Opinions Libres
le blog d'Olivier Ezratty

Using Google NotebookLM with "Understanding…"

I published the open-source book "**Understanding Quantum Technologies 2024**" late September 2024. Since then, I have updated it many times to follow the news, add a couple companies and also correct various spelling, presentation and other errors. I try to keep up with the scientific and other news until I publish this edition on arXiv and then on AWS in a paperback edition in 5 volumes (or parts), around November 2024.

The book length may be intimidating (1,552 pages) and you may have no time to read it entirely. I provide a 26-page zipped version which contains the key highlights from each part. But you may have many questions who need answers. Sometimes simple ones, but which require digging into many parts of the book to get a sound answer.

Some of you may have created a GPT with uploading the book or its five volumes (or "parts"). It can help you navigate in the book and ask questions about it. Recently, Google released an experimental tool named **NotebookLM**. With it, you can also create workspaces where you upload up to 20 documents and then can ask any question related to the uploaded documents. In the past, I had tried similar and more structured tools like **Perplexity** but I was less impressed.

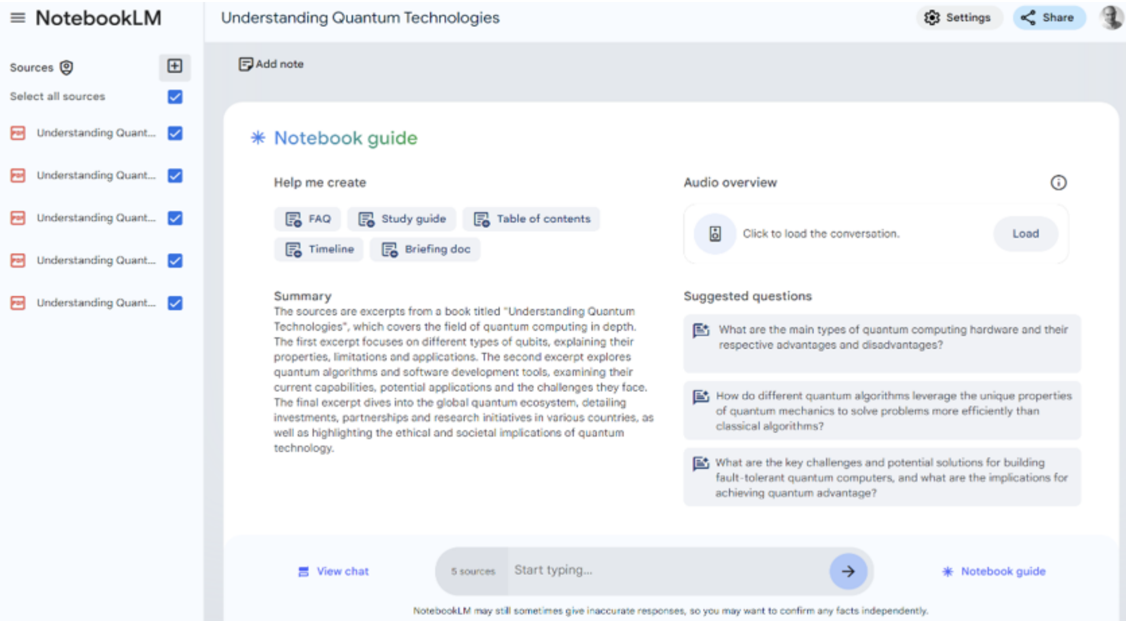
Its performance seems very good and better than with a GPT. It uses two main techniques: a RAG (Retrieval-Augmented Generation) to access the documents and an LLM to format your questions and the RAG answers. RAG vectorizes your document content in a database. NotebookLM provides responses to your questions with series of numbers pointing to the text extracts from your documents that were used to craft it. As a result, answers are traceable and explainable. If it makes mistakes, you'll see why.

