Institute for Artificial Intelligence and Fundamental Interactions (IAIFI)

and

Machine Learning for Amplitudes

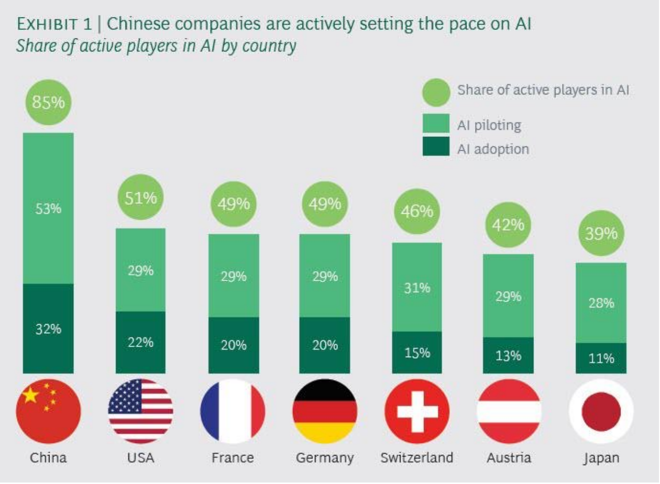
AI goes MAD

June 14, 2022

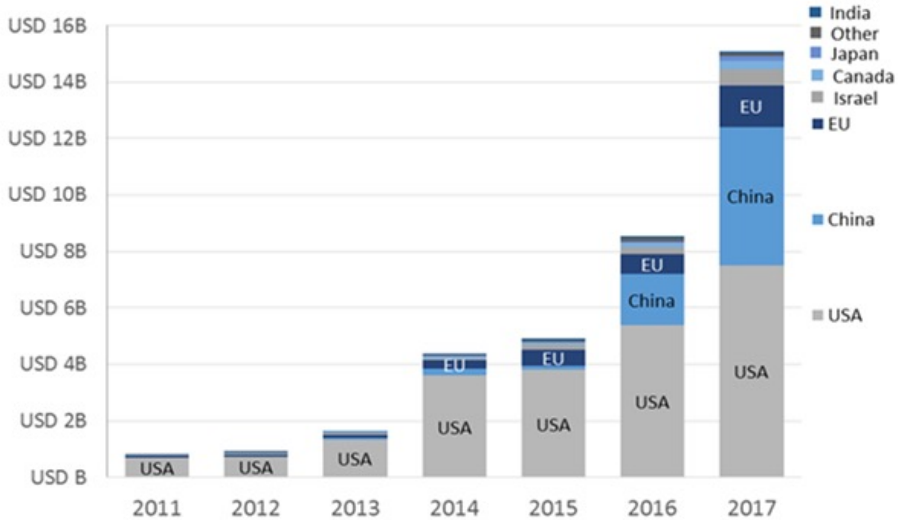
Matthew Schwartz

Harvard University

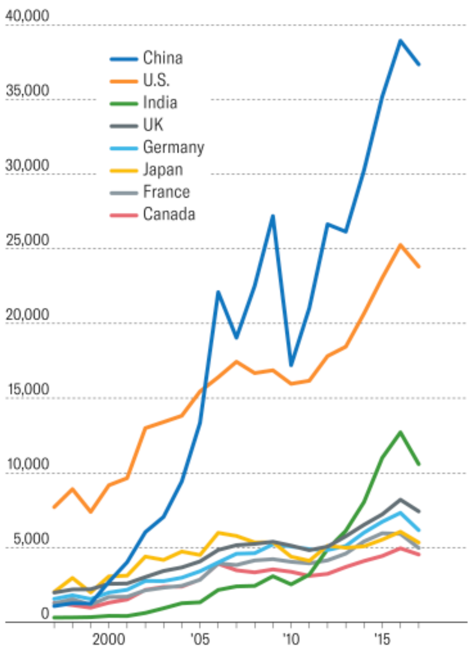
Context: Global Investment in Artificial intelligence

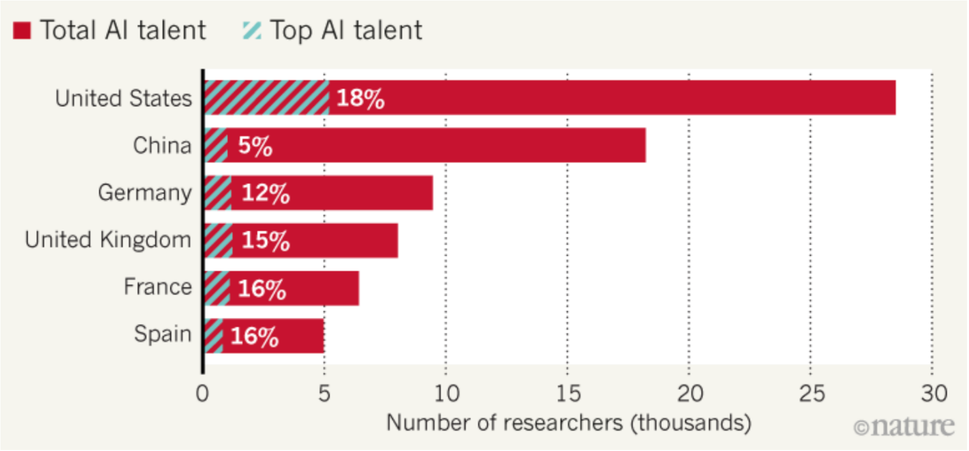


Total estimated investments in AI start-ups ($ billion), 2011-2017

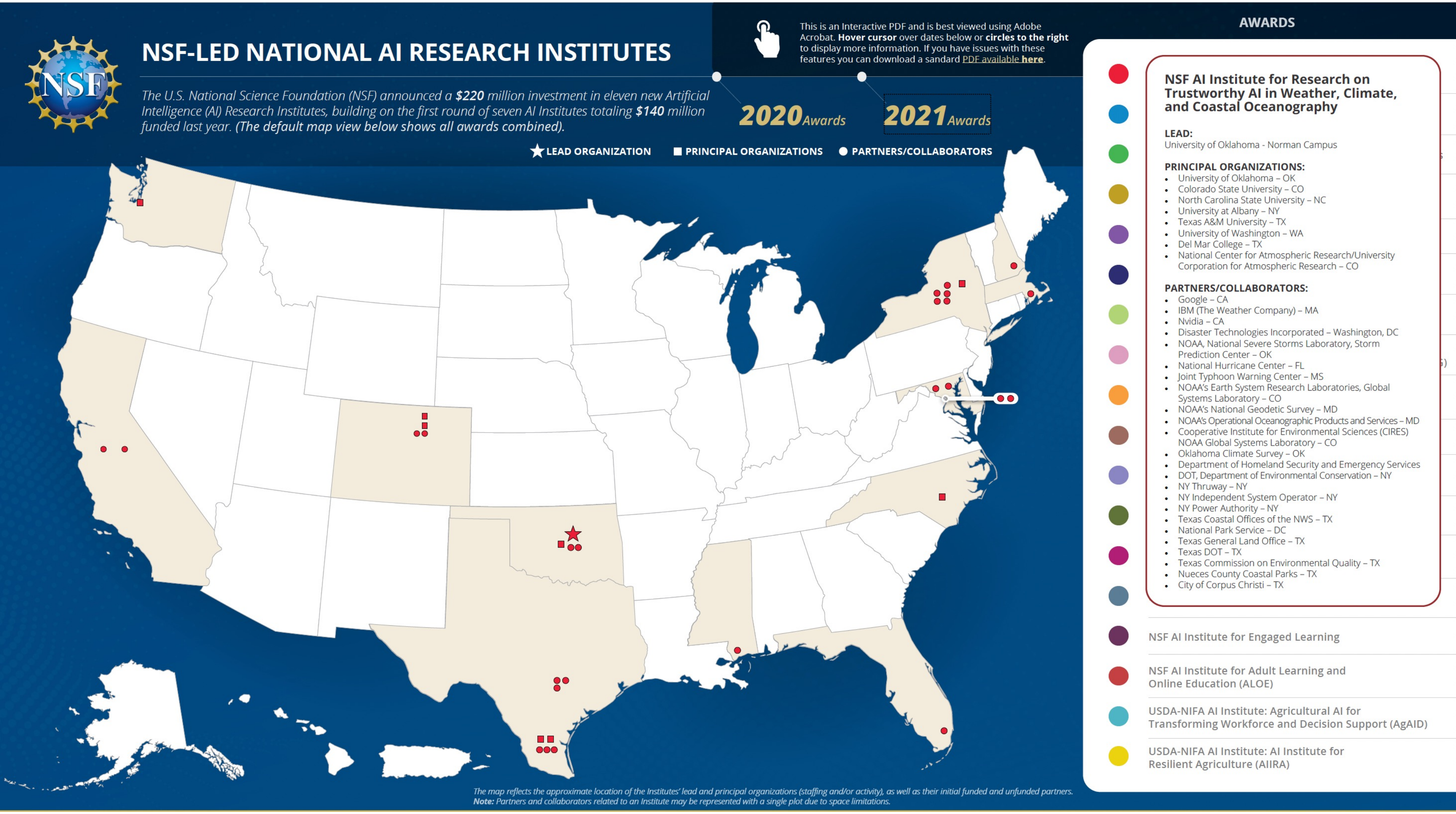


Global output of AI scientific papers





2020: US National Science Foundation funds 7 AI institutes for $140 million



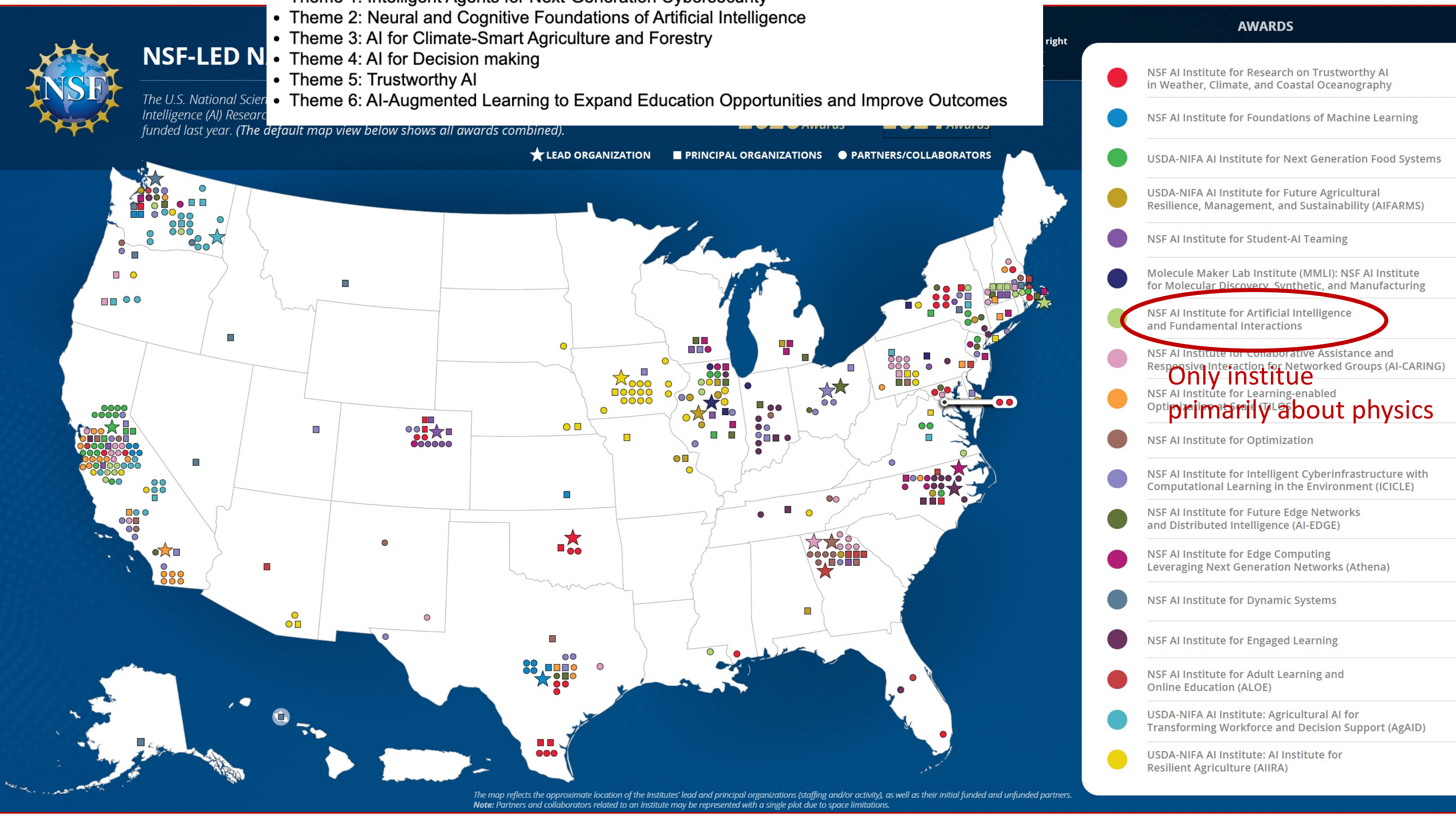
Matthew Schwartz

2020: US National Science Foundation funds 7 AI institutes for $140 million

2021: 11 more institutes funded (+ $220 million)

