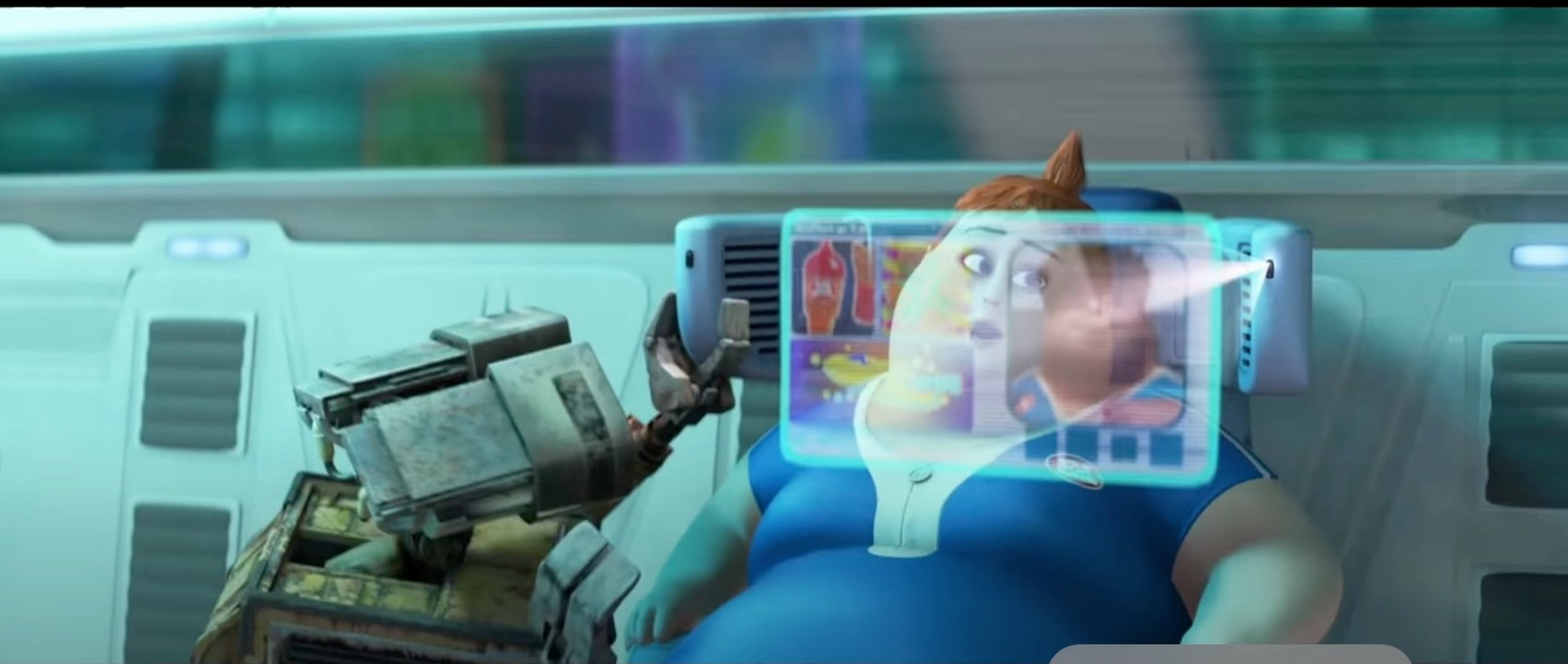


Sound check

13 THEY LIVE 23



0:31 / 2:05

⏪ 🔊 🔍 🗄️

Machine Learning and the Future of Particle Physics

DKPI Final Event
September 28, 2023

Matthew Schwartz
Harvard University



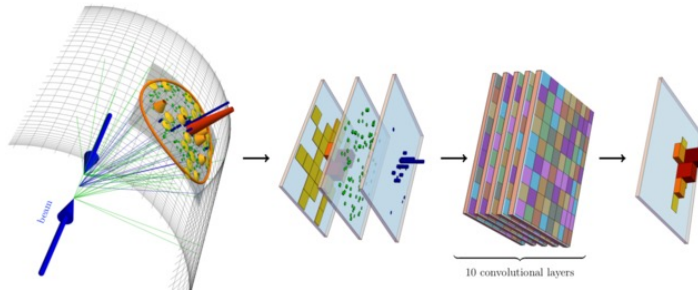
The NSF AI Institute for Artificial Intelligence and Fundamental Interactions (IAIFI)



Outline

Part 1: The present

- Machine learning in particle physics



Part 2: The future

- The revolution has arrived

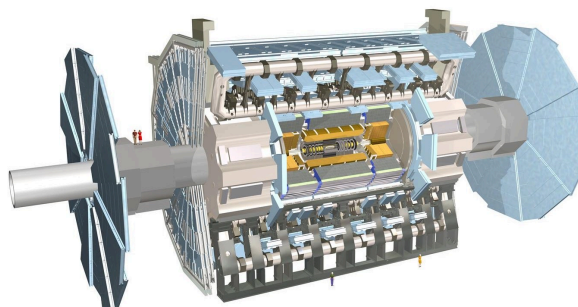


Particle physics is perfect for ML

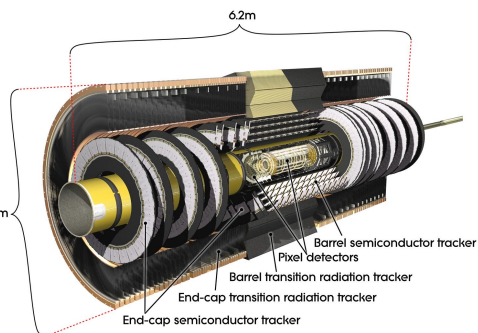


Large Hadron Collider 10^4 m

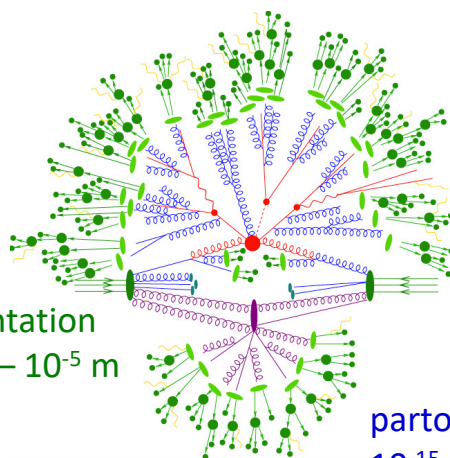
ATLAS detector 100 m



calorimeters 0.1 m – 10m

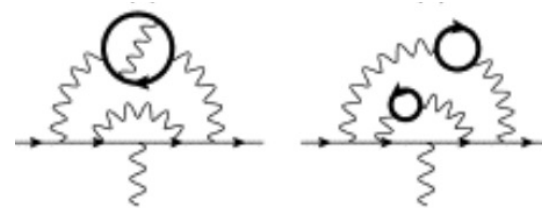


tracker 10^{-5} m – 10^{-1} m



fragmentation
 10^{-10} m – 10^{-5} m

parton shower
 10^{-15} m – 10^{-10} m



hard scattering
 10^{-20} m – 10^{-15} m

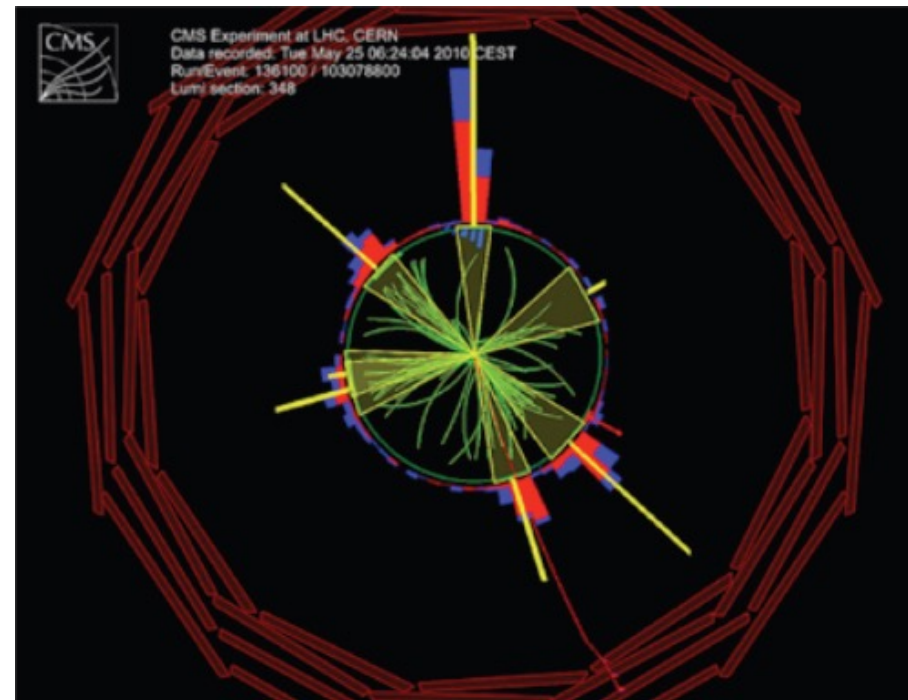
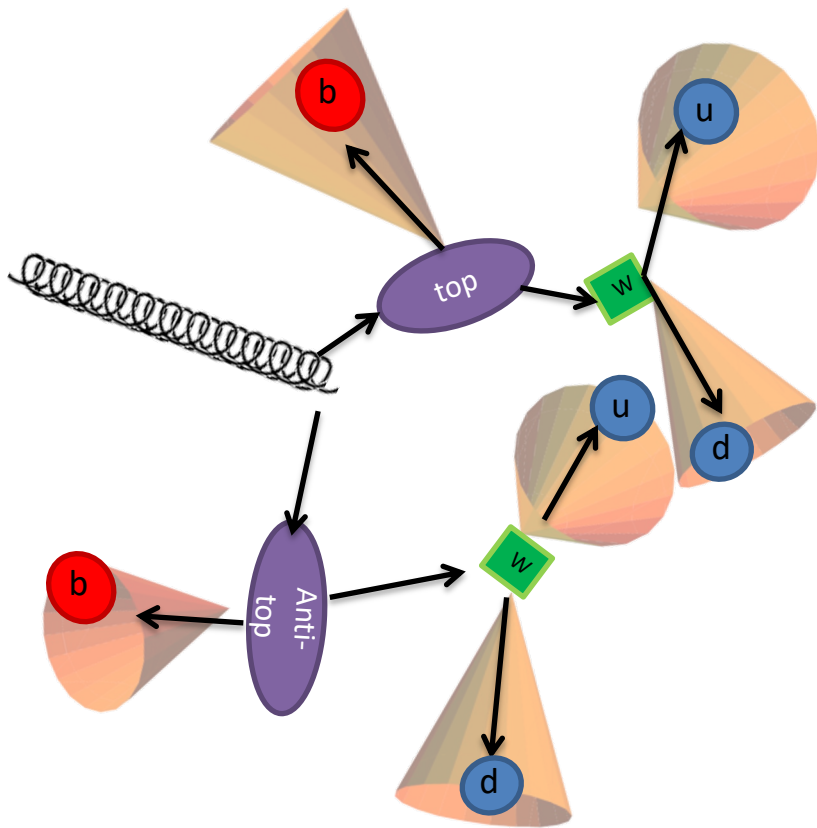
Simulations accurate over **24 orders of magnitude!**

Unparalleled environment for supervised ML

Ex. #1. Top quarks at the LHC

Top quark (mass = 170 GeV)

