



Entering the AI Era: STM TechTrends 2022

With

- **IJsbrand Jan Aalbersberg, Sr. Vice President, Research Integrity, Elsevier,**
 - **Gerry Grenier, Director of Publishing Technologies, IEEE**
 - **Phill Jones, CTO, Emerald Publishing**
- **Stacy Malyil, Director of Strategic Marketing, Wolters Kluwer**

**For podcast release
Monday 14 May 2018**

Recorded at STM US Conference 2018, Philadelphia

KENNEALLY: Last summer, we learned that a pair of computers at Facebook's offices had wandered sufficiently far off their coded scripts that engineers chose to shut them down. Journalists, not then on holiday at the time, had loads of fun writing terrifying stories about robots on the rise. Meanwhile, their editors searched the photo morgues for publicity stills from *2001: A Space Odyssey* showing HAL 9000, the cyclopean supercomputer with a sociopathic disposition.

Then, in the fall, no less a light than Stephen Hawking predicted for a *Wired* magazine report that AI, artificial intelligence, will eventually become so advanced, it will essentially be a new form of life that will outperform humans. The fiction, it seemed, was fading fast from sci-fi and morphing into fact.

Now, as it turns out, robotic chatbots frequently invent their own languages, creating a kind of shorthand that enables them to communicate with each other more efficiently. The Facebook story wasn't fake news. Not exactly. But it wasn't really much news either. In this case, too, the engineers decided to pull the plug not because they wanted to save humanity, but because the assignment was to construct chatbots that could speak with human beings rather than only communicate with other machines. As we've been learning from Eefke Smit, STM Tech Trends 2022 envisions a similar need in publishing to bridge human intelligence and artificial intelligence.

Entering the AI era, creative humans and smart machines imagines an alliance of tools and technicians making optimal use of available technology, while maintaining scholarly communications is very much a human endeavor, a creative craft and art in itself. Humanity and technology are inseparable partners and ever



have been. Nearly 2 million years ago, in fact, speech and technology emerged together on the African savanna. Those early ancestors of ours, the hominids, began to fashion the first stone tools at around the same time as they developed the necessary physical organ for speech, an unusually shaped tongue. Our inventions, our science, are bound up with our communication. And we are bound, it would seem, to communicate our inventions.

This morning, we're going to do more talking than inventing to explore the inspiration behind this very creative infographic. To help me do that, I have a wonderful panel here which I'd like to introduce. Moving from my left here, IJsbrand Jan Aalbersberg. Welcome.

IJsbrand Jan Aalbersberg is senior vice president of research integrity for Elsevier, where he's responsible for new technology initiatives that safeguard the integrity of content and the content-based products that Elsevier offers to the research community. In this role, he's also responsible for user privacy. IJsbrand Jan Aalbersberg earned a PhD in theoretical computer science at the University of Leiden. He has published scientific articles and hold patents in the areas of document retrieval, research-data linking, and user interfaces.

Immediately to my left is Gerry Grenier. Gerry, welcome.

GRENIER: Bonjour.

KENNEALLY: Bonjour. Ça va bien? Gerry Grenier is currently director of publishing technologies for IEEE, the Institute of Electrical and Electronics Engineers. He holds – or he leads, rather, a 40-person electronic publishing team responsible for development and operation of IEEE Xplore. He serves on the Boards of Crossref and the National Information Standards Organization.

And to my right is Stacy Malyil. Stacy, welcome. Stacy is director of strategic marketing for Wolters Kluwer's portfolio of texts, services, and online learning, reference, and practice solutions in medicine, nursing, and allied health. She has held marketing, business development, and product strategy positions at McGraw-Hill Education, Springer Nature, and Taylor & Francis.

And then finally, on the far right as I look, is Phil Jones. Phil, welcome. Phil joined Emerald Publishing as CTO earlier this month. He previously worked at Digital Science in a variety of positions. He holds a PhD in atomic and plasma physics and has held a faculty appointment in neurology at Harvard Medical School. So we will be asking you about some of those brain functions later on, Phil.



In fact, I'd like to start, because there was one item that caught my eye, which was this idea of Spotify for science. Thinking about playlists and mood music, and you were in that room in December that Eefke described so well for us, what was on the playlist? I think about it in terms of mood music. If the mood were sort of, on the one end of the scale, kind of nursery rhymes and lullabies, or on the other end of the scale, sort of post-apocalyptic punk, what was the playlist there?

