Why Modern AI Is So Successful, How to Make It Even Better, and How It Is All Related to Uncertainty Quantification, Physical Paradoxes, and Climate Models

Miroslav Svitek[1], Olga Kosheleva[2], and Vladik Kreinovich[3]
[1]Faculty of Transportation Sciences
Czech Technical University in Prague
Konviktska 20, 110 00 Praha 1, Czech Republic
miroslav.svitek@cvut.cz
Departments of [2]Teacher Education and [3]Computer Science
University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract

Our analysis of all these questions come from the fact that, according to most physicists, every model is approximate – there is no way to find the final Theory of Everything. With the current abundance of data, this leads to the phenomenon called *macronumeracy* – when statistical methods, based on the assumption that the model is correct, fail all proposed models. We explain why this happens, why AI largely avoids this problem – although not entirely – and how to make these models even better.

1 Can physics come to an end?

Why physics? Everything in the world – be it planets of living beings or rays of light – consists of elementary particles; see, e.g., [1, 9]. So, in principle, if we know how the states of elementary particles change with time – which is what physics studies – we will be able to predict how the state of the Universe changes, how everything changes. In this sense, physics is very important.

Because of this importance, physicists have been trying to find equations that described the particles' dynamics. In the past, this search has led to many practical successes, so societies allocate a lot of resources to fundamental physics.

Can physics come to an end? Sometimes, physics goes through a revolution, when new discoveries are constantly made – as it happened with Newton's laws, as it happened in the early 20 century with quantum physics and relativistic physics. But between the revolutions, there is usually a period where a progress is slow. This slowdown make some researchers think that there are fewer and fewer no physical laws to discover, we are getting closer and closer to the truth – and soon we will find the True Equations – what people call United Theory of Everything.

This happened in the late 19 century. When Max Planck, one of the future founders of quantum physics – after whom the main constant of quantum physics is named – informed his future professors that he wanted to study physics, he was strongly discouraged: according to these professors, physics is almost done, there are only a few small thinks to do.

Such an opinion was not limited to some obscure professors: when after the quantum and relativistic revolutions, progress started slowing down, no one else but the great Einstein believed that soon we will discover the United Theory of Everything. Not only he believed in this, this idea became his main research area, he spend the last several decades of his life searching for such a theory.

So can physics come to an end? Nowadays, when we have gone through many such boom-and-slowdown cycles in science, when hopes for the Theory of Everything were disrupted by a new revolutionary discovery, most physicists – at least most physicists whom we know and/or whose related papers we read – believe that such a theory is not possible, and physics will continue (although of course, some researchers are still trying to come with the Theory of Everything).

In other words, no matter what new equations appear, all the equations do not provide an exact description of reality – they are all *approximate*. From the practical viewpoint, these approximations may be good enough – for example, already Newtonian mechanics provides a very accurate description of many real-life phenomena: for example, it allows us to correctly predict solar eclipses hundreds of years ahead. But from the fundamental viewpoint, all current (and future) theories are approximate. They may be good now, but eventually, new phenomena are discovered for which a more accurate theory is needed.

With Newtonian physics, this happened when researchers got more information about phenomena both on the micro level – level of molecules, atoms, and elementary particles – and on the mega level – level of galaxies. To describe these phenomena, physics needed to come up with, corresponding, quantum and relativistic descriptions.

2 This leads to a – somewhat unexpected and somewhat paradoxical – important challenges

At first glance, this seems like a far-from-practice philosophical question. At first glance, when the theory is good enough for most current practical purposes, what is the big deal? From the philosophical viewpoint, whether the equations are exact already or whether a new phenomenon will appear hundreds of years from now – as it happened with Newtonian physics – may sound important, but it does not look like something practically important.

But nowadays, it has become important. Until recently, the above argument worked: engineers and scientists got great applications by using Newtonian physics, not by trying to find a more accurate theory. For example, space exploration was almost exclusively based on Newtonian mechanics. Nowadays, with GPS, we need to take into account relativistic effects to get more accurate location, but the original spectacular successes – missions to other planets, landing on the Moon – were based only on the Newton's (approximate) theory.

