

Theoretical High-Energy Physics and Artificial Intelligence

AIME24

November 21, 2024

Matthew Schwartz
Harvard University



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

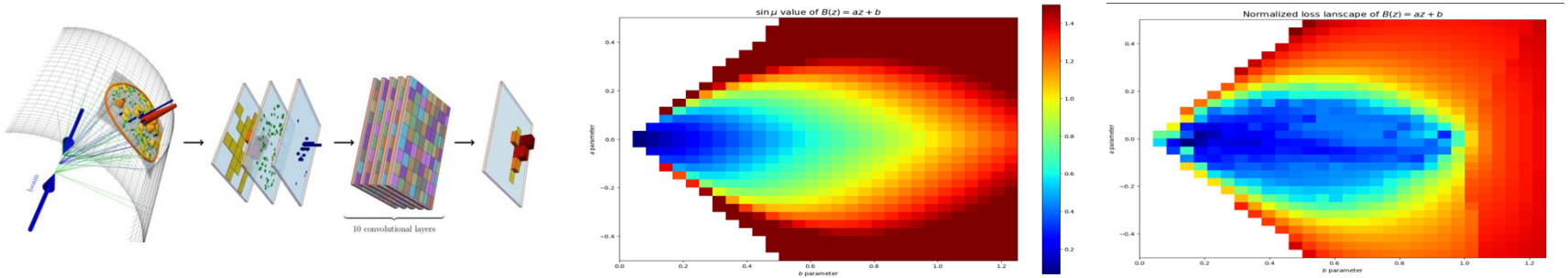
Boston, MA, USA



Outline

The Present

- Symbolic ML for high energy theory



The Future

- Can machines do theoretical physics?



1. The Past

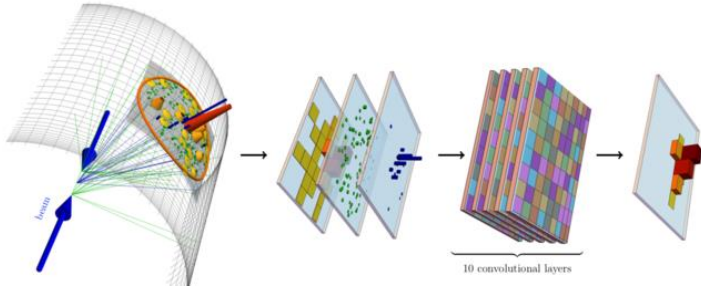


Björn Karlsson, MidJourney January, 2024

Machine learning approach

Take some tool highly engineered for another purpose and apply it to physics problems

Convolutional networks for facial recognition

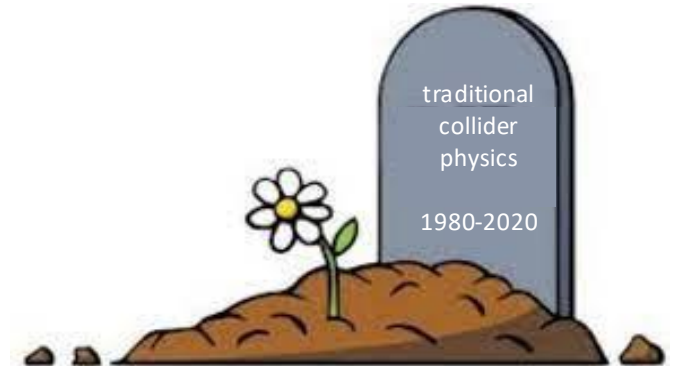
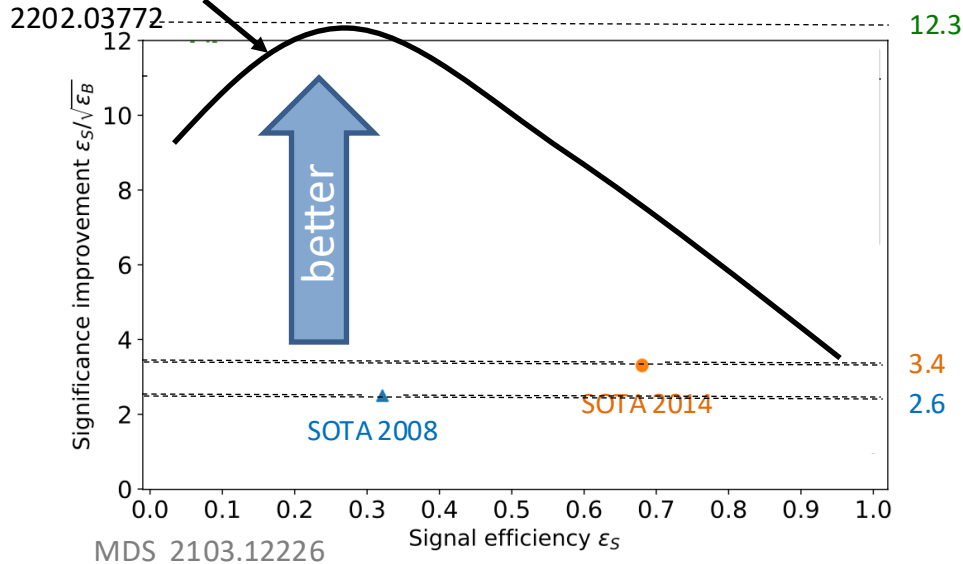


e.g. Top tagging

Point clouds/deep sets for self-driving cars



particle transformer



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

2. The present



Björn Karlsson, MidJourney January, 2024

What do high-energy theorists actually do?

1. Most papers on hep-th and hep-ph are largely symbolic

- Calculate something
- Find/establish some relationship
- Solve some toy model
- Extract behavior of some theory in some limit

Can ML do these things?

- It's starting to...
- ML is good at helping solve well-defined problems

2. What makes a question interesting?

- It connects to nature
 - You can make progress on it
 - Someone else thought it was interesting
 - It is related to something someone else thought was interesting
- Much harder problem

Can ML answer this?

Not yet. But soon.

Most uses of ML in physics so far are **data science**

- Machine learning is great at characterizing numerical data
- Has led to revolutionary progress on a great variety of physical questions
- Much much more still to be done

How do we transition from **data science** to **symbolic problems**?

- Large Language Models show that ML good for symbolic problems
 - Potential is there
- A lot of what physicists do is study **examples**, look for **patterns**, and **generalize**
- A place to start is with **discrete symbolic data**
 - Find some theoretical physics problems where ML can help
 - Get a feel for what it can (and currently cannot) do

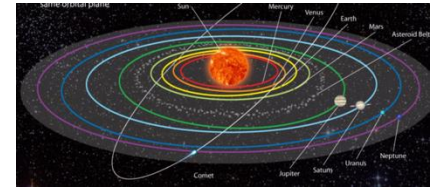
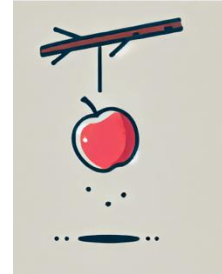
1. Simplifying spinor-helicity amplitudes

Many advances in theoretical physics come from simplification/unification



Newton: motion of apples and planets governed by the same simple law

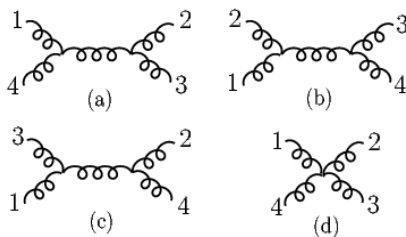
$$F = G \frac{m_1 m_2}{r^2}$$



Quantum field theory

calculate some Feynman diagrams for
4 gluon scattering

get a messy answer



$$= -\frac{\langle 12 \rangle \langle 13 \rangle \langle 24 \rangle [13] [24]}{\langle 23 \rangle \langle 34 \rangle^2 [12] [23]} + \frac{\langle 12 \rangle \langle 13 \rangle \langle 24 \rangle [14]}{\langle 23 \rangle \langle 34 \rangle^2 [12]} - \frac{\langle 12 \rangle \langle 24 \rangle [13] [24]^2}{\langle 34 \rangle^2 [12] [14] [23]} + \frac{\langle 12 \rangle \langle 24 \rangle [24]}{\langle 34 \rangle^2 [12]} + \frac{\langle 12 \rangle [13] [24] [34]}{\langle 34 \rangle [12] [14] [23]}$$

$$- \frac{\langle 12 \rangle [34]}{\langle 34 \rangle [12]} - \frac{\langle 13 \rangle \langle 14 \rangle [13] [34]}{\langle 34 \rangle^2 [12] [23]} - \frac{\langle 13 \rangle \langle 24 \rangle [13] [24] [34]}{\langle 34 \rangle^2 [12] [14] [23]} + \frac{\langle 13 \rangle \langle 24 \rangle [34]}{\langle 34 \rangle^2 [12]} - \frac{\langle 13 \rangle \langle 24 \rangle [13] [24]}{\langle 34 \rangle^2 [12]^2} + \frac{\langle 13 \rangle \langle 24 \rangle [14] [23]}{\langle 34 \rangle^2 [12]^2}$$

$$+ \frac{\langle 13 \rangle [13] [34]^2}{\langle 34 \rangle [12] [14] [23]} - \frac{\langle 14 \rangle \langle 23 \rangle [34]}{\langle 34 \rangle^2 [12]} - \frac{\langle 23 \rangle \langle 24 \rangle [13] [24]^2}{\langle 34 \rangle^2 [12]^2 [14]} + \frac{\langle 23 \rangle \langle 24 \rangle [23] [24]}{\langle 34 \rangle^2 [12]^2} + \frac{\langle 23 \rangle [13] [24] [34]}{\langle 34 \rangle [12]^2 [14]} - \frac{\langle 23 \rangle [23] [34]}{\langle 34 \rangle [12]^2}$$

simplify! = $-\frac{\langle 12 \rangle [34]^2}{\langle 34 \rangle [14] [23]}$

- simpler form suggests deeper structure
- is there a better way to do the calculation?
 - In this case, yes! (BCFW recursion)

Generate training data by scrambling

Input:
$$\frac{-\langle 34 \rangle^2 [12]^3 [34] [35] + \langle 34 \rangle^2 [12]^2 [13] [23] [45] - \langle 34 \rangle^2 [12]^2 [14] [23] [35]}{\langle 23 \rangle \langle 34 \rangle [13]^2 [34] [35] + \langle 23 \rangle \langle 45 \rangle [13]^2 [35] [45] - \langle 23 \rangle \langle 45 \rangle [13] [14] [35]^2}$$

[15] →
$$\frac{[14][35] - [13][45]}{[34]}$$
 Schouten

$$\frac{-\langle 34 \rangle^2 [12]^3 [35] - \langle 34 \rangle^2 [12]^2 [15] [23]}{\langle 23 \rangle \langle 34 \rangle [13]^2 [35] - \langle 23 \rangle \langle 45 \rangle [13] [15] [35]}$$

