The Best Numbers are in Sight. But Understanding?

Roald Hoffmann and Jean-Paul Malrieu[*]

[to be published in "Artificial Intelligence and Quantum Computing - Quantum Intelligence," edited by Greg Viggiano, 2021]

Abstract: The confluence of quantum computing and artificial intelligence (simulations using machine learning and neural networks) is going to bring us soon the most accurate solutions to Schrōdinger's equation for small molecules. Which, in our trade, we have been struggling to get approximately right for decades.

But those numbers -- now hard-won, soon easy -- will provide close to zero chemical understanding. To justify this provocative statement, we need to define understanding, a jewel of human thought for centuries. The ability to form explanations, to teach a student, play an important part of the definition. The potential social and personal abuses of an AI world are becoming clear to people. What concerns us is something more subtle – that the evolving mastery of quantum computing and artificial intelligence are part of a historical trend to assert the primacy of numbers, and deny the existence of "explanations".

Yet we cannot go back. AI and understanding, what we mean by Theory. will coexist; let's see if we can sketch a conjoined future for them.

The title voices our skepticism. And given that the authors are both in their eighties, it is easy to attribute it to their age and attendant creeping conservatism. But in a time of hyped enthusiasm for The New Jerusalem of IT, we thought there

[*] *RH: Dept. of Chemistry and Chemical Biology, Cornell University, Ithaca NY 14853, USA; J-PM: Laboratoire de Chimie et Physique Quantiques, Université de Toulouse 3, 118 route de Narbonne, 31062 Toulouse, France.*

might be a place to question the systematic confidence voiced and eventually define a more balanced, if oblique, perspective.[1]

Our response to what so-called "artificial intelligence" has done and will do to our lives is complicated. We want to rehearse with you the pretty obvious reasons why this is so, and then go on to what really worries us. This is the attack -- scientific, philosophical, and psychological – that artificial intelligence, augmented by quantum computing, might represent on a human jewel, the idea and processes of understanding

Attitudes: Quantum Computing

It's hard (and would be strange) to be against quantum computing. Quantum computing is marvelous on several accounts. To those of us who devoted their lives to solving approximately the wave equation of Quantum Mechanics, and connecting it up to the very tangible world of chemistry, to see superposition and entanglement turn into operational reality and precise numbers is astounding. We thought of those inherent quantum mechanical notions as philosophical quandaries, not for us, quantum mechanics that we were. And here, today and not tomorrow, these concepts are put to practical use. The design of the physical building blocks of a quantum computer – the qbits – also returns us to chemistry. Realizable qbits may be inorganic molecules, or an addressable defect in a crystal. One worries about state fidelity and decoherence, and so a new bridge between chemistry and computing forms. We are hiring young people in this specialty.

And Artificial Intelligence

Decades of life with computers have cleared the way for artificial intelligence to enter our lives. To search a library's holding for the title of a book it contains, to plot out a wave function – those are computing tasks, easily described for a human working with more primitive tools, easy to understand how the computer does it, since we wrote the algorithm. And it is always marvelous to see the computer's speed, and to smile at how easily our mistakes mislead the machine. On to machine translation from a foreign language, and our reaction is still unsullied wonder at how well Google Translate and commercial system can do it, using a neural network framework.

We go on to face recognition, and now things become murky. Not because of lack of admiration of how well the program does it, but because of the use to which human beings and machines may put these programs. Let us be specific: Cliff Kuang has described clearly a study by Michal Kosinski, using data from dating profiles, correlating sexual preference in individuals with self-identification, based on images of their faces and self-supplied profiles.[2] The data scientists achieved a much better correlation than human evaluators. Now that seems harmless (but could some institutions evaluate the suitability of a person for a job based on that "identification.") But the stretch to misuse in the treatment of certain groups, say the Uighurs in western China, as a prelude to "re-education", is not a long one.[3]

So, the problem is not in the machine, but in human abuse. Two comments on this:

1. OK, so human beings are fallible. Given any technological innovation, a certain fraction of the human users is bound to misuse it – for their own gain, to shame or harm others, or just for irresponsible creation of chaos. Surely the best way to counter this is to have the legal structures and strictures to limit and guide behavior on the computer.

