„We are creating a platform that provides GenAI and AI features for literally every product in our portfolio"

Interview mit Oren Beit-Arie, Senior Vice President beim britisch-amerikanischen Analyseunternehmen Clarivate™, zu dem u.a. ProQuest™ und ExLibris™ gehören

„Maschinen werden nie etwas ausschließlich aus Selbstzweck tun. Es gibt immer einen Menschen, der entweder eine Nachfrage oder ein Angebot oder beides schafft." Oren Beit-Arie ist davon überzeugt. Inmitten des Rummels um Künstliche Intelligenz / Artificial Intelligence (KI/AI) empfiehlt er: „Lasst die Maschine das tun, was sie gut kann, und lasst die Menschen neue Felder erschließen, in denen wir besser sein können." AI, so Beit-Arie, sei eine große Chance, die Wissenschaftskommunikation zu verbessern und gleichzeitig ein großes Risiko. Um gute Ergebnisse zu erzielen, ohne mit dem Trend der KI zu kollidieren, müsste die Informationsindustrie einen besseren Weg finden, mit den großen Technologieunternehmen zusammenzuarbeiten und wissenschaftliche Informationen in die Entwicklungen einbringen.

Der studierte Mathematiker und Computerwissenschaftler ist bei Clarivate als Senior Vice President verantwortlich für Strategie und Innovation im Geschäftssegment Academia & Government. Im Gespräch mit Dr. Rafael Ball, Direktor der ETH-Bibliothek und Chefredakteur dieser Zeitschrift, berichtet Beit-Arie unter anderem über die Strategie von Clarivate zur Integration von AI und Generative AI in die Produkte der Unternehmensmarken und über bereits dazu laufendende Entwicklungen wie den Web of Science™ Research Assistant, den ProQuest™ Research Assistant, AI für Primo sowie für automatisiertes Katalogisieren und nicht zuletzt die Clarivate Academic AI platform. Mit Web of Science™, Science Citation Index™, Alma™, Aleph, Primo und den Contentdatenbanken von ProQuest inklusive alter Features und neuer Fähigkeiten wie Document Insights ist die börsennotierte Unternehmensgruppe Clarivate in den Bibliotheken dieser Welt tief verwurzelt.

Mr. Beit-Arie, thank you very much for taking the time for this conversation. You have been in information industry for library services since your first job as a student of Computer science and mathematics at Hebrew University in Jerusalem. How did it happen?

❨ **Oren Beit-Arie** ❩ Computer sciences and mathematics was where I had interest and passion. However, after finishing my first degree, I had an opportunity to pop into a very small office at the university where there was a small team that developed a system to help libraries automate their processes. I got what I thought was a temporary student job before I decided what I was going to end up doing next in terms of education. Little did I know that this job would carry me to where I am today. I started as a developer.

What kind of project did you develop with this company?

❨ **Oren Beit-Arie** ❩ Aleph Integrated Library System. This was the late 1980s.

Okay, well.

❨ **Oren Beit-Arie** ❩ What turned it from perhaps a short-lived, summer job for a student, to becoming more a part of my life is that a year into my job, I had the opportunity to spend some time in a library in Copenhagen at the Danish Technical University.

Oh, how wonderful. Good address.

❨ **Oren Beit-Arie** ❩ Yes. The DTU was one of our first customers at the time. We were a small company with little presence in Europe. I met a lot of visionary people there. I spent a year there working in the library with the librarians and helped build the system to solve their problems. That really gave me the direction that there is a lot of things that we can do with the aid of technology to help libraries develop and deliver on their mission. And I would say there's also a personal note. My father used to be a national librarian.

So it's a genetic disposition, so to speak. Oren, coming to the use of technology, the aid of tools, and the aid of technology for the libraries and the librarians and services. What do you think: Do we skip out the people of this business, and will we end in a fully automated library? Or will we need people in the library of the future as well?

