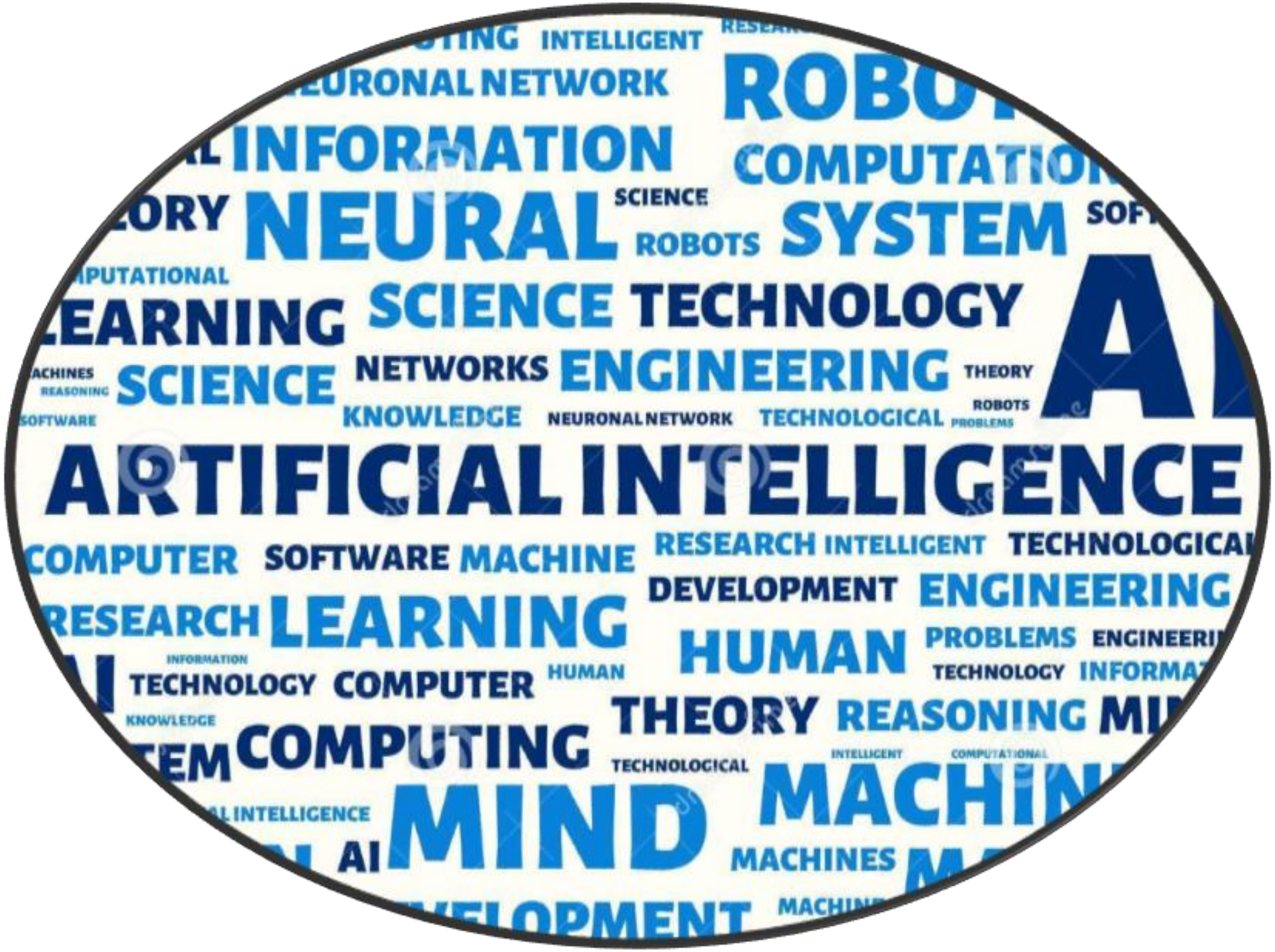
AI @ Startups – stories from the field

Madhu Shashanka

Why is Pune one of the best cities in India for startups?

Pune is considered one of the best cities in India for startups for a variety of reasons. Some of the factors that make Pune an attractive location for startups include:

- The availability of resources: Pune has a vibrant ecosystem of support for startups, with access to funding, networking opportunities, and a large pool of talent.
- The local market: Pune is home to a large and growing market for technology and innovation, with a strong presence of both domestic and international companies. This provides a favorable environment for startups to grow and thrive.
- The cost of living: Pune is known for its relatively low cost of living compared to other major cities in India, which can be attractive for startups looking to control expenses and maximize their resources.
- The quality of life: Pune is known for its high quality of life, with a pleasant climate, a rich cultural heritage, and a diverse population. This can be appealing to both entrepreneurs and employees looking to establish a startup in the city.



* https://thumbs.dreamstime.com/z/artificial-intelligence-image-words-associated-topic-artificial-intelligence-word-cloud-cube-letter-image-ill-94073519.jpg

The Unreasonable Effectiveness of Data

Alon Halevy, Peter Norvig, and Fernando Pereira, Google

Eugene Wigner's article "The Unreasonable Effectiveness of Mathematics in the Natural Sciences"[1] examines why so much of physics can be neatly explained with simple mathematical formulas

such as $f = ma$ or $e = mc^2$. Meanwhile, sciences that involve human beings rather than elementary particles have proven more resistant to elegant mathematics. Economists suffer from physics envy over

behavior. So, this corpus could serve as the basis of a complete model for certain tasks—if only we knew how to extract the model from the data.

Learning from Text at Web Scale

The biggest successes in natural-language-related machine learning have been statistical speech recognition and statistical machine translation. The reason for these successes is not that these tasks are easier than other tasks; they are in fact much harder

The unreasonable effectiveness of deep learning in artificial intelligence

COLLOQUIUM PAPER

Terrence J. Sejnowski[a,b,1]

[a]Computational Neurobiology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037; and [b]Division of Biological Sciences, University of California San Diego, La Jolla, CA 92093

Edited by David L. Donoho, Stanford University, Stanford, CA, and approved November 22, 2019 (received for review September 17, 2019)

Deep learning networks have been trained to recognize speech, caption photographs, and translate text between languages at high levels of performance. Although applications of deep learning networks to real-world problems have become ubiquitous, our understanding of why they are so effective is lacking. These empirical results should not be possible according to sample complexity in statistics and nonconvex optimization theory. However, paradoxes

conferences, I oversaw the remarkable evolution of a community that created modern machine learning. This conference has grown steadily and in 2019 attracted over 14,000 participants. Many intractable problems eventually became tractable, and today machine learning serves as a foundation for contemporary artificial intelligence (AI).

The early goals of machine learning were more modest than those of AI. Rather than aiming directly at general intelligence,

"… has become something of a scientific bandwagon.