The results are quite spectacular. Its capability to consolidate information scattered in different parts of the book is really cool. Sometimes, LLM-based chatbots give the impression they can reason and to understand what they are talking about. It is not (yet) the case, particularly with a RAG coupled with a LLM. Reasoning still exists in ChatGPT and the like due to their use of agents and chains of thoughts. The recent o1-preview from ChatGPT is able to decompose a question into sub-questions, showing basic reasoning skills. One must avoid projecting some anthropomorphic skills on these tools. These are machines using brute force to consolidate information from many sources.

Book ingestion

NotebookLM is a prototype solution that is free to use. You just need to login with your Google/Gmail account. Of course, it is not advised to upload confidential information in such a tool. My book being public, there is no such problem here.

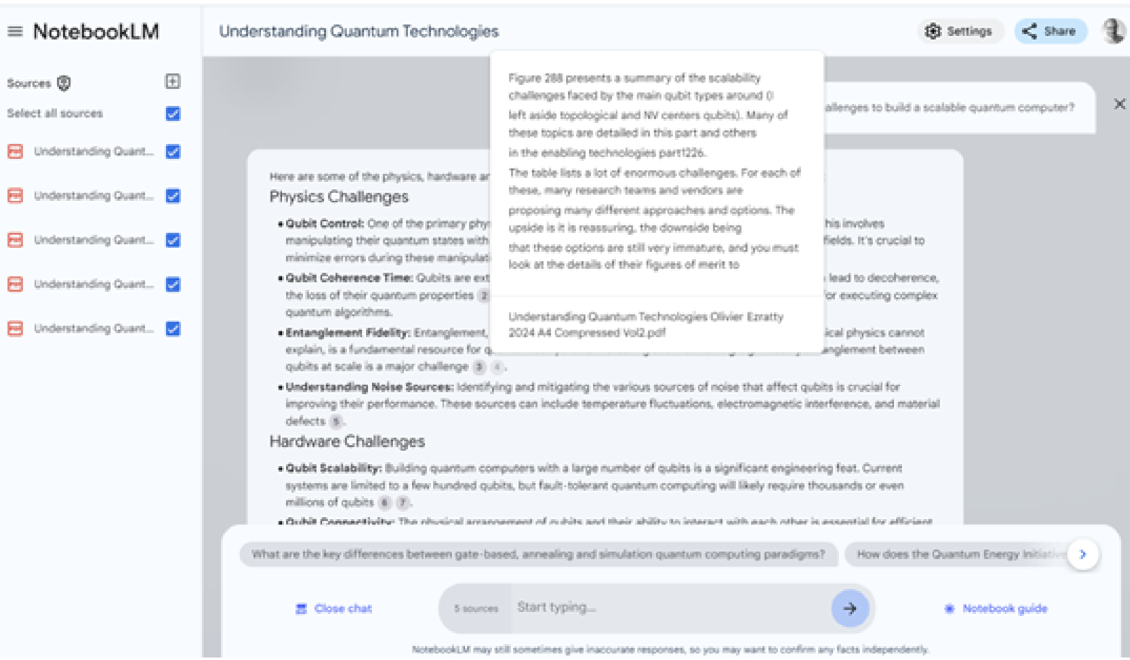
I uploaded the 5 volume PDFs from "Understanding Quantum Technologies" to see what it could achieve. The files ingestion lasted less than a minute.



Questioning "Understanding Quantum Technologies"

So, here are the first questions I tested with a rating from 1 (bad) to 5 (very good). You can try this yourself using **this link** which I hope will last long.

- **What are the physics, hardware and software challenges to build a scalable quantum computer?** (5) You can see below the tooltips showing the source of the information in the uploaded files. The response is fine, listing many relevant scientific and technological challenges.



- **List the academic labs that are the most frequently mentioned in the bibliography with their specialty** (4). The list is well sorted by qubit type with key academic labs but not complete.

