How can AI, LLMs and quantum science empower each other?

The field of generative artificial intelligence is currently dominated by the large language model chatbots like ChatGPT. This platform was first introduced to Internet users in late 2022 with its 3.5 version. Launched in spring 2023, the 4.0 version has since then attracted hundred million users.

My initial engagement with ChatGPT 3.5 yielded underwhelming results. However, an in-depth exploration of ChatGPT 4.0 and Google's Bard revealed a broader spectrum of applications in quantum science than I thought were possible. These AI tools were useful in very various contexts ranging from explaining quantum physics concepts, facilitating software coding tasks to preparing various content for my own teachings.

I thus wondered what the present and future implications of LLMs in quantum science and technologies were for academics, researchers, industry professionals and even policy makers. Despite all their well-documented flaws, these tools are bound to grow and create significant positive disruptions for all these roles.

I ended up publishing How AI, LLMs and quantum science can empower each other? (24 pages + 70 pages of supplemental materials including an index and extensive bibliography). This post presents the paper structure and thesis. It also adds some meta-observations on how LLMs may change the quantum science and technology landscape.

Paper outline

The paper looks at the intersection of machine learning and quantum science, with a focus on large language models chatbots. It investigates the various ways quantum scientists can benefit from these new tools. It describes the existing as well as future use cases for LLMs with the observation that they have a significant potential to impact research methodologies and pedagogical approaches in quantum science. It also explores the various other ways machine learning techniques and quantum technologies can help each other.

The intended audience of this paper are students and professionals in the broad quantum science and technology ecosystem who have a limited knowledge and understanding of machine learning and particularly, large language models chatbots. A preliminary understanding of neural network basics is however a plus here.
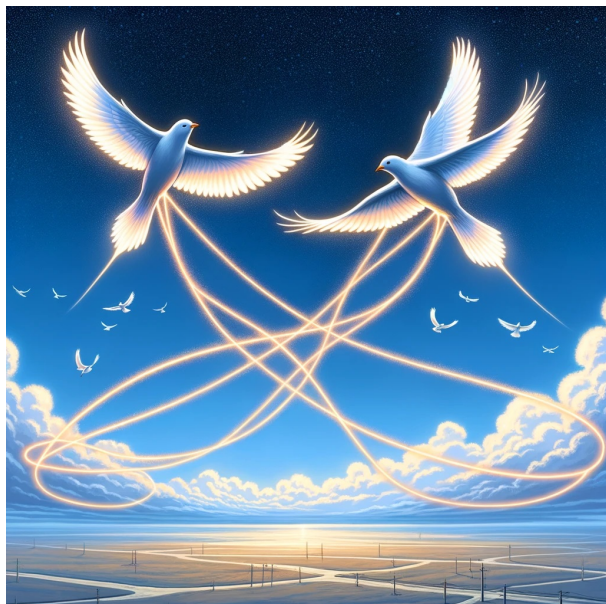
The first part describes how LLMs fit into the generative AI landscape, the key features of LLM-based chatbots, their size and their training data sources, their figures of merit and how to make best use of them with prompt engineering. It assesses the reasoning capabilities of these chatbots and shows how it will grow with showcasing emergent capabilities, on the road to so called "artificial general intelligence". This part describes how to create domain specific LLM chatbots using model fine tuning and adding retrieval-augmented generation (RAG) to query a library of documents (the chart below explains how a RAG is built and queried). It also suggests the launch of a quantum science domain specific LLM based chatbot using open source models.
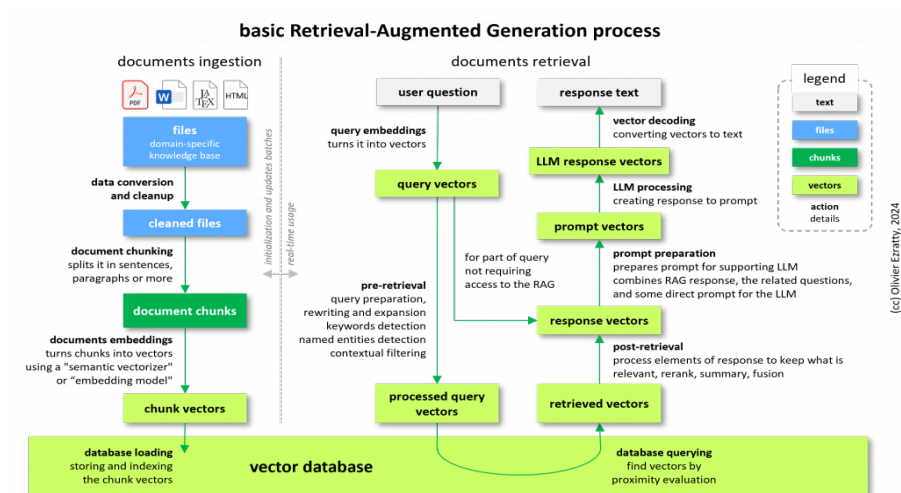
Then, it showcases how LLMs can be used in the context of quantum science by researchers, teachers, students, and industry professionals. I cover the process of learning and studying quantum science and technologies, reviewing, and writing scientific and other papers, doing classical and quantum software development, conducting research ideation and collaborative research, with existing and potential future capabilities.

