'ChatGPT' for Biology? A Dictatorship of Engineers

Analysis by Tessa Lena | September 01, 2023

STORY AT-A-GLANCE

> Some of the largest tech companies in the world are competing to develop a ChatGPT-like engine for biology, based on large language models (LLMs)

> One of the areas of using LLMs in biology is synthesis of completely new proteins

> Their long-term ambition is to generate not just new proteins but also cells, organs, and whole organisms

> A 2016 World Science Festival panel on human engineering is a great illustration of the bio engineers' mindset

> The participants of that panel, all renowned geneticists and bioethicists, discussed their ambition to manufacture people from scratch and to genetically engineer us to be smaller in size and allergic to meat, among other things

> An adjacent "hot" topic of human engineering is "memory modification." What can possibly go wrong …

66 **Hopefully, [the ultimate search engine Google was building] would never have a bug like HAL did where he killed the occupants of the spaceship. ~ Sergey Brin (NewsHour with Jim Lehrer, November 2002)**[1] 99

As if the ChatGPT craze weren't bad enough, the $$$$$ winds are blowing in the direction of trying to build a similar engine for biology — and on a large scale. Highly perched individuals with a technocratic vision are betting on AI that would surveil every nook and cranny in the body and then generate … well, something useful to them, they hope. On my end, I am afraid to think what kind of Frankenstein such AI can generate.

The idea, as usual, is to feed the AI as much data as possible (biological data, in this case), and hope that it will "understand" the "language of biology" — properties of different elements and the connections between them — and then "intelligently" build wondrous biological structures from scratch. Mommy, no.

A Few Thoughts About ChatGPT

Is generative AI's current ability to mimic natural language and spit out perfect English sentences on demand impressive? Yes, it's a cute inanimate parrot and information retriever, that generative AI.

But is it a reliable source of information? Nope! It makes things up unpredictably. It's a machine. An automaton. A Lego brick assembler. It does not think. It doesn't feel. It doesn't "know" anything. It doesn't "know" the meaning of the ones and zeros that it spits out.

It is prone to the so called "hallucinations," where the robot produces text that looks plausible — but the "facts" are simply made up. And I am not talking about intentional "lying" due to being programmed to propagandize — it does that, too — what I am talking about here is "lying" for no reason, with no benefit to anyone, just generating smooth-sounding "facts" that are made up and packing them alongside the statements that are factually correct.

Now let's imagine how it would work in biology. I think they've made horror films about this kind of thing, no?

Large Language Models for Biology

In July of this year, Forbes magazine published an [article](#) that provides some insight into the trend:

> "As DeepMind CEO/cofounder Demis Hassabis put it: "At its most fundamental level, I think biology can be thought of as an information processing system, albeit an extraordinarily complex and dynamic one. Just as mathematics turned out to be the right description language for physics, biology may turn out to be the perfect type of regime for the application of AI."

> Large language models are at their most powerful when they can feast on vast volumes of signal-rich data, inferring latent patterns and deep structure that go well beyond the capacity of any human to absorb. They can then use this intricate understanding of the subject matter to generate novel, breathtakingly sophisticated output.

> By ingesting all of the text on the internet, for instance, tools like ChatGPT have learned to converse with thoughtfulness and nuance on any imaginable topic. By ingesting billions of images, text-to-image models like Midjourney have learned to produce creative original imagery on demand.

> Pointing large language models at biological data — enabling them to learn the language of life — will unlock possibilities that will make natural language and images seem almost trivial by comparison … In the near term, the most compelling opportunity to apply large language models in the life sciences is to design novel proteins."

AI for Proteins

In late 2020, Alphabet's AI system called AlphaFold produced an alleged "[solution](#) to the protein folding problem." AlphaFold is [said](#) to have "correctly predicted proteins' three-dimensional shapes to within the width of about one atom, far outperforming any other method that humans had ever devised."

AlphaFold was not based on large language models but on an "older bioinformatics construct called multiple sequence alignment (MSA), in which a protein's sequence is compared to evolutionarily similar proteins in order to deduce its structure."

Recently, scientist started to explore using LLMs to predict protein structures. According to Forbes, "protein language models (LLMs trained on protein sequences) have demonstrated an astonishing ability to **intuit [emphasis mine]** the complex patterns and interrelationships between protein sequence, structure and function: say, how changing certain amino acids in certain parts of a protein's sequence will affect the shape that the protein folds into …

The idea of a protein language model dates back to the 2019 UniRep work out of George Church's lab at Harvard." Let's look at George Church and his work.

