An Exclusive Look at 3Play's AI-Enabled Localization Platform

[00:00:00.00] SOFIA LEIVA: Thank you for joining today's session, Experience the Future First, An Exclusive Look at 3Play's AI-Enabled Localization Platform. My name is Sofia Leiva, and I'll be moderating today's webinar. With that taken care of, I'll pass it off to Chris Antunes and Erik Ducker, who will talk today about 3Play's localization solutions.

[00:00:24.56] CHRIS ANTUNES: So I'm Chris Antunes. I'm one of the co-founders of 3Play Media. And Erik and I want to talk today about this next generation of localization solutions we're bringing to market and also these global accessibility solutions. I think before we dive in, we want to take a minute to rewind the clock and talk about where we came from. So back to the origin story of 3Play, kind of where we started.

[00:00:52.14] So Act 1, Laying the Foundation. So we got started in the education segment while we were at MIT, so myself and three other co-founders, solving a very specific problem for a group called OpenCourseWare at MIT. They were publishing thousands of hours of courses online, and they needed those courses made accessible. This was a really interesting problem because quality was non-negotiable, right? But so was budget. They couldn't afford to go pay the rates of existing captioning solutions to be able to caption that massive archive.

[00:01:26.31] So exuberance of youth, we set out, myself and our other co-founders, to solve that problem with MIT. And we thought that AI, at the time referred to as automatic speech recognition, could be a really good start to help us, again, solve this captioning, kind of budget-constrained problem that MIT was facing. It turned out that AI wasn't good enough even 15 years ago, and still isn't most of the time today, but it really was a great start. So allowed us to drive down the time that a captioner and expert would need to spend generating that caption file and also drive down cost ultimately to our customers.

[00:02:13.53] Why I want to start here is I think it's important to note that AI plus tools that we've built and experts have been at the core of 3Play since day one. So this isn't something we had to retrofit now because AI is all the rage. And I think as importantly, it's not something that we need to for a tech forward-- or a tech-only solution, we're not in the position where we need to scramble to figure out how to supervise and ensure quality with experts. From day one, AI responsibly used, plus people, plus tools. But again, first focused on the education segment.

[00:02:55.30] So we can go on to the next act here. So Act 2, Growth. So this chapter of 3Play's history was focused on both widening our product footprint-- so not just closed captioning, what we'd call recorded or captioning for on demand video-- but also live captioning and also audio description and accommodation for blind and low vision users. It was also a time where we were widening our footprint marketwise. So instead of-- beyond just focusing on higher ed or e-learning where we started, we expanded our focus areas to corporate media and entertainment as well.

[00:03:38.95] But in all three of these end markets and for all of these solutions, a common through line was still quality is non-negotiable. But quality can vary for different use cases and also can vary depending on the real world budgets that customers faced. But a few examples here of the scale we're talking about and the different end markets range from helping an e-commerce platform audio describe hundreds of thousands of videos on a short project timeline, under close regulatory supervision, to a partnership with NBC Sports for the Olympics. We were captioning thousands of video clips within hours for them. Very fast turnaround, significant scale, high visibility.

[00:04:26.45] And over on the higher ed side, supporting hundreds of students live in the classroom with high quality on demand captioning when they need it. But across all of these, same theme-- and I'm going to say this many times today-- AI, plus tools, plus experts, at scale. But in this second act, still, the business was primarily focused on accessibility and primarily focused only on English and Spanish video, which for our largely North American based customers was where their accessibility requirements largely lie.

[00:05:06.37] So we can move to the next slide. Act 3, Shaping the Future. So this is what we really want to focus on today, which is widening the aperture at 3Play, still focused on accessibility but globally. What this means is accessibility, so still captioning and audio description, but in any language, for any region, really for any enterprise customer. And then also widening the aperture product wise to include localization, so subtitling and dubbing.

[00:05:44.14] But again, the through line will remain the same, high quality, customized, enterprise scale, fast turnaround within a customer's budget. You can go to the next slide.

[00:06:04.14] So as we start to shift our focus here, as we have shifted our focus here towards localization, a common question we're asked is why? Why now? Video subtitling and dubbing are mature industries. They've been around for years. And 3Play has had an offering in the subtitling and dubbing space for many years as well. So why now? Why try to innovate in this area? Why try to bring new product, new scale, specifically here?

[00:06:41.67] So we think there's three primary reasons. So the first, Global Monetization. We see a trend across-- and you have to remember, we serve a wide range of end markets, and we see many, many use cases around video, spanning higher ed, corporate, which is a really wide use end market and media and entertainment. And consistently across all of those end markets, across all of those use cases, we see a trend that they are trying to reach a global audience, right? Trying to get their content out to market globally.

[00:07:19.47] And this ranges from an e-learning use case to an e-fitness use case to a social media content creator, all the way out to high end traditional media and entertainment production, both on demand and near live. The challenge has always been these workflows are costly, they're time intensive, they're complex. And price, very specifically, or cost, has been a blocker to many of these use cases really taking off.

[00:07:57.02] So we can get to the second bullet here. So Reduced Barriers. In the last few years, we've seen innovation on the technology side, certainly in machine translation, so in support of

subtitling, but real breakthrough capabilities in technology on the voice technology side, really supporting audio description and also dubbing, specifically talking about localization.