… it has received an extraordinary amount of publicity in the popular as well as scientific press.

… it has perhaps ballooned to an importance beyond its actual accomplishments. "

"… has become something of a scientific bandwagon.

… it has received an extraordinary amount of publicity in the popular as well as scientific press.

… it has perhaps ballooned to an importance beyond its actual accomplishments. "

- Claude Shannon, 1956
  on Information Theory

The Bandwagon

CLAUDE E. SHANNON

INFORMATION theory has, in the last few years, become something of a scientific bandwagon. Starting as a technical tool for the communication engineer, it has received an extraordinary amount of publicity in the popular as well as the scientific press. In part, this has been due to connections with such fashionable fields as computing machines, cybernetics, and automation; and in part, to the novelty of its subject matter. As a consequence, it has perhaps been ballooned to an importance beyond its actual accomplishments. Our fellow scientists in many different fields, attracted by the fanfare and by the new avenues opened to scientific analysis, are using these ideas in their own problems. Applications are being made to biology, psychology, linguistics, fundamental physics, economics, the theory of organization, and many others. In short, information theory is currently partaking of a somewhat heady draught of general popularity.

Although this wave of popularity is certainly pleasant and exciting for those of us working in the field, it carries at the same time an element of danger. While we feel that information theory is indeed a valuable tool in providing fundamental insights into the nature of communication problems and will continue to grow in importance, it is certainly no panacea for the communication engineer or, *a fortiori*, for anyone else. Seldom do more than a few of nature's secrets give way at one time. It will be all too easy for our somewhat artificial prosperity to collapse overnight when it is realized that the use of a few exciting words like *information*, *entropy*, *redundancy*, do not solve all our problems.

What can be done to inject a note of moderation in this situation? In the first place, workers in other fields should realize that the basic results of the subject are aimed in a very specific direction, a direction that is not necessarily relevant to such fields as psychology, economics, and other social sciences. Indeed, the hard core of information theory is, essentially, a branch of mathematics, a strictly deductive system. A thorough understanding of the mathematical foundation and its communication application is surely a prerequisite to other applications. I personally believe that many of the concepts of information theory will prove useful in these other fields—and, indeed, some results are already quite promising—but the establishing of such applications is not a trivial matter of translating words to a new domain, but rather the slow tedious process of hypothesis and experimental verification. If, for example, the human being acts in some situations like an ideal decoder, this is an experimental and not a mathematical fact, and as such must be tested under a wide variety of experimental situations.

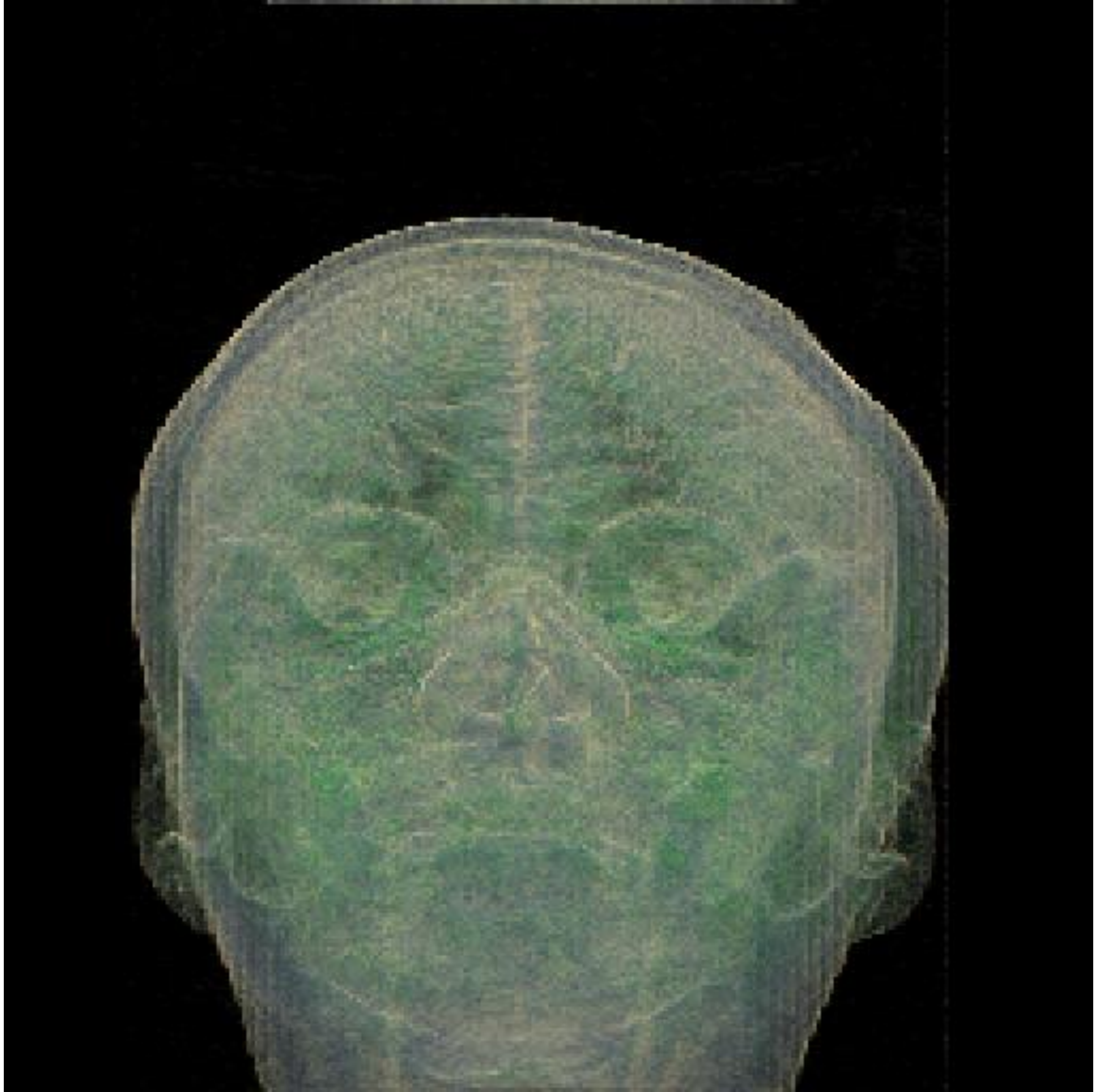
Secondly, we must keep our own house in first class order. The subject of information theory has certainly been sold, if not oversold. We should now turn our attention to the business of research and development at the highest scientific plane we can maintain. Research rather than exposition is the keynote, and our critical thresholds should be raised. Authors should submit only their best efforts, and these only after careful criticism by themselves and their colleagues. A few first rate research papers are preferable to a large number that are poorly conceived or half-finished. The latter are no credit to their writers and a waste of time to their readers. Only by maintaining a thoroughly scientific attitude can we achieve real progress in communication theory and consolidate our present position.

