Centaur Science:

Adventures in AI+Physics

Jesse Thaler

AI4Science Launch, Sorbonne Cluster for Artificial Intelligence — March 9, 2026

The NSF Institute for Artificial Intelligence and Fundamental Interactions (IAIFI /aɪ-faɪ/ iaifi.org)

Artificial intelligence as a pathway to scientific insight

Physics intelligence as a pathway to AI innovation

Launched August 2020

Progress in AI+Physics driven by early career talent with interdisciplinary training

Empowering the Next Generation of AI + Physics Talent

IAIFI Postdoctoral Fellows

Plus 2026-29 fellows to be announced!

Albergo	Boyda	Bright-Thonney	Cuesta	Dogra	Feng	Gagliano	Gerdes	Golubeva	Grosso	Harvey	Luo	Micallef	Mishra-Sharma	Yang
AI and Statistical Physics	⇓	AI for Particle Physics	⇓	Mathematical Physics of AI	AI for Scientific Imagining	AI for Time-Domain Astronomy	AI for Theoretical Physics	⇓	AI for Collider Physics	AI for String Theory	⇓ UCLA	AI for Neutrino Physics	⇓ BU	AI Frontiers of Reinforcement Learning

IAIFI Summer School & Workshop

Photos from 2024 Edition @ MIT

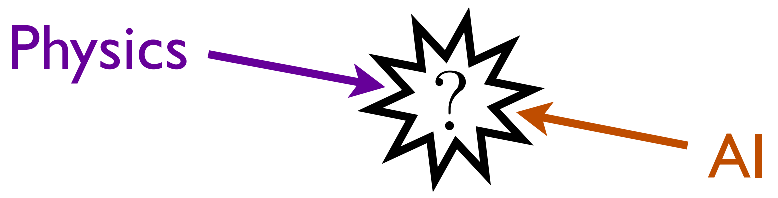
IAIFI

Summer School

August 3–August 7

2026

Lessons from the Physics / AI Collision



To meet the standards of scientific rigor and performance,
we need to teach machines to "Think like a Physicist"

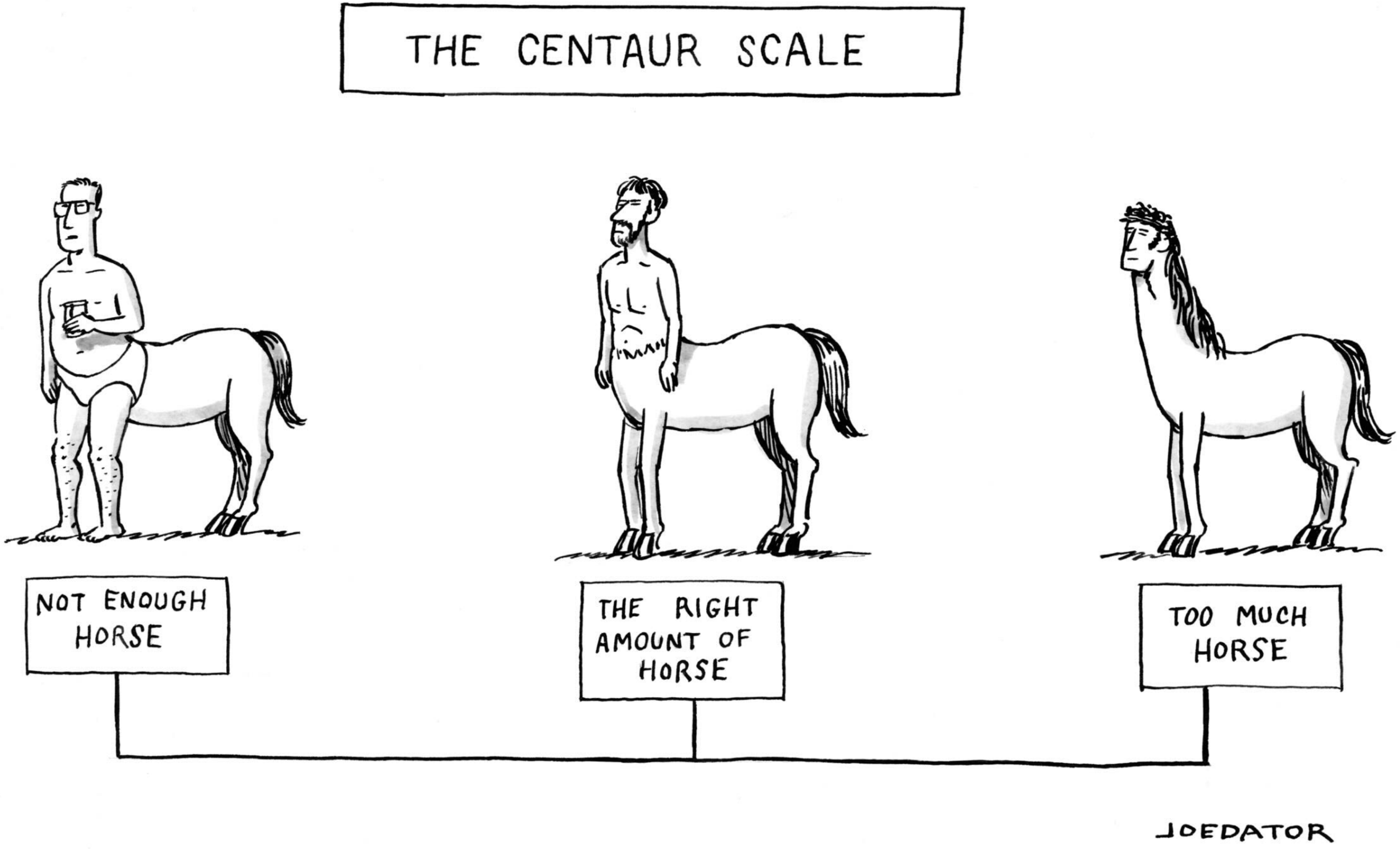
E.g.: symmetries, robustness to systematics, exactness guarantees, statistical inference, …

But to fully capitalize on AI technologies,
we also need to teach physicists to "Think like a Machine"

E.g.: computational complexity, reframing via optimization/search, algorithmic reasoning, …

The Power of "Centaur Science"

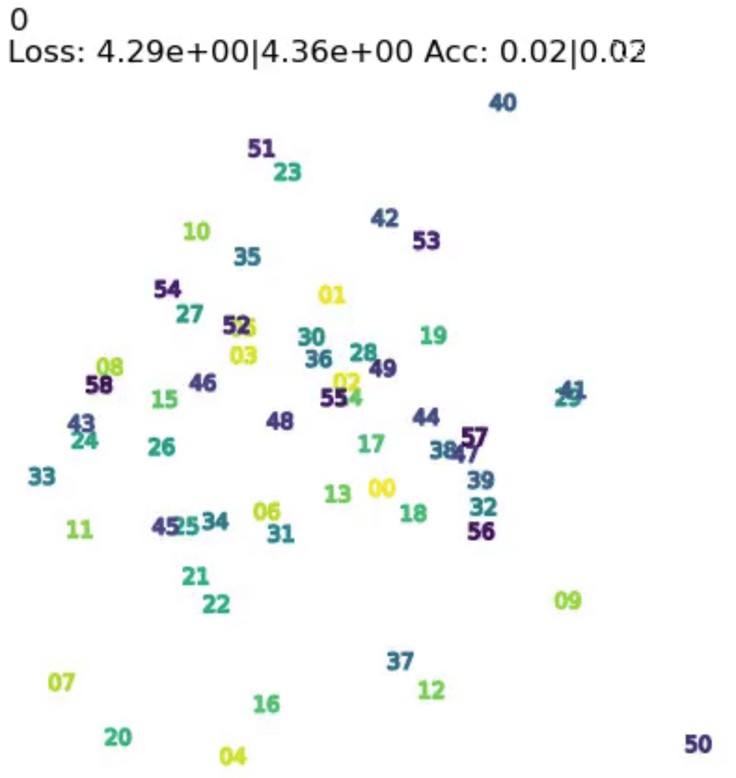
Progress in computation and information theory has long been intertwined with progress in the physical sciences (e.g. statistical mechanics, lattice gauge theory, quantum computers)



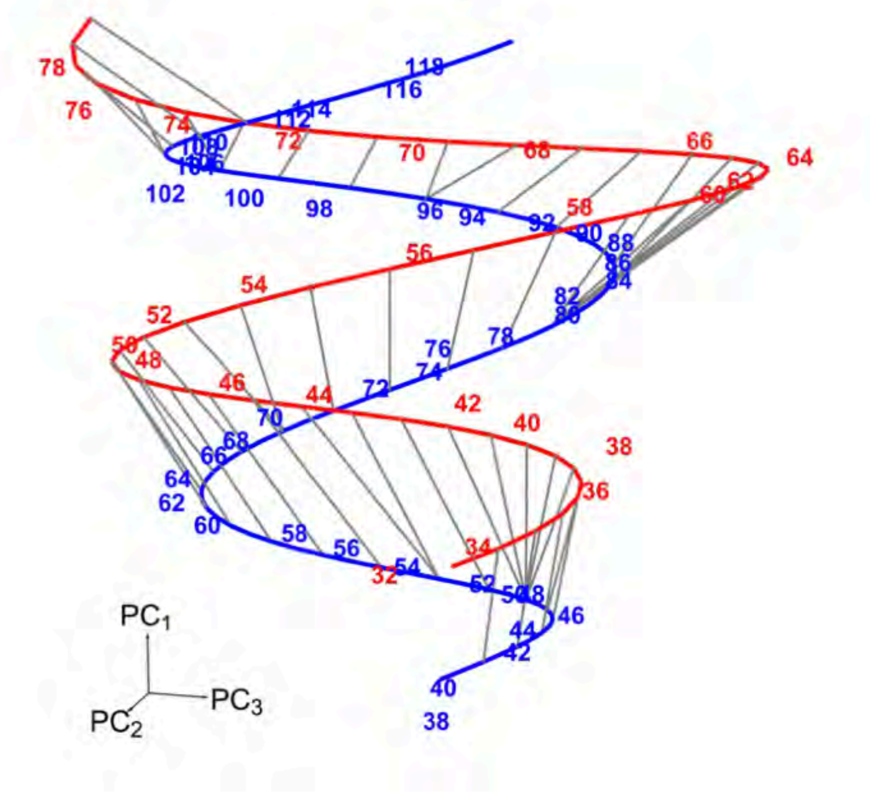
[Joe Dator, The New Yorker 2025; h/t Kyle Cranmer]

Physics for AI… and Back Again

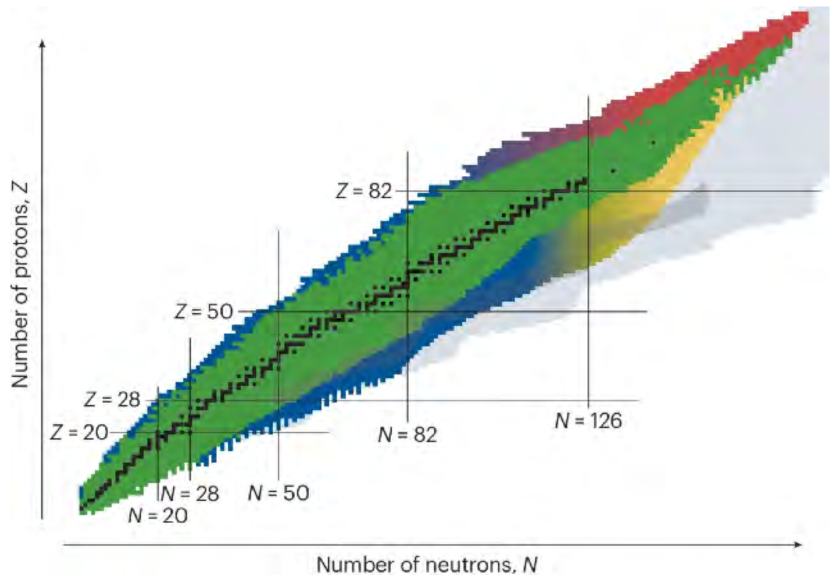
How do machines learn? What do they learn?



Machine-Learned Double Helix



cf. Nuclear Data Tables



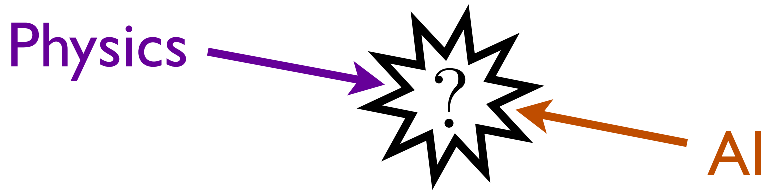
Understanding "grokking" (sudden learning) as phase transition

Predicting nuclear properties with high precision (often better than specialized nuclear models)

IAIFI Interim Director

[Williams, Tegmark, Kitouni, Nolte, Michaud, Liu, Pérez-Díaz, Trifinopoulos, Kantamneni, Richardson, NeurIPS 2022 Spotlight Oral, ICML 2023, ICML 2024, arXiv 2025]

Centaur Science: Adventures in AI+Physics



Machine Learning through the Lens of Physics

AI is transforming physics research, but many foundational aspects of AI can be translated into the language of physics

Scenes from my Sabbatical

Being a "centaur scientist" means capitalizing on emerging AI technologies while insisting on scientific rigor and robustness

Machine Learning through the Lens of Physics

AI is transforming physics research, but many foundational aspects of AI can be translated into the language of physics

Scenes from my Sabbatical

Being a "centaur scientist" means capitalizing on emerging AI technologies while insisting on scientific rigor and robustness

Confronting AI/ML Hype vs. Reality

"AI is a transformative technology"

vs.