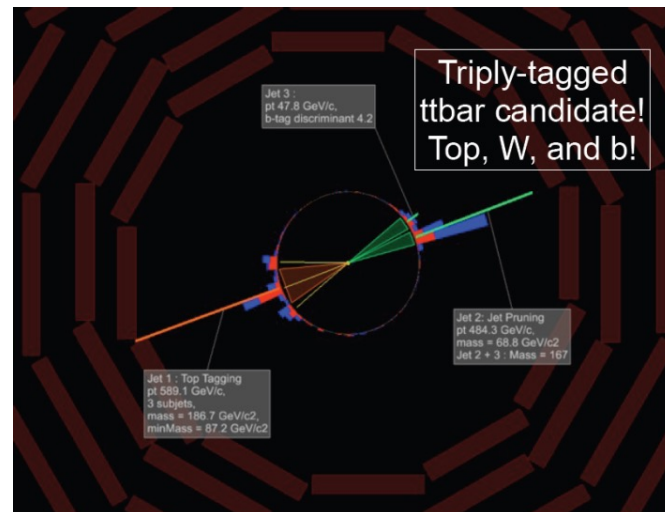
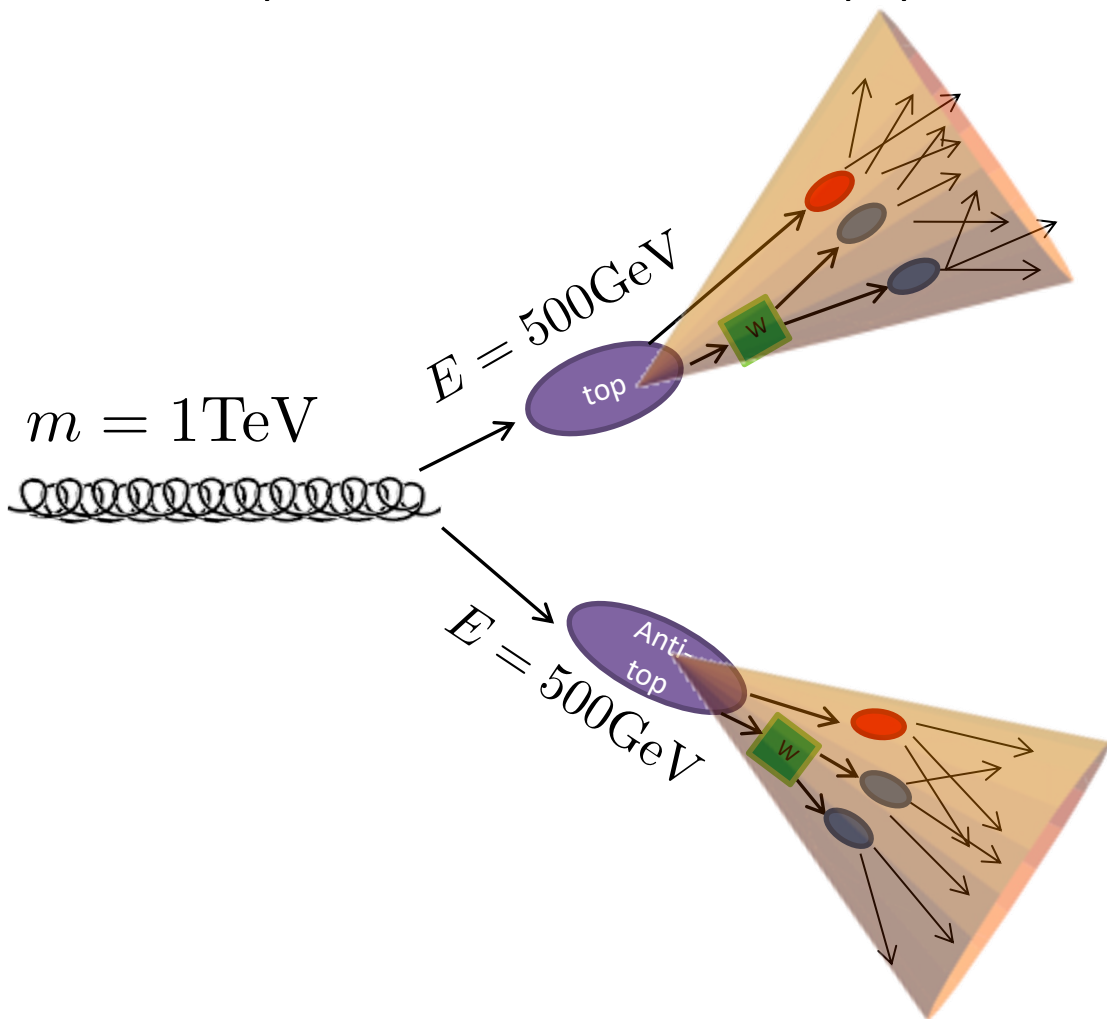
- Decays to $W+b$, then W decays to u and d quarks (mass = 0.010 GeV)
- Quarks look like “jets” at colliders



6 jet $t\bar{t}$ event

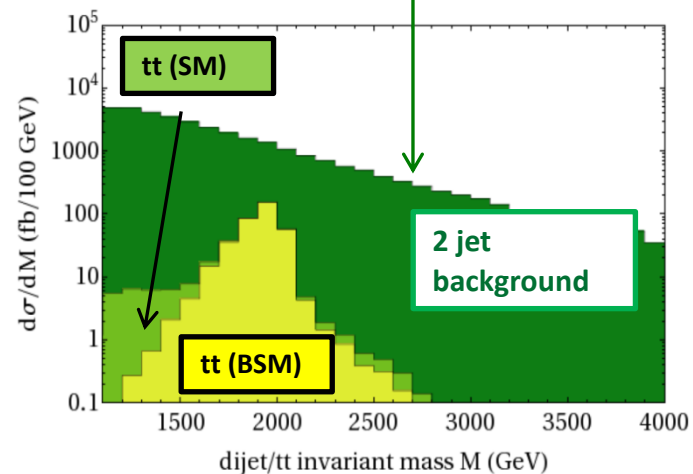
Boosted top quarks

- Models of beyond-the-standard model physics with very heavy new particles often produce **ultrarelativistic** top quarks



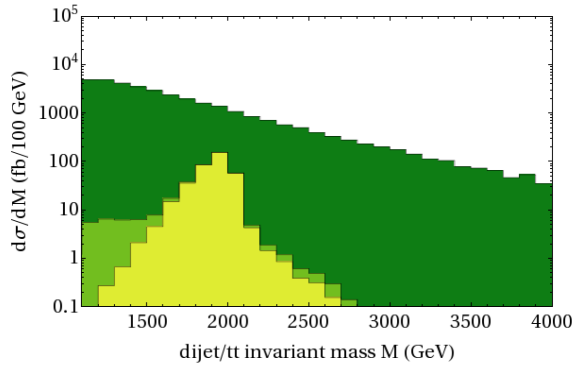
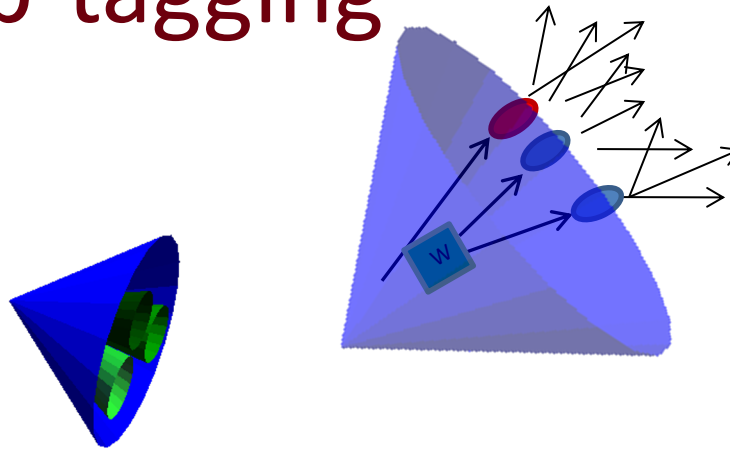
Backgrounds are now 2 jets not 6 jets

- 10^4 times bigger than signal

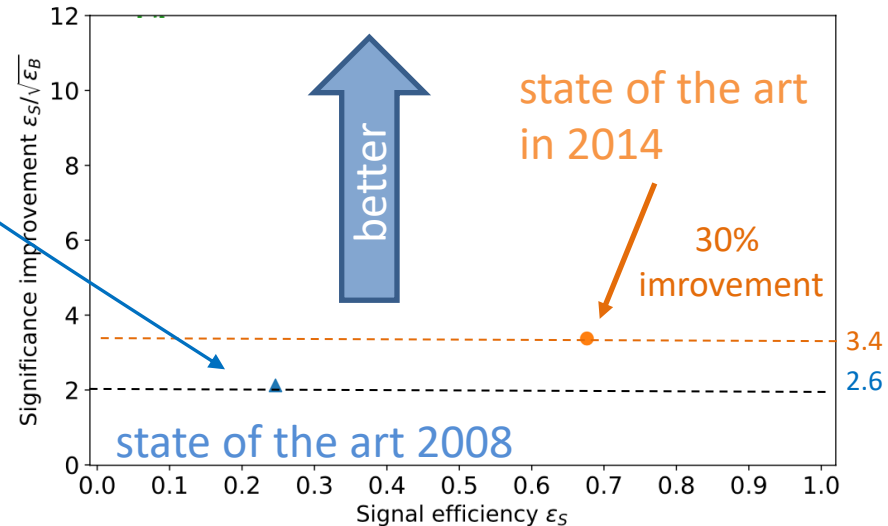
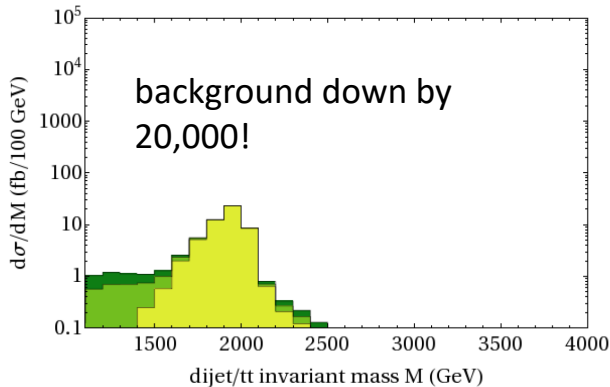


Conventional top-tagging

1. Look for big jets ($R = 1.2$)
2. with subjets within the jet
3. Analyze the subjets
 - look for W boson, W/top mass peak, helicity angle, etc.



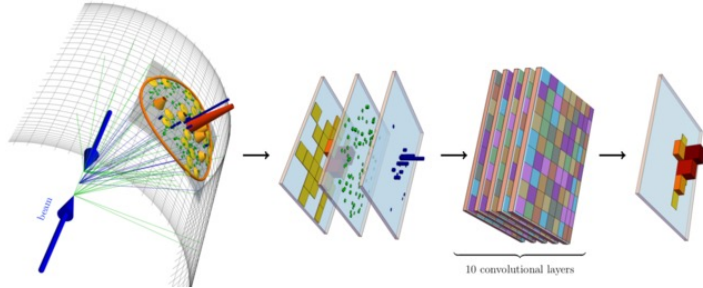
signal efficiency = 0.4
background efficiency = 0.006



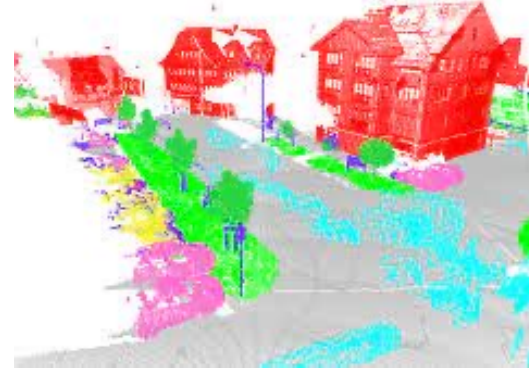
Machine learning approach

Take some tool highly engineered for another purpose and shoehorn it into physics

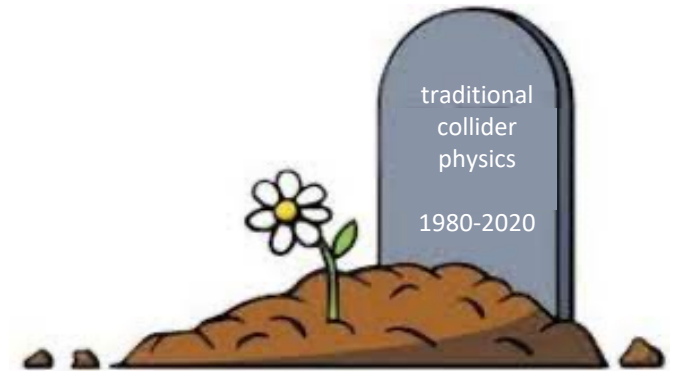
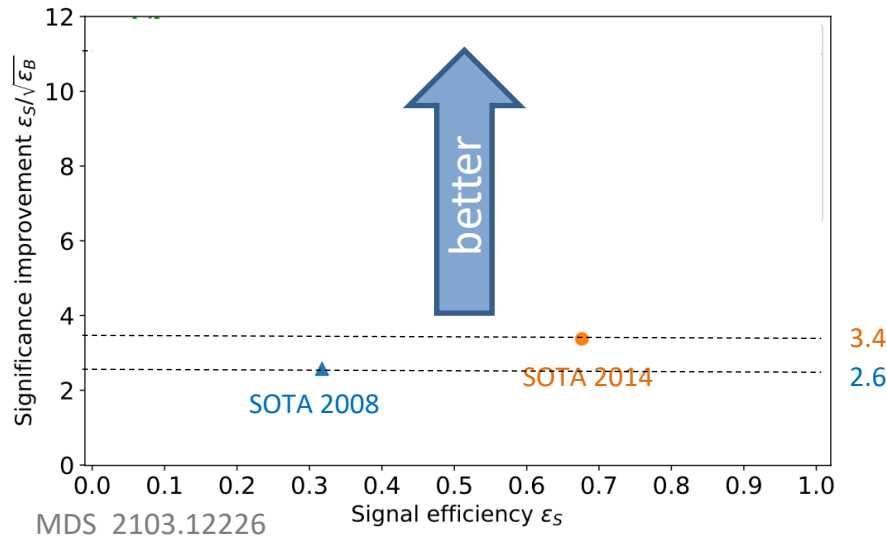
Convolutional networks for facial recognition



