Podcast

We Get AI for Work: An Exclusive Chat with Mark Zheng, Lead Corporate Counsel at Duolingo

By Eric J. Felsberg, Joseph J. Lazzarotti &

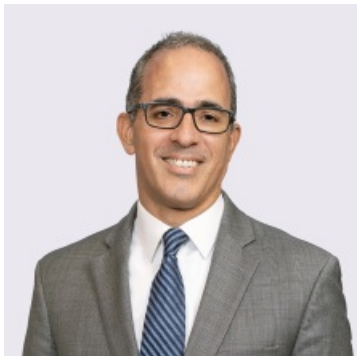November 21, 2024

Meet the Authors

Eric J. Felsberg
Principal
631-247-4640
Eric.Felsberg@jacksonlewis.com

Joseph J. Lazzarotti
Principal
908-795-5205
Joseph.Lazzarotti@jacksonlewis.com

Related Services

Artificial Intelligence & Automation

Details

November 21, 2024

For businesses adopting AI in the workplace, protecting intellectual property presents another legal consideration for employers.

▶ YouTube

Listen on Apple Podcasts

Listen on Spotify

Transcript

INTRO

For businesses adopting AI in the workplace, protecting intellectual property presents another legal consideration for employers.

On this episode of We get AI for work, we welcome our special guest, Mark Zheng, Lead Corporate Counsel at Duolingo. The company's flagship app has fast become the world's most popular way to learn languages. Mark provides actionable insights to help businesses safely operationalize and balance the palpable demand for AI implementation against a constantly evolving technology.

Today's hosts are Eric Felsberg, principal in Jackson Lewis's Long Island office, and Joe Lazzarotti, principal in the firm's Tampa office and co-leaders of the firm's AI Group.

Eric, Joe, and Mark, the question on everyone's mind today is: What intellectual property concerns should businesses be aware of when adopting AI, and how does that impact my business?

CONTENT

Eric J. Felsberg
Principal and Artificial Intelligence Co-Leader

Welcome everyone to our next episode of *We Get AI*. My name is Eric Felsberg, and I'm joined by my colleague Joe Lazzarotti. Joe, how's it going?

Joseph J. Lazzarotti
Principal and Privacy, Data and Cybersecurity Co-Leader

Going well, Eric.

Felsberg

So, Joe, we have a big episode today that we're both excited about. We are joined today by Mark Zheng. Mark is the lead corporate counsel at Duolingo. I know we're in for a treat today because we have a great discussion coming up. Mark, welcome to the podcast. We're thrilled to have you with us.

Mark Zheng
Lead Corporate Counsel, Duolingo

Thanks for having me, Eric.

Felsberg

Excellent. Let's get right into it. Mark, for those of our audience members who are not familiar with Duolingo, could you tell us a little bit about your company, your business and, specifically, your role at the organization?

Zheng

Absolutely. Duolingo is the leading language-learning app worldwide right now. We have about 40 million daily active users. Our mission is to make the best education in the world and make it universally available. And we definitely see AI as an important part of that. My position at Duolingo is I'm in house counsel. We have a pretty small legal team, which is just three people right now. So, I end up being pretty generalist, but I focus mostly on product issues, which would include AI in there. Part of my job is following up on these new developments in AI from a legal perspective and seeing what issues are coming up.

Lazzarotti

Just as an aside, I did use your product. I thought it was great. We wound up going to Italy and I can't say that I can speak Italian but I certainly love the product.

When we had our initial call on this, some things you mentioned that you're responsible for is IP and related issues. Maybe you can help the listeners and summarize some of the key legal developments in IP that businesses really need to think about when they're developing, deploying AI in their own organizations.

Zheng

Sure. The first thing that companies need to be thinking about is the use of training materials for AI, some of which are copyright protected. There have been a lot of AI companies that are just training their AIs by scraping things from the web, and there's been a ton of lawsuits about that: OpenAI has been sued by authors and journalists; Suno and Udio, which are two kind of music AI generators, have been sued by music companies; and then Mid Journey, Stable Diffusion, some of these art AI generators have also been sued by various artists.

It's really up in the air right now whether using these copyrighted materials to train is fair use. We haven't really gotten a straight court decision on that yet. That's definitely something that's going to be a really big deal because you can

imagine if AI companies end up having to license a lot of these materials that could add a lot to the difficulty of training the AI and their costs as well. So that's litigation that everyone is following.

