**Publishers, Copyright & AI: Taking Action**

*with*

**Sarah Fricker, IOP Publishing**
**Claire Harper, Sage**

**Recorded live at the 2024 London Book Fair**
**March 12, 2024**

**For podcast release**
**Monday, March 18, 2024**

KENNEALLY:  Welcome to Publishers, Copyright, and AI: Taking Action.  I'm Christopher Kenneally with CCC.  Thanks so much for joining us.

No prizes will be given for guessing the hot topic for publishers attending the London Book Fair this week.  Sorry.  Our panel this morning – and no exaggeration, the entire London Book Fair – stands on the intersection of publishing, copyright, and AI.

Artificial intelligence will transform the global economy, according to the International Monetary Fund.  And as they have done since the digital age dawned, scholarly and academic publishers are already embracing this latest technology.  While they explore how AI can improve their business practices for the benefit of researchers and information professionals, many publishers are also proactively asserting their rights when generative AI solutions that create text, images, and other media use copyrighted material to develop and train large language models.  Publishers recognize that gen AI depends on high-quality content for success, and they contend the tech industry must seek permission for its use.

What are the legal and regulatory precedents that really matter?  Which measures must be in place to support innovation, while facilitating responsible AI approaches by Big Tech?  This morning, starting off three days of discussion, we will hear from representatives of leading UK scholarly publishers to know what they are doing when it comes to facilitating permission for copyrighted works to train AI models.  I want to introduce my panel.  Sarah Fricker – Sarah, welcome.

FRICKER:  Thank you, Chris.  Very pleased to be here today.

KENNEALLY:  Very happy to have you.  Sarah Fricker is group head of legal for IOP Publishing – the Institute of Physics.  She leads a team providing legal advice to IOP

companies around the world on issues including copyright and licensing.  Sarah is a non-executive director with Publishers' Licensing Services and ALPSP.

Also joining me today is Claire Harper.  Claire, welcome.

HARPER:  Hi, Chris.  Thanks for having me.

KENNEALLY:  Very happy to have you as well.  Claire Harper is head of global rights and licensing at Sage UK, where she oversees licensing, permissions, and translations for Sage books and journals.

I think the appropriate place to start before we really dive into the copyright piece of this is to talk a little bit about AI and the ways that your publishers plan to use AI to improve the variety of services you provide to authors, researchers, and so forth, because it's important to establish that this is not an anti-AI perspective.  This is one that's very much about taking on AI, but doing so in a responsible way.  Sarah, if I can ask you first of all briefly about IOP's uses of AI and the way you're beginning to start to experiment with it?

FRICKER:  Yes, certainly.  We're spending a lot of time at the moment looking into how AI can be used to make our business processes more efficient, to help our authors, to make sure that our staff are kind of concentrating on the things that really matter.  So elements like routinely checking whether articles need editing or not, helping our staff use their time effectively day to day, also very important issues like research integrity, whether we can use AI tools to help with that side.  I think as you said, Chris, very rightly, we are not in any way anti-AI, and we can see that there are definite benefits when used in a responsible way.

KENNEALLY:  Research integrity – it's so important, and it's important for all kinds of reasons that have been around for hundreds of years.  But it's more important than ever, because one of the real threats around AI is the abuse of creating content to try to misinform and all that.  Talk about – when you said research integrity, you're talking about plagiarism detection, paper mill detection, that kind of thing?

FRICKER:  Yeah, certainly that side of it is part of it.  And I think also, as you said, misinformation is a very big concern, particularly as it was seen during the COVID pandemic.  Making sure that people can rely on the information that you're publishing is so important.  I think, therefore, when we take articles in for publication, we want to be sure that we have the warranties that we need from authors to be able to rely on that content – that it's original, it's been researched properly.  I think everyone would probably agree at the moment that the accuracy of some of the information coming out of the AI tools cannot always be relied upon.

KENNEALLY:  I'm sure NASA would check first, but if you get it wrong, the spacecraft goes right past the planet, right?  That's true with physics especially.  It has to be accurate.  You can't get it wrong.

FRICKER:  Absolutely.  And I think until – unless we can be absolutely certain that the research in all the articles that we publish has been tested and checked, it's very difficult to be able to pick up inaccuracies.  So we are relying on that human intervention with the publication process.

