



Government Policy, Science Research, and Machine Authors

**Dr. Christopher Tyler, University College London
Rachel Martin, Global Director of Sustainability, Elsevier**

**For podcast release
Monday, January 15, 2024**

KENNEALLY: Last spring, not long after OpenAI launched ChatGPT, an AI working group in the US House of Representatives obtained 40 licenses for the generative AI tool. ChatGPT and other available AI tools now conduct systematic review of scientific literature for government officials by searching millions of information sources. And the machines are expected to do much more in years ahead.

Welcome to CCC's podcast series. I'm Christopher Kenneally for Velocity of Content.

Members of government and public policymakers around the world rely on science and science publishing when shaping regulation and legislation. The responsibility to stay current on research is a formidable challenge for the public sector, especially as the volume of science publishing grows. Ethical concerns, of course, temper the enthusiasm over AI. Congressional staff, for example, must limit their ChatGPT use to research and evaluation only, and they should only input non-sensitive data.

Dr. Christopher Tyler in the Department of Science, Technology, Engineering, and Public Policy at University College London has written for *Nature* about the powerful potential of AI in developing science policy. He joins me from London. Welcome to the program, Dr. Tyler.

TYLER: Thank you for having me.

KENNEALLY: You identify two key areas that hold promise for improving the practice of policy guidance – synthesizing evidence and drafting briefing papers. Based on your research, how well do the AI tools currently perform at those tasks?

TYLER: That's a good question. They're not there yet. However, I did stop my research before Christmas, and it's now after Christmas, and things are moving extremely fast. There were certainly aspects of policy drafting that are completable in sections by tools like Bard and ChatGPT and so on, and I know from speaking to colleagues at Google and Elsevier and elsewhere that tools for evidence synthesis are coming on very, very rapidly as well. So to



say that they weren't ready in December doesn't mean that they won't be in use in February.

KENNEALLY: What are your concerns, Dr. Tyler, about using AI in public policy development?

TYLER: Well, I think the big one is around trust. There are concerns about the black box at the moment. The information goes in. There is a black box in which it's processed. And then it's spat out and used by a science advisor or a policymaker, whoever it might be. Developing trust and transparency in those what are currently black boxes is going to be essential.

Another thing that I think I'm worried about from sort of a professional perspective is that at the moment, a lot of policymakers are very reliant on science advisors for evidence synthesis. But as they get to do their own evidence synthesis rapidly with these forthcoming tools, we do run the risk of empowering much more policy-based evidence rather than evidence-based policy. That's certainly a risk that we need to be aware of. One of the ways that we'll be able to do that is to ensure that the science advisors get better at their jobs through the use of these new tools rather than sticking with a bunch of tools that they've been using for the last 40 years.

KENNEALLY: Dr. Tyler, what is important to know about the human element in this work when we are using these AI tools? Is it going to change the way that policy is created?

TYLER: I think that it will, but in a mechanistic kind of way. Ultimately, policy is a mechanism for making decisions for people by people. That isn't going to change. That component of society and humanity isn't going to change. One thing that's going to be really important is that there are people involved in the process all the way through – that we don't just hand over complicated policy decisions to a computer and hope for the best. That isn't going to provide the kinds of outputs we need, not least because ultimately it all comes down to politics, which is a human endeavor in the first place.

So we need to be thinking about these tools as ways to make policymaking (a) more efficient – can we get from A to B quicker with fewer human work hours put into that process – and (b) can we make better decisions? Can we, for example, come up with solutions that really achieve the goals that we're trying to achieve, and can we avoid as many of the unintended negative consequences as possible?

KENNEALLY: As the public becomes aware of the use of AI tools in policy creation, what does that do to their confidence in government? It's already at a low state. Will using machines to create policy only exacerbate this distrust?



TYLER: Yeah, I've heard that that is something that people are concerned about, and I think it relates to that black box issue I was referring to earlier. There's sort of an underlying assumption, I think, in a lot of public policy that more transparency equals more trust. I'm not sure that's entirely borne out by the evidence yet, but that's a separate question for a separate day.

I would also add that the nature of the epistemic underpinning of evidential policymaking is not something that the majority of people are disgruntled about government with. There's a much more complicated problem than that. So I don't see AI tools as being problematic in this at all. What we do need to be aware of, though, are these issues of trust and transparency and risk and making sure – government has to be both a user of these tools, but it also has to demand of the big tech providers of these kinds of tools that they do meet certain levels of transparency.

KENNEALLY: Early in your own career, Dr. Tyler, you were a science advisor in the House of Commons. Would you have wanted ChatGPT to help you do your work?

TYLER: Oh, a thousand times yes. It would have been fantastic. I can't tell you how long I used to spend doing things like scoping new inquiries for select committees, where I would have been able to just throw into ChatGPT a question like – as we did in the *Nature* paper, actually – give us 10 ways in which there is an evidence base for cutting crime rates. And it spat it out. Now, some of those examples were really obvious. Some I would have been able to come with myself. Some of them were wrong. And some of them are things that I didn't think of that were really valuable. That's just one example of the many ways in which as a tool, it can speed up some aspect of a process that has been slow for 40 years.

So what I think we'll probably find is that these kinds of tools will speed up a lot of kind of the donkey work component of science advice to enable people like me back in the day to spend more time face to face, more time crafting bespoke briefs for individuals, more time making sure that the evidence synthesis met the exact need of the policy questions that we're being asked, rather than just scrambling for information the entire time.

KENNEALLY: Dr. Christopher Tyler in the Department of Science, Technology, Engineering, and Public Policy at University College London, thanks for speaking with me.

TYLER: My pleasure. Thank you for having me.