- **Can you sort this list per country and provide more details on the labs whereabouts?** (4) The list is fine, not complete, and contains some industry vendors which shouldn't be mentioned. Extracting list of people, companies or whatever structures is not working well with these tools at this point. The results are never exhaustive. For example, when asking the list of the mentioned companies in the book, NBLM produces only about a hundred of them when there are over 750.

- **List the various quantum computers created in China** (5). They are all included.
- **What quantum computing case studies are in production in the industry?** (5). It answered: "*The sources state that, as of 2024, there are no quantum computing applications running live in production in any corporation. Most publicly available case studies are proof-of-concept projects, software prototypes, or theoretical explorations of potential applications. These case studies often highlight the potential future impact of quantum computing but do not reflect a currently deployable, production-ready state*". I did ask the same question to ChatGPT 4.0o and it said, well, yes, kind of, there are some in many fields, and if you want to know more, you can refer to McKinsey or BCG publications. Well well… !
- **What are the different techniques that can be used to read out the state of a superconducting qubit?** (5) It makes a good inventory of the various techniques that are currently used or tested.
- **Why do we need to cool down superconducting qubits at 15 mK**? (5) It gives a very good and complete response. Hint: it's not "just" because we are using a temperature enabling superconductivity.
- **Is it true that quantum computers using cold atoms, trapped ions and photons work at ambient temperature?** (5) The response is excellent, of course, based on all the engineering details provided in the book.
- **Can NV center qubits really operate at room temperature?** (5) Very good response, being "no", and with a lot of details.
- **What is the largest integer factorized by a quantum computer, and then, with Shor's integer factoring algorithm?** (3) It did not reconcile well the various records mentioned in the book although they are located in the same place. When pushing him with additional questions, it found the recent record obtained in China with a D-Wave and 20 and 50-bit numbers (not with Shor).
- **How is data loaded in a quantum computer?** (4). The response is correct but not very well documented.
- **When will a quantum computer with 4,000 logical qubits be available at at what price?** (4) It says it is a challenge to reach that goal, lists some of the things to do, but provides no specific numbers. Makes sense, since I am very cautious about it in the book.
- **What quantum algorithms are really bringing some exponential speedup?** (5) Response is concise and good, concluding: "*While a limited number of quantum algorithms, primarily those based on the QFT, offer the potential for exponential speedup, realising this potential in practice requires overcoming substantial technological challenges. The sources emphasize the importance of a holistic assessment that considers not only theoretical speedup but also practical limitations and advancements in classical computing. The field of quantum algorithms is constantly evolving, and future breakthroughs could lead to the discovery of even more powerful algorithms capable of achieving exponential speedups for a wider range of problems*".
- **How about quantum computers having an energetic advantage compared to classical computers?** (5) It answered pretty well the question, the conclusion being: "*The sources emphasize that while quantum computers offer a tantalising prospect of an energetic advantage, realising this potential will require overcoming significant hurdles. A comprehensive understanding of the energy demands of the entire*

quantum computing system, combined with ongoing efforts to improve the energy efficiency of classical components and quantum algorithms, will be crucial in determining whether quantum computers can deliver on their promise of energy-efficient computation“. This is the summary of the 24-pages section from my book on “Energetics” (in Quantum Computing Engineering, Volume 1).

- **Did you find any mistakes in the sources?** (1). It fared really badly here. It detected mistakes or inconsistencies that I am explaining, not creating. Its lack of reasoning probably explains it. It was the same when I asked it what was missing in the book. It mostly described things that… are in the book!
- **What being “quantum” means for a product or technology?** (5). It provided an excellent response.
- **Will quantum computers save the world (healthcare, climate change, …)**? (4). It provided a good response with many case studies but with some warning about their feasability.
- **What are the quantum computing resources estimates for key financial services applications?** (4) It makes a good assessment of the types of QPUs that be helpful and provides some resource estimates, particularly in the FTQC regime.

The detailed responses are in **this PDF**. At the end, I added a couple other questions that you can try on your own. It is now up to use your imagination… and information needs.