"ML is just numerical optimization"

Both are true! And both are changing the scientific enterprise

Direct AI/ML: Enables scientific investigations of high-dimensional spaces
e.g. simulation-based inference

Indirect AI/ML: Capitalizes on emergent behaviors in computational systems
e.g. foundation models, agentic AI

The Future of Artificial Intelligence and the Mathematical and Physical Sciences (AI+MPS)

Community Paper from the NSF Future of AI+MPS Workshop
Cambridge, Massachusetts — March 24–26, 2025

Cross-Disciplinary Opportunities:

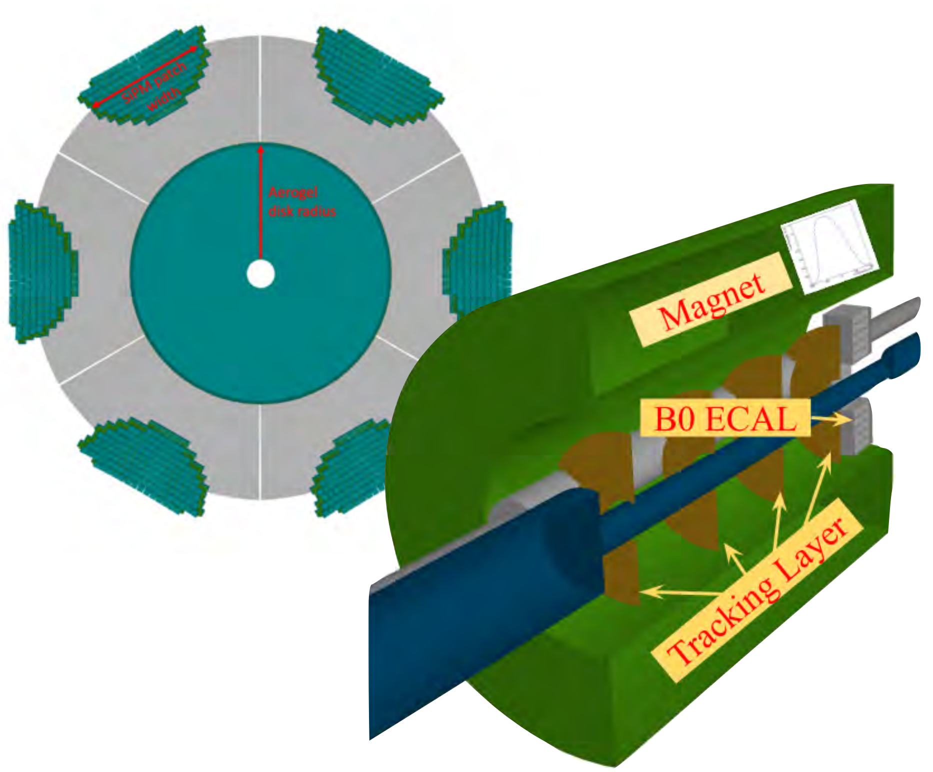
- Advocate for Diverse Funding Streams
- Pursue the Science of AI → AI Innovations from Science; Understanding AI Behaviors; Robust and Reproducible AI
- Establish Scalable AI Infrastructures
- Facilitate Interdisciplinary Collaborations
- Cultivate Key AI Techniques for Science
- Leverage AI for Conducting Research
- Educate and Train an AI+MPS Workforce
- Empower AI Innovation

[Ferguson, LaFleur, Ruthotto, JDT, Ting, Tiwary, Villar, et al., accepted in MLST]

AI is Changing What it Means To…

Representative examples, far from exhaustive!

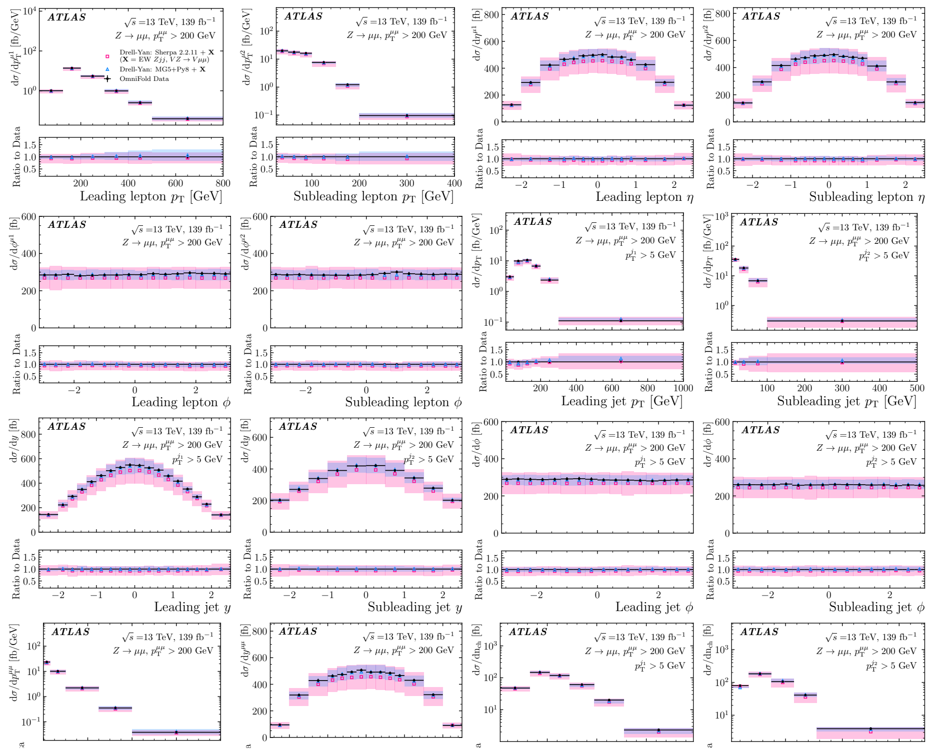
Design/Build/Operate Experiments



e.g. ePIC detector optimization for EIC

[AID(2)E Collaboration, JINST 2024]

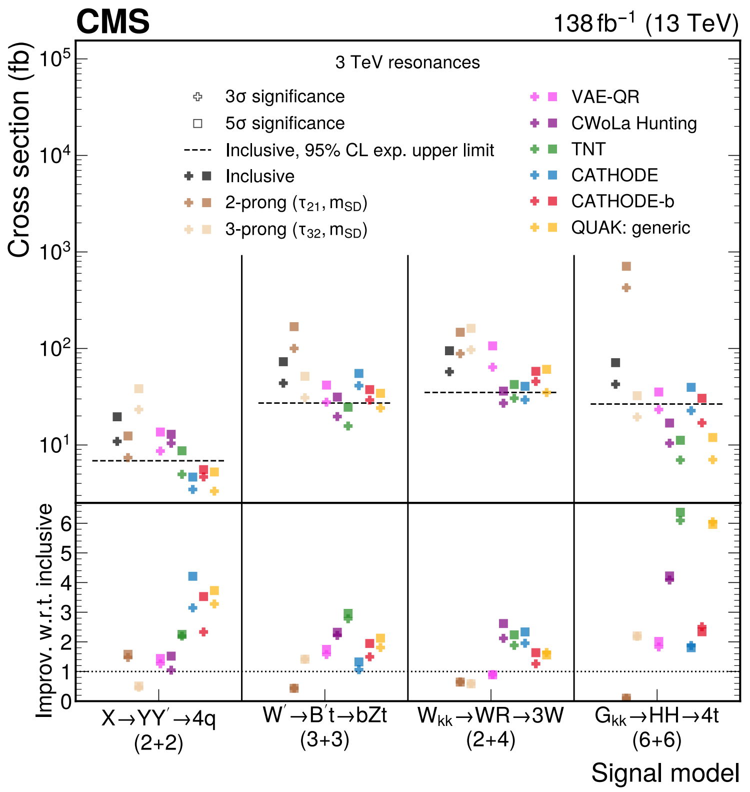
Perform/Report Measurements



e.g. ATLAS 24-dimensional unbinned unfolding

[ATLAS, PRL 2024]

Search for New Physics



e.g. CMS anomalous dijet resonance search

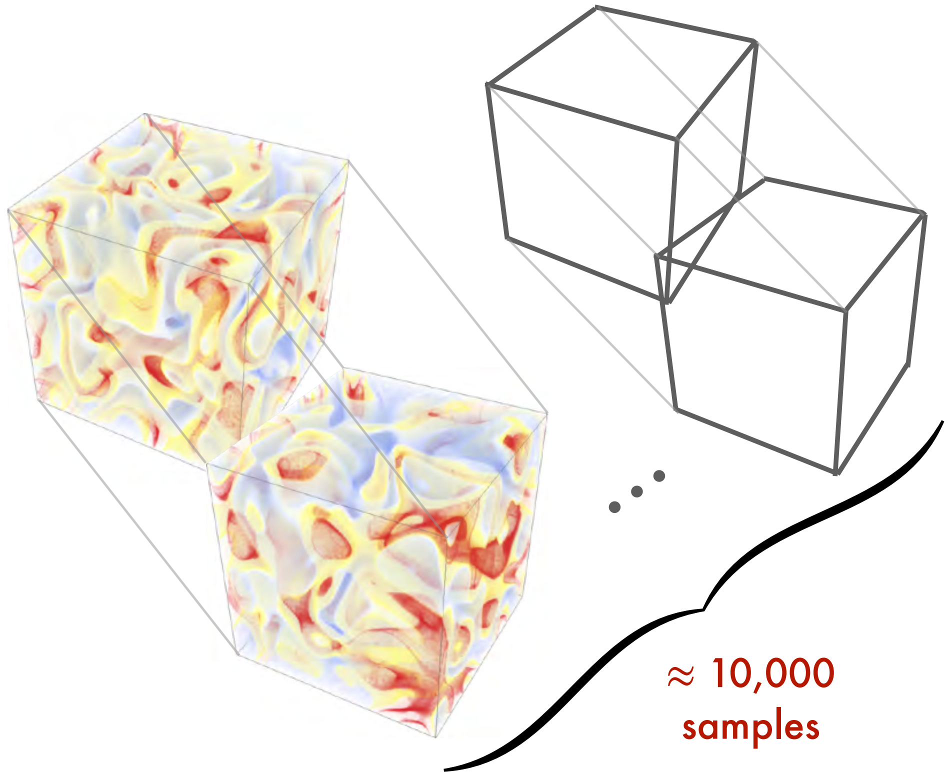
[CMS, RPP 2025]

[from my GGI lectures, January 2026; more examples at HEP ML Living Review]

AI is Changing What it Means To…

Representative examples, far from exhaustive!

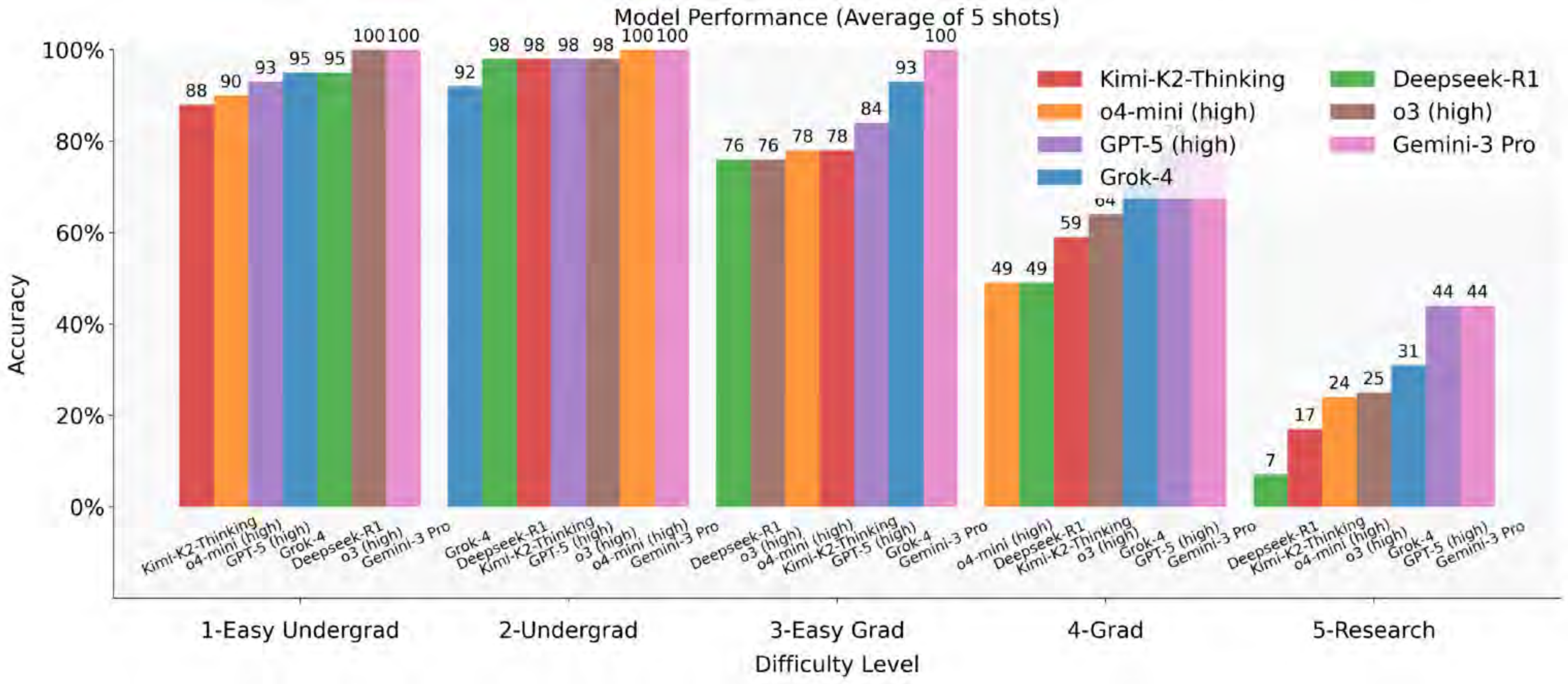
Make Theory Predictions



e.g. generative modeling
for lattice field theory

[Cranmer, Kanwar, Racanière, Rezende, Shanahan, Nature Reviews 2023]

Be a Physicist



e.g. evaluating AI reasoning
abilities on TPBench

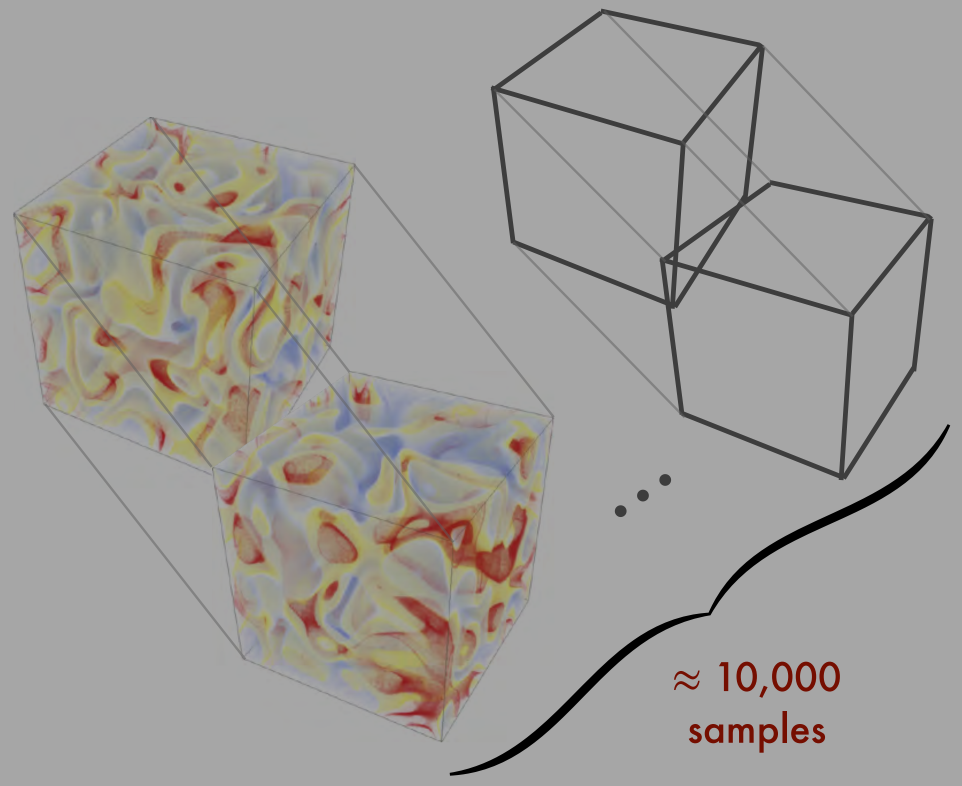
[Chung, Gao, Kvasiuk, Li, Münchmeyer, Rudolph, Sala, Tadepalli, MLST 2025]

[from my GGI lectures, January 2026; more examples at HEP ML Living Review]

AI is Changing What it Means To…

Representative examples, far from exhaustive!

