

第五届量子场论及其应用研讨会 @ PKU (Beijing)



NNU · 南京师范大学
NANJING NORMAL UNIVERSITY

Artificial Intelligence & High Energy Physics

Lei Wu

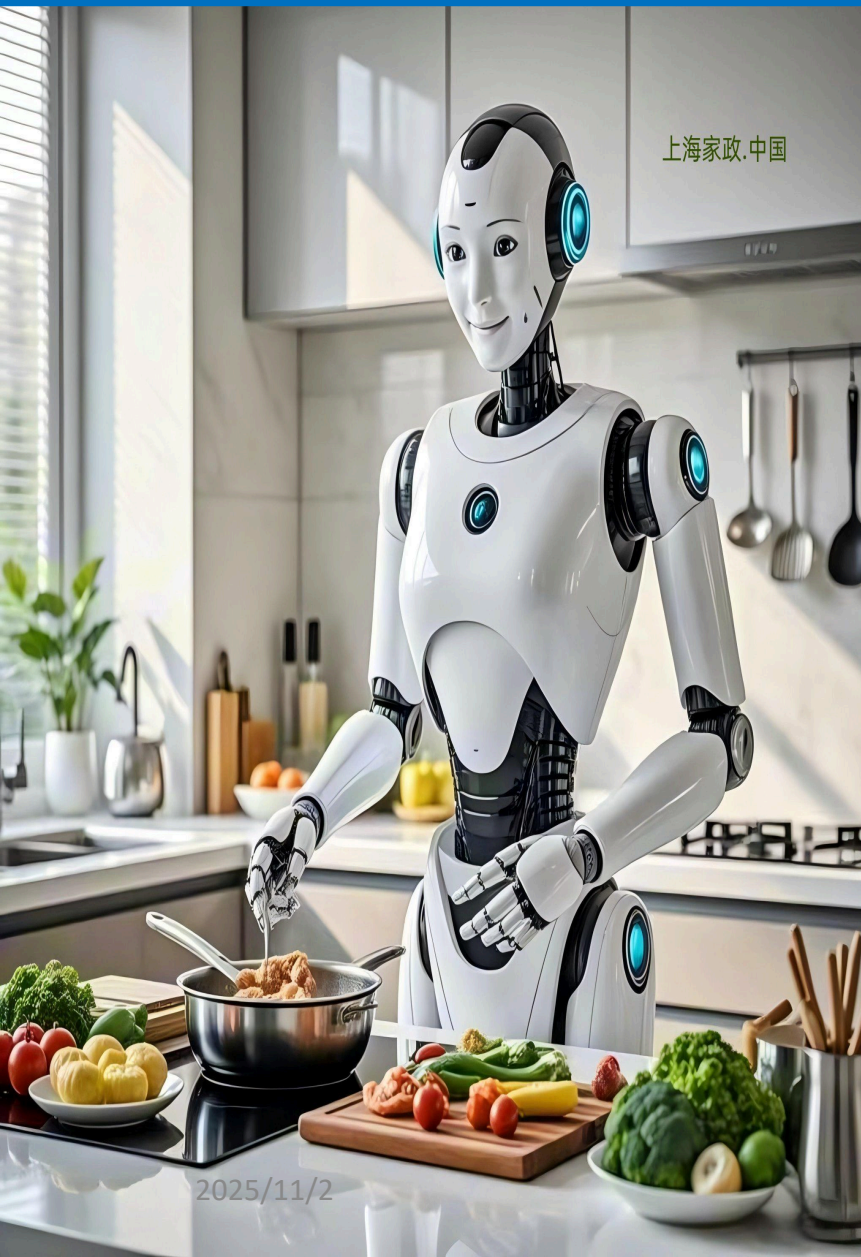
With H. Lv, A. Murat, J. Ren, D. Wang, J. M. Yang

2304.03237; 2203.14569
2106.07018; 2012.09874
1905.06047; 1901.05627
1807.09088; 1708.06615

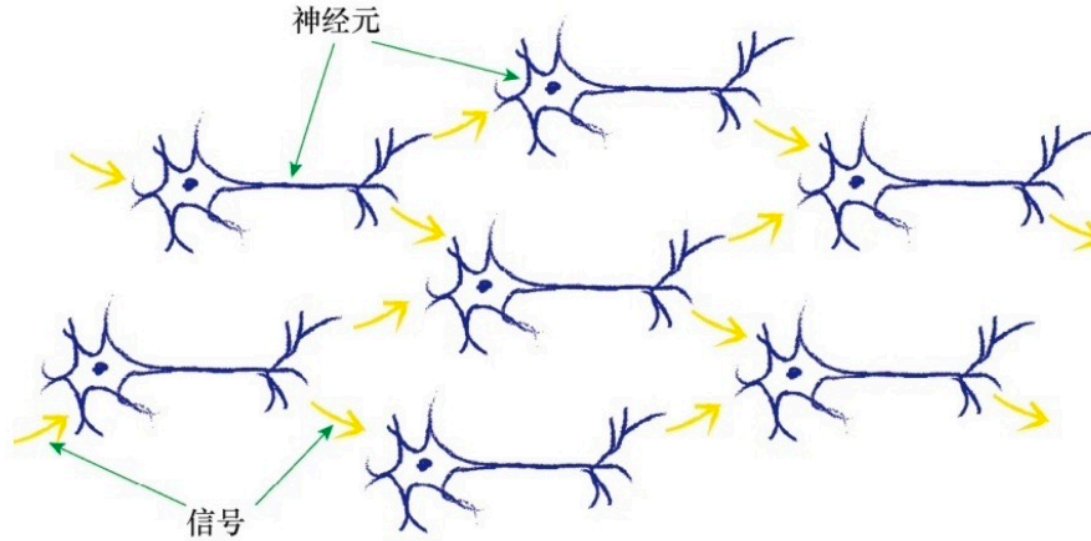
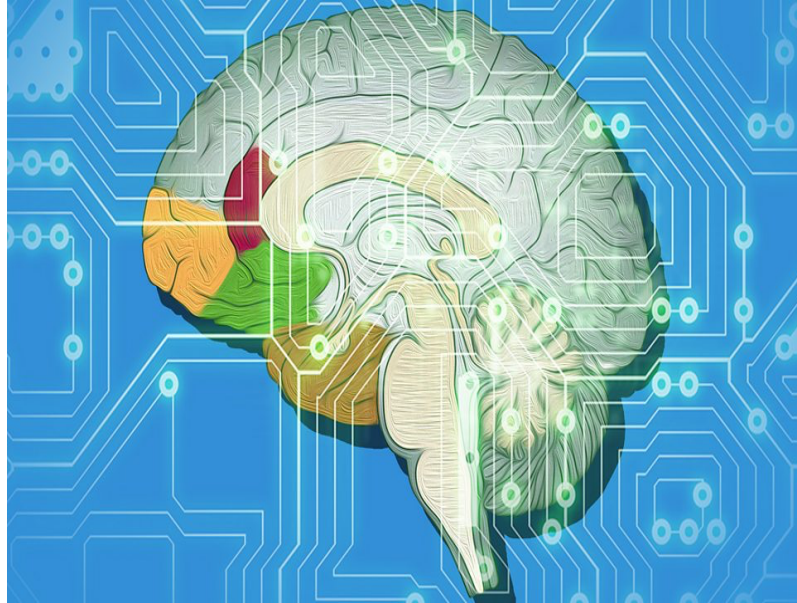
Outline

- **ABC of AI**
- **AI for HEP: from data to theory**
- **Prospects and Challenges**

1. A brief history of AI



1. A brief history of AI

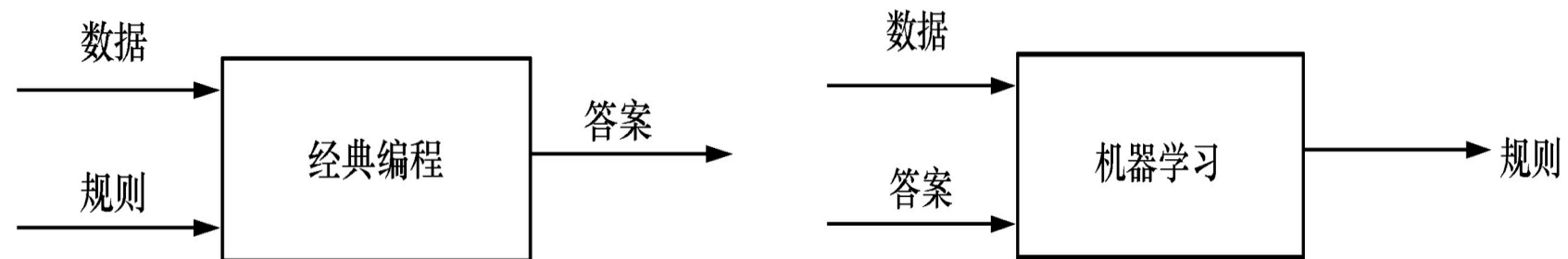
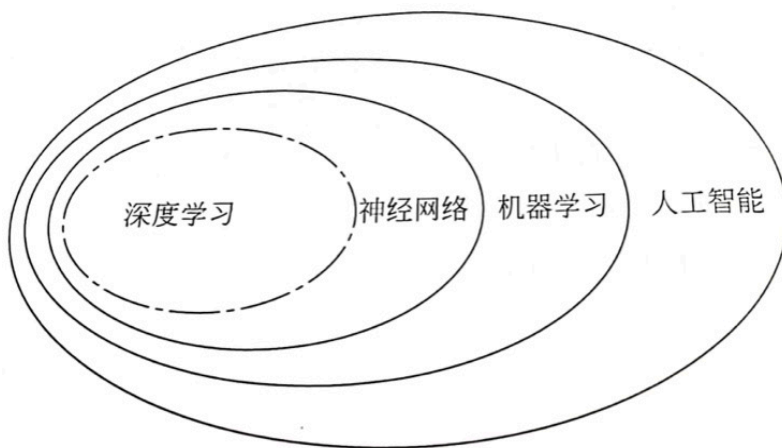