A Remarkable 2016 World Science Festival Panel

Remember the recently resurfaced short video clip from 2016 about "editing" humans to be intolerant to meat? The panel was from the 2016 World Science Festival. It featured a couple of renowned geneticists and bioethicists (George Church, Drew Endy, Gregory E. Kaebnick, S. Matthew Liao) and Amy Harmon, a journalist from the New York Times. (I wrote about it in detail here.)

The panelists talked about "manufacturing human DNA and whole new orphans people from scratch, about germline editing (introducing heritable genetic changes, which, they say, is already being done), about genetically editing people to be more compliant with the current thing empathetic, or to be allergic to meat and smaller in size 'for the planet,' etc."

George Church, now, is a very famous geneticist who has worked on age reversal, barcoding mammalian cells (see his work on barcoding the whole mouse), recreating the woolly mammoth, and "printing" DNA (with an implication of potentially "manufacturing" human beings) from scratch.

He is "Professor of Genetics at Harvard Medical School and Director of PersonalGenomes.org, which provides the world's only open-access information on human Genomic, Environmental & Trait data (GET). His 1984 Harvard PhD included the first methods for direct genome sequencing, molecular multiplexing & barcoding.

These led to the first genome sequence (pathogen, Helicobacter pylori) in 1994. His innovations have contributed to nearly all "next generation" DNA sequencing methods and companies (CGI-BGI, Life, Illumina, Nanopore).

This plus his lab's work on chip-DNA-synthesis, gene editing and stem cell engineering resulted in founding additional application-based companies spanning fields of medical diagnostics (Knome/PierianDx, Alacris, Nebula, Veritas) & synthetic biology / therapeutics (AbVitro/Juno, Gen9/enEvolv/Zymergen/Warpdrive/Gingko, Editas, Egenesis).

He has also pioneered new privacy, biosafety, ELSI, environmental & biosecurity policies. He was director of an IARPA BRAIN Project and 3 NIH Centers for Excellence in Genomic Science (2004-2020). His honors include election to NAS & NAE & Franklin Bower Laureate for Achievement in Science. He has coauthored 650 papers, 156 patent publications & a book (Regenesis)."

George Church has been working with DAPRA on various projects. For example, he has been a part of Safe Genes initiative, seeking to "develop systems to safeguard genomes by detecting, preventing, and ultimately reversing mutations that may arise from exposure to radiation."

That work was said to "involve creation of novel computational and molecular tools to enable the development of precise editors that can distinguish between highly similar genetic sequences. **The team also plans to screen the effectiveness of natural and synthetic drugs to inhibit gene editing activity [emphasis mine].**" Additionally, he was allegedly involved in DARPA's BRAIN Initiative.

As a side note, in 2019, he apologized for working with Epstein after the latter pleaded guilty, citing "nerd tunnel vision." Now, before we look at another notable World Science

Festival panelist, S. Mathew Liao, let's go back to large language models in biology and see what we got there.

Inventing New Proteins

"All the proteins that exist in the world today represent but an infinitesimally tiny fraction of all the proteins that could theoretically exist. Herein lies the opportunity," says Forbes.

I have one word for them: plastic. It was a wonderful invention at one time, and it sure changed our lives and added a lot of convenience to it — but then it turned out that it was not so great for our health, and now plastic can be found everywhere.

It can be found in the human brain, in placenta, and deep in the ocean — not to mention mountains of it at landfills. And that's just good ol' plastic, something that was invented during the "ancient times" of technological development, by the standards of today. But back to Forbes:

> "The total set of proteins that exist in the human body — the so-called 'human proteome' — is estimated to number somewhere between 80,000 and 400,000 proteins. Meanwhile, the number of proteins that could theoretically exist is in the neighborhood of $10^{1,300}$ — an unfathomably large number, many times greater than the number of atoms in the universe …

> An opportunity exists for us to improve upon nature. After all, as powerful of a force as it is, evolution by natural selection is not all-seeing; it does not plan ahead; it does not reason or optimize in top-down fashion. It unfolds randomly and opportunistically, propagating combinations that happen to work …

> Using AI, we can for the first time systematically and comprehensively explore the vast uncharted realms of protein space in order to design proteins unlike anything that has ever existed in nature, purpose-built for our medical and commercial needs."

What arrogance, dear God, just stop! The marketing brochure talks about curing diseases and "creating new classes of proteins with transformative applications in agriculture, industrials, materials science, environmental remediation and beyond." Methinks, it is going to be "transformative" alright but in what way, and for whose benefit? Not ours!