At this point, NotebookLM is “emulating” me to some extent but not entirely, hopefully. It is a useful tool to consolidate information coming from long documents or many documents. It helps uncover some myths and misconceptions, beyond the famous “quantum hype” that creates a reality distorsion field around the real and potential capabilities of quantum computers. But NotebookLM won’t really reason and it can make mistakes, lacking “judgment” in some situations. You need to know the rules of the game here and always fact-check what these tools are producing.

Extended version

The next step to this trial would be to create an extended version of this setup. How about uploading the 5600+ scientific references from the book’s bibliography? It is not possible with the end-user tools ala NotebookLM and ChatGPT GPTs. You need to create your custom solution and host it somewhere on the cloud and to use some open source LLM like Llama or Mistral. This is another, more complicated project.

Podcast generation

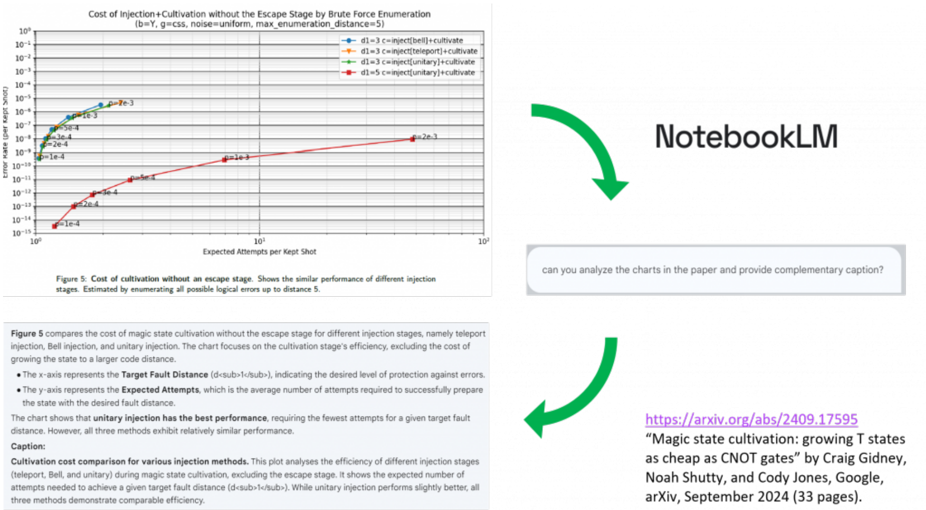
On top of that, NotebookLM can create a podcast with two (American) speakers summarizing key points from your documents in a very realistic way. I tried it with the book and the **result** is so-so regarding the content but spectacular with regards to its text-to-speech capability.

https://www.oezratty.net/Files/Publications/Understanding%20Quantum%20Technologies%20by%20NBLM.wav
You can even add a prompt when creating this audio. I used the following one: focus on the part on quantum computing hardware and software, be as technical as possible, describe the various challenges to build a scalable quantum computer, identify the challenges to do this for the main qubit modalities (superconducting, silicon, ions, cold atoms, photons), be specific on the key figures of merit (fidelities, number of qubits, etc.) and discuss about quantum error correction. It didn’t really follow all these instructions as you can hear.