JONES: I don't think it was either Mary Had a Little Lamb or Rage Against the Machine. I think it was perhaps somewhere in the middle there. But what I think is really interesting – looking back over several years of the Tech Trends and the Future Lab brainstorm – is that a lot of the technologies that we're talking about haven't changed all that much in terms of what the headline technology is, the name of the things that we're thinking about.

What I think has changed, the way we are evolving, is that it's gone from thinking about what are the technologies that are emerging into how those technologies are going to be applied and the relationship that those technologies have with the industry and with individuals and how human beings use and interact with those technologies. So I think it's really telling that at the center of the brain in the infographic we have the deep publishing knowledge, because technology isn't solving our problems for us and isn't going to solve our problems for us. We have to learn how to, as humans, use technology to our advantage to solve the specific problems. That's what I feel is kind of consolidating and emerging around the discussion.

KENNEALLY: What's consolidating and emerging? Is it more or less frightening than it might have been when we began to be aware of these things? Do you think that the disruption that we've been hearing about so much in scholarly publishing is continuing, is mushrooming, or people have grown accustomed to it and are now managing it better? How would you assess that?

JONES: I think the latter. I think the latter. It's really noticeable at meetings. A few years ago, when I was head of outreach at Digital Science, and I would talk about some of the technologies and the applications and workflow tools and data and analytics and all of these things, I would sense a certain amount of resistance from publishers, particularly smaller publishers in learned societies – a fear that this isn't what we do. Right? We don't do this technology. We publish content. This isn't our wheelhouse. That resistance to adopting newer types of technology, I feel, has reduced significantly over the past few years as people have gotten used to the idea, as people have learned more about it.



Now I feel that people are coming to technology companies and coming to platform companies. John Sachs (sp?) said this the other day, that people are coming to him asking him questions about technology rather than him explaining technology and people going, you know, I'm not so sure that that's really – so there's this getting used to it. And in that sense, I think it's a lot less scary for a lot of people.

KENNEALLY: Yeah. But as we look at the infographic, there's a lot of to-dos on that list, and they really are all about technology. To your point, since when did publishing become a technology business? It has been for some time.

JONES: It's always been a technology business.

KENNEALLY: Probably always. Exactly.

JONES: Right. And it's impossible to overstate how connected the technology of publishing is to the advancement of human knowledge. When scholarly publishing emerged back when *Philosophical Transactions of the Royal Society* was first published, that was a technology advancement. It was the idea that you could write down the proceedings of a meeting in a pamphlet and distribute it. That was technology. And that enabled researchers to communicate in a rapid fashion, whereas previously, they were writing whole books, and you had to wait until someone's whole career, or a big chunk of that career, had elapsed before you learnt what they were doing. So we have always used technology in order to accelerate the flow of information through the scholarly communication life cycle, and so in that sense, nothing's changed.

KENNEALLY: I think even, in fact, with Gutenberg, his technology of the printing press – wonderful invention, but I believe the man died a pauper. It was very difficult to find the business model, and there wasn't much of a business model in the books for quite some time after that.

IJsbrand Jan Aalbersberg, you were there in the room as well. How would you assess the mood music? And in particular, around artificial intelligence – I know at Elsevier, you're applying AI tools to a variety of pieces of the business, but would you say and would you agree with Phil that people are becoming a little bit more accustomed to all of this, and it's sort of become more partnering with technology than in fearing of it?

AALBERSBERG: Yes, I definitely agree. I think that the mood changed from being a little bit afraid, uncertain about a future with artificial intelligence, although I have to say that – yeah, well, let's say it was more in the past people were more afraid



about it. But now they talk about it more in a sense like how can it help? It was very telling that some people moved it from artificial intelligence to augmented intelligence, that it was really helping us, and people saw the opportunities. People were not afraid of it, but people saw it more – if you talk about the mood change, it was an upbeat mood that artificial, augmented intelligence is going to bring something. It's moving the scientific publishing, moving science ahead. And people were not afraid of that anymore. So definitely upbeat.

KENNEALLY: And upbeat because it's an opportunity to counter some of the concerns that all of publishing – not only scholarly publishing has, but all of publishing has, regarding how easy it is to manipulate, to fake, to create – if not artificial intelligence, artificial news, fake news. Talk, IJstrand Jan, about how you at Elsevier and in this community were using these AI tools to ensure research integrity better.