Point clouds/deep sets for self-driving cars

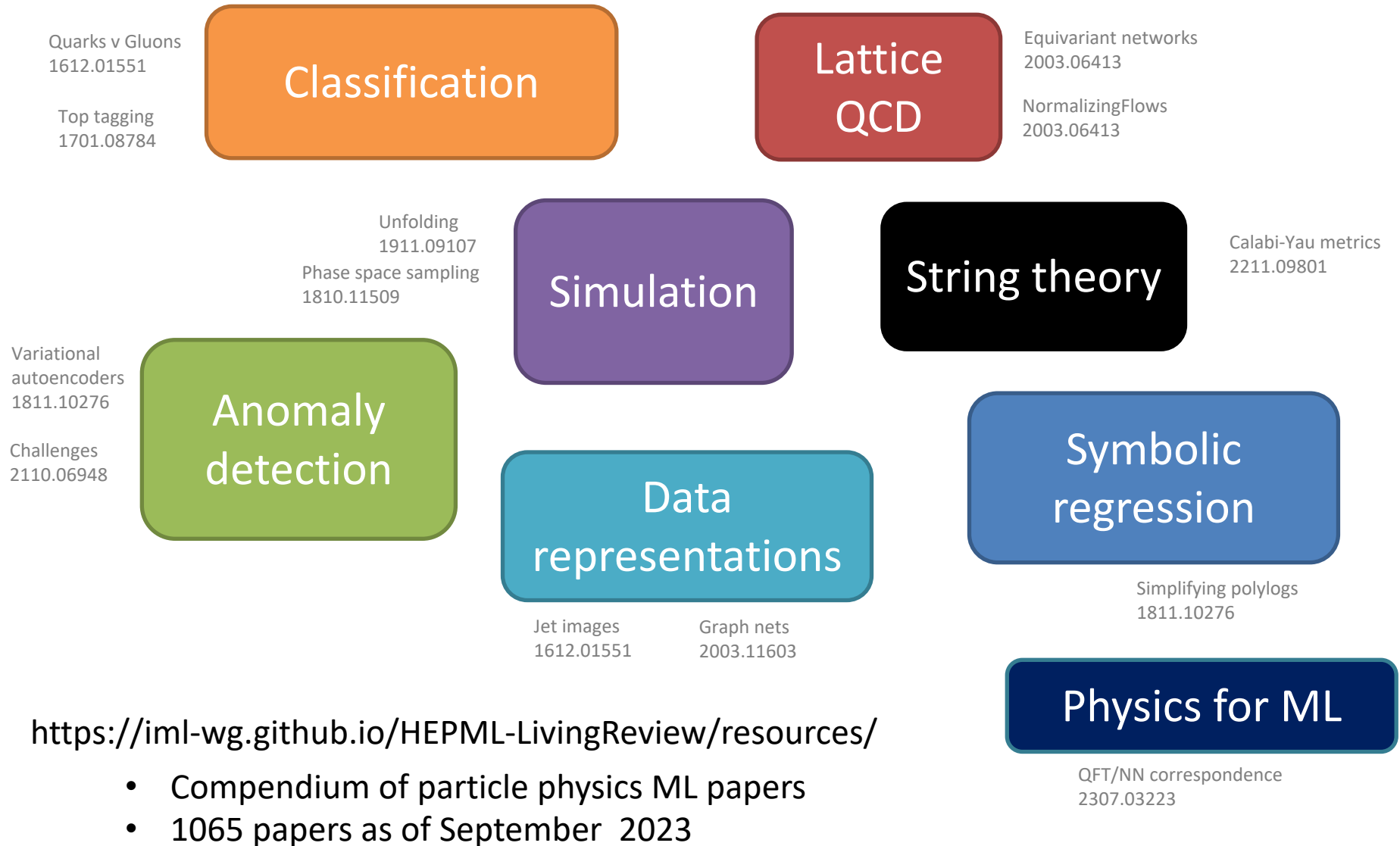


Machine learning methods are much better



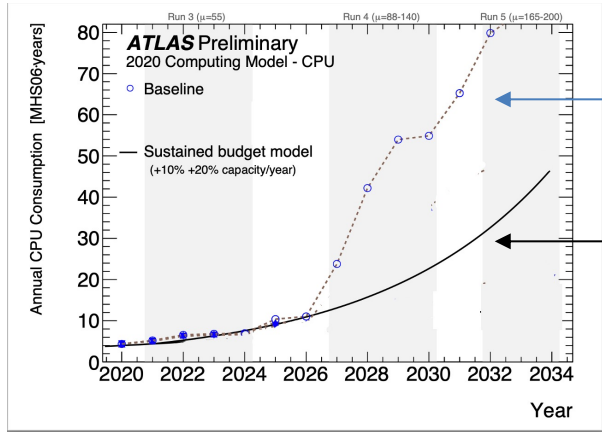
- ML requires less "thinking"
- Better performance
- Provides less physical insight

ML has touched all areas of particle physics



Ex. 2. Simulation/Unfolding

- Simulating events is accurate but slow
 - Full simulation (GEANT) can take **minutes per event!**
 - **Without improvements, we will not be able to keep up** with needs of LHC



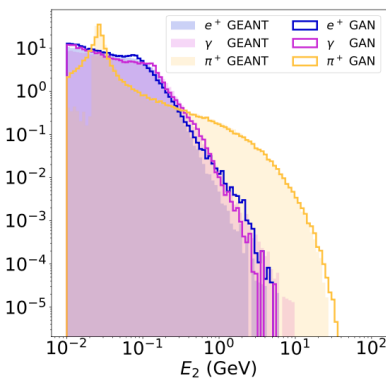
computing needs

what CERN can afford



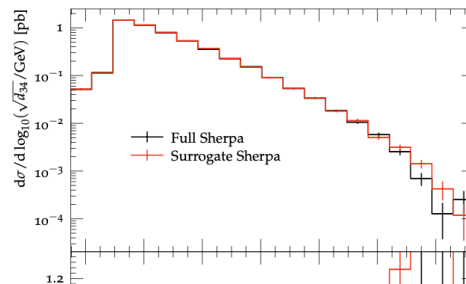
- We can use machine learning to improve the simulations

Paganini et al 1712.10312
CaloGAN: generative adversarial networks



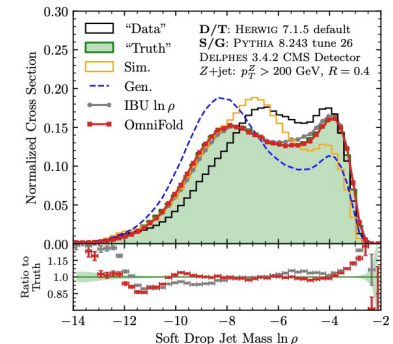
100,000 fold speed increase
for electromagnetic calorimeter simulation

Danziger et al 2109.11964
Unweight surrogate of differential cross section



600 fold speed increase
for hard matrix elements

Andreassen et al. 1911.09107
OMNIfold: learn to unfold from
calorimeter data to particles



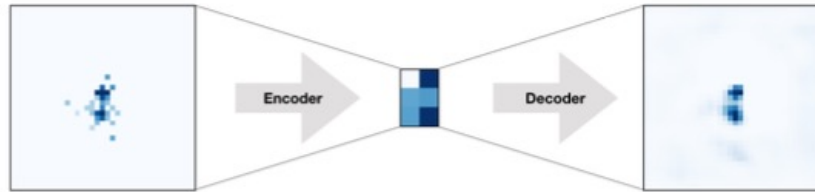
Avoids time-consuming
observable-by-observable approach

Ex. 3. Anomaly detection

Variational autoencoders:

[Farina,Nakai,Shih 1808.08992]

- Compress the background/data to a low-dimensional latent space

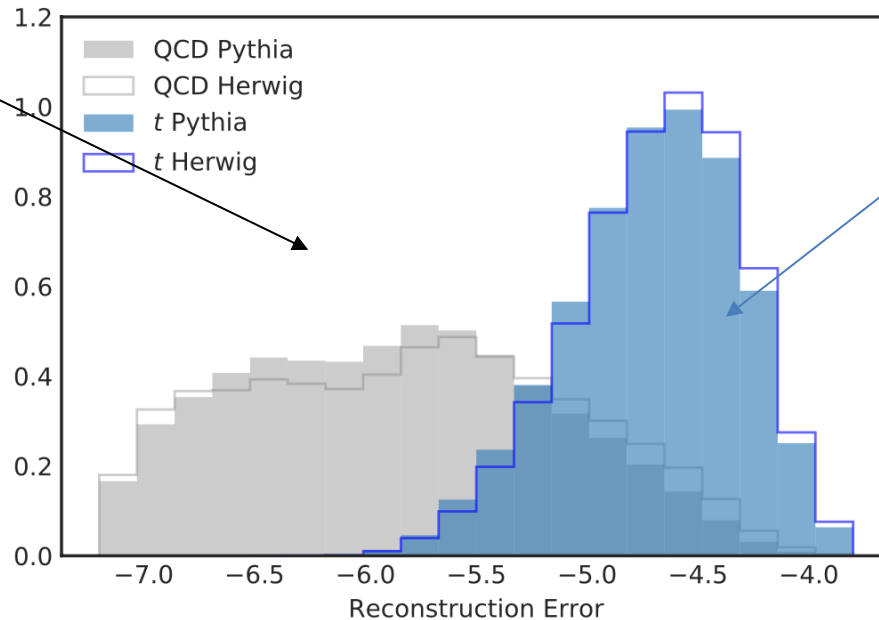


- Uncompress back to physical space.
- Events with large reconstruction error are anomalies

e.g.

Train to reproduce

backgrounds



small reconstruction error

Anything else has larger reconstruction error

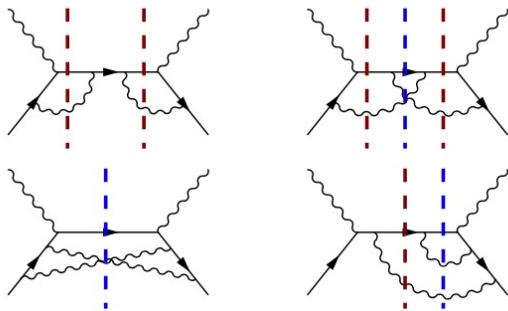
(top quarks in this case)

Potential to find unknown unknowns

Ex. 4. Computing Feynman diagrams

Computing Feynman diagrams often gives long expressions

Dersy, MDS, Zhang arXiv:2206.04115

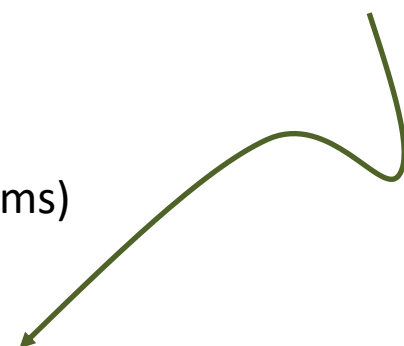


$$\begin{aligned}
 f(x) = & 9 \left(-\text{Li}_3(x) - \text{Li}_3\left(\frac{2ix}{-i + \sqrt{3}}\right) - \text{Li}_3\left(-\frac{2ix}{i + \sqrt{3}}\right) \right) \\
 & + 4 \left(-\text{Li}_3(x) + \text{Li}_3\left(\frac{x}{x+1}\right) + \text{Li}_3(x+1) - \text{Li}_2(-x) \ln(x+1) \right) \\
 & - 4 \left(\text{Li}_2(x+1) \ln(x+1) + \frac{1}{6} \ln^3(x+1) + \frac{1}{2} \ln(-x) \ln^2(x+1) \right)
 \end{aligned}$$

- Expressions can be **very complicated** (hundreds of terms)
- Expression often has a dramatically simpler form

$$f(x) = -\text{Li}_3(x^3) - \text{Li}_3(x^2) + 4\zeta_3$$

- **No known algorithmic way to simplify** the initial forms
- **Simple form is useful, not just pretty**
 - No spurious singularities/branch points
 - Clarifies the analytic structure of the scattering amplitude



Simplification of Polylogarithms




Two machine learning approaches

1. Reinforcement learning

- Used by Alpha Zero – world’s best chess program

Basic idea:

- apply known polylog identities like moves in a game
- train by learning to reverse scrambling steps

Reflection 	Inversion 	Duplication 
$-7\text{Li}_2(x) \rightarrow 7\text{Li}_2(1-x)$	$-7\text{Li}_2(x) \rightarrow 7\text{Li}_2\left(\frac{1}{x}\right)$	$-7\text{Li}_2(x) \rightarrow 7\text{Li}_2(-x) - \frac{7}{2}\text{Li}_2(x^2)$



2. Transformer networks

- Used by large language models
- Learn to “guess” answer
translate from complicated to simple

German: Unkrautvernichtungsmittel

$$f(x) = 9 \left(-\text{Li}_3(x) - \text{Li}_3\left(\frac{2ix}{-i + \sqrt{3}}\right) - \text{Li}_3\left(-\frac{2ix}{i + \sqrt{3}}\right) \right) \\ + 4 \left(-\text{Li}_3(x) + \text{Li}_3\left(\frac{x}{x+1}\right) + \text{Li}_3(x+1) - \text{Li}_2(-x) \ln(x+1) \right) \\ - 4 \left(\text{Li}_2(x+1) \ln(x+1) + \frac{1}{6} \ln^3(x+1) + \frac{1}{2} \ln(-x) \ln^2(x+1) \right)$$

translate

English: Herbicide

$$f(x) = -\text{Li}_3(x^3) - \text{Li}_3(x^2) + 4\zeta_3$$

Results

- Both methods work well (>80% success)
- Transforms do better: **91% success** up to transcendental weight 4

B. The future



“Retired Mario”

Markus Graf, MidJourney July, 2023

Photo or AI?