2022: 7 more institutes funded (+ $140 million)

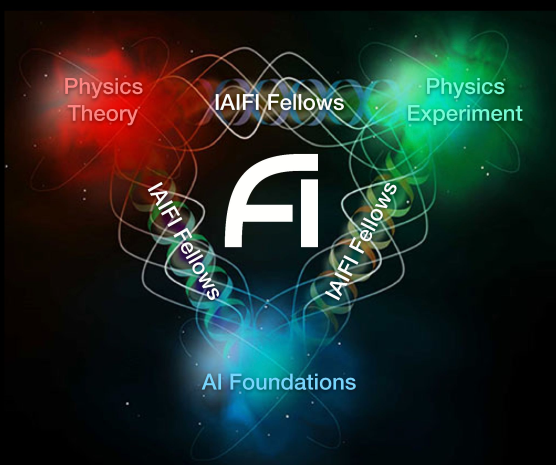
- Theme 1: Intelligent Agents for Next-Generation Cybersecurity
- Theme 2: Neural and Cognitive Foundations of Artificial Intelligence
- Theme 3: AI for Climate-Smart Agriculture and Forestry
- Theme 4: AI for Decision making
- Theme 5: Trustworthy AI
- Theme 6: AI-Augmented Learning to Expand Education Opportunities and Improve Outcomes

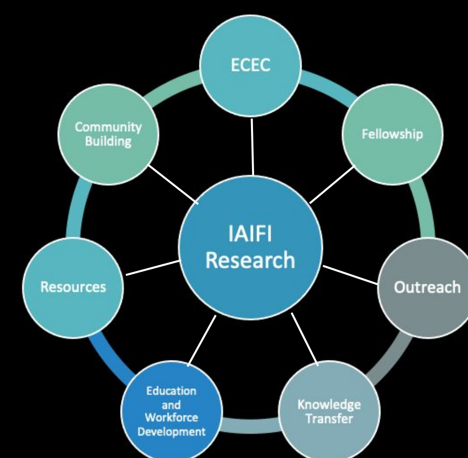


Matthew Schwartz

NSF AI Institute for Artificial Intelligence and Fundamental Interactions (IAIFI /aɪ-faɪ/)

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





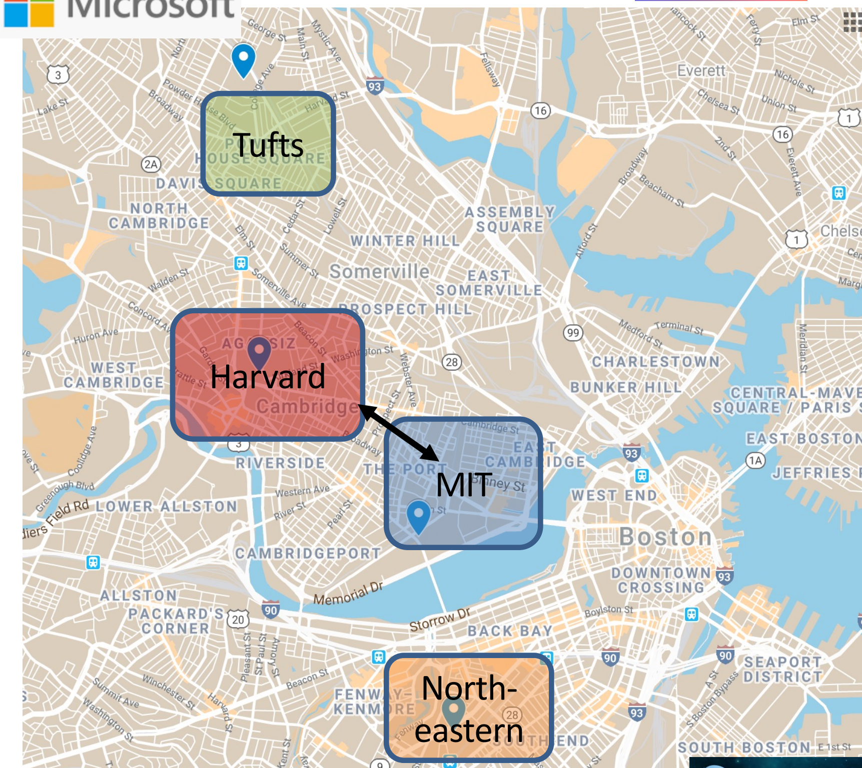
https://iaifi.org

IAIFI Postdoctoral Fellowship

iaifi-management@mit.edu

@iaifi-news

Boston area collaboration:
- MIT + Harvard + Northeastern + Tufts
- Connections to local (and distant) Industry partners



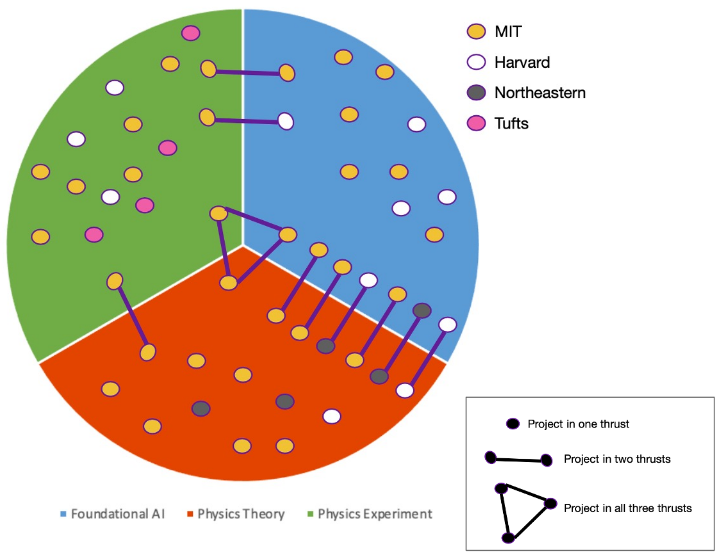
Central administration at MIT
- Dedicated space for CAIFI postdoctoral fellows
- Encourages cross-disciplinary communication among fellows

Activities and personnel distributed
- IAIFI space at each school

Regular seminars, lunch talks, meetings, social hours
- Hybrid seminars
- Journal clubs
- Social events
- Active Zulip discussion forum

IAIFI Mission

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



- ➢ Training, education & outreach at Physics/AI intersection
- ➢ Cultivate early-career talent (e.g. IAIFI Fellows)
- ➢ Foster connections to physics facilities and industry
- ➢ Build strong multidisciplinary collaborations
- ➢ Advocacy for shared solutions across subfields

Critical mass of Physics/AI expertise in Boston area

IAIFI Senior Scientists

Senior Investigators: 18 Physicists + 9 AI Experts + 11 IAIFI Affiliates **= 38 senior scientists**

Junior Investigators: ≈23 FTE PhD Students, ≈7 IAIFI Fellows in steady state

Pulkit Agrawal
Lisa Barsotti
Isaac Chuang
William Detmold
Bill Freeman
Philip Harris
Lina Necib
Kerstin Perez
Alexander Rakhlin
Dan Roberts

Phiala Shanahan
Tracy Slatyer
Tess Smidt
Marin Soljacic
Justin Solomon
Washington Taylor
Max Tegmark
Jesse Thaler
Mike Williams

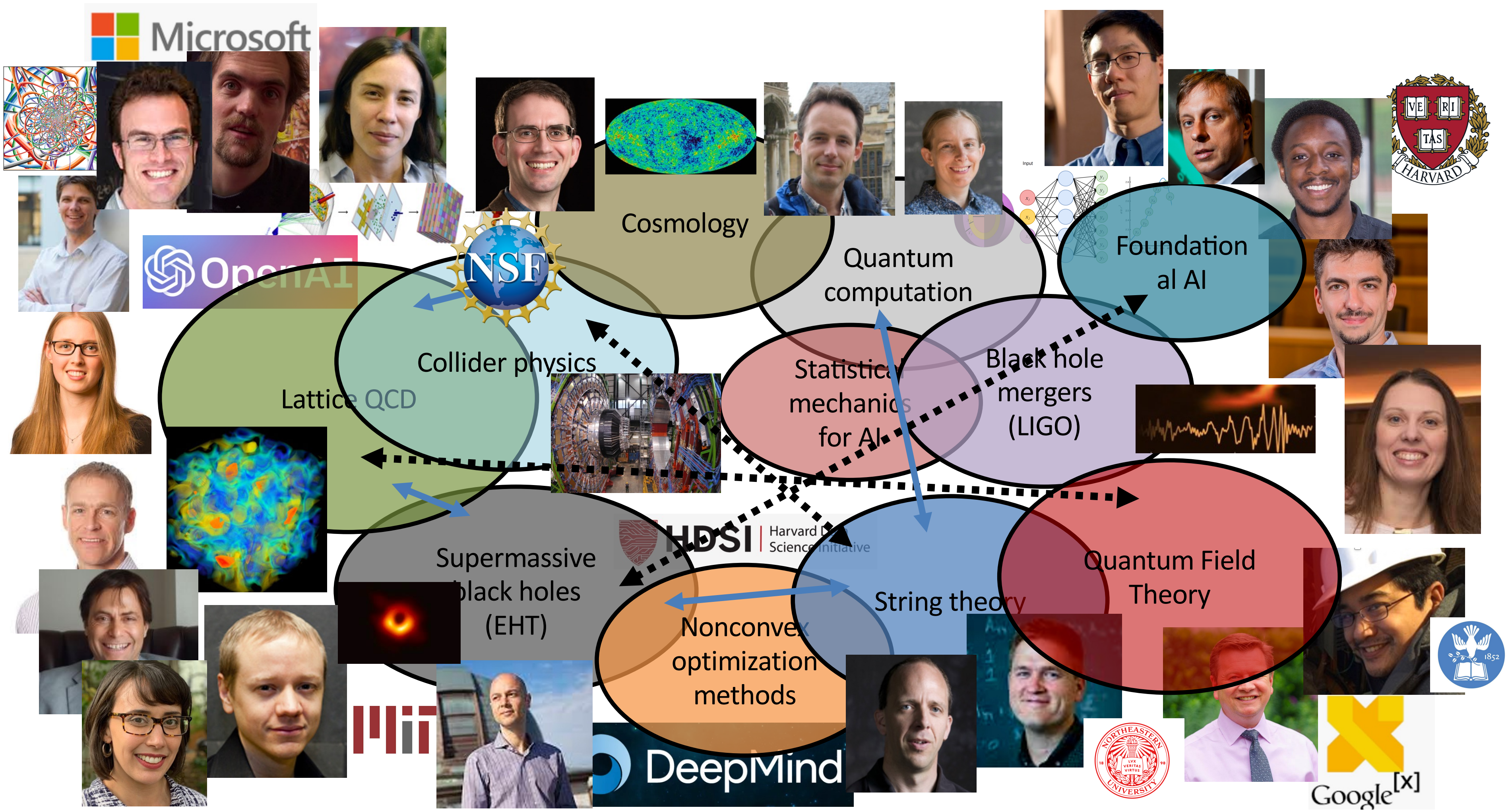
Carlos Argüelles
Demba Ba
Edo Berger
Mike Douglas
Cora Dvorkin
Daniel Eisenstein
Doug Finkbeiner
Cengiz Pehlevan

Artan Sheshmani
Haim Sompolinsky
Matthew Schwartz
Yaron Singer
Todd Zickler

Ning Bao
James Halverson
Brent Nelson
Fabian Ruehle

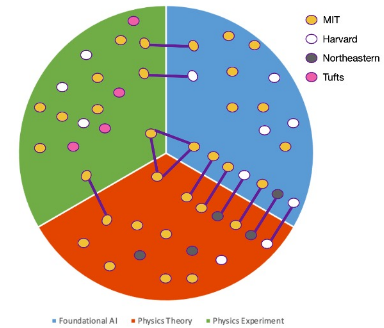
Shuchin Aeron
Taritree Wongjirad

Inter-institutional, inter-departmental, cross-disciplinary


Microsoft
OpenAI
NSF
Cosmology
Quantum computation
Foundational AI
Collider physics
Lattice QCD
Statistical mechanics for AI
Black hole mergers (LIGO)
Supermassive black holes (EHT)
HDSI | Harvard Data Science Initiative
String theory
Quantum Field Theory
Nonconvex optimization methods
DeepMind
Google[x]