AALBERSBERG: Yeah, it's a combination of how we, of course, think of the present day, how we can use the technology, but also an expectation of how we – and that's what Tech Trends is about. Tech Trends is also about 2022, that we hope in the future, it will be working in a certain way.

The expectation is that it will definitely help in detection of fraud, of manipulation. More specifically, yeah, we all know about plagiarism. That's already a problem. That problem is – at least the detection of plagiarism – is relatively solved. But the next step is, indeed, manipulation of images, manipulation of data, fabrication of data. That is where we do expect that AI and machine learning will definitely help our reviewers, our editors, the whole research community to keep the science and what's published proper and trustworthy.

In that sense, indeed, yeah, we are looking into those aspects. How can machine learning, how can also data and analytics, looking at what people are doing, but in all the different content – how can that help us in ensuring that the content can be trusted by our users?

KENNEALLY: Right. And fake science is surrounding this individual here who's doing all this deep thinking about deep publishing. How much fake science are you finding is coming into journals and into Elsevier? Is this something that is just beginning to be a problem, or has it grown to be a major problem? You mentioned data manipulation, image manipulation. Are we seeing a lot of that? Are you detecting a lot of that?

AALBERSBERG: Percentage-wise, it's relatively small. So most of the content is trustworthy, can be trusted across all publishers. But the effect of maybe one



article that is really manipulated or is fabricated – one single article that is wrong could have a major effect in science, in health, in medicine, in the social community. So it is not on how often it happens, but what could be the potential effect if it happens. That’s why it’s important that we have to, indeed, minimize that type of fraud and error in each and every case.

KENNEALLY: Gerry Grenier, you weren’t there in the room, but you have perspective on the potential for AI, and I understand you’re very positive about its potential.

GRENIER: Yeah, I (inaudible) deal with this little piece of technology. Yeah, I’m excited because AI is one of those general-purpose technologies. Through history, we’ve seen electricity, steam power, the internal combustion engine all classified as general-purpose technologies. And when they came along, it took imaginative people to exploit those technologies and bring new products and services to humankind.

Phil mentioned that publishing is a technology business, and yes, it is. Always has been. But I think it went through a couple of stages of involvement with technology, of influence by technology. At first, it was the printing press, and then there wasn’t much happening for a couple of hundred years probably. Then the ’80s and ’90s when we realized the potential of the internet, we went up another level of technology when we got into machine data conversion, converting Word files into XML, etc. – very almost mechanical kinds of technologies. But I think AI, it just throws us into another level altogether, just as it’s throwing civilization into another level of involvement with technology and influence by technology.

KENNEALLY: You mentioned the steam engine, and a colleague of mine was reading a book and sharing with me that he had read that in the very early days of the steam engines, there were boat-makers, manufacturers – they had sort of done the traditional clipper ship, and they thought, well, we could just throw a steam engine into the clipper ship and things will be up to date. But that really wasn’t going to be effective. You needed a purpose-built steam engine and a steamboat or steam-driven boat.

Is it true that as you described that evolution, we’ve gone from digitizing all of these texts to a really, truly digital era? Have we recognized that we have to be born digital, fully digital, we can’t just sort of be trying to couple this onto the existing structures?

GRENIER: No, I think that that’s where the human creativity comes into play in that sort of gray piece in the middle of the brain, the deep publishing knowledge that all of us bring to our business. I think that our challenge as executives and managers is



A Copyright Clearance Center Podcast

to help ease that transition. I don't think there'll be a step change. I think we'll see gradual changes over time. So no, I think that we won't throw out the old. Everything's a transition. We have to do that business-wise. I think that we can't just stop doing something today and have a switching period where we lose revenue.

KENNEALLY: We will get back to that organizational structure issue in just a moment. But Stacy Malyil at Wolters Kluwer, you're working on the customer-facing side of the business, so you see the applications of the technology in the marketplace. Talk about the ways that you see AI helping today with the work you do.

MALYIL: That's true. I'm not so much on the product development side, but my team is focused on what happens with the user experience and customer experience. Something we've seen in our education business is the use of machine learning with adaptive learning and quizzing. We've made some acquisitions in this area, acquired companies that do adaptive learning for high-stakes exams. You used to have students that studied really hard but didn't necessarily study very smart. And with machine learning, you can really analyze their mastery level of what they need to learn for high-stakes exams and present them better questions, better study plans that are really individualized and personalized. And it doesn't matter how you really developed your content up front, as long as on the delivery end you're taking into account what they're strong at and what they're not.

