# Status Quo Engines: Algorithms, AI tools, and the Need to Keep Research in Touch with Reality

Ilana Goldowitz, 2025

Our modern, data-driven society has a great predilection to base new programs and solutions on existing situations, even if those situations are pathological.

Some examples:

1. Breastfeeding has health benefits, but breastfed babies typically gain weight more slowly than formula-fed babies. The CDC's percentile charts that track infant weight were based on data from a 1977 study in which almost all included babies were formula-fed. Pediatricians used these charts to gauge the growth of most US babies born from the late 1970s onward, thus making the faster growth rate of formula-fed babies into a standard. A 2006 panel convened by the CDC and other groups found that some doctors were inappropriately advising parents to switch from breastfeeding to formula feeding to help their babies "catch up" with the CDC growth charts. In 2010, these concerns led the CDC to recommend that pediatricians switch to using newer WHO growth charts, which are based on breastfed babies.[1]

2. In 2009, the National Oceanic and Atmospheric Administration (NOAA) rolled out new "catch-share" regulations that were intended to solve problems with overfishing in the New Bedford, MA fishery. Under the regulations, the total fish harvest would be capped, and each fishing business would be allotted a piece of the pie based on its catch of each vulnerable species in past years. The result was that companies that had already been doing the most damage to vulnerable species were given the highest allocations, and those companies were able to further increase their shares by driving more responsible fishermen out of business.[2]

3. In hospitals across the US, a healthcare algorithm was tasked with determining which patients' health conditions were severe enough that they should be offered extra care. The algorithm influenced the allocation of care to millions of patients beginning in year. Then in 2019, a group of researchers noticed that the decision-making algorithm was much less likely to offer extra care to black patients compared to equally sick white patients. This was happening because the algorithm used the money that had been spent on each patient in past years as a proxy for which patients were "sicker." Thus, the algorithm biased decision-making against patients who had been undertreated due to poor healthcare access. Because the algorithm allocated services away from patients for whom money and access had

> been limiting factors in their health, and toward patients for whom they had not been, it ensured that money was spent in the *least efficient way possible*.[3]

Algorithms and AIs didn't originate this "status quo bias," but because these tools typically rely on past data to make predictions, basing the future on the past or present state is something they're naturally good at. Several research groups and organizations are working on reducing AI and algorithmic bias in healthcare, hiring, and other areas, but the problem might be inherent to any system that relies on making associations in existing data rather than going out and investigating the causes of problems.[4,5]

Science-based policies shouldn't reward the irresponsible or help entrench existing pathological states. But this can too easily happen if scientists crunch data, and then medical associations, regulators, or company executives go forward with the idea that the data represent the way things should be.

AI training typically depends on being able to tell the machine what the correct answer is. This creates a fraught situation for using AI tools in scientific research. In research, we rarely know what "the correct answer" is. If we think we do know this, we may be wrong in cases with different context from the ones we've considered — even if we don't know there's a difference. And the gap between the "correct answers" we think we know and the actual situation in nature is greatest in the most innovative or under-explored areas of study.

A related limitation of AI tools is that, lacking the ability to go out into the world and collect their own data, they only have access to the data we choose to feed them. This data may have a scope that's too narrow for our intended purpose. It may consist of the data that were the most accessible and easiest to collect, not the most relevant to the answer we are seeking.

**Case study 1**

Understanding the three-dimensional structures of proteins is important for the progress of many biological subfields as well as for drug discovery, but protein structure determination traditionally relied on painstaking laboratory work by molecular biologists and crystallographers. In recent decades, a growing array of computational tools have offered scientists the ability to predict protein structures. In 2018, the AI system AlphaFold beat out 97 competitors when it accurately predicted the structures of 25 out of 43 proteins whose structures had recently been solved in the lab, thus winning that year's Critical Assessment of Structure Prediction (CASP) competition.

AlphaFold was trained on the Protein Data Bank (PDB), a large database of proteins whose structures were determined mostly through crystallography. This AI system puts multiple recent discoveries in protein biology and molecular evolution to work and has gained recognition as an important advance in structural biology. Still, there are important limitations.

The first issue is that the solved structures in PDB represent a biased sample of proteins. Rather than a random sample from every protein found in nature, they consist of proteins that a) scientists were able to successfully produce, purify and crystallize and that b) come from species humans were interested in or that were easy to studying in the lab. We don't really know how much we can extrapolate from these known protein structures to, say, the proteins of extremophiles (organisms that live under extremely high temperature, pressure, acidity, salinity, or other extreme conditions), or proteins with low sequence similarity to any protein in PDB, or proteins whose structures haven't been determined because they tend to kill the bacterial cells in which we try to grow them to generate enough protein for study.

Another issue with using PDB is that not all proteins adopt a single, static structure in nature. The up-and-coming fields of intrinsically disordered proteins (IDPs), intrinsically disordered regions (IDRs), and bi-stable "fold-switching" proteins are revealing that many proteins are fully or partly unstructured or can switch between two different structures. The older view of proteins as mostly static is giving way to an understanding that many proteins are in constant motion within cells, and that their interactions are less like the traditional view of a key turning in a lock than like a wrestling match between two strands of cooked spaghetti. Current estimates are that over 50% of eukaryotic proteins contain at least one long disordered segment, and these features appear to be key to many proteins' functions.