I propose that intelligence is the ability to do things humans do.

The first step to answering this question is to ask: what are humans doing? Humans have existed for about 200,000 years, and for all but about the last 5,000 of those years, humans have spent most of their time doing one thing: sitting around doing nothing. For the vast majority of human history, humans have done nothing but sit around, day after day, week after week, year after year.

So what does it mean to be intelligent? It means to be able to do nothing.

* https://arr.am/2020/07/31/human-intelligence-an-ai-op-ed/



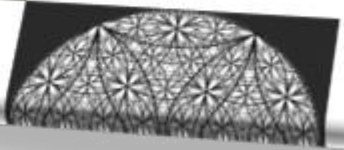
FORMAL THEORY OF FUN & CREATIVITY EXPLAINS SCIENCE, ART, MUSIC, HUMOR

Formal Theory of Creativity & Fun & Intrinsic Motivation
(1990-2010)
by Jürgen Schmidhuber

Since 1990 JS has built curious, creative agents that may be viewed as simple artificial scientists & artists with an intrinsic desire to explore the world by continually inventing new experiments. They never

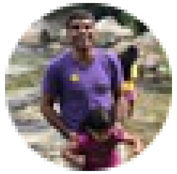
Left: JS giving a talk on creativity theory & art & science & humor at the Singularity Summit 2009 in New York City. **Videos: 10min** (excerpts at YouTube), 40min (original at Vimeo), 20min. JS' theory was also subject of a TV documentary (BR "Faszination Wissen", 29 May 2008; several repeats on other channels). Compare H+ interview and slashdot article.

How the Theory Explains Art. Artists (and observers of art) get rewarded for making (and observing) **novel patterns**: data that is **neither**



LOW-COMPLEXITY ART:
FEMME FRACTALE
J. SCHMIDHUBER
1997-2010

I argue that data becomes temporarily interesting by itself to some self-improving, but computationally limited, subjective observer once he learns to predict or compress the data in a better way, thus making it subjectively simpler and more beautiful. Curiosity is the desire to create or discover more non-random, non-arbitrary, regular data that is novel and surprising not in the traditional sense of Boltzmann and Shannon but in the sense that it allows for compression progress because its regularity was not yet known. This drive maximizes interestingness, the first derivative of subjective beauty or compressibility, that is, the steepness of the learning curve. It motivates exploring infants, pure mathematicians, composers, artists, dancers, comedians, yourself, and (since 1990) artificial systems.



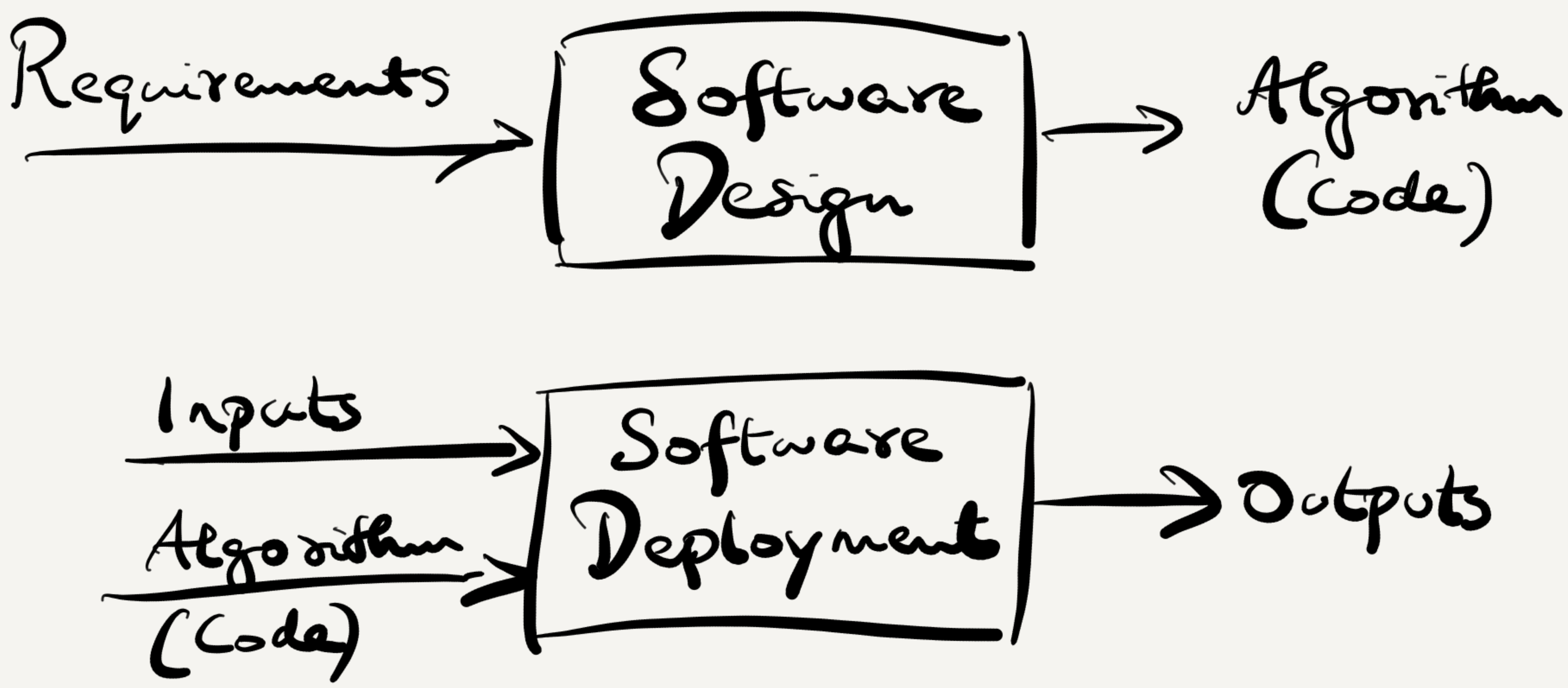
Madhu Shashanka
@ShashankaMadhu

To begin with, the field has no precise or clear definitions for “artificial” or “intelligence” but together it’s supposed to make sense!? It is good for marketing but not for an academic discipline.