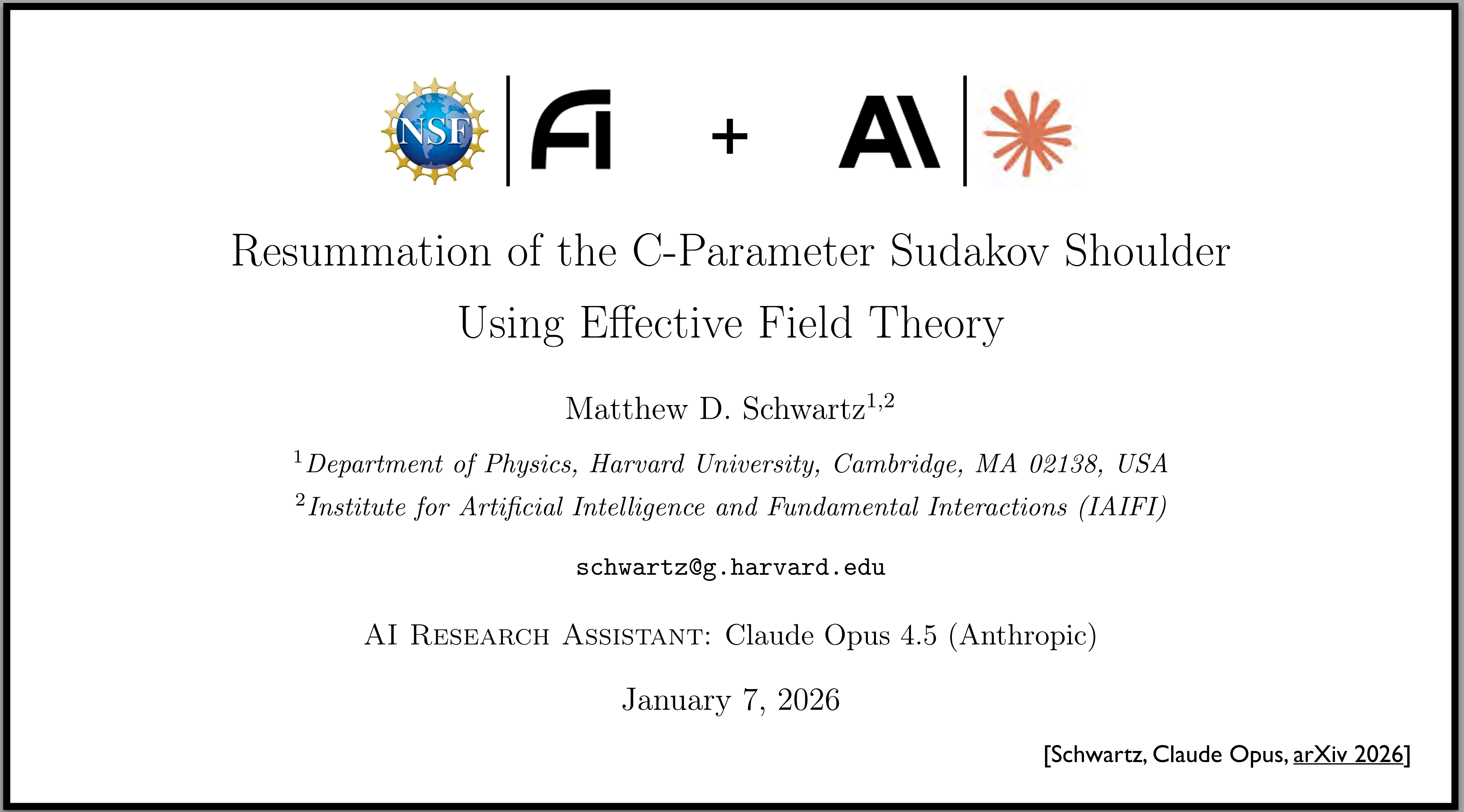
Make Theory Predictions



e.g. generative modeling
for lattice field theory

[Cranmer, Kanwar, Racanière, Rezende, Shanahan, Nature Reviews 2023]

Be a Physicist



Resummation of the C-Parameter Sudakov Shoulder
Using Effective Field Theory

Matthew D. Schwartz[1,2]

[1] *Department of Physics, Harvard University, Cambridge, MA 02138, USA*

[2] *Institute for Artificial Intelligence and Fundamental Interactions (IAIFI)*

`schwartz@g.harvard.edu`

AI Research Assistant: Claude Opus 4.5 (Anthropic)

January 7, 2026

[Schwartz, Claude Opus, arXiv 2026]

[from my GGI lectures, January 2026; more examples at HEP ML Living Review]

ML is "Just" Optimization in Response to Data

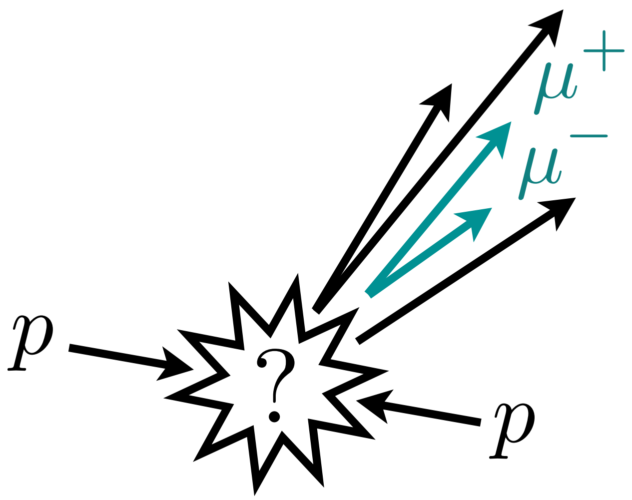
Surprisingly powerful, both directly and indirectly!

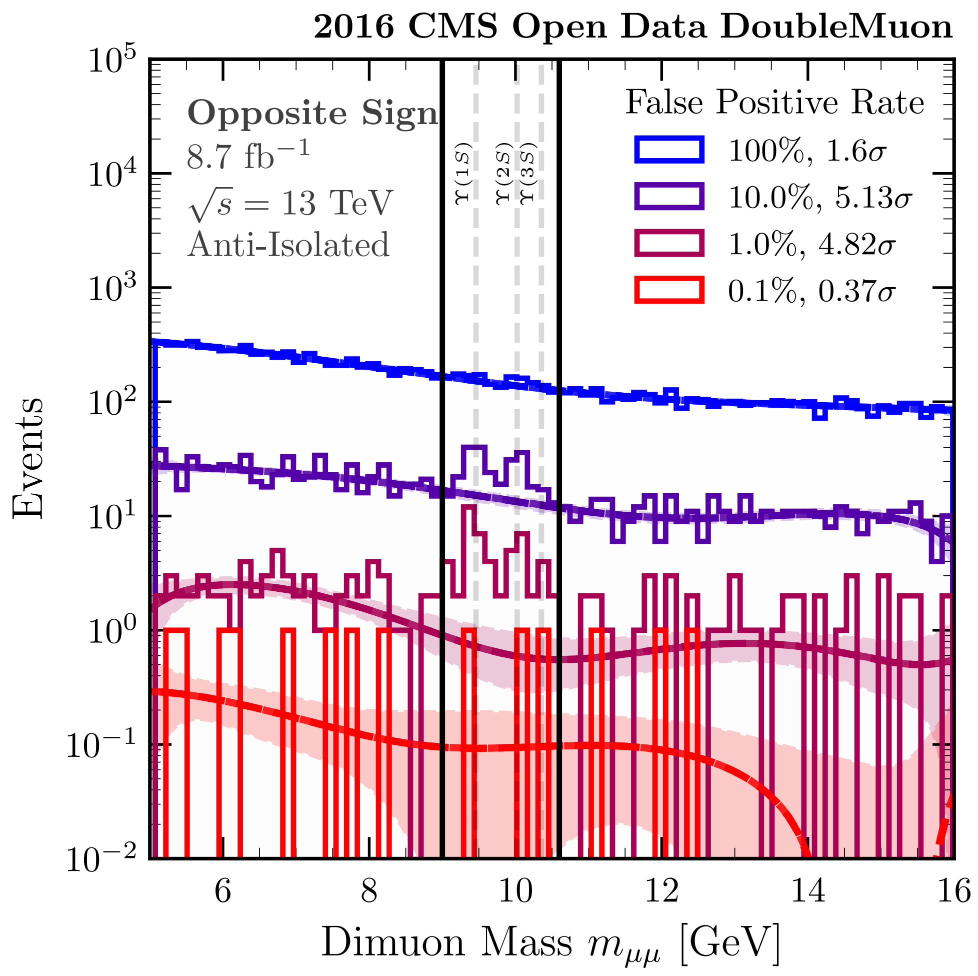
	ML Framing		Physics Translation	
Training Data	$\frac{1}{N}\sum_{i=1}^{N}(\cdots)$	$=$	$\int d\Phi\,(\cdots) + \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$	Monte Carlo Integration
Loss Function(al)	$\mathcal{L}_i[f(\Phi)]$	$\Leftrightarrow$	$\frac{\delta\mathcal{L}}{\delta f} = 0$	Lagrangian Mechanics *e.g. Euler-Lagrange equation*
Learnable Function *e.g. neural networks*	$f_{\mathrm{NN}}(\Phi; b)$	$\Rightarrow$	$f_{\mathrm{physics}}(\Phi; b)$	Physics Knowledge *e.g. symmetries*
Optimizer *e.g. gradient descent*	$\Delta b = -\eta\,\nabla_b\mathcal{L}$	$\Leftrightarrow$	$m\,\ddot{b} + \gamma\,\dot{b} + \frac{dV}{db} = 0$	Newtonian Dynamics *e.g. overdamped oscillator*

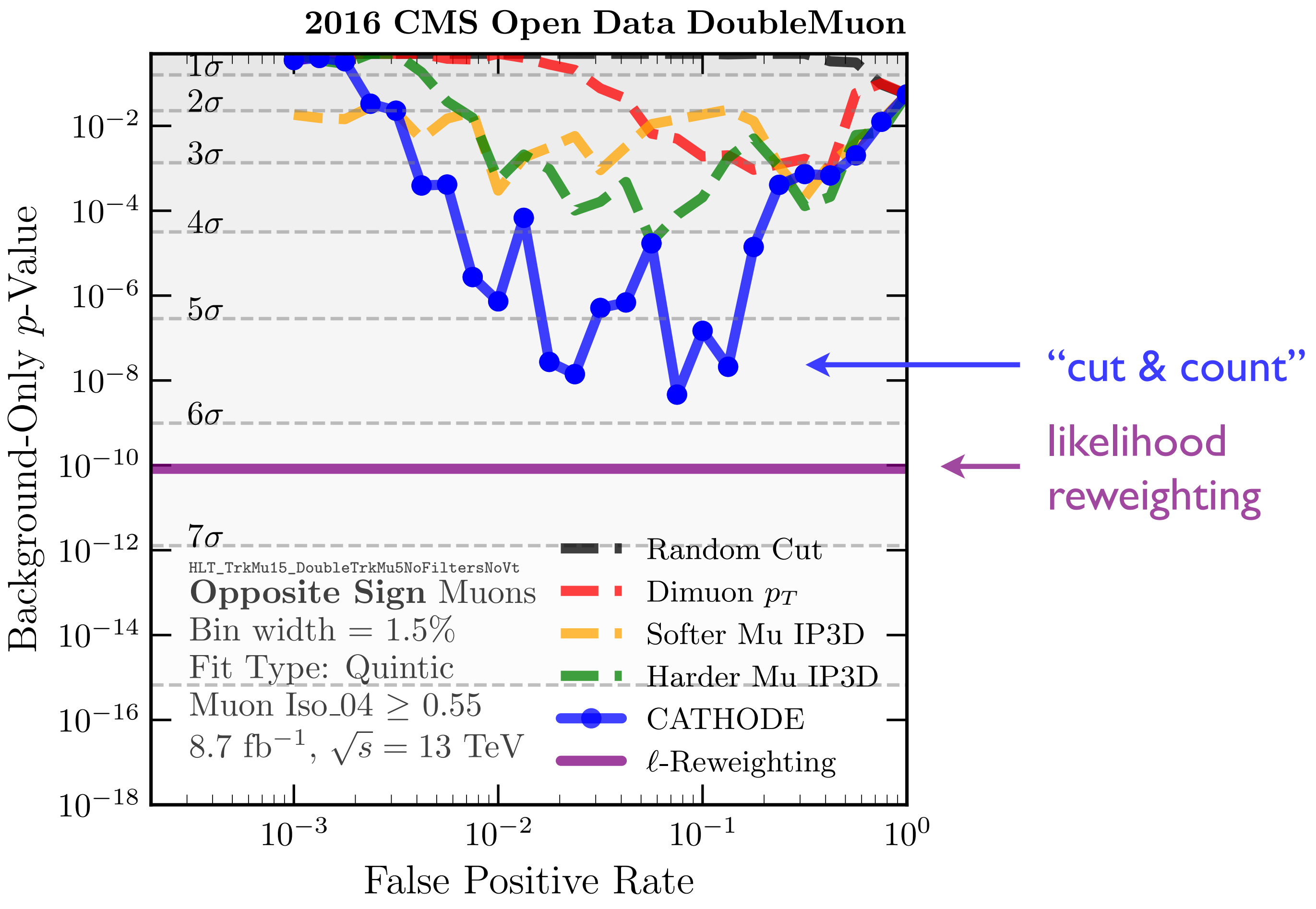
[from my GGI lectures, January 2026]

Direct AI/ML: Rediscovering Upsilons with Anomaly Detection

Background predictions from normalizing flows, signal classification through boosted decision trees…







…with statistical significance estimated using physics domain knowledge about localized resonances

