Post #2 on AI. Question regarding COVID-19 vaccines. Date 2024-8-25

Carlos Alegria

In my previous post on critical thinking on AI systems, I used the example of our intensive 3-year research project on mapping out the human and economic cost of the COVID-19 pandemic, under our "Humanity Projects"[1]. Our research led us to an in-depth knowledge of the pandemic, the adverse effects of the novel COVID-19 vaccines on human health and its economic impact. With a deep understanding of the data, we tested OpenAI's ChatGPT4 with the question: **Are COVID-19 vaccines safe and effective?** The answer and its critical analysis can be read in my previous post.

In this post I extend the analysis of the possible dangers of AI-based systems for decision making processes.

1. How are AI-based systems biased?

I ended my previous post by explaining how AI systems can be biased depending on the input data that is fed into the learning algorithms. In essence, "garbage in, garbage out". In general, one can achieve a bias by **volume** of certain selected content, or by ranking the content into **quality** metrics, for instance by ranking authoritative sources more highly than "untrusted" sources.

With the generalised usage of more sophisticated AI chat boxes to produce content (academic, journalistic, fiction, etc) in the format of books and articles, AI is rapidly replacing traditional human jobs and AI productivity in terms of volume of content is unchallenged by humans. This has implications that new generation AI systems will be feeding into their algorithms content that is mostly generated by previous AI versions.

A case in hand comes from a recent Lex Friedman interview[2] on the 2nd of August of 2024, where Elon Musk mentioned that one of the biggest problems they are currently facing is how to filter the data that is "fed" into the AI model. Elon Musk's AI company[3] (xAI) is a strong competitor to OpenAI and is currently working on a chatbox (that is named Grok) which is comparable to ChatGPT4, that I used in my previous article to ask a question regarding the COVID-19 vaccines.

In the interview, Elon Musk mentioned that his AI team had to feed the source data into a purposefully built AI-system that is designed to identify AI-generated content and filter it out from the data that is fed into their most sophisticated Grok 2 chat box. Inclusively, in the interview, Elon Musk mentioned that source data from 2023 onwards was eliminated due to being compromised (which will lead to obvious biases).

2. Can AI ever be un-biased?

By now, it is clear that achieving an un-biased AI is akin to mission impossible. The process of attempting to remove systematic bias in the source data, will lead to a biased system as it will not represent the available data sources relative to either volume of information or "quality" of information. Ultimately, one would have to build thousand of AIs that target specific contexts and that are subject to a prior research process (human intelligence) to understand the possible factors that lead to different biases.

After this is achieved, the AI needs constant monitoring from any slippage towards biases, that can creep into the source data. One such example of bias creep into the source data can occur by the control of the language that is allowed to be used within the data sources. For example, censorship by means of qualifying certain type of language as "dangerous" or "offensive" can lead to language adaptations by individuals that would alter the form of the language, but not its intent.

---

[1] https://phinancetechnologies.com/HumanityProjects/Projects.htm

[2] The interview: https://www.youtube.com/watch?v=Kbk9BiPhm7o
The interview transcript: https://lexfridman.com/elon-musk-and-neuralink-team-transcript

[3] https://x.ai/

In the case of the question I posed "are Covid-19 vaccines safe and effective?", academic sources are under extreme pressure relative to the language used within the published papers. Comments that suggest that there is a link between Covid-19 vaccines and adverse effects, must be followed by statements such as "vaccines are an essential tool to manage pandemics, and their benefits outweigh any possible adverse effects". The control of the language that is permitted for publications to be accepted by peer review or avoid fact-checking agencies will lead to bias creep of the source data.

How can one obtain un-biased AI systems?

My answer is: Very carefully, with great effort and constant monitoring.

To understand the importance in building un-biased AI systems and how they could be achieved, let's look at the extremes:

3. When the "consensus" is an extreme view.

It is "well established" that vaccines have helped eradicate numerous diseases over time, improving human health. The push for the reach of global vaccination that started in the 1970s and increased exponentially until current times, has been unquestioned, with the exception of a minority of scientists and other sceptics that are labelled extremists, anti-vaxxers, anti-science and other pejorative terms.

Over time, in order to protect populations' health from the undue influence of extreme viewpoints on the usage of vaccines, policy makers made the target of policy to combat vaccine hesitancy. With the Covid-19 pandemic, authorities went a step further and equated the spreading of vaccine mis-information and dis-information as information terrorism, which falls under the same legal umbrella as actual terrorism.

With the Covid-19 vaccine rollout, governments were advised by psychological operations units, that would target "vaccine hesitancy" by nudging individuals' behaviours towards vaccine uptake (steps 3 and 4 in Figure 1). The psychological operations work with a combination of sticks and carrots, exploiting social interaction patterns in order to "convince" people to take the vaccines "voluntarily".