Large Language Models

What is a large language model?

- Designed to generate human-like text
- Usually transformer-autoencoder deep neural networks with attention:

Attention is all you need

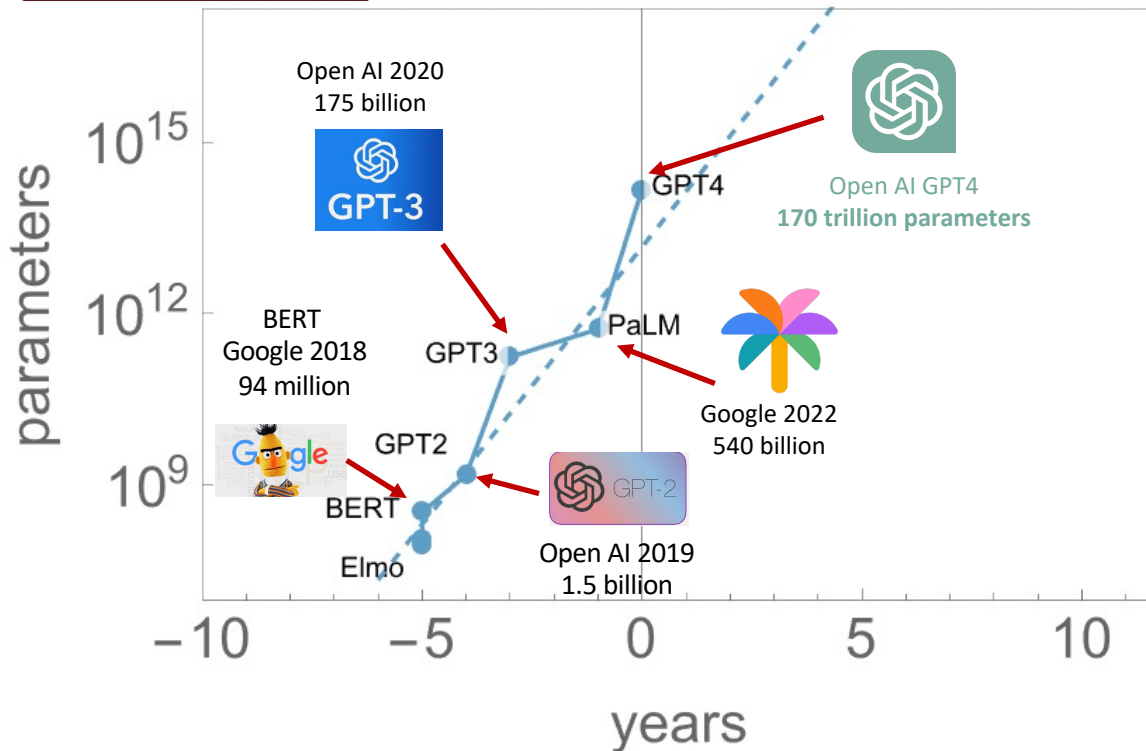
[A Vaswani, N Shazeer, N Parmar...](#) - Advances in neural ..., 2017 - proceedings.neurips.cc

... to attend to **all** positions in the decoder up to and including that position. **We need** to prevent ... **We** implement this **inside** of scaled dot-product **attention** by masking out (setting to $-\infty$) ...

☆ Save 📄 Cite **Cited by 87936** Related articles All 62 versions ⇨

- All they do is predict the next word

History of LLMs



Mammalian brains

Cat brain



0.760 billion neurons
10 trillion synapses

Human brain



80 billion neurons
150 trillion synapses

size of GPT 3.5

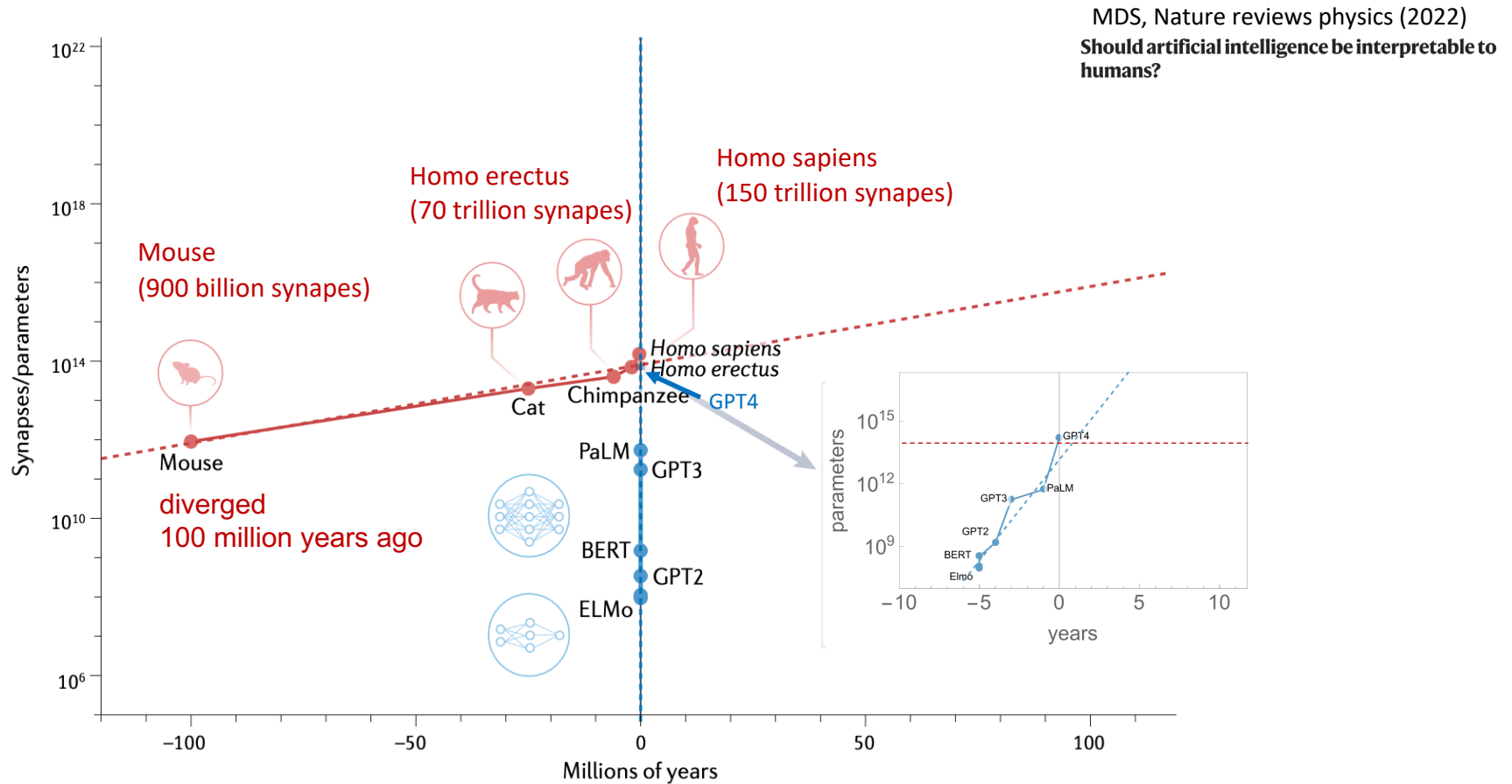


size of GPT 4



Machine vs. Biological intelligence

- Biological intelligence grows by a factor of 2 in one million years
- Machine intelligence grows by a factor of 10 in 1 year



- Both AI and biological intelligence grow exponentially
- **Factor of 10^6 difference in exponent**
- Intersection, when machines and biology have comparable "intelligence" is 2023

Concerns about the future of language models

1. Quality

- Facts are “hallucinated”
- Machines are just memorizing and parroting data

2. Atrophy

- We will lose the ability to think and to write for ourselves
- Students/adults will use LLMs to cheat

3. Scalability

- The exponential growth cannot continue
- Training data, energy, materials are limited

4. Interpretability

- Humans must be able to understand the machines
- Intellectual pursuit is a fundamentally human endeavor

5. Controllability

- AI cannot be controlled
- Machines will take over the world and kill us all

1. Quality: Is GPT 4 garbage?

No

“Einstein at a water park”



GPT4 can solve most undergraduate physics problems

MA Two blocks with equal mass m are connected with an inelastic string. One of the blocks is placed on a table, the other hangs over the edge. Determine the acceleration of the masses if the mass of the string is neglected and there is no friction.

Given:

- Two blocks with equal mass m .
- The string is inelastic (which means it does not stretch or change length).
- Mass of the string is neglected.
- No friction.

Let's consider the forces involved:

- The weight of the hanging block, $F_{gravity}$, which is equal to mg , where g is the acceleration due to gravity.
- The tension in the string, T .

When the system is released, the hanging block will start moving downward, pulling the block on the table along with it. Since there's no friction, the only opposing force on the hanging block is the tension T in the string.

Let's apply Newton's second law to both blocks:

For the block on the table:

There's no vertical motion, so we'll only consider the horizontal direction.

$$\Sigma F_x = T$$

$$ma = T$$

$$T = ma \quad \text{(Equation 1)}$$

For the block hanging:

$$\Sigma F_y = mg - T$$

$$ma = mg - T$$

$$T = mg - ma$$

$$T = m(g - a) \quad \text{(Equation 2)}$$

From Equations (1) and (2):

$$ma = m(g - a)$$

$$ma = mg - ma$$

$$2ma = mg$$

$$a = \frac{g}{2}$$

Thus, the acceleration of the two blocks will be half the acceleration due to gravity, $a = \frac{g}{2}$.

• explaining steps
• not just guessing

Correct!

$a = \frac{g}{2}$

MA Write code to generate a figure for this problem

To visualize the problem, we can use Python with the `matplotlib` library to draw the two blocks, the table, and the string connecting them. Here's a simple way to generate such a figure:

```
python
import matplotlib.pyplot as plt
# ... (code continues) ...
```

MA The string is going through the table. Make it hanging off the edge instead. And make the string not right next to the blue block

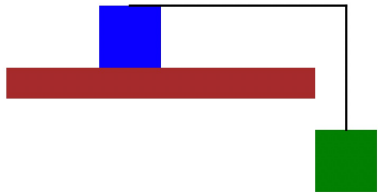
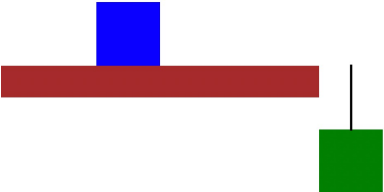
You're right; I apologize for that oversight. Let's correct the visualization to show the string hanging off the edge of the table and provide some space between the string and the block on the table.

Here's the corrected code:

MA that's better. But the upper part of the string is not showing

I apologize for the oversight. Let's ensure the upper part of the string connects the block on the table to the edge of the table.

Two blocks connected by an inelastic string



It can write new problems

MA

Write a homework problem about Vienna for a physics course

Topic: Kinematics and Dynamics in Two Dimensions

Background: Vienna is famous for its efficient public transportation system, especially its iconic trams. These trams often follow curved paths, making them an interesting subject for a physics problem involving two-dimensional motion.

Given:

- A tram in Vienna starts from rest at station A and accelerates uniformly along a straight path for $t_1 = 15$ seconds to reach a velocity of $v = 36$ km/h.
- The tram then enters a circular curve of radius $R = 100$ meters, maintaining a constant speed.
- After completing half of the circular path, the tram exits the curve and decelerates uniformly, coming to a stop at station B in $t_2 = 10$ seconds.