❨ **Oren Beit-Arie** ❩ Even in the realm of AI now, there's a very important term, „the human in the loop" which we strongly believe in. The human in the loop has an important role to play: they trigger the work with the AI models, guide, test and check its work and performance and then, of course, also benefits from the output of that. I think over time we'll have more and more automation, but I think we'll have more and more things to automate

– we will expand the scope and range of things we could do in support of institutional missions. Humans will have a central role in directing and triggering work in those new domains. As much as I know that we all like to think about the future in which machines take our place, I don't see this happening anytime soon. In fact, what I have seen over time, certainly in the last 20 years when automation advancement happened, was the introduction of the ubiquitousness of the internet, mobile, and other technologies. Roles change. We need to change our skill sets. We need to learn new things, but we haven't been replaced.

This is a hopeful view, isn't it? We will come back to AI later on. Before that, one question about the Clarivate mission and self-assessment. On the website landing page Clarivate states: „Let our intelligence move you however you want to change the world. We'll get you there." From our perspective, this feels a bit of high claim. If I am your customer, what would you offer me to fulfill this claim?

❰ **Oren Beit-Arie** ❱ I think at the core here, this reflects our belief that research, education and knowledge have the power to transform society. I think it's a key critical ingredient in what makes societies better and what makes us develop and create a better future. In that context, what makes this happen is the accessibility of the knowledge to really drive the innovation cycle and the knowledge cycle, among other things. You need tools, you need access to resources and you need tools to make you more efficient as you go through those processes, whether it's a process of research, a process of teaching or a process of learning, and that's what we do. This is our role as a company.

So you are saying that science and knowledge transform society and play a very important role forever in societies and your company will be part of this transformation of this knowledge creation in the world. Web of Science plays an important role in analyzing material and analyzing the output of research. As a librarian I highly appreciate this quantification of science. On the other hand, we have a movement in the world to focus on the evaluation of science on quality rather than quantity. Do you think it could be the end of quantifying systems?

❰ **Oren Beit-Arie** ❱ I don't think it's a binary question or a binary answer. I think the importance of citations is not diminished by the fact that we have more ways to view and more needs to assess the quality of science. At Clarivate, we always took the position that, for example, the Journal Impact Factor™ is a journal-level metric and our Journal Citation Reports contain a whole variety of met-

Clarivate SVP Oren Beit-Arie (l.) und Rafael Ball (r.) in Hamburg, wo sie sich Anfang Juni am Rande der 112. BiblioCon zum Interview trafen.

rics, visualizations and descriptive data. In other words, I believe that the existence and the rise of new evaluation criteria, both quantitative and qualitative, are natural. But I don't think we are in a position where we say we don't need the others anymore. It's just a matter of building more insights and driving more insights on how to evaluate science. There is a rise in the need to evaluate, provide, and look at research's economic and societal impact. A lot of that is still loosely defined.

I totally agree with you.

"GenAI tools – large language models – are very good at their language capabilities. They're not very good about their knowledge of the world."

❰ **Oren Beit-Arie** ❱ I think we will end up with a healthy mix of quality indicators that will help evaluate the products of science. Some of them will be more quantitative, and some of them perhaps will be more qualitative.

Another point that is touching on your products as well: We are living in a sharing economy in the open world. Open science is not only a buzzword but even a really heavy demand from the political side. At the BiblioCon24 conference, there were a few talks on the new approach of open databases, OpenAlex as the alternative to the proprietary system, such as Web of Science or Scopus or Dimensions. What is your argument that a proprietary system, a commercial system perhaps is better than the OpenAlex system?

❰ **Oren Beit-Arie** ❱ I think the approach of openness is a very healthy one. I think it incorporates a lot of compo-

ZUR PERSON:

Oren Beit-Arie zeichnet seit August 2023 als Senior Vice President Strategy & Innovation im Segment Academia and Government für die strategische Ausrichtung und Weiterentwicklung der Produkte der Clarivate[1] Unternehmensmarken verantwortlich.