So an example with adaptive learning – you answer a certain bank of questions wrong, well, we're not going to keep giving you harder questions. If we can automate that with machine learning, a lot of questions – your average student will take 2,000 to 3,000 questions before their next high-stakes exams. That's a high volume of data that we're sitting on – usage data. So let's make them study smarter and not harder. Don't add another 2,000 questions in the same vein. Deliver five more that calibrate according to their mastery level. That's really where the power of machine learning has come into play with Wolters Kluwer in some of the acquisitions we've made in adaptive learning.

But I also think on the human and AI collaboration piece, you asked about the mood. And I wasn't there in December, but we're all having these conversations in our companies. Everyone's looking to figure out, should we be hiring teams focused on AI? Should we be hiring data scientists? At Wolters Kluwer, we actually have a team of data scientists looking at what are the opportunities of the point of care, point of research and point of learning, and how could we improve patient outcomes? How can we reduce medical errors? How can we deliver smarter learning tools?



A Copyright Clearance Center Podcast

I think there's more of a level of excitement now, because we realize what a human is still needed for. There is deep domain knowledge that we as humans have in all of our companies, but also the user intimacy. And if you don't understand what problem you're solving for your user, it doesn't really matter what the machine is doing, because they need to know the context of what the user will do with that answer that the machine spits out. Right? So is it to be used in research? Is it to be used for their training purposes? Is it to help guide them answering a clinical question or how to treat a patient? That's where human expertise has to be combined with what the machine can learn over time, because the machine can deliver an answer, but the human delivers the context. I think that's what we're excited about being able to apply at Wolters Kluwer.

KENNEALLY: In fact, one of those acquisitions that you referred to was for Firecracker, a Boston-based company. Tell us about how that came to be, because I think it's an interesting example of this notion of expertise. Because the founders were themselves physicians, as I understand.

MALYIL: Right, and there's that idea of – a lot of startups out there are founded by the users themselves, in their verticals. So we are seeing in medicine, a lot of startups are done by students – medical students who didn't like the solutions that were out there that unfortunately we as big, traditional publishers produced. So how do we compete? We should probably look to join forces or where do we partner in an effective way?

What they've taught us, some of these startups that are user-built, is the users should be part of the curation process. We as publishers have been the great curator for how many years? That's what we sell. We tell our customers there's a wealth and sea of information and evidence and research out there. You'll never get to all of it. How do you weed out things like the fake and the bad science? We'll do it for you. That's what you come to us to do is the curation.

But I think there's an expectation of users now that they are part of the curation process. Don't just tell me what you think I should know. Look at what I'm doing on your sites and within your products and know me a little better. That ties into the smart services and personalization. So it's a partnership now that we can achieve – subject matter expertise that the publishers have, that deep domain knowledge, coupled with what are we watching our users do?

If you give a search results page of content and you have video, articles, chapters, topic summaries, animations, podcasts, and every time that user, that individual, unique user goes for the video, their expectation is we're going to auto-rank that higher. We're going to present it more prominently, because that is their



A Copyright Clearance Center Podcast

expectation. They're getting curation done in all of their other consumer activities, so for their professional activities, they're expecting it from us as well. But they expect that we're using their behavior to drive our curation more than just what we think is the best thing for them to be reading or interacting with.

KENNEALLY: And this notion of machine learning, is sort of at the heart of all of that. The kind of curation you're talking about is data-driven. It's about gathering data, understanding what it's trying to tell you. Even as you're learning in that environment, the machine's learning you. And this partnership that you described seemed so sort of – it's game-changing for publishing, it sounds like.

MALYIL: It is. In the brain here, the user-oriented publishing, it's not just publishing knowing your users, but it's also continuing to deliver what you've already published in a way that meets their expectations. Again, that's what we're focused on in the more customer/user experience side. But all of these trends, when it comes down to it, if you don't have a human that understands the user's needs and expectations, the machine may be able to learn it over time, but I don't think it can do it up front.

I don't know if others agree, but I do think that's where our employees, our staff bring – they know what's at stake for the user at the end of what you deliver them if it's a search result or if it's an article or if it's a topic summary. We know why they're looking for it, and we have to make sure that collaboration with the technology is strong. Because if we're just tagging images and doing image recognition and things – yeah, it spits it out, but if you're not using it for a specific purpose, it's not helpful. That's where this partnership with our users is so important.