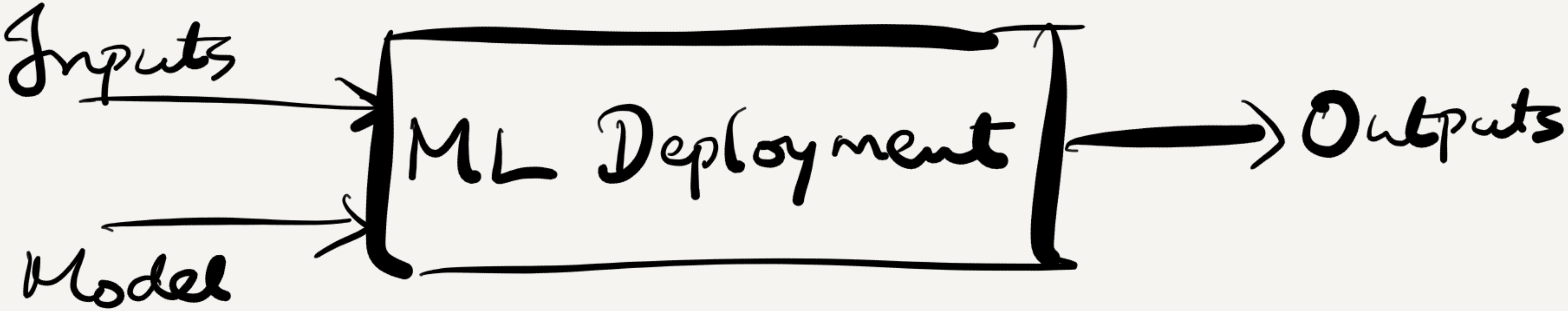
>  **Talia Ringer** 🌻🪬 @TaliaRinger · Sep 5
>
> I often think that AI would be a way better field if it abandoned its name and went with something like "automation" or whatever
>
> Show this thread

Yogi Jaeger
@yoginho

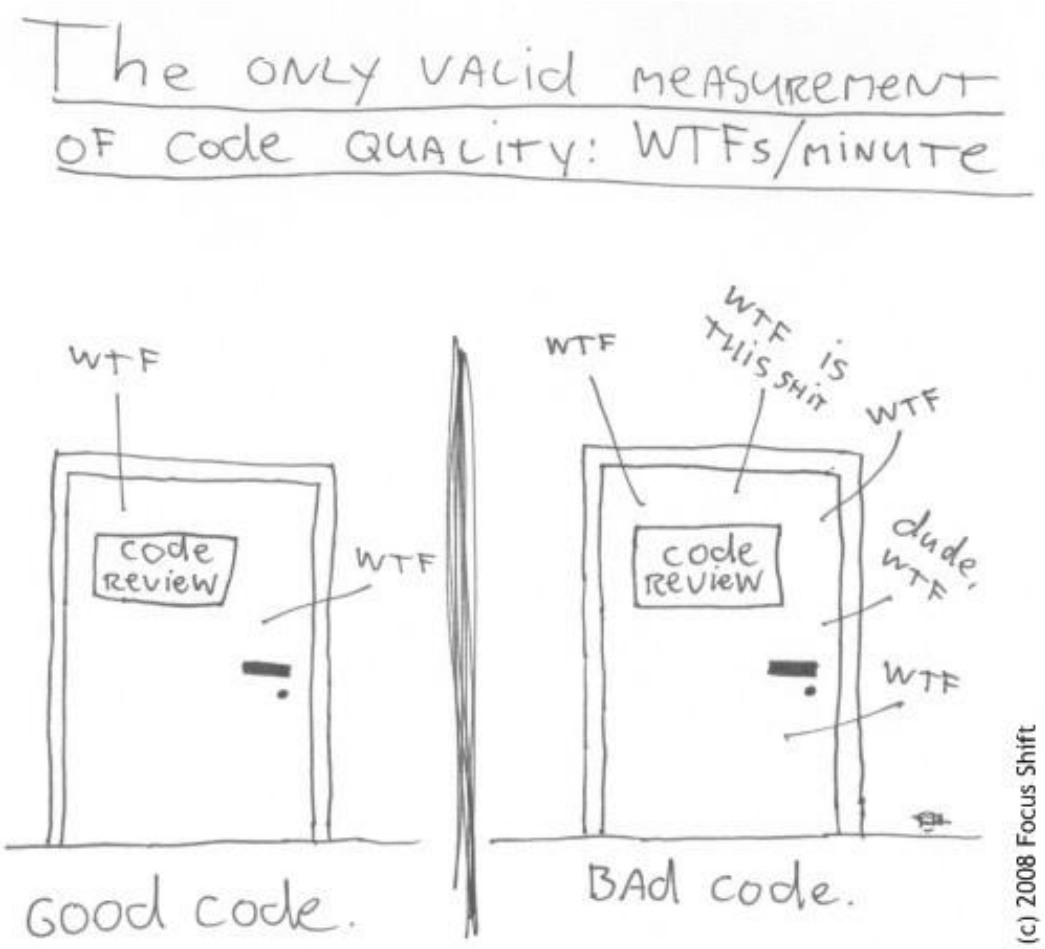
Why not be honest and call "artificial intelligence" (#AI) "algorithmic mimicry" (AM) instead? It's been a misnomer (and AI researchers always knew that) from the very beginning. It would help us so much to properly understand & contextualize its capabilities & limitations.



SOFTWARE AUTOMATION



ML/AI AUTOMATION



In all seriousness, I think this is actually a really good measurement of code quality. It's basically an indicator of how well something conforms to the principle of least surprise.