Of course, we must try. But realistically, human misuse (whether of chemistry in synthesizing to order new, addictive opioids, or plagiarism) always outpaces in its ingenuity and scope the rules and regulations we impose. The mental, not physical, energy that powers trolls is hard to match in a defensive stance.

2. So maybe humans should get out of this. No, the ethical situation is not resolved by leaving the choice in the hands of machines. Remarkably, this argument has been made in another context:

> *"Several military experts and roboticists have argued that autonomous weapons systems should not only be regarded as morally acceptable but also that they would in fact be ethically preferable to human fighters. For example, roboticist Ronald C. Arkin believes autonomous*

> *robots in the future will be able to act more "humanely" on the battlefield for a number of reasons, including that they do not need to be programmed with a self-preservation instinct, potentially eliminating the need for a "shoot-first, ask questions later" attitude. The judgments of autonomous weapons systems will not be clouded by emotions such as fear or hysteria, and the systems will be able to process much more incoming sensory information than humans without discarding or distorting it to fit preconceived notions."*[4]

This argument is weak: the robot may be programmed to survive, since they are expensive, and may be employed in the next task that requires no judgment.

Taking a life, even just hurting someone, cannot be delegated to an algorithm. With the action comes moral responsibility, our deepest one. Were we to pass to that world – of machines making decisions of life or death – then our society has failed us, and we it.

The Wave

Actually, what bothers us about the wave of machine learning and neural networks, of artificial intelligence, that is breaking over our heads, is that AI makes epistemological claims. The claim is that artificial intelligence provides real understanding. Because one needs to get the numbers (the energy of a molecule) or the face right in a recognition task, to say that one really understands. Even if this were true (see next section), does the correctness of the answer provide an explanation? As René Thom said "Prédire n'est pas expliquer / To predict is not to explain."[5]

Let us provide the background for this assertion by telling you what we mean by understanding, and place us in a line of theory that has that as a goal. And then the attack of AI on understanding will be clarified. We will also outline a psychological aspect of the struggle.

Understanding

Elsewhere we have gone at some length into the attributes of understanding.[1] It is often tacit, a state of mind. And most of the time qualitative, though it can certainly have a quantitative aspect to it (and that is where Quantum Computing comes in). In physics and chemistry understanding usually resides in Theory and in interpretative models, a practice of analysis of any observable in terms of the possible physical or chemical mechanisms (causes, elementary actions) that could lead to the observable. And making an order of magnitude estimate for the role of their contributions.

There is a strong pedagogic aspect to understanding. In fact one of the defining aspects of understanding is that it is that which can be taught. To an intelligent graduate student (forget the professor; often they are unteachable), in words or concepts, in equations.

This quality, the teachability of understanding, is what we use all the time to keep honest our colleagues who are aficionados of machine learning or neural network. We probe "What have you learned from your calculation that you can teach me?"

There are some parts of physics where the gross overall understanding is there, but small effects – which may be measurable – are missing. We think of the Lamb shift in the excited states of a hydrogen atom.[6] No one found fault with Bohr's theory of atomic spectra because it didn't have the Lamb splitting in it. That detail, while it was definitely a part of understanding hydrogen, was not the major concern. It could be and was analyzed, and its understanding was important to the overall structure of physics. Yet it was not the main fact about hydrogen.

Where We Came From

Something must be responsible for the skepticism about AI of two theoretical chemists, whose whole life was enabled by the progress of electronic computers.

We come from quantum chemistry. 90 years ago, a sage of our tribe. P. A. M. Dirac, wrote:

"The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation."[7]

We have been trying. But far from brute force applied mathematics, we have tried to construct frameworks that allow us privileged passages to forming explanations. And, in another direction, we have tried to connect up to qualitative ideas that chemists have formed, about the ability of atoms to share electrons, and of the varying propensity of nuclei to hold on to them. And to insights from quantum mechanics. Here especially useful was perturbation theory, a time-honored way of getting physical insight (understanding!) of equations that could be solved only approximately.

