A practical guide to AI tools for life scientists

Dr Michael Kuiper, CSIRO

8 May 2024

Questions asked during the webinar

Question	Michael's Answer
As individual researchers, as research institutes, as a country, how are we going to keep up with the big players in terms of the input and also the costs of subscribing to multiple agents?	We think about that a lot, too. It's incredibly difficult to create something new, it costs 10s of millions of dollars, if not hundreds of millions of dollars to make these large language models. It depends what you're trying to do. It's really hard to compete. You have to really pick your lane in this area. The key is keeping up to date with the advances. Being agile. If you're trying to implement it, I think it's probably easier to implement what's already out there. So have a subscription model or perhaps you could talk to your IT people and get a local model delivered as well. So the Llama models are all open source. So if you need that capability locally, you can do it. It won't be the cutting edge performance, but it'll be pretty good. There's a new Llama three 400 billion parameter coming out. So that'll be probably around GPT four level.
While you were giving your chat GPT 4 demo you uploaded some CSIRO data. Where does that data go? And is it publicly available? Or because it's your own agent is it hidden away?	When I launch data to ChatGPT-4, it gets uploaded into their model, and then gets processed there and is shared in the cloud. You can run local models to avoid this but it takes a lot more hardware.

	The data set I used is already publicly available from the CSIRO data portal so I wasn't worried about sharing it with the model as it's probably already trained on it.
So how are we possibly going to keep up? It's changing faster than we can learn. And where would you go for ideas around governance in this space?	That's a good question. Certainly the National Artificial Intelligence Centre, they have a lot of discussion around that.

Find out what's happening locally and find people to talk to and share ideas with. For example at CSIRO we have Tuesday morning AI, where we share our experiences with using various tools. |
| A major challenge in using AI is how do we handle the blackbox nature? How can we check the validity of complex methods that it's using to generate the results? How can we maintain our academic rigour? | I don't think I have a really good answer on that. I think there's going to be certain test cases like a test suite that we can reiterate over, just basically to check that it's giving the good responses, but it's going to be really hard unless it's an open model. For example, the Llama3 model - we won't actually know what's inside or how it works.

If this is really important to you I would say stick to the open models. |
| In what ways can LLMs or AI be used in drug development or discovery ? | Many new ways! LLMs significantly enhance drug development and discovery by streamlining target identification, designing effective molecules, and screening compounds to name a few. |
| How do you weigh up costs of LLM agents, when subscriptions are required to a handful if not dozens of different tools?

Do you see institutions moving to adopting business subscriptions to any of these? | Good question, - I'd do an experiment. Try out a paid version, and see how much more productive you are with it. It doesn't help everyone. But for a dollar or so a day I find these things greatly help my programming for example. It saves me hours

The business case could be made to institutions after the experiment related to productivity increases. One thought I had is, people can subscribe month by month. Those that use it will renew. Those that don't won't thus saving dollars where there is no usage. |

chatGPT3.5 often creates false citations. Is there any prompt we should use to avoid this?	True. I am turning more to agent based modelling where agents check each other's work. Also use other for-purpose tools like Perplexity and Elicit to find references. GPTs are actually chatbots, not librarians, though they can do that job. We shouldn't expect the older models to be good at everything. :)
Can we use AI to predict the binding site of small molecules?

For instance, if you have a small molecule that inhibits HIV RT in vitro but you don't know where it binds (i.e. X-ray crystallography is challenging). | And more so. Look at Alphafold 3, (which came out the day after the webinar) alphafoldserver.com

Isomorphic labs are doing that right now. |
| Are there rules or best practices in terms of using AI for coding? For instance, while using Google collab, sometimes it auto-completes the code and also provides error explanations and solutions.
Then, it says something like "these suggestions may be subject to license." How does the licensing clause influence the use of gen AI as an assistant in analysing biomedical data (interpretation, report/paper writing, etc)? | Good question. Generally I try to use AI as a guide, but try to write most myself, or at least make sure I've understood everything it writes. It will be harder to spot over time. Can you tell if someone used a spell checker or calculator for their work for instance?

If a licence is required you can always ask for alternatives from the AI or state to specifically not use licensed material. |
| Do you have any advice for someone with bioinformatics experience who would like to transition to working within the AI space? | Play! Get yourself a paid account and push it to the limits. Practice prompting and conversational AI. Get out of search engine thinking and chat with these as a really knowledgeable colleague. Don't assume you know anything about your topic and get the AI to explain things to you. This builds up the context window and unlocks adjacent knowledge domains. |
| How does use of AI tools impact innovation in young scientists? | It can be double edged as many technologies are.

If you are lazy, I suspect AI will dull your actual understanding of things, but you may think you are informed. (See this Nature paper I mention towards the end of the webinar)

If you use it to broaden your knowledge, and test your understanding, it may lead you to think more creatively. Innovation is |

| | often about combining diverse ideas in new ways.

Just don't be a lazy thinker. Check sources and truely try to gain an understanding of things. :) |
|---|---|