大约1000亿个神经元

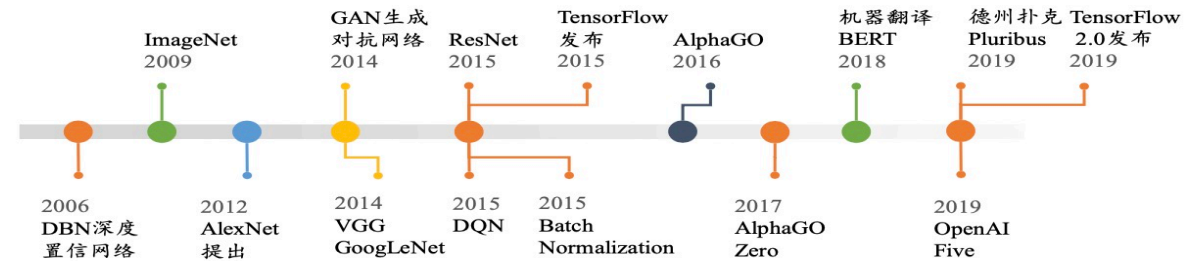
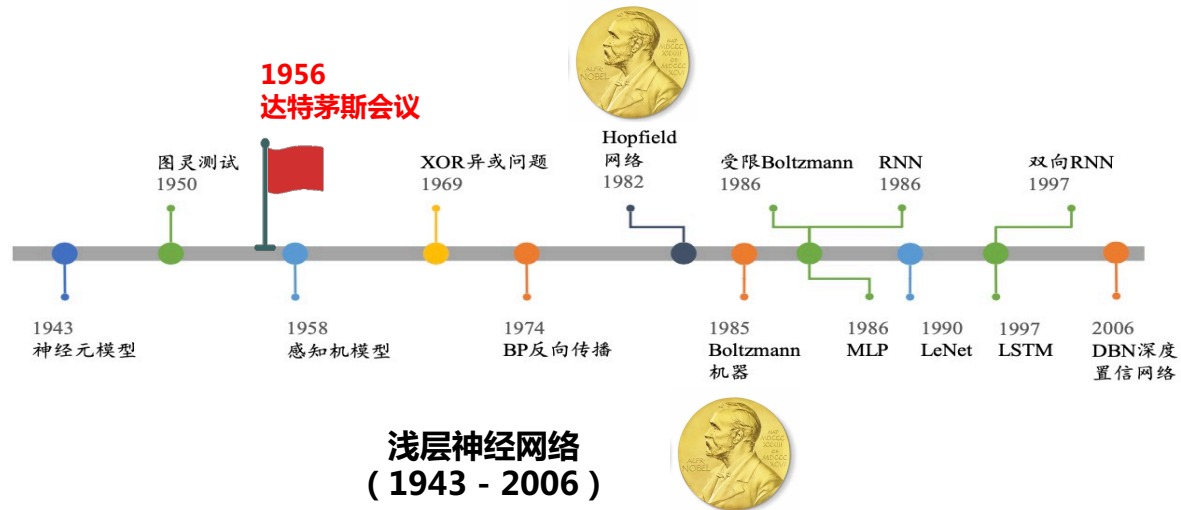
突触传递信号，
形成神经网络

学习、迁移、创造

大脑耗能：0.5度/天，ChatGPT：30w度/天

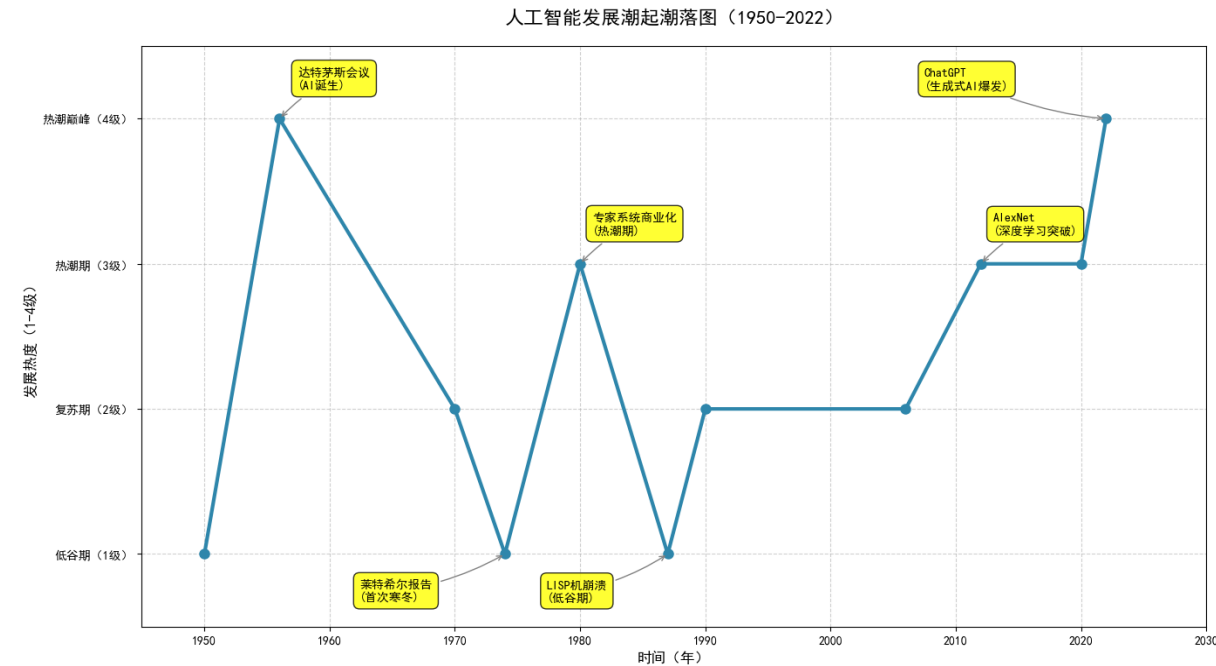


1. A brief history of AI



网络结构创新与优化
(2006 - 2015)

强化学习与自然语言处理的突破
(2016 -)



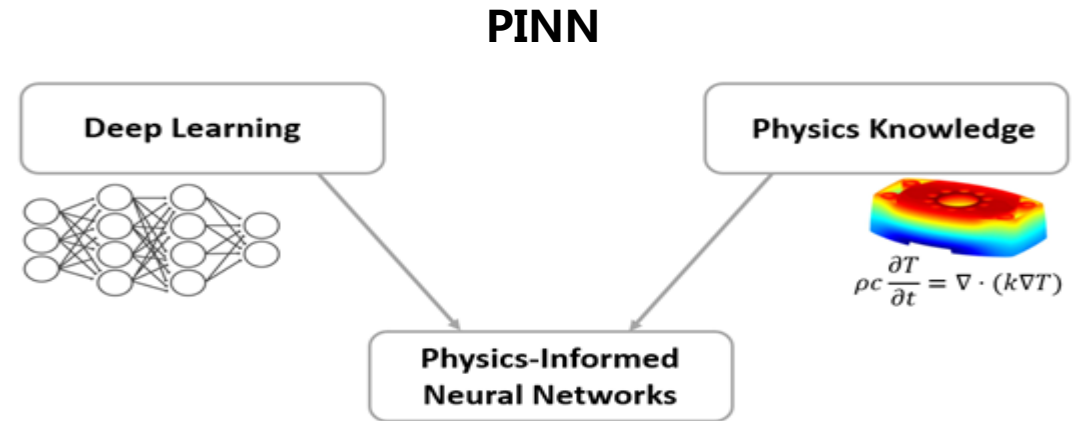
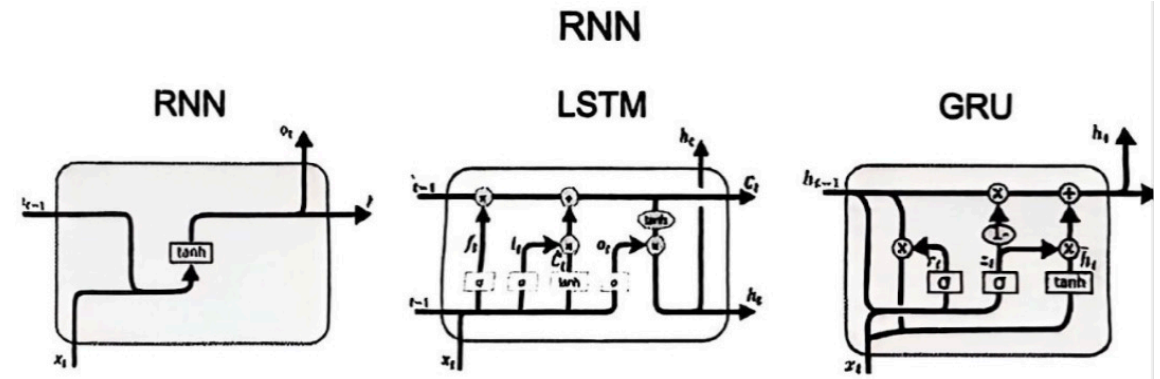
计算机视觉、自然语言处理、音频处理