The sound of the old key punch is in our ears. As is Fortran. We became pretty good at what we do, meaning that we could calculate some numbers approximately, interpret (understanding!) the numbers as factors influencing a real outcome, and, ultimately, construct explanations. Chemists, ever so talented at making molecules that effected a small change of atomic composition (an H here replaced by a F, and NO_2 group by a CN), could then test our qualitative prediction. So ensued the most wonderful aspect of science, the dance back and forth between theory and experiment.

When it worked (and it didn't some of the time), people understood.

The machine understands?

Machine learning and neural networks, new engines of simulation arrive. We, who solve Dirac's equations approximately (but do much more) are deemed to have been replaced. Not entirely, for we are needed to get exactly right the energies of a goodly set of molecules involving the same elements. This is the "training set." In machine learning, a theoretician might also design the indicators which the AI machine uses to make its correlations – these will the identity of the atoms, their positions in space, other characteristics. And then the program, whether it is in supervised or unsupervised machine learning, is set loose to find its own way to the best fit of, say, the energy of the training set molecules.

The outcome (it would not be publishable if it were different) is a predicted energy for an unknown molecule, one outside the training set, that is lower (better) than our quantum mechanical approach can give us.

You can see what the journalist writing the press release on that work will say. “AI now understands the molecule better than any calculation”. The scientist often does not make the claim as patently, but when you mix the justifiably enthusiastic scientist with his or her institution’s journalist, and the news-worthiness-hungry editors of Nature or Science, you get the perfect storm for (to put it mildly) exaggerated epistemological claims.

Psychological Factors

Roald thinks another danger lurks in the practice of simulation, which creates a block to even imagining that explanations, in the time-honored way of theory, might even exist. And that derives from the psychology of human-machine interactions.

Computer programs are naturally complex, made up sometimes of hundreds of lines of code. There always problems in getting them to work; the chore of debugging is an experience all of us have shared. If all that work has to be done to get a number, surely there cannot be a simpler way to get it, even approximately?

There is no question that computers are so much more efficient than human beings in pattern recognition. And it could be that human beings fall for simplistic explanation in order to avoid complexity. In their lives, and in Nature. Sometimes we think that the current wide range of belief in conspiracy theories, no matter how cockeyed they might seem, comes from that desire for a simple, ordered world. Nevertheless, we think the process of getting computer programs to work predisposes the programmers to depart from the hard path of finding a theoretical explanation.

And What Will Quantum Computing Do for This?

For small molecules, quantum computing has solved those governing equations more accurately than any other calculation. It will soon do it for larger molecules. The best numbers will be provided by quantum computing. And no understanding.

Let us claw our way back

Are we done maligning and moaning? First, the problem of AI providing numbers but not explanations was recognized early on. Part of the field is moving on to crafting the programs to tell us how an AI implementation (machine learning or neural networks) learns, how it does what it does. This is an outstanding problem, the field of "explainable AI".

For instance, DARPA, the US Defense Advanced Research Projects Agency, has had for several years a program (sponsoring university research) on under this name. Program Director David Gunning's textual exposition of the program is couched in Defense Dept. lingo, but clearly the goal is more general, as this graphic[8] demonstrates:

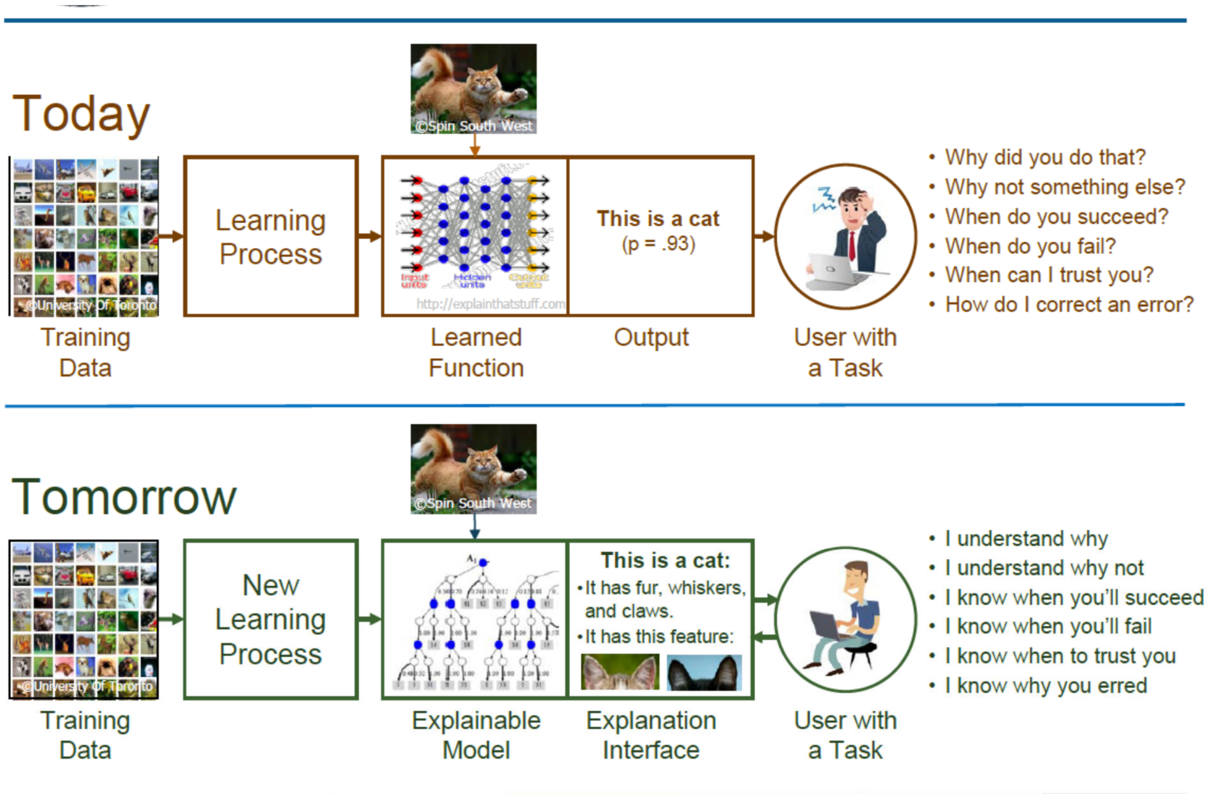

Figure. A conception of the present and intended state of AI. *From a presentation by David Gunning on Explainable Artificial Intelligence, DARPA*

Classifying dogs rather than cats might be more of a challenge.

There are several productive directions in the field – one is simply exploring what happens inside the box – to know what the computer does at the structural level, and perhaps to learn from it something in the ways of discernment or proof strategies.

Coming up with regularities in numbers is easy. Defining the regularity in terms of a classical theorem (say one of Ramanujan's incredible series) is a challenge. Finding a theorem that is "interesting" to a human mathematician is hard. But there is progress even on this.

Explainable quantum computing? The beauties and constraints of error correction, the effects of decoherence can be explained. But the result, a number, seems to be inherently mute. Until a human being (or maybe an explainable AI program) builds a story out of several or many such numbers.

Seeking Numbers, Forming Theories

Like it or not, the future will hold much more simulation and AI than we would like. At the same time, both of us have confidence in our students, more than in ourselves – they will find a *modus vivendi* in theoretical chemistry in that future world. We'd like to think about what that world would be like, with at least a partially open mind. And perhaps find something special in that future.