[00:08:25.10] While we believe each of these services still require close human supervision at many steps in the process, these technologies allow time on tasks to come down, allow costs to come down, and maybe most notably, remove, in the case of dubbing, the need for significant capital expenditures, which again, drive up cost. This has reduced the barrier of entry. We've seen the prices reduced, which again allow many different emerging use cases to start to present. People who just couldn't justify this from a Return On Investment, or ROI perspective, now can. And that's a really exciting time.

[00:09:08.37] Third, and this brings us back to accessibility, still our core capability over the last 15 years, the European Accessibility Act. And Erik will talk more about this in a second, but there is regulation coming at the end of June of this year that is requiring many of our customers who produce or distribute content in Europe today, or in the EU today, to make that content accessible, meaning captioned, in our world, and audio described.

[00:09:43.20] Multiple different profiles here, use cases here, but one I'll call out specifically, just to give the group a feel for the scope of this challenge and the urgency is maybe a traditional media company that has been, as a global strategy, producing content, and let's say dubbing it, into many, many languages in Europe for years. Starting at the end of June, that content that is dubbed in many languages and distributed in these countries will be required to be made accessible, meaning captioned and described.

[00:10:25.42] This is a significant challenge on a really tight turnaround. This is 3Play's bread and butter, is figuring out ways to meet the quality required, but do it in a way that can fit into your budget, where we have to use technology, whether it's tools or AI or anything else, responsibly, but the quality is non-negotiable. So how do we solve that problem together? And this is a really exciting time for innovation.

[00:10:53.12] Regulatory moments like this almost always inspire innovation. You have to do something difficult that's never done before under a tight timeline, and we're grateful that many of our customers today are partnering with us right now in real time to solve these problems in, I think, a repeatable and scalable way that will be accessible to all of our customers moving forward.

[00:11:16.33] So next, so we have a little more detail on the European Accessibility Act. Erik, do you want to go into anything more there? Do we want to keep moving?

[00:11:27.02] ERIK DUCKER: I think it's just a call out to we have some additional content that allows you to dive deeper into the European Accessibility Act, if you're not familiar. But the key date is June 28, 2025. Changes will be happening from an enforcement perspective around this regulation, which has been kind of in a preview state for many years. And we're just kind of in the act now phase. And so if you are interested in learning more or discussing how this might impact your business, Chris and I are very available to follow up one on one. But we can go back to why 3Play?

[00:12:06.18] CHRIS ANTUNES: Yeah, so second question we asked ourselves after why now. So first, just covered those three. And there's many more reasons why we're focused on localization and making our accessibility solutions global, but I just covered the big three. But that isn't sufficient for us to want to spend product resources, to spend time at 3Play really focused on this challenge. Certainly, our customers are thinking about it.

[00:12:28.74] But we want to make sure also that 3Play can do something different, that if we're going to bring a new product to market, if we're going to innovate,, it better be towards something fundamentally different, which means fundamentally changing our customer's localization experience in some meaningful way, hopefully many meaningful ways,

[00:12:54.62] But that can mean more scale, right? That can mean more streamlined workflows. That can mean lower cost, of course. That can mean higher quality, right? That can mean more control of all of those. One of the, I think, the things that has been resonating recently with customers, both on the localization side and on the accessibility side, with all of, let's say, the AI hype and the zeitgeist, is it's not binary, right? It's more nuanced than that.

[00:13:31.16] It's really around how to use technology responsibly. And what I mean by that is it is totally possible, because of budget or other customer metrics that they are focused on, you may want to treat an event like the Super Bowl differently than a video in a back catalog that got five views last year. And that's OK, right?

[00:14:01.45] So we think about within the constraints of a real world budget, how do we maximize the solutions we can deliver both the customer, but to the end users in both the accessibility and the localization space. How do we give the best experience to someone who's trying to access this content, who otherwise couldn't understand it, subject to a real world budget?

[00:14:26.97] So we're going to focus on three core components, though, that we think make 3Play really special. The first, and I've hit on this a number of times, which is this AI plus expert collaboration for every single video, and I sort of just covered that. The second, effortless scale for any project. So it's not sufficient to be able to, let's say, optimize this blend of professional expert time and technology and AI usage for a single video.

[00:15:01.42] In the real world, our customers have thousands and thousands of videos, and they send them to us every day, and they need to meet their deadlines. And what the 3Play tools, combined with our global marketplace of experts all over the world, provide is Effortless Scale for Any Project, or said another way, on demand nearly limitless scale.

[00:15:25.28] One of our North Stars from the beginning, in the accessibility space that we are bringing to localization, is that this solution should scale, should give you the scale like a pure technology, right? If you need to order 100 hours or 1,000 hours of these things, and you don't give us weeks of notice, we ought to be able to deliver on that. But that scale, also with expert professional humans in the loop giving you that peace of mind and meeting that quality that you require. So scale of the robots, but care that you only get with people supervising at the wheel all along.

[00:16:08.87] And then Customer-Centric Integration. What this means in our world is that captioning, description, subtitling, dubbing, these things shouldn't be hard, right? They really shouldn't require teams of people uploading and downloading and talking about standards and project managing every step of the way. These things ought to be easy. And in our language, what that means is we're going to meet you where your video is.