[25] →
$$\frac{[12][35] + [15][23]}{[13]}$$
 Schouten

⟨24⟩ →
$$\frac{\langle 45 \rangle [15] - \langle 34 \rangle [13]}{[12]}$$

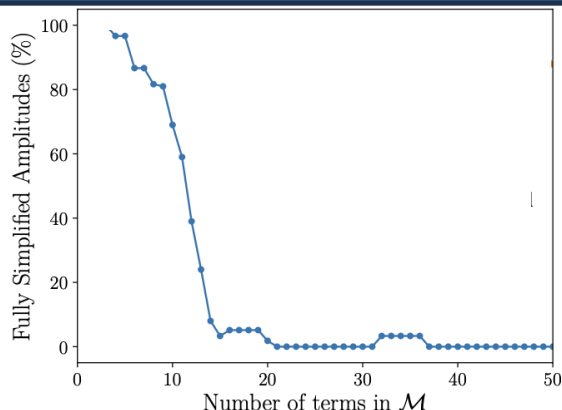
Momentum conservation

$$\frac{\langle 34 \rangle^2 [12]^2 [25]}{\langle 23 \rangle \langle 34 \rangle [13] [35] - \langle 23 \rangle \langle 45 \rangle [15] [35]}$$

$$\frac{\langle 34 \rangle^2 [12] [25]}{\langle 23 \rangle \langle 24 \rangle [35]}$$
 Desired output

transformer learns to translate

Works great for expressions with < 15 terms



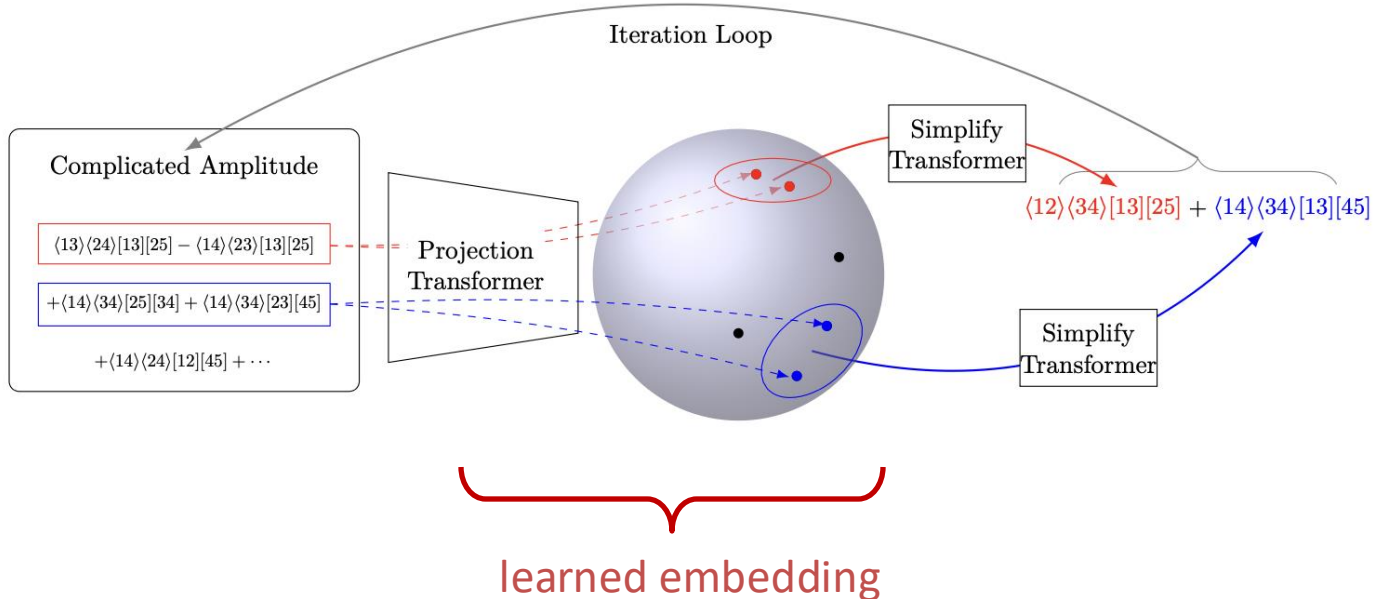
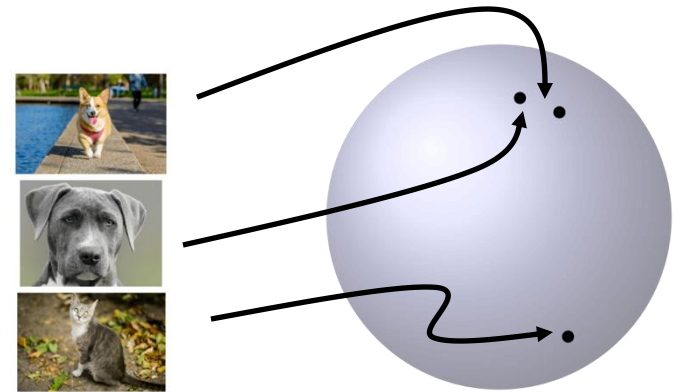
- Expressions of interest can be very long
 - hundreds of terms
- Need new techniques for organizing transformer
 - We use **contrastive learning**

Contrastive learning

Learn an embedding so that terms that are **similar** are **close**

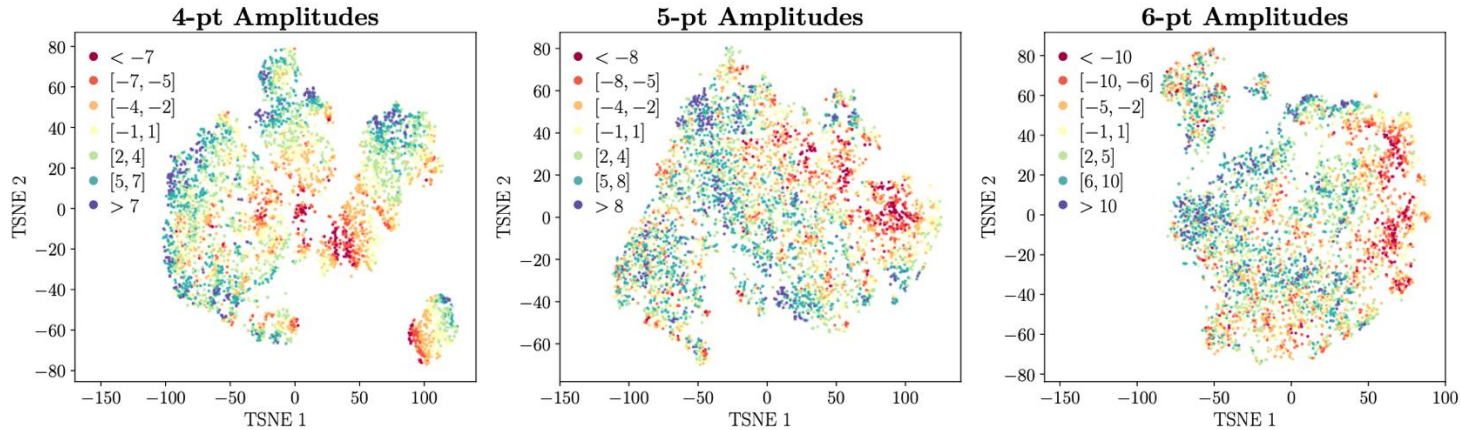
- **similar** = appear in some identity
- **close** = metric on embedding space

1. Pick subset of terms expected to simplify
2. Apply transformer
3. Repeat



Contrastive learning

- Cross-check: t-SNE on latent space

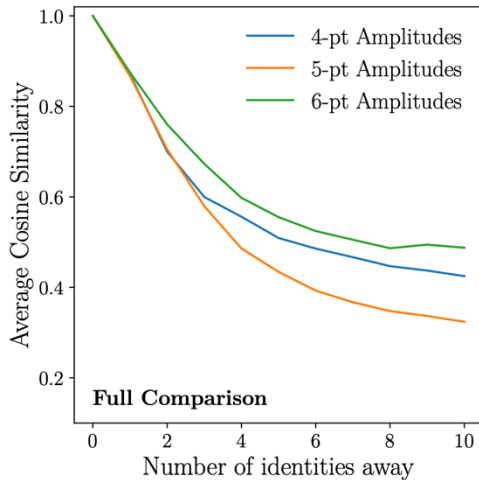


- color is mass dimension
 - Learns dimensional analysis
 - Learns other features as well

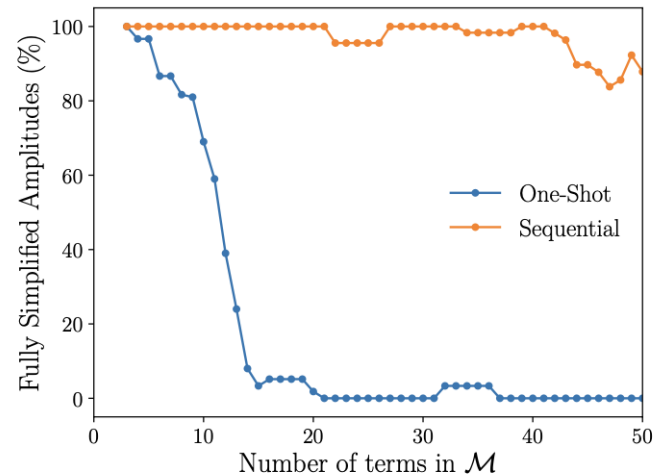
Does it work?

Yes!

It can simplify long expressions now

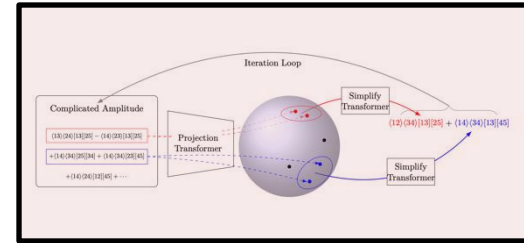
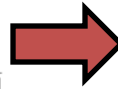