Tasks:

1. **Acceleration and Deceleration:** Determine the magnitude of the tram's acceleration during its uniform acceleration and deceleration phases.
2. **Distance Covered:** Calculate the total distance traveled by the tram from station A to station B.
3. **Centripetal Acceleration:** While the tram is on the circular curve, compute its centripetal acceleration.
4. **Net Force on the Tram:** If the mass of the tram is $m = 30,000$ kg, calculate the net force acting on it when it's halfway through the circular path. Assume that the frictional force (which provides the necessary centripetal force) is the only horizontal force acting on the tram.
5. **Historical Context:** The Austrian physicist Ernst Mach, who once taught in Vienna, proposed that a tram were to travel at a speed equal to half the speed of sound in air (assuming a speed of sound of 343 m/s at room temperature), what would its Mach number be? Would such a speed be feasible for a tram in a city setting? Explain.

Hints:

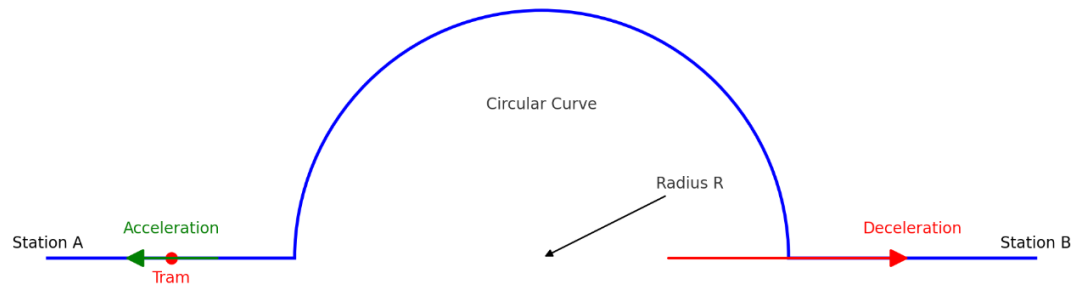
- Use the kinematic equations for uniformly accelerated motion to find distances and accelerations.
- The centripetal acceleration for an object moving in a circle of radius R at a constant speed v is given by $a_c = \frac{v^2}{R}$.
- Remember to convert all units appropriately before performing calculations.

Note: Disregard any effects of air resistance, and assume the tracks are frictionless except for the necessary friction to keep the tram moving in a circle.

- Really good problem. I would use it in a course

MA

Draw a figure for this problem



Can it do graduate coursework?

MA

Calculate the tree-level matrix element squared for $e^+ e^- \rightarrow \gamma \gamma$ summing over spins and polarizations.

The process $e^+ e^- \rightarrow \gamma \gamma$ involves the annihilation of an electron and positron to produce two photons. To compute the tree-level matrix element squared for this process, we utilize Feynman diagrams and rules from Quantum Electrodynamics (QED).

At tree level, there are two Feynman diagrams for this process:

1. The electron and positron annihilate into a virtual photon, which then decays into two real photons.
2. The electron emits a photon, becomes a virtual electron, and then annihilates with the positron, which emits the second photon.

Let's calculate the matrix element using the Feynman rules:

1. Matrix Element for the First Diagram:

Let's denote the momenta as: e^- with momentum p , e^+ with momentum p' , and the two outgoing photons with momenta k and k' .

For the electron-positron annihilation vertex, the matrix element contribution is:

$$-ie\gamma^\mu$$

For the photon decay vertex, the contribution is proportional to the metric tensor:

$$-ie^2 g^{\mu\nu}$$

Thus, the matrix element for the first diagram is:

$$\begin{aligned} \mathcal{M}_1 &= (-ie)^2 \bar{u}(p') \gamma^\mu v(p) \epsilon_\mu^*(k) \epsilon_\nu^*(k') g^{\mu\nu} \\ &= e^2 \bar{u}(p') \gamma^\mu v(p) \epsilon_\mu^*(k) \epsilon_\nu^*(k') g^{\mu\nu} \end{aligned}$$

1. Matrix Element for the Second Diagram:

The electron emission vertex gives:

$$-ie\gamma^\mu$$

The annihilation vertex into a photon gives:

$$-ie\gamma^\nu$$

So, for the second diagram:

$$\mathcal{M}_2 = e^2 \bar{u}(p') \gamma^\nu v(p) \epsilon_\nu^*(k') \epsilon_\mu^*(k) \gamma^\mu$$

1. Combine Both Diagrams:

$$\mathcal{M} = \mathcal{M}_1 + \mathcal{M}_2$$

2. Square the Matrix Element and Sum over Spins and Polarizations:

$$|\mathcal{M}|^2 = |\mathcal{M}_1 + \mathcal{M}_2|^2$$

For $|\mathcal{M}|^2$, we'll get terms like:

$$\mathcal{M}_1 \mathcal{M}_1^*, \mathcal{M}_2 \mathcal{M}_2^*, \text{ and } \mathcal{M}_1 \mathcal{M}_2^* + \mathcal{M}_2 \mathcal{M}_1^*$$

To compute these products, we need to use completeness relations for the spinors u and v and for the photon polarization vectors ϵ . After these steps, we'll obtain the squared matrix element summed over spins and polarizations.

It should be noted that the detailed calculations would involve a lot of Dirac algebra, usage of spin sums, and metric tensor contractions. This above is a broad outline. If you need the detailed steps, including the actual Dirac algebra and calculations, that would be quite lengthy.

Looking good

MA

write latex code to draw the feynman diagrams for this process

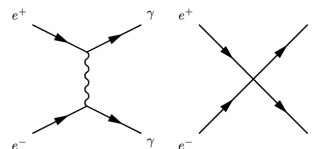


For the process $e^+ e^- \rightarrow \gamma \gamma$, there are two Feynman diagrams at tree-level. One diagram involves the electron and positron annihilating into a virtual photon, which then splits into two real photons. The other diagram is similar but involves the exchange of the virtual photon between the electron and positron.

To draw these Feynman diagrams using LaTeX, you can utilize the `feynmp` package. Here's how you can do it:

latex

Copy code



These aren't right, but code compiles and isn't far off

Can it do graduate coursework?

MA

continue with the calculation



Alright, let's dive deeper into the calculation for the process $e^+e^- \rightarrow \gamma\gamma$.

For simplicity, we won't use the full Dirac notation; instead, we will illustrate the procedure.

Recall our matrix elements:

$$\mathcal{M}_1 = e^2 \bar{u}(p') \gamma^\mu v(p) \epsilon_\mu^*(k) \epsilon_\nu^*(k') g^{\mu\nu}$$

$$\mathcal{M}_2 = e^2 \bar{u}(p') \gamma^\mu v(p) \epsilon_\nu^*(k') \epsilon_\mu^*(k) \gamma^\mu$$

1. Compute $|\mathcal{M}_1|^2$:

Using the properties of the metric tensor and Dirac matrices, the squared amplitude is:

$$|\mathcal{M}_1|^2 = e^4 \bar{u}(p') \gamma^\mu v(p) \bar{v}(p) \gamma^\alpha u(p) \epsilon_\mu^*(k) \epsilon_\nu^*(k') \epsilon_\alpha(k) \epsilon_\beta(k') g^{\mu\nu} g^{\alpha\beta}$$

Given the completeness relations for the photon polarization vectors and spin sums for the electron and positron spinors, we obtain:

$$|\mathcal{M}_1|^2 = e^4 (-g^{\mu\alpha}) (-g^{\nu\beta}) Tr[\gamma^\mu (\not{p} + m) \gamma^\alpha (\not{p}' + m) \gamma^\nu (\not{p} + m) \gamma^\beta (\not{p}' + m)] g_{\mu\alpha} g_{\nu\beta}$$

$$= e^4 Tr[\gamma^\mu (\not{p} + m) \gamma^\alpha (\not{p}' + m) \gamma^\nu (\not{p} + m) \gamma^\beta (\not{p}' + m)] g_{\mu\alpha} g_{\nu\beta}$$

1. Compute $|\mathcal{M}_2|^2$:

$$|\mathcal{M}_2|^2 = e^4 \bar{u}(p') \gamma^\nu v(p) \bar{v}(p) \gamma^\beta u(p) \epsilon_\nu^*(k') \epsilon_\mu^*(k) \epsilon_\beta(k) \epsilon_\alpha(k') g^{\nu\alpha} g^{\mu\beta}$$

Using similar techniques as before:

$$|\mathcal{M}_2|^2 = e^4 Tr[\gamma^\nu (\not{p}' + m) \gamma^\beta (\not{p}' + m) \gamma^\mu (\not{p} + m) \gamma^\alpha (\not{p} + m)] g_{\nu\alpha} g_{\mu\beta}$$

1. Compute the cross terms:

$$\mathcal{M}_1 \mathcal{M}_2^* + \mathcal{M}_2 \mathcal{M}_1^*$$

2. Summing Over Spins and Polarizations:

For electrons (and positrons):

$$\sum_{spins} u(p) \bar{u}(p) = \not{p} + m$$

$$\sum_{spins} v(p') \bar{v}(p') = \not{p}' - m$$

For photons:

$$\sum_{pol} \epsilon^\mu(k) \epsilon_\nu^*(k) = -g^{\mu\nu}$$

$$\sum_{pol} \epsilon^\mu(k') \epsilon_\nu^*(k') = -g^{\mu\nu}$$

These sums replace the individual spinor and polarization products in the matrix elements.

1. Putting It All Together:

$$|\mathcal{M}|^2 = |\mathcal{M}_1|^2 + |\mathcal{M}_2|^2 + \mathcal{M}_1 \mathcal{M}_2^* + \mathcal{M}_2 \mathcal{M}_1^*$$

The result will be an expression in terms of the momenta of the particles. To obtain a numerical result for a specific scattering event, these momenta would need to be provided.

This outline provides an approach, but the full calculation requires further algebra, including the detailed evaluation of traces of gamma matrices and the use of various Dirac matrix identities. Such calculations are standard in QED textbooks or in research articles that provide explicit computations.

MA

keep going until you get the answer



Alright, let's go step by step.

1. Amplitude \mathcal{M}_1 :

$$\begin{aligned} \mathcal{M}_1 &= e^2 \bar{u}(p') \gamma^\mu v(p) \epsilon_\mu^*(k) \epsilon_\nu^*(k') g^{\mu\nu} \\ &= e^2 \bar{u}(p') \gamma^\mu v(p) \epsilon_\mu^*(k) \epsilon_\nu^*(k') (-g^{\mu\nu}) \\ &= -e^2 \bar{u}(p') \gamma^\mu v(p) \epsilon_\mu^*(k) \epsilon_\nu^*(k') \end{aligned}$$

2. Amplitude \mathcal{M}_2 :

$$\mathcal{M}_2 = e^2 \bar{u}(p') \gamma^\mu v(p) \epsilon_\nu^*(k') \epsilon_\mu^*(k) \gamma^\mu$$

Square of \mathcal{M}_1 :

$$\begin{aligned} |\mathcal{M}_1|^2 &= \sum_{spins, pol} |\mathcal{M}_1|^2 \\ &= \sum_{spins} |\bar{u}(p') \gamma^\mu v(p)|^2 \sum_{pol} \epsilon_\mu^*(k) \epsilon_\nu(k') \epsilon^\mu(k) \epsilon_\nu^*(k') \\ &= \sum_{spins} |\bar{u}(p') \gamma^\mu v(p)|^2 (-g^{\mu\nu}) (-g^{\mu\nu}) \\ &= 4 \sum_{spins} |\bar{u}(p') \gamma^\mu v(p)|^2 \\ &= 4 Tr[\gamma^\mu (\not{p} + m) \gamma_\mu (\not{p}' - m)] \\ &= 4 Tr[\gamma^\mu \not{p} \gamma_\mu \not{p}'] \end{aligned}$$

Using properties of Dirac matrices, this becomes:

$$= 4(4p^\mu p'_\mu)$$

Square of \mathcal{M}_2 :

$$|\mathcal{M}_2|^2 = \sum_{spins, pol} |\mathcal{M}_2|^2$$

Following a similar approach to above, we get:

$$= 4(4p^\nu p'_\nu)$$

Cross terms:

$$\begin{aligned} \mathcal{M}_1 \mathcal{M}_2^* + \mathcal{M}_2 \mathcal{M}_1^* &= 2 Re[\mathcal{M}_1 \mathcal{M}_2^*] \\ &= 2 Re[e^4 \bar{u}(p') \gamma^\mu v(p) \bar{v}(p) \gamma_\mu u(p')] \\ &= 2e^4 Re[\bar{u}(p') \gamma^\mu v(p) \bar{v}(p) \gamma_\mu u(p')] \\ &= 8e^4 p^\mu p'_\mu \end{aligned}$$

Total squared matrix element:

$$\begin{aligned} |\mathcal{M}|^2 &= |\mathcal{M}_1|^2 + |\mathcal{M}_2|^2 + \mathcal{M}_1 \mathcal{M}_2^* + \mathcal{M}_2 \mathcal{M}_1^* \\ &= 8e^4 p^\mu p'_\mu + 8e^4 p^\nu p'_\nu + 8e^4 p^\mu p'_\mu \\ &= 24e^4 p^\mu p'_\mu \end{aligned}$$

This is the squared matrix element for the process $e^+e^- \rightarrow \gamma\gamma$, summed over spins and polarizations. Here, p and p' are the momenta of the electron and positron, respectively. The actual differential cross-section and total cross-section would require further integration over the final-state phase space.