KENNEALLY: Right. And it sounds to me also that there's opportunities in all of this for human beings. I imagine that the work you're talking about requires having teams of data scientists and others who really understand these issues as part of the people who are in the room in publishing.

MALYIL: That's correct. I think a lot of companies, probably in this room, are investigating how to hire data scientists, what you want them working on. In Wolters Kluwer, we are singularly focused in our vertical on healthcare. So we are looking for people with data science backgrounds, but who also understand the workflow of healthcare, so again, they can apply the intelligence to – this is a point of care use case. This is a training use case. This is a research use case. That way, they can build the technology but also understand at the endgame what that user is going to use it for. So I do think it's important in looking for data science help and personnel, what else can they bring in terms of vertical knowledge?



KENNEALLY: Phil Jones, you've been nodding throughout a lot of that. I understand for you that you also see AI as opportunity for human beings. There's a headline out there every day about how robots are taking over and doing away with jobs, but you see opportunities. You see the kinds of jobs emerging in publishing that Stacy was just describing – data scientists and others. That puts publishing, if we are entering this AI era, in a competition for talent that might otherwise be working at a startup or in Silicon Valley and places like that. So how is that changing publishing from your perspective, the competition for talent? Talk about that.

JONES: Yes, you're absolutely right. I'm not quite yet building the bunker in the back yard for fear of the machines starting Armageddon. To go to your point of the human-AI collaboration, I think it's important to note that there's a lot of things said about artificial intelligence that perhaps can be a little bit misleading. There's a lot of misunderstandings, a lot of miscomprehension. AI isn't like C-3PO or HAL from *2001*, at least not yet. It's a suite of algorithms and tools that allow you to ask and answer specific questions. Certainly, I think, the ones that I've come across and I've found very interesting are things like topic modeling, assertion mapping, and those kinds of semantic things, where you can take a corpus of literature, where you can take a bunch of words, and you can do things like word frequency analyses and semantic analyses, relational things like that, and you can extract the ideas and you can kind of quantify the ideas from that content.

What's interesting about that, for example, is you might compare that practice to the old-fashioned way of producing a taxonomy. So when you create a taxonomy traditionally, you get a bunch of subject matter experts in the room and you say, let's talk about what are the fields and subfields within a research discipline, or within all of research or whatever? That's a great way to do it, and there are some fantastic taxonomies that have been built that way, but the fundamental limitation is that it's limited by the imagination and the experience of the domain experts in the room, and they're always talking about what's already happened and what's emerged as specific types of fields and disciplines and ideas that are fully fleshed out, because that's what everybody knows.

When you look at it the other way and you apply those techniques to the content itself, you allow the words that have been written to create the taxonomy. So instead of imposing one on a set of literature, you're allowing that set of literature to create that taxonomy, to tell you what's going on rather than you telling it. What I think is fascinating about that is it allows you to find things before they've been identified. It allows you to give – it gives you a few on what's going to happen next and what the emerging trends are.



The flipside is that once you've gotten that taxonomy out, or once you've gotten those bunch of topics out, they need to be curated by a person. And they need to be curated by a person who understands the research or has at least an ability to understand what each of the words means and flesh it out.

The last thing I was doing at Digital Science before I moved to Emerald was in the consultancy division, and I was working with publishers and with funders and with institutions looking at bodies of content – either awarded grants or published articles or any other kind of sets of data – and I was extracting the topics from them and then saying this is the stuff that's getting more funding, here's the regions that you need to be looking at, here are the growth areas, here are the shrinkage areas. What that then allows you to do is really interesting, strategic decision-making or the support of really interesting, strategic decision-making that's more far-sighted and less guesswork than has previously been done.

KENNEALLY: IJbrand Jan Aalbersberg, for your remit at Elsevier, it includes user privacy issues. We have, as Eefke referred to, been living in a news cycle that has raised the concern level about the data that we voluntarily, or perhaps without any great awareness, give to many of the social media platforms, the kind of data that Stacy and Phil have been talking about collecting. How do researchers feel when they understand that their data is being collected and used? What are their concerns? How do you manage those privacy concerns at Elsevier?

AALBERSBERG: That's indeed a good question, because that's one thing that I think every publisher these days has to struggle with because of the GDPR.

KENNEALLY: Not only because of GDPR.