KENNEALLY:  Claire Harper at Sage, you're a very different kind of publisher, very different kind of content.  How are you approaching AI?

HARPER:  Yeah, I think it's interesting.  When we talk about AI from a publisher's perspective, we quite often talk about challenges, but there's obviously opportunities as well.  At Sage, we've got an internal steering group which is looking at AI across the business – so how are using AI?  How do we want to use AI?  Looking at best practices.  Maybe there's some processes that we could improve with the use of AI.  I think obviously we want to prepare for the future with these extra tools, but we also need to think about how we're protecting our content.  That's a really important part, and we want to make sure we're still – kind of what Sarah said – we want to still be using responsible and ethical publishing practices.  So trying to get a balance of using AI to save time, but how do we make sure it's still responsible and ethical?

KENNEALLY:  Right.  You mentioned the steering group.  I believe that a number of publishers are using that kind of process to really look at AI thoroughly and to understand the implications.  It's difficult.  It's a moving target.  Things are always changing.  But to have that group of all the various stakeholders in the company is really important.

HARPER:  Yeah, definitely.  Even from a licensing perspective as well, that group is so helpful to me to – you know, I can go to them and say we've had these opportunities.  We ought to talk about it.  Because we don't really know what we're doing, right?  So we all need to just talk about it and see what the opportunities are.  And when they're looking at processes, that group is really such an important group, because like you say, they're across the whole business.  So they really know all the ins and outs of the different aspects.

KENNEALLY:  Right.  You told me, Claire, that Sage is already using AI tools to really make things a lot easier for authors, because you're checking the submissions at the beginning for what?  You're not checking for the integrity piece, but something else.

HARPER:  I think the idea is that we would be checking manuscripts post-submission to make sure that they adhere to the formatting guidelines.  I think that's quite a good opportunity with AI, because it will save someone time, but it won't really affect the integrity of the content, because it's just formatting guidelines.

KENNEALLY:  But that's a great tool for authors and for you as publishers, and I see you really in agreement with that, Sarah.

FRICKER:  Absolutely.  I think we're already using it for editing checks to see whether – when articles are submitted, how much of an editing process they need to go through.  That does very dramatically, and I think that's where AI can also help to actually really reduce the time needed to identify that quanta of articles that need the most work and those that can go ahead and publish, which is a benefit then to the authors, because those that don't need a lot of editing can go ahead and be published perhaps more quickly.

KENNEALLY:  Right.  We're here in the Tech Theatre at London Book Fair, and it's really clear from the way you both are talking that publishing is a technology business, like nearly every business at this point, but it's particularly important for publishing to have a relationship with the other tech providers and to talk with them about the way they're using your content.  Why is it important for this audience to understand the need for permission when it comes to training these large language models?  And when we say large language models, we mean enormous.  They're consuming incredible amounts of content.

FRICKER:  I think if you look at a number of the legal cases that are going through at the moment in the US and the UK, this is something that's a concern not only to publishers, but to the authors themselves.  The authors are – there's multiple class-actions being taken by authors, because this is work that's been created, that's original, that people have spent a lot of time putting together and publishing, and it only to me seems fair and reasonable that if these companies want to use it to improve their tools that they get the permissions they need from the copyright owners to do that.

And there are benefits on both sides.  The big companies can be sure they've got the right version of the article, which as you said is so important.  But equally, there's value placed on that work, and the copyright owner is getting recompense for that value.  I think that is a very important balance.

KENNEALLY:  And the authors you have – they're researchers.  They're pure science.  But they really care about this as well.  Have they talked to you about their concerns around this?

FRICKER:  I've heard anecdotally – not so much through IOP, but in the industry – that there are a number of authors who are very concerned about this, not least about their content

being misrepresented. There's a number of stories about fiction authors in particular where their books are being sort of rewritten by AI in a way in their style, where obviously they're not getting any credit or any recompense for that, either. So I think there are authors out there who are genuinely concerned about this, and as I said, are looking to their publishers to help them to get that kind of recognition.

KENNEALLY: Claire Harper, what about Sage? It's a very special kind of company. As we say, it's in the social sciences. That's a far cry from the pure science of physics. How about your business – how it views the need for permission? Why is that important to you?

HARPER: Well, I think the first thing to say is the content that we publish is copyrighted content. So if you're using it without permission, that's copyright infringement. (laughter)

KENNEALLY: Says the woman who runs the permissions.