But now the situation has changed, and the reason for this is fundamental. How is science progresses? In a nutshell, first, new phenomena appear that show that the current theory is not perfect. Then someone – Newton, Einstein, others – come up with new equations that fit all the data. And then comes the ultimate test – the new theory is compared with new observations, and if it fits Nobel prizes are given. this sounds natural and reasonable, but if one thinks about it – the testing is based, in effect, on the hypothesis that the new theory is exact. But in reality, as we have mentioned according to the prevailing physicists' view, the new theory is inevitably approximate too – more accurate that the current one but still approximate.

We may not feel this approximate character, this approximation error if it is very small – and this is what happened until now. In general, if we have N observations with approximation accuracy ε, them, according to statistics (see, e.g., [8]), we can determine the theory with accuracy $\varepsilon/\sqrt{N}$. In the past, N was small. For example, in the beginning, we had very few observations confirming general relativity, to the extent that modern statistical analysis shows that famous original Eddington's 1919 confirmation of General Relativity theory was not statistically significant – and would not be publishable now.

This was the problem in the past – and in some areas, it is still a problem – that getting each experimental result is difficult, and we only have few of them. But in many other areas, technological progress allows us to generate many experimental data – and with the resulting large N, the accuracy $\varepsilon/\sqrt{N}$ with which we can determine the equations is so small that it is smaller than the accuracy of this model – we can see that this new model is only approximate. As a result, whatever new theory we propose, we cannot get a statistically perfect fit.

This is a paradoxical situation. In the past, progress of theoretical physics was limited by the experimental abilities – we need to wait for new data until we can come up with a new theory. Now, getting new data is often relativity easy, sensors and cheaper and more accurate than ever. Space telescopes, accelerators, etc. provide huge amounts of data – it is worth remembering that the current speedy Internet was largely designed to handle the need to process data produced by CERN. That should be helpful, but in reality, this technological progress clips with wings of theoretical science – no new theory can be confirmed.

This may sound paradoxical but it can easily explained. For example, Ohm's

Law – a classical example of a linear dependence – is only approximately linear. Ohm was lucky that his measuring instruments were not very accurate, as a result of which his data was a perfect fit for a linear dependence – otherwise, it would have been very difficult for him to find a relation between current and voltage. Newton would most probably not have been able to discover the laws of mechanics is the instruments of that time were accurate enough to detect relativistic and quantum effects. These are examples of what is called *macronumeracy*: the curse of having too much data; see, e.g., [3, 6, 7].

This phenomenon explains the problem with climate science. There is a lot of discussion about different climate models, and a lot of skepticism about them – because with a large amount of weather data, none of these models fits the data in the usual statistical sense, we can see that all these models are approximate.

This phenomenon explains similar phenomena in psychology. Macronumeracy is ubiquitous. It also explains why first impressions are often the most accurate – because at the time of the first impression, we do not yet have too much data.

This is a problem not only in science: for example, a skilled psychotherapist can analyze and cure others, but therapists are usually unable to analyze and cure themselves – they know too much about themselves, and this prevents them from a good understanding of their own behavior.

3 But, surprisingly, AI techniques are successful where traditional methods are not

One would expect that AI would suffer from the same problem. Modern AI tools (see, e.g., [2]) are trained on thousands and millions of data points, they use more or less traditional statistical criteria for success – so it is reasonable to expect that they would suffer from the same macronumeracy problem as more traditional data processing techniques.

But, somewhat paradoxically, AI methods are often very successful. In contrast to these pessimistic expectations, AI models are often very successful: based on all these data point, they often provide very good predictions of the corresponding phenomena. For example, they can detect cancer based on the X-ray and other images as successfully (and sometimes more successfully) that best human doctors. How can we explain this paradoxical situation?

Why is AI successful: our explanation. So why is modern AI so successful? In deep learning, in particular, in Large Language Models like ChatGPT, we use huge amounts of data and still get very impressive results.