On top of LLM use cases, the paper also inventories the various applications of other types of machine learning tools in quantum science with research and technology developments in quantum physics and hardware, quantum software and tools and other domains like quantum sensing and quantum communications and cryptography.
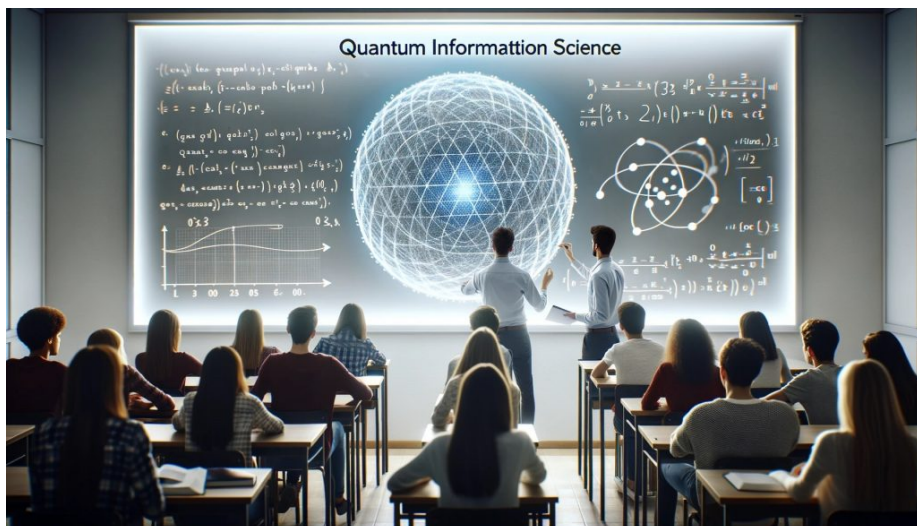
It quickly covers the potential use cases of quantum computing in machine learning applications including LLMs with a cautionary note. We are far from letting AI really benefit from quantum computing. Quantum computing will probably not help in the scaling of LLMs. However, quantum inspired algorithms will help classical machine learning.

The paper supplemental materials contain a wealth of prompts and results covering various fields of quantum science, including learning, summarizing the state of the art, listing figures of merits for technologies like single photon sources and detectors, circulators, traveling wave parametric amplifiers, analyzing a paper, producing a position paper, and creating a quiz. If you have not toyed yet with ChatGPT 4.0 in the field of quantum science, you will probably be astonished by its capabilities. Given ChatGPT 5.0 is around the corner and may be released in a couple months!

At last, the paper contains extensive bibliography for readers willing to dig into the way LLM and RAG operate.

Meta-observations

Here are some insights from this multi-month journey discovering the potential impact of LLMs and generative AI on quantum science.