算法

数据

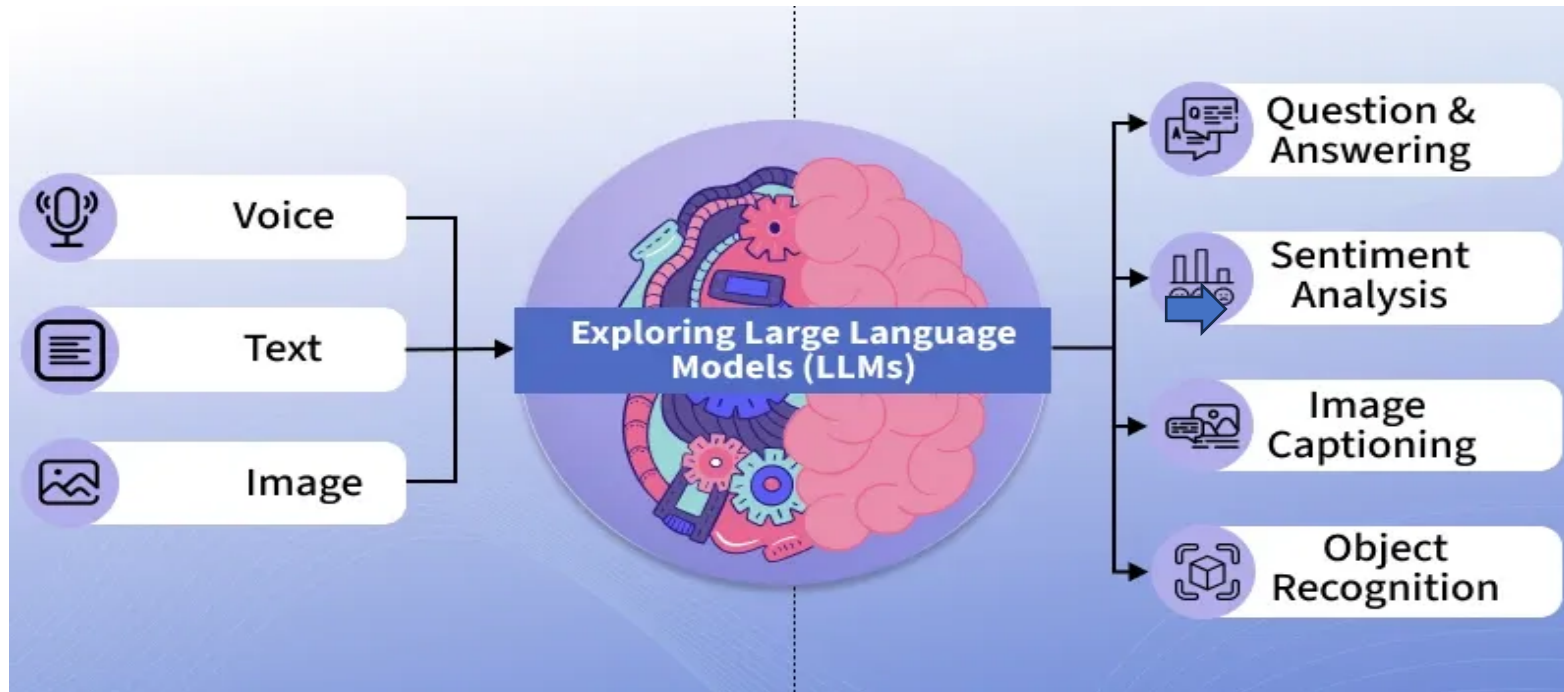


算力



1. A brief history of AI

大语言模型（LLM）：通过海量文本数据训练，具备理解、生成和推理文本的能力。



Data ➡ **Tokens (vector)** ➡ **Training (weighted tokens)** ➡ **Model**

1. 局部相关性



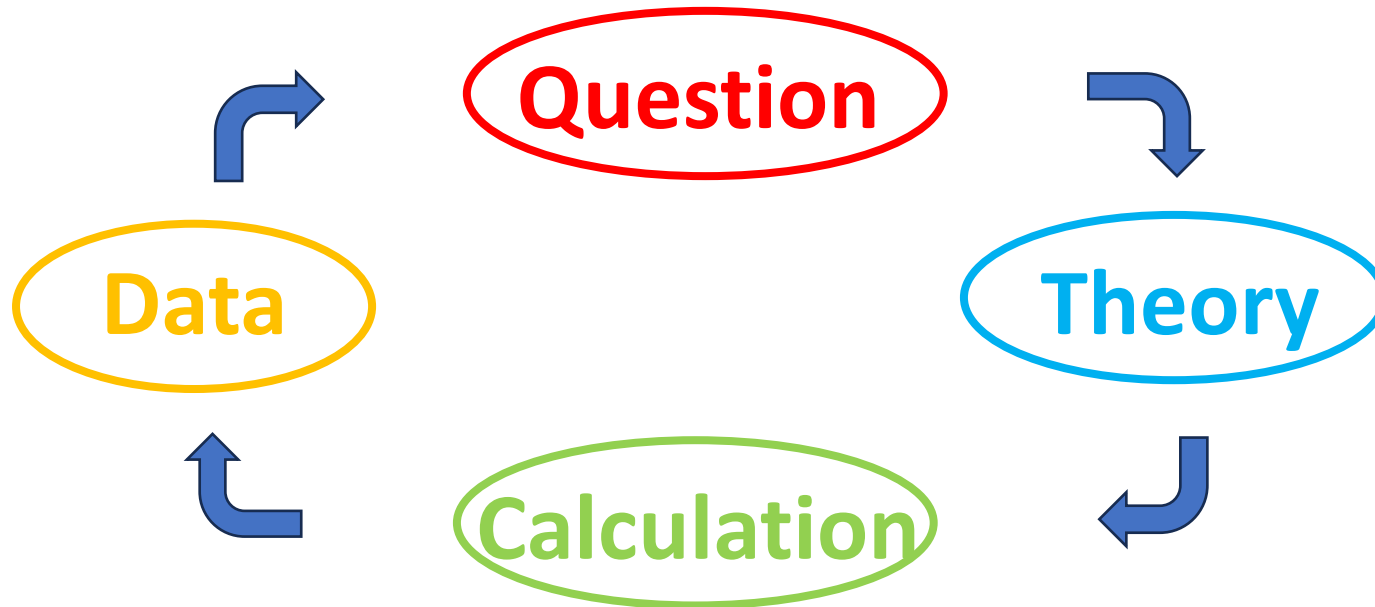
2. RNN & LSTM



3. Transformer

**DeepSeek、Claude、
ChatGPT、Copilot、Llama、
Gemini ...**

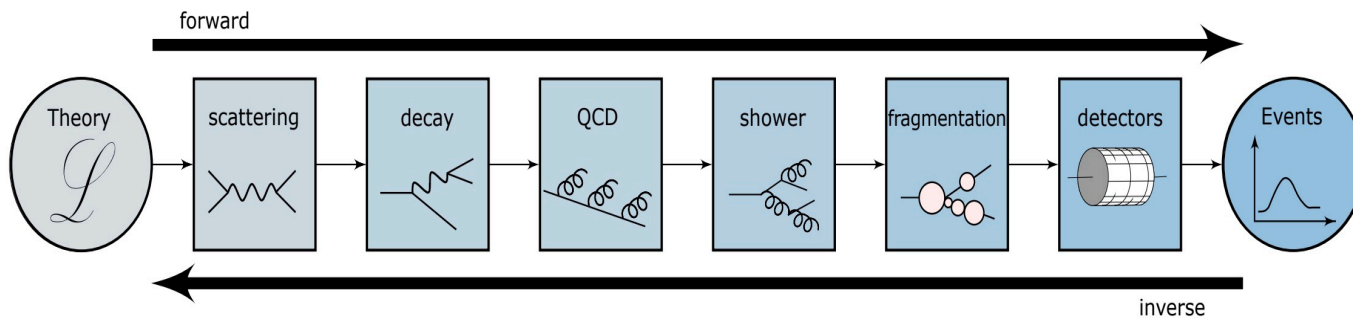
2. AI for HEP: from data to theory



数据密集型科学
(PB, EB)

Classical papers

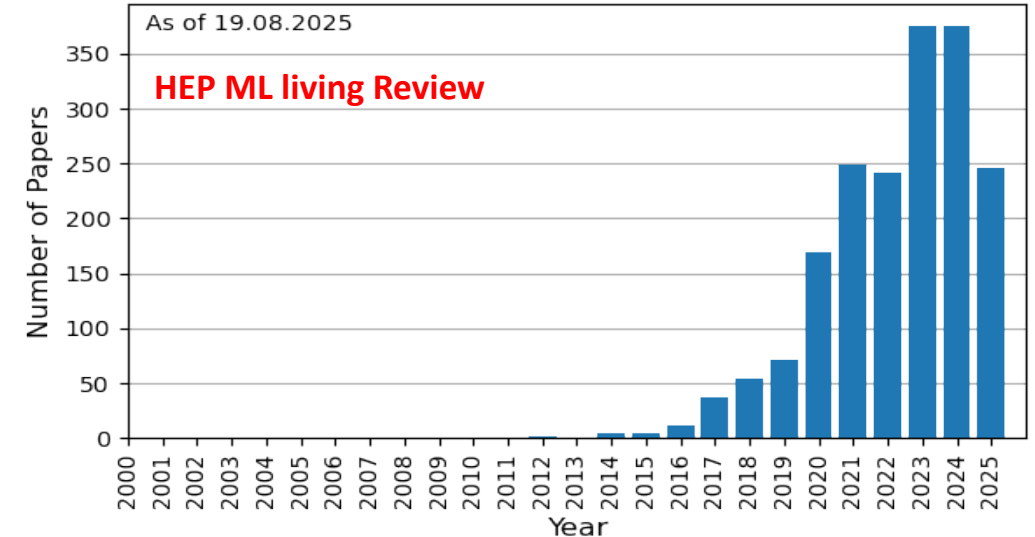
- Finding Gluon Jets With a Neural Trigger (1990)
- Neural Networks and Cellular Automata in Experimental High-energy Physics (1988)



2211.01421

2025/11/2

Number of HEP-ML Papers by Year



2. AI for HEP: from data to theory

1. Boosted Decision Tree (BDT)

Boosting: combining many "trees" into a strong classifier.