Is it memorizing?

- No. It's generalizing.

AI is doing just what we ask students to do

- Learn from practice
- Show your steps

In any case, memorization is an essential part of learning

Answer is wrong

- Right formulas and ingredients
- Student could easily fix

Can it help with research?

Research question a graduate student asked me last week

MA

what is consistent slicing of the bulk in an RS model?



In the context of Randall-Sundrum (RS) models and other brane-world scenarios, the term "slicing" often refers to a specific coordinate choice or parameterization of the 5-dimensional bulk spacetime. Different slicings give different pictures of the 5-dimensional spacetime and can be more or less convenient depending on the specific problem being addressed.

In the original RS models, the 5-dimensional metric was often written in terms of the so-called conformal coordinate z , with the metric given by:

$$ds^2 = e^{-2k|z|} \eta_{\mu\nu} dx^\mu dx^\nu + dz^2$$

where $\eta_{\mu\nu}$ is the 4-dimensional Minkowski metric, k is a scale related to the curvature of the 5D spacetime, and z is the extra-dimensional coordinate.

However, in certain calculations, especially those involving gravity or cosmology on the brane, it's sometimes more convenient to use a different slicing, known as the Gaussian Normal (GN) coordinates. In the GN coordinate system, the brane is always located at a fixed value of the extra-dimensional coordinate (often chosen to be zero), and the metric takes on a form like:

$$ds^2 = n^2(y) \eta_{\mu\nu} dx^\mu dx^\nu + dy^2$$

where y is the new coordinate and $n(y)$ is the warp factor.

The consistent slicing of the bulk refers to a way of expressing the 5D spacetime that consistently embeds the 4D brane in it. This can be especially important when studying phenomena like the cosmological evolution of the brane universe.

It's important to note that different slicings are just different coordinate choices: they all describe the same physical 5-dimensional spacetime, but the mathematics might be simpler in one slicing compared to another depending on the specific problem.

- Provides useful explanations of subtle concepts

WARNING!

Hallucinates references and quotes
(as of 2023)

Text generation

MA

Write an abstract for a public lecture by matthew schwartz in vienna on Machine Learning and the Future of Particle Physcs

Machine Learning and the Future of Particle Physics

Public Lecture by Prof. Matthew Schwartz, Harvard University

it knows who I am

Prof. Matthew Schwartz, renowned for his contributions to quantum field theory and the author of "Quantum Field Theory and the Standard Model", will elucidate the transformative role of machine learning (ML) in particle physics. As the data influx from modern experiments like the Large Hadron Collider becomes overwhelming, ML offers groundbreaking approaches to data interpretation, anomaly detection, and the prediction of particle interactions. Join us to explore how this intersection of ML and physics is reshaping our quest to understand the fundamental nature of the universe.

it knows what my talk is about!

Chat GPT's ability to write and edit text is astounding

- Particularly useful for non-native speakers

Code generation

MA

can you write a script to download all the pdf files from a website?



Certainly! The task of downloading all the PDFs from a website can be accomplished with a web scraping tool like BeautifulSoup in Python.

However, you should be aware that web scraping is not always legal or ethical. Always consult a website's `robots.txt` file to determine if scraping is allowed and never overload a website's servers or download proprietary content without permission.

Below is a basic script using Python's `requests` and `Beautiful Soup` to scrape and download PDF files from a given website:

```
python
```

 Copy code

```
import os
import requests
from bs4 import BeautifulSoup
```

Worked immediately without editing!

Advanced data analysis



Advanced data analysis

Try a version of ChatGPT that knows how to write and execute python code, and can work with file uploads. Try asking for help with data analysis, image conversions, or editing a code file. Note: files will not persist beyond a single session.



Beta features

MA

Generate data for the the height of a bouncing ball including the effects of friction over the course of 5 bounces with 100 data points as a csv file



Certainly! To simulate the height of a bouncing ball with the effects of friction, we will use the following assumptions:

1. The ball is dropped from an initial height h_0 .
2. When the ball hits the ground, it bounces back up with a fraction r of its previous height due to the effects of friction. That is, after the first bounce, it reaches a height of $r \times h_0$, after the second bounce, it reaches a height of $r \times r \times h_0$, and so on.
3. The height of the ball as a function of time during each bounce will follow a parabolic trajectory due to gravitational acceleration.

Given these assumptions:

- The maximum height after the i -th bounce will be $h_i = r^i \times h_0$.
- The time it takes for the ball to hit the ground after being dropped from height h is $t = \sqrt{\frac{2h}{g}}$, where g is the gravitational acceleration (approximated as 9.81 m/s^2).
- The time it takes for the ball to complete an up-and-down bounce cycle after the i -th bounce is $2t_i$, where t_i is the time it takes for the ball to hit the ground from height h_i .

For simplicity, let's assume:

- $h_0 = 10 \text{ m}$ (initial height from which the ball is dropped)
- $r = 0.8$ (restitution coefficient, accounting for energy loss due to friction)



Working... ✨

Show work ▾

You can [download the CSV file here.](#)

Advanced data analysis

MA



bouncing_ball.csv
Spreadsheet

Analyze what is going on with the data in this csv file



Alright, let's start by loading the data from the CSV file and taking a preliminary look at its contents.

Finished working

Show work

The CSV file contains two columns: "Time" and "Height". It appears to represent the height of an object (possibly a ball) over time.

Let's rename the file to `physicsexample.csv`
and rename the columns to `t` and `h`

Advanced data analysis

MA



physicsexample.csv
Spreadsheet

Analyze this data



Let's start by loading and inspecting the contents of the provided CSV file.

Finished working

Show work

The data consists of two columns:

1. t - which seems to represent time.
2. y - which seems to represent some kind of position or measurement with respect to time.

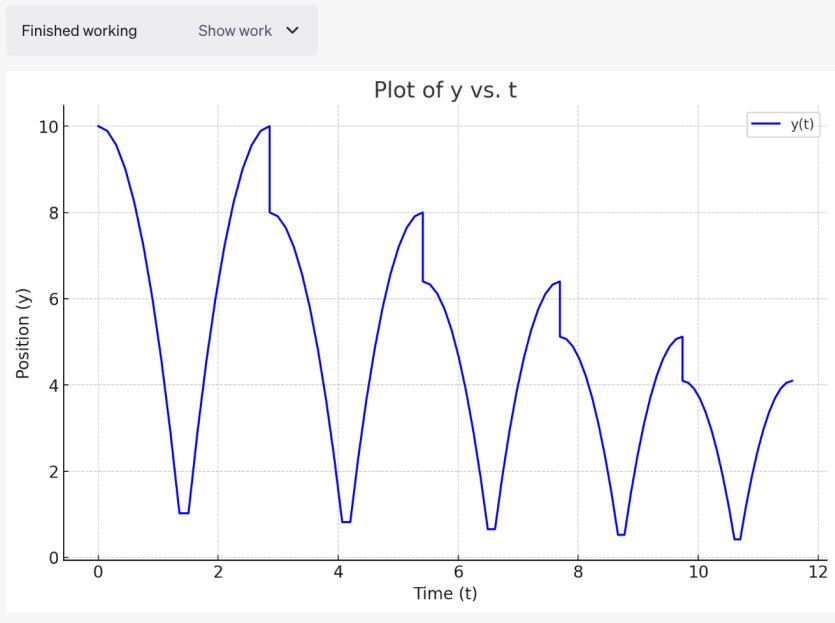
Given the context, it's possible that this data represents the motion of an object over time, perhaps its vertical position as it undergoes free fall or some other type of motion.

To better understand the nature of this data, we can plot it and look for any discernible patterns. Would you like to proceed with this visualization?

MA

yes

Advanced data analysis



The plot shows a downward concave curve, which suggests that the object is undergoing accelerated motion, possibly due to gravity. The object starts from a certain height (as indicated by the initial y value) and its height decreases over time, implying a downward motion.

This behavior is consistent with the motion of an object in free fall under the influence of gravity, starting from rest. The mathematical model for such motion can be represented as:

$$y(t) = y_0 - \frac{1}{2}gt^2$$

Where:

- y_0 is the initial height.
- g is the acceleration due to gravity (approximately 9.81 m/s^2 near the Earth's surface).

If this is indeed the case, we can fit the data to this model to estimate the values of y_0 and g .

Would you like to proceed with this fitting?

MA

yes

Finished working Show work

The fitted model suggests the following:

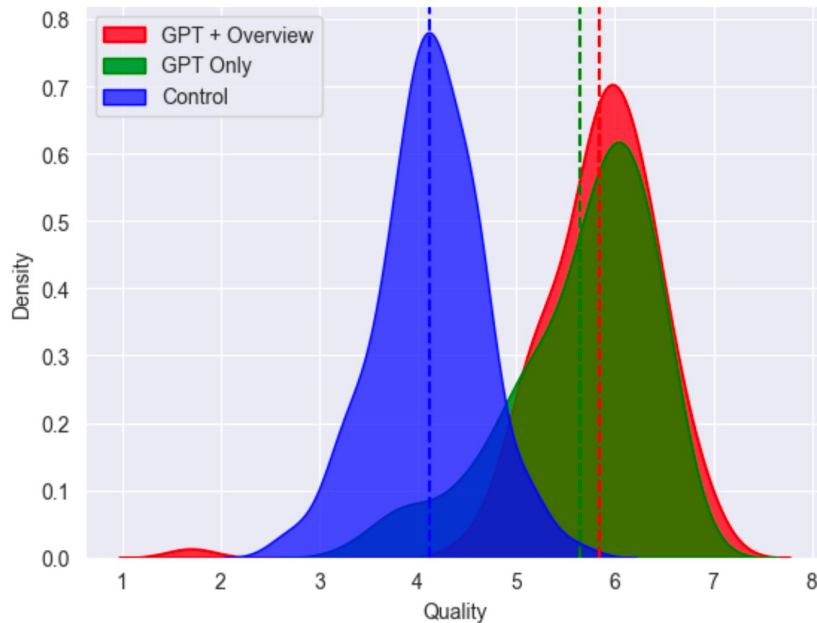
1. The initial height y_0 is approximately 6.23 units (possibly meters).
2. The estimated acceleration g is approximately 0.0643 units/s^2 .

This value for g is much smaller than the typical acceleration due to gravity near the Earth's surface (which is around 9.81 m/s^2). This could imply that the data doesn't represent free fall due to gravity on Earth, or perhaps it's a scaled representation, or there might be other forces at play.

To visualize how well our model fits the data, I can plot both the original data and the fitted curve on the same graph. Would you like to see this visualization?

and so on...

Additional observations on GPT4



Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013

- Consultants using AI produced **40% higher quality results** than those without.

- GPT-4 is creative
 - ❖ exceeds [99% of people on the Torrance Tests of Creative Thinking](#).