Capabilities. LLM based chatbot are amazing in their capability's breadth. They can consolidate a wealth of information in a structured way like the figures of merit of various technologies. I tested it with single photon sources and detectors, circulators and the TWPA used in superconducting qubit readout. They can create a quiz and help teachers save time. They can create memos on various topics, including proposing a National Quantum Initiative, and then translate it in multiple languages. They can provide feedback on your own texts and help you improve it. You can even ask an LLM to improve a prompt so that it better answers your questions! And this is bound to improve. These tools are making significant progress year after year in a pace that looks like being much faster than, say, quantum computing. We are not far from obtaining some forms of AGI (artificial general intelligence) although the notion is not well defined.

Hybridization. LLMs can showcase apparent forms of reasoning and sometimes explain their results. They don't support real dynamic reasoning but embed rudimentary forms of static reasoning thanks to model fine tuning, instruction tuning, reinforcement learning with human feedback and with "chain of thoughts" and other specific agents that structure the way LLMs respond to your questions. Some forms of reasoning are generated that come from a probabilistic exploitation of the texts used in training the models but also from symbolic AI tools driving the LLM. LLM-based chatbots are already highly hybridized forms of AI. On top of that, prompt engineering techniques is becoming an art, which helps obtain better responses.

Size. The size of the latest generic large language models, their training data and time, and the involved

hardware resources is gigormous. It is respectively expressed in trillions of parameters, terabytes of data, months, and tens of thousands of GPGPU in the likes of those from Nvidia. We can wonder about their related energy consumption, which is not well documented.

Opportunity: there is no domain-specific chatbot for quantum science and it would make a lot of sense to create one, at a reasonable cost. It could be hosted on a standalone web site using its own open-source LLM or be a plugin for ChatGPT 4.0. For various security in intellectual property reasons, the first solution is preferable to the latter. But creating such chatbots will require the licensing of a wealth of content coming from publishers like Springer (Nature, EPJ) and the American Physical Society (PRX Quantum, PRL, PRA, PRB, …).
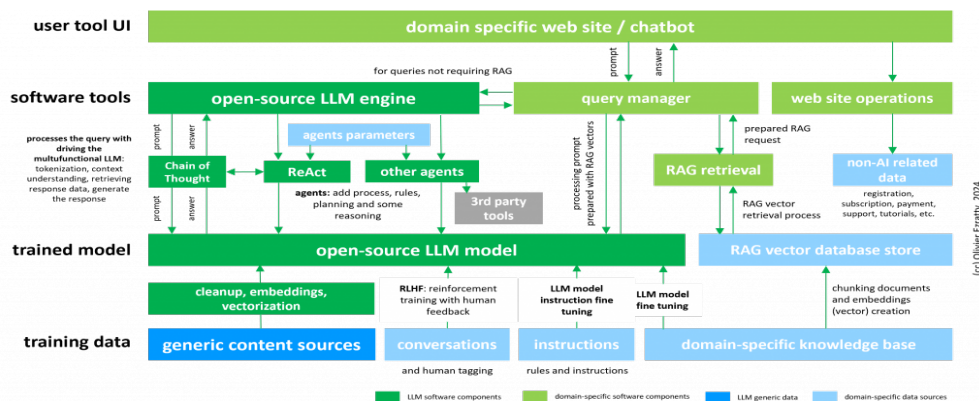
Issues. Yes, LLMs invent stuff that is not true or does not exist ("hallucinations") and do not produce explainable results. But these problems can be mitigated when addressed properly, particularly with the many tools surrounding LLMs like "chains of thoughts" and model fine-tuning. RAG knowledge base, although also imperfect, bring the capability to source the elements of chatbot responses. Other issues pertain to overusing LLMs in daily work. At some point, you can be overwhelmed by their production and over rely on it, having less time to brainstorm, really learn and memorize and be creative. We will have to learn to make the best use of these tools.

Charts. On top of various forms of reasoning, what LLM and generative AI lack is the capacity to interpret charts in scientific papers and to safely consolidate unstructured data. However, these problems are being addressed by several ongoing research projects. Other needs will show up by experimenting with them. I list many of them in the paper like the ability to analyze a scientific paper with their metadata (authors track-record), weighing the proposed advances vs the state of the art and quickly identifying any missing figures of merit.