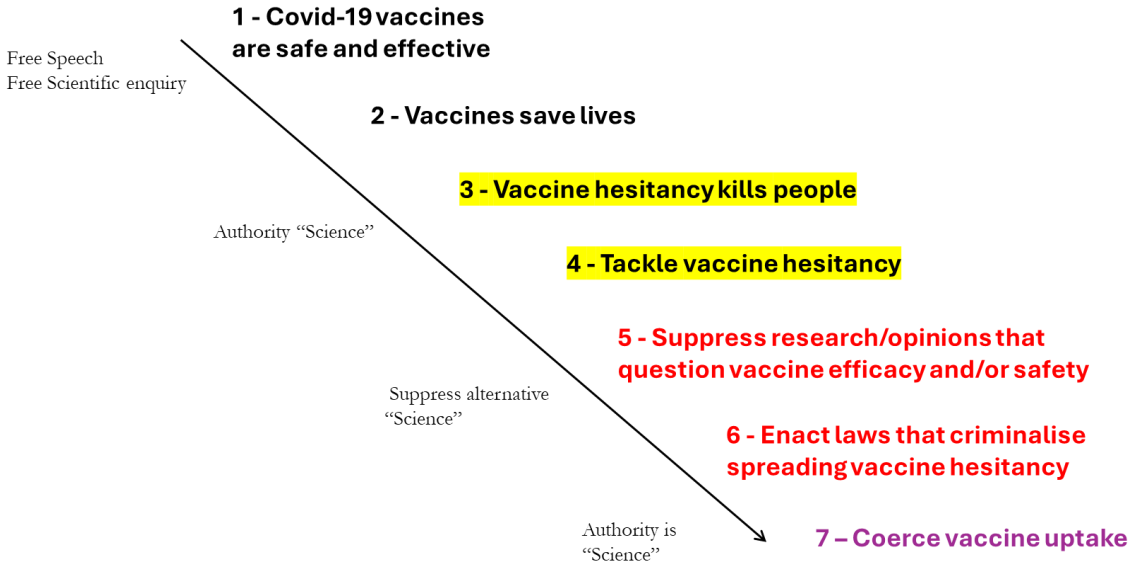

Figure 1 - Logical policy-making steps

The whole sequence of events starts at the assumption that vaccines are "safe and effective". The logical sequence of steps described in Figure 1 illustrate the current path taken by health authorities and governments worldwide. Each logical step build on each other, which in a very logical and rational way leads to a totalitarian implementation of a policy that is "for the greater good".

To any actual scientist this is a clear red flag as "science" is based upon asking questions, examining reality, by using the scientific method. Scientific truths are constantly being disproved, updated and refined as new evidence emerges.

How does this example relate to biases in AI-systems?

The policy actions described above lead to the suppression of alternative research and the promotion of information that confirms the consensus, which biases the information landscape. These biases are picked up by AI systems and then reinforced by the AI regurgitating back the consensus view. As I've stated in my previous article, "biased sources in -> biased response out".

The consensus view becomes extreme

The first two logical statements in the diagram above are not extreme. The extreme view comes from walking down the logical steps towards an absolute view that cannot be challenged. The policy statements from 3 to 6 are the steps that lead to the implementation and enforcement of what **could** become an extreme view. In the case of the Covid-19 vaccines research has now shown that the underlying assumption is wrong and consequently its implementation using logical steps 3 to 6 has indeed led to a dystopian scenario where there is in place an almost totalitarian implementation of a wrong assumption.

The previous logical process can be applied to many of today's societal issues. Controversial issues such as climate change, food preferences, and many others will be subject to the same logical steps described above and will necessarily lead to an extreme view, irrespective of the underlying assumption being true or false. All that is needed to verge towards an extreme consensus view is the implementation of the logical steps from 3 onwards.

4. When moderation is found in the extremes

Moderation comes from incentivising the analysis of the main assumption in the logical diagram shown in Figure 1, by independent researchers, who do not have conflicts of interest, and are provided the data resources in order to perform an independent validation of the assumptions.

These independent views should suffer no form of coercion or censorship that would put barriers to their work. This alternative research would then enter the information landscape and be subject to scrutiny by other individuals. Policy makers would then have a richer information landscape to make the recommendations that are deemed as an adequate policy response and individuals could make their decisions accordingly. The process is the basis of informed consent[4] for any policy recommendation.

Unfortunately, the moderation I describe above is currently considered an extreme view when it comes to questioning vaccines in general, and Covid-19 vaccines in particular, using the argumentation that these independent sources could lead to "vaccine hesitancy".