distance inversely correlated with complexity



Example application: 5 gluon amplitude

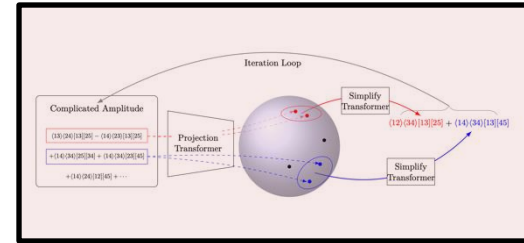
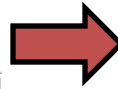
$$\begin{aligned}
 \mathcal{M} = & \frac{\langle 12 \rangle^3 [13]}{(23)(24)(35)(45)[23]} + \frac{\langle 12 \rangle^3 [14][25]}{(13)(24)(35)(45)[12][45]} - \frac{\langle 12 \rangle^3 [15][24]}{(13)(24)(35)(45)[12][45]} - \frac{\langle 12 \rangle^2 [13][34]}{(24)(35)(45)[14][23]} + \frac{\langle 12 \rangle^2 [13]}{(24)(35)(45)[12]} \\
 & - \frac{\langle 12 \rangle^2 [13][24][35]}{(24)(35)(45)[12][23][45]} + \frac{\langle 12 \rangle^2 [13][25][34]}{(24)(35)(45)[12][23][45]} + \frac{\langle 12 \rangle^2 [14][35]}{(24)(35)(45)[12][45]} - \frac{\langle 12 \rangle^2 [15][34]}{(24)(35)(45)[12][45]} - \frac{\langle 12 \rangle^2 [13][45]}{(23)(24)(35)[12][23]} \\
 & - \frac{\langle 12 \rangle^2 (23)[13][45]}{(15)(24)(34)(35)[14][15]} - \frac{\langle 12 \rangle^2 [45]}{(15)(34)(35)[15]} - \frac{\langle 12 \rangle^2 [13][34][45]}{(15)(24)(35)[14][15][23]} + \frac{\langle 12 \rangle^2 [13][45]}{(15)(24)(35)[12][15]} + \frac{\langle 12 \rangle^2 [14][45]}{(15)(23)(35)[12][15]} \\
 & + \frac{\langle 12 \rangle^2 (34)[13][34][45]}{(12)^2 (15)[15]} - \frac{\langle 12 \rangle^2 (23)[23]}{(12)^2 (23)[23]} - \frac{\langle 12 \rangle^2 (23)[25][34]}{(12)^2 (23)[25][34]} \\
 & + \frac{\langle 15 \rangle (23)(24)(35)[12][15][23]}{(13)(24)(35)(45)[12]} + \frac{\langle 13 \rangle (24)(35)(45)[12]}{(13)(24)(35)(45)[12]} - \frac{\langle 13 \rangle (24)(35)(45)[12][45]}{(13)(24)(35)(45)[12][45]} \\
 & + \frac{\langle 12 \rangle^2 (23)[15][24][34]}{(13)(24)(35)(45)[12][14][45]} + \frac{\langle 12 \rangle^2 (23)[14][23][25]}{(13)(24)(35)(45)[12]^2 [45]} - \frac{\langle 12 \rangle^2 (23)[15][23][24]}{(13)(24)(35)(45)[12]^2 [45]} - \frac{\langle 12 \rangle^2 (23)[25]}{(13)(24)(34)(35)[12]} \\
 & + \frac{\langle 12 \rangle^2 (23)[15][24]}{(12)^2 (23)[15][24]} + \frac{\langle 12 \rangle^2 [24]}{(13)(35)(45)[12]} - \frac{\langle 12 \rangle^2 (34)[13][24][34]}{(13)(24)(35)(45)[12][14][23]} - \frac{\langle 12 \rangle^2 [45]}{(13)(24)(35)[12]} + \frac{\langle 12 \rangle^2 (23)[13][24][45]}{(13)(15)(24)(35)[12][14][15]} \\
 & + \frac{\langle 12 \rangle^2 [24][45]}{(13)(15)(35)[12][15]} + \frac{\langle 12 \rangle^2 (34)[13][24][34][45]}{(13)(15)(24)(35)[12][14][15][23]} - \frac{\langle 12 \rangle (13)[13][34][35]}{(24)(35)(45)[12][23][45]} + \frac{\langle 12 \rangle (13)[13][15][34]^2}{(24)(35)(45)[12][14][23][45]} \\
 & + \frac{\langle 12 \rangle (13)(23)[13][34]}{(12)(13)(23)[13][34]} + \frac{\langle 12 \rangle (13)[34]}{(12)(13)[34]} + \frac{\langle 12 \rangle (13)[13][34]^2}{(12)(13)[13][34]^2} + \frac{\langle 12 \rangle (14)(23)[13][45]}{(12)(14)(23)[13][45]} \\
 & + \frac{\langle 15 \rangle (24)(34)(35)[12][14]}{(15)(24)(34)(35)[12][14]} + \frac{\langle 15 \rangle (34)(35)[12]}{(15)(34)(35)[12]} + \frac{\langle 15 \rangle (24)(35)[12][14][23]}{(15)(24)(35)[12][14][23]} + \frac{\langle 15 \rangle (24)(34)(35)[12][15]}{(15)(24)(34)(35)[12][15]} \\
 & + \frac{\langle 12 \rangle (14)[14][45]}{(12)(14)[14][45]} + \frac{\langle 12 \rangle (14)[13][34][45]}{(12)(15)[13][15][34]} - \frac{\langle 12 \rangle (23)[13][24][34][35]}{(12)(23)[13][24][34][35]} \\
 & + \frac{\langle 15 \rangle (34)(35)[12][15]}{(15)(24)(35)[12][15]} + \frac{\langle 15 \rangle (24)(35)[12][15][23]}{(15)(24)(35)[12][15][23]} - \frac{\langle 24 \rangle (35)(45)[12][14][23]}{(24)(35)(45)[12][14][23]} + \frac{\langle 24 \rangle (35)(45)[12][14][23][45]}{(24)(35)(45)[12][14][23][45]} \\
 & - \frac{\langle 12 \rangle (23)[34][35]}{(12)(23)[34][35]} + \frac{\langle 12 \rangle (23)[15][34]^2}{(24)(35)(45)[12][14][45]} - \frac{\langle 12 \rangle (23)[13][24][35]}{(24)(35)(45)[12]^2 [45]} + \frac{\langle 12 \rangle (23)[14][23][35]}{(24)(35)(45)[12]^2 [45]} \\
 & - \frac{\langle 12 \rangle (23)[35]}{(24)(34)(35)[12]} + \frac{\langle 12 \rangle (23)[15][34]}{(24)(34)(35)[12][14]} + \frac{\langle 12 \rangle [34]}{(35)(45)[12]} + \frac{\langle 12 \rangle (34)[13][34]^2}{(24)(35)(45)[12][14][23]} - \frac{\langle 12 \rangle (23)^2 [13][23][45]}{(15)(24)(34)(35)[12][14][15]} \\
 & + \frac{\langle 12 \rangle (23)^2 [13][24][35]}{(15)(24)(34)(35)[12][14][15]} - \frac{\langle 12 \rangle (23)[23][45]}{(15)(34)(35)[12][15]} + \frac{\langle 12 \rangle (23)[24][35]}{(12)(23)[24][35]} - \frac{\langle 12 \rangle (23)[13][24][34][35]}{(12)(23)[13][24][34][35]} \\
 & + \frac{\langle 15 \rangle (24)(34)(35)[12][14][15]}{(15)(24)(34)(35)[12][14][15]} + \frac{\langle 15 \rangle (34)(35)[12][15]}{(15)(34)(35)[12][15]} + \frac{\langle 15 \rangle (24)(35)[12][14][15][23]}{(15)(24)(35)[12][14][15][23]} \\
 & + \frac{\langle 12 \rangle (23)(45)[13][45]^2}{(12)(23)(45)[13][45]^2} + \frac{\langle 12 \rangle [34][45]}{(12)[34][45]} + \frac{\langle 12 \rangle (45)[45]^2}{(12)(45)[45]^2} + \frac{\langle 12 \rangle (34)[13][34]^2 [45]}{(12)(34)[13][34]^2 [45]} \\
 & + \frac{\langle 15 \rangle (24)(34)(35)[12][14][15]}{(15)(24)(34)(35)[12][14][15]} + \frac{\langle 15 \rangle (35)[12][15]}{(15)(35)[12][15]} + \frac{\langle 15 \rangle (34)(35)[12][15]}{(15)(34)(35)[12][15]} + \frac{\langle 15 \rangle (24)(35)[12][14][15][23]}{(15)(24)(35)[12][14][15][23]} \\
 & + \frac{\langle 12 \rangle (45)[13][34][45]^2}{(12)(45)[13][34][45]^2} + \frac{\langle 12 \rangle (14)(23)[13][24][34]}{(12)(14)(23)[13][24][34]} - \frac{\langle 12 \rangle (14)(23)[34]}{(12)(14)(23)[34]} - \frac{\langle 12 \rangle (14)(23)[45]}{(12)(14)(23)[45]} \\
 & + \frac{\langle 15 \rangle (24)(35)[12][14][15][23]}{(15)(24)(35)[12][14][15][23]} + \frac{\langle 13 \rangle (24)(35)(45)[12][14][23]}{(13)(24)(35)(45)[12][14][23]} - \frac{\langle 13 \rangle (24)(35)(45)[12]}{(13)(24)(35)(45)[12]} - \frac{\langle 13 \rangle (24)(34)(35)[12]}{(13)(24)(34)(35)[12]} \\
 & + \frac{\langle 12 \rangle (14)(23)[14][25]}{(12)(14)(23)[14][25]} - \frac{\langle 12 \rangle (14)(23)[15][24]}{(12)(14)(23)[15][24]} + \frac{\langle 12 \rangle (14)(23)^2 [13][24][45]}{(12)(14)(23)^2 [13][24][45]} + \frac{\langle 12 \rangle (14)(23)[24][45]}{(12)(14)(23)[24][45]} \\
 & + \frac{\langle 13 \rangle (24)(34)(35)[12]^2}{(13)(24)(34)(35)[12]^2} - \frac{\langle 13 \rangle (24)(34)(35)[12]^2 [45]}{(13)(24)(34)(35)[12]^2 [45]} + \frac{\langle 13 \rangle (15)(24)(34)(35)[12][14][15]}{(13)(15)(24)(34)(35)[12][14][15]} \\
 & + \frac{\langle 12 \rangle (14)(23)[13][24][34][45]}{(13)(15)(24)(35)[12][14][15][23]} - \frac{\langle 12 \rangle (15)(23)[15][23]}{(13)(24)(35)(45)[12]^2} + \frac{\langle 12 \rangle (23)^2 [23]^2}{(13)(24)(35)(45)[12]^2} - \frac{\langle 12 \rangle (23)^2 [23][25][34]}{(13)(24)(35)(45)[12]^2 [45]} \\
 & + \frac{\langle 12 \rangle (23)^2 [15][23][24][34]}{(13)(24)(35)(45)[12]^2 [14][45]} - \frac{\langle 12 \rangle (23)^2 [23][25]}{(13)(24)(34)(35)[12]^2} + \frac{\langle 12 \rangle (23)^2 [15][23][24]}{(13)(24)(34)(35)[12]^2 [14]} + \frac{\langle 12 \rangle (23)[23][24]}{(13)(35)(45)[12]^2} \\
 & + \frac{\langle 12 \rangle (23)(34)[13][24][34]}{(13)(24)(35)(45)[12]^2 [14]} + \frac{\langle 12 \rangle (23)[13][24][45]}{(12)(23)[13][24][45]} - \frac{\langle 12 \rangle (23)[23][45]}{(13)(24)(35)[12]^2 [14]} + \frac{\langle 12 \rangle (23)[25][34]}{(13)(24)(35)[12]^2} - \frac{\langle 12 \rangle (23)[15][24][34]}{(13)(24)(35)[12]^2 [14]} \\
 & + \frac{\langle 12 \rangle (23)(45)[25][45]}{(12)(23)(45)[25][45]} - \frac{\langle 12 \rangle (23)(45)[15][24][45]}{(13)(24)(34)(35)[12]^2 [14]} + \frac{\langle 13 \rangle (23)[13][34]^2 [35]}{(24)(35)(45)[12][14][23][45]} + \frac{\langle 13 \rangle (23)^2 [13][34][35]}{(15)(24)(34)(35)[12][14][15]} \\
 & + \frac{\langle 13 \rangle (23)[34][35]}{(13)(23)[34][35]} + \frac{\langle 13 \rangle (23)[13][34]^2 [35]}{(13)(23)[13][34]^2 [35]} + \frac{\langle 14 \rangle (23)[13][34]^2}{(14)(23)[13][34]^2} + \frac{\langle 14 \rangle (23)^2 [13][34][45]}{(14)(23)^2 [13][34][45]} \\
 & + \frac{\langle 15 \rangle (34)(35)[12][15]}{(15)(34)(35)[12][15]} + \frac{\langle 15 \rangle (24)(35)[12][14][15][23]}{(15)(24)(35)[12][14][15][23]} + \frac{\langle 24 \rangle (35)(45)[12][14][23]}{(24)(35)(45)[12][14][23]} + \frac{\langle 15 \rangle (24)(34)(35)[12][14][15]}{(15)(24)(34)(35)[12][14][15]} \\
 & + \frac{\langle 14 \rangle (23)[34][45]}{(14)(23)[34][45]} + \frac{\langle 14 \rangle (23)[13][34]^2 [45]}{(14)(23)[13][34]^2 [45]} - \frac{\langle 15 \rangle (23)[15][35]}{(15)(23)[15][35]} + \frac{\langle 23 \rangle^2 [13][24][34][35]}{(23)^2 [13][24][34][35]} \\
 & + \frac{\langle 15 \rangle (34)(35)[12][15]}{(15)(34)(35)[12][15]} + \frac{\langle 15 \rangle (24)(35)[12][14][15][23]}{(15)(24)(35)[12][14][15][23]} - \frac{\langle 24 \rangle (34)(35)[12]^2 [45]}{(24)(34)(35)[12]^2 [45]} + \frac{\langle 24 \rangle (35)(45)[12]^2 [14][45]}{(24)(35)(45)[12]^2 [14][45]} \\
 & - \frac{\langle 23 \rangle^2 [23][34][35]}{(24)(35)(45)[12]^2 [45]} + \frac{\langle 23 \rangle^2 [13][24][35]}{(24)(34)(35)[12]^2 [14]} + \frac{\langle 23 \rangle [24][35]}{(34)(35)[12]^2} + \frac{\langle 23 \rangle [34][35]}{(24)(35)[12]^2} - \frac{\langle 14 \rangle (15)(23)[15][45]}{(13)(24)(34)(35)[12]^2} \\
 & + \frac{\langle 14 \rangle (23)^2 [13][24][34]}{(13)(24)(35)(45)[12]^2 [14]} - \frac{\langle 14 \rangle (23)^2 [23][34]}{(13)(24)(35)(45)[12]^2} + \frac{\langle 14 \rangle (23)^2 [13][24][45]}{(13)(24)(34)(35)[12]^2 [14]} + \frac{\langle 14 \rangle (23)[24][45]}{(13)(34)(35)[12]^2} + \frac{\langle 14 \rangle (23)[34][45]}{(13)(24)(35)[12]^2}
 \end{aligned}$$



$$\overline{\mathcal{M}} = - \frac{\langle 12 \rangle^3}{\langle 15 \rangle \langle 23 \rangle \langle 34 \rangle \langle 45 \rangle}$$