HARPER: Yeah, I would say that, wouldn't I? But I think also kind of to Sarah's point, we want to make sure that the content going in is reliable, because – I mean, you've probably heard the saying garbage in, garbage out. Are we going to trust what an AI tool is telling us if it's just pulling from a random source on the internet? The internet is full of fake news. We know that. So if we can license the content in, we can be a bit more assured that the content coming out is going to be more reliable and accurate.

KENNEALLY: Sage particularly is concerned about this because of the kind of company, the kind of business that it is. Tell us about that.

HARPER: We're an independent company, and we are quite – we say fiercely independent. We have this saying inside the company that's we're free to think long-term, because it means that we can really take the time to make decisions that are best for us and our authors and our societies. We don't have to rush a decision that's best for stakeholders. So we can really take the time and make sure that decision we make is for the long term. Really, it's about protecting our authors and our societies.

KENNEALLY: As we were sort of joking, of course you'd say that. You're the permissions director for Sage. But how do you respond when the companies who are creating these large language models say it's too hard to get permission? Or we can get into this in a moment, but also, well, it's just fair use. But let's start with the it's too hard. How do you respond to that?

HARPER: I'd probably say, have they tried? Because like Sarah said, there's multiple lawsuits that we can point to that indicate they didn't try to get permission first. So I think as

publishers, we're trying really hard to make sure that we get streamlined processes, we get ways of licensing this to make it as easy as possible, because that's in our interest as well. We want licensed content to go into it. We want to be remunerated for that content. So we want to make it as easy as possible. So I personally wouldn't say it's hard. (laughter)

KENNEALLY: And you're not saying we want to block this from happening. You're saying we want to facilitate it.

HARPER: Yeah, absolutely. We don't want to block it at all. I think from our perspective, it's better to license it than to block it. And I'm sure in time, we will see some licensing options become available. I think having a collective license would be great. We've seen it in the past, right? So photocopying came about. It was very disruptive to the industry. But we found a way to license that. So I feel like we could do that in this case as well.

KENNEALLY: Do we have to stop and explain what photocopying is to some people?

HARPER: Old-school, right?

KENNEALLY: Everybody here knows what a large language model is, but they may not know what photocopying means. Sarah, what about that? The responses of these companies has been it's just too hard, as we say. Talk about that.

FRICKER: Yeah, I would totally agree with what Claire just said. I think it's easy to say that. I would be very interested to know what these companies have actually done to try and get licenses. Because I think the industry is trying. It's a very complicated area, as we've said. It's very fast-moving. Nobody wants to get it wrong if they can avoid it. So coming to a license – or a collective license is not an easy process. But I do think that maybe individual publishers having their own licenses together with the collective licenses to me is the best way forward, because then all parties know what they can and can't do. It's very clear. Nobody has to rely on a copyright exception that may or may not cover them. It's much easier than to move ahead. But I do think that the argument that it's too hard, I can't do it, doesn't really hold up at the moment.

KENNEALLY: I do want to talk about licensing and the difference between sort of the bilateral licenses that may exist and the collective licensing. But maybe first, we can get to the point that some people are talking about having exceptions in the law that would allow for this in a statutory way. I know you're an attorney. You would not be for that. You don't like that. But there are real reasons for that – not just for the business sake, but because those kind of exceptions really never work in the end.

FRICKER:  No, and I think a copyright exception is only really meant to be introduced if it's kind of the last resort, if you like.  If the industry can't manage something themselves, then an exception has to come in.  I think it's too early to say that we can't manage it ourselves.  And I think at the moment, doing a copyright exception would be putting a line in the sand and saying this is what the exception needs to cover.  We all know the AI models are moving so, so quickly, an exception could very well be obsolete tomorrow.  So I think we need to just try and see first, is an exception needed?  Let's explore other options first.

KENNEALLY:  Claire Harper, is an exception needed as far as Sage is concerned?

HARPER:  No, I don't think so.  I think Sarah's right.  It should be a last resort.  I think we're so early in the process, really.  AI is really quite new.  Like I say, in time, I think licenses will be available.  And I think that's a really key part about making sure we avoid an exception.  Because if we can point to the fact that they can get a license, then there's really no need for a copyright exception.