AALBERSBERG: Not only because of the GDPR. Definitely not. From that perspective at Elsevier, we have been already for many years looking at how can we be as safe with respect to the data, with respect to privacy data, personal data, and to also the attitude how to convince the user of our products that we are really caring about their product. From that perspective, we do see the GDPR not that much as a problem, but actually as an opportunity, because it gives us at Elsevier the possibility to show how we do care about the user and about his or her personal data.

How we look at the data that we collect at Elsevier – one of the things that we try to do is store as little as possible really personal data. We try to anonymize and aggregate as quickly as possible. And a lot of data analytics can indeed be done by anonymized and aggregated data. For the recommender systems, for example, for predictable systems, you don't need to know who did what. You just need to know



what is a certain pattern in behavior. It's not relevant to know that it was that particular user or that particular user who did it. There is a pattern of the behavior, and somebody in that world or many people in the world had that behavior. That's where we learn from.

So one of the first challenges is indeed trying to convince the user that we collect their data, but we immediately, as much as possible, anonymize, aggregate it, and we don't store it if we don't really need to store it for the personal use of the user itself. I think that is a challenge that we all have, all publishers have, in these days – how to personalize, how to aggregate, but also how to tell the user that we don't store anything more than really is at the benefit for that particular user. That is our goals and our challenge.

KENNEALLY: Right. But even as you learn about the user, you may realize things about the users collectively, anonymized, and so forth, that they will want to know about themselves as well, right? Users of any kind of online platform sometimes aren't aware of the things that they are doing that the AI, the algorithm can discover. And that's important in research, because these researchers do care about the impact of what they're doing.

AALBERSBERG: Yeah, and one of the examples that we really try to do – and we are not there yet. But as much as possible across all our products, at every point of the data collection, we really want to say we collect this type of data for that particular reason, and if you say that we are allowed to collect your data, then we will give you this in return. And we will try to be very specific about that, really explaining about you give us this, and then in return, we are able to do that. So that relation is very crucial, and yeah, we have to tell that to our user – and not in a big, large privacy statement, but at the spot where we are indeed sort of communicating with the user, where the user access our pages.

KENNEALLY: Gerry Grenier at IEEE, obviously you're collecting data as well – not at the level that Facebook is about what'd you have for dinner or where you went for your vacation. How are you meeting this challenge, not only of GDPR, but of people's growing sensitivity around their data? And how are you doing what IJsbrand Jan was just describing, which is sort of making an exchange and having them understand there's real value here?

GRENIER: Yeah, I think it's important to remember one major element of GDPR, and it's sprinkled throughout the document itself, is the phrase collecting data for legitimate business purposes and establishing a relationship with a customer. GDPR, to me, is all about consent. It's all about consent, and as IJsbrand said, transparency, and conveying to the user the return that they get from the collection



of their information and the improvements in the product and the delivery of information that we can make to them. So it's about, again, consent, transparency, security of their data, and that's been the focus of our GDPR task force at the IEEE.

KENNEALLY: So this is obviously creating a greater need for those data scientists that we've been hearing about, and I guess, in your positive view of this AI era that we are entering. So we should not think jobs go away when we hear AI. There's opportunity here.

GRENIER: Oh, absolutely. I think that I've got a five-point plan for introducing AI into any company – not just the IEEE – but recommendations I would have. Five simple things. Scout AI technology. Look out there for partners. You can't build this overnight. There are plenty of partners out there, there are plenty of startups out there that we can all upon to not only teach us a few things, but also help us build products.

The second thing is to maybe work with those small AI firms to do some modest pilots. I'm thinking about – for example, we do plagiarism checking through iThenticate, through CrossCheck – the product formerly known as CrossCheck. I think it's called Similarity Check now. At IEEE, the way that we handle plagiarism checking is we assign a score – a similarity score. And then when the articles that hit those thresholds are surfaced, we have a human actually then look at that article before we send it back to the author and make the egregious claim that they're a plagiarizer. I think that we can take the plagiarism checking one step further and take that human element out of it. The human element just doesn't scale in plagiarism checking, certainly even at a publishing operation as small as – well, as medium-sized as the IEEE.

Second, develop AI experts and evangelists within your company. Encourage people with that publishing knowledge to run to meetups in your city if there are AI meetups. I live just outside New York City. The ACM does great AI and data science meetups, and it's great to meet people, and it's allowed me to go back to my company and be that evangelist. I by no means am an AI expert, but I can at least talk the talk and bring those ideas back to the company.