7 IAIFI fellows (+3 next year)

- Inter-institutional postdocs
- 3 year positions
- Ideally cross physics/ML boundaries



Anna Golubeva

IAIFI Fellow, PhD Summer School Committee Member

Theory of Deep Learning, Condensed Matter Theory

Di Luo

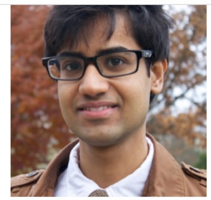
IAIFI Fellow, Speaker Selection Committee Member

quantum algorithms and machine learning for condensed matter physics, high energy physics, and quantum information science

Ge Yang

IAIFI Fellow

reinforcement learning, planning, optimal transport, robotics



Siddharth Mishra-Sharma

IAIFI Fellow, IAIFI Early Career and Equity Committee Member

particle astrophysics, cosmology, simulation-based inference, probabilistic programming

Denis Boyda

Incoming IAIFI Fellow

lattice field theory, generative models, Markov Chain Monte Carlo, high performance computing

Carolina Cuesta

Incoming IAIFI Fellow

smology and AI, ML models, statistics

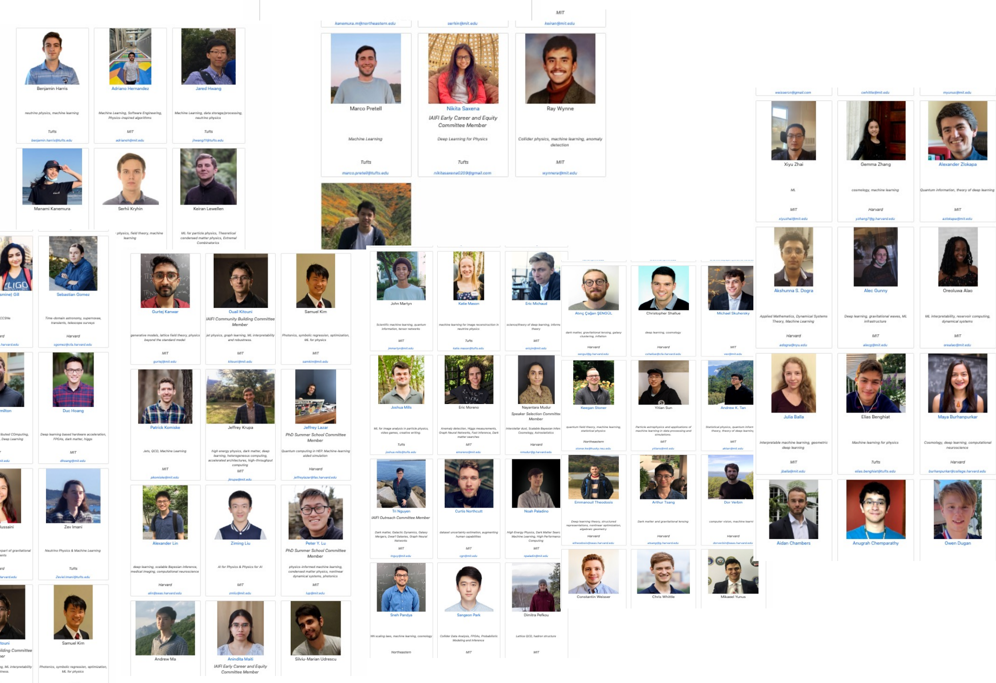
Jessie Micallef

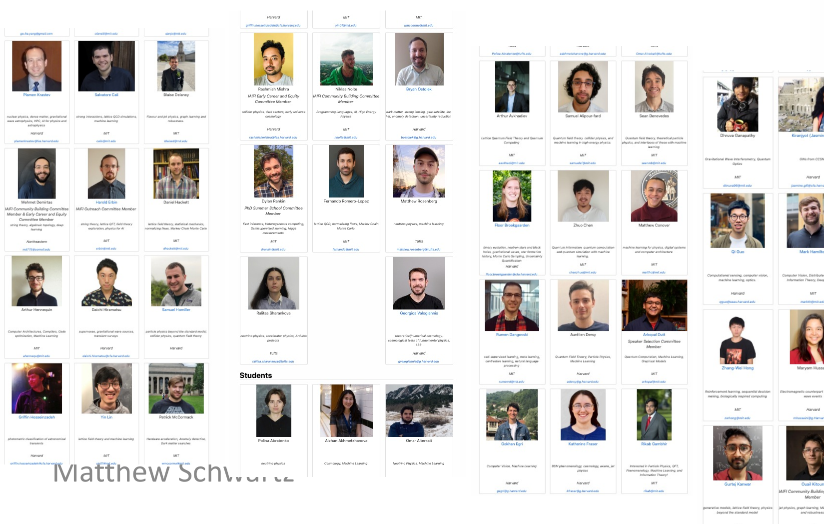
Incoming IAIFI Fellow

achine learning, particle physics experiments, neutrinos

+22 affiliated postdocs

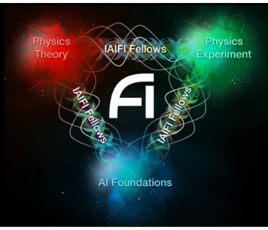
+86 graduate students





Lots of activities!

IAIFI in March 2022



IAIFI Research

Theoretical Physics
- Nuclear & Particle Physics
- String Theory/Physical Mathematics
- Astroparticle Physics
- Automated Discovery of Models

Experimental Physics
- Particle Physics Experiments
- Gravitational Wave Interferometry
- (Multi-Messenger) Astrophysics

Foundational AI
- Symmetries & Invariance
- Speeding up Control & Inference
- Physics-Informed Architectures
- Neural Networks Theory

IAIFI Colloquia
- Biweekly talks from leaders in AI and Physics
- Broadcast live on IAIFI YouTube Channel
- Fall 2021/Spring 2022: every other Friday at 2 pm

IAIFI Fellowship Program
- Three-year postdoctoral appointment
- Freedom in pursuing research and collaborations
- Applications for 2023-2025 open Summer 2022

Interdisciplinary PhD Program at MIT
- Physics, Statistics, and Data Science
- Take 4 classes, 1 each in the areas of Probability, Statistics, Computation & Statistics, and Data Analysis
- Submit and defend a PhD thesis that involves the utilization of statistical methods in a substantial way

IAIFI Affiliate Program
- Senior researchers in the Boston area contributing to IAIFI mission
- Must include nomination from existing IAIFI Senior Investigator

IAIFI Early Career and Equity Committee
- Serves as advisory board to IAIFI Management on aspects related to early career researchers and diversity, equity, and inclusion (DEI)
- Developed a Code of Conduct for IAIFI
- Established and monitors anonymous form for feedback

IAIFI Internal Events
- Includes IAIFI Internal Discussion Seminars, Journal Club, and social/networking events
- Open to IAIFI Investigators and affiliated junior and senior researchers in the Boston area

IAIFI Computing Resources
- Conducted a survey of IAIFI members regarding their computing needs
- Plan to purchase 8 Lenovo GPU nodes, each with 4x nVidia A100 GPUs (~$540k)
- Will be stored and operated through Harvard Cannon

MITx course
- Developing digital course based on IAP course: “Computational Data Science in Physics”
- 12 weeks of content at the undergraduate/graduate level
- Received a $72,000 grant from MIT for development

Research Overview

Foundational AI Research

Galvanize AI innovation by incorporating physics intelligence into artificial intelligence

Symmetries & invariances		Efficient algorithms
Physics-informed architectures	→	Algorithm speed
Statistical physics		Inverse problems
Field theory techniques		Neural network dynamics

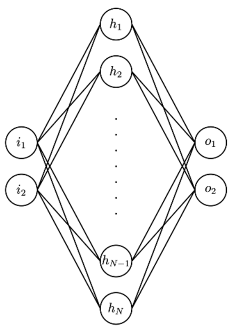
Foundational AI Overview

Contributing to Research Objectives:

- Breaking down barriers between AI and Physics
- Developing methods that can be applied to real-world tasks and generalized to previous unseen domains
- Identifying ways to solve problems faster and more accurately

- **Intelligent Clustering for High-Energy Collider Physics: Demba Ba, Electrical Engineering and Bioengineering, Harvard**
- **Group Sparse Autoencoders: Demba Ba, Electrical Engineering and Bioengineering, Harvard**
- **Understanding the Generalization Gap in Visual Reinforcement Learning: Pulkit Agrawal, EECS, MIT**
- **Learning Task Informed Abstractions: Pulkit Agrawal, EECS, MIT**
- **Light Field Networks: William Freeman, EECS, MIT**
- **Generalization in Overparametrized Models: Alexander Rakhlin, Brain and Cognitive Sciences, MIT**
- **Visual Grouping with a Field of Junctions: Todd Zickler, Engineering and Applied Sciences, Harvard**
- **Scalable Differentiable Models for Task-Specific Inverse Optical Design: Todd Zickler, Engineering and Applied Sciences, Harvard**
- **Learning Pointcloud Representations: Pulkit Agrawal, EECS, MIT**
- **The Principles of Deep Learning Theory: Dan Roberts, Physics, Salesforce (Affiliate)**

IAIFI Senior Investigator: Jim Halverson



Neural Network (NN):

Theory of random functions

Defined by their construction

Quantum Field Theory (QFT):

Theory of random functions

Defined by their distribution

$$Z = \int D\phi \, e^{-S[\phi]}$$

$$S[\phi] = \int d^dx \, \phi(x)(\Box + m^2)\phi(x)$$

NN-QFT correspondence: Progress by considering one from the other's perspective

QFT ideas for NNs: [Halverson, Maiti, Stoner] 2020
Modeling NN Densities

- Non-Gaussian phenomenological model of NN density.
- Compute NN ensemble correlations with Feynman diagrams.
- RG flow arises in some density models.
- Agreement with NN experiments.

QFT ideas for NNs: [Halverson, Maiti, Stoner] 2021
Symmetry-via-Duality

- Deduce symmetries of NN actions by study of correlations computed in parameter space.
- Input / output symmetries of NN are analog of spacetime / internal symmetries in QFT.
- Both continuous and discrete symmetries, Abelian and non-abelian.

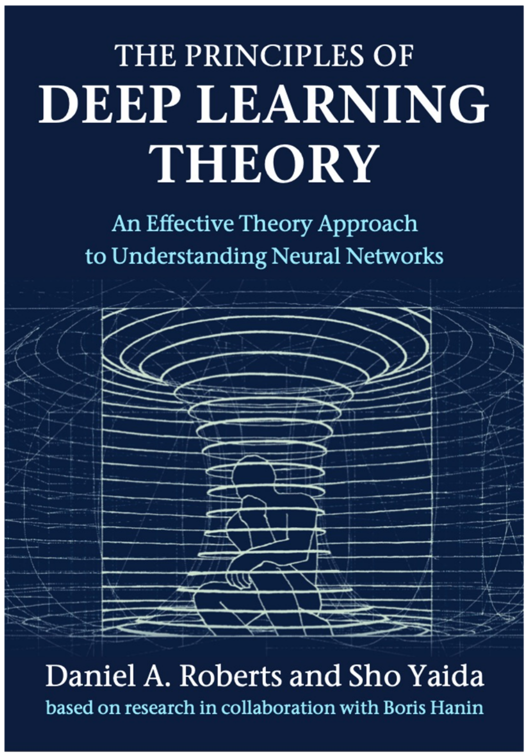
NN ideas for QFTs: [Halverson] in progress
Building Quantum Fields out of Neurons

- Reframe randomness of QFs in parameter-space; How we build fields, not how we draw them.
- Use NNs to define Lorentz-invariant, unitary QFTs.
- Explains prevalence near-Gaussianity in QFT.

Other work progress:
[Gukov, Halverson], [Halverson, Maiti, Stoner, Schwartz]
See also: *[Roberts, Yaida], [Erbin, Lahoche, Samary]*

IAIFI Affiliate: Dan Roberts

With Sho Yaida (Facebook AI Research), Boris Hanin (Princeton)



The Principles of Deep Learning Theory

A new monograph/textbook on ***deep learning theory*** inspired by ideas from ***physics***.
Available online now (arxiv:2106.10165), to be published by *Cambridge University Press*.

- Puts forth a set of principles that enable us to theoretically analyze deep neural networks of practical relevance.
- Based on the "***effective theory***" framework of physics, draws on (i) the Wilsonian *renormalization group*, (ii) *criticality* and *universality*, and (iii) the *1/n expansion*.
- Develops tools for understanding the *statistics* of wide and deep networks at initialization as well as for understanding the training *dynamics* when learning from data.
- *Representation learning* in deep networks can be understood in terms of the interactions of neurons that occur in realistic networks.