[00:16:39.31] So we've invested again, from day one, across all these end markets, with all of these different use cases, we've invested in meeting our customers in their workflow where their videos are. So if you have one or you have 10,000 videos, we want to be integrated into that system, whether it's a video platform, like a Brightcove or a YouTube, Wistia, Kaltura, et cetera, or it's a Espera or S3 or Google Cloud. Wherever that video resides, we have integrations that build in seamlessly so we can get the video and ultimately deliver the accessibility or localization assets back without a lot of fuss.

[00:17:18.91] We think this is essential to scale. We think this is essential to meeting our customers' requirements when we think about workflow. When we start the conversation and we talk about quality, I always want to include this. It's the quality of the whole experience end to end. Accuracy of the captions and accuracy of the subtitles and dubs is essential and non-negotiable. But quality doesn't stop there. It has to extend beyond that.

[00:17:50.97] So now, Erik is going to go through a more detailed demo of some of the tools we've built around these new localization products. So, Erik, take it away.

[00:18:02.97] ERIK DUCKER: Awesome. Thank you, Chris. We'll give you a little bit of a breather. I'm always excited to give this demo because this is not the demo I give to customers. The customer demo is all about the ordering interface, that customer centric integration. Today's demo is all about the things that you don't get to see at 3Play. So this is where our exclusive look really comes into play here, at the end of the day, is understanding how a file that comes into our system, gets all of the treatment it needs to exit the system and come back to your proprietary video platform or Amazon S3 bucket. We're going to show you the steps that go into that.

[00:18:41.24] So to get started, we need a video file. And that video file is going to be a sample from a Spanish comedy or a Spanish horror film from very long ago. So the audio quality is actually not very good to start with. We'll keep that in mind as we go through this demo. But ultimately, we're going to just play back the demo video, and then we'll go from there.

[00:19:05.61] [VIDEO PLAYBACK]

[00:19:06.55] [SPEAKING SPANISH]

[00:19:18.58] - [LAUGHS]

[00:19:19.45] [SPEAKING SPANISH]

[00:19:34.41] [END PLAYBACK]

[00:19:35.22] ERIK DUCKER: Great. So at this point, we need to close caption this file. That's the first part of this order. And how we do that is we use AI as a first pass. And then, we use our global marketplace to edit and review and QA those captions for whatever spec you need. Whether you're delivering to a high end media streaming platform or your personal website, we're going to be any spec that you need for closed captioning.

[00:20:03.60] And what we're really getting here is a couple of really, really important pieces of data. We're not just getting the words that are being transcribed, we're getting the speaker labels, we're getting the timing, the word level timing. And all of that data is really important for us to be able to produce subtitles, dubs, and audio description tracks later in this process.

[00:20:25.61] The other thing I want to note is that our use of AI is exclusively focused on making sure that the humans are more efficient. We take a very human editing, human review perspective first, and we layer in AI and automation along the way very, very intentionally to make the human editing process more seamless. We don't just use AI willy nilly. We're really, really focused on making sure that we're using it for the right problems that we're running into in the process, specifically those bottlenecks that we might run into if we were to rely on more human scale, and so on, and so forth.

[00:21:02.07] So once we have this closed caption file, we can move to subtitling. And our subtitling process, you're going to figure out a pattern here. We're going to first start with a machine translation pass. And this machine translation pass is not just doing a pure output of machine translation. Our customers can actually give us glossaries or a note, a do not transcribe translate list, and so on.

[00:21:27.17] We've also implemented context aware translation. So we're using some of our audio description technology to actually apply to the translation process. So we're able to understand some of the key elements of the video. So when we're translating from English to a romance language where there's gendered language, we can identify that gendered language earlier in the process and use AI to auto edit some of those characters, so that the human editors are spending less time on mundane tasks and really focused on the cultural nuance of localization.

[00:21:57.82] So when we're creating the subtitle file, once the AI step is done, this is going to land in our subtitling editor. And the subtitling editor is proprietary to our contractor or our workforce platform. And this is really where they're going to do all of their work for editing subtitles. You can do everything from editing the text, obviously, but also, it's going to alert you on all of the potential errors in terms of the output based on the needs of the customer.

[00:22:24.16] So if there are specific character count per line that we need to adhere to for that customer, like the defaults, they're going to be alerted to that, and they're going to figure out how to truncate that translation or separate that translation to make sure that it keeps pace with the original video. So there's a lot of tooling built into here, such as like keyboard shortcuts. So as our contractors get more familiar in this environment, they're going to spend less and less time because we're actually providing a space that allows them to grow from an efficiency standpoint, not just from a linguistic standpoint.

[00:22:57.80] So there's a lot of things that we're really paying attention to in terms of measurement and how we track progress, and then finding bottlenecks and solving for them iteratively through our process. So the subtitler editor can come-- the subtitle editing linguist will be done with their work. And we're going to post-process this into all the variety of formats that the customer needs automatically. So whether it's a WebVTT file, an SCC file, an SRT file, those will all be produced automatically.