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS

* xkcd

Automation needs predictability

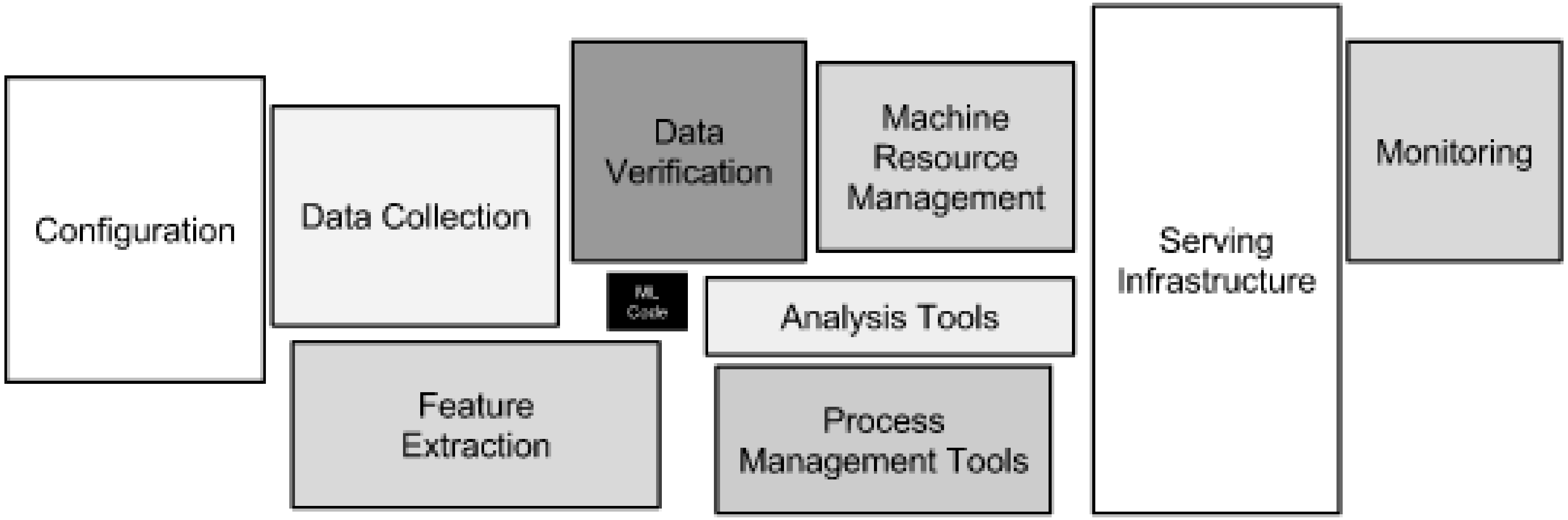


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

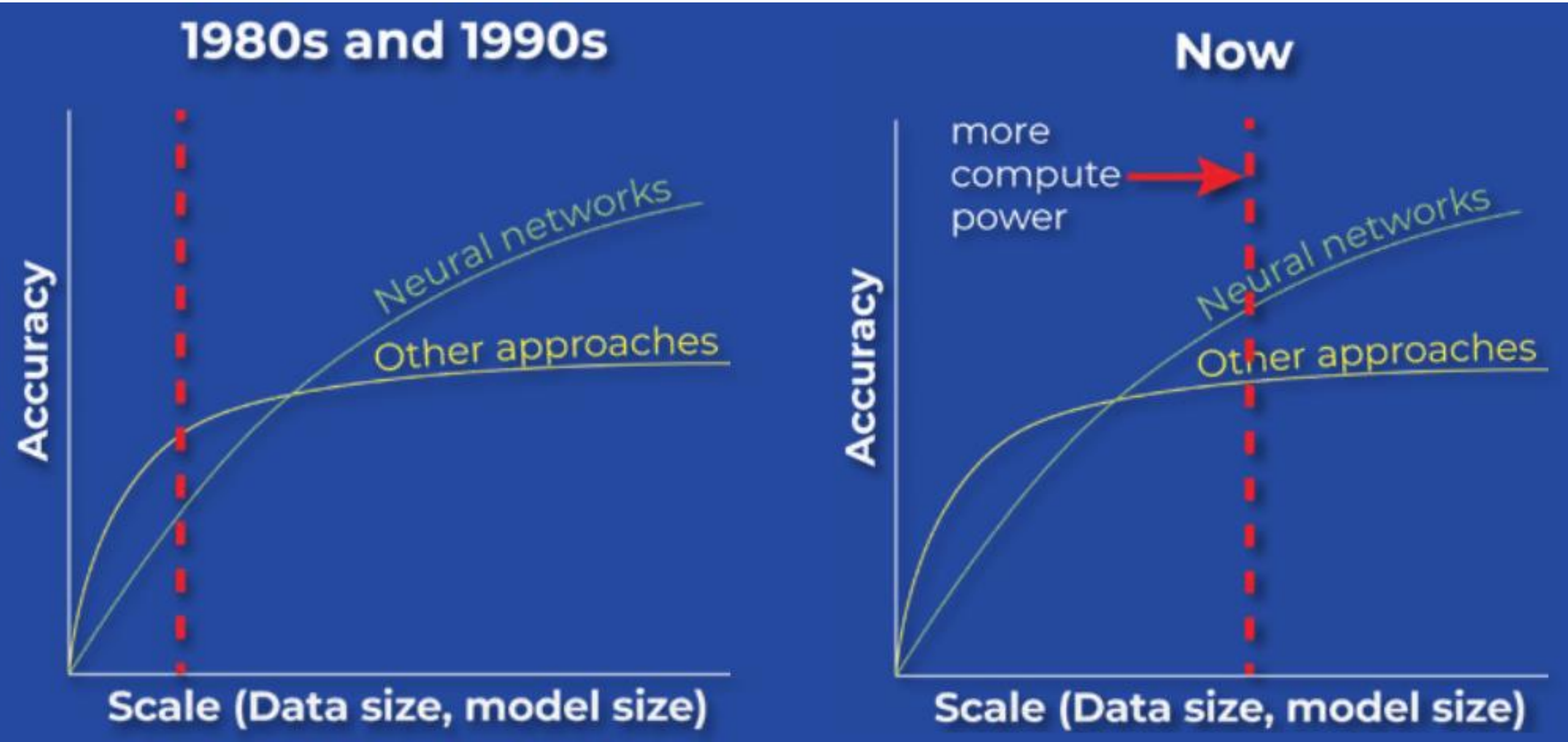
* Hidden Technical Debt in Machine Learning Systems, 2015