Physics Theory Research

Enable (theoretical) physics discoveries by developing and deploying the next generation of AI technologies

Point clouds

Normalizing flows

Uncertainty quantification

Symbolic regression

→

Provably-exact ab-initio theory calculations

Discovery of physical features, symmetries, correlations

Interpretable automated algorithm design

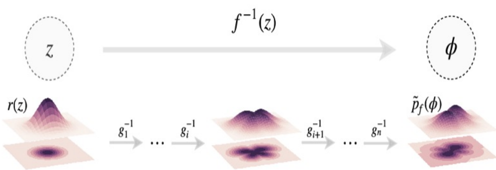
Physics Theory Overview

- New applications for ML in theoretical physics and also theoretical physics for ML. **ML for ab initio calculations!**
- **Examples:** 1) develop ML architectures for ab initio calculations and/or the resulting datasets
  2) lessons from math and physics for ML, as well as the converse.
  e.g. of both: SU(N)-equivariant normalizing flows for lattice QCD

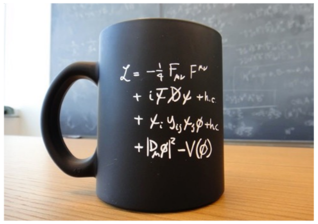
- **Generative Flow Models to Accelerate Lattice Quantum Field Theory Calculations: Phiala Shanahan, Physics, MIT**
- **AI Preconditioners for Dirac Matrix Inversion: Phiala Shanahan, Physics, MIT**
- **Efficient Variational Calculations for Nuclear Theory with AI: Phiala Shanahan, Physics, MIT & William Detmold, Physics, MIT**
- **Point Cloud Learning with Energy Flow: Jesse Thaler, Physics, MIT & Justin Solomon, EECS, MIT**
- **Infinite Networks for Self-Generative Learning: Jim Halverson, Physics, Northeastern**
- **Machine Learning for Topology: Knot Theory: Jim Halverson, Physics, Northeastern**
- **NN-QFT Correspondence: Jim Halverson, Physics, Northeastern**
- **An architecture to extract the dark matter signal: Siddharth Mishra-Sharma (Fellow), Physics, MIT**
- **Discovering Sparse Interpretable Dynamics from Partial Observations: Marin Soljacic, Physics, MIT**
- **Path-Integral Contour Deformation for Estimation of Noisy Observables in Lattice Field Theory: William Detmold, Physics, MIT**
- **Jet Metrics and Autoencoders: Matthew Schwartz, Physics, Harvard**
- **Exploring Dual Moduli Spaces via Topological Data Analysis: Brent Nelson, Physics, Northeastern**
- **Machine-Learning Invariance & Invariants: Max Tegmark, Physics, MIT**
- **ML and Calabi-Yau Geometry: Washington Taylor, Physics, MIT**
- **Topological Obstructions to Autoencoding: Dan Roberts, Physics, Salesforce**
- **Discoveries from applying neural networks to QFT and string theory: Harold Erbin (Postdoc), Physics, MIT**

Generative Flow Models to Accelerate Lattice Quantum Field Theory Calculations

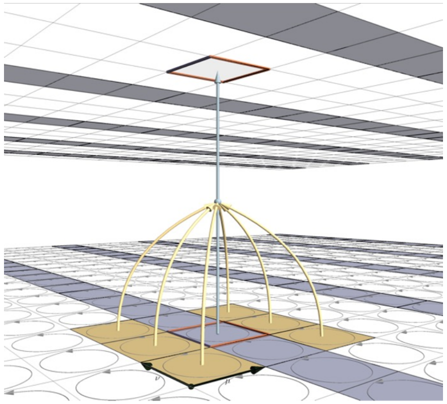
Phiala Shanahan, Daniel Hackett, Gurtej Kanwar (MIT), Denis Boyda (MIT & ANL), Michael S. Albergo (NYU, CCPP), Sébasten Racanière, Danilo J. Rezende (DeepMind), Julian M. Urban (U. Heidelberg, ITP), Kyle Cranmer (NYU, CCPP)



Generative models based on normalizing flows



Symmetry structure of the Standard Model of particle physics

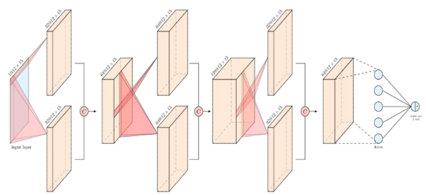


Development of machine learning frameworks for efficient sampling in lattice quantum field theory

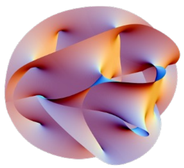
- Series of papers developing generative flow models on compact domains, and on U(n) and SU(n) Lie group variables
- Proof-of-principle demonstration of orders-of-magnitude acceleration over traditional sampling approaches
- Roadmap to QCD for state-of-the-art nuclear/particle physics studies
  - 🤵🏾‍♀️ Architectures for compact variables
  - 🤵🏾‍♀️ Incorporation of gauge symmetry
  - 🤵🏾‍♀️ Non-Abelian groups
  - 🤵🏾‍♀️ Fermions
  - ☐ Scaling to state-of-the-art, exascale hardware

[PRD 103, 074504 (2020), PRL 125, 121601 (2020), ICML, PMLR 8083-8092 (2020), 2107.00734 (2021), PRD 104, 114507 (2021), 2101.08176 (2021), 2202.11712 (2022)]

IAIFI Postdoc: Harold Erbin

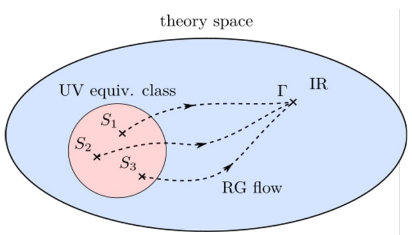


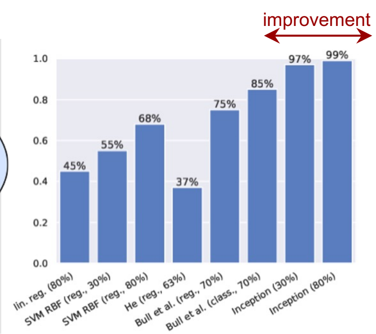
Neural Networks



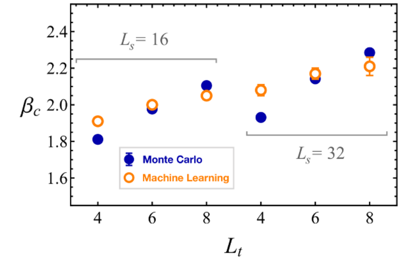
QFT and String Theory

Discoveries from applying neural networks to QFT and string theory



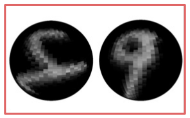


- QFT and renormalization flow for neural networks (NN-QFT): RG scale = weight standard deviation
- Inception neural networks for algebraic topological data (Hodge numbers of Calabi-Yau 3- and 4-folds)
- Volume extrapolation of phase transition for 3d compact QED using neural networks
- Neural network computation of Casimir energy for 2d and 3d lattice scalar field theories



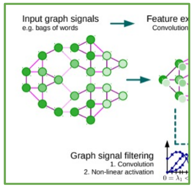
IAIFI Fellow: Siddharth Mishra-Sharma

Cohen et al [ICLR 2018]



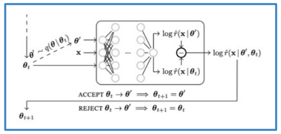
Spherical CNNs

Defferard et al [NIPS 2016, ICLR 2020]

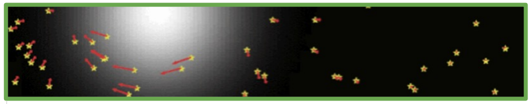


Graph signal processing

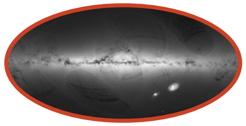
Hermans et al [ICML 2020]

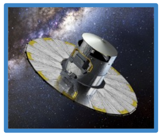


Likelihood-free inference



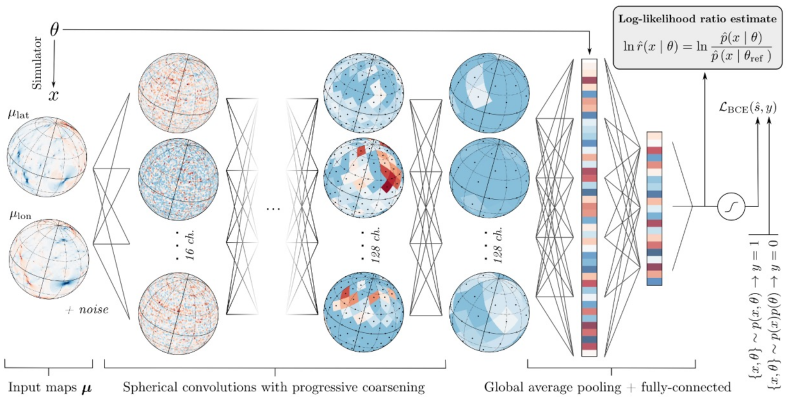
Gravitational lensing of background stars due to dark matter clumps as measured by e.g. the *Gaia* satellite



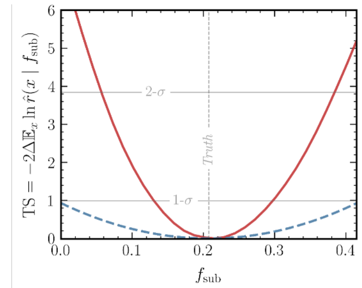


An architecture to extract the dark matter signal

- ★ Account for **structure of data domain** (observations on the celestial sphere)
- ★ Account for **physical symmetries** expected in signal
  - ○ Rotational equivariance *(signal features are similar across the sky)*
  - ○ Rotational invariance *(there is no preferred direction)*
- ★ Infer **statistically-meaningful** quantity: likelihood ratio



Significantly greater sensitivity to dark matter compared to traditional methods



Reference: Mishra-Sharma [MLST 2022; arXiv:2110.01620]

Physics Experiment Research

Enable (experimental) physics discoveries by developing and deploying the next generation of AI technologies

Reinforcement learning

Optimal transport

Normalizing flows

Robust/Interpretable ML

Deep Learning Compression

Data Augmentation

→

Data Reconstruction Algorithms that extract parameters from higher dimensional information

Ultra Fast Real Time data processing

Real Time Detector Controls

Reconciling data and simulation

Discovering New Physics without an underlying physics model

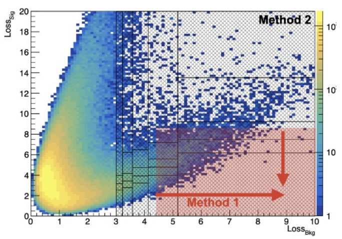
Physics Experiment Overview

- Aiming to improve the operations and enhance the physics potential of various experiments, including the **Large Hadron Collider (LHC)**, the **Deep Underground Neutrino Experiment (DUNE)**, and the **Laser Interferometer Gravitational Wave Observatory (LIGO)**
- Using machine learning techniques for prediction and classification across astrophysics fields
- Applying neural network techniques for imaging and mapping in astronomy and cosmology

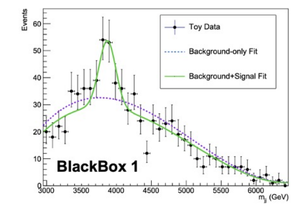
- **Robust AI for Real-Time Applications: Mike Williams, Physics, MIT**
- **High-Dimensional Uncertainty Quantification for Collider Physics: Jesse Thaler, Physics, MIT**
- **Fast Machine Learning: Phil Harris, Physics, MIT**
- **Semi-supervised Anomaly Detection for Physics Discovery: Phil Harris, Physics, MIT**
- **Physics Factorized AI for Measurements of Fundamental Physics Properties: Phil Harris, Physics, MIT**
- **Using Generative Networks in the Reconstruction and Analysis of Neutrino Interactions: Taritree Wongjirad, Physics and Astronomy, Tufts**
- **Galaxy Surveys and CMB Foreground Removal in Cosmology: Douglas Finkbeiner, Astronomy, Harvard**
- **Machine Learning Classification of Optical Transients and Multi-Messenger Astrophysics: Edo Berger, Astronomy, Harvard**
- **Emulating Energy Injection Effects in the Early Universe: Tracy Slatyer, Physics, MIT**
- **Discerning Line-of-Sight Halos from Substructure with Machine Learning: Cora Dvorkin, Physics, Harvard**
- **Improving the Performance of the LIGO Instruments with AI, An Exploration: Lisa Barsotti, Astrophysics and Space Research, MIT**
- **Reconstructing the Primordial Density Field from Galaxy Redshift Surveys: Daniel Eisenstein, Astronomy, Harvard**
- **A Compound Poisson Generator Approach to Point Source Inference in Astrophysics: Kerstin Perez, Physics, MIT**

