing yet? Yes. Oh, she's bringing the mic.

Volunteer [00:40:20] My question is in terms of, for example, if we're taking it a lot of screenshots, we're doing visual investigation, right? And we are looking at emojis. Does the sidebar also filter by emoji used, for example, or items that aren't necessarily written, but they are pertinent in terms of researchers or journalists looking at the use of emojis for certain words online, to evade content moderation or things like that.

Mary Nahorniak [00:40:50] That is really interesting. I don't know the answer to that.

Ashley Edwards [00:40:53] I will follow up with the Pinpoint team because a lot of and if you start using, please leave feedback because a lot of the features that have been built were based on feedback we've gotten from the tool, but that's a really good one because I know that's a popular thing on social media, on Instagram and TikTok, to use emojis to replace words to get past like the content filtering. So I will definitely follow up with that.

Mary Nahorniak [00:41:16] So what it won't do to the first part of your question is it won't break those out in the sidebar because they're not names, organizations, or places. Those are the things that breaks out on its own in the analysis. But the other way that you can use Pinpoint is to search for something, a particular phrase, block of text, a word. And so I think that's probably something that's easily testable is a photo with some emoji on it. Search through for that emoji, see if it comes up I don't know. That is a great question and thank you for the use case. I want to show y'all. We only have five minutes left, so I want to show you the data extraction and the generative AI. So I'm just going to do this one from the beginning. So I'm going to start with a new private collection. This is what it looks like just to build one and I'm going to bring in our documents. So these are all truly exciting files, which you'll see when I open them up. Which are, you know, when you search for a country on Google and you get like the CIA Country Factbook. That's what these are. So they're like the second result that comes up. So there's nothing newsy in them at all. But they are all structured the same. They're being uploaded right now. They're coming in. You can see the analysis kind of happening just as we go. So I have ten of them. I might not have actually highlighted them all. Oh. There they are. Okay, so that's how fast it uploads. I got my ten documents and then the thing that I'm going to do here is this beta. This is a public beta, everyone with a Pinpoint account has this. Now extract structured data. So

I'm telling Pinpoint. Yes, all of these are things I want to extract and extract the same. So it's going to just work on understanding what's in the documents. This also only takes a minute or two. And then we're going to annotate one of the documents. If you are using this and I'm thinking of this person's question from earlier. If you begin by annotating one document and you later realize there's a better one, that would be what we call the golden document. You can just switch that so you don't lose hours of annotation or extraction. You can just kind of work within. You don't have to start again from scratch. So I'm starting by annotating. It chooses a document at random, every time I do this, it does it differently. But we've got US right here. So I'm just going to start by drawing boxes around stuff, which again, is fun to watch other people do, I know. And it's it's working on understanding. I'm only annotating the value and it thinks that this is the key. So I'm just going to fix it and say no it goes with country report. So there's lots you can do here. And I'm just going to add a few more in. I'm drawing the boxes all the way to the right, because some of these documents have words that just run longer than these do. Like, the word Afghanistan probably takes up more space than United States. And so what's happened here? Just as I've done that, I'm going to bring up the bottom bar, if it's taken all of that similar information off of all the documents at the same time. So they're all right here. And so again, the information is what it is. But I didn't have to annotate ten documents, I only had to do one. And it's pulling it off at the same time. I'm going to also do a table and then we'll extract it and take a look at it. So I'm going to do this table called Largest Cities because that's what it is and I'm going to give it a name. And this table is bigger than it looks here. And so I have lots of options as I go. But I'm just going to make it bigger. I'm going to save it and it's going to do all of that for these ten PDFs at the same time. So if I click on my largest cities tab here, the largest cities in Afghanistan. Brazil, etc. which it all pulled off just as I was working on the one. And so when this is in good enough shape, you have those options like clear your annotations, Start again, reprocess, etc. they're all here. I can just extract this. This is what we saw on the little gif on the slides. I'm downloading it. They go into a zip file which I will unzip. And then I'm just going to bring those into a sheet. I'm just going to bring one of them in. And now I'm going really fast, so I'll tell you all that. In the cheat sheet. There's a link to the video for how to do this specifically. Oh, hang on. I've gone to the wrong file. I can't talk and do this at the same times. Every time. ISOJ, that is what we're looking for. Here's my CSV that just extracted. I'm going to bring all this in. So here's what we got. All that data is now in workable format. Everything we pulled off of those PDFs, which we didn't actually even pull it off of. Afghanistan pinpointed that we pulled it off of the US. And so I can. I can work with this, right? I can now start sorting it. I can make sense of it. What I really, really like is this validation link. This is going to take you to the exact point in your collection where any of this information came from. So you can always go look at it in context, especially if you go that one looks weird. I think it's missing a zero or, you know, things that stick out or I just now I know where my story is, right? I'm going to go look at everything that surrounds that. So the validation link makes a lot of sense. Okay. We are at time. So I'm going to lastly just show you that generative AI feature. And then I'll tell you that Ashley and I will be here all afternoon. So we're happy to take your questions. So these are also the very newsy Federalist papers, just a few hundred years old. And so I'm asking what are key themes? In this collection. And so this is a flyout sidebar. It's just on top of my extraction sidebar that's still right here. Once you're in the beta, you will have this. This is where it is, where it works. And so this, LLM has given me what it thinks are the key themes of the Federalist Papers

and if they're underlined in purple, I can go exactly to where I got that information from because I have show sources turned on. So I'm just going to click on the first one. And it's taking me to here's where we got this. Here's what we think The Federalist Papers are about. So we can go check it out for ourselves. And I can open any one of these and learn a little bit more about it. I can also ask questions of just the paper. So I'm going to ask who is the author, even though we can see it right there. Sometimes that's buried in 56 pages of Senate testimony. So that's working on it. If I can also share that this is now switched from me talking, asking questions of a collection to asking questions of a document so it's just calling that out that I've done that and it says, oh, the author is John Jay. Here's where we got this information. So I think the possibilities for this are really interesting. The collection level Gen AI query is brand new. And so I would love to know if you do end up using this how you use it, what's helpful about it. I'm only thinking of questions like what are key themes and who is the author? But when you're working with live documents and you're doing reporting, the questions you're going to come up with, they're going to be really, really interesting. Okay. That's all the time that we have. This is a survey for what you thought of the training. If you wouldn't mind taking it. We would appreciate it. And I'll bring up the cheat sheet QR code in about two minutes when I see these phones go down.

Ashley Edwards [00:49:01] Yes, please give us your feedback. Be as honest as you need. We don't get easily offended. It's fine. So.

Mary Nahorniak [00:49:09] Yeah. How's that? Nope. Oh, it's because it's in Zoom. Let me do this. How about that? Need a little space at the bottom. Okay. So this is the survey. And again I'll bring up the cheat sheet QR code in just a minute or two. Thanks everyone. I really appreciate you spending your lunch hour with us. QR code is on the screen for the cheat sheet.