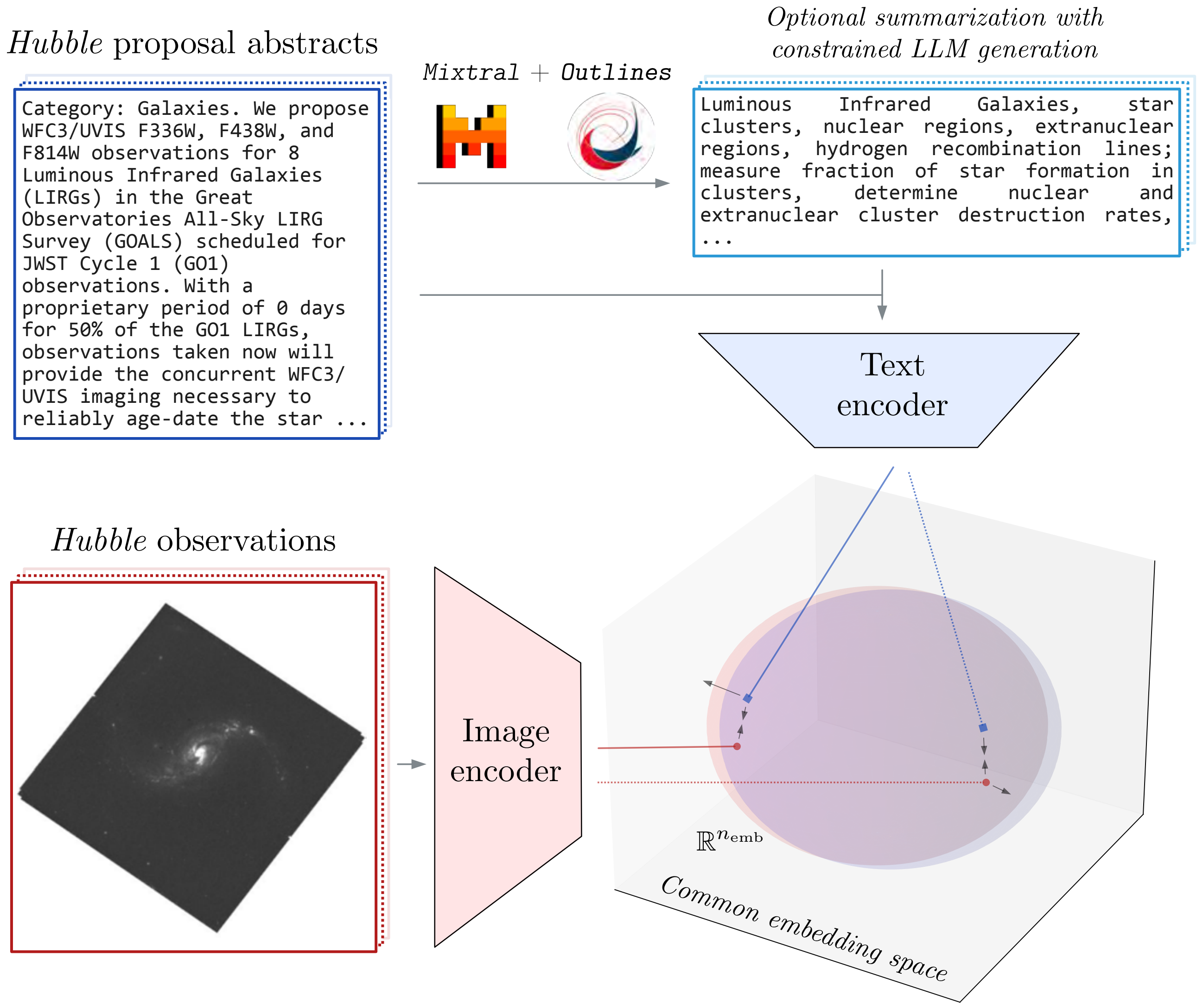
[Gambhir, Mastandrea, Nachman, JDT, PRL 2025;
using CATHODE in Hallin, Isaacson, Kasieczka, Krause, Nachman, Quadfasel, Schlaffer, Shih, Sommerhalder, PRD 2022]



Indirect AI/ML: Contrastive Learning for Hubble Data

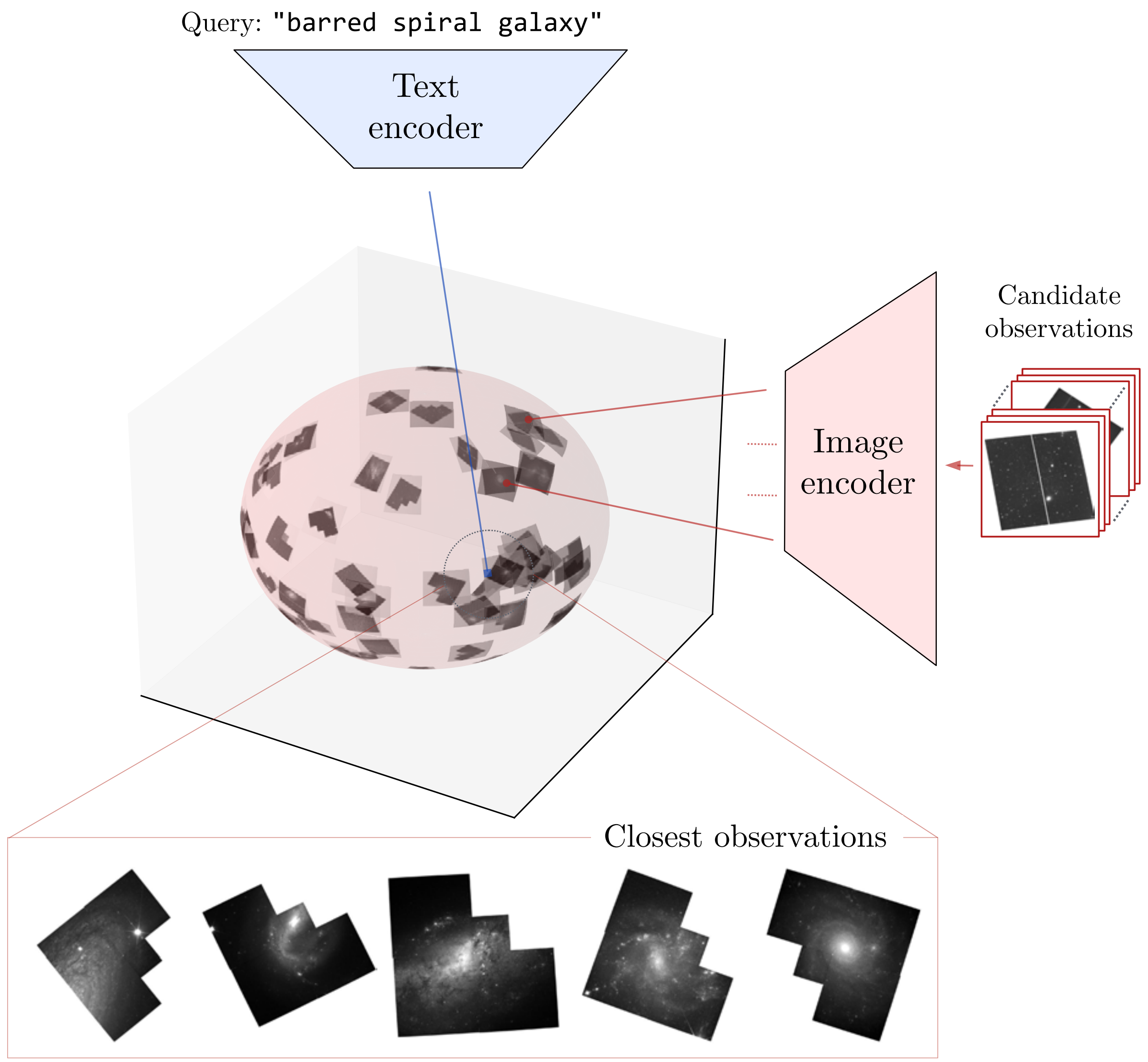
CLIP fine-tuning

With Hubble observation–proposal abstract pairs



Downstream task: observation retrieval

Given natural language text query



[PAPERCLIP in Mishra-Sharma, Song, JDT, COLM 2024;
based on CLIP in Radford, Kim, Hallacy, Ramesh, Goh, Agarwal, Sastry, Askell, Mishkin, Clark, Krueger, Sutskever, PMLR 2021;
see also AstroCLIP in Parker, Lanusse, Golkar, Sarra, Cranmer, Bietti, Eickenberg, Krawezik, McCabe, Ohana, Pettee, Regaldo-Saint Blancard, Tesileanu, Cho, Ho, MNRAS 2024]

Indirect AI/ML: Contrastive Learning for Hubble Data

"What is this a picture of?"

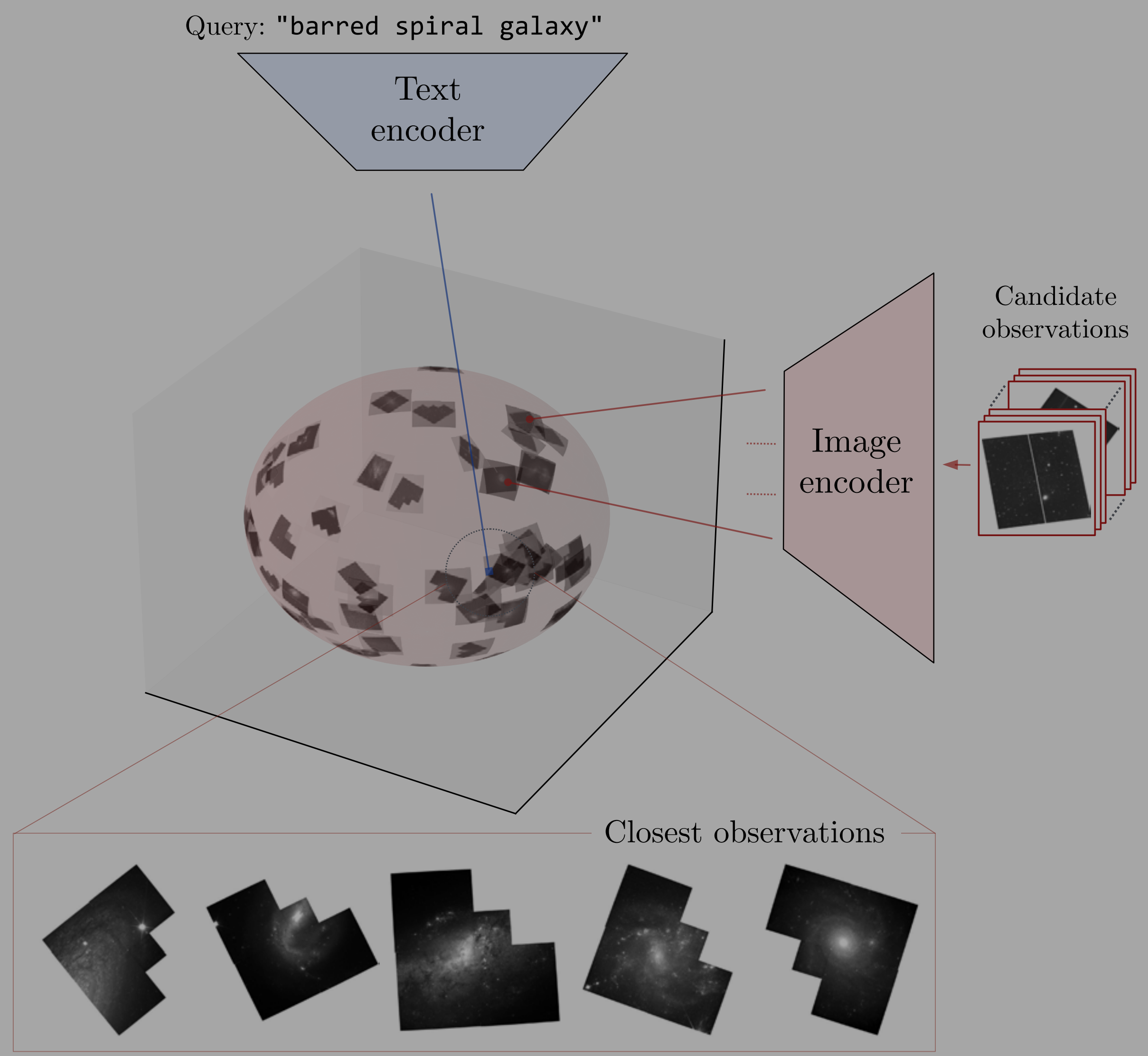
CLIP-ViT-B/16: "gravitational lens"

Fine-Tuned: "supernova remnant"

[image from Supernova 1987A INTensive Survey, HST Proposal 11653]

Downstream task: observation retrieval

Given natural language text query



[PAPERCLIP in Mishra-Sharma, Song, JDT, COLM 2024;
based on CLIP in Radford, Kim, Hallacy, Ramesh, Goh, Agarwal, Sastry, Askell, Mishkin, Clark, Krueger, Sutskever, PMLR 2021;
see also AstroCLIP in Parker, Lanusse, Golkar, Sarra, Cranmer, Bietti, Eickenberg, Krawezik, McCabe, Ohana, Pettee, Regaldo-Saint Blancard, Tesileanu, Cho, Ho, MNRAS 2024]

The rise of centaur scientists who synthesize physics intelligence and artificial intelligence

Machine Learning through the Lens of Physics

AI is transforming physics research, but many foundational aspects of AI can be translated into the language of physics

Scenes from my Sabbatical

Being a "centaur scientist" means capitalizing on emerging AI technologies while insisting on scientific rigor and robustness

Question I've been wrestling with on sabbatical:
What does it mean to do robust discovery science with AI?

Exactness guarantees?
e.g. lattice sampling

Explicit verification?
e.g. symbolic engines

Statistical verification?
e.g. calibrated confidence intervals

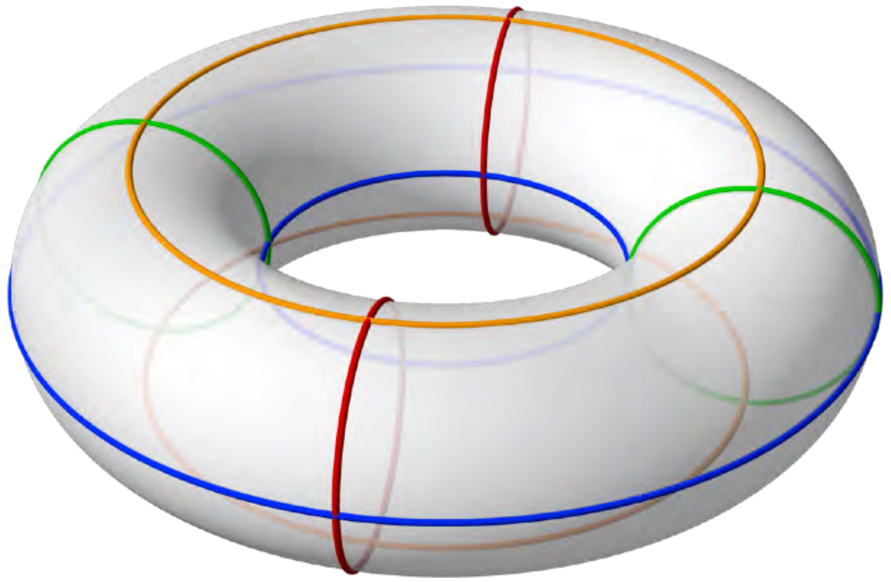
[e.g. wifi ensembles in Benevedes, JDT, PRD 2025]

Case study: Searching for "new theoretical physics" through optimization

Caution: Results are very preliminary, but I want to show my current thinking (which continues to evolve)

Case Study: Two-Dimensional Conformal Field Theories

Highly constrained mathematical framework that we know a lot about
(though not my area of expertise…)



Posing a research question to Claude

Please identify a two-dimensional conformal field theory with central charge in the range of 1.00 to 1.15.

The $\mathbf{Z_5}$ **parafermion CFT** (equivalently, the $SU(2)_5/U(1)$ coset) has central charge $\mathbf{c = 8/7 \approx 1.143}$.

[Fateev, Zamolodchikov, Sov. Phys. JETP 1985]

Only four decades too late…

[h/t to Siddharth Mishra-Sharma for the Claude credits!]