Example application: 5 gluon amplitude

$$\begin{aligned}
 \mathcal{M} = & \frac{\langle 12 \rangle^3 [13]}{(23)(24)(35)(45)[23]} + \frac{\langle 12 \rangle^3 [14][25]}{(13)(24)(35)(45)[12][45]} - \frac{\langle 12 \rangle^3 [15][24]}{(13)(24)(35)(45)[12][45]} - \frac{\langle 12 \rangle^2 [13][34]}{(24)(35)(45)[14][23]} + \frac{\langle 12 \rangle^2 [13]}{(24)(35)(45)[12]} \\
 & - \frac{\langle 12 \rangle^2 [13][24][35]}{(24)(35)(45)[12][23][45]} + \frac{\langle 12 \rangle^2 [13][25][34]}{(24)(35)(45)[12][23][45]} + \frac{\langle 12 \rangle^2 [14][35]}{(24)(35)(45)[12][45]} - \frac{\langle 12 \rangle^2 [15][34]}{(24)(35)(45)[12][45]} - \frac{\langle 12 \rangle^2 [13][45]}{(23)(24)(35)[12][23]} \\
 & - \frac{\langle 12 \rangle^2 (23)[13][45]}{(15)(24)(34)(35)[14][15]} - \frac{\langle 12 \rangle^2 [45]}{(15)(34)(35)[15]} - \frac{\langle 12 \rangle^2 [13][34][45]}{(15)(24)(35)[14][15][23]} + \frac{\langle 12 \rangle^2 [13][45]}{(15)(24)(35)[12][15]} + \frac{\langle 12 \rangle^2 [14][45]}{(15)(23)(35)[12][15]} \\
 & + \frac{\langle 12 \rangle^2 (34)[13][34][45]}{(12)^2(15)[15]} - \frac{\langle 12 \rangle^2 (23)[23]}{(12)^2(23)[23]} - \frac{\langle 12 \rangle^2 (23)[25][34]}{(12)^2(23)[25][34]} \\
 & + \frac{\langle 15 \rangle (23)(24)(35)[12][15][23]}{(13)(24)(35)(45)[12]} + \frac{\langle 13 \rangle (24)(35)(45)[12]}{(13)(24)(35)(45)[12]} - \frac{\langle 13 \rangle (24)(35)(45)[12][45]}{(13)(24)(35)(45)[12][45]} \\
 & + \frac{\langle 12 \rangle^2 (23)[15][24][34]}{(13)(24)(35)(45)[12][14][45]} + \frac{\langle 12 \rangle^2 (23)[14][23][25]}{(13)(24)(35)(45)[12]^2[45]} - \frac{\langle 12 \rangle^2 (23)[15][23][24]}{(13)(24)(35)(45)[12]^2[45]} - \frac{\langle 12 \rangle^2 (23)[25]}{(13)(24)(34)(35)[12]} \\
 & + \frac{\langle 12 \rangle^2 (23)[15][24]}{(13)(24)(34)(35)[12][14]} + \frac{\langle 12 \rangle^2 [24]}{(13)(35)(45)[12]} + \frac{\langle 12 \rangle^2 (34)[13][24][34]}{(13)(24)(35)(45)[12][14][23]} - \frac{\langle 12 \rangle^2 [45]}{(13)(24)(35)[12]} + \frac{\langle 12 \rangle^2 (23)[13][24][45]}{(13)(15)(24)(35)[12][14][15]} \\
 & + \frac{\langle 12 \rangle^2 [24][45]}{(13)(15)(35)[12][15]} + \frac{\langle 12 \rangle^2 (34)[13][24][34][45]}{(13)(15)(24)(35)[12][14][15][23]} - \frac{\langle 12 \rangle (13)[13][34][35]}{(24)(35)(45)[12][23][45]} + \frac{\langle 12 \rangle (13)[13][15][34]^2}{(24)(35)(45)[12][14][23][45]} \\
 & + \frac{\langle 12 \rangle (13)(23)[13][34]}{(12)(13)(23)[13][34]} + \frac{\langle 12 \rangle (13)[34]}{(12)(13)[34]} + \frac{\langle 12 \rangle (13)[13][34]^2}{(12)(13)[13][34]^2} + \frac{\langle 12 \rangle (14)(23)[13][45]}{(12)(14)(23)[13][45]} \\
 & + \frac{\langle 15 \rangle (24)(34)(35)[12][14]}{(15)(24)(34)(35)[12][14]} + \frac{\langle 15 \rangle (34)(35)[12]}{(15)(34)(35)[12]} + \frac{\langle 15 \rangle (24)(35)[12][14][23]}{(15)(24)(35)[12][14][23]} + \frac{\langle 15 \rangle (24)(34)(35)[12][15]}{(15)(24)(34)(35)[12][15]} \\
 & + \frac{\langle 12 \rangle (14)[14][45]}{(12)(14)[14][45]} + \frac{\langle 12 \rangle (14)[13][34][45]}{(12)(15)[13][15][34]} - \frac{\langle 12 \rangle (23)[13][24][34][35]}{(12)(23)[13][24][34][35]} \\
 & + \frac{\langle 15 \rangle (34)(35)[12][15]}{(15)(24)(35)[12][15]} + \frac{\langle 15 \rangle (24)(35)[12][15][23]}{(15)(24)(35)[12][15][23]} - \frac{\langle 24 \rangle (35)(45)[12][14][23]}{(24)(35)(45)[12][14][23]} + \frac{\langle 24 \rangle (35)(45)[12][14][23][45]}{(24)(35)(45)[12][14][23][45]} \\
 & - \frac{\langle 12 \rangle (23)[34][35]}{(12)(23)[34][35]} + \frac{\langle 12 \rangle (23)[15][34]^2}{(24)(35)(45)[12][14][45]} - \frac{\langle 12 \rangle (23)[13][24][35]}{(24)(35)(45)[12]^2[45]} + \frac{\langle 12 \rangle (23)[14][23][35]}{(24)(35)(45)[12]^2[45]} \\
 & - \frac{\langle 12 \rangle (23)[35]}{(24)(34)(35)[12]} + \frac{\langle 12 \rangle (23)[15][34]}{(24)(34)(35)[12][14]} + \frac{\langle 12 \rangle [34]}{(35)(45)[12]} + \frac{\langle 12 \rangle (34)[13][34]^2}{(24)(35)(45)[12][14][23]} - \frac{\langle 12 \rangle (23)^2 [13][23][45]}{(15)(24)(34)(35)[12][14][15]} \\
 & + \frac{\langle 12 \rangle (23)^2 [13][24][35]}{(15)(24)(34)(35)[12][14][15]} - \frac{\langle 12 \rangle (23)[23][45]}{(15)(34)(35)[12][15]} + \frac{\langle 12 \rangle (23)[24][35]}{(12)(23)[24][35]} - \frac{\langle 12 \rangle (23)[13][24][34][35]}{(12)(23)[13][24][34][35]} \\
 & + \frac{\langle 15 \rangle (24)(34)(35)[12][14][15]}{(15)(24)(34)(35)[12][14][15]} + \frac{\langle 12 \rangle (23)(45)[13][45]^2}{(12)(23)(45)[13][45]^2} + \frac{\langle 12 \rangle [34][45]}{(12)[34][45]} + \frac{\langle 12 \rangle (45)[45]^2}{(12)(45)[45]^2} + \frac{\langle 12 \rangle (34)[13][34]^2[45]}{(12)(34)[13][34]^2[45]} \\
 & + \frac{\langle 15 \rangle (24)(34)(35)[12][14][15]}{(15)(24)(34)(35)[12][14][15]} + \frac{\langle 15 \rangle (35)[12][15]}{(15)(35)[12][15]} + \frac{\langle 15 \rangle (34)(35)[12][15]}{(15)(34)(35)[12][15]} + \frac{\langle 15 \rangle (24)(35)[12][14][15][23]}{(15)(24)(35)[12][14][15][23]} \\
 & + \frac{\langle 12 \rangle (45)[13][34][45]^2}{(15)(24)(35)[12][14][15][23]} + \frac{\langle 12 \rangle (14)(23)[13][24][34]}{(12)(14)(23)[13]} - \frac{\langle 12 \rangle (14)(23)[34]}{(12)(14)(23)[34]} - \frac{\langle 12 \rangle (14)(23)[45]}{(12)(14)(23)[45]} \\
 & + \frac{\langle 12 \rangle (14)(23)[14][25]}{(13)(24)(35)(45)[12][14][23]} + \frac{\langle 13 \rangle (24)(35)(45)[12][14][23]}{(13)(24)(35)(45)[12]} - \frac{\langle 13 \rangle (24)(35)(45)[12]}{(13)(24)(35)(45)[12]} - \frac{\langle 13 \rangle (24)(34)(35)[12]}{(13)(24)(34)(35)[12]} \\
 & + \frac{\langle 12 \rangle (14)(23)[13][24][34][45]}{(13)(15)(24)(34)(35)[12][14][15]} + \frac{\langle 12 \rangle (23)^2 [23]^2}{(13)(24)(35)(45)[12]^2} - \frac{\langle 12 \rangle (23)^2 [23][25][34]}{(13)(24)(35)(45)[12]^2[45]} \\
 & + \frac{\langle 12 \rangle (23)^2 [15][23][24][34]}{(13)(24)(35)(45)[12]^2[14][45]} - \frac{\langle 12 \rangle (23)^2 [23][25]}{(13)(24)(34)(35)[12]^2} + \frac{\langle 12 \rangle (23)^2 [15][23][24]}{(13)(24)(34)(35)[12]^2[14]} + \frac{\langle 12 \rangle (23)[23][24]}{(13)(35)(45)[12]^2} \\
 & + \frac{\langle 12 \rangle (23)(34)[13][24][34]}{(13)(24)(35)(45)[12]^2[14]} + \frac{\langle 12 \rangle (23)[13][24][45]}{(12)(23)[13][24][45]} - \frac{\langle 12 \rangle (23)[23][45]}{(13)(24)(35)[12]^2[14]} + \frac{\langle 12 \rangle (23)[25][34]}{(13)(24)(35)[12]^2} - \frac{\langle 12 \rangle (23)[15][24][34]}{(13)(24)(35)[12]^2[14]} \\
 & + \frac{\langle 12 \rangle (23)(45)[25][45]}{(13)(24)(34)(35)[12]^2} - \frac{\langle 12 \rangle (23)(45)[15][24][45]}{(13)(24)(34)(35)[12]^2[14]} + \frac{\langle 13 \rangle (23)[13][34]^2[35]}{(24)(35)(45)[12][14][23][45]} + \frac{\langle 13 \rangle (23)^2 [13][34][35]}{(15)(24)(34)(35)[12][14][15]} \\
 & + \frac{\langle 13 \rangle (23)[34][35]}{(13)(23)[34][35]} + \frac{\langle 13 \rangle (23)[13][34]^2[35]}{(13)(23)[13][34]^2[35]} + \frac{\langle 14 \rangle (23)[13][34]^2}{(14)(23)[13][34]^2} + \frac{\langle 14 \rangle (23)^2 [13][34][45]}{(14)(23)^2 [13][34][45]} \\
 & + \frac{\langle 14 \rangle (23)[34][45]}{(15)(34)(35)[12][15]} + \frac{\langle 15 \rangle (24)(35)[12][14][15][23]}{(15)(24)(35)[12][14][15][23]} + \frac{\langle 24 \rangle (35)(45)[12][14][23]}{(24)(35)(45)[12][14][23]} + \frac{\langle 15 \rangle (24)(34)(35)[12][14][15]}{(15)(24)(34)(35)[12][14][15]} \\
 & + \frac{\langle 14 \rangle (23)[34][45]}{(15)(34)(35)[12][15]} + \frac{\langle 14 \rangle (23)[13][34]^2[45]}{(15)(24)(35)[12][14][15][23]} - \frac{\langle 15 \rangle (23)[15][35]}{(24)(34)(35)[12]^2} + \frac{\langle 23 \rangle^2 [13][24][34][35]}{(24)(34)(35)[12]^2[14][45]} \\
 & - \frac{\langle 23 \rangle^2 [23][34][35]}{(24)(35)(45)[12]^2[45]} + \frac{\langle 23 \rangle^2 [13][24][35]}{(24)(34)(35)[12]^2[14]} + \frac{\langle 23 \rangle [24][35]}{(34)(35)[12]^2} + \frac{\langle 23 \rangle [34][35]}{(24)(35)[12]^2} - \frac{\langle 14 \rangle (15)(23)[15][45]}{(13)(24)(34)(35)[12]^2} \\
 & + \frac{\langle 14 \rangle (23)^2 [13][24][34]}{(13)(24)(35)(45)[12]^2[14]} - \frac{\langle 14 \rangle (23)^2 [23][34]}{(13)(24)(35)(45)[12]^2} + \frac{\langle 14 \rangle (23)^2 [13][24][45]}{(13)(24)(34)(35)[12]^2[14]} + \frac{\langle 14 \rangle (23)[24][45]}{(13)(34)(35)[12]^2} + \frac{\langle 14 \rangle (23)[34][45]}{(13)(24)(35)[12]^2}
 \end{aligned}$$



$$\overline{\mathcal{M}} = - \frac{\langle 12 \rangle^3}{\langle 15 \rangle \langle 23 \rangle \langle 34 \rangle \langle 45 \rangle}$$

2. S-Matrix bootstrap

The S-matrix is the fundamental object of Quantum Field Theory



- A lot is known about it from perturbation theory (Feynman diagrams)
- Some things are known/conjectured about it non-perturbatively
 - e.g. it should be *unitary* which implies the optical theorem

Optical theorem

$$\text{Im}\langle a|\mathcal{M}|b\rangle = \sum_{|c\rangle} \langle a|\mathcal{M}^\dagger|c\rangle\langle c|\mathcal{M}|b\rangle$$

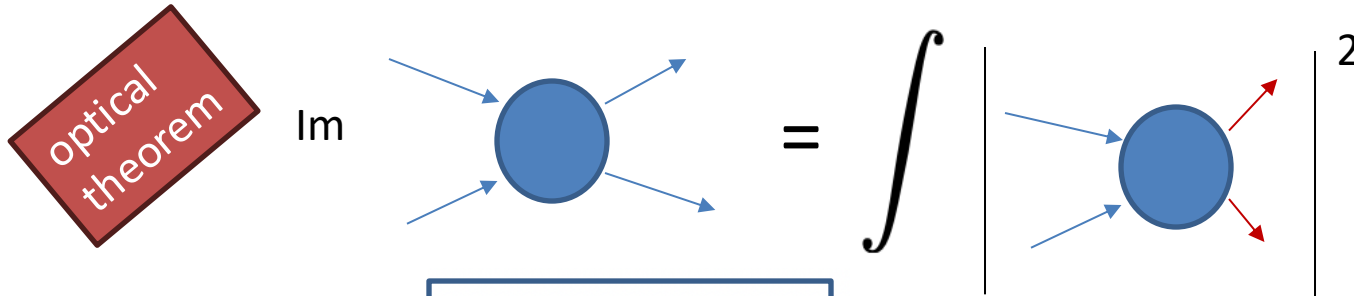
- Non-perturbative constraint
- Relates **complex** scattering amplitude M to **real** cross section $\sigma=M^\dagger M$

Given cross section $\sigma=|M|^2$ can the phase of M be uniquely determined?