Our explanation is that modern AI techniques avoid the macronumeracy effects because they drastically decrease the amount of data. Most deep neural networks do it by using transformers – T in GPT – that significantly reduce the data's dimension. Convolutional neural networks reduce the amount of data by compressing several values from neighboring pixels into a single data point.

And in both cases, huge computations become possible because instead of using all the digits of each data value, they only use a few first digits and ignore the rest – e.g., by using 8-bit computations; see, e.g., [5].

4 How can we make AI tools even better?

Remaining challenge. Modern AI tools are often good (and sometimes spectacularly good), but they are not perfect: sometimes, they produce nonsense results, e.g., classifying a cat as a bus. How can we make them better?

The above arguments seem to indicate that for this, we need to further decrease the aount of data, but then we will have not enough data to make conclusions. So what can we do?

Two approaches. If we use current data processing methods, then we need to supplement the insufficient data with some other knowledge. Alternatively, we can try to come up with different data processing methods which are better fit to process smaller amount of data – and ideally, we should use both approaches.

Let us briefly describe both options.

First option: using additional knowledge. In practice, to make predictions, we not only use previous data, we also use logical reasoning.

Experience of physics shows that even when a new theory appears because of the new data, it later turns out that it could have been developed based on logic, since the original theory is inconsistent: has *paradoxes.* For example, the switch from Newtonian to quantum physics was motivated by experiments, but it turned out that, e.g., non-quantum physics cannot explain the existence of atoms: in pre-quantum physics, they would have collapsed in a few microseconds [1], since in this physics, an accelerating charged body emits radiation and thus, loses energy.

Other examples include black body radiation that in pre-quantum physics leads to physically meaningless infinite energy. General Relativity could be motivated by the fact that in Newtonian physics, the sky at night would have been almost as bright as in daytime.

Second option: modifying data processing. How can we modify data processing? Modern analysis of scientific data relies on traditional statistics: we form a model, we test whether the data is consistent with the model, and then we apply this model to make predictions and to recommend actions. This paradigm works when we have a reasonably small amount of data, but when we have a large amount of data, then, since the model is approximate, testing techniques will reveal that the data does not fit.

As we have mentioned, this problem is most acute in climate models, where so far, all proposed models are not statistically consistent with the data – which is one of the reasons why there is so much controversy about these models. Actually, we can appropriately modify current statistical methods – specifically, we need to take into account interval uncertainty, one of the basic uncertainty quantification techniques. Let us elaborate; see, e.g., [3].

5 How to modify statistical methods to deal with macronumeracy

Formulation of the problem: reminder. A usual statistical approach to processing data $x_1, \ldots, x_n$ is to come up with a model – e.g., based on the training part of the data – and then test whether this model adequately describes the remaining testing part of the data.

For example, if it turns out that the observations are consistent with a normal distribution with mean m and standard deviation σ, then we can use the chi-square criterion and check whether we have

$$\chi^2_{n,1-\alpha} \le \sum_{i=1}^{n} \frac{(x_i - m)^2}{\sigma^2} \le \chi^2_{n,\alpha}$$

for appropriate values $\chi^2_{n,1-\alpha}$ and $\chi^2_{n,\alpha}$.

These tests are designed in such a way that:

- when the actual distributions is the assumed one, this test returns "true" with frequency close to 1, while
- when the actual distribution is different from the assumed one, for sufficiently large n – above a certain threshold n_0 – the corresponding test fails with frequency close to 1.

This traditional statistical approach has worked successfully for more than a century. However, with the emergence of big data, when we have millions and even billions of data points, this traditional statistical approach often fails. The reason for this failure is clear (see, e.g., [3, 6, 7]): in most applications areas – e.g., in econometrics – all statistical models are approximate. When n was reasonably small, much smaller than the threshold value n_0, the tests still worked. However, big data often means $n > n_0$, so the tests fail. As a result, either we cannot find a model that fits the training data, or, if such a model is found, we cannot show that it fits the testing data. This phenomenon is known as *macronumerosity.*

This is well-known problem in many application areas – e.g., in modeling climate change.