一棵树的输出

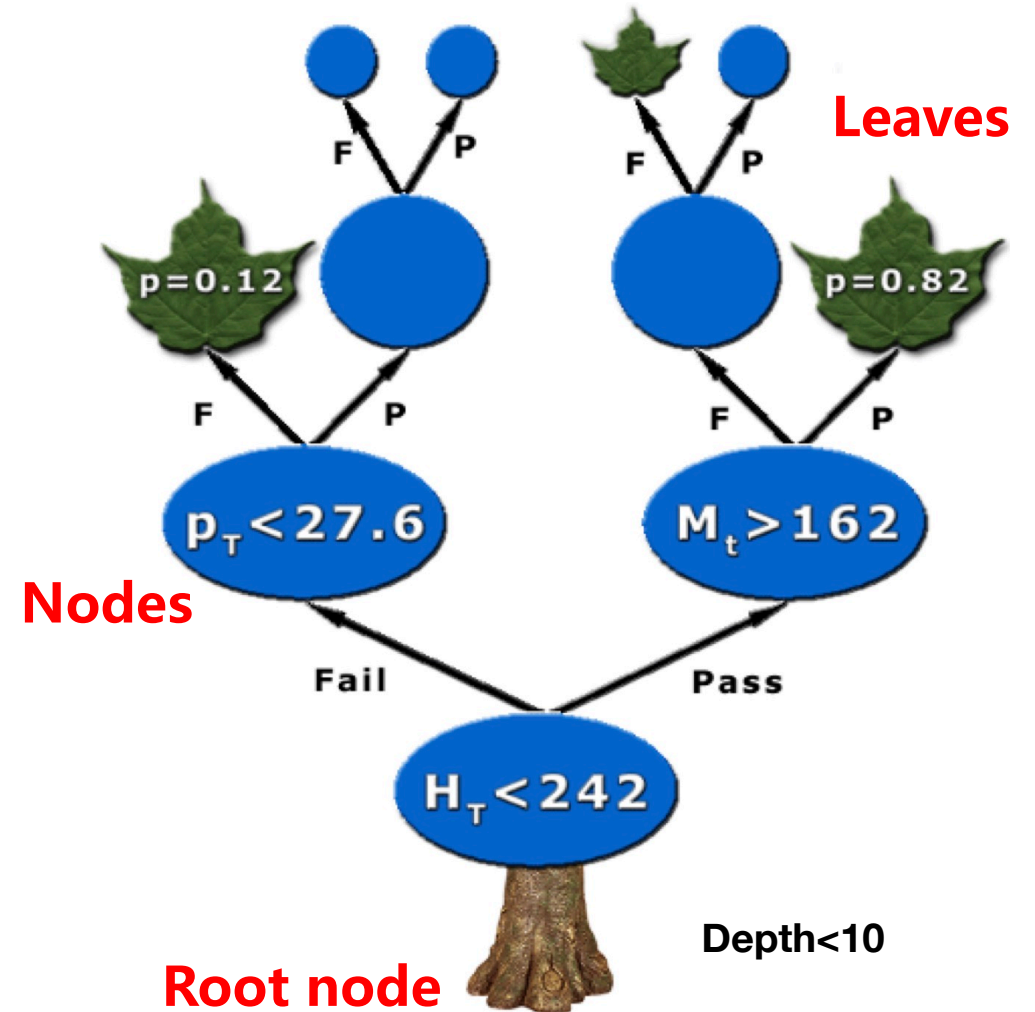
$$\hat{y}(x) = \sum_t w_t h_t(x)$$

权重，随迭代变化

The goal is to minimize an objective function

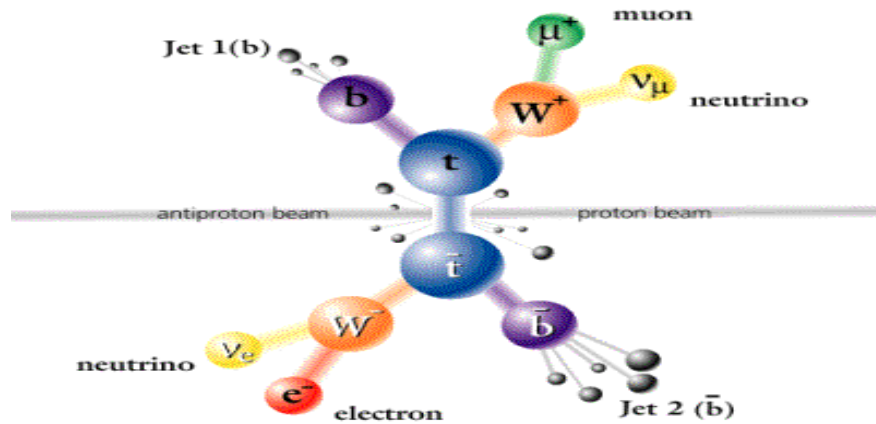
$$O(x) = \sum_i l(\hat{y}_i, y_i) + \sum_t \Omega(f_t)$$

误差函数 正规化函数



2. AI for HEP: from data to theory

Improved Top Quark Pair Production



PHYSICAL REVIEW D

VOLUME 47, NUMBER 5

1 MARCH 1993

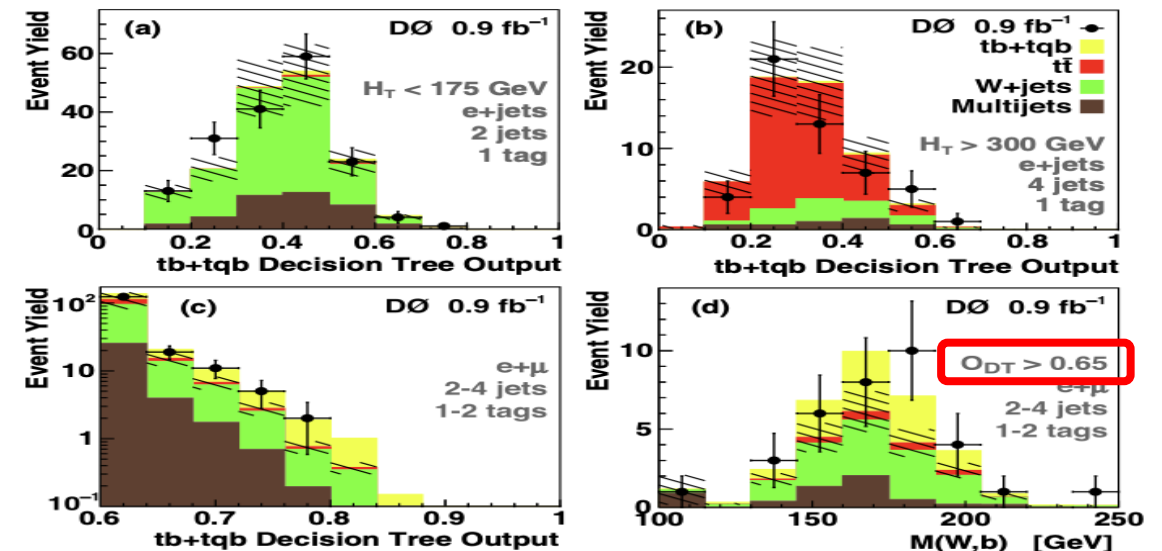
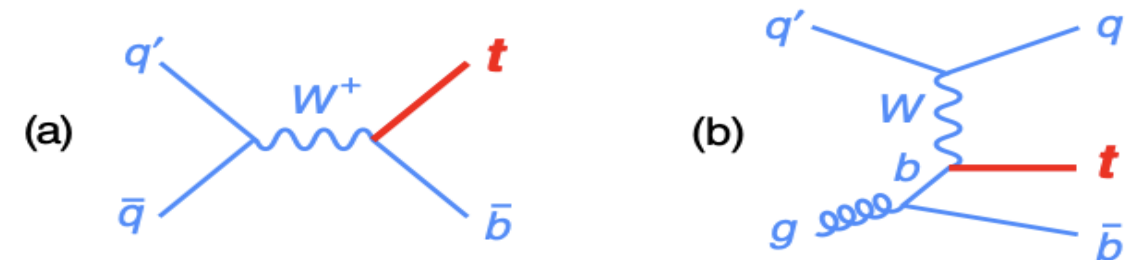
Comparison of the use of binary decision trees and neural networks in top-quark detection

David Bowser-Chao* and Debra L. Dzialo†
Center for Particle Physics, University of Texas at Austin, Austin, Texas 78712
(Received 11 September 1992)

The use of neural networks for signal versus background discrimination in high-energy physics experiment has been investigated and has compared favorably with the efficiency of traditional kinematic cuts. Recent work in top-quark identification produced a neural network that, for a given top-quark mass, yielded a higher signal-to-background ratio in Monte Carlo simulation than a corresponding set of conventional cuts. In this article we discuss another pattern-recognition algorithm, the binary decision tree. We apply a binary decision tree to top-quark identification at the Fermilab Tevatron and find it to be comparable in performance to the neural network. Furthermore, reservations about the "black box" nature of neural network discriminators do not apply to binary decision trees; a binary decision tree may be reduced to a set of kinematic cuts subject to conventional error analysis.