5. Wisdom of crowds, madness of crowds.

AI-systems such as chatGPT4 are seen by most enthusiasts to represent the wisdom of the marketplace of ideas. This is an extension of the idea that individuals interacting in a free market, by acting on their own opinions and ideas, contribute to a decentralised system that provides a wisdom that goes beyond that of any individual expert. The wisdom of crowds[5] was popularised in an exciting book by James Surowiecki who provides many examples of how the "collective" opinion surpasses the opinion of an expert.

As a mostly free-market advocate I agree with this approach for individuals to "find" solutions for complex problems, through the creation of an invisible collective "brain". However, we must not forget that the underlying assumption for such a system to function, the individuals that compose the system should be independent (free thinkers) and not restrained, censored or coerced. In technical terms, if each individual is

[4] https://en.wikipedia.org/wiki/Nuremberg_Code
[5] https://www.amazon.com/Wisdom-Crowds-James-Surowiecki/dp/0385721706

considered as a source of information regarding a given subject, by aggregating uncorrelated (independent) sources of information the system should result in an improved signal to noise ratio than any individual expert.

However, as individuals in a social environment we are clearly subject to all sorts of biases, patterns of behaviour and social pressures. These biases can and are exploited to nudge[6] individuals preferences in different desired directions, which in the logical diagram shown in Figure 1 are implicit within the logical steps from 3 to 6. These social behaviour management steps[7] are designed to create a feedback loop that reinforces the "consensus" narrative of the day by reducing the pool of available ideas (in volume and quality). In effect, these steps reduce the independence of thought between the various individuals that participate in the "free" market of ideas.

When such conditions occur, we can easily be led into extreme outcomes such as the "Vaccines are safe and effective" that I described above. These mass formations[8] via a manufactured "consensus" lead to crowd behaviours such as those described by Charles Mackay" in "The Madness of Crowds"[9] (published in 1841). This is clearly not a new phenomenon, and AI is just another technology that can be used to achieve such outcomes.

The way to protect from the pitfalls of hive mind and to achieve continued progress in human development, free societies should focus on being on constant alert of the steps in the logical diagram above, that act as feedback mechanisms that lead to an entrenched "consensus", so that there is a free flow of ideas that can challenge (or re-enforce) the underlying assumptions. This is the core idea behind "freedom of speech" and also the scientific method of enquiry.

6. Applications for AI systems.

One of the many "promises" of AI systems are using them to uncover emerging trends in different phenomena and find relationships that are not obvious to the human mind. In essence, some AI enthusiasts believe that AI systems could replace human creativity.

Personally, I doubt that AI systems will replace human research and creativity that is performed through processes using deduction, induction and abduction. Perhaps, at best, AI systems could be employed to point to emerging trends within existing human-based research publications. But for this to be done effectively one would need a rigorous analysis of biases in AI systems, and by consequence the biases in the source data.

The example I put forward above is a clear example where "factual" information regarding vaccine harms would be hidden within the vast volume of data sources claiming that vaccines are "safe and effective", and in order to find an answer that is closer to actual reality (e.g. true information), one would have to feed the AI systems with data sources that lay in the extremes. By filtering out the vast volume of consensus information (noise), one could then assess the underlying patterns (possible signals).

The reasons why private corporations would want to build such un-biased AI systems would be to have a competitive advantage in understanding reality and then based on that information perform better decision making. The fact that AI systems do not find the "true" answer to "Are Covid-19 vaccines safe and effective?" can lead to decision making that proves to be quite costly.

Corporations being subject to highly competitive market forces which have "skin in the game" should be highly concerned when using AI-systems for decision making processes. Ultimately, progress into breakthrough ideas, methods and technologies will arise from human-based research and the corporations that do not push these boundaries will be akin to automated large scale marketing operations.

[6] https://www.amazon.com/Nudge-Improving-Decisions-Health-Happiness/dp/014311526X
[7] These marketing methods can be employed by governments or corporations.
[8] https://www.amazon.com/Psychology-Totalitarianism-Mattias-Desmet/dp/1645021726
[9] https://www.amazon.com/Extraordinary-Popular-Delusions-Madness-Crowds/dp/1463740514

7. For those who are interested…

By carefully designing the context of the question (prompt engineering) that is posed to ChatGPT one can obtain less biased answers. The following substack post asks ChatGPT a sequence of questions that leads to the inevitable conclusion, that the precautionary principal should be applied to the novel mRNA technology.

https://thenobodywhoknowseverybody.substack.com/p/ask-chat-gpt-about-mrna

The sequence of questions posed to ChatGPT, was performed in light of the report recently released by means of a FOI request, regarding the biodistribution of the mRNA[10] (see page 40), submitted to the Australian TGA[11] in January of 2021, by Pfizer.

[10] foi-2389-06.pdf (tga.gov.au)

[11] Therapeutic Goods Administration, Department of Health, Government of Australia