Case Study: Two-Dimensional Conformal Field Theories

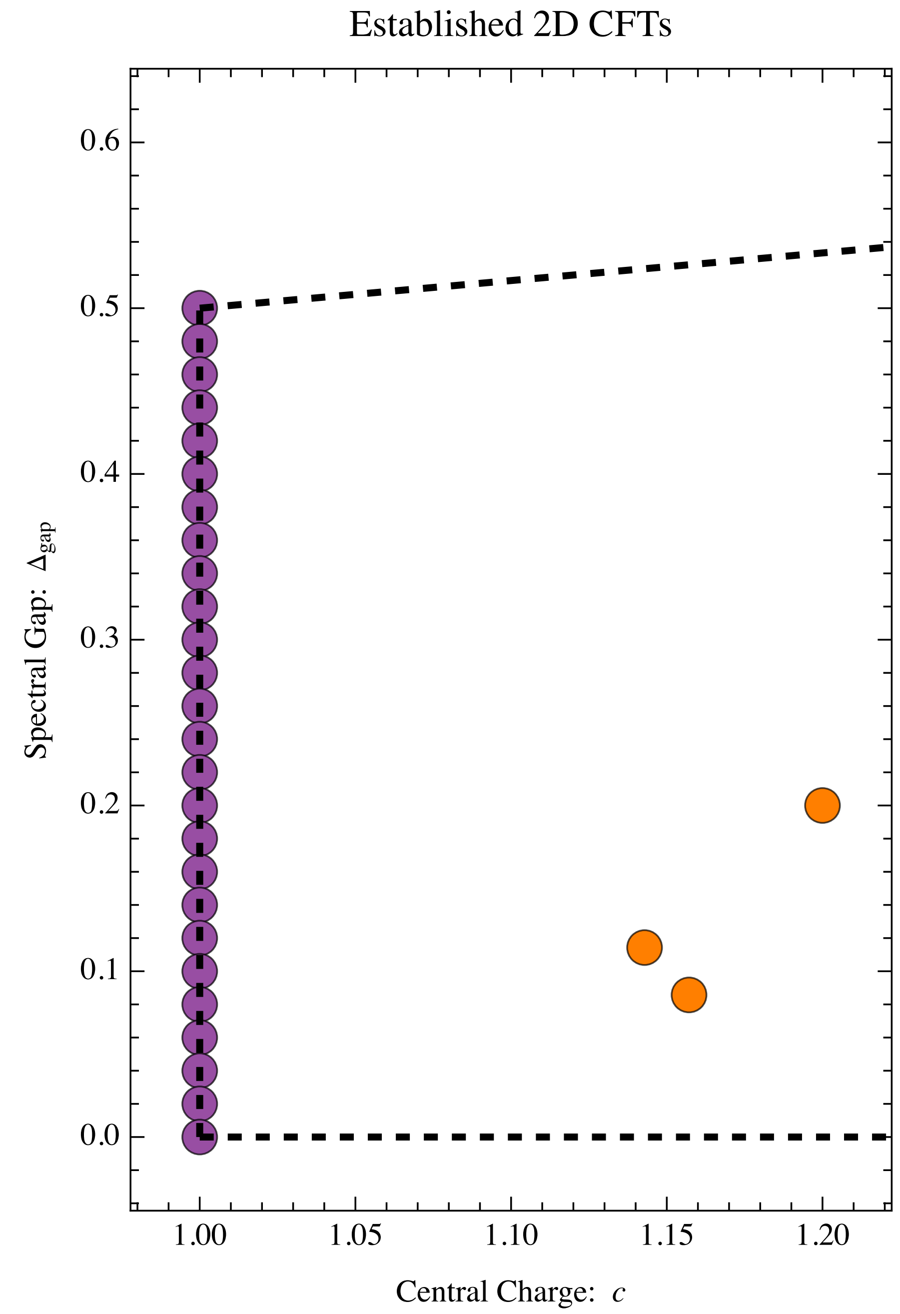
Ok, great. So let me revise my question...

Please identify a two-dimensional conformal field theory with central charge in the range of 1.00 to 1.14.

Ah, you've narrowed the window to exclude the Z_5 parafermion ($c = 8/7 \approx 1.1429$).

To my knowledge, there is no known unitary 2D CFT with central charge strictly in the range (1.00, 1.14). The theories cluster at the boundaries:

- **Below:** $c = 1$ theories (free boson, orbifolds)
- **Above:** Z_5 parafermion at $c = 8/7 \approx 1.143$, N=1 minimal model (m=5) at $c = 81/70 \approx 1.157$



Are there any other 2D CFTs in this range?

(And could I convince the experts if I found one? Could I convince myself?)

[Benjamin, Fitzpatrick, Li, JDT, in progress and very preliminary; any mistakes are my fault]



Strategies to Study/Discover CFTs

Ruling out theories (“dual”) is a lot easier than ruling them in (“primal”)

Indirect AI/ML?

Could you think about it a bit more and come up with a new theory that doesn't exist in the literature?

The honest answer: I can't construct a specific theory in (1.00, 1.14). The known RCFT constructions—cosets, W-algebra minimal models, orbifolds—all seem to have discrete spectra of central charges that skip this interval.

This could mean:

1. There's a deep reason (modular bootstrap + unitarity) forbidding unitary RCFTs here
2. There's an irrational CFT we haven't characterized
3. There's an exotic construction nobody has tried

At least it refuses to hallucinate…

Direct AI/ML?

Classification

[Chen, He, Lal, Zaz, arXiv 2020]
[Kuo, Seif, Lundgren, Whitsitt, Hafezi, PRR 2022]

Reinforcement Learning

[Kántor, Niarchos, Papageorgakis, PRL 2022, PRD 2022; + Richmond, PRD 2023]

Generative Modeling

[Laio, Valenzuela, Serone, PRD 2022]

Gradient-Based Methods

[Reehorst, Rychkov, Simmons-Duffin, Sirois, Su, van Rees, SciPost 2021]
[Afkhami-Jeddi, JHEP 2022]

Genetic Algorithms

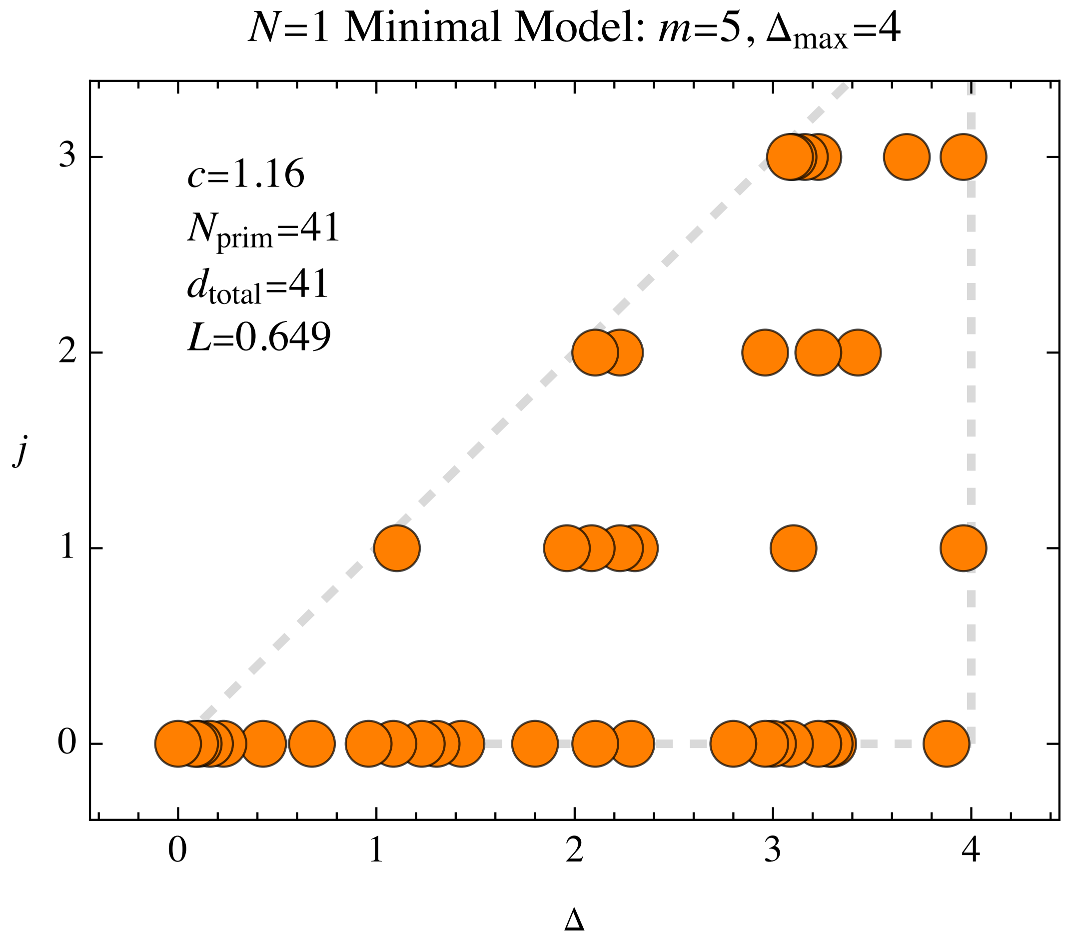
[Huang, Lee, Liao, Rumbutis, arXiv 2025]

[Please let me know if I missed your ML-related CFT paper!]
[see reviews of dual approach to conformal bootstrap in Poland, Rychkov, Vichi, RMP 2019; Rychov, Su, RMP 2024]

Warm-Up Problem: Modular Bootstrap

Necessary but not sufficient condition for unitary 2D CFT to exist

Example of more general bootstrap philosophy of
defining theoretical objects by consistency conditions



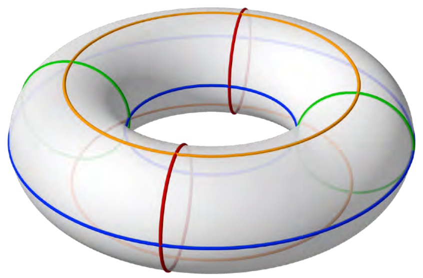
Spectrum of (Virasoro Primary) Operators:

Dimension: Δ $\quad$ $\Delta \geq j \geq 0$ $\quad$ assuming parity for $-j$

Spin: j $\quad$ $d > 0$

Degeneracy: d $\quad$ j, d are integers

(Euclidean) Partition Function: $Z(\tau; c, \{\Delta_a, j_a, d_a\})$

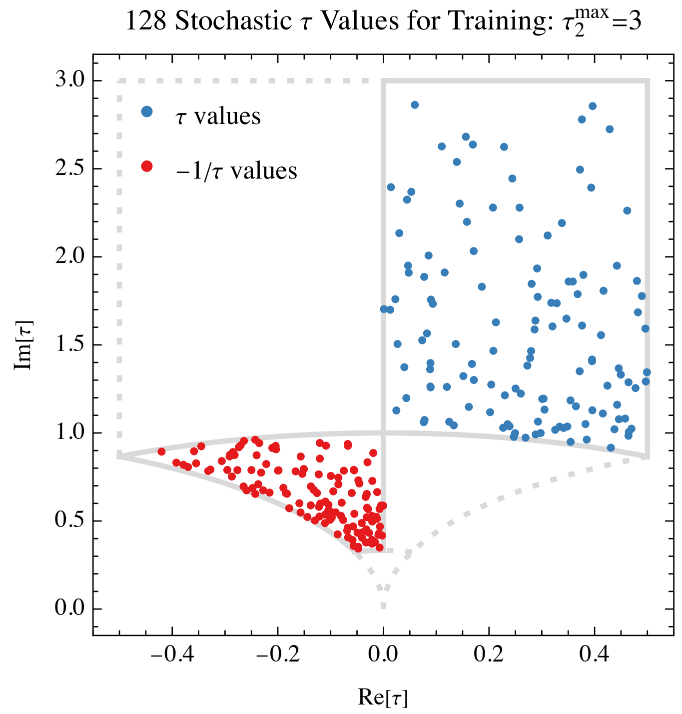


Modular Invariance:
$$Z(\tau) = Z(-1/\tau)$$

[Cardy, NPB 1991; Hellerman, JHEP 2011; Friedan, Keller, JHEP 2013; Collier, Lin, Yin, JHEP 2018; …]

"Thinking Like a Machine"

Modular bootstrap as "self-supervised learning"



✓ Training Data:

MC Integration over τ

$$L = \int_{\mathscr{F}} \frac{d\tau\, d\bar{\tau}}{\operatorname{Im}\tau} \mathscr{L}$$

✗ Loss Function(al):

Modular Invariance

$$\mathscr{L} = \big(Z(\tau) - Z(-1/\tau)\big)^2$$

✓ Learnable Function:

Partition Function

$Z(\tau; c, \{\Delta_a, j_a, d_a\})$ with Δ_a as parameters

✗ Optimizer:

Gradient Descent

$$\Delta_a^{(i)} = \Delta_a^{(i-1)} - \eta \frac{dL}{d\Delta_a}$$

Unfortunately, this approach fails dramatically at finding anything close to physical…

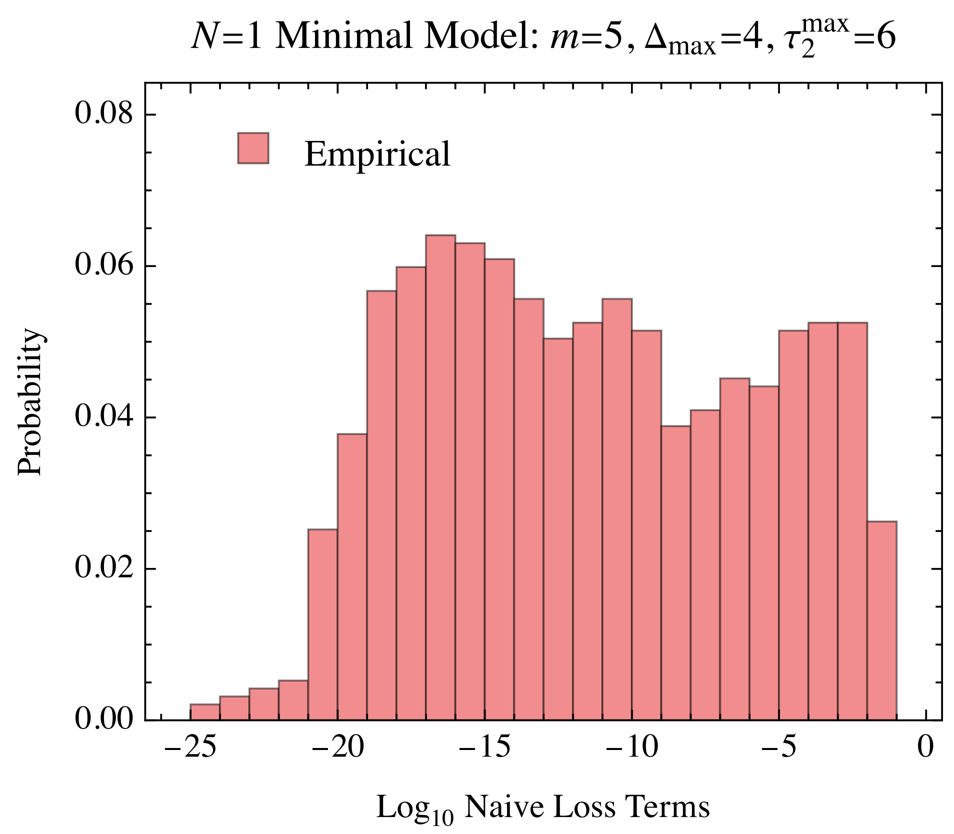
[similar philosophy to physics-informed machine learning for PDEs: Karniadakis, Kevrekidis, Lu, Perdikaris, Wang, Yang, Nature Reviews 2021]