Evidence for Single Top Quark Production

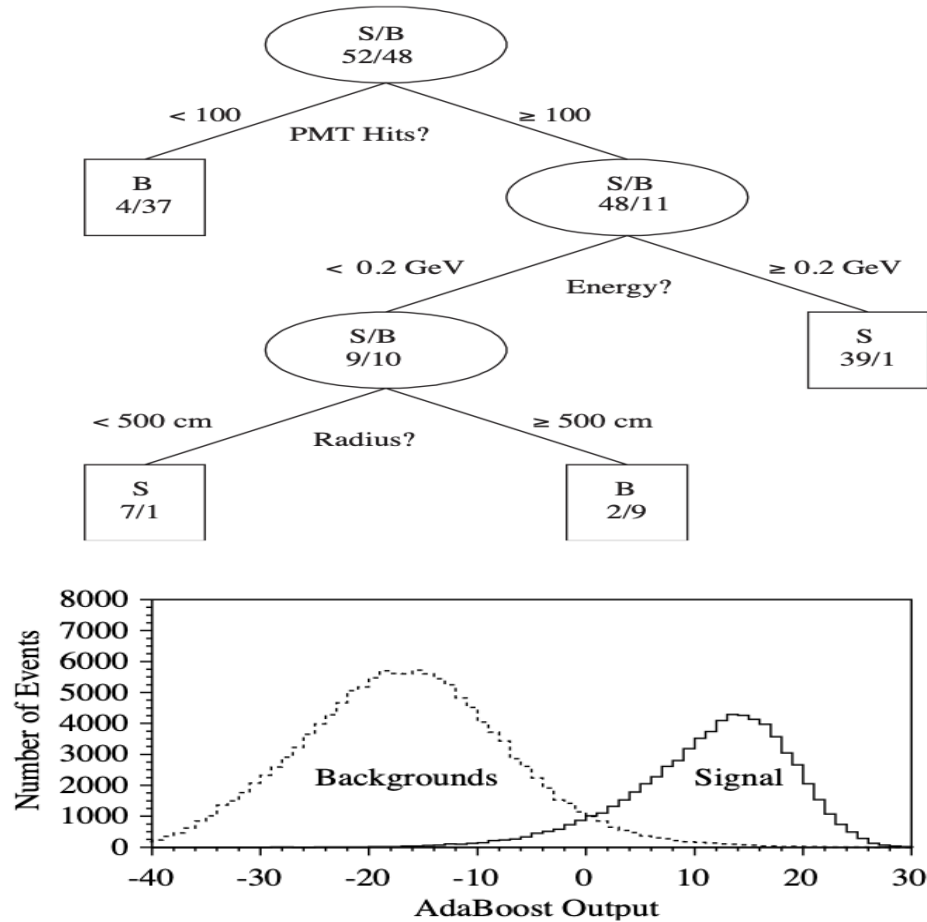
D0 collaboration
PRL 98 (2007) 181802



2. AI for HEP: from data to theory

MiniBooNE particle identification

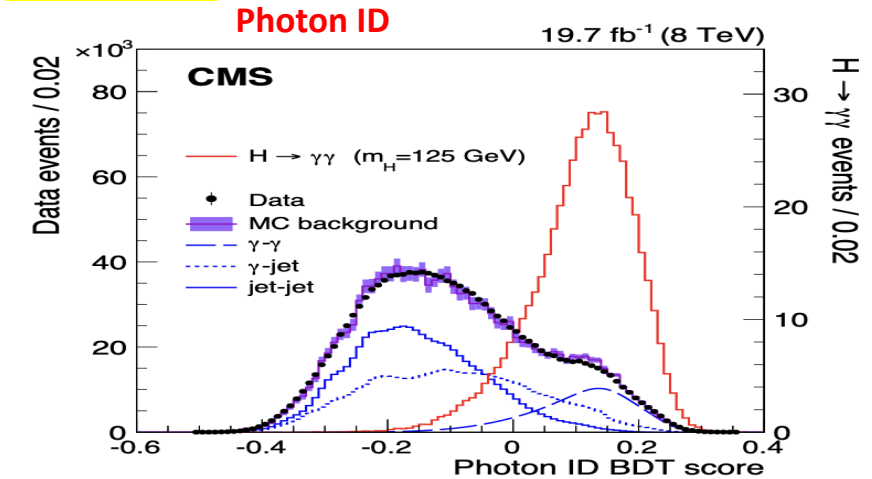
Roe, Yang, , Zhu, Liu, Stancu, NIMA 543 (2005) 577; Yang, Roe, and Zhu, NIMA 555 (2005) 370



Measurements of Higgs boson

$$h \rightarrow \gamma\gamma$$

arXiv:1407.0558 (5.1+19.7 fb⁻¹)



$$h \rightarrow \mu^+\mu^-$$

CMS: a **23%** increase in sensitivity ($\simeq$ **50% more data**).

arXiv:1807.06325 (35.fb⁻¹)

ATLAS: a **50%** increase in sensitivity

ATLAS-CONF-2019-028 (Full run-2)

$$h \rightarrow c\bar{c}, h\bar{t}t\text{.....}$$

2. AI for HEP: from data to theory

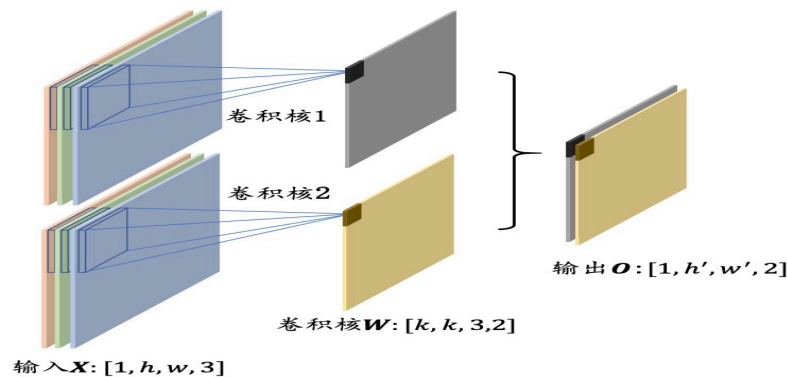
2. Conventional Neural Networks (CNN)

2203.14569

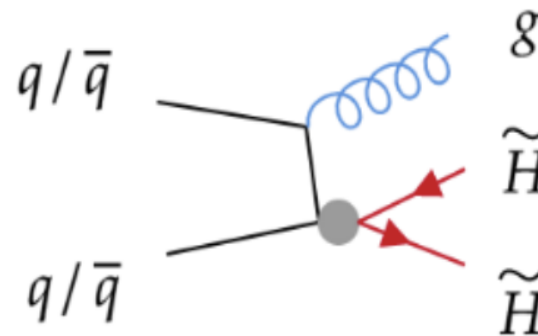


局部相关性

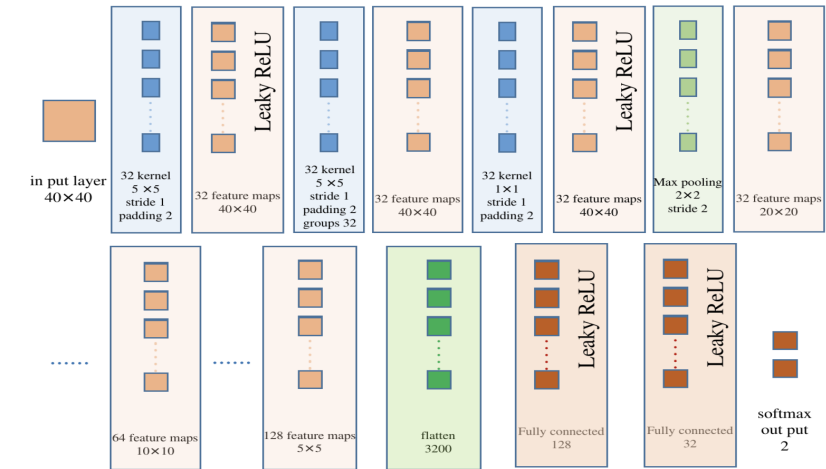
权值共享



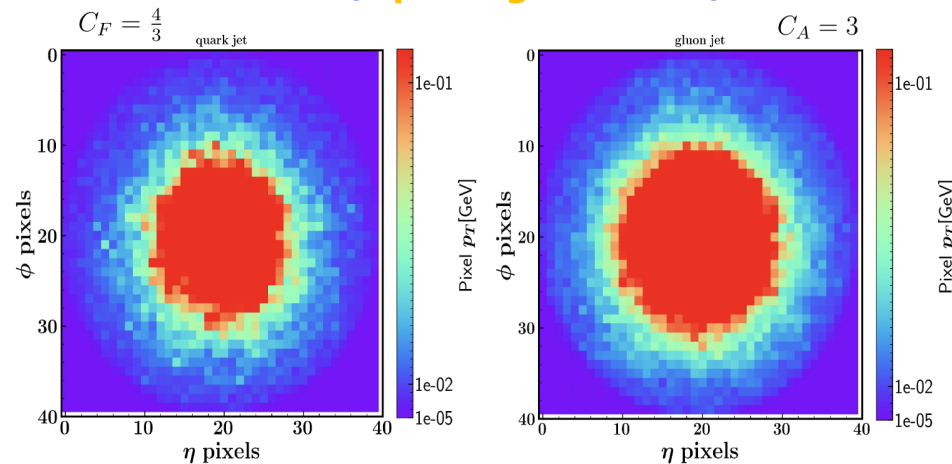
单喷注产生



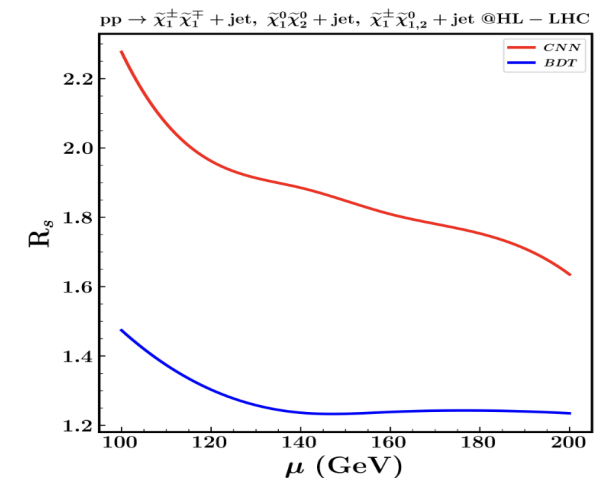
40x40灰度图像，单通道，3层卷积



喷注图像 (Splitting function)