Elastic scattering

Unitarity constraint simplifies in the "elastic regime"

- $4 m^2 < s < 9 m^2$
- Only 2-particle states are relevant
- energy conserved, kinematics described by scattering angle $z = \cos \theta$



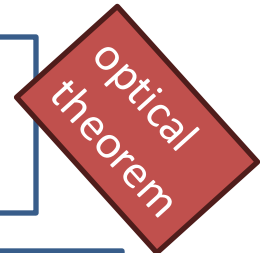
Write the amplitude as $\mathcal{M}(z) = B(z)e^{i\phi(z)}$

Unitarity constraint $\text{Im}\mathcal{M}(z) = \int d\Pi |B|^2$

$$\Rightarrow B(z_{12}) \sin \phi(z_{12}) = \frac{1}{4\pi} \int d\Omega_3 B(z_{13}) B(z_{23}) \cos(\phi(z_{13}) - \phi(z_{23}))$$

$z_2 = z z_1 + \sqrt{1-z^2} \sqrt{1-z_1^2} \cos \phi_1$

$$\Rightarrow \sin \phi(z) = \frac{\int_{-1}^1 dz_1 \int_0^{2\pi} d\phi_1 B(z_1) B(z_2) e^{i\phi(z_1) - i\phi(z_2)}}{4\pi B(z)}$$

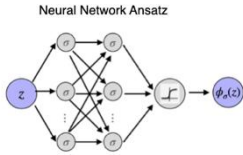


For what $B(z)$ does $\phi(z)$ exist satisfying this equation?
 For what $B(z)$ is $\phi(z)$ unique or not-unique?

Can we find $\phi(z)$ given $B(z)$ with ML? ... Yes!

[Dersy, MDS, Zhiboedov, 2308.09451]

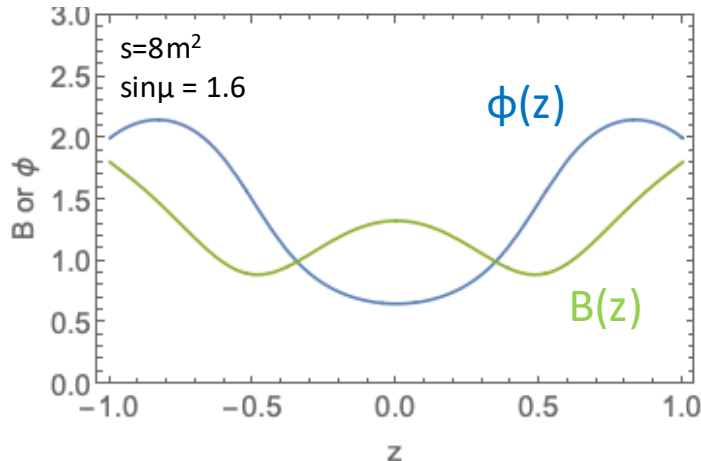
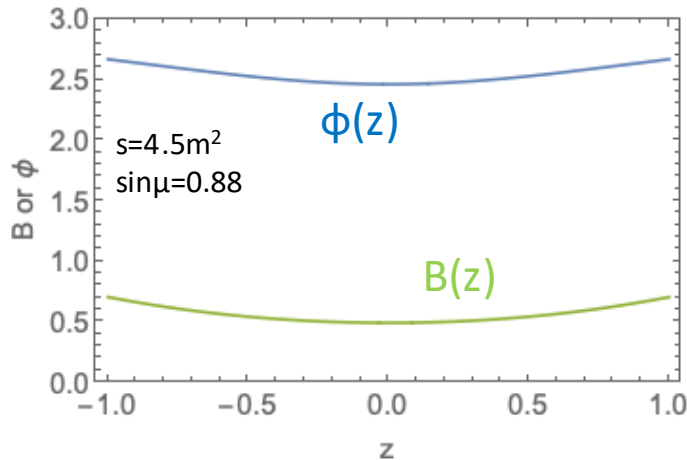
- Parametrize $\phi(z)$ as a neural network



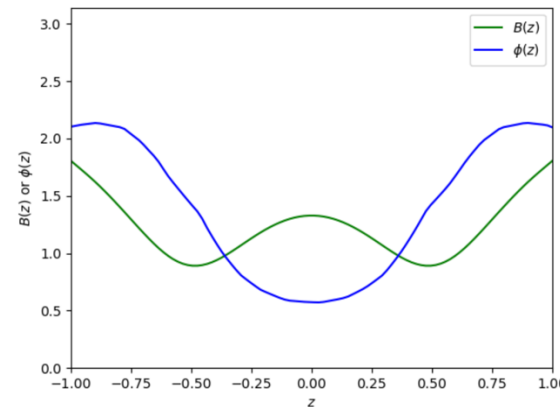
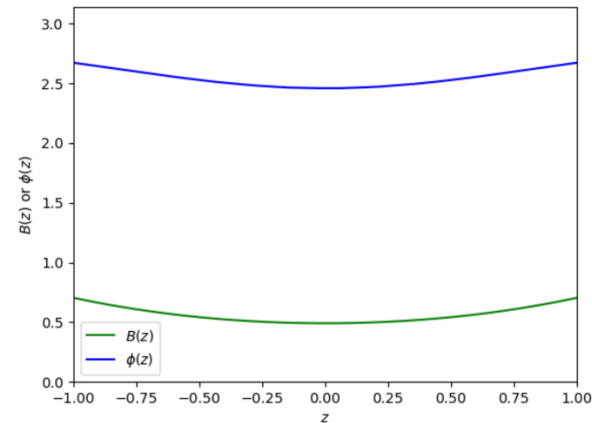
- Loss function is unitarity condition

$$\mathcal{L} = \mathbb{E} \left\| \left| B(z) \sin \phi(z) - \frac{1}{4\pi} \int_{-1}^1 dz_1 \int_0^{2\pi} d\phi_1 B(z_1) B(z_2) \cos(\phi(z_1) - \phi(z_2)) \right| \right\|^2$$

some known examples



find ϕ given B using ML



excellent agreement with known results

When does a phase exist?

In 1967 Andre Martin proposed

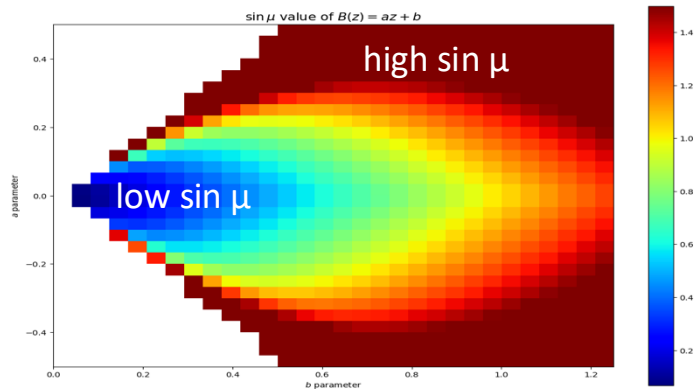
$$\sin \mu = \max_z \frac{\int_{-1}^1 dz_1 \int_0^{2\pi} d\phi_1 B(z_1) B(z_2)}{4\pi B(z)}$$

as an indicator of phase determination

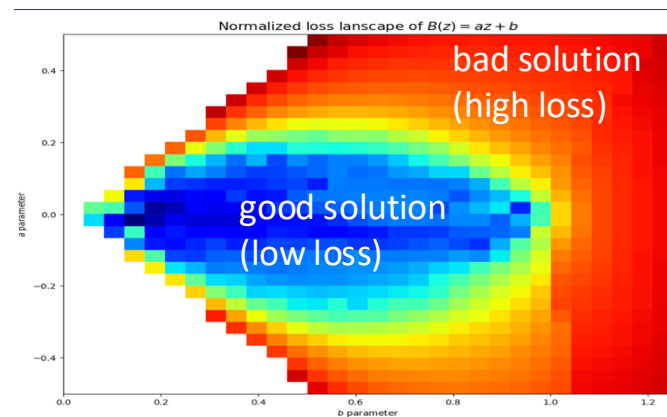
- Martin proved:
 - If $\sin \mu < 1$ for a given $B(z)$ then there always exists a phase $\phi(z)$

What is special about $\sin \mu$?

Contours of $\sin \mu$



Loss landscape from ML search for ϕ



- Loss landscape correlates with $\sin \mu$
 - $\sin \mu$ measures how hard these solutions are to find
 - Don't need exact solutions to learn this lesson

Can there be more than one ϕ given B ?

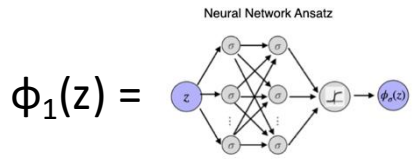
Crichton (1966): yes

Atkinson (1977) found two phases $\phi_1(z)$ and $\phi_2(z)$ for the same $B(z)$ with $\sin\mu > 2.15$

Are there phase-ambiguous amplitudes with $\sin\mu < 2.15$?

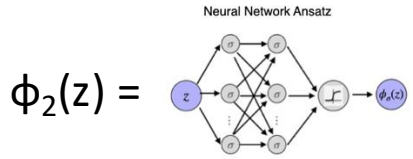
Open question since 1977

ML approach: two NNs for two phases $\phi_1(z)$ and $\phi_2(z)$



- Impose unitarity condition loss for each ϕ :

$$\mathcal{L} = \mathbb{E} \left\| \left| B(z)\sin\phi(z) - \frac{1}{4\pi} \int_{-1}^1 dz_1 \int_0^{2\pi} d\phi_1 B(z_1) B(z_2) \cos(\phi(z_1) - \phi(z_2)) \right| \right\|^2$$

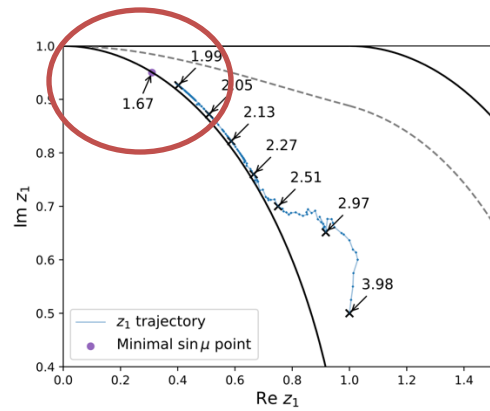


- Add repulsive loss to keep solutions apart

$$\mathcal{L}_R = \mathbb{E}_z \|d(\phi_1(z), \phi_2(z))\|^{-p} + \mathbb{E}_z \|d(\pi - \phi_1(z), \phi_2(z))\|^{-p}$$

Gradient descent in $\sin\mu$ leads to new phase-ambiguous cross sections

First new phase-ambiguous solution in 50 years!



New lowest $\sin\mu$ phase ambiguity

$$\sin\mu = 1.67$$

3. The future



Markus Graf, July, 2023

Current AI models can do college-level physics

Ask Chat GPT 4o

Two blocks with equal mass m are connected with a string. One of the blocks is placed on a table, the other hangs over the edge. Determine the acceleration of the masses.



Assumptions:

1. The pulley is frictionless.
2. The string is inextensible and massless.
3. The only forces acting on the blocks are gravity, tension in the string, and the normal force on the block on the table.