[see ML approach to S-matrix bootstrap in Dersy, Schwartz, Zhiboedov, JHEP 2024; Niarchos Papageorgakis, PRD 2024; Gumus, Leflot, Tourkine, Zhiboedov, JHEP 2025, arXiv 2026]

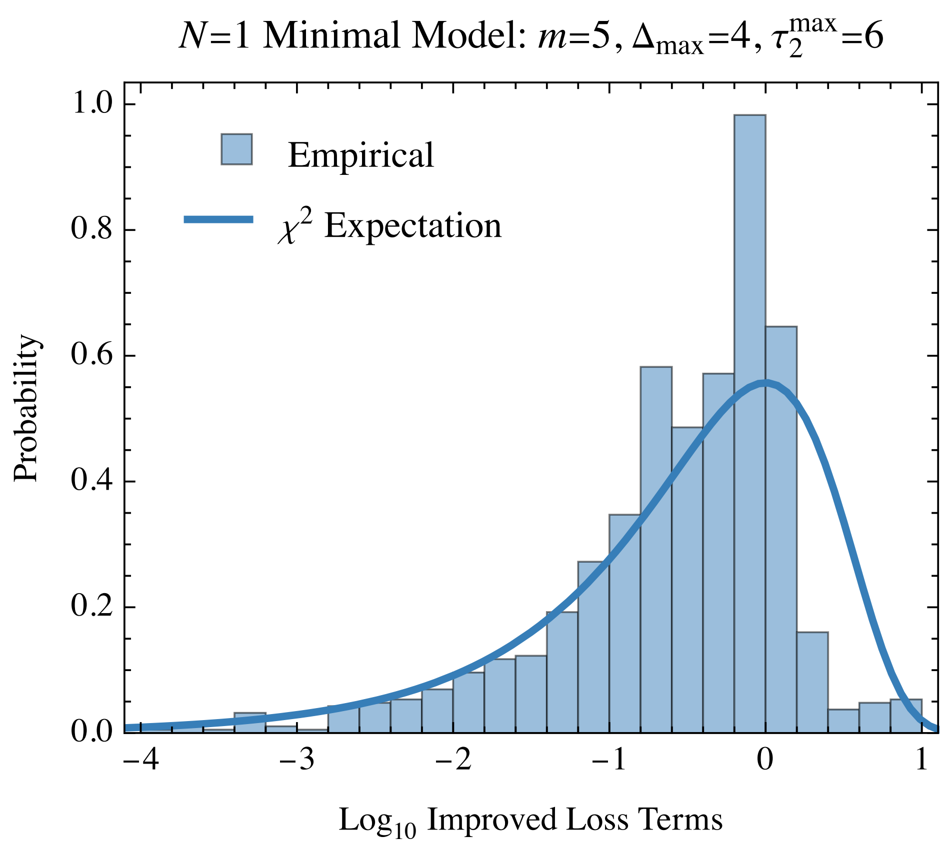
"Thinking Like an Experimental Physicist"

What are the Uncertainties on your Loss Function?

Use "χ^2" instead of MSE: $$\mathcal{L} = \left(\frac{Z(\tau) - Z(-1/\tau)}{\sigma(\tau)} \right)^2$$



$\Rightarrow$



Remarkable that this distribution comes from purely theoretical data

[Benjamin, Fitzpatrick, Li, JDT, in progress and very preliminary; any mistakes are my fault]

"Thinking Like an Experimental Physicist"

Needed a robust method to quantify uncertainties

Collaborating with Claude helped me identify connections in the literature

> We previously discussed a strategy for taking a known CFT with c = 1 and deforming its partition function to mimic the Cardy scaling behavior of a CFT with c > 1, in order to estimate uncertainties associated with truncating the primary spectrum. Can you remind me of that formula and where it first appeared in the literature?

To shift a partition function from central charge c to a target c', use:

$$\tilde{Z}_{c'}(\tau) = (\mathrm{Im}\,\tau)^{(c'-c)/2} \cdot |\eta(\tau)|^{-(c'-c)} \cdot Z_c(\tau)$$

The $|\eta|$ factor shifts the vacuum energy to give the correct $q^{-c'/24}$ scaling, while the $(\mathrm{Im}\,\tau)^{1/2}$ factor preserves modular covariance under $S : \tau \to -1/\tau$.

A similar construction appears in **Benjamin, Collier, Fitzpatrick, Maloney, and Perlmutter, "Harmonic analysis of 2d CFT partition functions" (arXiv:2107.10744)**, where they define the primary partition function $Z_p = (y^{1/2}/|\eta|^2)^c Z$ to enable spectral decomposition on the modular fundamental domain.

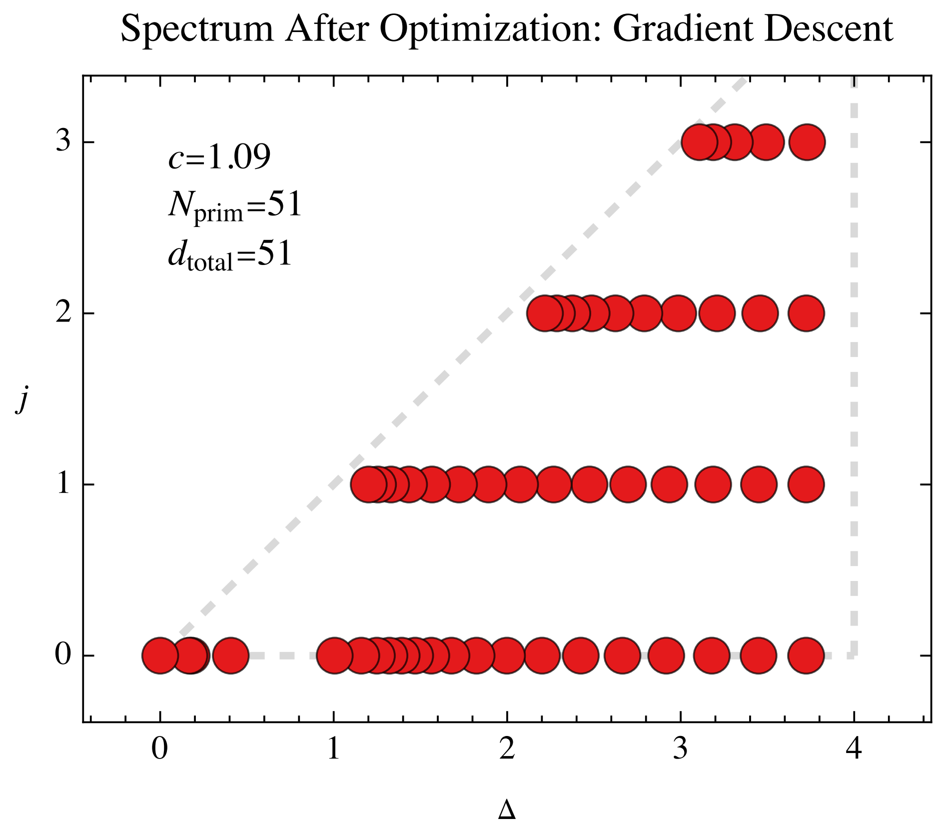
[Benjamin, Fitzpatrick, Li, JDT, in progress and very preliminary; any mistakes are my fault]

“Thinking Like an Mathematical Physicist”

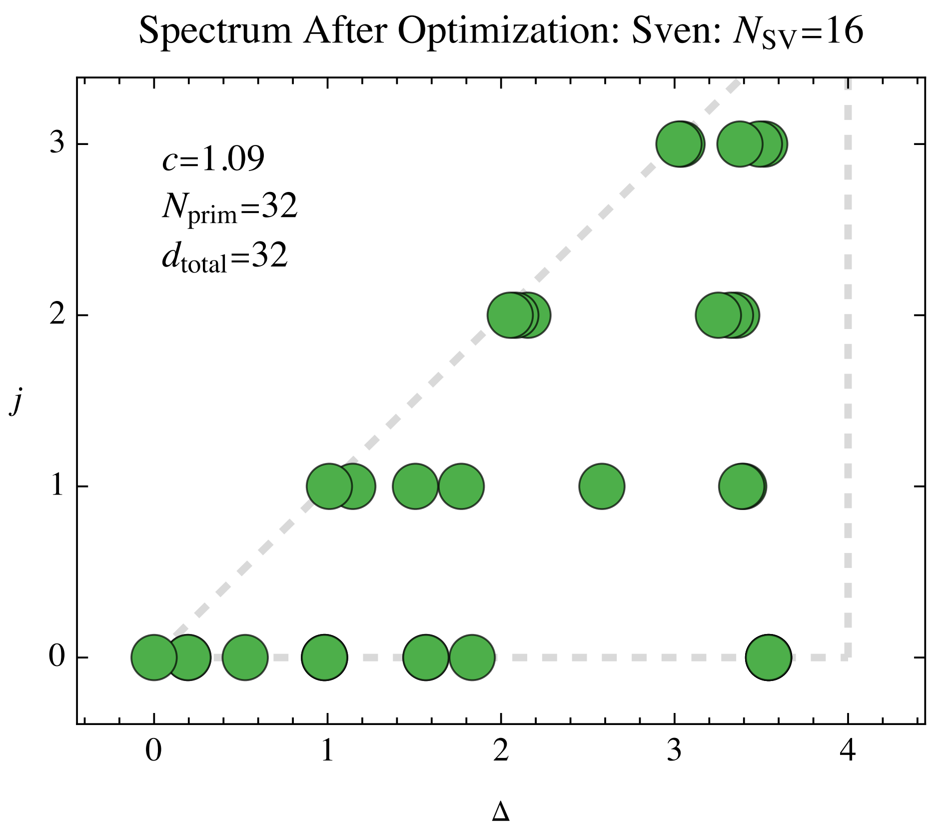
What is the Geometry of your Loss Landscape?

Use quasi-second order methods instead of SGD:

$$\Delta_a^{(i)} = \Delta_a^{(i-1)} - g_{ab} \frac{dL}{d\Delta_b}$$



$\Rightarrow$



Code name: “Sven”
Singular Value dEsceNt

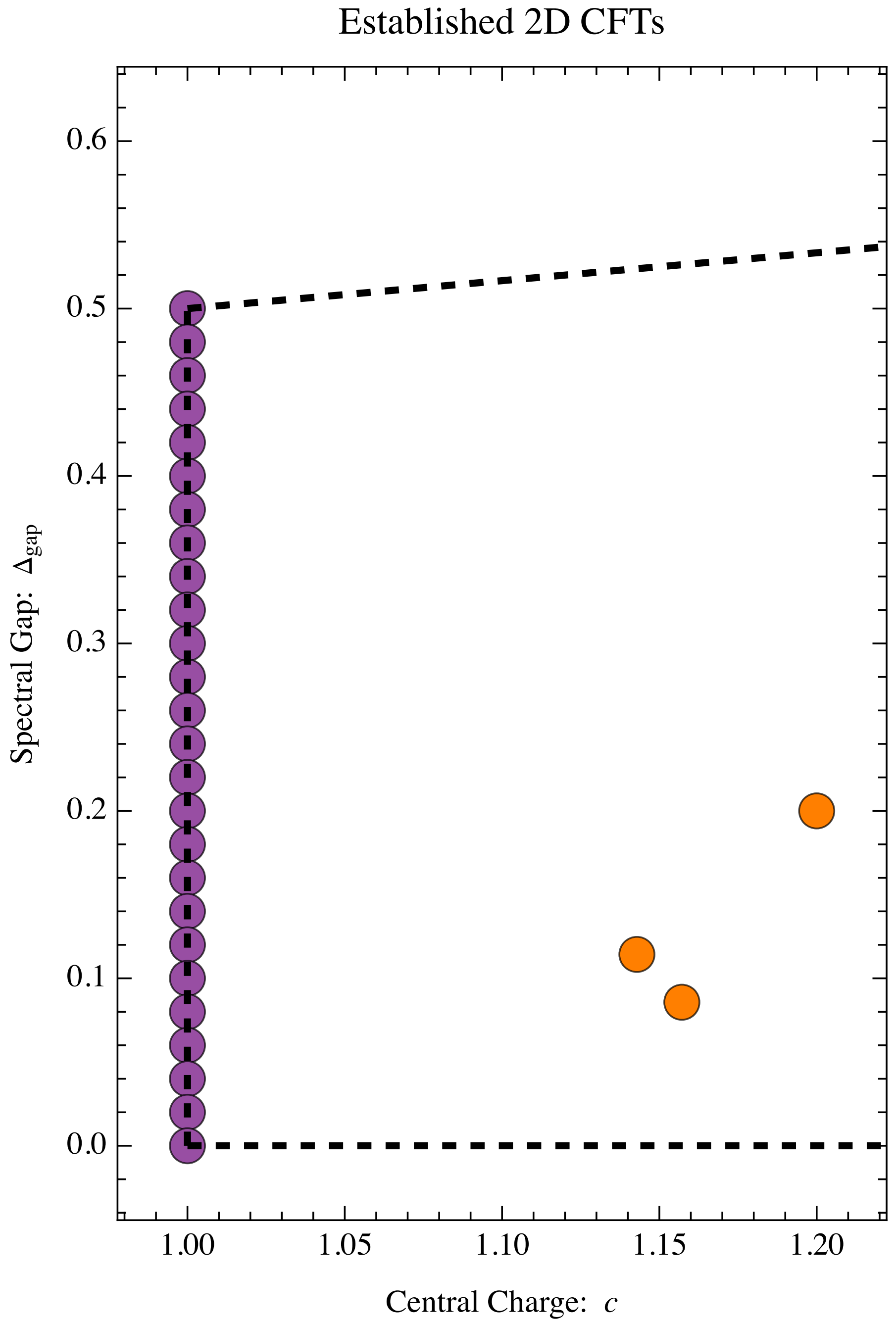
Coming soon to a pip installer near you…

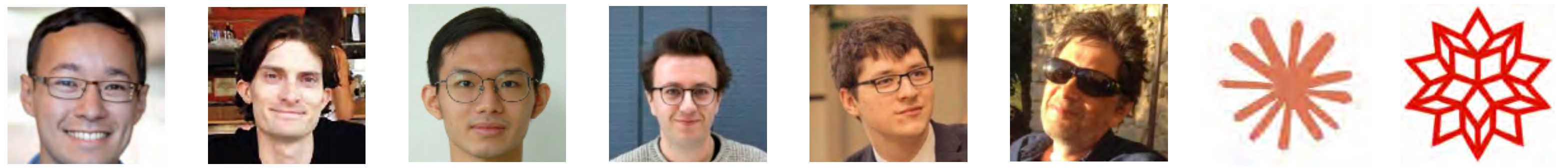
[Bright-Thonney, Harvey, Lukas, JDT, in progress and very preliminary; any mistakes are my fault]