It's more of a big deal for these big foundational AI companies like OpenAI. Your average company that's just using AI probably doesn't have to worry about these kinds of inputs quite as much. But if you're ending up doing a lot of fine tuning or something with copyrighted data, that's something that you want to be thinking about as well. Just as one example: We do an English test, the Duolingo English test. We do use some academic articles for training AIs to help with that and we do have license agreements in place with the owners of those kind of academic corpuses. So that's definitely something that you should think about if you're fine tuning your AI models until you get more legal clarity.

In addition to training, the other big issue that's probably more relevant for a lot of companies is AI outputs. The question is whether these AI outputs are infringing. AI outputs are potentially a huge opportunity for various companies. You can suddenly generate your own music, generate your own art. You can use it in advertisements. We use AI to generate a lot of our course content now as well, which has really helped us build a lot more course content than before. So, the question of how copyright applies to AI outputs is pretty interesting.

Right now, the Copyright Office has basically said that AI created works cannot be copyrighted because there's insufficient human involvement. Now, you can copyright the way that you compile the AI works together. There's a case about a comic book where the pictures were AI generated. And that was said that *that* could not be copyrighted, but the way you arrange the pictures and the captions could be. So, there are some copyrightability issues.

But for now, a lot of AI generated content is going to be in the public domain and that's going to create a lot of interesting opportunities for companies to think about alternative ways to monetize that content besides copyright, as well as the ability to kind of generate their own content, a lot more than they were before.

Now, theoretically, AI outputs could still be considered infringing if they're too substantially similar to existing outputs. So, a lot of the AI companies have tried to put safeguards into their models, where if you ask them to generate something that's basically a copy of an existing copyrighted work, the models won't do it — so that does give you some protection. In the past, a lot of these AI companies also basically created indemnity clauses that said if you got sued for using their outputs and you weren't intentionally trying to create infringing works, then they would indemnify you from any lawsuits. Some companies have reduced those indemnities a little bit. But right now, I think that's not really a huge concern.

Then, there's also questions about rights of publicity — if you want to use AI to clone someone's voice or create a so-called deep fake of someone. That's been a big issue. There's been a lot of state legislation on that issue. And that also runs against some pretty interesting First Amendment concerns because the right of publicity is kind of limited by other people's right of freedom of expression. So that's another area where the law is going to be evolving in the future.

But if you're a company and you're just kind of using AI internally, I wouldn't worry about this stuff quite as much. Right now, the lawsuits are really focused on the training of AI and, potentially, there will be attacks on people using AI publicly in a way that arguably infringes copyrights or the rights of publicity. But just for internal use, we haven't really seen any lawsuits about that.

Lazzarotti

It's interesting. Bringing the scope in a little bit in terms of organizations' use and how their employees engage with the technology, even just thinking about what you said about rights of publicity and use of someone's image or likeness: Sometimes organizations may have employees that will use their voice or their image in company advertisements and whatnot. Thinking about that, how are you operationalizing in terms of how employees look at those outputs and follow the steps that you want to take in the policies? How are you doing that from a governance perspective to minimize risk in that area?

Zheng

We definitely want to look at the contracts that we have in place; if we're using employees or if we're using other contractors to produce voices that we might want to clone with AI. We would definitely want to make sure that our contracts allow whatever AI usage that we want to do. Ultimately, these are things that can be resolved by contract.

In terms of governing our employees, we're basically allowing them to use AI to create a lot of things. But we also have important guidelines like: Don't use AI to try to recreate a copyrighted work; don't use AI to try to clone somebody's voice or anything without their consent. In terms of governance, just trying to avoid this sort of intentional infringement is probably enough at this point. I wouldn't worry too much about the unintentional stuff because a lot of these AI models do have safeguards built in to prevent unintentional infringement and the rights holders are really not going after unintentional infringement at this point.

Felsberg

Mark, just to follow up on that: We've been speaking with employers in virtually every industry about this issue of governance that you mentioned. The series of issues that come up are pretty uniform: First is who should own this. Who in the organization should have their arms wrapped around this issue of AI? How do we avoid different stakeholders giving maybe different and inconsistent impact? Should we have a policy? If so, how do we go ahead and implement that policy?

Talking about different various use cases — you mentioned, earlier, vendor assessments. I'm curious if you could talk a little bit more about how your organization oversees the use of AI to make sure everyone is on the same page and feels comfortable with how it's being used. If you could touch on some of those issues, that would be great.

Zheng

Yes. I think it's really important for an organization to have a uniform AI policy

because if you have different parts of an organization doing different things that could really conflict with each other, especially in terms of what data you're giving to these AI vendors. So, it's important to have a uniform policy in place. To the extent that you're dealing with consumers on AI, you're also going to need to have a uniform policy as expressed in the terms of service for your product and things like that. So, it'd be pretty hard to have one group doing something different than other groups.