And then have those evangelists educate the entire organization as to the value of AI. Right now, AI is a hard sell. Semantic enrichment is not quite AI, but semantic enrichment, for me, has been a terrifically hard sell within my company to spend \$200,000, \$300,000 to semantically enrich the content. So I think AI requires a level of education within the organization to gain trust among the executive level, among the people that hand out the money.



And then fifth is attract and retain AI talent. So begin to clear the decks, I like to say. Look at your current operation. What efficiencies can you gain in your current operation to create that economic headroom for you to create a luxury of bringing on some AI people into your organization – junior AI people, at this point – to again, work with those in-house publishing evangelists and trade that publishing knowledge back to the AI people to again begin to imagine those new products.

So again, create a spider graph for yourself, right? That great spider graph with these five points. Scout the technology. Implement small AI pilots. Develop AI experts and evangelists. Educate the entire organization. Attract and retain talent. And see where you are in three to five years. If you've got some symmetry amongst those five points, you're on the right track.

KENNEALLY: What about that point, Stacy Malyil at Wolters Kluwer, about the sell on AI? How hard a sell is it? What are the points that you have to raise when you're in a meeting trying to get people persuaded that this is something we need to do? What are the things that really work?

MALYIL: I think efficiency. I think there's two levels of efficiency. The way you develop your products. How efficient can that process be by using and leveraging AI and machine learning? And then how efficient can you be in delivering right answers to your users? We have users at point of care. They can't afford to be wrong. It's life or death. But if there's so much human intervention to make sure the answers are right, where can technology help make the answers we provide our clinicians more predictable?

But getting back to the product development idea, one thing that I think has resonated in a lot of companies is that investment in tagging and taxonomies. I was talking to Rich Coppola (sp?) earlier this morning about this. As we are – the STM industry, there's a lot of mergers, there's a lot of acquisitions. We're starting to bring in bigger and smaller companies together. How do we make our products interoperable? The only way you do that is by normalizing all of your datasets so that they can work together. And the user expects this. They don't want to go from one Wolters Kluwer platform to another and have a disjointed experience. They want to be able to pull the content that we tout and we've asked them to pay for, but they want it to work across all the products that they have.

We sell to a lot of libraries, a lot of hospitals and health systems. They share access across different sites, across different patrons. They don't want it to feel like Groundhog Day every time they go from one product into another. So that



investment in tagging and taxonomies and normalizing that so that you can be interoperable – that justification goes a long way. Because at the end of the day, from a sales point of view, if you want customers to take on multiple products from you, they have to operate well together.

That's where most businesses' future revenue is from is we're in a finite market. We're not selling soap to consumers. We are selling to academic institutions, research institutions, and the employees that are housed there. We have to be able to sell them more than what they have today. So the only way you can do that is making sure the next thing you sell them works nicely with what they already have, and that is grounded in the tagging and taxonomy work that technology enables.

KENNEALLY: So the challenge isn't only selling them a variety of products, but you're selling it to a variety of markets as well. The market in the US and Europe will have certain needs and expectations that may be different. In Asia, research is growing tremendously fast in Asia, China particularly. How does that globalization make a difference, and how does the gathering of data about the researchers and your customers in those marketplaces help you meet those challenges?

MALYIL: Well, I think it's where we find commonalities in regions, so how does research – we focus on how it differs in countries, but where is it the same? I'm focused on healthcare training of physicians, nurses, and allied health professionals. A lot of differences. The scope of practice is very different around the world. But where is it the same? Because when you build it once, you can scale it globally, and that's a very compelling argument for an investment is your ability to scale globally. The only way you know what you have in common between users in the UK versus Singapore versus Mexico versus the US is by you looking at the data and having a very robust way of tracking user data so you can find those user commonalities and say if we build this or we invest in this, it can be deployed and scaled globally. But you don't know that unless you have the data to prove it. So I think that's really where it starts.

KENNEALLY: Phil Jones – someone with a research background, PhD in neurology, appointment at Harvard Medical, all of that – how do researchers feel about the gathering of their data as we've been discussing this here? Are they conservative about that? Are they cautious? They certainly want to share science. That's an important point. But when it comes to their data, how concerned are they about the publishers and others who are looking at it and knowing about them things that they may not even know themselves?



A Copyright Clearance Center Podcast

JONES: Yeah. Well, there are multiple sorts of data, of course. You mentioned experimental research data, and that's one class of data. Researchers, by and large today, want to share that information. They want to share that data. And they want us as publishers to facilitate the sharing of our data, the linking of it to articles, and all that sort of thing.