> "The first work to use transformer-based LLMs to design de novo proteins was ProGen, published by Salesforce Research in 2020. The original ProGen model was 1.2 billion parameters …
>
> Another intriguing early-stage startup applying LLMs to design novel protein therapeutics is Nabla Bio. Spun out of George Church's lab at Harvard and led by the team behind UniRep, Nabla is focused specifically on antibodies.
>
> Given that 60% of all protein therapeutics today are antibodies and that the two highest-selling drugs in the world are antibody therapeutics, it is hardly a surprising choice Nabla has decided not to develop its own therapeutics but rather to offer its cutting-edge technology to biopharma partners as a tool to help them develop their own drugs."

"The Road Ahead"

Still Forbes:

> "In her acceptance speech for the 2018 Nobel Prize in Chemistry, Frances Arnold said: 'Today we can for all practical purposes read, write, and edit any sequence of DNA, but we cannot compose it. The code of life is a symphony, guiding intricate and beautiful parts performed by an untold number of players and instruments.
>
> Maybe we can cut and paste pieces from nature's compositions, but we do not know how to write the bars for a single enzymic passage.'

As recently as five years ago, this was true. But AI may give us the ability, for the first time in the history of life, to actually compose entirely new proteins (and their associated genetic code) from scratch, purpose-built for our needs. It is an awe-inspiring possibility."

Mommy, no!!

"Yet over the long run, few market applications of AI hold greater promise … Language models can be used to generate other classes of biomolecules, notably nucleic acids. A buzzy startup named Inceptive, for example, is applying LLMs to generate novel RNA therapeutics.

Other groups have even broader aspirations, aiming to build generalized "foundation models for biology" that can fuse diverse data types spanning genomics, protein sequences, cellular structures, epigenetic states, cell images, mass spectrometry, spatial transcriptomics and beyond.

The ultimate goal is to move beyond modeling an individual molecule like a protein to modeling proteins' interactions with other molecules, then to modeling whole cells, then tissues, then organs — and eventually entire organisms. [Emphasis mine.]"

The crazies are truly running the asylum at the moment. How many times do the arrogant scientists have to hurt the world in order to wake up? What will it take for them to wake up? When they personally grow a third leg?!

S. Matthew Liao, the Bioethicist

Now let's talk about the ambitions to engineer people on order to make them smaller and allergic to meat — and to erase undesirable memories. Meet the renowned bioethicist, a strange person, S. Matthew Liao.

S. Matthew Liao "holds the Arthur Zitrin Chair in Bioethics and is the Director for The Center for Bioethics at New York University. From 2006 to 2009, he was the Deputy

Director and James Martin Senior Research Fellow in the Program on the Ethics of the New Biosciences in the Faculty of Philosophy at Oxford University.

He was the Harold T. Shapiro Research Fellow in the University Center for Human Values at Princeton University in 2003–2004, and a Greenwall Research Fellow at Johns Hopkins University and a Visiting Researcher at the Kennedy Institute of Ethics at Georgetown University from 2004–2006. In May 2007, he founded Ethics Etc, a group blog for discussing contemporary philosophical issues in ethics and related areas."

His scholarly works make me wonder about his life. I certainly wish him well but the topics make me wonder. Here's one, "The Right to Be Loved":

> "S. Matthew Liao argues here that children have a right to be loved … His proposal is that all human beings have rights to the fundamental conditions for pursuing a good life; therefore, as human beings, children have human rights to the fundamental conditions for pursuing a good life. Since being loved is one of those fundamental conditions, children thus have a right to be loved."

Here's another: "The normativity of memory modification"

> "We first point out that those developing desirable memory modifying technologies should keep in mind certain technical and user-limitation issues. We next discuss certain normative issues that the use of these technologies can raise such as truthfulness, appropriate moral reaction, self-knowledge, agency, and moral obligations.

> Finally, we propose that as long as individuals using these technologies do not harm others and themselves in certain ways, and as long as there is no prima facie duty to retain particular memories, it is up to individuals to determine the permissibility of particular uses of these technologies."

Speaking of, here is his talk about memory modification:

And just as I was wrapping this article up, I got a newsletter from Open to Debate, titled, "Should we erase bad memories?" featuring Nita Farahany, "agenda contributor" at the

WEF. (My answer to that question, by the way, is a resounding NO.)

Conclusion

I will end this story with a short quote from my recent article:

> *"They are trying. They are likely going to create a lot of unnecessary, stupid, cruel suffering. But in the end, they are not even going to end up with "I am afraid I can't do it, Dave." They are going to end up with this."*

About the Author

To find more of Tessa Lena's work, be sure to check out her bio, **Tessa Fights Robots**.

Sources and References

- [1] americanarchive.org