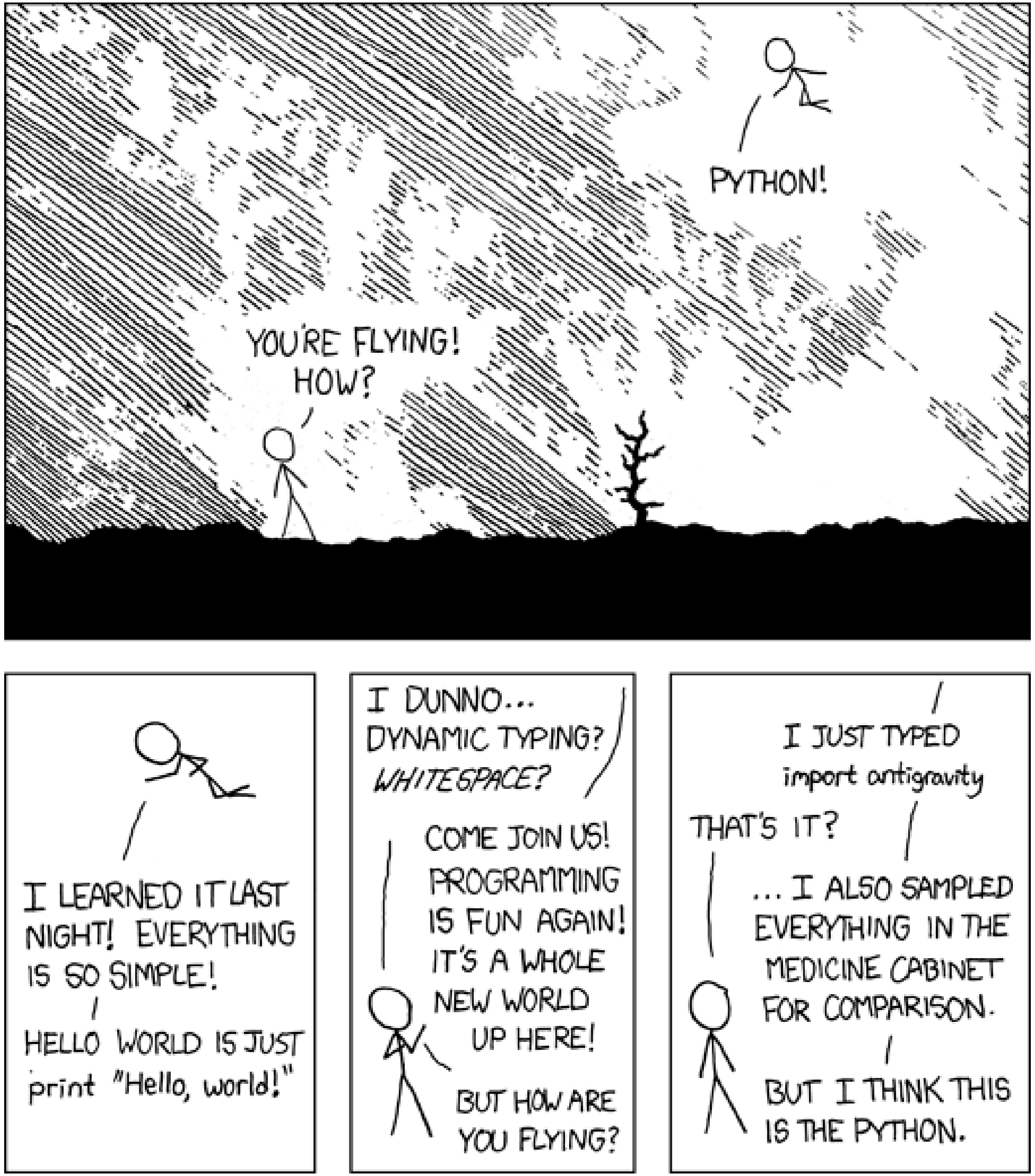
Deep learning (DL) training algorithms utilize nondeterminism to improve models' accuracy and training efficiency. Hence, multiple identical training runs (e.g., identical training data, algorithm, and network) produce different models with different accuracies and training times. In addition to these algorithmic factors, DL libraries (e.g., TensorFlow and cuDNN) introduce additional variance (referred to as implementation-level variance) due to parallelism, optimization, and floating-point computation.

This work is the first to study the variance of DL systems and the awareness of this variance among researchers and practitioners. Our experiments on three datasets with six popular networks show large overall accuracy differences among identical training runs. Even after excluding weak models, the accuracy difference is 10.8%. In addition, implementation-level factors alone cause the accuracy difference across identical training runs to be up to 2.9%, the per-class accuracy difference to be up to 52.4%, and the training time difference to be up to 145.3%. All core libraries (TensorFlow, CNTK, and Theano) and low-level libraries (e.g., cuDNN) exhibit implementation-level variance across all evaluated versions.

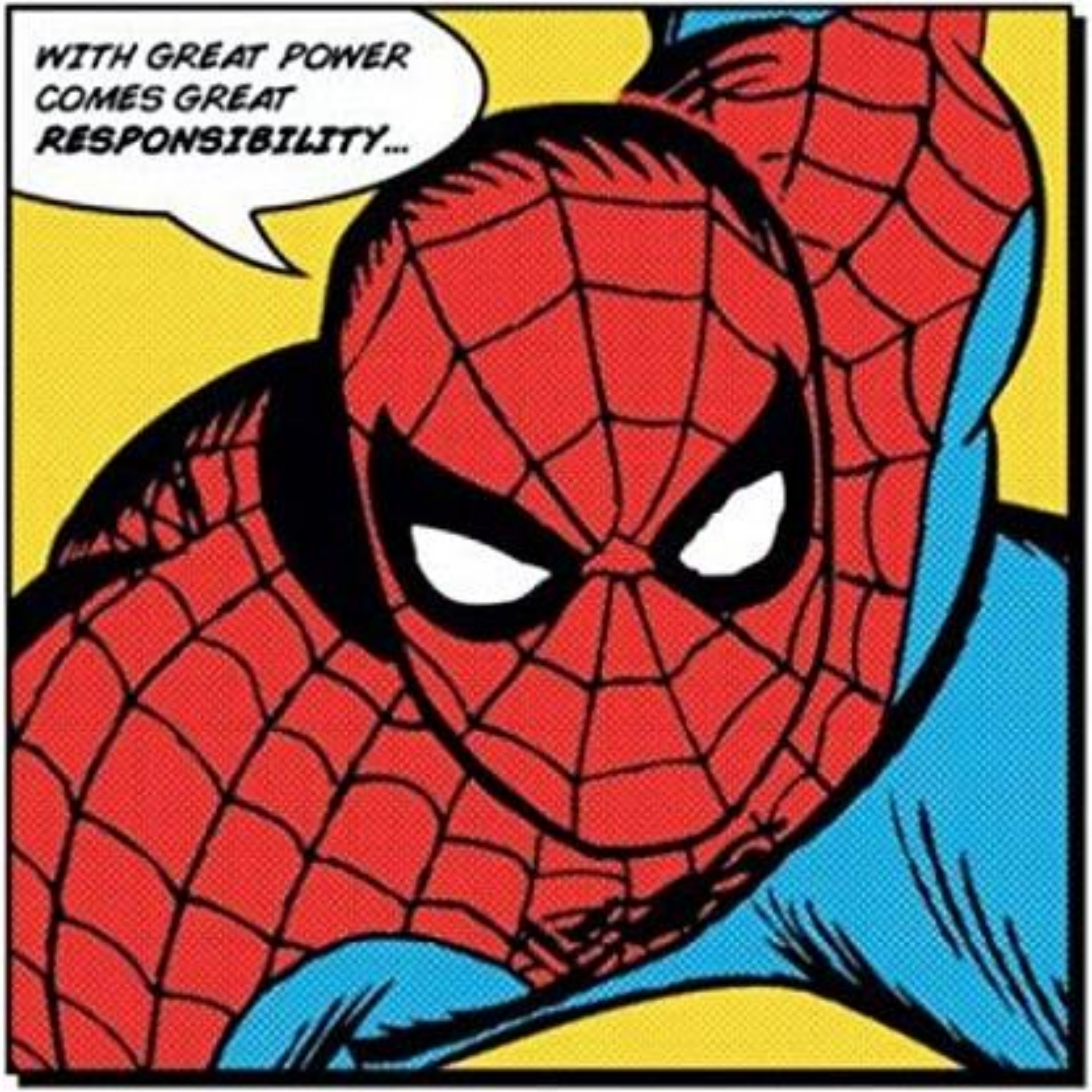
Our researcher and practitioner survey shows that 83.8% of the

* Problems and Opportunities in Training Deep Learning Software Systems: An Analysis of Variance, 2020


1980s and 1990s
Now
more
compute
power
Neural networks
Other approaches
Accuracy
Scale (Data size, model size)
Neural networks
Other approaches
Accuracy
Scale (Data size, model size)



PYTHON!
YOU'RE FLYING! HOW?
I LEARNED IT LAST NIGHT! EVERYTHING IS SO SIMPLE!
HELLO WORLD IS JUST print "Hello, world!"
I DUNNO... DYNAMIC TYPING? WHITESPACE?
COME JOIN US! PROGRAMMING IS FUN AGAIN! IT'S A WHOLE NEW WORLD UP HERE!
BUT HOW ARE YOU FLYING?
I JUST TYPED import antigravity
THAT'S IT?
... I ALSO SAMPLED EVERYTHING IN THE MEDICINE CABINET FOR COMPARISON.
BUT I THINK THIS IS THE PYTHON.