Beit-Arie hat an der Hebrew University of Jerusalem Mathematik und Computerwissenschaft studiert, dort seinen Master of Arts (MA) 1988 cum laude abgelegt. Das Masterstudium in Theoretischer Linguistik an der Tel-Aviv University schloss er Summa cum laude ab (1994). Bereits sein erster Studienjob führte ihn in die Welt der Bibliotheken. Er arbeitete als Programmierer am Aleph Integrated Libray Systeme. Seither befasst er sich mit der Entwicklung innovativer Produkte und Services für Bibliotheken und Unternehmen. Von September 1998 bis Juni 2018 war Beit-Arie 19 Jahre und 10 Monate bei Ex Libris, die letzten 12,5 Jahre davon als Chief Strategy Officer. Nach der Übernahme von Ex Libris durch ProQuest war er von Juni 2018 bis März 2022 zunächst Chief Strategy Officer, später zudem President Books bei ProQuest. Im Zuge der Übernahme von ProQuest kam Beit-Arie im April 2022 zu Clarivate.

1) https://clarivate.com/about-us/

nents that we are very clear in supporting. I think there's great value in transparency in making sure that we share common beliefs about how we, for example, select and evaluate research. Again, this is another example of how I think that for us to be successful as a society, we should not decide on one approach. This is not going to get us better. I think this is going to get us into a narrower thinking.

The way that we at Clarivate are thinking about this is that there's great value that we believe we continue to bring into that ecosystem. We're doing a lot of work, and putting a lot of focus on the curatorial work and the selection of the content that gets into our core collection. We put a lot of focus on our editorial processes, making sure that we enrich that content with the necessary interconnections that enable you to create networks of information, not just static databases. We use a combination of technology and people. I think we're one of the only out there who believe that a healthy combination of automation and editorial expertise is key to bring forth the

1   https://clarivate.com/pulse-of-the-library/

best solutions. We are also participating in that realm of openness in a number of ways. For example, we contributed the code, from Researcher ID, as a founding sponsor of ORCID. We are a part of the movement that looks into, in a transparent way, issues around integrity. We're very active in the community, but we still strongly believe that editorial and curatorial processes play an important role. That's what we do at the core of the Web of Science.

What are your concrete plans and ideas for implementing AI in your products? I'm sure there is artificial intelligence already working in the Web of Science as well as with ProQuest and Ex Libris products and services. As a librarian, I would be very happy if our library system, Ex Libris, would use artificial intelligence even more effectively than it is now and to make the search process even more sophisticated than it is at the moment. Is your company doing anything in this regard?

❰ **Oren Beit-Arie** ❱ Absolutely! In fact, we are in the middle of a meeting with a group of directors, and the main topic is artificial intelligence, so there is a lot to share here. First of all, let me start by saying artificial intelligence is not new. It's been around for decades. I'm kind of telling you my personal story. I did a master's degree in theoretical linguistics with a focus on computational linguistics, which at the time incorporated methodology.

This was artificial intelligence, wasn't it? Data mining and so on.

❰ **Oren Beit-Arie** ❱ Exactly. Neural networks, if you recall. And for a long time Clarivate, as well as Ex Libris, used artificial intelligence in multiple use cases. Of course, the difference certainly in the last year and a half since the public introduction of ChatGPT is the accessibility and the ubiquitousness and the common use of generative AI models and tools by everyone, including students, faculty, and researchers. What's new now is the concepts of generative AI and the availability of large language models. The way that we are thinking about this is that this has the potential to transform the work that we're doing, the mission of institutions, and the way that they accomplish this mission. You have asked about our plan. I want to contextualize our plan by the recognition that we spend a lot of time talking to many institutions globally. We are present in 99 percent of the top 400 institutions in academia and governments around the world. More than 26,000 libraries use at least one of our products. We have great global exposure to academic and research institutions. We are having many conversations with them and we just published a Pulse of the Library report[1] which captured their thoughts on the opportunities and concerns about AI.