Semi-Supervised Anomaly Detection for Physics Discovery

Phil Harris, Sang Eon Park, Mikaeel Yunus, Patrick McCormack (MIT),
Matt Schwartz, Bryan Ostdiek, Katie Fraser (Harvard)

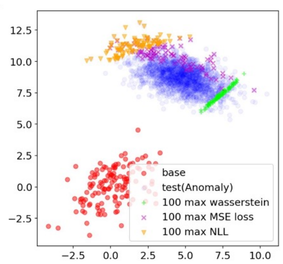


Normalizing Flow Models to identify anomalies



Search for new physics with limited injected priors

New approach to search for anomalous physics models

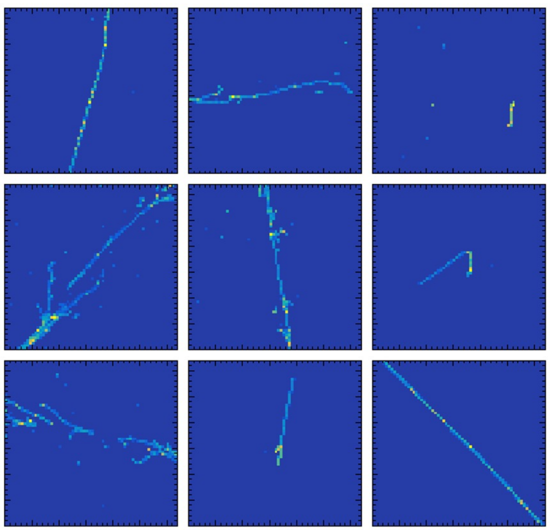


- Approach builds on semi-supervised learning
- Exploring new ways to search for anomalous new physics
- First version of algorithm demonstrated on LHC Olympics
- Further exploration with Optimal transport ideas

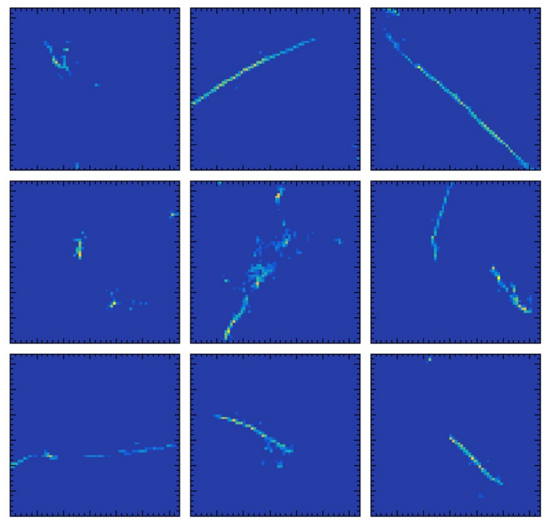
Using Generative Networks in the Reconstruction and Analysis of Neutrino Interactions

Taritree Wongjirad (Tufts), Paul Lutkus (Tufts), Nikita Saxena (Tufts), Jared Hwang (Tufts), Shuchin Aeron (Tufts)

from simulation



from generative network



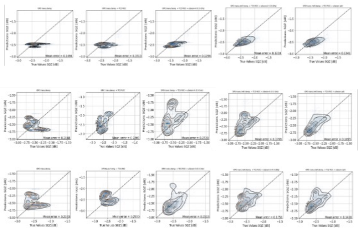
Project Goals

- Generate images of single particles in liquid argon time projection chambers (LArTPCs) conditioned on particle type and momentum.
- Use the generator to improve the reconstruction of neutrino interactions in LArTPCs.

Demonstrated method for quantifying quality of generated images to guide further exploration

Improving the Performance of the LIGO Instruments with AI

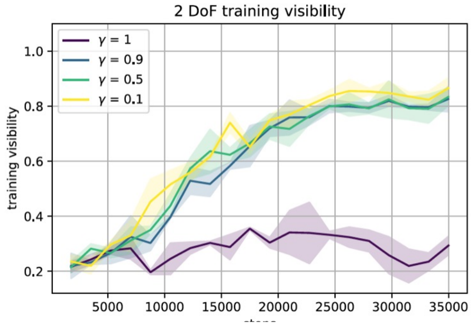
Lisa Barsotti, Christopher Whittle, Dhruva Ganapathy, Ge Yang, Pulkit Agrawal, Matthew Evans (MIT)



Reinforcement learning (RL) agents able to optimally tune systems

Squeezed light to improve LIGO sensitivity to gravitational waves

New RL methods with potential to optimize LIGO squeezing performance over long periods of time

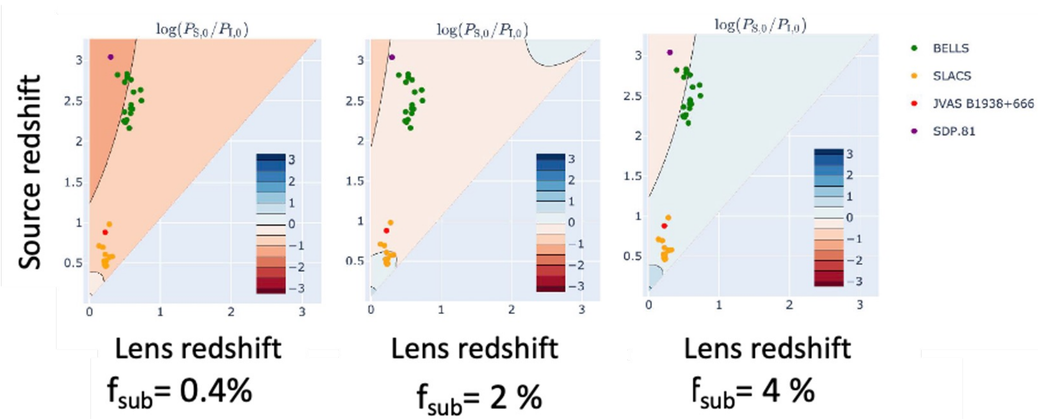


- Used deep learning to understand causes of squeezing degradation on past data
- Made progress in predicting squeezing levels based on a large set of auxiliary channels, selecting those with predictive power
- Built a simulation environment to train a Deep-Q-Network (DQN) RL Agent on simple versions of the optimization problem, specifically how to optimize mirror alignment with 2 and 4 degrees of freedom
- The far reaching goal is to create an agent able to observe the system and react to keep it on its optimal state by modifying a variety of parameters

Discerning Line-of-Sight Halos from Substructure with Machine Learning

Cora Dvorkin, Arthur Tsan, Bryan Ostdiek, Cagan Sengul (Harvard)

Subhalos versus line-of-sight halos

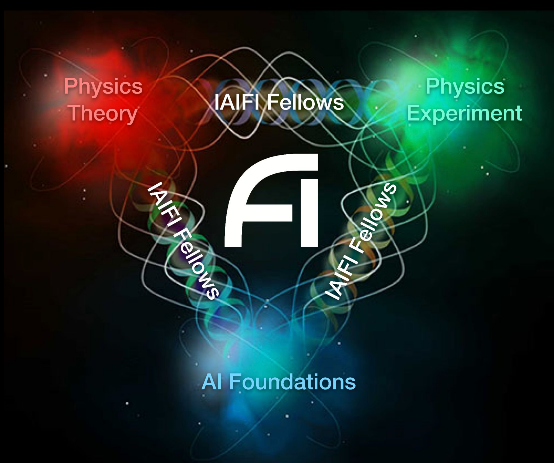


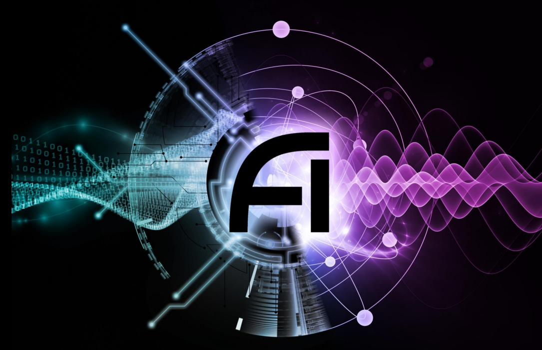
Fraction of dark matter halo mass in substructure (fsub)

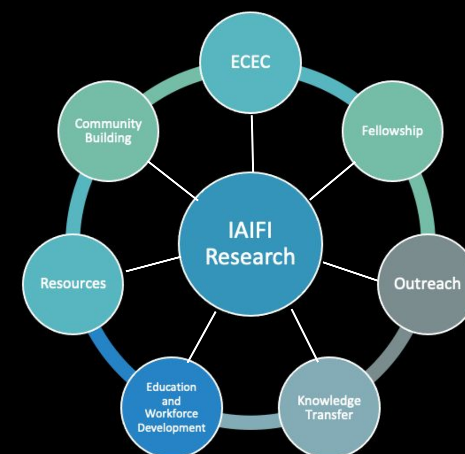
- Direct detection of substructure is computationally very expensive.
- Can we **speed up the process of analyzing the huge number of lensed galaxies** expected with near-future surveys?
- We can predict with high accuracy the curl component of the deflection field caused by halos along the line of sight for a problem without any additional noise.

NSF AI Institute for Artificial Intelligence and Fundamental Interactions (IAIFI /aɪ-faɪ/)

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







https://iaifi.org

iaifi-management@mit.edu

IAIFI Postdoctoral Fellowship

@iaifi-news

Part 2:
Machine learning for Amplitudes

Based on

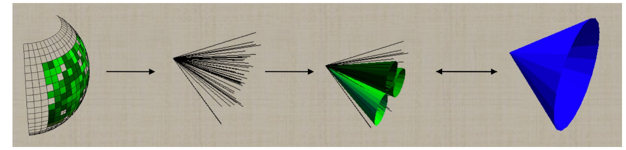
arXiv:2206.04115

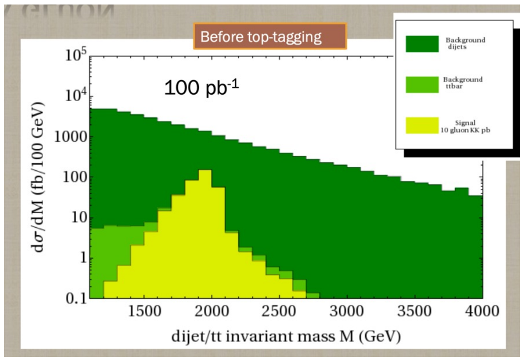
Simplifying Polylogarithms with Machine Learning

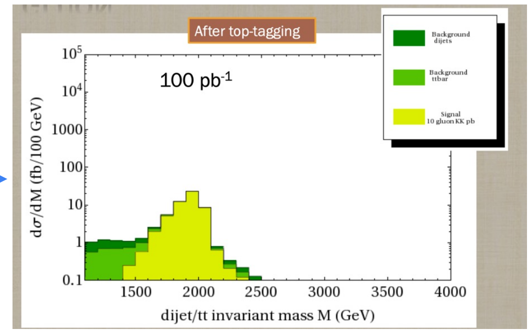
Aurélien Dersy, Matthew D. Schwartz, Xiaoyuan Zhang

Machine learning has transformed collider physics

Top tagging
- Jet substructure approach (2008-2017)
- Mass drop, helicity angle, W subjet

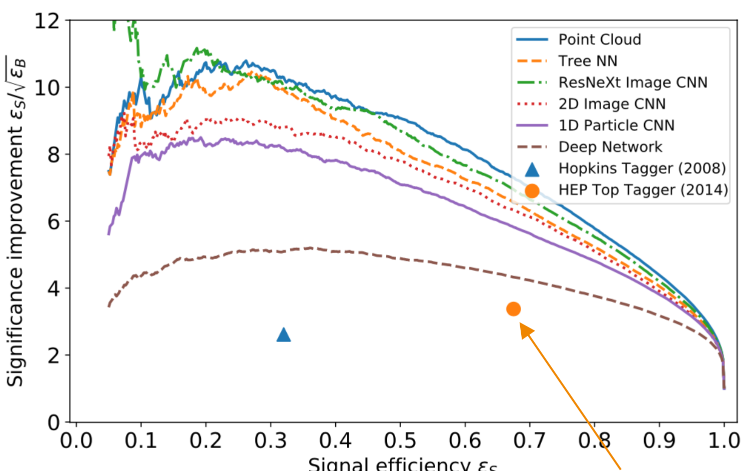






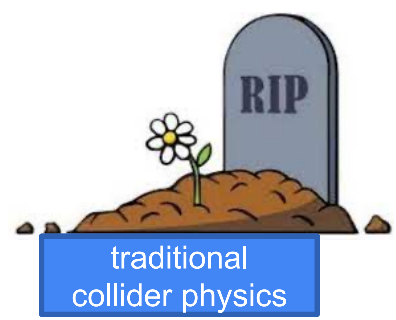
- 30% signal efficiency
- 1% bg efficiency
- Revolutionary at the time

Machine learning methods are much better



best jet subsubstructure algorihm

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



What about the rest of hep-ph? hep-th? What field ML make obsolete next?