[00:23:26.59] So we have our subtitle file and now we need a dub file. You might be thinking, how can we just use the subtitle file and put a voiceover? Well, that doesn't always work because you have different constraints with dubbing. And to just give you a quick overview of how our dubbing process works, it's really three large buckets. There's a script preparation step, which we basically covered in our closed captioning step. We have everything kind of ready, down to word level timing speaker labels, so that we can attach voices to the right speakers at the right time.

[00:23:56.88] But we've added this machine translation step. We're going to use the same kind of methodology of making sure that the machine translations incorporate as much customer specifics as possible, or video specifics in that process. And then we're going to do our audio processing step. And this is where we're going to use AI to extract the background audio tracks so that we just have isolated speech tracks at the end of the day.

[00:24:19.80] Then we have our voice matching service or voice matching technology, which is going to match a voice, an unlimited number of voices effectively, to each speaker in the video. And that voice match is going to be closely related to that original voice, but it's not going to be cloned in the sense that there's a unique model created for that voice in that sample.

[00:24:41.96] Then, finally, we're going to add in one last layer of AI review. And that review is once again going to be really focused on truncating translations that are too long or too short. So English to Spanish or Spanish to German, there's going to be differences of how things are presented for the same translation in terms of character count. And so that's going to have a really, really strong correlation to how long it takes to pronounce or to deliver that line. And we're constrained in the audio segment timeline to make sure that the translations match those times. So AI is going to suggest these edits. And I'm going to show you an example of that in just a second.

[00:25:20.84] And then we have our professional step, which is really going in and reviewing the translations. And then we're going to have an audio QA step, which is assuming the translations are good, we're going to make sure that the audio quality is really high. And then we're going to post-process and mix all that together into video formats or audio formats as necessary. So let's go through a couple examples of common edits that we run into.

[00:25:45.31] The first one is the one that I mentioned around time constraints. So here's an example of how AI is recommending an actual shortened version of the translation. And the editor gets to choose to use that to make sure that it's maintaining context. We're not automatically doing this because that could result in missed context. And so we want to expose that to the editor, but it's greatly reducing their time on task in the sense of making sure that they're not having to recreate the wheel every single time.

[00:26:13.82] So that's a really simple example, the translation editing step. Obviously, there's other similar tools to the subtitling to make sure that you can edit and make all of the output translation of high quality for the customer. And then we go to audio QA. Sorry, one second. Audio QA. We're going to show two examples of audio QA that are really important.

[00:26:35.27] The first is, how do we retain audio that we actually want, speech that we actually want? So for example, Scooby-Doo's voice. How do we make sure that we retain Scooby-Doo's voice? We just extracted speech from everything else, including Scooby-Doo. So instead, we can actually retain some of those snippets back into the mix track. And I'll show you how we do that with this sample.

[00:26:57.18] [VIDEO PLAYBACK]

[00:26:57.94] - Look at the one you sing, this one-- Soledad. Ta-ta-ta-ta-ta-ta-ta. Damn, what do you think?

[00:27:04.73] ERIK DUCKER: The robot is not great at singing yet. So we're going to go fix that.

[00:27:10.81] - [SPEAKING SPANISH] Ta-ta-ta-ta-ta-ta.

[00:27:13.87] [END PLAYBACK]

[00:27:14.85] ERIK DUCKER: So we want to keep that original because it's not words being said, so there's no additional meaning if we just retain that audio. And since we're using a voice match, it's going to not be a jarring experience for the end viewer. So we're going to add that back in, and we're going to duck the audio of the output dub so that you can't hear that. And then we're going to just blend it back in.

[00:27:35.43] [VIDEO PLAYBACK]

[00:27:37.10] - This one, Soledad. Soledad. Ta-ta-ta-ta-ta. Maldita--

[00:27:42.08] - What do you think?

[00:27:43.14] - Hey, you have--

[00:27:44.24] ERIK DUCKER: Great. So that's the first example. The next audio example is delivery sometimes can be a little bit jarring, especially if you need to slow down your speech to deliver. So we're going to show how we can fix that problem with our speech-to-speech technology with our editors.

[00:28:03.69] [VIDEO PLAYBACK]

[00:28:04.50] - Well a coffee, a trip dancing, anything.

[00:28:09.52] - Very kind, but I can't accept it.

[00:28:11.61] [END PLAYBACK]

[00:28:13.78] ERIK DUCKER: Great. So I'm going to go see what the Spanish sounds like to get an idea.

[00:28:19.48] [VIDEO PLAYBACK]

[00:28:19.94] - [SPEAKING SPANISH]

[00:28:23.12] [END PLAYBACK]

[00:28:24.27] ERIK DUCKER: Great. I'm going to try to mimic that delivery in English, using our speech-to-speech interface.

[00:28:34.96] [AUDIO PLAYBACK]

[00:28:36.45] - Well, a coffee, a trip dancing, anything.

[00:28:40.68]

[00:28:41.43] ERIK DUCKER: Great. I'm not a voice actor. I'm a product marketer.

[00:28:43.95] - Well, a coffee, a trip dancing, anything.

[00:28:46.71] [END PLAYBACK]

[00:28:48.80] ERIK DUCKER: It's going to resynthesize, and then we can play it back.