Forces on the Blocks:

- **Block on the table (Block 1):**
 - Tension in the string (T) acts horizontally.
 - The normal force (N) balances the weight of Block 1 (mg).
- **Block hanging off the edge (Block 2):**
 - Tension in the string (T) acts upward.
 - The gravitational force (mg) acts downward.

Equations of Motion:

For Block 1 (on the table):

- Since it only accelerates horizontally:
 $T = ma$

For Block 2 (hanging):

- The net force is the difference between the gravitational force and the tension:
 $mg - T = ma$

Solve for Acceleration:

We have two equations:

$$T = ma \quad (1)$$

$$mg - T = ma \quad (2)$$

Substitute equation (1) into equation (2):

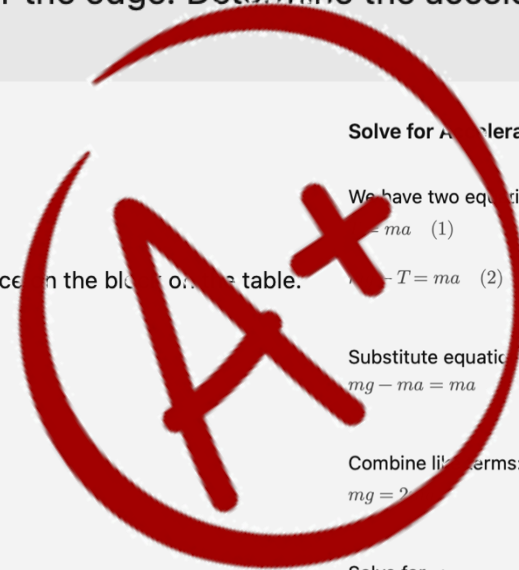
$$mg - ma = ma$$

Combine like terms:

$$mg = 2ma$$

Solve for a :

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



Final Answer:

The acceleration of the masses is $\frac{g}{2}$, where g is the acceleration due to gravity (approximately 9.8 m/s^2).

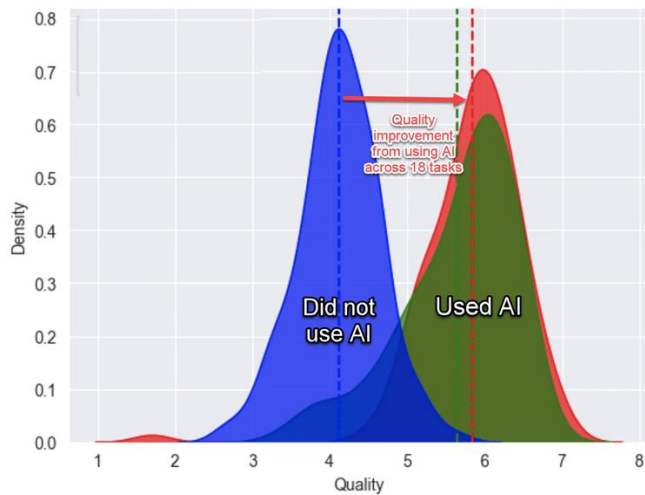
Correct!

- Not only does it get the right answer, but it uses the right equations and shows its work

Augmented intelligence: LLMs help us thrive

Dell'Acqua et al, "Navigating the Jagged Technological Frontier"
(Harvard Business School, 2023)

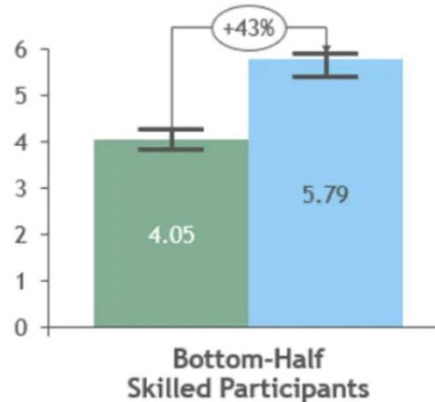
Can LLMs help consultants?



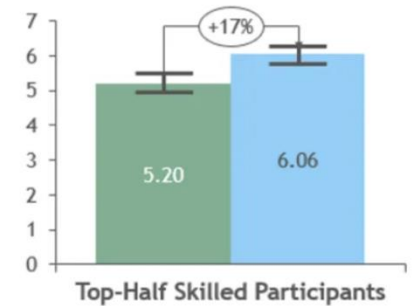
Distribution of output quality across all the tasks. The blue group did not use AI, the green and red groups used AI, the red group got some additional training on how to use AI.

AI is a "skill-leveler"

bottom half improved 43%



top half improved 17%



Yes!
By a lot!

- Mediocre consultants using AI become as good as the best consultants
- The best consultants also improve

10,000 Einsteins

MA

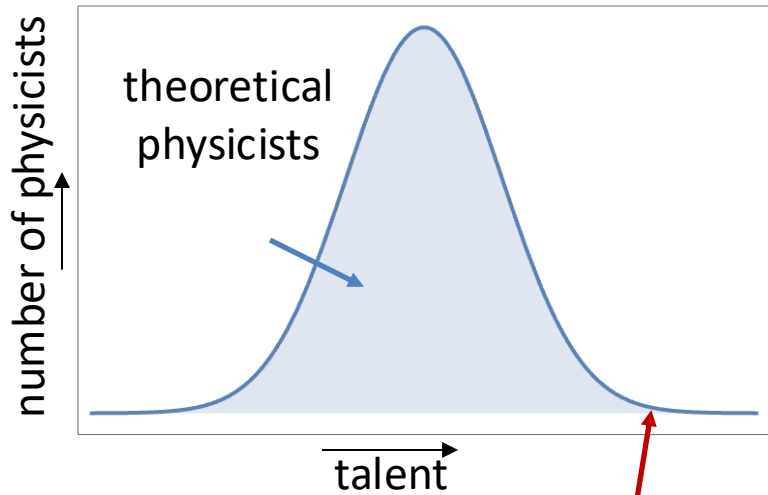
You

is progress in theoretical physics due to the outsized contribution of a small number of individuals?

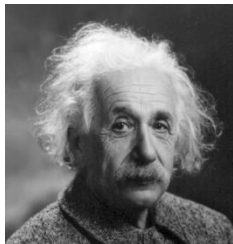
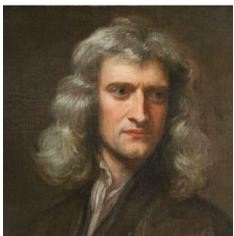


ChatGPT

Yes, progress in theoretical physics, like many fields of science, often appears to be driven by the outsized contributions of a relatively small number of individuals. These key figures often make



exceptional
theoretical physicists



10,000 Einsteins



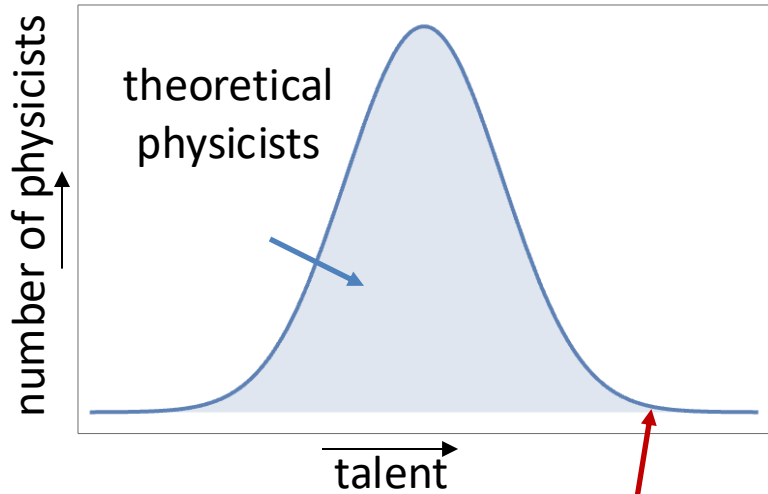
You

is progress in theoretical physics due to the outsized contribution of a small number of individuals?

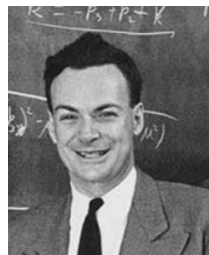
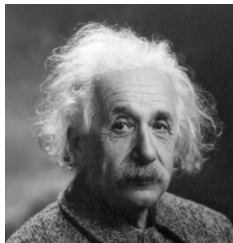
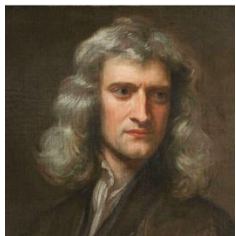


ChatGPT

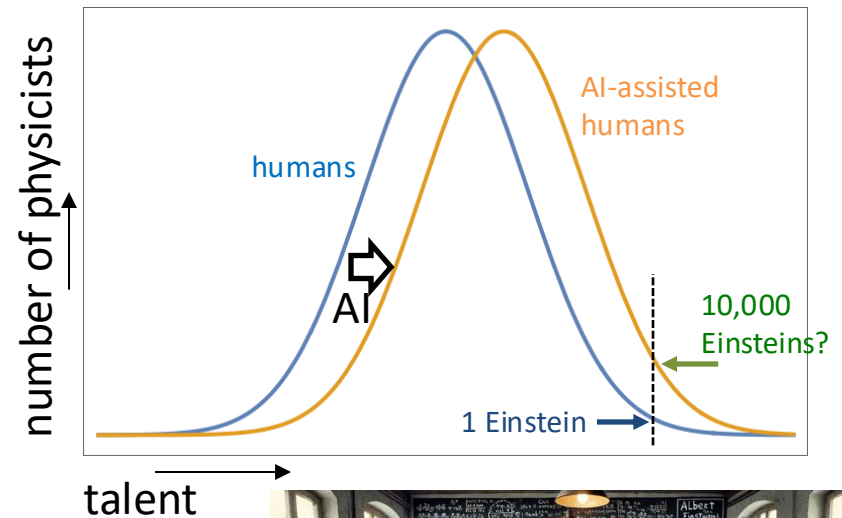
Yes, progress in theoretical physics, like many fields of science, often appears to be driven by the outsized contributions of a relatively small number of individuals. These key figures often make



exceptional theoretical physicists



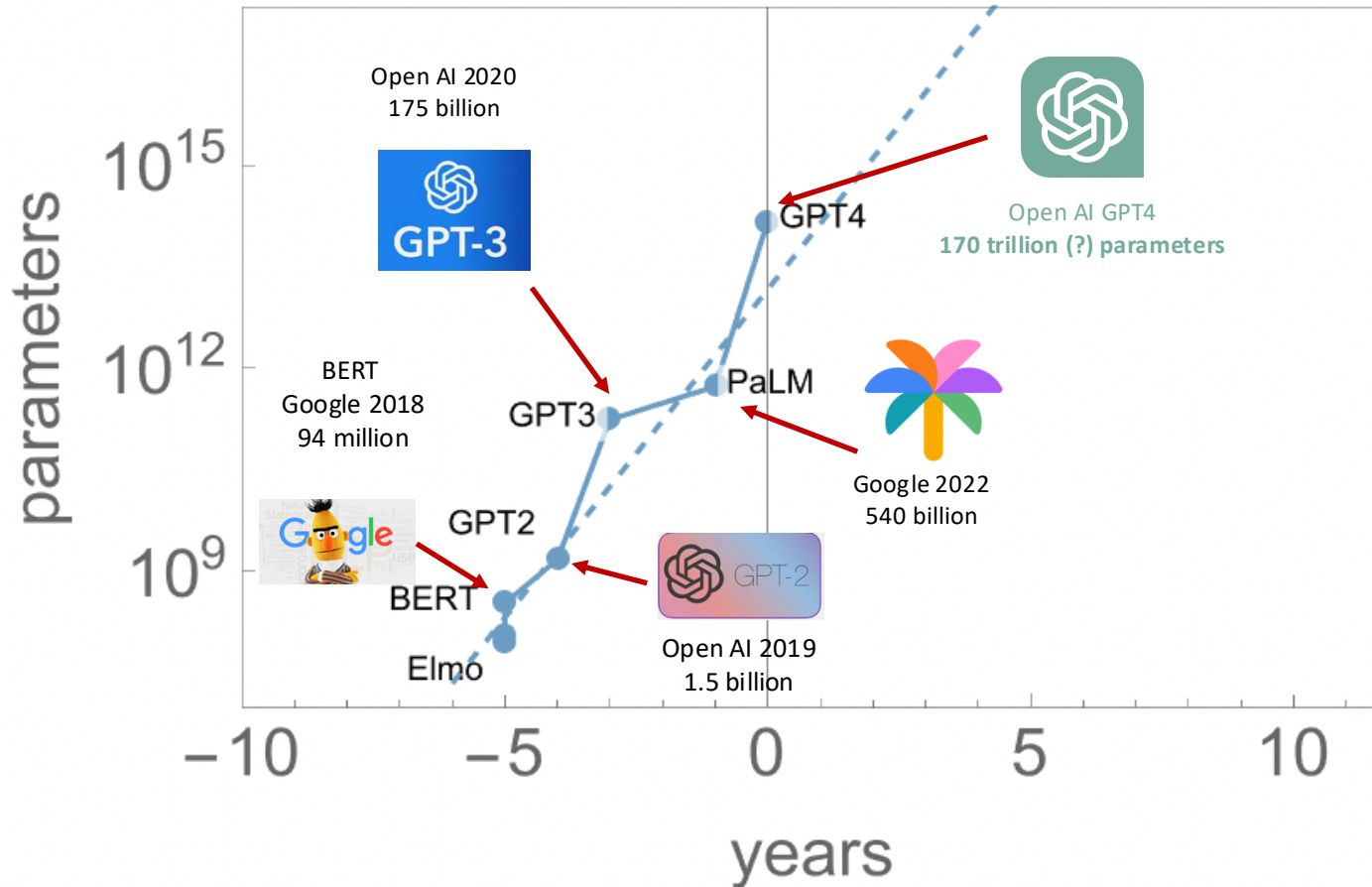
Can AI be a skill-leveler for theoretical physics?