WITH GREAT POWER COMES GREAT RESPONSIBILITY...

Madhu Shashanka
@ShashankaMadhu

Couldn't agree more. But the key is empathy, what's simple for you maybe too complex and hard to use for most other people. You need to know what simple means for the people you are building for.

> **François Chollet** @fchollet · Apr 20, 2021
> Technologists shouldn't just make things possible. They should make them simple. In many ways it's far more difficult. To something possible you just need to be clever. To make it simple you need vision. Intelligence is common, vision is rare.
> Show this thread

Madhu Shashanka
@ShashankaMadhu

Defining precise use-cases and having a clear understanding of the value of solving those use-cases are necessary for real-world success. This is where domain expertise comes into play. Machine learning / deep learning hammers alone will not give you that.

The importance of stupidity in scientific research

Martin A. Schwartz

Department of Microbiology, UVA Health System, University of Virginia, Charlottesville, VA 22908, USA
e-mail: maschwartz@virginia.edu

Accepted 9 April 2008
Journal of Cell Science 121, 1771 Published by The Company of Biologists 2008
doi:10.1242/jcs.033340

I recently saw an old friend for the first time in many years. We had been Ph.D. students at the same time, both studying science, although in different areas. She later dropped out of graduate school, went to Harvard Law School and is now a senior lawyer for a major environmental organization. At some point, the conversation turned to why she had left graduate school. To my utter astonishment, she said it was because it made her feel stupid. After a couple of years of feeling stupid every day, she was ready to do something else.

I had thought of her as one of the brightest people I knew and her subsequent career supports that view. What she said bothered me. I kept thinking about it; sometime the next day, it hit me. Science makes me feel stupid too. It's just that I've gotten used to it. So used to it, in fact, that I actively seek out new opportunities to feel

I'd like to suggest that our Ph.D. programs often do students a disservice in two ways. First, I don't think students are made to understand how hard it is to do research. And how very, very hard it is to do important research. It's a lot harder than taking even very demanding courses. What makes it difficult is that research is immersion in the unknown. We just don't know what we're doing. We can't be sure whether we're asking the right question or doing the right experiment until we get the answer or the result. Admittedly, science is made harder by competition for grants and space in top journals. But apart from all of that, doing significant research is intrinsically hard and changing departmental, institutional or national policies will not succeed in lessening its intrinsic difficulty.

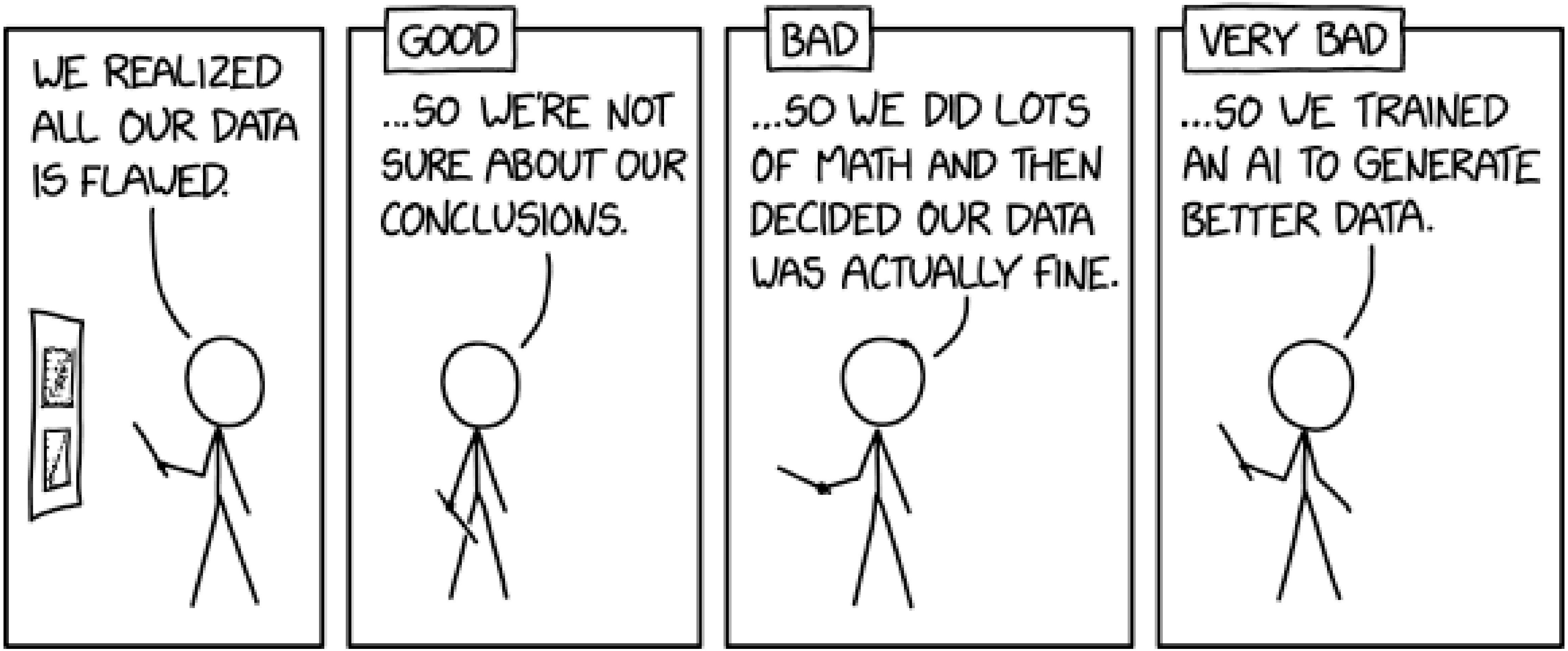
Guiding principles

While AI expertise is a key strength and differentiator for Concentric, what sets us apart is that data security is in our DNA.