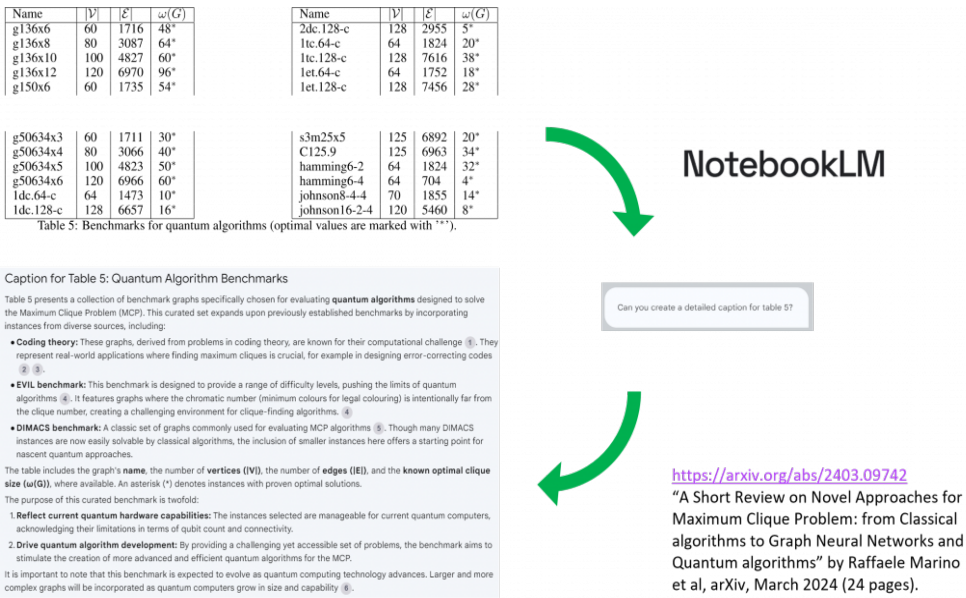
Figures captioning

I tried NotebookLM in other situations like with digging into the history of quantum foundations and comparing quantum physics interpretations (Qbism, RQM, MWI, CSM, etc) and it fared very well. To do that, I extracted from my book about 85 PDFs from the related chapter bibliography. I consolidated it into 4 PDF files using a PDF file merger tool, due to the NotebookLM file upload limit of 50 items.

I also discovered recently that it can create clear captions for cryptic charts and figures in scientific papers. I tried it on "**Magic state cultivation: growing T states as cheap as CNOT gates**" by Craig Gidney, Noah Shutty, and Cody Jones, Google, arXiv, September 2024 (33 pages). The result is very convincing. It seems that NBLM consolidates information from the paper text, its caption and even its own content. This is a game changer.



To be sure I was not just lucky once, I tried it several times, like with the following table that is in my book as an example of bad captioning. The source is **A Short Review on Novel Approaches for Maximum Clique Problem: from Classical algorithms to Graph Neural Networks and Quantum algorithms** by Raffaele Marino et al, arXiv, March 2024 (24 pages). And here is NotebookLM's response. It is really good, again!



Name	$\|\mathcal{V}\|$	$\|\mathcal{E}\|$	$\omega(G)$	Name	$\|\mathcal{V}\|$	$\|\mathcal{E}\|$	$\omega(G)$
g136x6	60	1716	48*	2dc.128-c	128	2955	5*
g136x8	80	3087	64*	1tc.64-c	64	1824	20*
g136x10	100	4827	60*	1tc.128-c	128	7616	38*
g136x12	120	6970	96*	1et.64-c	64	1752	18*
g150x6	60	1735	54*	1et.128-c	128	7456	28*
g50634x3	60	1711	30*	s3m25x5	125	6892	20*
g50634x4	80	3066	40*	C125.9	125	6963	34*
g50634x5	100	4823	50*	hamming6-2	64	1824	32*
g50634x6	120	6966	60*	hamming6-4	64	704	4*
1dc.64-c	64	1473	10*	johnson8-4-4	70	1855	14*
1dc.128-c	128	6657	16*	johnson16-2-4	120	5460	8*

Table 5: Benchmarks for quantum algorithms (optimal values are marked with '*').

So not only these tools can summarize long arXiv and other scientific papers, but it can help you decipher their cryptic content. It is a boon for all scientists!

More on this at Quantum+AI

That was kind of a teaser for my presentation on October 29th, 2024, in New York City at the conference **Quantum+AI**. I will present some insights on how LLM-based tools are already changing the life of scientists and others in the quantum ecosystem and where we may be going. My presentation will be a follow-up to the paper I published on this topic in January 2024: **How can AI, LLMs and quantum science empower each other?**.

See you there, maybe!