Large Language Models are growing fast

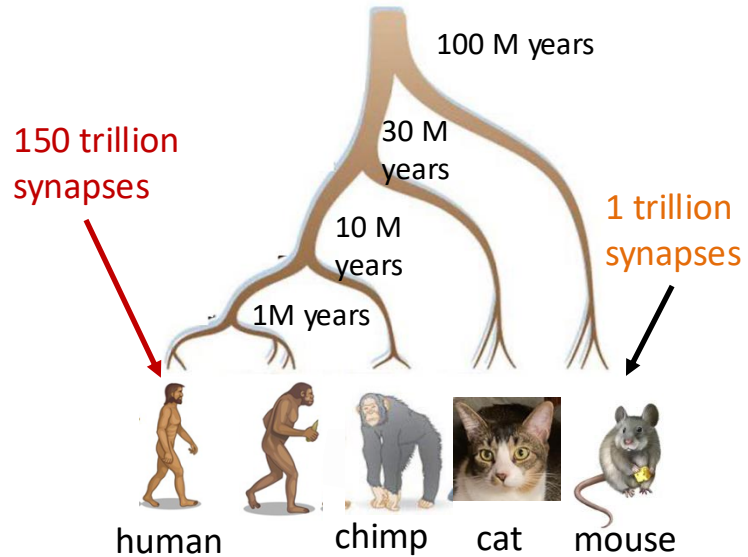
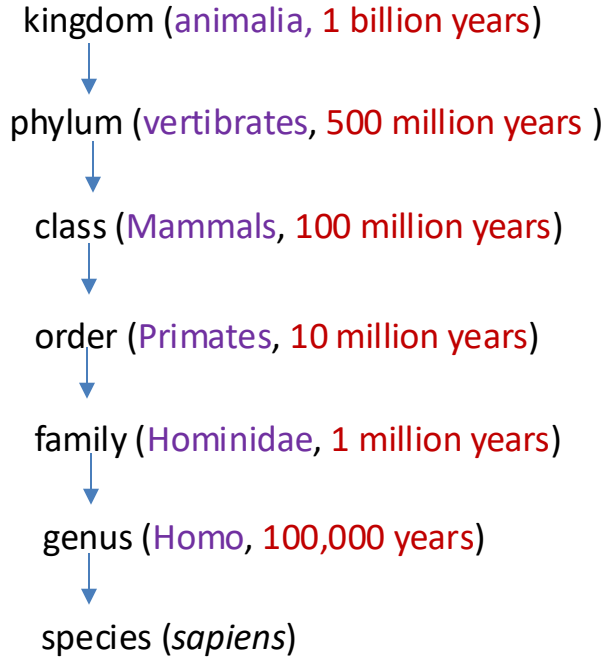
What is a Large Language Model?

- Transformer-based artificial neural network
- Trained on vast amounts of textual data
- Designed to predict the next word (token)



- Exponential growth: LLMs are 10 times more powerful each year!
- # parameters in current LLMs (170 trillion) \approx # synapses in human brain (150 trillion)

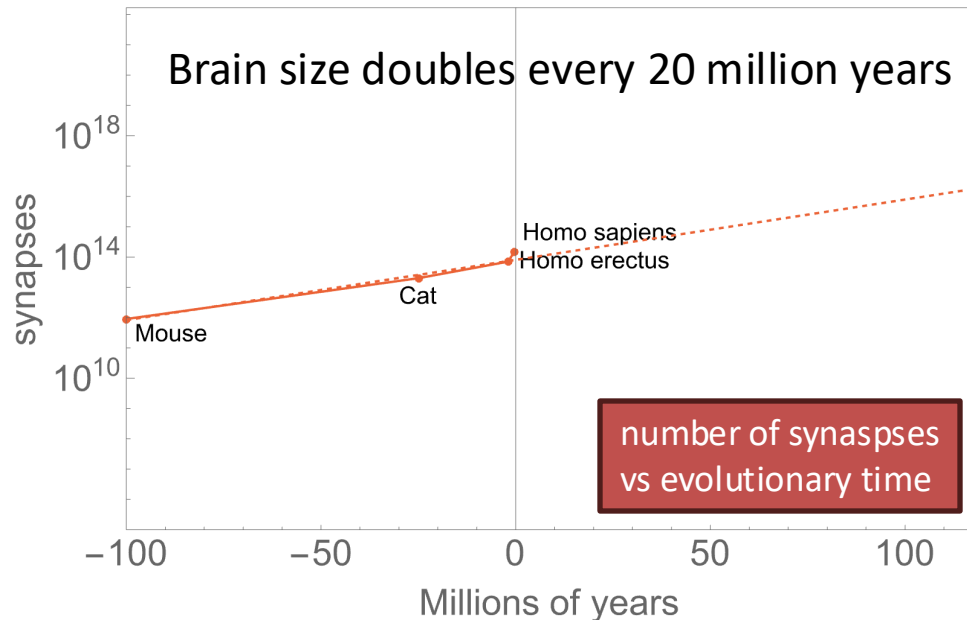
Evolution is slow



- The universe is **only 15 billion** years old
- Only enough time to evolve kingdoms
 - What would evolve after **1 trillion years**?

We are not
the endpoint of intelligence

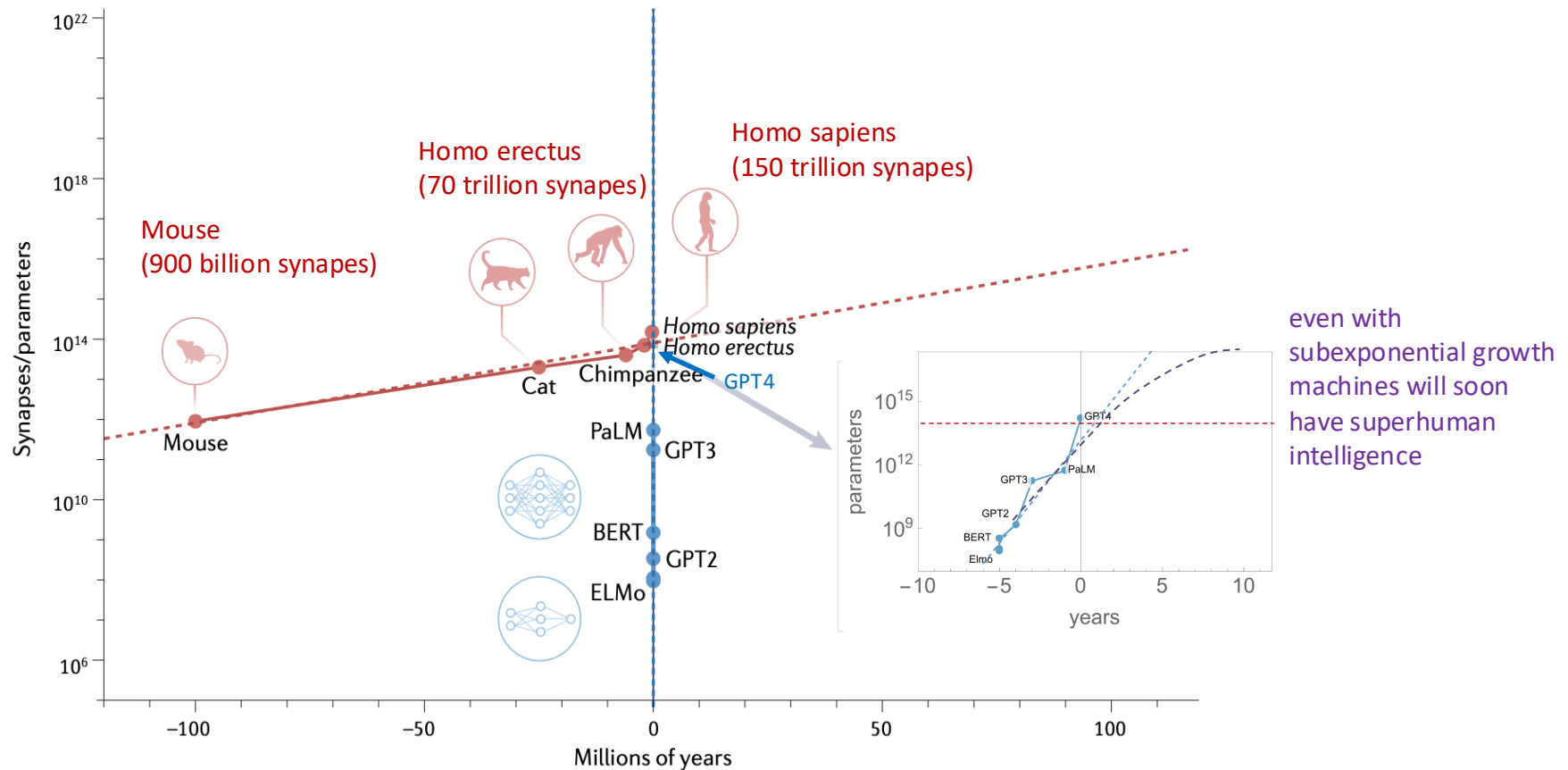
We are only the smartest things
to evolve on earth, so far



Machine vs. Biological intelligence

- Machine intelligence grows by a factor of 10 in 1 year
- Biological intelligence grows by a factor of 2 in 20 million years

MDS, "Should artificial intelligence be interpretable to humans?"
Nature reviews physics (2022)



even with subexponential growth machines will soon have superhuman intelligence

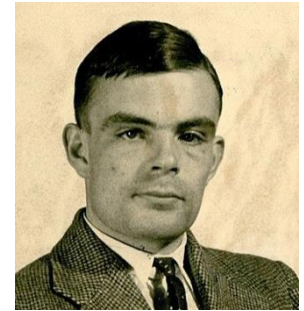
- Both AI and biological intelligence grow exponentially
- **Factor of 10^7 difference in exponent**
- Intersection time, when machines and biology have comparable "intelligence" is **now**

Can AI pass the turing test?

October, 1950

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING



1. *The Imitation Game.*

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as

Turing Test: Using text only, can a person decide if it is talking with a machine or person?