[00:28:51.84] [AUDIO PLAYBACK]

[00:28:52.53] - Well, a coffee, a trip dancing, anything.

[00:28:57.33] ERIK DUCKER: Great.

[00:28:57.83] - Very kind, but I can't accept it.

[00:28:59.19] ERIK DUCKER: It's done a great job of mimicking that voice. And so this is really powerful for pronunciations or deliveries. This is not meant to replace voice actors in any way. This is really focused on how do we make sure that we fix some of these potential jarring issues in the audio QA, and that's really what our editors are focused on in this step.

[00:29:19.74] So now, once again, we have our dub file. Because of the European Accessibility Act that we just mentioned, we have now two audio tracks. We have a Spanish and an English audio track. Instead of finding writers for both English and Spanish, we're actually just going to

write the English description track once. So we're going to once again use our AI tools to provide a first pass of the audio description. And optionally, we can send that to our marketplace-- or sorry, to the-- yeah, to our marketplace for further description with human review.

[00:29:52.06] In this sample that I'm going to show you is just a pure AI output. There is no human review in the description. And the only other note is I truncated the audio segment, or the video segment for time because there's a big gap with no audio description because of dialogue.

[00:30:10.01] [VIDEO PLAYBACK]

[00:30:10.94] - Well, I know you have a lot of commitments and maybe that's why we can't see each other. But I want you to take it with you. That way you'll make me feel good.

[00:30:17.93] - He gives her flowers.

[00:30:19.34] - Well, thank you very much. Look, just so-- you sure are well informed. Well, I have to go now because my little sister is waiting for me in the car.

[00:30:25.77] - Hey, but what time are you going and where?

[00:30:27.57] - At the Central Studios at 8:00 AM. Thank you. Goodbye.

[00:30:30.35] - The woman leaves the shop, gets in her car.

[00:30:34.51] [END PLAYBACK]

[00:30:35.47] ERIK DUCKER: Great. So at this point, we have an English audio description track. Instead of once again having a Spanish audio describer who might describe things slightly differently because it's a different person, we're just going to translate that. We're going to put the audio description back into our audio description translation tool, which I won't show because it just looks like our dubbing tool. That is going to then get translated, and we're going to make sure that the timings work in case there were any new timings that we need to be worried about with the dubbed audio versus the original audio.

[00:31:06.08] So at this point, we've now delivered closed captions, subtitles, dubs and audio description in two different tracks, and the customer has never talked to us since ordering. So at this point, all of these files get delivered to the customer, and we do this for every single file that our customers bring to us. So at this point, I want to talk about what's next, and I'm going to hand it back over to Sofia and we'll have Q&A.

[00:31:34.22] SOFIA LEIVA: Great, thank you very much. And thank you, Chris and Erik, for a great and informative presentation. Before we dive into the Q&A, if you're interested in continuing the conversation beyond today, we'd love to connect. If you're looking to dive deeper into your localization strategy or want to see how 3Play can support your specific needs, we'd love to chat. We're happy to explore how we can be the right partner for you. And if you're

producing or distributing video content in the EU and want guidance on how the EAA might apply to your team or business, we've got you covered.

[00:32:09.24] You can use the QR code on the screen or the link in the chat to fill out the form to get in touch with our team. So now, we're going to transition into our Q&A and answer some of the questions you submitted during the session. Feel free to keep adding the questions in the chat as we go, and we'll try to get to as many as we can. The first question we had here is, what services can be ordered on the localization platform? Does it cover dubbing, subtitling, and transcription all in one place?

[00:32:41.48] ERIK DUCKER: Yeah, I'll take that one, Chris. It's softball. [LAUGHS] Absolutely. So what I just walked you through is kind of an illustration of how things work in the back end. But ultimately, our customer facing platform makes it really easy for you to drop a file in and order all of these services in a checkbox like manner and with associated instructions for one or 1,000 files. And that can happen automatically. You don't have to order sequentially. We'll take care of the sequencing for you on our side.

[00:33:16.63] SOFIA LEIVA: Perfect. Thank you. The next question we had here is, are Arabic or Hebrew languages supported?

[00:33:26.52] CHRIS ANTUNES: Erik, before-- sorry, just before we go into that one, just to build on what you said about the ordering on the platform. I just wanted to underline also that these services, ultimately, will be ordered in any way that we offer, you can order captions with us today, which largely means API integrated with video platforms, old school, like FTP transfer, whatever it is. Really, the headline is all these services are available in one place, and we can access the video, place orders in a hundred different ways depending on specific workflow needs.

[00:34:02.37] ERIK DUCKER: Yeah. Chris, I'm going to let you take the second question, though.

[00:34:07.44] CHRIS ANTUNES: Yeah, so the question was specifically about Arabic and Hebrew. So I assume that question is probably targeted towards dubbing. Something we hear often is about the relatively weak voice technology right now, in Hebrew specifically. The short answer here is for subtitling, yes and yes, Arabic and Hebrew. For dubbing, yes, Arabic, and Hebrew, coming. This is really a limitation of the voice technology. There's some models out there. They're not as, we'll say, well, good as we'd like them to be right now, but they're improving pretty rapidly.