Unlike newcomers in the field who see every challenge as a nail for the proverbial AI hammer, our team has decades of lived experience confronting and overcoming real-world cybersecurity challenges.

We've been thinking about the data security problem for quite some time, long before Concentric's founding back in 2018.

1. **First-principles thinking**. Focus on solving the challenges at hand, ignoring prevailing conventional wisdom of what's possible or not while avoiding unnecessary complexity. It starts with truly understanding the real pain-points of practitioners and identifying where you can add tangible value.
2. **Relentless focus on precision**. Any product that produces novel insights must confront the issue of false-positives. We had a strict requirement that anything we built kept false positives to a minimum (also known as "high-precision"). This is critical in a field like cybersecurity, which is chronically under-staffed. Practitioners cannot afford to chase after alerts that turn out to be benign.
3. **Enterprise-grade scale**. Whatever we built had to meet enterprise-level scale and prove to be bullet-proof enough to support enterprise environments. A corollary is that the technology stack chosen should be amenable to being operationalized in a cost-sustainable manner. There is a history of AI startups ignoring this at their peril, releasing seemingly impressive demos only to eventually fold at large data volume thresholds.


WE REALIZED ALL OUR DATA IS FLAWED.
GOOD
...SO WE'RE NOT SURE ABOUT OUR CONCLUSIONS.
BAD
...SO WE DID LOTS OF MATH AND THEN DECIDED OUR DATA WAS ACTUALLY FINE.
VERY BAD
...SO WE TRAINED AN AI TO GENERATE BETTER DATA.

* xkcd

160k+ high school students will only graduate if a statistical model allows them to

Jun 17, 2020

tl;dr *: Due to curriculum disruptions the International Baccalaureate (IB) is going to use a statistical model to assign grades for > 160,000 high school students. These grades have a very significant impact on students' lives (for better or for worse). This is an inappropriate use of 'big data' and a terrible idea for a plethora of different reasons.*

Alan Mackworth
@AlanMackworth

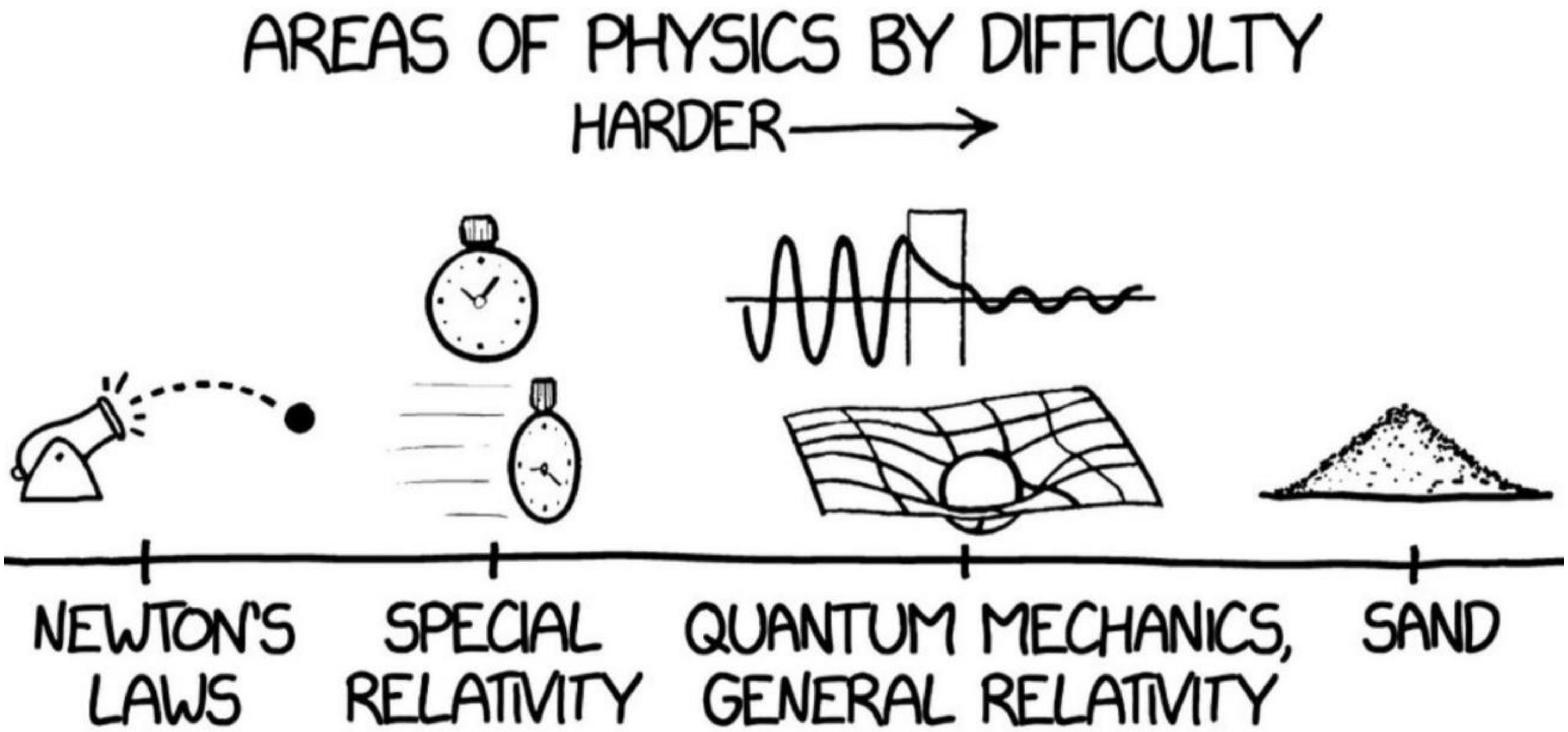
"It's completely obvious that within 5 years deep learning will do better than radiologists", Geoff Hinton (2016) youtu.be/2HMPRXstSvQ

> **Alan Mackworth** @AlanMackworth · Mar 22, 2021
> "none of the models identified are of potential clinical use due to methodological flaws and/or underlying biases." t.co/SRhLG21PD9

Madhu Shashanka
@ShashankaMadhu

perfect use-cases for automation via ML are those where the cost of making mistakes are low

> **Andrew Whitby** @EconAndrew · Jan 26
>
> Maybe machine learning in radiology is on the wrong track.
>
> Rather than trying to replace radiologists in high-stakes work like diagnosing cancer, maybe it should target areas where radiology would currently be unaffordable, like routine ultrasound for minor soft tissue injuries.
>
> Show this thread


AREAS OF PHYSICS BY DIFFICULTY
HARDER
NEWTON'S LAWS
SPECIAL RELATIVITY
QUANTUM MECHANICS, GENERAL RELATIVITY
SAND

* xkcd

hardmaru
@hardmaru

Machines see objects

Humans see ideology