预期结果

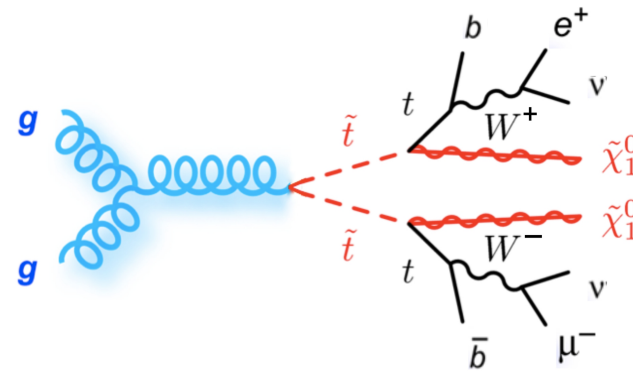


2. AI for HEP: from data to theory

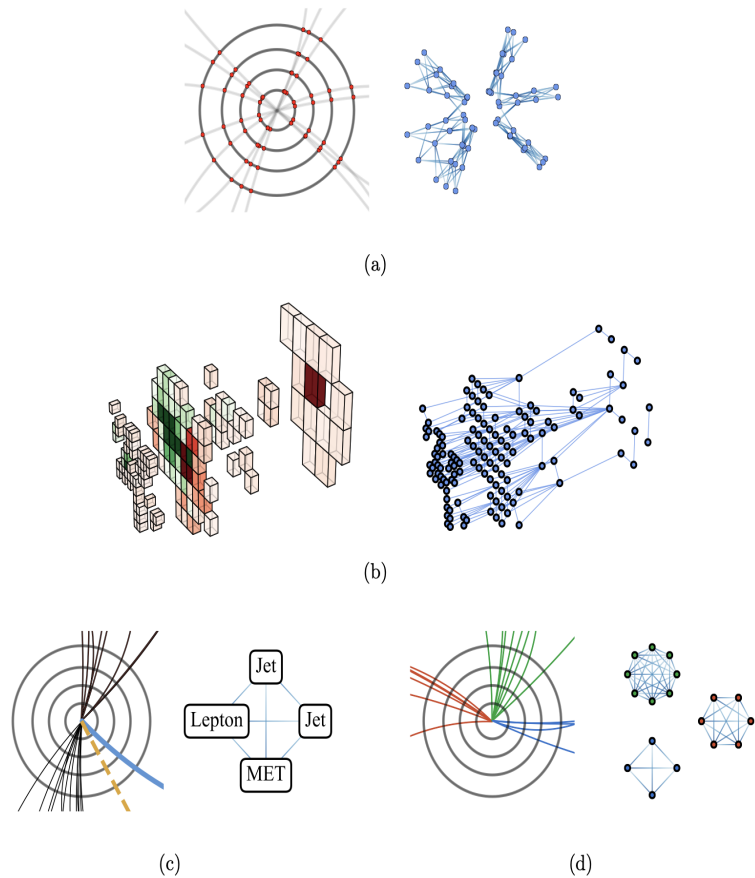
3. Graph Neural Networks (GNN)

1807.09088

LHC上top-squark对产生



事件图的构建



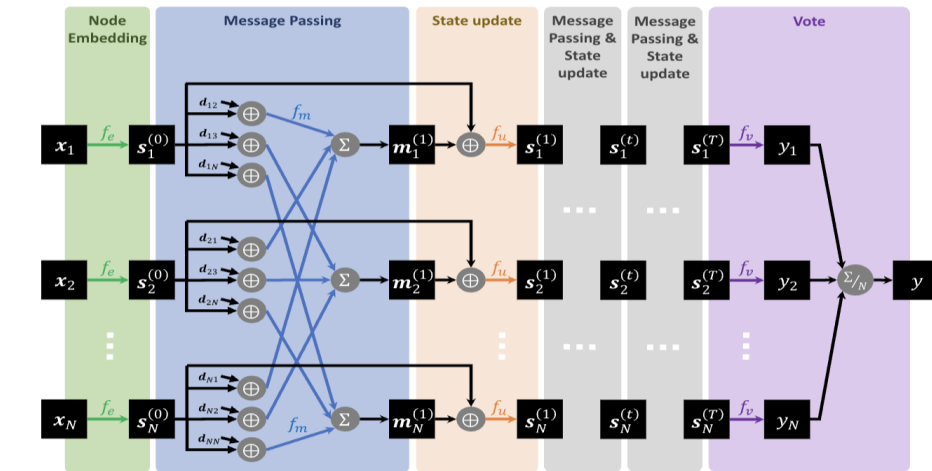
$$\mathbf{x}$$

	Photon	lepton charge	b-jet or light jet	MET	p_T (TeV)	E (TeV)	m (TeV)
$\mathbf{x}_1 =$	0	-1	0	0	0.0229	0.0289	0.0000
$\mathbf{x}_2 =$	0	0	-1	0	0.2637	0.3304	0.0373
$\mathbf{x}_3 =$	0	0	-1	0	0.1003	0.1888	0.0091
$\mathbf{x}_4 =$	0	0	1	0	0.0980	0.1146	0.0133
$\mathbf{x}_5 =$	0	0	1	0	0.0689	0.0773	0.0062
$\mathbf{x}_6 =$	0	0	0	1	0.2107	0.2107	0.0000

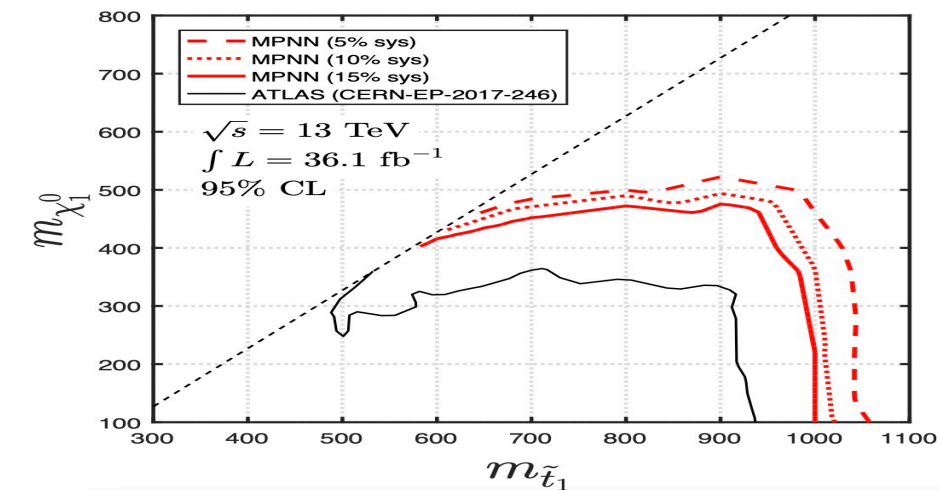
$$\mathbf{d}$$

	1	2	3	4	5	6
1	0	1.3971	2.5649	1.2801	3.2752	3.0312
2	1.3971	0	1.9019	1.6688	3.0871	3.1717
3	2.5649	1.9019	0	3.4440	1.5805	1.7831
4	1.2801	1.6688	3.4440	0	2.2175	2.1387
5	3.2752	3.0871	1.5805	2.2175	0	0.4912
6	3.0312	3.1717	1.7831	2.1387	0.4912	0

MPNN架构，适用于变长输入，不依赖粒子数量和顺序



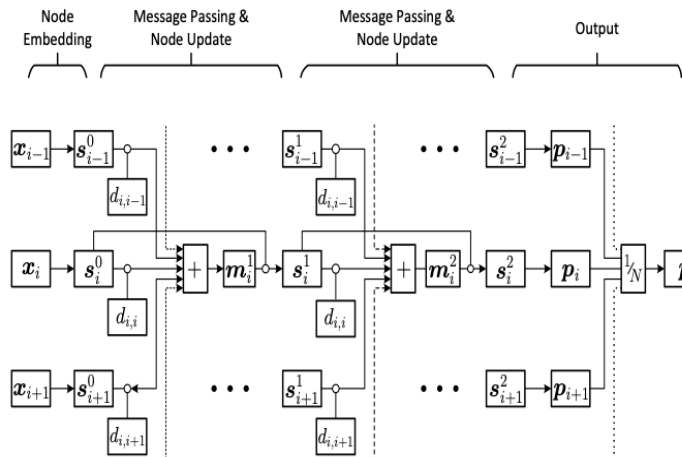
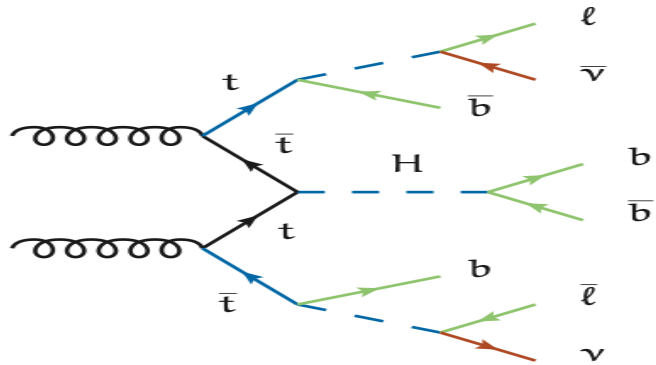
预期结果