Our own experience with numerical calculation (call it simple simulation) and the building of theories, gives us a clue. If we look at it abstractly, we see for both us, quite different scientists because of our education and history, a similar dynamic. We alternated spells of highly detailed quantum chemical calculations, with the construction of simplified explanations that involved our game pieces, orbitals. These orbitals, or rather the electrons in them, were/are involved in very specific interactions. These we puzzled out, sometimes directly, sometimes over twenty years of work, through specifically tailored probing calculations -- simulations with an aim to learn something from them, not necessarily to simulate reality.

If you allow us the conceit of thinking of our numerical calculations as simulations, and grace them with the AI label (how we would have blanched at the thought!) then we alternated periods of AI simulation with theory building. It worked! At the end of it was a double satisfaction – a qualitative theory to explain the chemistry, and a prescription for the level of theory needed to get specific experimental numbers right.

Narrative

We built a story out of numbers and theory. Our students will do this better, dazzling us how they jump in and out of riffs of computation.

Stories are ancient, stories are in the human psyche. When the wooly mammoth was killed, the hunters told in reliable geographic detail where the beast was found (cross the river, turn right at the giant oak, go up beside the cliff). And, we suspect without a pause they recounted the fierceness of the cornered giant, the courage of the hunters it trampled.

Theories are stories. They share with fictional tales temporality – a calm beginning, a problem to be solved in a tense middle, where you don't know if this technique applied will work, and stumble toward another. And an ending, which itself carries an inherent tension -- one does want to give the impression that something significant was learned, and yet must leave the reader/listener with a feeling that there remain mysteries to be solved. Theories clearly ascribe causation – they are the most deterministic of narratives. [9]

Thinking one has seen all the causes, excessive rationality, is actually a problem for science.

And do scientific stories have human interest? Oh, they do. Sometimes we need to be privy to the cognitive structures of the field to appreciate them – as in Einstein's remarkable use of the entropy of radiation in his classic 1905 photoelectric effect paper. Sometimes we need the rivalry of competitive theories, real people pushing them on. Sometimes a theory is so compelling that it needs little new experimental support – we think Darwin's theory of evolution, the greatest story ever told, was like that.

Futuristic

We think that AI and quantum computing will enter the chemistry of the future in two ways. The first is in their simple utility – in perfecting the technological capabilities of the chemist to make any specific molecule/material, or to design and make a molecule with certain desired properties. The aim should be to serve humanity, of course. Even as we are aware that abuse of that wonderful synthetic capability of human beings is, sadly, common.

But our spirit wants more. We want to understand. And quantum computations will remain mute. But explainable AI has a future – a theoretician will use it to find regularities worth thinking about, and will explore their origins. He or she will come up with better frameworks for understanding. Which can be taught.

World-building

We imagine a world in chemistry and other sciences as moving toward near infinite capacity in the Techne of searching for facts and properties. And also a world where we hope our students, and not just a privileged few, will experience, working hard all the way, the Sophia of making sense of things.

And now a flight of fancy, an excursion into the imaginative realm, appropriate to the origins of this book and our debt to Jules Verne and Ursula LeGuin.

Elsewhere we have imagined a Museum of Science of the future, no doubt virtual, as the current pandemic has gotten us used to. We describe the objects in it:

> *"In one room hangs the Cassini mission image of the lakes of ethane on the surface of Titan. In another, the discovery of archaea. And of how their lipids differ from ours. In a third room we see Onsager's solution of the two-dimensional Ising model, in a fourth room the synthesis of coenzyme-B_{12}."*[10]

This is a sacred space, these are artistic and scientific achievements, of pervasive <u>spiritual</u> value.

Does that seem a long jump, from coping with AI to the spiritual? Or just two scientists going soft? Let us trace the chain of thinking that moved us there.

Our lives and our science have been transformed by Information Technologies, with more change on the way. And there are dangers around us, from real anthropogenic climate change, to totalitarian tendencies masquerading as populism. The two of us differ in our degree of optimism about the human condition. Will we in fact be sage, build safeguards, and form more just societies? Or will too many perish along the way; humanity will survive, but at what cost?