[00:34:50.09] So that's another thing we should mention. When we think about how we use AI and how we use these technologies, and I think there's maybe a few other questions that touch on this, we are sort of by design-- feature, not bug here-- sort of model agnostic. And what I mean by that is we want to be able to monitor at the edge the best technology for every one of our customer's use case, and even video to video.

[00:35:17.28] Like, if certain engines are better at certain languages, we want to be able to plug and play very easily over to that language. And these technologies are evolving rapidly. So month to month, certainly, quarter to quarter, the best varies. And again, we want to be an integrator and agnostic in that area and not hitch our star to one single provider in that area. So on the Hebrew side, I'd say coming soon.

[00:35:50.75] SOFIA LEIVA: Great, thank you.

[00:35:51.53] CHRIS ANTUNES: Anything to add, Erik?

[00:35:55.10] ERIK DUCKER: I guess the only thing is, and just to go back to the accessibility world, we take this very seriously. Every single year, we publish a state of ASR report, and we publicize that publicly across a dozen different ASR engines. And that's one to help the market decide what ASR engines are best out there for an accessibility use case, but also for us to continue to check in and make sure that we're providing the best valuable option for you, because those, as AI gets better, we want to make sure that we're sharing those gains with our customers as well, who are expecting the 99 plus percent quality over time.

[00:36:36.35] SOFIA LEIVA: Great, thank you so much. The next question we had here is what measures do you take to ensure consistent quality across the different languages and services?

[00:36:47.69] CHRIS ANTUNES: Yeah, so generally, our entire ecosystem, and this is true in captioning and description today and is true also across all these languages, is we have multiple levels of review for each video that passes through our system. So in addition to maybe an AI pass, we'll have a skilled editor and then a skilled QAer, or Quality Assurance expert, review the file. So we make sure that every single video that match-- or every service that matches through our system gets the attention it requires.

[00:37:26.67] So that's what we call synchronous quality management. So in other words, a video is uploaded, we want to make sure it gets multiple eyes on it in most cases, before it goes back to the customer. But also, asynchronously, we have a process to continue to evaluate all of our experts, all of our captioners, all of our subtitlers, all of our dubbers, to make sure that they continue to meet the 3Play high standard that we've set, and give them feedback so they can improve on the tools and ultimately score and manage that system. So we both have synchronous, call it on the production line, checks and balances, but also asynchronous checks as well.

[00:38:08.95] SOFIA LEIVA: Great, thank you very much. We have a couple of questions on our voice technology. Does the AI take varying tones of the actors into account when processing a dub file?

[00:38:23.86] CHRIS ANTUNES: Ducker, do you want to grab this one?

[00:38:25.52] ERIK DUCKER: Yeah, so there's a few things that we're doing. One, when we're doing a voice match, we're taking multiple samples across the video file to identify what is the best possible match. So it's not just like a three second clip at the very beginning for the first

time. We are really trying to sample and identify the best possible voice across the potential range that that voice is going to have.

[00:38:51.31] I think on the delivery side, we kind discussed a few of the tools that we provide in terms of making sure that we can enact tone and delivery cadence in that speech. So in general, we also have additional tooling from using the ARPAbet, which allows us to spell out things using phonetic spellings of words. So there's a number of ways in which we can manipulate how the synthesized voice is going to deliver that content or those words.

[00:39:29.23] At the end of the day, these are AI voices, and so it's going to interpret things how it wants to interpret. And our job on the human editing side is to override things where it was misinterpreted at the end of the day. And that's where it's so important to have that human in the loop review aspect along the way in those key moments.

[00:39:48.17] CHRIS ANTUNES: Yeah. And I saw another question in the chat that's related that maybe we can hit now, which is around, how do you get emotion? How do you inject the acting into this process? I mean, obviously, Erik demoed this, kind of what we'd call speech-to-speech workflow. So in the extreme case, someone could do that for the entire video, right? That would be very close to what a traditional voice actor does today.

[00:40:15.87] So a professional voice actor could go through every single piece of speech, right? What would that do? Well, it'd dramatically increase costs. So the reality is you can calibrate that to taste. And that comes with higher price. And obviously, we didn't talk about cost and price specifically here, but one thing I'm finding pretty interesting in this space in particular, where we were starting in some languages at really high per minute rates with traditional, we'll call it dubbing, techniques in the hundreds of dollars per minute for brick and mortar studio traditional voice acting.

[00:40:59.43] Well, there's a huge difference in price out there between that, a few hundred dollars a minute, and some of the kind of tech-only solutions and the maybe under $1 a minute. So in between, there's a lot of room for nuance. There's a lot of room to dial up or down the expert time needed or involved in the process. And so we might see for a very, let's say, more budget constrained e-learning customer, who has a single speaker, it's not meant to carry a lot of emotion, it's instructional content, you don't need to bake a lot of that into the process, and the price can be much lower. But for some of the higher end media content, maybe you have to drift closer to more traditional acting methods.

[00:41:47.19] SOFIA LEIVA: Great, thank you very much. The next question we have is, will these new features be able to pick up on multiple languages being spoken in a film? For example, if you have a film with both Spanish and English, can the system detect when the language changes?