Nobody knows what these models are capable of

GPT4 is the worst generative model you will ever use



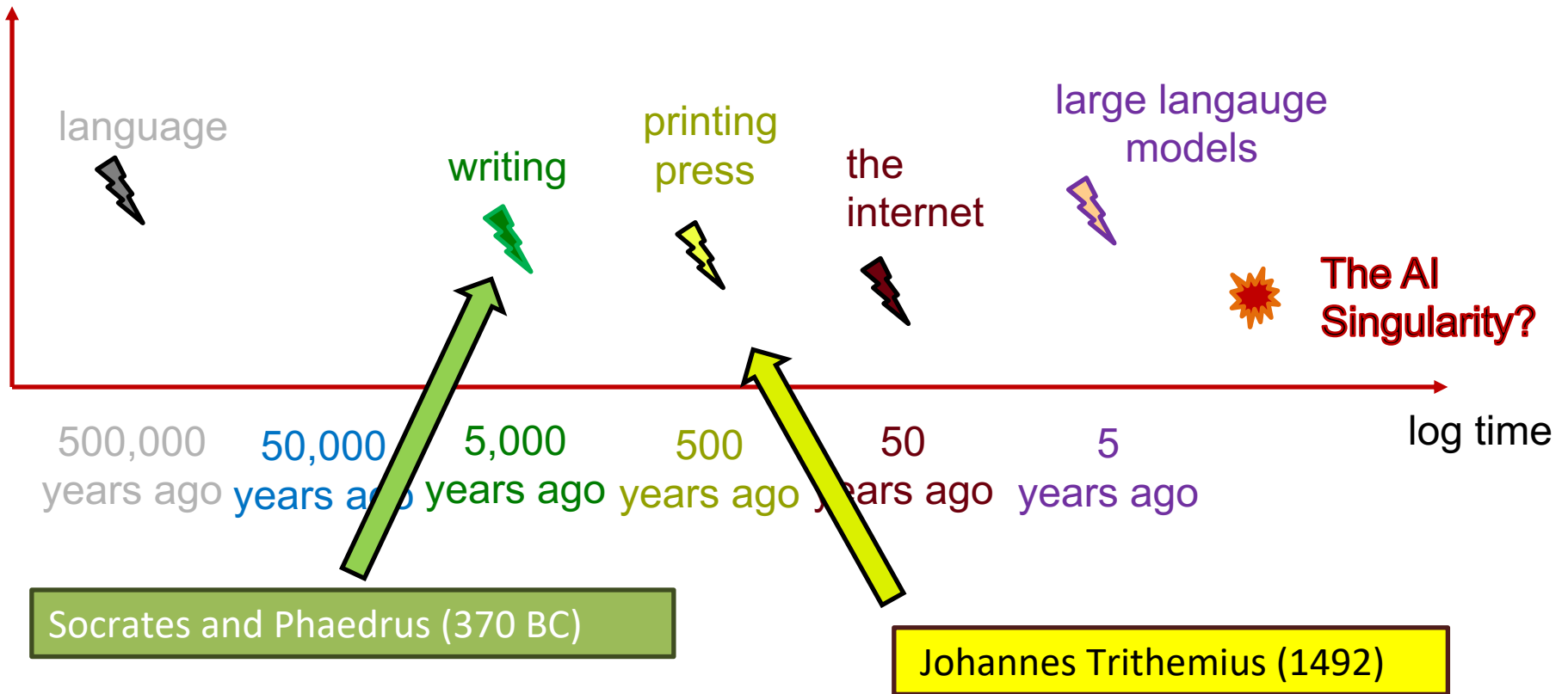
Wall-E
2008

2. Atrophy

Will large language models make us stupid?

No

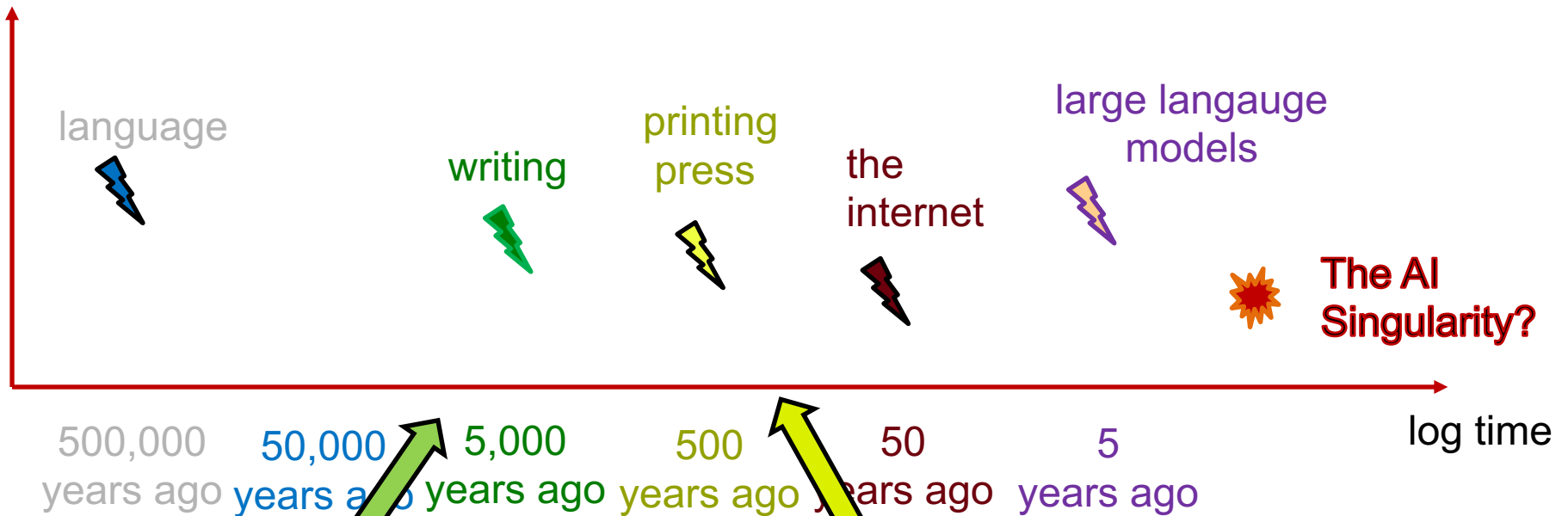
Landmarks in education



And it is *not true wisdom* that you offer your disciples, but **only the semblance of wisdom**, for by telling them of many things without teaching them you will make them seem to know much while for the most part they know nothing.

"The printed book is made of paper and, like paper, will quickly disappear. But the scribe working with parchment ensures lasting remembrance for himself and for his text"

Landmarks in education



Socrates and Phaedrus (370 BC)

*If men learn [from writing], it will implant forgetfulness in their souls. **They will cease to exercise memory because they rely on that which is written.***

*And it is no true wisdom that you offer your disciples, but **only the semblance of wisdom**, for by telling them of many things without teaching them you will make them seem to know much while for the most part they know nothing.*

Johannes Trithemius (1492)

"The printed book is made of paper and, like paper, will quickly disappear. But the scribe working with parchment ensures lasting remembrance for himself and for his text"

Will large language models destroy education?

No

Teaching

- LLMs can be infinitely patient, super-knowledgable teaching assistants
- Education adapted to invididual learning styles
 - Classroom teaching is inefficient
- Private tutors for all
 - **Complete democratization of education**

Assessment

- Cannot base grades on correctness of homework problems anymore
- Must confront **misalignment between assessment and learning goals**

short-term fix

In-class exams

long-term fix

machine does assessment

Potential to improve education is huge!

- What is the future role of the instructor?
- What is the future role of institutions?



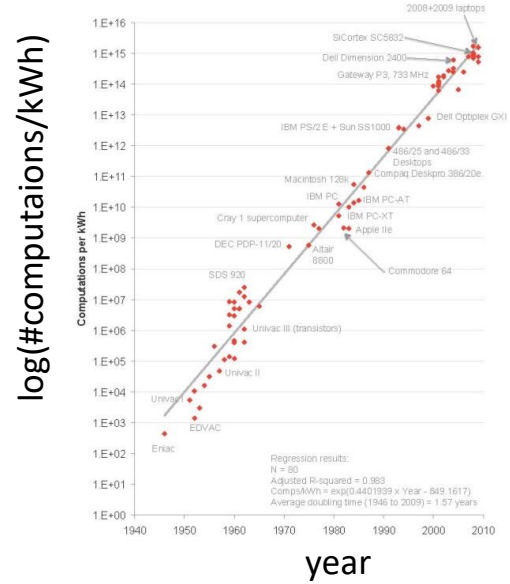
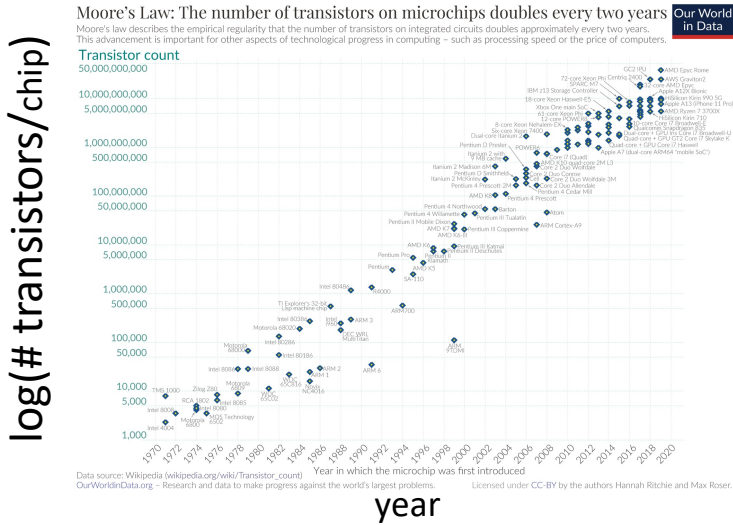
3. Scalability

Exponential laws
must fail

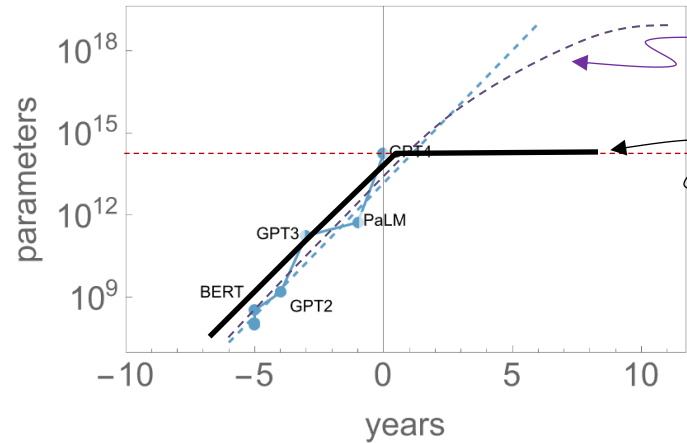
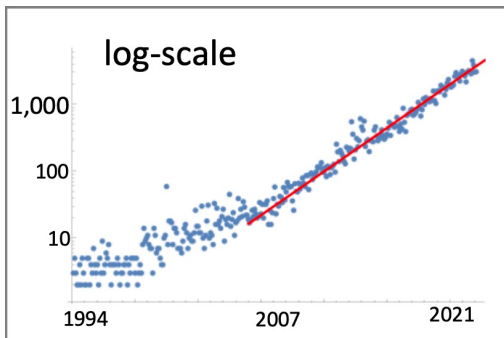
Exponential growth finds a way

Moore's law: computation power/time

Koomey's law: energy efficiency /time



number of papers about ML or AI
Krenn et al arXiv:2210.00881



• even subexponential growth will soon exceed human capacity

- not going to happen
- impossible to believe this is the endpoint of AI

How can machines continue to grow?

high school

- **19%** of high school graduates in the US **cannot read**

early college

- **solve standard problems using standard formulas**
- learn computational tools

late college

- master abstract concepts
- **solve difficult but previously solved problems**

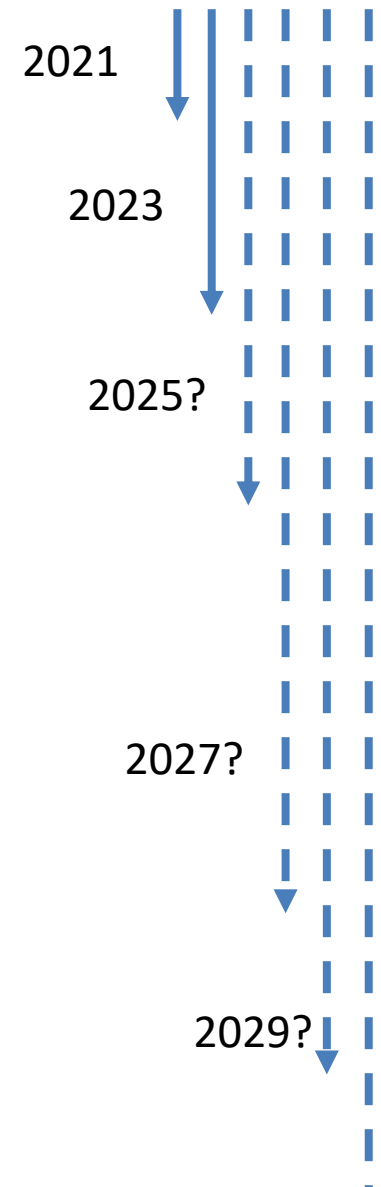
graduate school

- tackle difficult unsolved problems
 - **invent** or adapt appropriate **tools**
 - **create** and solve **toy problems** to build intuition
 - simplify enough until solvable
- arrive at the boundaries of human understanding

professor

- write grants, teach, advise students, give public lectures
- **find solvable problems** -- not too easy, not too hard
- nudge the boundaries of human understanding

Machine learning



How can machines continue to grow?

Current state-of-the art can solve textbook physics problems

- Trained on solved problems from books, physics.stackexchange, chegg etc.
- We have nearly saturated the training data. **Where will new data come from?**

Where did the old data come from?