Then there's also metadata around articles and objects and things of that nature. And it's really important that that metadata is able to travel between systems in a very efficient and effective and seamless way, more to come to your point there.

And then there's the user data – things like usage statistics, behavioral tracking, and all of those sorts of things. And I think that last class, people are ambivalent about it. Because they do like those features when they work seamlessly. They do like the recommendation engine that points them to the correct paper that they're going to read that's really useful to their research. And they do like the fact that they see that video that's really useful and apropos to the thing that they're going to do next or whatever it is. But they're uncomfortable and concerned about their data being further analyzed or aggregated or shared. They worry about their email address getting into the wrong hands so they get spam. They're worried about people tracking their behavior and understanding things about them in ways that they don't fully understand. The Facebook and Cambridge Analytica is an example of people being afraid of a data activity, a big data activity, that seems to have been quite harmful, and at the same time, a lot of people are really unsure what's actually gone on.

So when that happens, people say, well, if that's what Cambridge Analytica and Facebook have done with this data, I had no way to predict that that was going to happen, and I'm not quite sure how it's affected me, what's happening elsewhere? That's a fear that we need to – first of all, we need to fully understand what's going on and what those risks are and benefits, and we need to communicate that to our users, to our researchers, and to our customers so that they can feel reassured that they are not accidentally giving us permission to do things that they don't want us to do – even though at this point, they're not quite sure and we're not quite sure what those things might even be.

KENNEALLY: It occurs to me, looking at the infographic here, the deep publishing knowledge at the center in this AI era, there are so many challenges. We see it on the slide – Brexit and net neutrality issues and GDPR and the rest of it. Does the world sort of turn to publishers expecting publishers to solve all of these problems? It seems like quite a challenge. No matter how much deep publishing knowledge there is, there's just more problems than there are solutions.



A Copyright Clearance Center Podcast

JONES: It feels like that, doesn't it, sometimes? Yeah, I think everybody in this space – publishers, librarians, funders – all feel like everybody's looking to them to solve the problem. I don't think it's literally true, as true as it might sometimes feel, that all the other players in this space are blaming us solely. I think all of these issues, whether it's around data sharing or privacy or research integrity, there is a sense that at least some of those areas, there's no firm leadership that has emerged within the ecosystem. And as a result, everybody's kind of looking to everybody else to show some leadership and to try and solve some of these problems.

The way I would put it is not so much that people are holding us responsible for something that perhaps we might not want to feel completely responsible for, but perhaps there is an opportunity to show leadership here within the larger community and add further value. That's what publishers are all about, isn't it, is adding value to the flow of information.

KENNEALLY: IJbrand Jan Aalbersberg, I suppose that's a message that you would want to echo as well, because I think you are an advocate for collaboration among publishers to address some of these issues.

AALBERSBERG: Yeah, I completely agree. There are a lot of issues. And in some cases indeed, publishers are blamed. Sometimes they are also to blame for some of the issues that are there. But I think that many of the issues that we have here, like indeed resource access, like indeed responsible sharing or sharing in a generic way, like indeed research integrity – they need to be addressed as the community of publishers, as all publishers together. And we have to work together to have one message to both our customers and our users that might be, depending on the stakeholder, different messages, but definitely aligned messages – one message to the user and to the customer that we as an industry want to secure access to science, that we want to secure that science will be communicated, that everybody has the availability to get to their science material, but also that we support the different objects in science – it can be the videos, research datasets, or the codes. But also that we take the responsibility for doing, as an industry, our utmost best to make sure that whatever we do can be trusted.

I think the key word – and then we get back to also the theme of the day – the key word is the trust, that we need to make sure that the users and the customers can trust us. And yeah, we need all means in doing that. Indeed, artificial intelligence can help us if we do it all and we use it in the proper way.

KENNEALLY: It's certainly clear that our tools are more robust than ever. The demand for answers, solutions, the demand for more data is greater than ever. The machines can crunch the numbers all day long, and they do it at lightning speed.



A Copyright Clearance Center Podcast

But so far, no computer's ever been able to create an original idea, and I think we've heard a few original ideas on this panel.

I want to thank our panelists, IJsbrand Jan Aalbersberg at Elsevier, Gerry Grenier at IEEE, Stacy Malyil from Wolters Kluwer, and Phil Jones with Emerald Publishing. I want to thank them very much. Let's give them a round of applause.

(applause)

END OF FILE