In terms of what that policy might look like, you would definitely want to have some policies about what data is okay to send to AI companies versus what data is not okay. That's a big one. You might want to have requirements on what you can use AI for versus what you cannot use AI for. And then, the extent to which you want to have AI be consumer facing — that's also something that you should decide at an organizational level.

Lazzarotti

Those are all really helpful, Mark. One last thing I know we wanted to try to touch on a little bit is just managing third parties. A lot of organizations are kind of challenged by that, working with lot of different partners to find the best solution and then to deploy it and whatnot. Can you talk a little bit about how you and your team think about that and assess and procure vendors to help support initiatives you have using AI?

Zheng

Yes, we definitely want to review all of our vendor agreements whenever AI is involved for a couple of things. One, we want to see what their rules for using AI are. Right now, most vendors, at least for enterprise agreements, say that they won't use your stuff for AI training. So that's something we look for. Sometimes we might have individual employees who might be using various services, so we definitely want to make them aware of when their stuff might be used for AI training versus not. Another thing we would want to do is cybersecurity reviews of different vendors. Even if their agreements say they're not using their stuff for AI training, there's always a risk of them being hacked or having a data breach or something like that. So, we want to be aware of what the risks are from that perspective.

We also want to have clear rules about who can own the outputs from AI and what we can do with them because some companies in their contract will have limits on what you can do with the AI outputs that they generate. So, we want to make sure that we have the flexibility to use those outputs and that we have full ownership of them to the extent that we want to put them in the product or make them consumer facing.

And then we also want to have a risk-based approach because there's always a probability that a vendor isn't going to follow their own terms of service. So, we want to think about: Is there any data where it would kind of be catastrophic for that data to be leaked? And then if there is, then we probably just want to have a rule that we can't use AI with that. That [rule] could also apply to data that's kind of limited within the company. If we don't want people all across the company to

know about something, then we probably don't want to use that in AI. Because most of these AI companies, if you have a model for your whole company, any data that goes in is potentially going to be available to anyone.

Felsberg

Mark, we've been asking all of our guests at the end of our discussions to name the top three takeaways that you think listeners should walk away from our discussion today thinking about.

Zheng

Absolutely. I would say that the top three takeaways are:

1. You want to be thinking about IP issues when you're using AI. It is unclear whether training AI on copyrighted data is infringement or fair use at this point. That is a legal gray area. And, also, for the AI generated outputs: It is possible for those to be infringing, and AI models are trying to think about ways to reduce the infringement risk there. So, that's something to think about.
2. Make sure to think about data and confidentiality when you're dealing with AI. Whatever data that you give to these AI vendors, you want to make sure that that's covered by your terms of service and your privacy policy to the extent you're using consumer data. And you want to make sure that you're not giving these AI companies any kind of extremely sensitive data.
3. The last thing is make sure that you definitely consider public relations. AI is a very controversial thing. A lot of people are also concerned about it replacing jobs and things like that. A couple of companies had mini-PR scandals last year where their terms of service were broader than they needed to be. People thought that they were using more stuff for AI training than needed to be. So that's something to be wary of. I know traditionally we kind of draft terms of service to be as broad as possible, but people are actually looking at them now.

   On the jobs front, I think that's another PR issue. Duolingo had an issue where we didn't renew some contractors because we were using kind of more AI and this got turned into a big PR thing where it was like, "Duolingo has mass layoffs," which is totally not the case. But it's definitely something that people are sensitive about. You want to make sure that the way you're using AI is clear, that you're using it to do things that you wouldn't be able to do before and not just kind of replacing people.

Lazzarotti

Well, listen, this has been great, Mark. We really appreciate you giving us some time here. I think all of this is going to be really helpful for folks listening in, particularly on the IP front and some of the governance issues you addressed. So, thank you for taking the time to be with us today.

For any of the listeners that have any questions or have any ideas about what they'd like to see on an upcoming episode, please email us at ai@JacksonLewis.com.

Felsberg

Very helpful. Thank you, Mark.

Zheng

Yes. Thank you, Eric.

OUTRO

Thank you for joining us on We get work™. Please tune into our next program where we will continue to tell you not only what's legal, but what is effective. We get work™ is available to stream and subscribe to on Apple Podcasts, Libsyn, SoundCloud, Spotify and YouTube. For more information on today's topic, our presenters and other Jackson Lewis resources, visit jacksonlewis.com.

As a reminder, this material is provided for informational purposes only. It is not intended to constitute legal advice, nor does it create a client-lawyer relationship between Jackson Lewis and any recipient.