The concerns revolve around two main areas. One is the quality of AI. The other category is the appropriateness or the integrity. Under the tagline quality I would incorporate the hallucinations, made-up facts, biases, and lack of attributions. They're really not on par with our notions of academic and scientific work. The quality is a real concern certainly when you use tools such as ChatGPT, et cetera. But even if the quality was 100 percent, which it would never be, questions arise like "Is it appropriate to use the tools?" That gives rise to issues of, for example, plagiarism and paper mills, where factories generate research outputs for publications that are fraudulent.

We have worked on AI before, but since last year, we have a very concentrated focus on GenAI in particular, not just as a feature, but as a fundamentally new technological backbone that we're building across our portfolio. We're doing this by introducing the Clarivate Academic AI platform. This platform provides consistent and reliable implementation of GenAI designed for academic use cases that can be used and embedded in literally every product in our portfolio. We're building this as an infrastructure that is a service available to all our products.

So, you are saying this will be part of all your products?
❰ **Oren Beit-Arie** ❱ It's a new infrastructure backbone that empowers GenAI and AI in general into our portfolio. What does that mean? We work with large language models, and we enable a secure and private endpoint for those large language models and use RAG, Retrieval-Augmented Generation architecture, to ground all GenAI features in our curated content rather than the LLM knowledge. We enable capabilities such as conversational discovery, recommendations, translation, article key takeaways and concepts, and other GenAI features that can be serviced to all our products. This is a focus I started in my position so that we could introduce it in different use cases across our portfolio in products as diverse as Web of Science and Alma. Before I give you the product examples, I just want to add one more point about our strategy. We see AI as important in three categories: One is that we believe it's a new technology that will be incorporated into almost every product. We'll use natural language conversation more and more. It will be just an evolutionary path that we will take. That's one pillar of our work and our strategy to introduce AI into products. The second pillar of our strategy, and the important one, is that we believe that now with GenAI in particular, we can start solving problems that were very hard to solve before, and we can do it in a better way and on a greater scale. Again, all this is about solving problems for researchers, students, and librarians. For example, we are introducing a series of research assistants, such as the Web of Science Research Assistant. What it does is, it gives you the ability as a researcher or a

research administrator, not only to discover the wealth of the Web of Science collection in a different way through, for example, conversational, like the chat discovery, but it also helps guide the researcher through multiple tasks. For example, if you are a researcher and perhaps you're interested in a novel kind of topic or interested in understanding what the trend looks like, in terms of publication on a certain topic, we're going to guide you through visualization and guided workflows to conduct that inquiry. Another example is if you're interested in doing a literature review, the Web of Science Research Assistant will not only help you perform the discovery but also the careful discovery of literature on a specific topic with the aid of things like topic maps, citation networks, visualization of

those concepts and so forth. This gives us the possibility to do things that we couldn't do before. We're going to introduce other use cases of GenAI as well.

Let me give you another example related to the ProQuest platform. As you know, we have primary resources collections of journals and books. We've introduced a new GenAI based capability that we refer to as Document Insights, which means that you can get insights at an article, book or dissertation level. You will be able to use generative AI to answer questions like, „What are the key findings of this article?"

So the Document Insights Assistant or feature gives you an abstract?
❰ **Oren Beit-Arie** ❱ It gives you article-specific insights such as key takeaways, key concepts, suggested research topics, other recommended articles, and so on. We also plan to introduce it in Primo for library discovery.

Let me point out two more areas. One other area is in operations. We think that AI, in general, and particularly GenAI, can help with productivity and enable efficiencies in processes, such as cataloging. We are now working internally on developing an AI-based tool that will help to enrich catalog data, for example, to scan a book cover or digital image and extract metadata, such as language and the Library of Congress Subject Headings.

Do we need human-driven cataloging anymore?