Whatever happens, artificial intelligence will be an essential part of it. And … there will remain human beings who long to experience and be motivated by more than ultimate efficiency and accurate numbers. Who will seek understanding. The space of understanding in us is the same space that is touched by music, poetry, all the arts. It is a sacred, deeply human space.

The theoreticians of the future will be listened to because they will have learned how to tell their stories of theoretical discovery in a convincing narrative. It is not out of place that some will take Carl Sagan's pointer and use fiction. And

-- wishful thinking – some of these theoreticians will also be master teachers, recognizing the intertwined ways of understanding and teaching.

The aesthetic aspects of chemistry will always be there – the stinks, bangs, and vivid color changes that attracted Primo Levi[11] and Oliver Sacks[12] will be brought up to date by the heirs of Theodore Gray[13] and Yan Liang.[14] The sheer variety of properties and function in the denumerable yet astronomically (no, chemically) large set of possible molecules will pull in people, as will symmetry and the lack of it, the power of frontier orbital reasoning, and the sharp logic of organic synthesis.

Just thinking about all that beauty in variety sends shivers through us. Which is perhaps the evidence that the logical has caught sight of the spiritual. The husk covering the sublime has been breached.

We see the future theorist making use of AI and quantum computing to play infinite games. The games' purpose is not their reliable numbers, but the stories that the theorist can assemble from all those explorations of the "What if?" world. The understanding gained will be eminently teachable; though it doesn't hurt if there is some incentive in the system for the pedagogic. The best stories that emerge from this directed roaming will please you and make you shiver. Perhaps then the distinction between art and science, those different ways of knowing this world, will become less important. Because both touch the spiritual in us.

[1] The arguments in this paper are set forth in greater detail in three essays: R. Hoffmann and J.-P. Malrieu, Simulation vs Understanding: A Tension, in Quantum Chemistry and Beyond. Part A. Stage Setting, Angewandte Chemie, 59, 12590-125610 (2020); B. The March of Simulation, for Better or Worse, Angewandte Chemie., 59, 13156-13178 (2020); C. Toward Consilience, Angewandte Chemie. 59, 13694-13710 (2020).

[2] "Can A.I. Be Taught to Explain Itself?": C. Kuang, New York Times, November 26, 2017, on Page MM46 of the Sunday Magazine.

[3] "Facial Scans Tighten China's Grip on a Minority," P. Mozur, The New York Times, April 15, 2019, p. 1, A8.

[4] "The Pros and Cons of Autonomous Weapons Systems, A. Etzioni, O. Etzioni, Military Review, May-June 2017, 2017, 72-80. The internal reference to Arkin is Ronald C. Arkin, "The Case for Ethical Autonomy in Unmanned Systems," Journal of Military Ethics 9, no. 4 (2010): 332–41.

[5] R. Thom, É. Noël, "Predire n'est pas expliquer," Flammarion, 2009.

[6] https://en.wikipedia.org/wiki/Lamb_shift

[7] P. A. M. Dirac, Proceedings of the Royal Society of London. Series A, 1929, Vol. 123, 714-733.

[8] From a presentation by David Gunning on Explainable Artificial Intelligence, DARPA. https://www.darpa.mil/program/explainable-artificial-intelligence.

---

[9] For more on storytelling in science, see Roald Hoffmann "Narrative," American Scientist, **88**(4), 310-313 (2000); "Storied Theory," American Scientist **93**(4), 308-312 (2005); "The Tensions of Scientific Storytelling," American Scientist, **102**, 250-253 (2014).

[10] Ref. 1, part C.

[11] Primo Levi, The Periodic Table

[12] Oliver Sacks, Uncle Tungsten. Memories of a Chemical Boyhood,

[13] Theodore W. Gray, The elements : a visual exploration of every known atom in the universe, New York : Black Dog & Leventhal Publishers. 2009; Molecules : the elements and the architecture of everything; New York, New York : Black Dog & Leventhal Publishers, 2018.

[14] Yan Liang, https://www.beautifulchemistry.net/; https://www.l2molecule.com/about; https://www.envisioningchemistry.com/