Updating the Space of Candidate 2D CFTs

Remember: Modular invariance is necessary but not sufficient

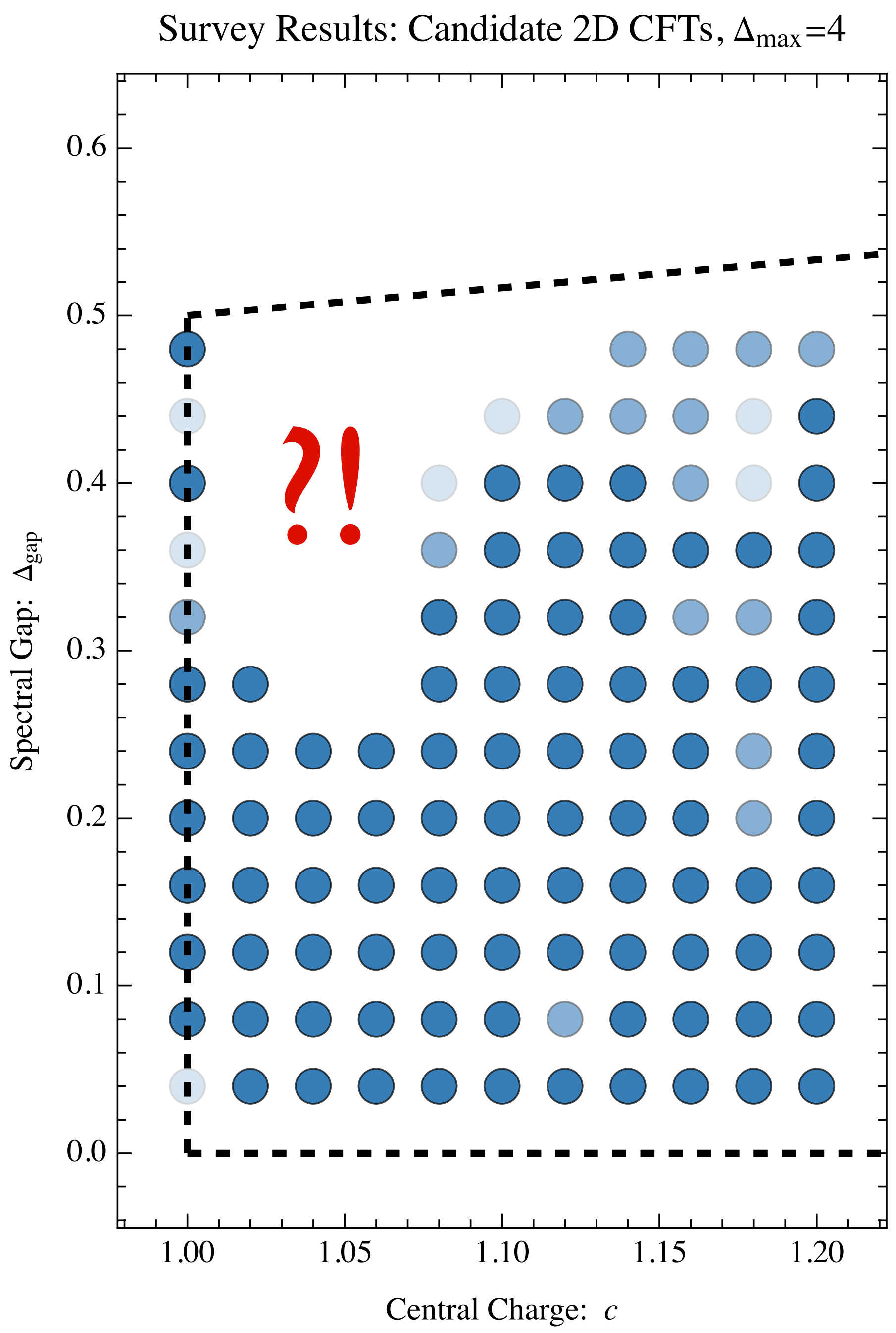




[Benjamin, Fitzpatrick, Li, JDT, in progress; Bright-Thonney, Harvey, Lukas, JDT, in progress, both still very preliminary; any mistakes are my fault]

Updating the Space of Candidate 2D CFTs

Remember: Modular invariance is necessary but not sufficient



I don't know how I would have gotten this far without AI/ML methods!

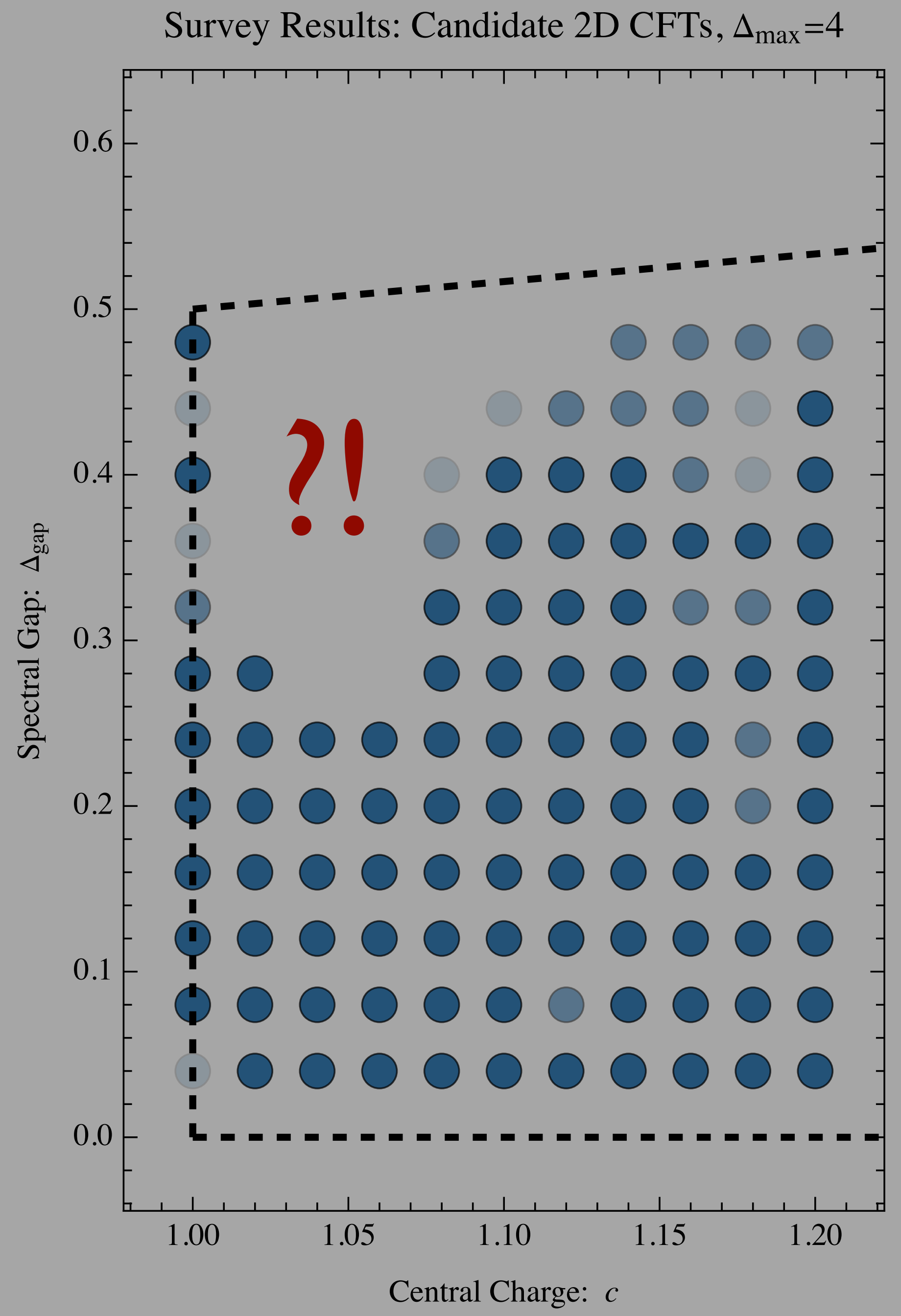
Direct AI/ML: Reframe modular bootstrap in optimization language

Indirect AI/ML: Reason through conceptual issue of estimating uncertainties

[Benjamin, Fitzpatrick, Li, JDT, in progress; Bright-Thonney, Harvey, Lukas, JDT, in progress, both still very preliminary; any mistakes are my fault]

Updating the Space of Candidate 2D CFTs

Remember: Modular invariance is necessary but not sufficient

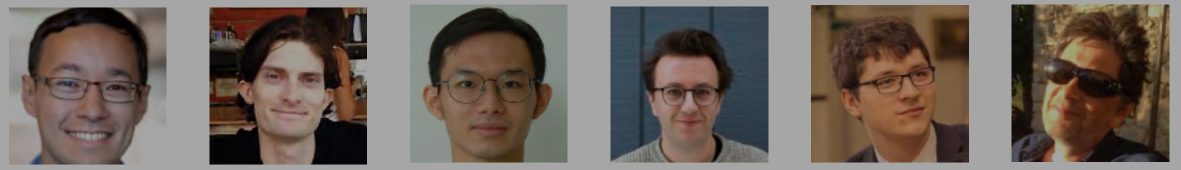


But I also don't know how I would have gotten this far without physics input!

Identify tractable problem of relevance to experts in the field

Develop new optimization algorithms based on physics intuition

Pursue statistical robustness even in a purely theoretical setting



[Benjamin, Fitzpatrick, Li, JDT, in progress; Bright-Thonney, Harvey, Lukas, JDT, in progress, both still very preliminary; any mistakes are my fault]

Advancing (theoretical) physics through a centaur-style merging of physical reasoning and computational algorithms

Centaur Science: Adventures in AI+Physics

Physics → ? ← AI

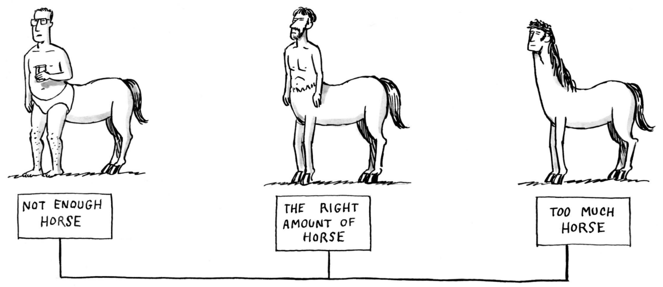
Machine Learning through the Lens of Physics

AI is transforming physics research, but many foundational aspects of AI can be translated into the language of physics

Scenes from my Sabbatical

Being a "centaur scientist" means capitalizing on emerging AI technologies while insisting on scientific rigor and robustness

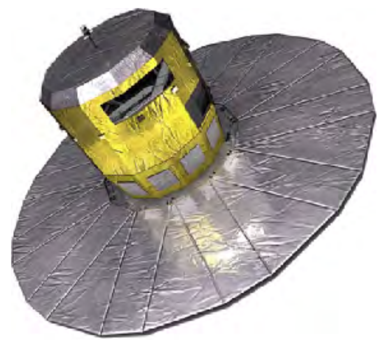
Progress in AI+Physics driven by early career talent with interdisciplinary training



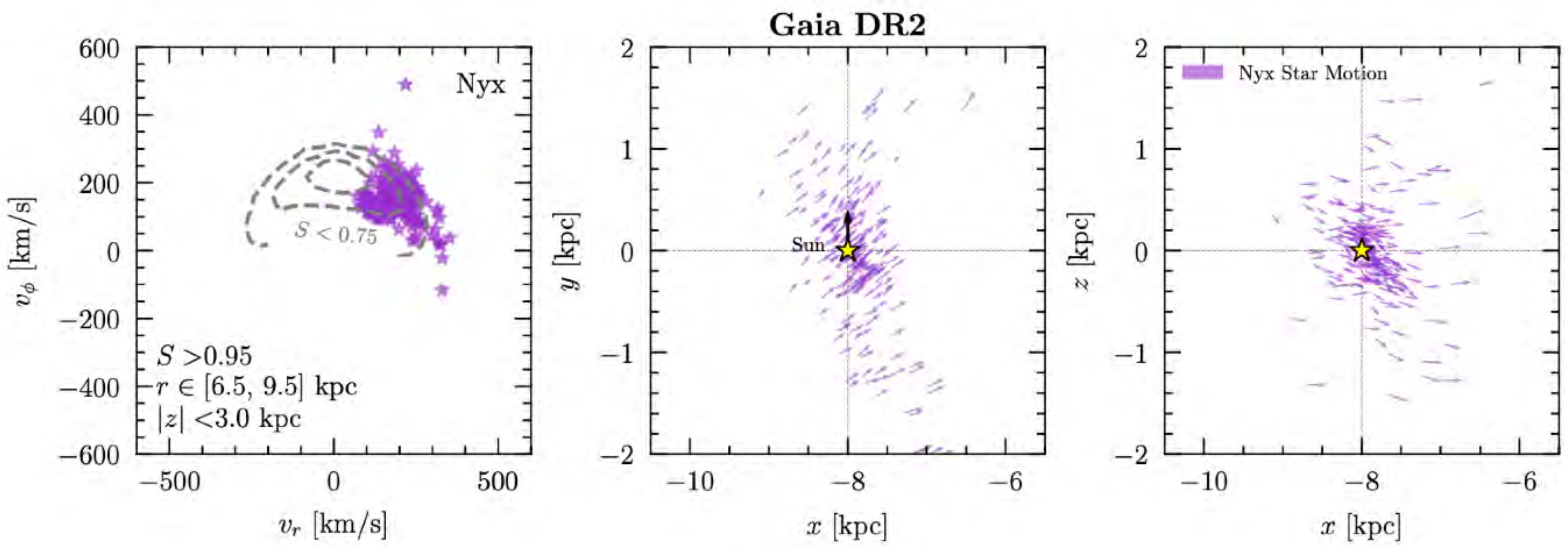
Backup Slides

AI for Astrophysics: Galaxy Formation and Dark Matter

Scrutinizing stellar kinematics from Gaia to reconstruct history of the Milky Way



Discovery of Nyx!



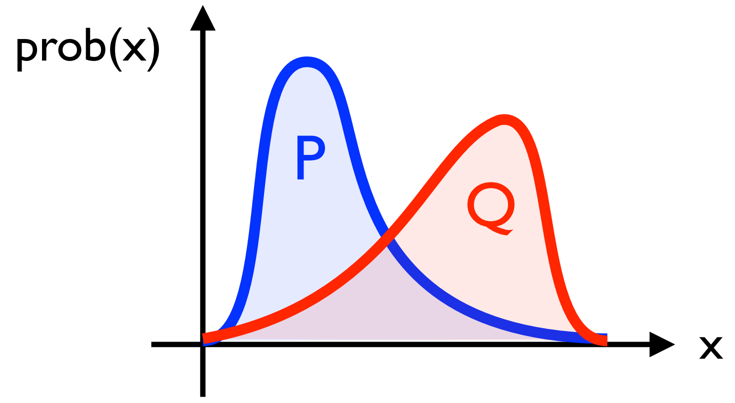
My former PhD student
(and current MIT colleague)

[Necib, et al., Nature Astronomy 2020; Ostdiek, et al., A&A 2020]

Machine Learning 101: Likelihood Ratio Trick

Many HEP problems can be expressed in this form!