A natural solution. That the model is approximate means that there are some close values $\widetilde{x}_i \approx x_i$ that fit this model. A typical statistical idea would be to find the distribution for the approximation error $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$, but in this case, we would still assume some exact distribution for x_i, so this brings us back to the same problem.

A natural solution out of this seemingly vicious circle is not to assume any specific distribution for Δx_i, but instead to use interval uncertainty, i.e., to assume that all the values Δx_i are within an interval $[-\Delta, \Delta]$. In this case, for each model and each corresponding test $C(x_1, \ldots, x_n) \le C_0$ that is not satisfied for the actual data, we can describe the degree to which data fits the model by the smallest Δ for which some Δ-close values $\widetilde{x}_i$ satisfy the test. Usually, the values Δx_i are small, so we can ignore terms quadratic in Δx_i; in this linear approximation, we have:

$$\Delta = \frac{|f(x_1, \ldots, x_n) - C|}{\sum\limits_{i=1}^{n} |C_{,i}|},$$

where $C_{,i}$ are partial derivatives of C with respect to x_i.

Then, e.g., between two models with the same number of parameters, we can select the model with the smallest Δ.

6 Interestingly, there are also unexpected positive consequences of the all-models-are-approximate idea

So far, we have been talking about challenges related to the idea that all models are approximate. Interestingly, while at first glance, the all-models-are-approximate idea is somewhat negative – we cannot hope to find the True Equations of Physics – it has positive consequences: under this idea, by using generic measurement results we can, in principle, speed up computations; see, e.g., [4].

This may be the reason why, when one gets stuck in trying to solve a problem, a distraction can help.

Acknowledgments

This work was supported in part by the US National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), HRD-1834620 and HRD-2034030 (CAHSI Includes), EAR-2225395 (Center for Collective Impact in Earthquake Science C-CIES), and by the AT&T Fellowship in Information Technology.

It was also supported by a grant from the Hungarian National Research, Development and Innovation Office (NRDI), by the Institute for Risk and Reliability, Leibniz Universitaet Hannover, Germany, by the European Union under the project ROBOPROX (No. CZ.02.01.01/00/22 008/0004590), and by the European Union under the project ROBOPROX (reg. no. CZ.02.01.01/00/22 008/0004590).

References

[1] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.

[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, Massachusetts, 2016.

[3] O. Kosheleva and V. Kreinovich, "For statistical analysis of big data, interval uncertainty is needed", In: E. Auer, M. Lahme, and A. Rauh (eds.), *Abstracts of the 20th International Symposium on Scientific Computing, Computer Arithmetic, and Verified Numerical Computations SCAN 2025*, Oldenburg, Germany, September 22–26, 2025, pp. 36–37.

[4] O. Kosheleva, M. Zakharevich, and V. Kreinovich, "If Many Physicists Are Right and No Physical Theory Is Perfect, Then by Using Physical Observations, We Can Feasibly Solve Almost All Instances of Each NP-Complete Problem", *Mathematical Structures and Modeling*, 2014, Vol. 31, pp. 4–17.

[5] C. Q. Lauter and A. Volkova, "A Framework for Semi-Automatic Precision and Accuracy Analysis for Fast and Rigorous Deep Learning", *Proceedings of the 27th IEEE Symposium on Computer Arithmetic ARITH 2020*, Portland, Oregon, USA, June 7–10, 2020, pp. 103–110.

[6] A. Nichols and M. E. Schaffer, "Practical steps to improve specification testing", In: N. N. Thach, D. T. Ha, N. D. Trung, and V. Kreinovich (eds.), *Prediction and Causality in Econometrics and Related Topics*, Springer, Cham, Switzerland, 2022, pp. 75–88.

[7] M. E. Schaffer, "Null hypothesis misspecification testing revisited: how (not) to test orthogonality conditions", In: V. Kreinovich, W. Yamaka, and S. Leurcharusmee (eds.), *Data Science for Econometrics and Related Topics*, Springer, Cham, Switzerland, to appear.

[8] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.

[9] K. S. Thorne and R. D. Blandford, *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics*, Princeton University Press, Princeton, New Jersey, 2021.