KENNEALLY:  To your point, to both of you, you're saying let's try to do some licensing first and see how that works before we get into any exceptions.  Talk about licensing, particularly around collective licensing, which means various publishers getting together to provide kind of a repertory for a license that facilitates all this permission.  Why is that a useful model for you?

HARPER:  I think if we're talking about trying to avoid a copyright exception and this notion that it's too hard to get permission, then collective licensing is a good tool for that, because they can just have one license with CCC, for example, that covers all different publishers.  There's no having to go to individual publishers to get the content.  It's just all in one place.  That in my mind is quite an easy way for them to license, and I just think it makes a lot of sense.

KENNEALLY:  Sarah, collective licensing – what's the value to IOP of that?

FRICKER:  I think for us as not a small publisher, but not one of the really large publishers, to be able to put our content into a collective license means that we are able to get our content out to the companies who perhaps might not come to us directly to begin with, reach those companies, have an agreement with them, and be able to move forward.  So for us, there are so many advantages to collective licensing.  From the actual users' point of view, they get the same rights for all of the content that they license.  We know it's being handled correctly, and therefore we can kind of step away and let the license do its work.

KENNEALLY:  I think the really important piece about licensing for all publishers, but particularly for a publisher like IOP and maybe for Sage, is it empowers you.  It gives you

a role in this process.  And I think a lot of publishers – certainly at the beginning, when OpenAI launched ChatGPT, thought oh my goodness, they've left us out of that.  But licensing puts you right at the center of it.

FRICKER:  Yeah, absolutely.  And I think we know that our content is probably very useful to these companies, so it feels now the right time to be able to say, OK, here's the content.  Here's how you can use it.  And then these companies can come and get the license that they need.

KENNEALLY:  Claire, what about that part?  Sage wants to have a real part in this revolution, because it's going to change publishing as much as it's going to change how we interact with retail, with our healthcare, and all the rest.

HARPER:  Yeah, absolutely.  I completely agree with everything that Sarah said.  I think we want to license the content.  Like we said, we don't want to block this.  I think direct licensing would be another option, I guess.  I think like Sarah said, some companies are not going to come directly to us, or they want content from both of us, so they don't want to go individually.  Going through a collective license just makes it streamlined, and we then still get remunerated for that content.  I think this too hard thing – that's one way to make it easy, so let's do it.

KENNEALLY:  It's all part of what everyone's calling governance around AI.  Your organizations have set out for researchers, for authors, the various uses and the kinds of ways that they can use AI.  So tell us briefly about that, because this is important to them.

HARPER:  We do have a section in our author guidelines that kind of states that the authors need to disclose if they're using AI within the content.  That's kind of where we're at at the minute is just trying to get everyone to disclose if they are using AI in the content, and then obviously we're getting questions from the authors about how we're going to protect the copyright of their content.  So that's something we're just really working hard internally at the minute to try and make sure that we do still protect their content.

KENNEALLY:  Sarah, the same?

FRICKER:  Yeah, that's exactly the same, really.  I think, as you said, we want the authors to tell us if they're using AI.  And yeah, as Claire just said, to actually – in terms of then authors are turning to their publishers, turning to us and saying, how are you going to help us protect our content from the AI models?  What steps are you taking?

KENNEALLY:  Can ChatGPT be a coauthor?

FRICKER:  No, at the moment, if they use any form of AI, they would need to declare it on their article.  But no, we're not having articles that ChatGPT are the sole author.  No.

KENNEALLY:  It's kind of a fun question to end the discussion with, and we'll take some questions from the audience.  But there's an issue, too, around whether that output is copyrightable with these machines.

FRICKER:  Yeah, there is.  I think this is where some of the complex legal arguments are going to come in, because the UK and the US have always had very different rules about whether computer-generated content is copyright-protected or not.  So yeah, that's another issue yet to be determined.

KENNEALLY:  My colleagues know that I'm fond of saying that when it comes to copyright, if you're confused, you're beginning to understand the problem.  This is exactly it.  But you've both helped me to understand things better around copyright and AI, and I want to thank you for that discussion.  Sarah Fricker with IOP and Claire Harper with Sage, thank you both very much indeed.

For more information about copyright and AI, go to copyright.com/ai.  I'm Christopher Kenneally with CCC.  Thanks for joining us.

(applause)

END OF FILE