KENNEALLY: Rachel Martin, Elsevier's global director of sustainability, served on a team that developed a proof of concept project testing the suitability of gen AI narratives for advisors



and their clients in government. She joins me now from Elsevier in Amsterdam. Welcome back to the program, Rachel.

MARTIN: Thanks, Chris. Thanks for having me.

KENNEALLY: What research and narrative elements were determined necessary for the proof of concept?

MARTIN: So this is a super-interesting question, because when we sort of approached it as our team, I think there was an assumption with our data experts that there was some sort of magic generic policy template that existed and that we just basically needed to pull out the snippets and we would be able to write a machine-readable piece of paper or a policy brief. Of course, that's not how it kind of works. So in terms of the research elements, we started talking and we started looking around, and there's some best practices. But what we found was actually, there isn't one policy brief. They're not even called policy briefs in many different scenarios. They differ per country, per ministry, per office, and even per person on their own personal preference. So we kind of really had to think about, OK, what would be – as Chris also highlighted – that initial input, and in a generic way that you could then help speed up that process of vetting?

We eventually came to the idea that you needed an executive summary. You needed some recommendations and main text. You needed some policy implications. And probably most important of all, you needed a conclusion, and of course, the references. Even that was sort of an interesting dive into the world of policy advice, which baffled us, I think, from a data perspective, because we always – you know, we're a publisher. We're used to abstract, methods, results, conclusions. This seemed very simple, but it wasn't necessarily that process.

And when we sort of thought about the narrative side – so what kind of machine-generated report did we want – it was kind of interesting, because we assumed that we would want sort of a professional – we wanted to make sure that it read really well. We wanted to use active language, because nobody wants to hear something that isn't interesting to read. And we also said let's make it layperson. Let's make it simple. Let's also get rid of jargon and abbreviations.

So we kind of defined this, and we said, OK, all of that needs to be packaged within like four to six pages, because these summaries – you only have to look at the IPCC. That's 100-odd pages. Some of these briefs are 20 pages. Some are one and a half pages. So we thought, OK, four to six pages seems like a good kind of input criteria, basically. That's what we defined for the proof of concept. And we did it on a specific subject. We did it



on lithium battery supply in the EU, which was a topic being discussed in the EU at that time. So we wanted to kind of make it as realistically as possible.

KENNEALLY: Working with artificial intelligence tools, of course, the choice of data source is going to be critical. Talk about that. Talk about why that's so important and what kind of choice you made.

MARTIN: Yeah, exactly. I think as you said at the beginning, ChatGPT is a black box. There's lots of different information sources on there. What we did in our proof of concept was just to run it on peer-reviewed articles. They're an amazing resource. They're already peer-reviewed. They're vetted. Publishers have worked very hard to make sure they're incredibly structured. So any data scientists working on publishing data, they love it, because it's super-structured. That was really an important part for us is to make sure that this was using just science. We weren't using tweets or blogs or other information. It was just scientific articles.

The other thing I think really is important to think about is that when we were thinking about the data sources, we also had already done some work two years prior when we produced something called the Clean Energy Report. We had already defined with a set of experts key words defining that area of renewable energy. So when we were looking at lithium battery supply, we'd already done a lot of the hard work. So this was something that we could reuse – again, with that human element coming in, letting us know that these keywords relate, that this is going to happen.

And I have my funniest story. So Marius, who's on our team – he tried it. He built the pipeline. And I got this email all in caps going, oh my God, it's not bad! (laughter) I was like, that's amazing, right? It was because we had done all of that basic work. It was trusted. It was peer-reviewed science. What came out really quite quickly was really good, better than we thought. And at that particular stage, we were very excited about the possibilities.

KENNEALLY: Better than you thought it would – well, that's a pretty good rating. But what did readers think of the machine-written documents, Rachel Martin?

MARTIN: Oh, that was even more fascinating. And again, I think this was a journey on challenging our own assumptions. One of the biggest things was that everybody said it reads well. Nobody thought, oh my God, a machine has written this at all.

I think what the main feedback came was it's too layperson. So I think as publishers, and particularly in areas like sustainability or climate change, we're like, oh, we need to make it more simple. But actually, it really kind of made us think that we have to challenge



those assumptions. The people working in the policy offices have a scientific background. They actually really do want that detail, and they want it in a lot more complication, a lot more of what we would classify as jargon. That was a big learning experience for us, and I thought it was really interesting. Eventually, we used something around 18,000 records, and one of the interviewees said to us, for 18,000 records, I expect a lot more detail in there. It was this idea that if you have this huge body of knowledge, this must be like an IPCC report. So there's some expectation-matching which I think will be interesting as this evolves as well about how people receive the information and the level of detail that's there.

I think the final thing that everybody talked about, and Chris alluded to this a little bit, is around the pinpointing of certain aspects. People wanted data. They wanted a clear number. And they wanted that to be citable. They wanted to be able to click on that and go to that document and be able to say, OK, this study says that it's this number. In the interviews, what came back there was at the moment, again, with social media, ChatGPT, trust in science, trusting governments, sometimes these political discussions – you know, 95% sure isn't good enough in some situations. That 5% can undermine a whole debate. So the idea of where do these facts come from – and again, back to the data source – having open data or having the article linked to the data – you know, all of these elements come into it, and you suddenly realize that this is a lot more complicated. It isn't just a simple, hey, ChatGPT, please write my Christmas menu. This is far more detailed and far more nuanced if it's going to work and work at scale.

KENNEALLY: Rachel Martin, Elsevier's global director of sustainability, thank you for joining me on Velocity of Content.

MARTIN: Thanks so much.

KENNEALLY: That's all for now. Our producer is Jeremy Brieske of Burst Marketing. You can subscribe to the program wherever you go for podcasts, and you can find Velocity of Content on YouTube as part of the CCC channel. I'm Christopher Kenneally. Thanks for joining me.

END OF FILE