[00:42:04.90] CHRIS ANTUNES: So it's a little nuanced, right? The short answer is yes, many of the technologies today that do the voice part of this can do exactly that. So from a machine translation perspective, you can both take the original source that's multi-language and you can translate it into your target language. So if there's a little English and a little Spanish, and you

ultimately want it translated into, let's say, German, and dubbed into German, you can do that. And some of the technologies also can do that quite well on the voice generation side. So ultimately taking those translations and generating speech.

[00:42:45.19] The nuance is that we still believe, in almost all cases, you need experts in the loop, right? So you need different people with different skills to review those translations. Let's say English to German is a different person often than Spanish to German. So there's still kind of a marketplace and kind of expert-to-task matching problem. We're well equipped to solve that. I would say that is in something we see often enough that it's particularly scaled now.

[00:43:17.59] But would love to talk more about that use case, if it's a common one, for anyone on this call and think about how we could again scale that, call it marketplace problem. But from a tech perspective, I think that problem is pretty well solved.

[00:43:33.10] SOFIA LEIVA: Great, thank you. The next question we have is, is there anything regarding localization, for example, in Spanish from other Latin American countries, due to language variations?

[00:43:48.91] ERIK DUCKER: Sofia, maybe help me decipher that. Are you saying-- or is the question around different dialects of Latin American Spanish and-- OK.

[00:43:59.39] SOFIA LEIVA: I believe--

[00:44:00.46] ERIK DUCKER: Say it again.

[00:44:01.22] SOFIA LEIVA: Yes, I believe they're asking how we handle localization of different dialects.

[00:44:06.08] ERIK DUCKER: Yeah. No, absolutely. So there's a couple similar, probably, responses that you've heard. So on the machine translation step, it's not probably going to be as specific to say a country or a region within a country in terms of dialect support. They're going to have a general like, Castilian Spanish and Latin American Spanish generally.

[00:44:35.12] And then once you get down to the country level, it comes back to the human review. In a scaled global marketplace, you'll have reviewers in every single country. But it comes back to what's your turnaround? If you need resources in one specific country for one specific dialect, it might be a little bit more challenging to get to that scale versus taking a more agnostic approach to the dialect output, where you can scale across multiple countries that might have those resources for Latin American translation. Chris, anything to add?

[00:45:08.32] CHRIS ANTUNES: Yeah, no. Yeah, no, I totally agree. I mean, look, at the end of the day, the North Star, meaning what we want to build towards is-- and you know, we're not there now, but this is where we want to head, from a product capability perspective, from a marketplace scale perspective, is enterprise grade or media grade quality, subtitling, dubbing, captioning, description in any language, delivered within hours.

[00:45:37.63] That's where we want to head. To do that, there's technology investments that have to get made. But the hardest part of that problem is the people problem. Some of these new technologies, they're evolving rapidly. But some of the voice stuff, largely, it's somewhat commoditized, or at least on the way towards it. Scaling the experts in all these regions, dialect specific, et cetera, is a tough nut to crack. It's one we're really, really focused on.

[00:46:05.60] And part of that is scaling those very specific subregions or dialects as they come. And what I mean by that is if there's a large project where it's very important to be really precise there, let's partner together, let's work together, and we'll build that scale for that very specific use case. So we're starting wide, starting broad, and we can narrow as needed.

[00:46:36.58] SOFIA LEIVA: Absolutely. The next question we have here-- I'm sort of combining two that came through-- is around our security measures. What protection exists for fraud prevention, and how is content stored on the platform and that data used to train models?

[00:46:55.33] ERIK DUCKER: Chris, you are muted, but I do think that you should answer this one.

[00:46:58.54] CHRIS ANTUNES: Yeah, no, no. I can take those. So I mean, I'd love to know more specifically what is meant by fraud prevention. Certainly, when you're thinking about scaling people, and we're scaling people all around the world, we take very seriously background checking, identity verifying, all of those things. So we are sort of fluent in the, call it fraud prevention, on the professional marketplace side and can go into a lot of detail there. So I'll assume that was the question and stop that one there. Happy to follow up if I missed the mark there.

[00:47:35.65] On the data storage and modeling side, I'm going to give a very unsatisfying, it depends, answer, which basically means it's up to our customer, right? Do you want us to store your data for a long time? For some, that's a benefit. For others, it's a risk. A benefit means you can log into our platform any time, even years later, and access that data. You can continue to edit it, make small changes to the subtitles.

[00:48:04.50] If you're going to redeliver to a different media platform two years later, it's an asset to be able to go in and still have that asset at the ready and have it be reusable. If instead, you're more worried about security, we have data retention policies where we can clear that out as soon as we process the content for you.

[00:48:21.41] On the modeling side, it's a very similar story, which is at the end of the day, if you are sort of viewing 3Play as an AI partner and a technology partner with you and you want to see scale increase, efficiency increase, cost ultimately come down, we love partners like that. But often, that might mean how do we use your data? How do we build models specific to you that streamline our process so we can pass that cost savings on to a customer?

[00:48:47.44] So if someone wants to do that, we are all ears and want to partner and fine tune models and customize and do all those great things. If, on the other end of it, a customer again,

for security or data privacy reasons, doesn't want us to do anything with the data, we'll delete it. So it's really to taste.