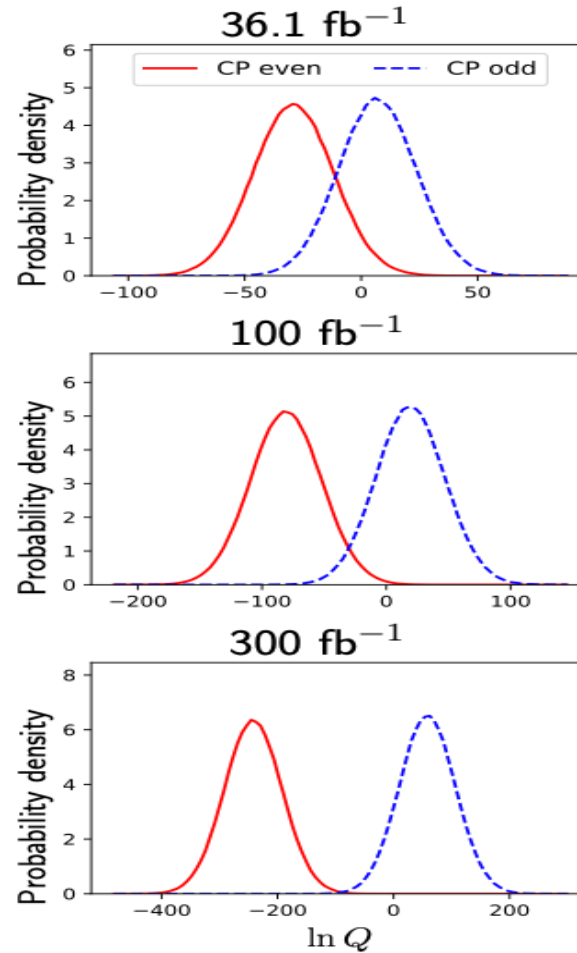
2. AI for HEP: from data to theory

3. Graph Neural Networks (GNN)

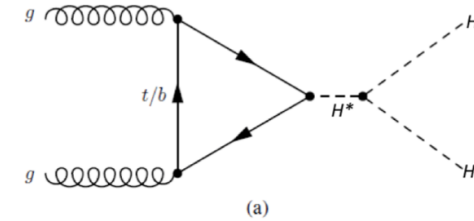
1901.05627



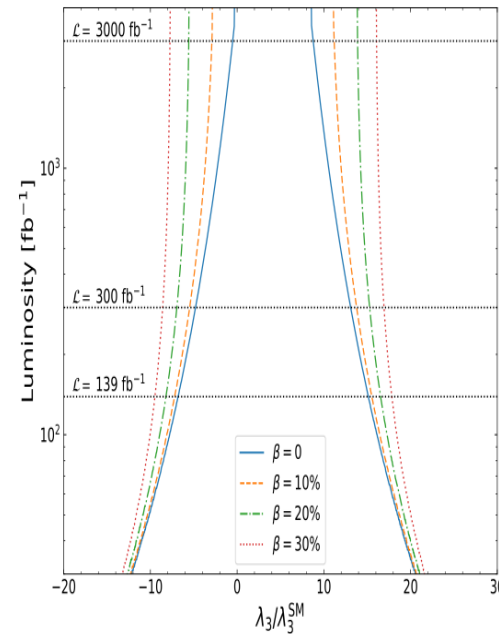
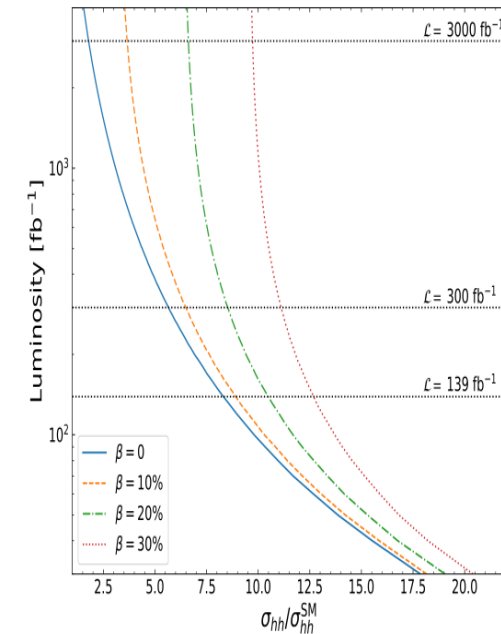
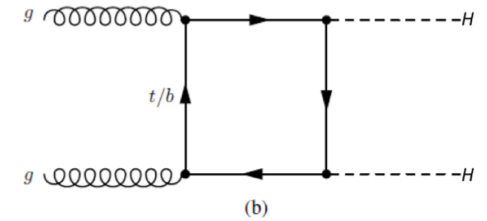
2025/11/2