- Most ML in physics is highly **numerical**
  - Collider physics applications involve millions of events
  - Data is real numbers
  - Approximate answers are ok

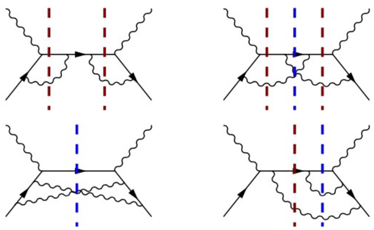
- Most hep-ph and hep-th papers are **symbolic**
  - Model building
  - Approximate but exact solutions to some equation
  - Analytic understanding of some simplified system
  - Loop calculations
    - Analytical computation of loops
    - Numerical implementation for precision physics

Symbolic ML methods will be key to future progress in HEP

Simplifying scattering amplitudes

e.g. Compton scattering at NLO [Lee, Schwartz, Zhang *Phys.Rev.Lett.* 126 (2021)]

Compute some
Feynman diagrams



Reduce to iterated integrals

$$I(s) = \int_1^s \frac{ds'}{\sqrt{(s'-1)(s'+3)}} \int_1^{s'} \frac{ds''}{s''}$$

Do integrals
(not so easy)

$$\begin{aligned} I(s) &= \int_1^{r(s)} d\ln r' \int_1^{r'} d\ln \frac{r''^2 - r'' + 1}{r''} \\ &= \int_1^{r(s)} \frac{dr'}{r'} \int_1^{r'} dr'' \left(-\frac{1}{r''} + \frac{1}{r'' - z_1} + \frac{1}{r'' - z_2} \right) \\ &= \frac{2\pi^2}{9} - \frac{\ln^2 r}{2} + \left(-\frac{i\pi}{3} + \ln r\right) \ln(r - z_1) + \ln(r z_1)\ln(r - z_2) \\ &\quad - \mathrm{Li}_2(1 - z_1) + \mathrm{Li}_2(1 - r z_1) - \mathrm{Li}_2(1 - z_2) + \mathrm{Li}_2(1 - r z_2) \end{aligned}$$

$$z_1 = \frac{1+i\sqrt{3}}{2}$$

$$z_2 = \frac{1-i\sqrt{3}}{2}$$

- result is complicated
- simplifying result ~ 1 month

$$I(s) = -\frac{1}{3}\mathrm{Li}_2(-r^3) + \mathrm{Li}_2(-r) - \frac{1}{2}\ln^2 r + \frac{\pi^2}{18}$$

Why is simplification important?

- Removes unphysical singularities
- A lot of physics in analytic structure
- Simple form indicates deeper structure
- Simplification at intermediate steps make full calculation tractable
- In a sense, **all of science is simplification**

How do we simplify polylogarithms?

Logarithm is easy: only one identity

$$\ln(xy) = \ln x + \ln y$$

Dilogarithms have lots of identities

$$(\textit{inversion}) \quad \mathrm{Li}_2(x) = -\mathrm{Li}_2\left(\frac{1}{x}\right) - \frac{\pi^2}{6} - \frac{\log^2(-x)}{2}$$

$$(\textit{reflection}) \quad \mathrm{Li}_2(x) = -\mathrm{Li}_2(1-x) + \frac{\pi^2}{6} - \log(x)\log(1-x)$$

$$(\textit{duplication}) \quad \mathrm{Li}_2(x) = -\mathrm{Li}_2(-x) + \frac{1}{2}\mathrm{Li}_2(x^2)$$

5-term identity:

$$\mathrm{Li}_2(x) + \mathrm{Li}_2(y) + \mathrm{Li}_2\left(\frac{1-x}{1-xy}\right) + \mathrm{Li}_2(1-xy) + \mathrm{Li}_2\left(\frac{1-y}{1-xy}\right)$$
$$= \frac{\pi^2}{2} - \ln(x)\ln(1-x) - \ln(y)\ln(1-y) - \ln\left(\frac{1-x}{1-xy}\right)\ln\left(\frac{1-y}{1-xy}\right)$$

Li_3, Li_4, etc. have identities too (complete set not known)

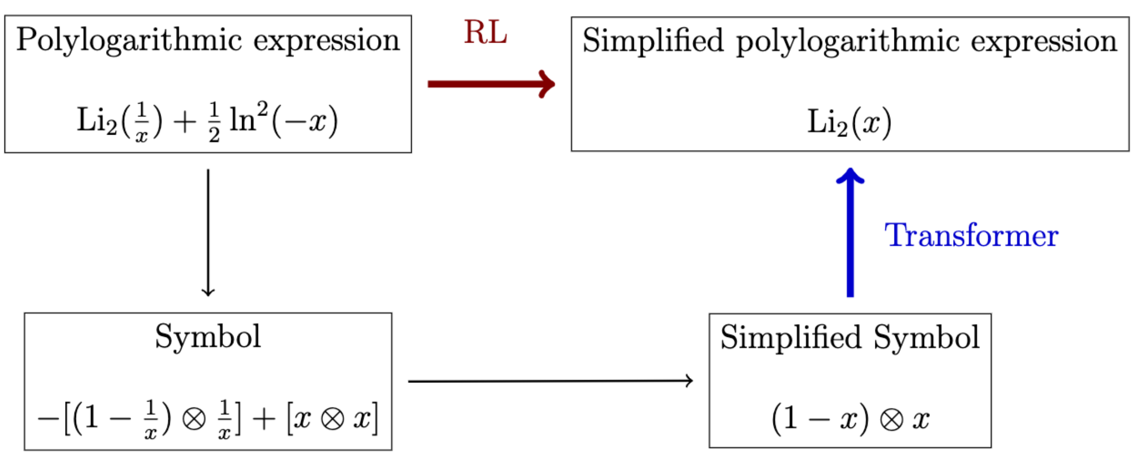
$$\mathrm{Li}_3(x) = \mathrm{Li}_3\left(\frac{1}{x}\right) - \frac{1}{6}\ln^3(-x) - \frac{\pi^2}{6}\ln(-x)$$

Problem statement:

Given some polylogarithmic expression:

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

1. What is its simplest form?
2. Does it simplify to zero?
3. What identities do we apply in what order to simplify it?



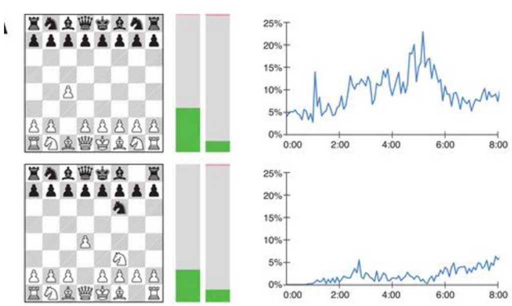
Two approaches
1. Reinforcement learning
2. Transformer networks

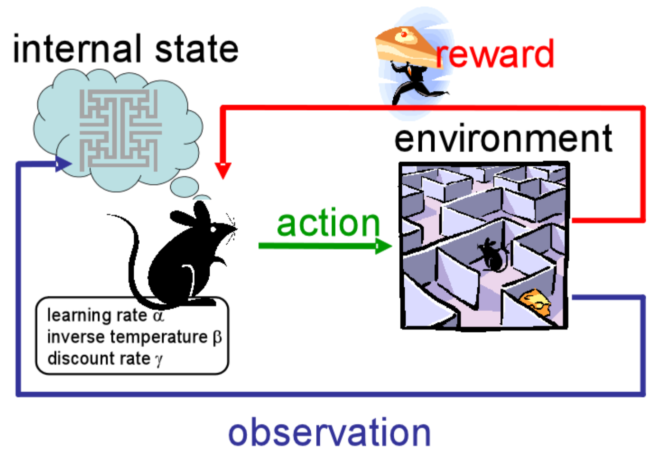
Reinforcement learning

- Applying identities can be viewed as moves in a game

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

Killer Ap: AlphaZero
(Deepmind, 2017)





Tries action to take based on state
- Success reinforces good choices
- Learns best move given state

Reinforcement learning

Reward $r_t = \begin{cases} 1 & \text{if } N_t^{\text{dilogs}} < N_{t'}^{\text{dilogs}} \quad \forall\, t' < t \\ 0 & \text{else} \end{cases}$ (# dilogs goes down)

- Simple reward works better than more sophisticated ones (depending on expression length for example)
- Analog of "taking the king" in chess
  - RL learns more sophisticted value function during training

Dataset: Linear combinations of dilogarithms that reduce to 0

$$0 = 2\,\mathrm{Li}(1-x) - 2\,\mathrm{Li}(1-x)$$

Duplication

$$-2\,\mathrm{Li}(x-1) + \mathrm{Li}(x^2) - 2\,\mathrm{Li}(1-x)$$

Inversion

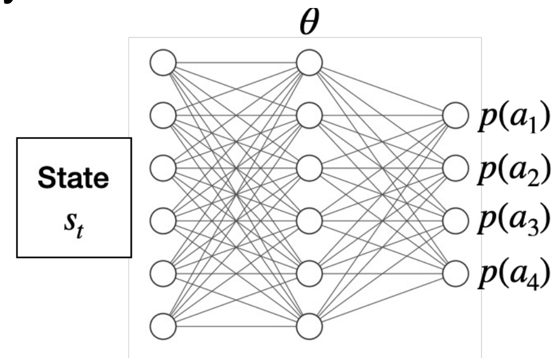
$$2\,\mathrm{Li}\left(\frac{1}{x-1}\right) + \mathrm{Li}(x^2) - 2\,\mathrm{Li}(1-x)$$

Generate 13,500 training expressions + 1,500 testing expressions

Reinforcement learning

Policy and value: learned by RL as neural networks

$\pi_{\varphi,\theta}(a|s)$ $V_{\varphi,\theta}(s)$

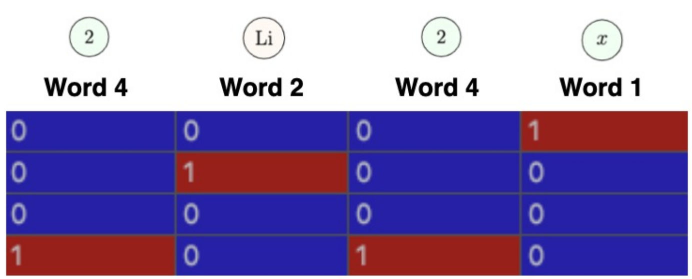


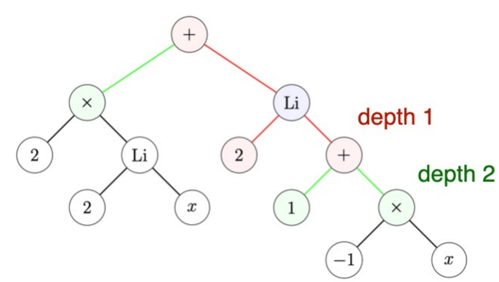
Agent: We use Trust Region Policy Optimization (**TRPO**) [Schulman et al, 2017]

- Ensures the policy updates stay close
- Maximizes the advantage of the new policy over the old policy
- (also tried Proximal Policy Optimization (PPO), but not as good)

Sentence embedding:

- Explored both **one-hot encoding** and **graph neural networks**



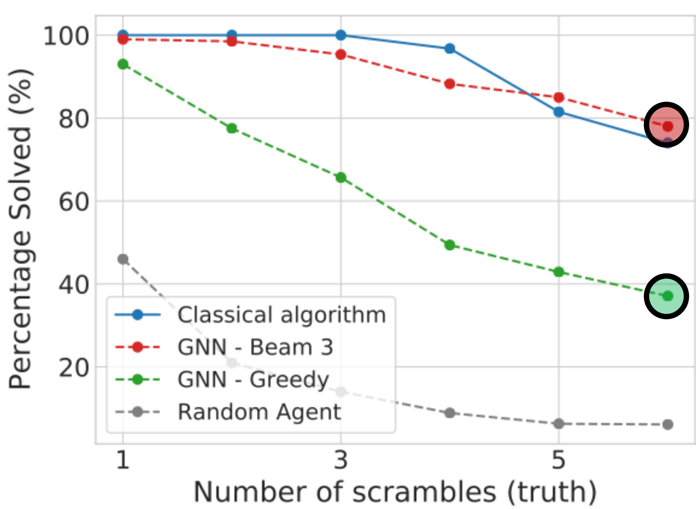


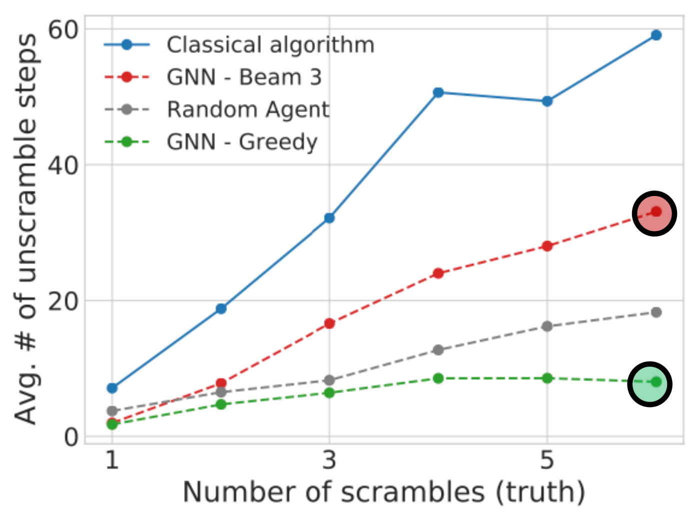
Results

Greedy: best guess only

Beam 3: try best 3 choices

Reward	Agents	Greedy		Beam size 3	
		Solved (%)	# Unscramble steps	Solved (%)	# Unscramble steps
No penalty	One-hot	50 %	5.3	78 %	14.7
	GNN	56 %	6.4	80 %	15.9
Δr_t penalty	One-hot	59 %	7.4	85 %	19.2
	GNN	53 %	8.7	**89 %**	**20.3**
	Random	13 %	8.7		
	Classical	**91 %**	**39.3**		





Does as well as classical algorithm with many fewer steps

- Has learned something!