Scientists who work with IDPs/IDRs and fold-switching proteins caution that AlphaFold (and its successor AlphaFold2) may predict spurious structures for these proteins.[5] Scientists working on proteins that switch between two or more structures report that AlphaFold will typically find just one of the structures, or an intermediate structure, but the prediction will be given with high confidence. Scientists *are* using AlphaFold to (carefully) study IDPs and IDRs, and AlphaFold seems to do well at marking these regions as "low-confidence" predictions. However, the authors of a 2021 paper warn that those inexperienced with IDPs/IDRs who use AlphaFold may be tempted to make incorrect inferences that have "zero physical meaning" and that the AlphaFold process can introduce artifacts that are especially severe for IDRs.[7] Because IDPs/IDRs are still a lesser-known field, some scientists today may be looking at AlphaFold's static outputs with less than the appropriate skepticism.

We can use AlphaFold and similar AI tools well by understanding where they perform better and worse, where nobody knows the "ground truth," where we may be wrong about it, or where we may be extrapolating too broadly.

If humans had turned large portions of the task of investigating protein structures over to an AI (or to scientists who rely heavily on an AI) before we had discovered IDPs and bi-stable proteins, we may never have known about these important phenomena. It took human intelligence to make the discoveries leading to the recognition of IDPs/IDRs: the hands-on work of trying and failing to crystallize some groups of proteins, and the innovative thinking needed to notice where the dominant paradigm was breaking down.

**Case study 2**

Machine learning programs for medical image analysis have been under development for decades. A few of these programs are currently used in the clinic, though reception has been mixed. To the extent that we can define what constitutes a “normal” or “abnormal” imaging result, and to the extent that representative and accurately-labeled imaging datasets are available, it should be possible to eventually develop useful AIs that replicate physicians’ judgments regarding images.

But image evaluation is only one step in a larger diagnostic process, the same step filled by laboratory blood tests and the like. Because doctors who understand a lab test’s limitations can make better use of it, clinical leaders like Catherine Lucey, MD of UC San Francisco are advocating for doctors to place less emphasis on the tests alone by placing test and imaging results within a Bayesian framework (such as by using Fagan nomograms, also called “Bayesian ladders”), with the goal of improving medical decision-making.

From this perspective, if AI tools eventually become capable of replacing doctors in specific image analysis tasks, allowing AI tools to take over *diagnosis* would be a step in the wrong direction. Instead, the advent of radiology AIs would be analogous to the mid-1800s replacement of doctors skilled at tasting urine by laboratory tests for glucose in diagnosing diabetes. The laboratory glucose tests eventually proved more accurate than the human tongue, and most human doctors probably didn’t mind giving up that role.

But any laboratory test has its limitations, and lab results shouldn’t be taken as “the answer.” Doctors need to consider the prior probability that a patient has a particular disease based on an exam and patient interview, the possibilities for other diseases that may explain the patient’s symptoms, available tests’ specificity and sensitivity, the consequences of a false positive or false negative, the uncertainty around the definition of “normal,” differences among populations, and cases where the test will fail, in order to decide whether to apply the test in a particular patient’s case. The same considerations, plus more, apply to AI analyses.

Tools used in research contexts require the same careful consideration of context and where these new results fit in with our other sources of information. If researchers don’t consider and report all the relevant information they have, including information on why the data were generated and the limitations of the tool used, we risk flattening complex information into a conclusion that may or may not be true, or that only applies in a narrow range of cases.

***

If AIs can’t be made to think outside the data, some advanced AIs may just be better status quo engines, skilled at propagating harmful situations forward in time. Others may help us convince ourselves we have everything we need to get the right answer to a scientific question, when in reality we need to continue broaden our scope of data collection and consider additional context. In the corporate world, AI could easily feed right into many business leaders’ inclination to throw new, heavily advertised technology at a problem instead of understanding it.

Alignment between AI systems and the goals of their developers is a much-discussed problem, but it is not enough. As with traditional product design, the way developers understand real-world environments and the needs of users is often far from reality, at least until they go out and talk to people. The Lean Manufacturing movement and the broader Lean Thinking movement have facilitated greater value for customers precisely by teaching people to open up communication, investigate the real roots of problems hands-on, and seek alignment of goals and actions not only within a company but with customers and vendors. To improve our thinking and action, we should be learning to break out of our status-quo-is-correct thinking patterns instead of reinforcing them and making them harder to detect.

## References and Notes

1. https://www.cdc.gov/mmwr/preview/mmwrhtml/rr5909a1.htm
2. https://hakaimagazine.com/features/last-trial-codfather/ This allocation system probably increased overfishing. It turns out that the companies most willing to overfish vulnerable populations for short-term gain were also willing to deceive regulators. The companies owned by fraudster Carlos Rafael, dubbed The Codfather, circumvented the new regulations (while contaminating data that fisheries scientists use for forecasting) by mislabeling fish and misreporting how much they were catching. Carlos Rafael had gained ownership of a large proportion of the New Bedford fishery by the time he was caught.
3. https://www.science.org/doi/10.1126/science.aax2342
4. http://ziadobermeyer.com/wp-content/uploads/2019/09/measurement_aer.pdf
5. http://ziadobermeyer.com/wp-content/uploads/2021/08/Predicting-A-While-Hoping-for-B.pdf
6. https://www.nlm.nih.gov/research/researchstaff/labs/porter/pdfs/chakravarty_AF2.pdf
7. https://www.sciencedirect.com/science/article/pii/S0022283621004411