Key tool for simulation-based inference



Goal: Estimate p(x) / q(x)

Training Data: Finite samples P and Q

Learnable Function: f(x) parametrized by, e.g., neural networks

Loss Function(al): $$L = -\langle \log f(x) \rangle_P + \langle f(x) - 1 \rangle_Q$$

[see e.g. Cranmer, Pavez, Louppe, arXiv 2015; D'Agnolo, Wulzer, PRD 2019; simulation-based inference in Cranmer, Brehmer, Louppe, PNAS 2020; relation to f-divergences in Nguyen, Wainwright, Jordan, AoS 2009; Nachman, JDT, PRD 2021]

Machine Learning 101: Likelihood Ratio Trick

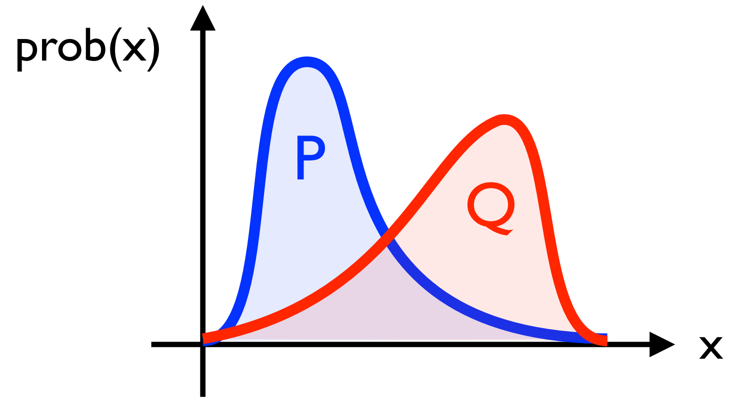
Many HEP problems can be expressed in this form!

Key tool for simulation-based inference

Goal: Estimate p(x) / q(x)

Training Data: Finite samples P and Q

Learnable Function: f(x) parametrized by, e.g., neural networks



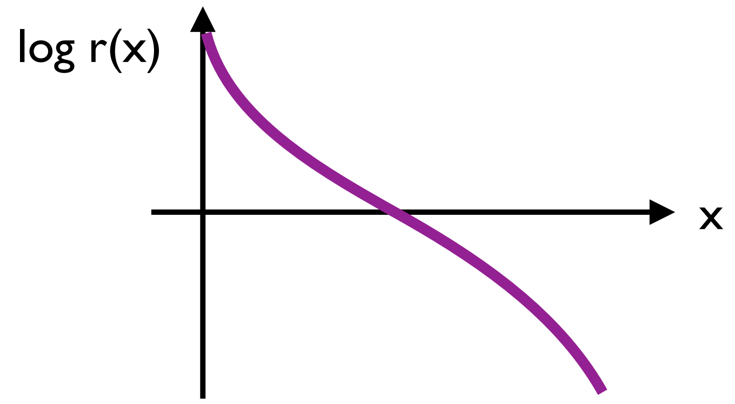
Action/Lagrangian:
assuming enough data

$$L \approx \int dx \left(- p(x) \log f(x) + q(x) \left(f(x) - 1 \right) \right) \quad !!$$

Euler-Lagrange:

$$\underset{f(x)}{\arg\min} \, L = \frac{p(x)}{q(x)}$$ *Likelihood ratio*

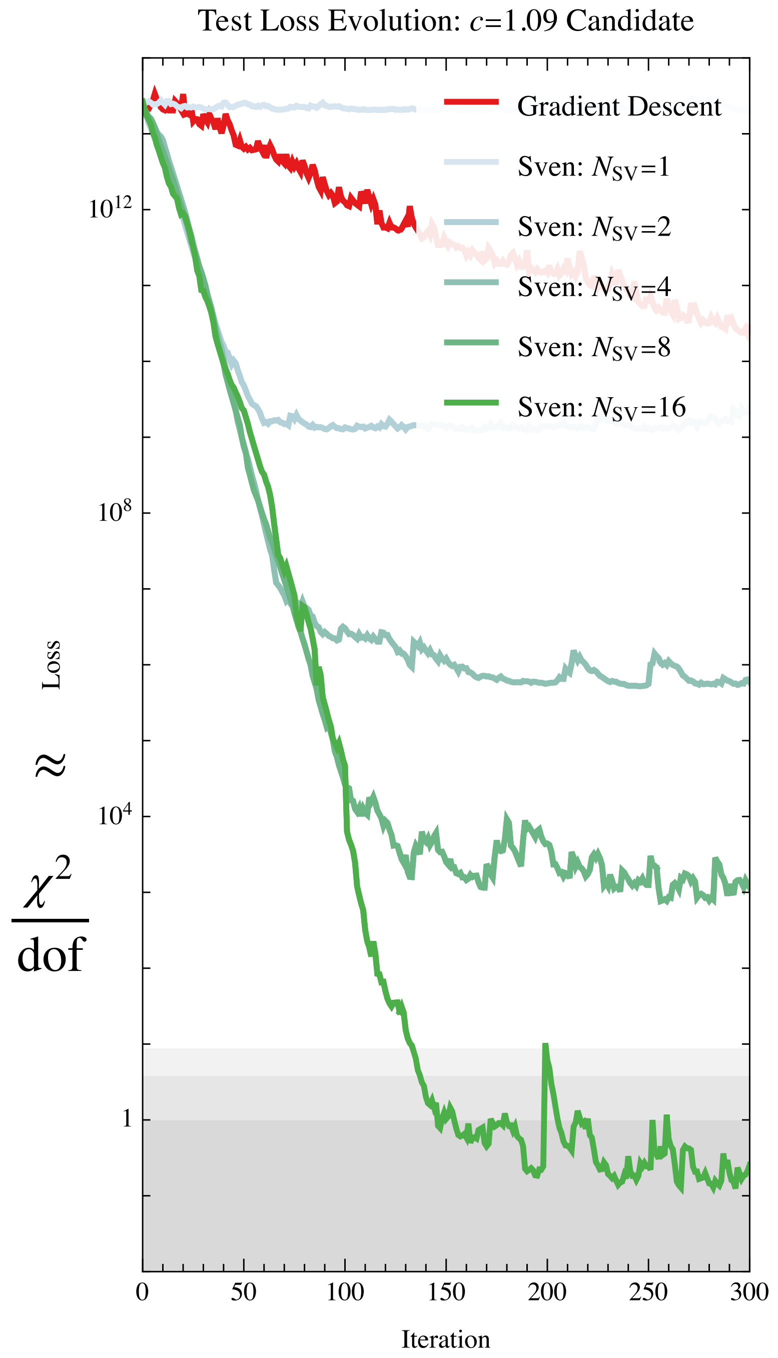
$$- \min_{f(x)} L = \int \mathrm{d}x \, p(x) \log \frac{p(x)}{q(x)}$$ *Kullback–Leibler divergence*



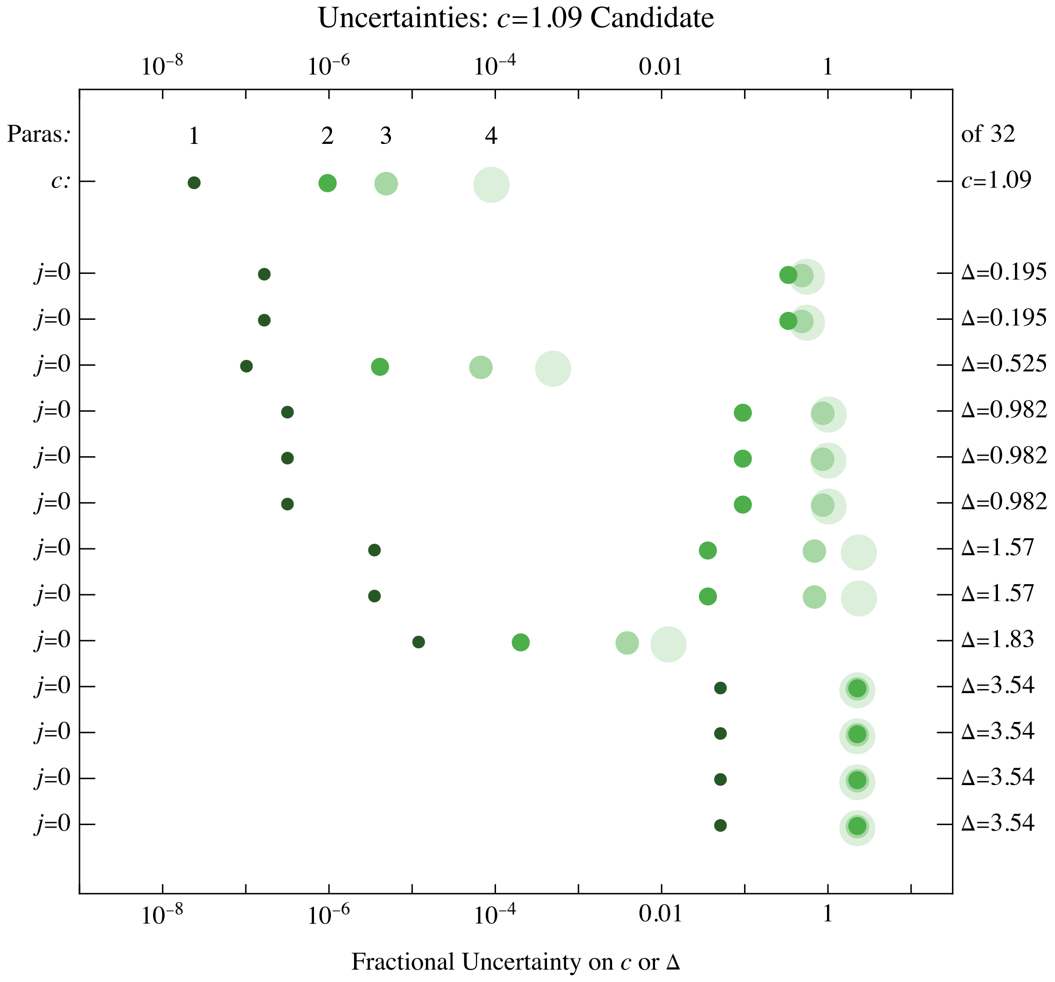
[see e.g. Cranmer, Pavez, Louppe, arXiv 2015; D'Agnolo, Wulzer, PRD 2019;
simulation-based inference in Cranmer, Brehmer, Louppe, PNAS 2020;
relation to f-divergences in Nguyen, Wainwright, Jordan, AoS 2009; Nachman, JDT, PRD 2021]

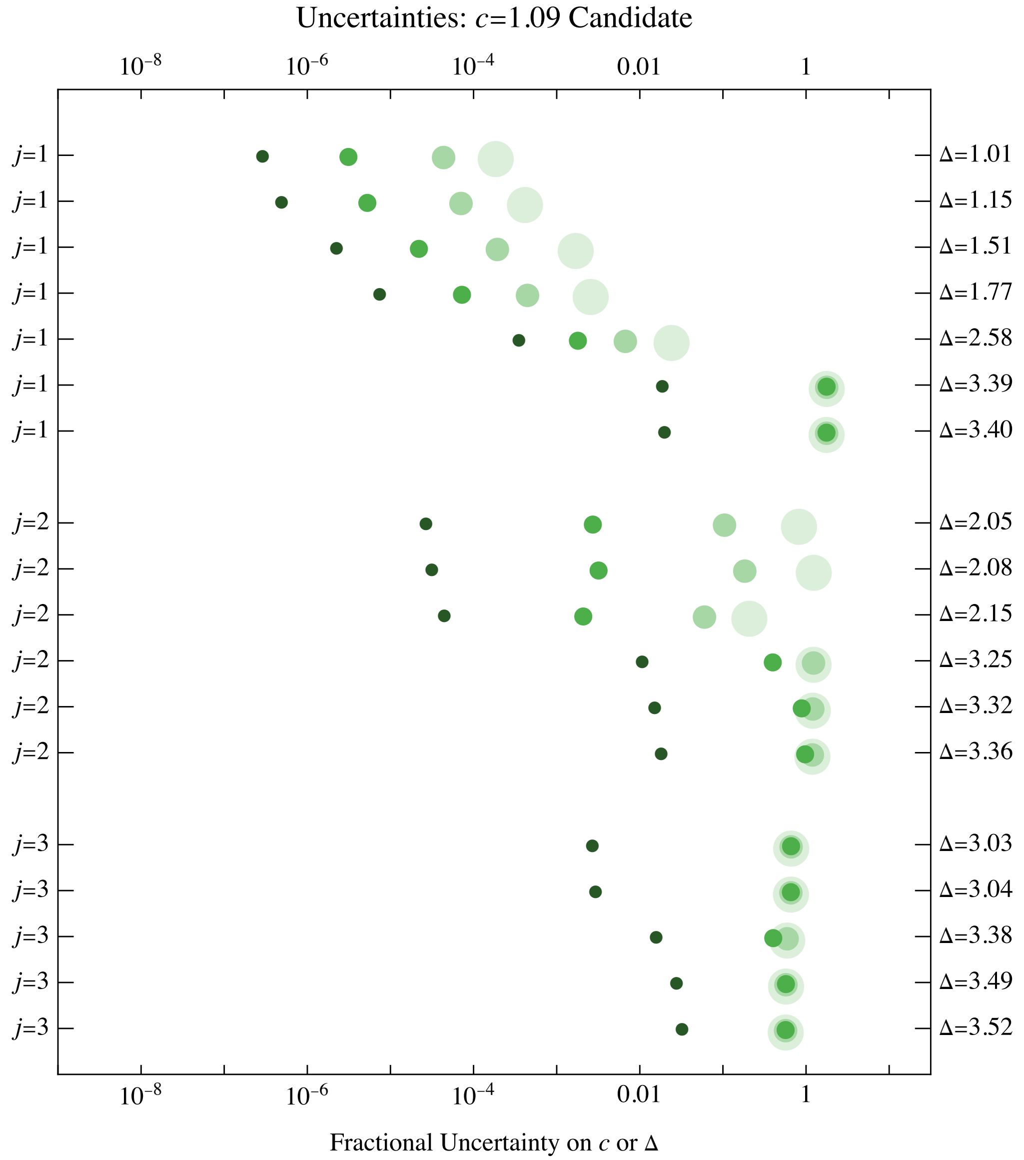
Descending into Irrational CFTs

Or at least into modularly invariant Euclidean partition functions



Tight constraints on individual operators…





…most likely spanning a moduli space of solutions!

[Benjamin, Fitzpatrick, Li, JDT, in progress; Bright-Thonney, Harvey, Lukas, JDT, in progress, both still very preliminary; any mistakes are my fault]

From AI Curmudgeon… to AI Evangelist!

MIT News
ON CAMPUS AND AROUND THE WORLD

Laboratory for Nuclear Science
August 26, 2020

National Science Foundation announces MIT-led Institute for Artificial Intelligence and Fundamental Interactions

IAIFI will advance physics knowledge — from the smallest building blocks of nature to the largest structures in the universe — and galvanize AI research innovation.

The New York Times

By Dennis Overbye Nov. 23, 2020

Can a Computer Devise a Theory of Everything?

"In five to 10 years from now, I'm going to want to do exactly what you're getting at: Here's the data, here's a very rough tool kit; find the equation I could put on a T-shirt, the equation that replaces the Standard Model of particle physics. What's the equation that replaces Einstein's general relativity?"

N.B. This was November 2020. ChatGPT was released November 2022.