Transformer networks

An alternative approach to simplifying polylogarithms is to translate them

Attention is all you need

A **Vaswani**, N **Shazeer**, N **Parmar**… - Advances in neural …, 2017 - proceedings.neurips.cc

The dominant sequence transduction models are based on complex recurrent orconvolutional neural networks in an encoder and decoder configuration. The best performing such …

☆ Save 99 Cite Cited by 43724 Related articles All 35 versions

- 13 cites in hep-ph
- 15 cites in hep-th

```
This Nondisclosure Agreement (the "Agreement") is entered into by and between __
____________ with its principal offices at ______________, ("Disclosing Part
y") and ______________, located at _____________ ("Receiving Party") for the
purpose of preventing the unauthorized disclosure of Confidential Information as
defined below. The parties agree to enter into a confidential relationship conce
rning the disclosure of certain proprietary and confidential information ("Confi
dential Information"). 1. Definition of Confidential Information. For purposes o
f this Agreement, "Confidential Information" shall include all information or ma
terial that has or could have commercial value or other utility in the business
in which Disclosing Party is engaged. If Confidential Information is in written
form, the Disclosing Party shall label or stamp the materials with the word "Con
fidential" or some similar warning. If Confidential Information is transmitted o
rally, the Disclosing Party shall promptly provide writing indicating that such
oral communication constituted Confidential Information. 2. Exclusions from Conf
idential Information. Receiving Party's obligations under this Agreement do not
extend to information that is: (a) publicly known at the time of disclosure or s
ubsequently becomes publicly known through no fault of the Receiving Party; (b)
discovered or created by the Receiving Party before disclosure by Disclosing Par
ty; (c) learned by the Receiving Party through legitimate means other than from
the Disclosing Party or Disclosing Party's representatives; or (d) is disclosed
by Receiving Party with Disclosing Party's prior written approval.
```

GPT-3 $\longrightarrow$ Translate for 2nd grader

This means that if you tell anyone else something I told you in private, I can sue you and make you pay me lots of money. It also means that if I tell you something in private, I can sue you if you tell anyone else. It also means that if I tell you something in private and then I change my mind and want to tell everyone else about it, I can sue you if you tell anyone else first.

What are 5 key points I should know when studying multiple polylogarithms ?

How can you simplify dilogarithms?

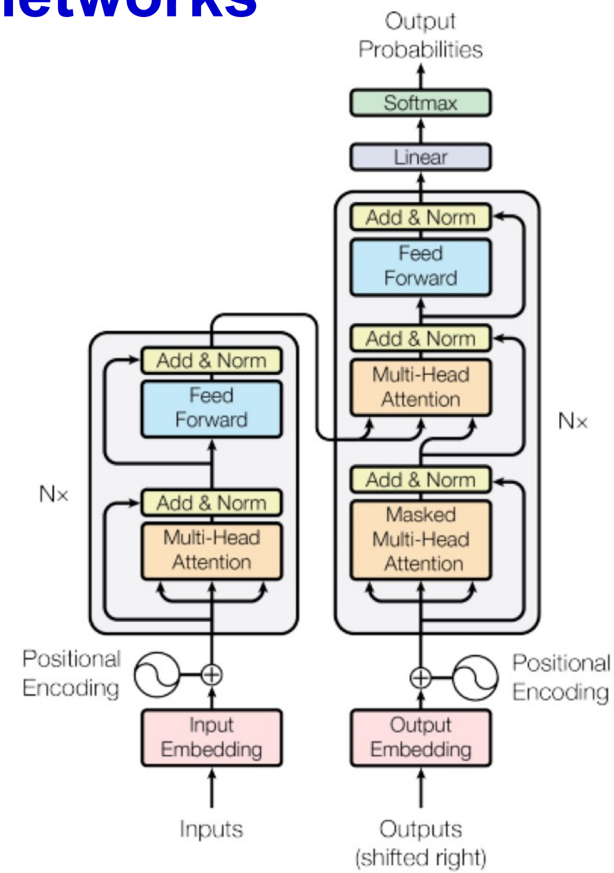
There is no general way to simplify dilogarithms. However, there are some special cases where dilogarithms can be simplified. For example, if the argument of the dilogarithm is 1, then the dilogarithm can be simplified to 0.

```
In[1]:= PolyLog[2, 1]
```

Out[1]= $\frac{\pi^2}{6}$

Hmm...

Transformer networks



Key Facts

- Sequence to sequence model
- Encoder and Decoder structure
- Used in translation tasks
- Uses ***attention blocks***

Attention allows it to appreciate context

The animal didn't cross the street because it was too tired.
L'animal n'a pas traversé la rue parce qu'il était trop fatigué.

The animal didn't cross the street because it was too wide.
L'animal n'a pas traversé la rue parce qu'elle était trop large.

How can we use transfomers?

Treat polylogarithmic expressions as sentences

$$\mathrm{Li}_2\left(\tfrac{1}{x}\right) + \tfrac{1}{2}\ln^2(-x) \longrightarrow \mathrm{Li}_2(x)$$

English — French

Data generation

- Can no longer only consider functions that simplify to zero

$$f(x) = \sum_{i=1}^{n_s} a_i \,\mathrm{Li}_2(g_i(x)) + \sum_{i=1}^{n_t} \left[c_j \,\mathrm{Li}_2(h_j(x)) - c_j \,\mathrm{Li}_2(h_j(x))\right]$$

Scramble with dilog identities

e.g.

$$6\,\mathrm{Li}_2(1-x) - 4\,\mathrm{Li}_2\left(-\frac{1}{x^2-x+1}\right)$$

Scramble

$$\xrightarrow{\hspace{2em}}$$

$$-4\,\mathrm{Li}_2\left(x^2-x+1\right) + 3\,\mathrm{Li}_2\left(x^2-2x+1\right) - 3\,\mathrm{Li}_2\left(\frac{1}{x^2-4x+4}\right)$$
$$+\frac{5}{2}\,\mathrm{Li}_2\left(\frac{x^2+2x+1}{4x^2+4x+1}\right) - 4\,\mathrm{Li}_2(-x-1) - 5\,\mathrm{Li}_2\left(\frac{x+1}{2x+1}\right)$$
$$+2\,\mathrm{Li}_2\left(x^4-2x^3+3x^2-2x+1\right) + 6\,\mathrm{Li}_2\left(\frac{1}{x-2}\right) - 4\,\mathrm{Li}_2(x+2)$$
$$-3\,\mathrm{Li}_2\left(\frac{1}{2x+1}\right) - 5\,\mathrm{Li}_2\left(\frac{-x-1}{2x+1}\right) - 3\,\mathrm{Li}_2\left(\frac{2x}{2x+1}\right)$$

Transformer vs RL

- For RL, 0 is unique correct answer
  - RL tries to find path to 0
  - Knows when it succeeds
- For transformer, many equivalent simple expressions
  - Path not determined, just result
  - Not guaranteed to be correct

- Can be many equally simple translations

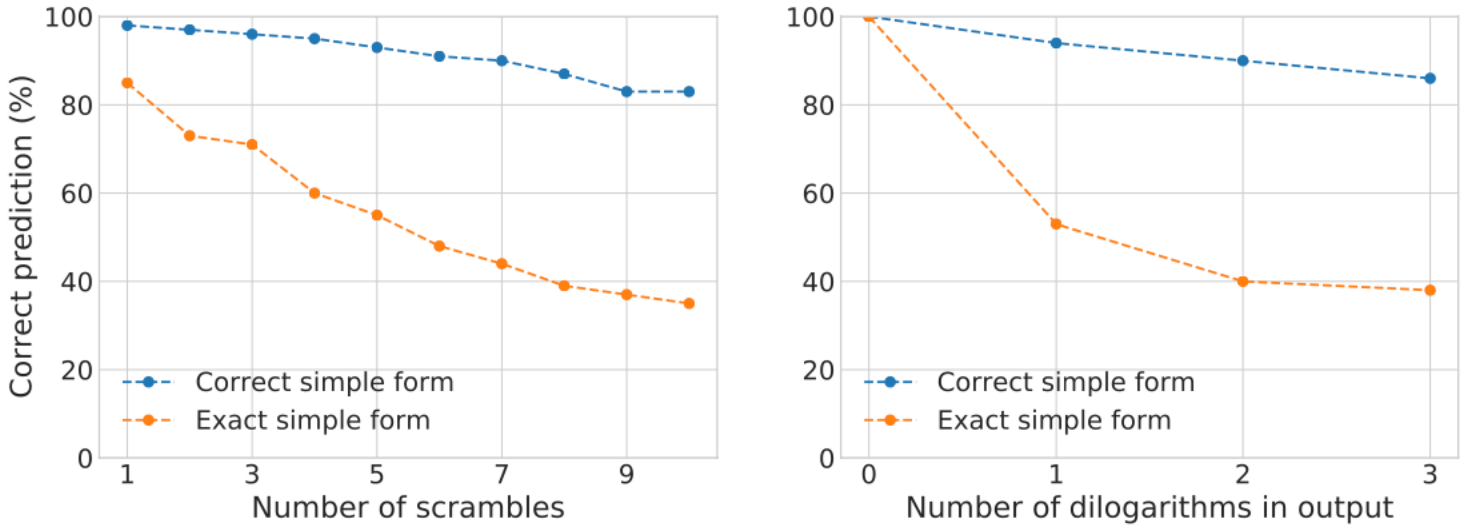
Input :	$-\mathrm{Li}_2\left(-2x^2\right)+\mathrm{Li}_2\left(-\frac{1}{2x^2}\right)$ $-7\mathrm{Li}_2(-2x)-7\mathrm{Li}_2\left(-\frac{1}{2x}\right)$ $4\mathrm{Li}_2\left(x^2-2x+2\right)$
Target :	$4\mathrm{Li}_2\left(-x^2+2x-1\right)$

$\xrightarrow{\text{transformer}}$

Hypothesis	Valid ?
$-4\mathrm{Li}_2\left(x^2-2x+2\right)$	$\checkmark^\star$
$-4\mathrm{Li}_2\left(-\frac{1}{x^2-2x+1}\right)$	$\checkmark^\star$
$4\mathrm{Li}_2\left(-x^2+2x-1\right)$	$\checkmark$
$4\mathrm{Li}_2\left(\frac{1}{x^2-2x+2}\right)$	$\checkmark^\star$
$-4\mathrm{Li}_2\left(-\frac{1}{x^2+2x+1}\right)$	$\times$

Beam 5

Results



Predicts a correct simplified answer 91% of the time!

Symbols

An alternative approach to simplifying polylogarithms is with the **symbol**

[Goncharov, Spradlin, Vergu, Volovich PRL 2010]

- Used symbol to simplify 17 page 2-loop 6 point amplitude to a few lines

Polylogarithms are iterated integrals

$$\mathrm{Li}_1(s) \equiv \int_0^s \frac{dx}{1-x} = -\ln(\qquad \mathcal{S}\left[\frac{1}{n!}\ln^n x\right] = \underbrace{x \otimes \cdots \otimes x}_{n}$$

etc.