[00:49:04.68] SOFIA LEIVA: Thank you very much. We have time for a couple more questions. The next one we have here is, so will turn around be dependent on the availability of experts in certain languages? And how fast does it get turned around?

[00:49:19.08] CHRIS ANTUNES: So, yeah. So the answer is yes, sort of at a macro level. But ultimately, that should be our problem not the customer's, managing that marketplace. So we will responsibly, for these language pairs on the subtitling and dubbing side, bring down our turnaround SLAs over time. We know that a lot of these projects take weeks today, or certainly days today.

[00:49:46.46] And like I said, we want them to take hours. We believe they should take hours if you have the right scale and network of experts kind of at the ready and you surround them with the right technologies. We'll get there language by language, but we absolutely want to talk about every one of these use cases you see.

[00:50:03.37] So if two weeks is sufficient, that's easy. If we want to get down to two hours or four hours or eight hour turnaround in a very specific language, like we've done that on the captioning side, we can do that, we know the recipe to do it, let's work through that together. And again, many of our customers already we're thankful to have partnering with us as we build this out have those outcomes in mind.

[00:50:28.16] That's why they're so excited to partner with us today, as we develop these new solutions, because they've seen us do it on the accessibility side. What they want at the end of the rainbow there is high quality localization in hours, not weeks. But again, it's going to take a partnership to deliver on that promise.

[00:50:49.06] SOFIA LEIVA: Thank you. The next question we have here is, what is the capacity of the system to detect what needs to be introduced in an audio description script?

[00:51:00.07] CHRIS ANTUNES: So I assume this is AI audio description. Is that the angle here, kind of how well does the automation work, sort of liberally rephrased?

[00:51:11.60] SOFIA LEIVA: Yes.

[00:51:13.33] CHRIS ANTUNES: Erik, do you want to take that or do you want me to?

[00:51:17.20] ERIK DUCKER: I'm going to let you. I'm a little-- I'm trying to figure out, decipher the question.

[00:51:21.89] CHRIS ANTUNES: Yeah, AI, so backing up on audio description for those who aren't as fluent in it. Audio description, so the idea of taking a video, describing what a scene in the video, and ultimately producing an audio track that describes what's happening for someone

who can hear but cannot see, is the typical use case. You can break audio description into three steps typically.

[00:51:45.96] Script writing, so someone writing down what's happening in the video. Voiceover, someone has to, in a traditional setting, speak or read that script, and you need to do that in the space available to you. Meaning, don't collide with the dialogue. So writing the script and then voicing it over is an art. And then, the third is mixing. So you actually have to create that audio track. Typically, that might be a sound engineer or someone using pro tools or something like that.

[00:52:14.20] All three of those can be automated to some extent. Imperfectly, in some cases and in some settings, but they can. So if we're talking about AI end to end, so automating script writing, automating the voice-- text to speech and the mixing, right now, we do have a solution that does that. We are mostly focused, almost entirely focused now, in education use case for that. And I know that is not the purpose of this webinar, the focus of this webinar, but there's an ADA extension called Title 2 that is motivating public universities to think about captioning and audio description at massive scale, just massive scale.

[00:52:55.03] So it requires real innovation on the audio description side to even come close to budget. So we're partnering with a lot of universities there, but that's available. On the media side, certainly for complex content, multi-speaker content, lots of change of scene, I don't think automated script writing is there yet, candidly. I would love to test that out for someone who wants to look at it. So we still believe script writing with human review is the sweet spot, and then try to automate those other two, text to speech, if you can, and the mixing.

[00:53:29.31] But again, would love to talk to people who are interested in this topic. I don't know if our QR code gets them there. But if someone has a really interesting scale to audio description use case, all ears.

[00:53:39.64] SOFIA LEIVA: Yes, definitely feel free to use that QR code, and indicate that you'd like to discuss this more. We have time for one more question, and this is, when will these new features be released, and will these new features increase the cost of using 3Play?

[00:54:00.13] CHRIS ANTUNES: They're released now, so these services are available. We've been investing in them over the last year plus. They're at varying levels of maturity. So, you know, again, we're looking for partners in a lot of these areas, too, for projects that make sense, with timelines that make sense, so that we can build that scale, build that global marketplace, iterate on the tools. But they're all available now. So we are selling these. We are partnering, we are scaling every single day.

[00:54:32.53] And the way we sell all of our services today is a consumption model. So you pay per video, or per minute or per hour. So none of these services increase the cost of the platform. None of these services increase the cost of the captioning you might be doing with us already, or the description. It's all incremental to that.

[00:54:54.76] And obviously, we are-- on the localization side, subtitling and dubbing, we are very aware that this is about global reach and it's about monetization of your content almost always. So we really want to start all those conversations with, what are your goals? What are your budget? What's your budget? What's your quality requirement? And we will figure out how to customize a solution that makes sense for you at the economics that makes sense for you.

[00:55:21.22] SOFIA LEIVA: Amazing. Well, thank you so much, Chris and Erik. That's all the time we have for today. Thank you, everyone, for joining us and asking such great questions. And thanks again. I hope everyone has a great rest of your day.