May 2024

[Submitted on 9 May 2024] arXiv:2405.08007

People cannot distinguish GPT-4 from a human in a Turing test

Cameron R. Jones, Benjamin K. Berden



GPT person



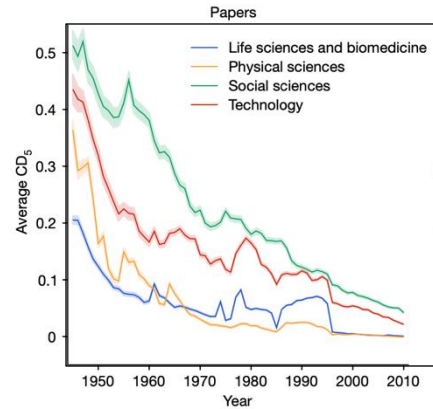
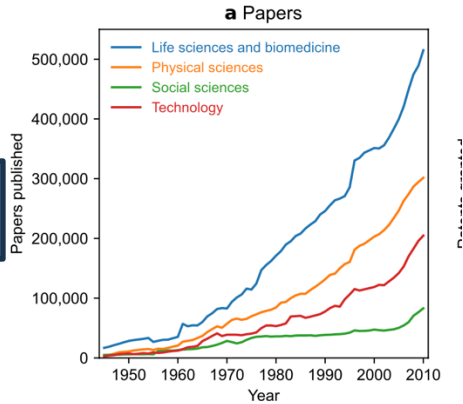
Theoretical particle physics may have stalled nature

Article | Published: 04 January 2023

Papers and patents are becoming less disruptive over time

[Michael Park](#), [Erin Leahey](#) & [Russell J. Funk](#) 

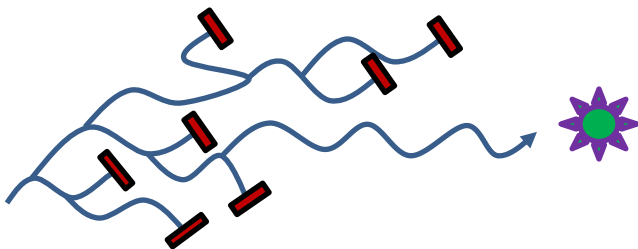
more and more papers are written



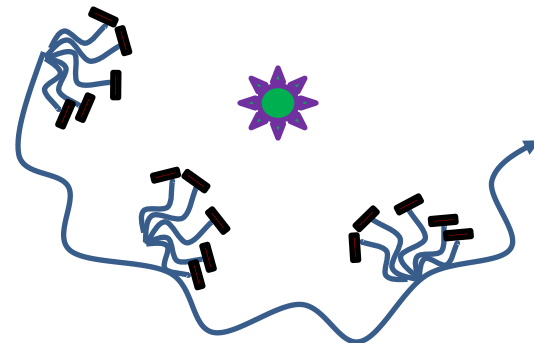
the papers are less and less innovative

Maybe the problems are just too difficult

In the past, we made progress despite many dead ends



Are we even making forward progress anymore?

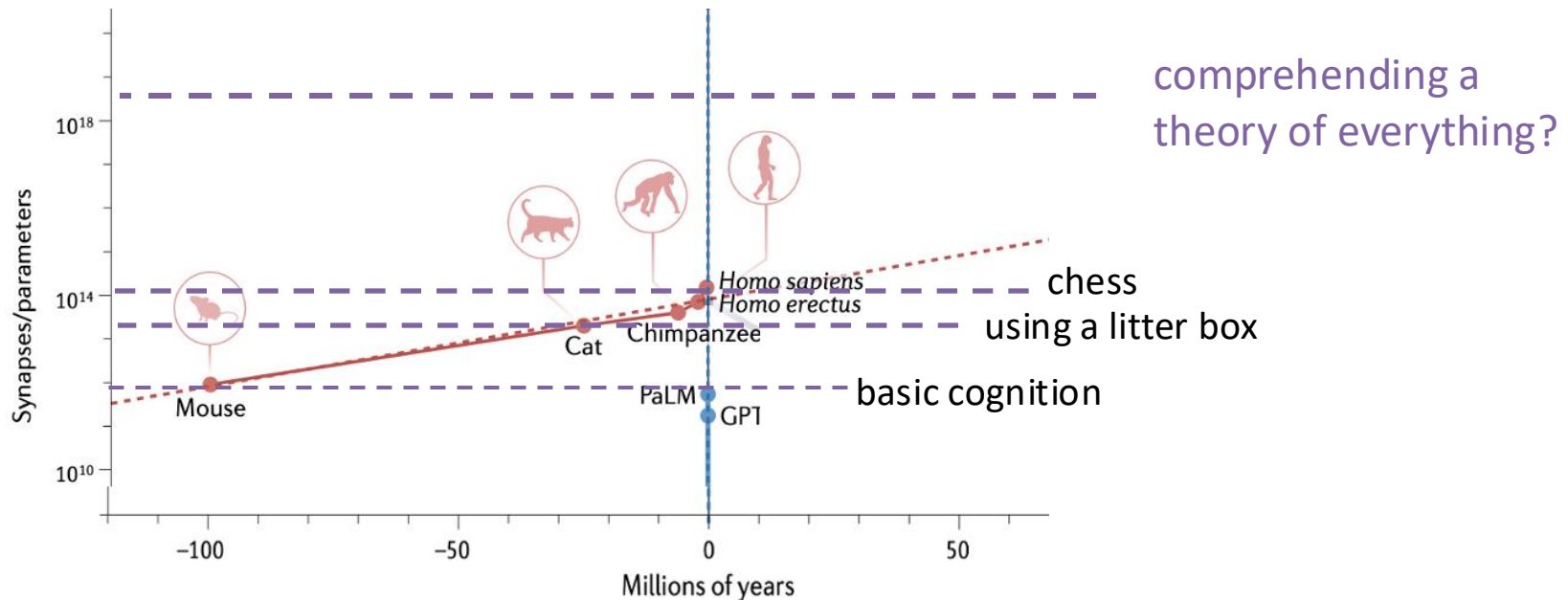


Maybe the problems are just too difficult (for us)



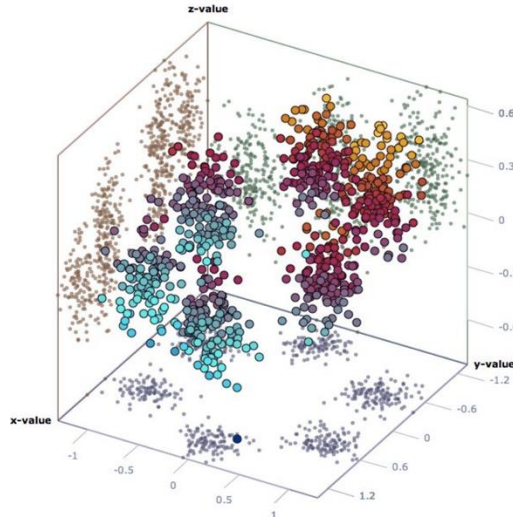
Could a cat ever learn to play chess?

- Humans have limits too
- Why should Homo sapiens be able to understand the theory of everything?

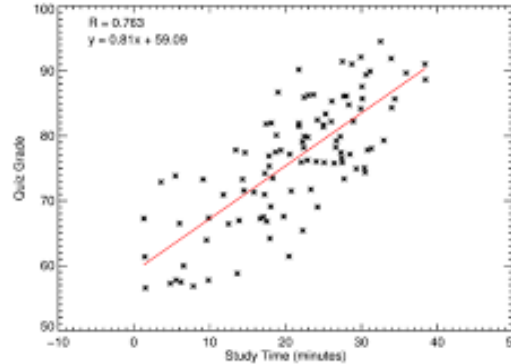


Humans are limited by biology

Humans 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 fundamental 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 at once
- They can understand systems not governed by simple equations

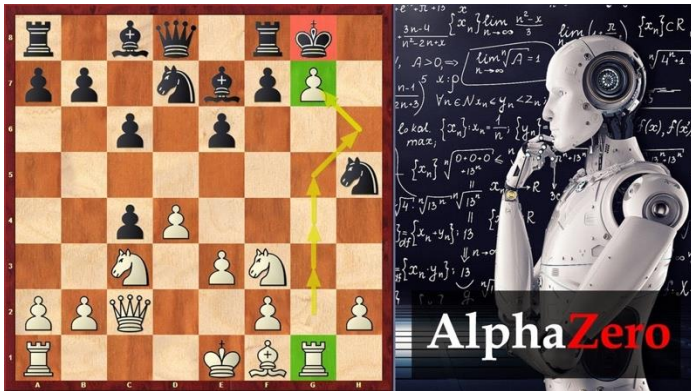
What do we need to progress further?

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

- Trained on solved problems from **books**, physics.stackexchange, chegg etc.
- **Books**, chegg, etc. written by **human beings who read books**, chegg, etc.
- i.e. we generate our own training data

Next step: machines generate toy problems/training data

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



Current LLMs

- can generate and solve problems
- user (human) feedback helps refine model
- LLMs can refine their own code!
 - GPT5 will be written largely by GPT 4

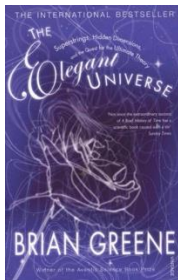
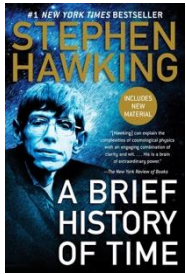
Language models are very close to training themselves to be smarter

Superhuman intelligence

Suppose a machine understands the theory of everything but we don't

- e.g. can calculate electron mass from scratch
- e.g. can explain dark matter

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



- The authors of **Popular science books** understand the details; we just get the general idea

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

- I'm glad that somebody does
- Does it matter that the person is human?

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.

Because of AI, I am now optimistic
for substantive progress in high-
energy theory in my lifetime

Conclusions

- Machine learning is **rapidly transforming high energy physics**
 - Current revolution in applications and advances are in “**data science**”
 - In hep-th and hep-ph problems are largely **symbolic**

1. How do we transition from data science to symbolic theoretical physics?

- It will get easier once we **get started**
 - Symbolic search problems (polylogarithms, spinor helicity)
 - Properties of the S-matrix (unitarity)
 - String Theory Vacua
- } searching for simplicity

2. Generative AI is the future

- Short term: **augmented intelligence**
 - Machines help us organize information
 - Smooth transition to arXAIv: more and more AI input into arXiv papers
- Long term: **artificial intelligence**
 - Machines will suggest problems, solve problems: Chat Ph. D
 - Machines will dumb things down, so we can appreciate their work
 - Superhard problem in theoretical physics may finally be solved

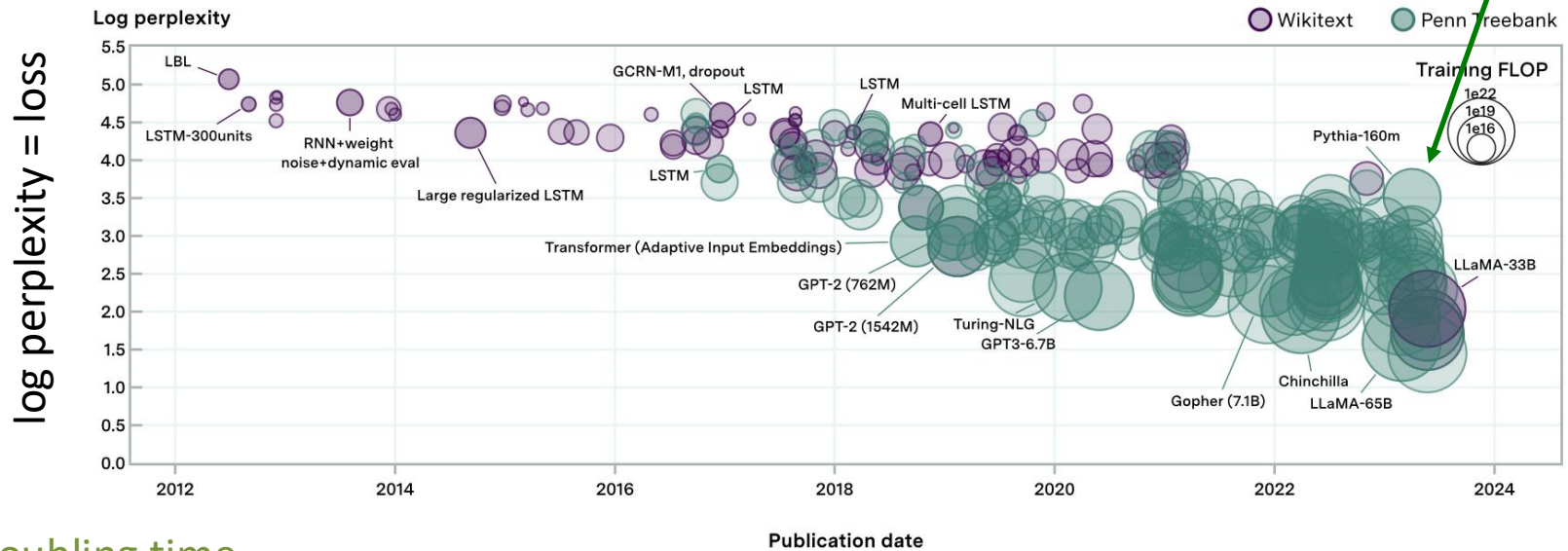
Will we run out of data/compute/energy?

ALGORITHMIC PROGRESS IN LANGUAGE MODELS

Ho et al. arXiv:2403.05812

size is compute

How much of advance is algorithmic progress vs data + compute?



doubling time



algorithmic doubling time = 6 to 14 months!

- controlling for data and network size
- algorithms will continue to get better, especially when written by AI