Symbol is a map that extracts the dlog forms

$$\mathcal{S}\left[\int_a^b d\ln R_1 \circ \cdots \circ d\ln R_n\right] = R_1 \otimes \cdots \otimes R_n$$

$$\mathcal{S}[\mathrm{Li}_n(x)] = -(1-x) \otimes \underbrace{x \otimes \cdots \otimes x}_{n-1}$$

$$\mathcal{S}\left[\frac{1}{n!}\ln^n x\right] = \underbrace{x \otimes \cdots \otimes x}_{n}$$

Symbols

Symbol satisfies the product rule (and other identities, but this is the most powerful)

$$\cdots \otimes f(x)g(y) \otimes \cdots = \cdots \otimes f(x) \otimes \cdots + \cdots \otimes g(y) \otimes \cdots$$

- Reduces simplifying $\mathrm{Li}_n(x\,y)$ to simplifying $\log(xy)$

e.g.

$$[(1-\frac{1}{x})\otimes x] = [-\frac{1-x}{x}\otimes x] = [(1-x)\otimes x] - [x\otimes x]$$

$$\mathcal{S}\left[\mathrm{Li}_2(x)+\mathrm{Li}_2\left(\frac{1}{x}\right)+\frac{1}{2}\ln^2(-x)\right] = -[(1-x)\otimes x] - [(1-\frac{1}{x})\otimes\frac{1}{x}] + [x\otimes x] = 0$$

No free lunch: integating the symbol back to polylogarithms not trivial

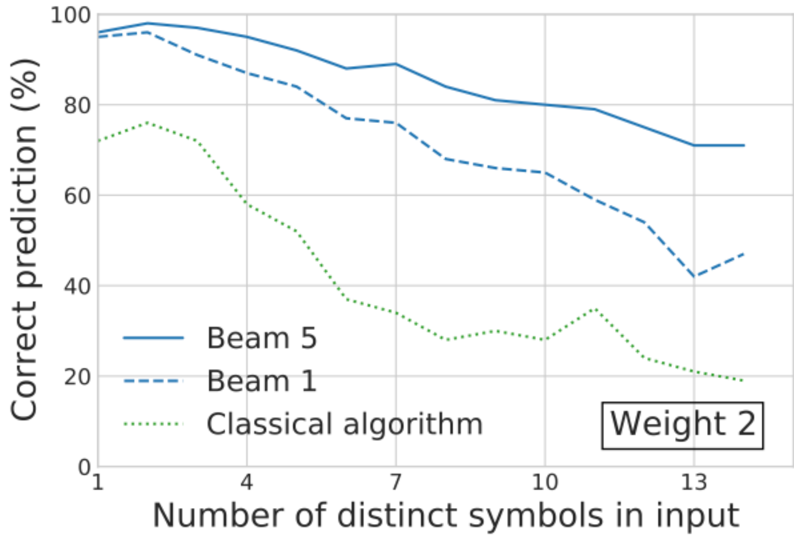
$$\begin{tikzcd}
\boxed{\begin{array}{c}\text{Polylogarithmic expression}\\[1em] \mathrm{Li}_2(\frac{1}{x})+\frac{1}{2}\ln^2(-x)\end{array}} \arrow[r, "\text{RL}", "\text{or transformer}"'] \arrow[d] & \boxed{\begin{array}{c}\text{Simplified polylogarithmic expression}\\[1em] \mathrm{Li}_2(x)\end{array}} \\
\boxed{\begin{array}{c}\text{Symbol}\\[1em] -[(1-\frac{1}{x})\otimes\frac{1}{x}]+[x\otimes x]\end{array}} \arrow[r] & \boxed{\begin{array}{c}\text{Simplified Symbol}\\[1em] (1-x)\otimes x\end{array}} \arrow[u, "\text{Transformer}"']
\end{tikzcd}$$

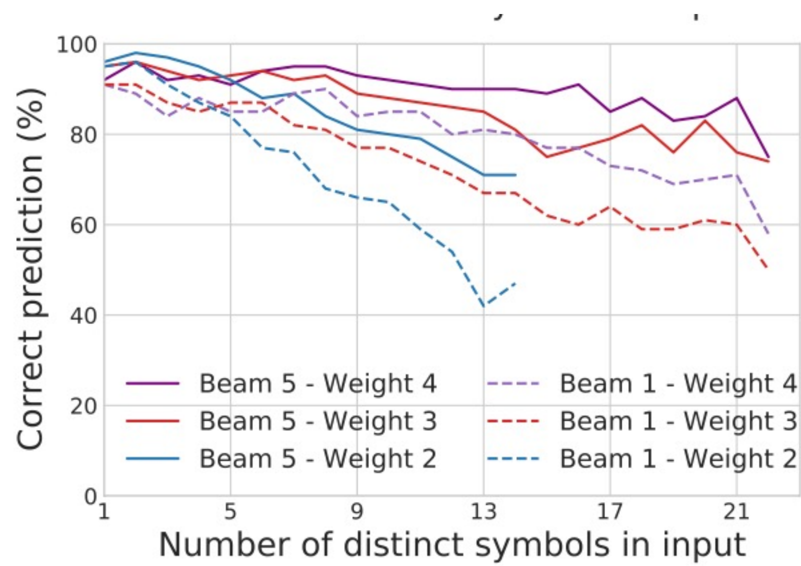
Example symbol integration

Training data looks like this

Input symbol $\mathcal{S}_i$	Simple expression F_i
$-\left(-x^2-x+1\right)\otimes(1-x)+\left(-x^2-x+1\right)\otimes x$ $-\left(-x^2-x+1\right)\otimes(x+1)+x\otimes(1-x)-x\otimes x+x\otimes(x+1)$	$\mathrm{Li}_2\left(\frac{(1-x)(x+1)}{x}\right)$
$-\frac{23}{4}(1-x)\otimes x-6\left(x^2+x+1\right)\otimes x$	$2\,\mathrm{Li}_2\left(x^3\right)-\frac{1}{4}\,\mathrm{Li}_2(x)$
$-40\left(6-x^2\right)\otimes\left(6-x^2\right)-3(1-x)\otimes\left(-x^6-x^2+3\right)$ $-3(x+1)\otimes\left(-x^6-x^2+3\right)-3\left(x^4+x^2+2\right)\otimes\left(-x^6-x^2+3\right)$ $+\frac{1}{4}(5-2x)\otimes(2-x)$	$3\mathrm{Li}_2\left(-x^6-x^2+3\right)$ $-\frac{1}{4}\,\mathrm{Li}_2(2x-4)$ $-20\ln^2\left(x^2-6\right)$
$8\,\frac{x^2-x-1}{x-1}\otimes x-8\left((x+1)\left(x^2-x-1\right)\right)\otimes x$ $+8(1-x)\otimes\left(-x^3+x^2-x-1\right)-8\frac{1}{x-1}\otimes x$ $-8(1-x)\otimes\left(x\left(x^3-x^2+x+1\right)\right)$	$4\,\mathrm{Li}_2\left(x^2\right)$

Results





		Beam Size 1	Beam Size 5
Weight 2	Transformer	82%	**91%**
	Classical Algorithm	59%	59%
Weight 3	Transformer	78%	**88%**
Weight 4	Transformer	80%	**89%**

Weight 2	Weight 3	Weight 4
$\mathrm{Li}_2(x)$	$\mathrm{Li}_3(x)$	$\mathrm{Li}_4(x)$
$\ln(x)\ln(y)$	$\mathrm{Li}_2(x)\ln(y)$	$\mathrm{Li}_3(x)\ln(y)$
	$\ln(x)\ln(y)\ln(z)$	$\mathrm{Li}_2(x)\mathrm{Li}_2(y)$
		$\mathrm{Li}_2(x)\ln(y)\ln(z)$
		$\ln(w)\ln(x)\ln(y)\ln(z)$

Can simplify complicated expressions with Li_2, Li_3, Li_{4},...

- Limited by compute (network size, training time)

Final example

1. Scattering amplitude gives some function of GPLs with complex arguments

$$f(x) = 4\zeta_3 + 9\left[G(0,0,1,x) + G\left(0,0,\frac{-1-\sqrt{3}i}{2},x\right) + G\left(0,0,\frac{-1+\sqrt{3}i}{2},x\right)\right]$$
$$+4\left[-G(-1,-1,-1,x) + G(-1,0,-1,x) + G(0,-1,-1,x) + G(0,0,1,x) - G\left(0,0,1,\frac{x}{x+1}\right)\right]$$

2. Express in terms of classical polylogs

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

3. Compute the symbol and simplify

$$\mathcal{S}[f(x)] = 9\left(x^2+x+1\right)\otimes x\otimes x + 13(1-x)\otimes x\otimes x + 4(x+1)\otimes x\otimes x$$

4. Integate the symbol with a transformer network

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

Summary

We considered the problem of simplifying polylogs with ML

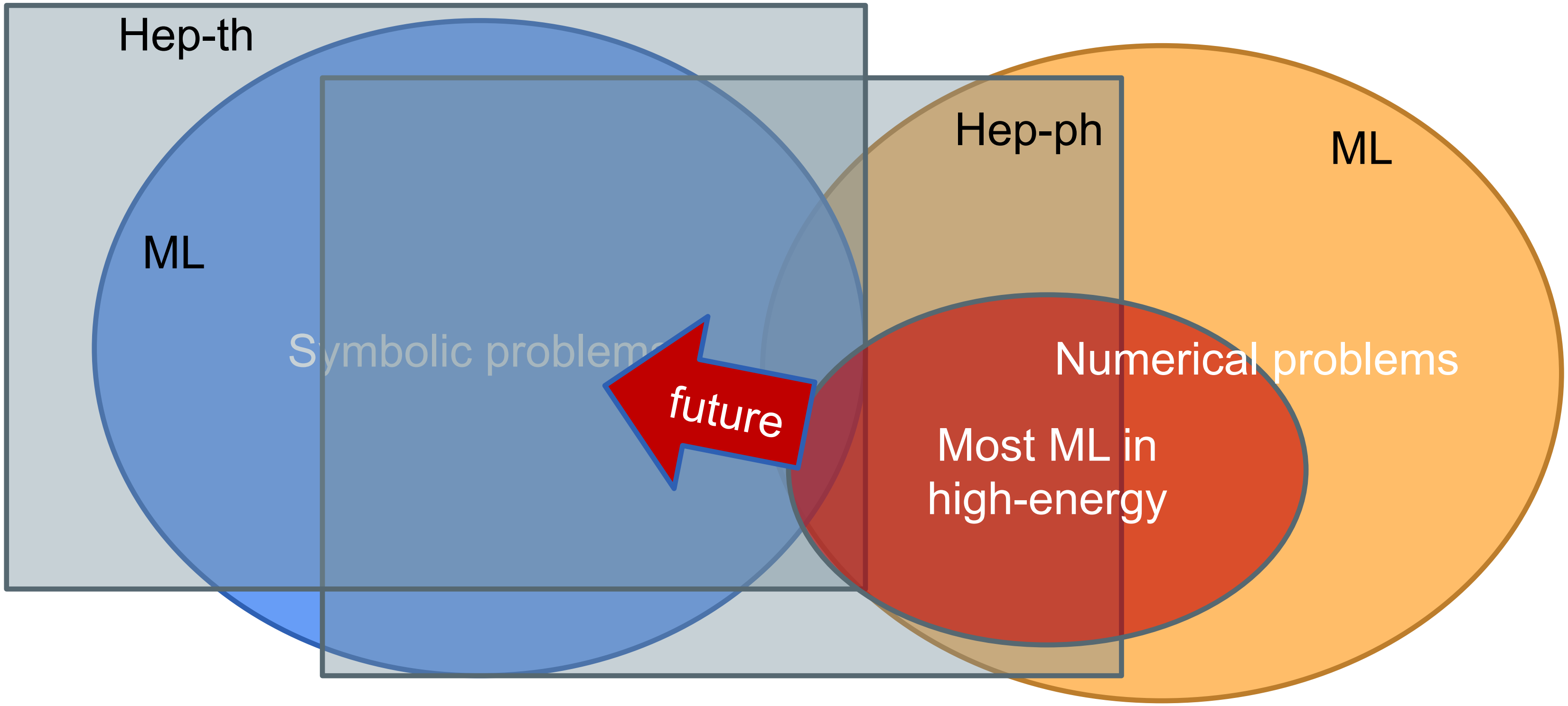
$$-\mathrm{Li}_2\left(-2x^2\right)+\mathrm{Li}_2\left(-\frac{1}{2x^2}\right) -7\mathrm{Li}_2(-2x)-7\mathrm{Li}_2\left(-\frac{1}{2x}\right) 4\mathrm{Li}_2\left(x^2-2x+2\right) \longrightarrow 4\mathrm{Li}_2\left(-x^2+2x-1\right)$$

- Reinforcement learning
  - Applies identities like moves in a game
  - Learns which identity to apply

- Transformer network
  - Guesses the answer like in langauge translation
  - Can be used to integrate the symbol
  - Powerful method works for high-weight polylogs

- Methods around 90% accurate
- Perform non-trivial task better than any public algorithm

Outlook



- Machine learning has revolutionaized data-driven particle physics
- Most work in high energy theory is symbolic
- Machine learning can help with that too!