❰ **Oren Beit-Arie** ❱ Going back to the concept of a human in the loop, I think what humans will continue to do is, control, guide and test the quality of that process. As I said before, I think it will also help us direct and refocus our skill sets to do things that we couldn't do before because we didn't have the time or resources. Let the machine do what the machine can do well, and let the humans start developing new fields in which we can do better. Let me add one other category. I think for the whole area of research integrity that I mentioned earlier, introducing AI is an important component. We're going to use AI to help identify integrity issues Also, in a reverse way we started by talking about the new metrics that are being developed, societal metrics. AI will be very helpful for us both to gather information and create narratives, use cases, case studies, et cetera.

We could go on discussing AI developments for hours, but we have to come to an end, at least for this interview. So please tell us: What is your vision for scholarly communication in the future? Do you think there will be a fundamental change in the next 10 to 15 years?

❰ **Oren Beit-Arie** ❱ I hope so. I think we're at a very interesting point. One of the things that I didn't talk much about is that our implementation of GenAI is based on the notion that GenAI tools, large language models, are very good at their language capabilities. They're not very good about their knowledge of the world. We started by talking about the problems that we are hearing in academia and research, and that is hallucinations, made-up facts and inaccuracies. The way that we solve this is by not using those language models for their knowledge, but rather for their language only. The way we do that is by applying the language models to our core trusted authoritative collection; for example, the Web of Science Core Collection™.

We are using RAG architecture which enables the best of both worlds. You use the language model for the language capabilities, but you basically ignore its knowledge of the world, which often introduces what's referred to as "hallucinations". We only use our trusted scholarly content to inform the answering of the questions. That's a very important point. In the context of your current question, one of the things that I think is we're still in a very early stage, and let's remember that in many ways, the point in time that we are right now with regard to AI, to take an analogy of the mobile revolution is similar to the days of the iPod, if you remember, before even the iPhone. There is a lot more that is going to come in terms of technology advancement. GenAI also introduces opportunities for us to benefit from new tools that will advance research, be more efficient and create outputs that are being dissem-

inated more effectively. Also, technology companies who develop generative language models and applications realize more and more the importance of trusted and authoritative knowledge to the quality of their outputs. We must work to turn it into an opportunity that could benefit both research and scholarly communication.

One last time back to the basics of scholarly communication. What about books and journal articles? Do you think scientists will, in 20 years, also publish a seven-page standard article in the journal, which is branded by Elsevier, Springer Nature, or something else, and then it's distributed and disseminated? Or will there be some new technology-driven explication, well, let's say, of the knowledge in the output of scientific developments? Do you have any visions for this? Or do you think it is quite really acceptable to go on writing books, journal articles, and make database entries?

❰ **Oren Beit-Arie** ❱ I don't think that traditional publishing will entirely go away, but I think it will have to evolve. I think that we'll use new technologies for production, dissemination, and consumption. We'll have more variety of publishing models and more blended modalities of publishing, not just textual or just video. I think it will be more incorporated and combined. We will be more relaxed about the aid role of technology in producing some of the content, but I still think human beings will be in charge of putting this together and signing off on the results.

It's very interesting that you talk about production, dissemination, and consuming information and knowledge and not about writing, dissemination, and reading. Writing and reading are only one format or one technology of the future, but there will be different and new formats of production and consumption, on the other hand.

❰ **Oren Beit-Arie** ❱ Right. I think the writing will be more collaborative in many ways, and we'll use more technology assistance. As you say, it's a subset of the production. I think that there will be more human reading but also more machine reading, and we need to work together on the copyright issues to enable the consumption of materials to machines, the interoperability.

My vision of a dystopia of academics communication is the production of knowledge by machines, for consumption by machines, and humans are kicked out.

❰ **Oren Beit-Arie** ❱ Machines won't do it just for their own good. There's always a human that either creates a demand or the supply or both. But I think that there certainly is going to be more machines in the value chain.

Absolutely. Oren, thank you very much for the insights.