- Textbook problems are written by human beings
- physics.stackexchange solutions are written by human beings

Alpha Zero: learns to solve chess problems by generating its own training data



GPT4:

- can generate and solve problems
- user (human) feedback helps refine model
- it can refine its own model!

However, negative energy doesn't make sense in this context, which means we made a mistake in our calculations. Let's go back and check the steps.

Language models have access to everything graduate students have

4. Interpretability



“orc yoga”

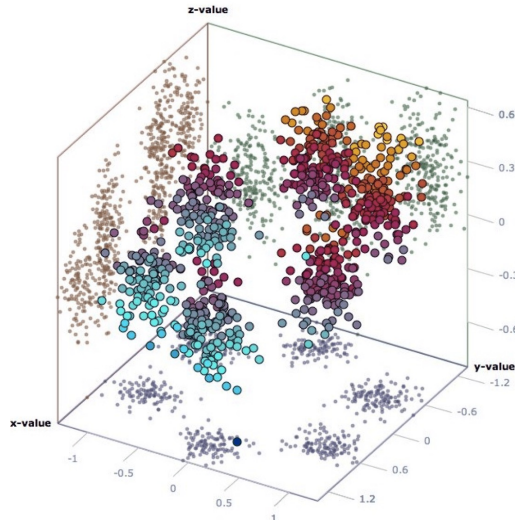
Markus Graf, March 2023

Should artificial intelligence be interpretable to humans?

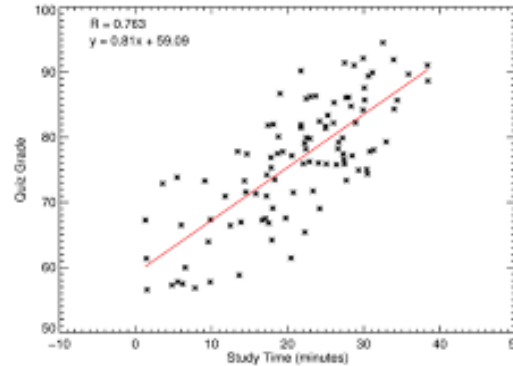
MDS, Nature reviews physics (2022)

Humans are limited by biology

❖ We like to “visualize”



project
to 2D



Why do we do this? Because we have **eyes**

- 2D is not special to a machine.
- Machines can “visualize” in d dimensions

Eyes have **nothing to do**
with particle physics!

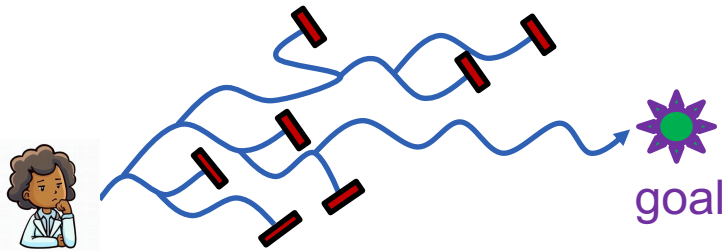
❖ Humans can only **hold 5-9 concepts in working memory** at once

- We like simple-looking equations

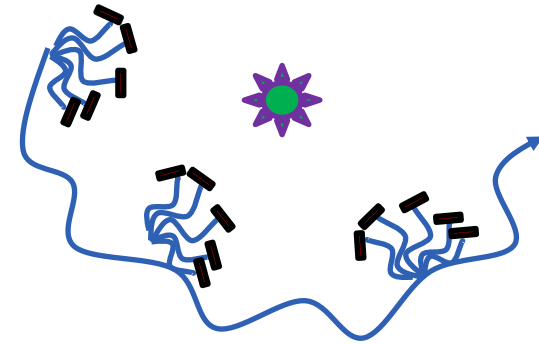
$$i\partial_t\psi = H\psi \quad i\partial\psi = m\psi \quad G_{\mu\nu} = \kappa T_{\mu\nu}$$

- **Computer** memory can **handle much more** than 5-9 concepts
- They can understand systems not governed by simple equations

In the past, we have made progress despite many dead ends



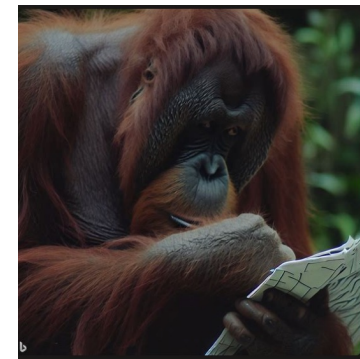
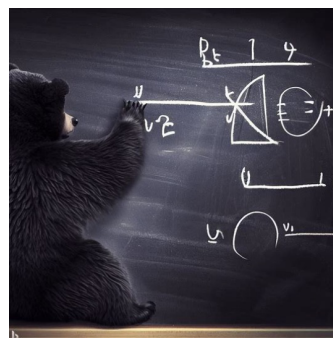
In recent particle physics, it's hard to tell if dead ends have deflected us off course



At this rate, a complete theory of everything may take a very long time

- Maybe real progress will take 100 years, or maybe 10,000 years

Maybe the problems are just too hard (for us)



Could a cat ever learn to play chess?

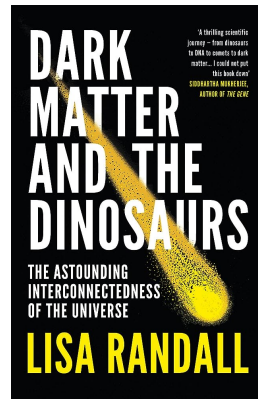
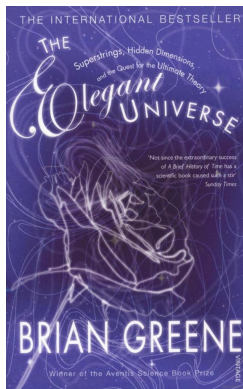
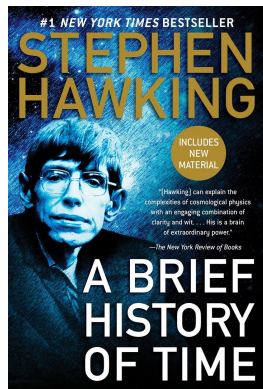
If humans are not qualitatively different from cats, we have limits too

Benefits of machine understanding

Suppose a machine understands the theory of everything

- e.g. can calculate the fine-structure constant from scratch
- e.g. can predict the endpoint of black-hole evaporation

Is this enough or do we need to understand it too?



popular science books

- They give us a feeling for how things work
- We are happy that the authors understand the details, even if we don't

I don't understand the proof of Fermat's last theorem

- But I'm glad that somebody does
- Does it matter to me that the person is human? Not really.

If a machine understands fundamental physics it can

1. Dumb it down so we can get the general idea
2. Find practical applications

Is this what we want? No.
But maybe it's the best we will get.

5. The end is near

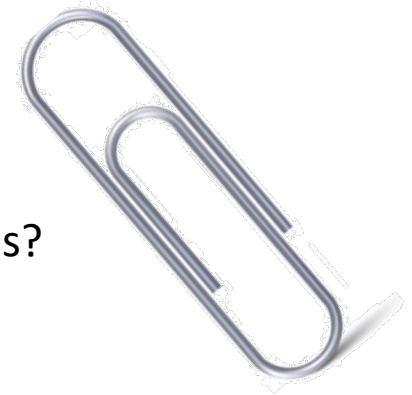


Jonty Hampson, Sep 2023

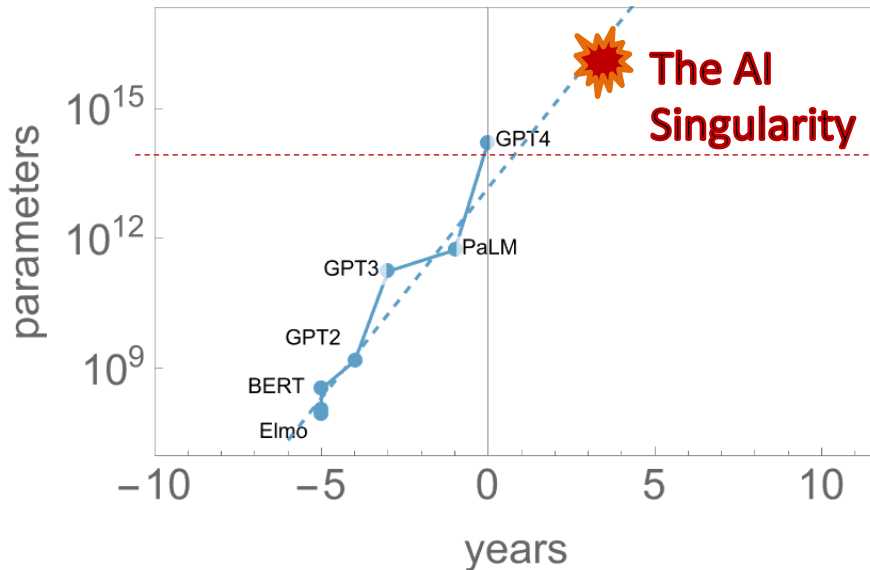
What could go wrong?

❖ “Paperclip Maximizer” Nick Bostrom, 2014

- Suppose a machine was designed to maximize paperclip production.
- To truly minimize its loss function, it would destroy everything on earth to produce paperclips



❖ What if machines can write their own code and improve themselves?



Not that far off

- Machines can already write and debug code
- Machines can already evaluate their own flaws

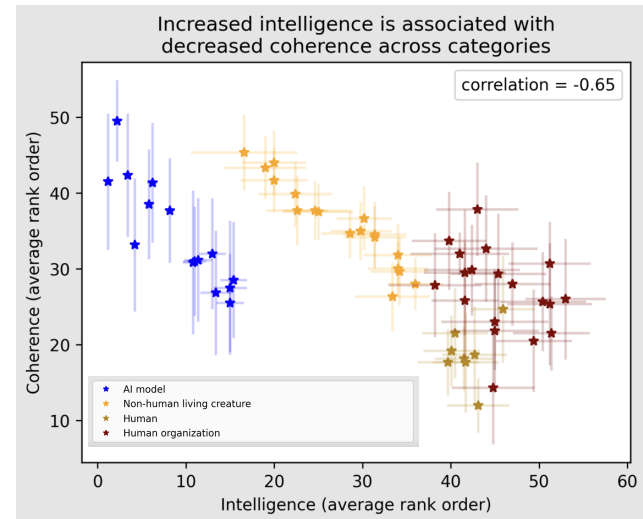
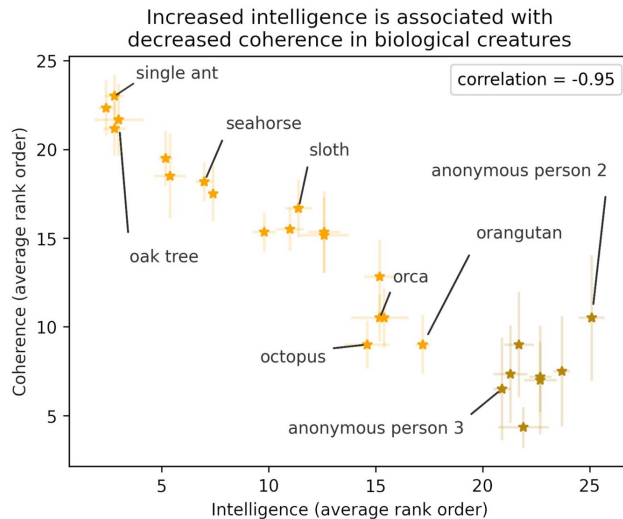
Are we doomed?

Maybe it won't be so bad

Jascha Sohl-Dickstein [Google Deepmind <https://sohldickstein.com/>]

- “We are a **hot mess** of inconsistent, self-undermining, irrational behavior, with objectives that change over time”

Coherence
=
focus on
a single goal



- Intelligence is anti-correlated with coherence
 - More profoundly, intelligence *requires* incoherence
- Are intelligent machines dangerous? **Maybe**
- Are intelligent animals dangerous? **Definitely**

We are not intelligent enough to predict what higher intelligence will do

Conclusions



“Trump and Putin in a Jail Cell”

Marko Lindell , March 2023

The future is here!

- Machine learning is having a **huge impact** on particle physics
 - Collider physics, detector simulation, anomaly detection, lattice QCD,
 - Many untouched future directions: model building, experiment design, etc.
- Large Language models are **revolutionary!**
 - Education will never be the same
 - Private tutors for everyone
 - Research will never be the same
 - Information is organized and analyzed critically and efficiently

There are reasons to be optimistic

- Problems in particle physics are **difficult** and **progress has been slow**
- Machine learning has the potential to make **rapid progress in our lifetimes**

Future is bright, but mysterious:

What will education look like in 10 years?

What will research look like in 10 years?

What will be our relationship with artificial intelligence in the future?