Analogies. We can create some analogies comparing NISQ vs FTQC and current AI vs AGI. Current LLM-based chatbots use hybridization techniques with other machine learning tools and symbolic AI to embed rules and static reasoning, a bit like NISQ is based on hybrid variational algorithms. Other proposed avenues use different methods like what Yann LeCun is working on, akin to FTQC in quantum computing but with a longer term perspective. It is a battle between an unperfect present and a far-fetch better future.

Time scales. AI and quantum computing seem to operate on both different and similar time scales. It looks like AGI is around the corner in a few years. And one can wonder about the capabilities of ChatGPT 5.0 that will show up this very year. Classical computing power is still growing, particularly thanks to GPGPUs (general purpose GPUs ala Nvidia) and their related distributed architecture. Still, the AI term was coined by John McCarthy, Nataniel Rochester, Marvin Minsky and Claude Shannon in a **memo** written in August 1955. We are approaching the 70th year milestone and we are not there yet with "artificial general intelligence" capable of generic reasoning, but the achievements of AIs are still significant and already delivering value. The progress in the next few years may be enormous. The equivalent to the Dartmouth 1955 memo in quantum computing are papers published by Paul Benioff, Yuri Manin and Richard Feynman between 1979 and 1981. Thus far, "prototypes" of quantum computers are available who are short from delivering significant value compared to their classical peers. But progress is looming and at least two platforms may deliver about 100 logical qubits in less than a decade (cat-qubits from Alice&Bob, and neutral atoms from QuEra).

Platforms. Despite quantum computers have not really reached the utility threshold, quantum software and cloud platforms are already competing for the hearts and minds of quantum developers (IBM Quantum Experience, AWS Braket, Microsoft Azure, …). The LLM chatbot landscape is also becoming a platform battle

with the ChatGPT Store and its plugins and looming competition from Google and Meta. I wonder how these advanced AI systems will consolidate access to scientific information, like Google did with search engines. ChatGPT is already positioned as a content access platform with its Store. Domain-specific information providers will be faced with a tough choice: either create a plugin for the dominant AI/AGI platforms like ChatGPT and Google Bard or create their own website, without benefiting from the genericity of these platforms. Publishing groups like Springer (Nature, EPJ, …) and the American Physical Society (PRX Quantum, PRL, PRA, PRB, …), on top of arXiv (Cornell University) will also have to decide how and where they license access to their content or if they want to create their own LLM chatbot, given most physicists will prefer an AI service that is trained on all scientific sources and not just one.

Winters. AI already went through two winters in the early 1970s and the mid-1990s. Some are expecting quantum computing to live its first quantum winter soon. The comparison stops there at this point since this winter is not there yet. Below is a first-degree representation of the quantum winter by DALL-E.

Fun. ChatGPT 4.0 can show an excellent sense of humour thanks to its ability to create fancy analogies. When I asked it to explain Shor's algorithm ala Donald Trump, it generated "*Now, let me tell you, nobody knows algorithms better than me. And this Shor, he's a smart guy, really smart. He figured out how to use these tiny, tiny things called qubits. Qubits, unlike regular bits, they can be in* **multiple states at once** – *it's like,* **they can be in New York and Florida at the same time**. *Can you believe it?*". It made my day (the full response is in the paper). Never thought of it! For the boomers who can read French, I also asked an explanation of quantum physics ala Raymond Devos. The **result** is quite funny as well.

Research practices. I suspect that today and future LLM chatbots will soon be used daily by quantum science researchers. It will progressively change research practices in many dimensions: preparing research work, investigating the state of the art, finding and comparing technology vendors, preparing experiments, developing code for experiments, and data analysis, preparing reports, papers, and thesis, creating charts, preparing presentations and so on. Some learning in on the way!

What are your dreams and fears there?

Cet article a été publié le 26 janvier 2024 et édité en PDF le 15 mars 2024.