第五届量子场论及其应用研讨会



2005.11086



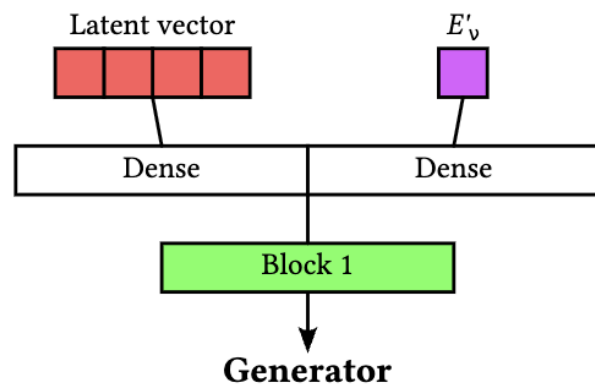
14

2. AI for HEP: from data to theory

4. Generative adversarial neural networks (GAN)

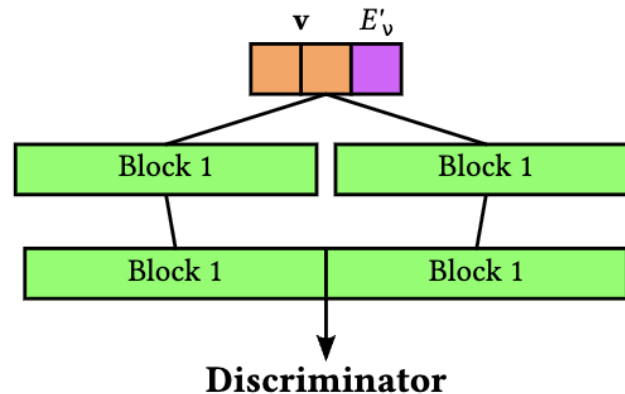
2502.20244

GAN代替MC产生器NuWro，生成中微子-碳核散射事件的 μ 子运动学参数（能量 E_μ 、散射角 θ ）



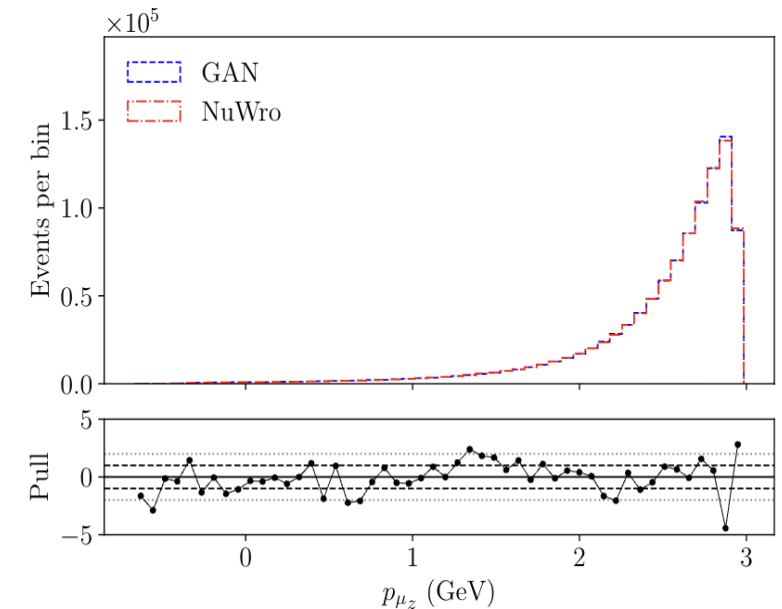
输入 E'_ν + latent 向量

输出符合物理分布的 E'_μ 、 θ



输入样本（真实 / NuWro 或生成 / GAN）+ E'_ν

输出样本真实概率



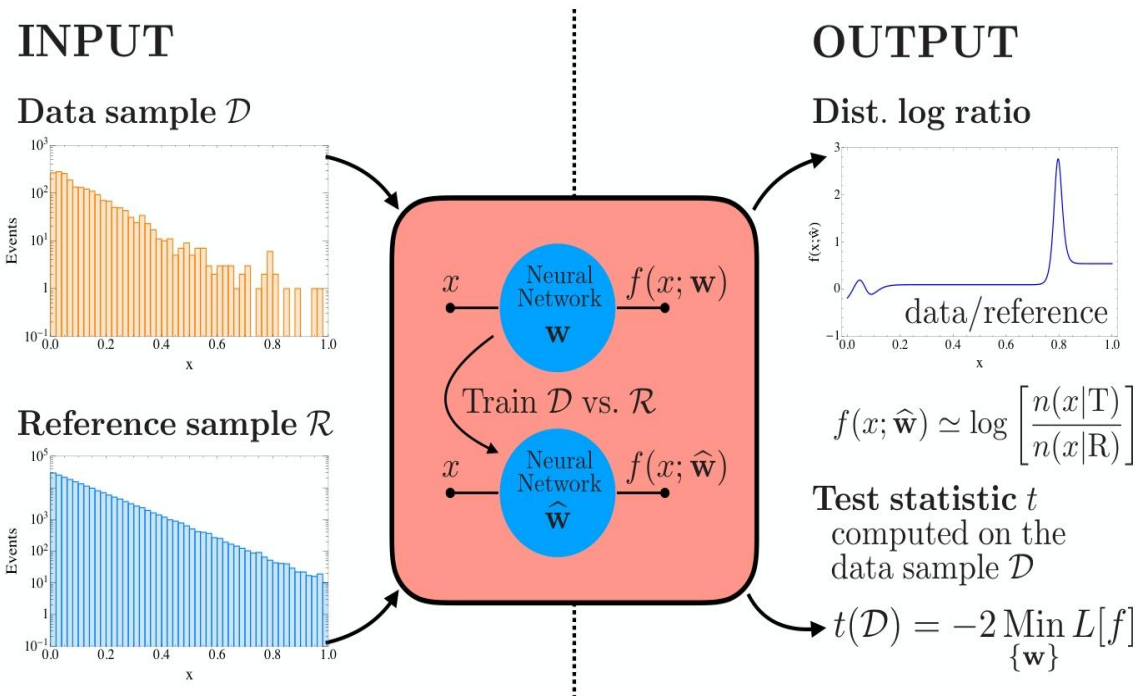
生成工具	生成 100 万事件时间（单核 CPU）	生成 100 万事件时间（RTX 4080 GPU）
NuWro（全模拟）	~12 分钟	-
GAN	~28 秒	~3 秒

2. AI for HEP: from data to theory

5. Anomaly Detection (AD)

1806.02350

模型无关的检测数据与参考模型 (SM) 的偏离



似然比检验，新物理模型为网络输出

数据: $D=\{x_i\}$ 参考模型分布: $n(x|R)$ 数据分布: $n(x|T)$

替代模型分布: $n(x|w)$: $n(x|w) = n(x|R) e^{f(x;w)}$

$$\log \left[\frac{L(w)}{L(R)} \right] = \sum_{x \in D} \log \left[\frac{n(x|w)}{n(x|R)} \right] - [N(w) - N(R)]$$

核心挑战：

- 集中在低概率相空间（事件占比少但效应可能大）
- 分散在大区域（效应微弱），且无法预先确定

2. AI for HEP: from data to theory

6. Physics Information Neuron Network (PINN)

视费曼积分为一阶偏微分方程组 (PDEs) 的解 2312.02067

$$\frac{\partial}{\partial v_i} \vec{F}(\vec{v}; \epsilon) = A_{v_i}(\vec{v}; \epsilon) \cdot \vec{F}(\vec{v}; \epsilon), \forall i = 1, \dots, n_v$$

独立运动学变量

主积分构成的向量

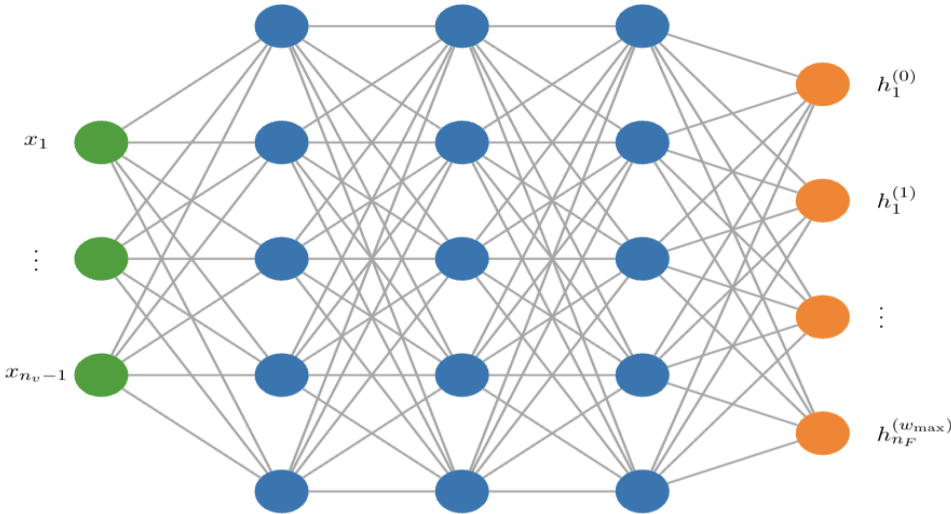
维数正则化参数

连接矩阵

输入：运动学变量

输出：所有主积分在 ϵ 展开的各个阶次的系数

激活函数：GELU，保证二阶可导，便于自动微分



用神经网络近似解，损失函数含微分方程约束+边界值约束

$$L_{DE} = \sum \left[\partial_{x_l} h_j^{(w)} - \sum_k A_{x_l}^{(k)} \cdot h_r^{(w-k)} \right]^2 \quad L_b = \sum \left[h_j^{(w)}(\vec{x}^{(i)}) - g_j^{(w)}(\vec{x}^{(i)}) \right]^2$$

实部、虚部：两个全连接前馈神经网络独立训练

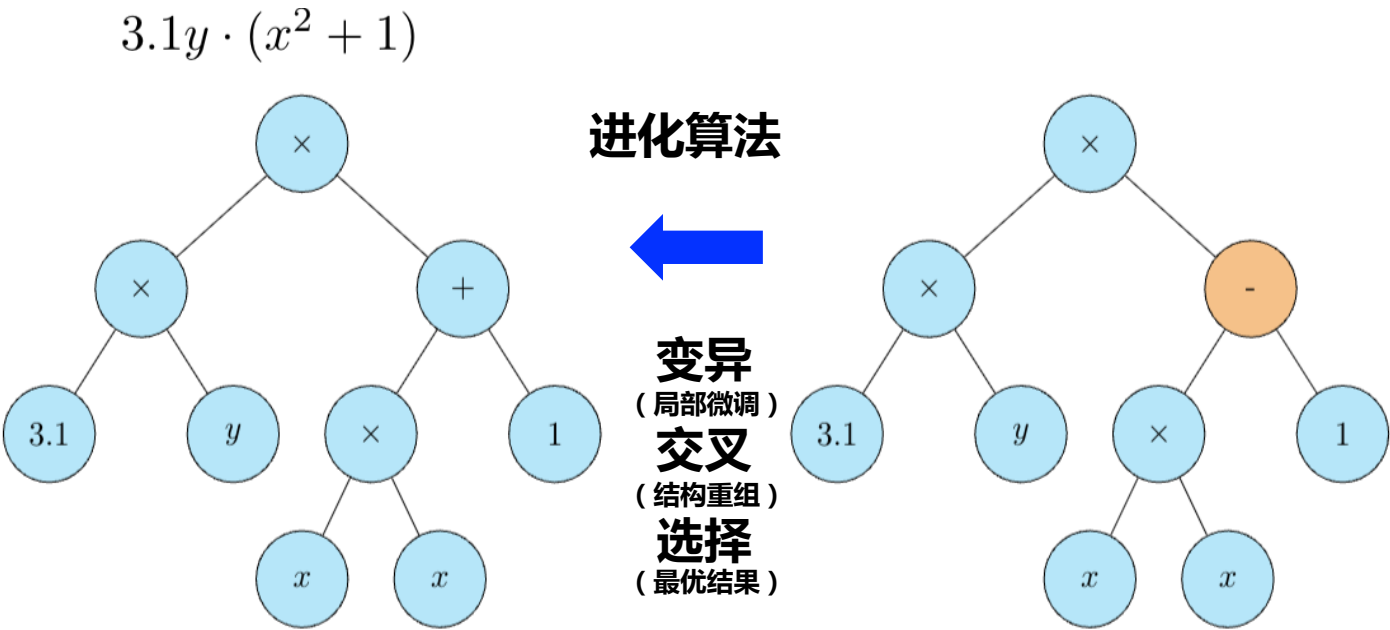
实例类别	对比方法	核心精度 (MMRD)	关键时间指标	数据需求
1 圈无质量盒积分	传统解析法 (handyG)	$\sim 10^{-8}$	推导：数周；单点点评估： $\sim 10^{-3}$ 秒 (CPU)	无
	本文神经网络 (PIDL)	2.2×10^{-5}	训练：16 分钟 (GPU)；单点点评估： $\sim 10^{-6}$ 秒 (GPU)	2 个边界点
2 圈单质量双盒积分	传统解析法 (文献 [108])	$\sim 10^{-8}$	推导：数月；单点点评估： $\sim 10^{-3}$ 秒 (CPU)	无
	现有机器学习 (文献 [55])	3%	训练：4 小时 (GPU)；单点点评估： $\sim 10^{-4}$ 秒 (GPU)	10^4 个标注点
2 圈重交叉盒积分	本文神经网络 (PIDL)	1.1×10^{-2}	训练：53 分钟 (GPU)；单点点评估： $\sim 10^{-6}$ 秒 (GPU)	6 个边界点
	传统数值法 (AMFlow)	$\sim 10^{-4}$	单点点评估： ~ 2 秒 (CPU)	无 (逐点计算)
2 圈顶双盒积分	本文神经网络 (PIDL)	7.3×10^{-3}	训练：75 分钟 (GPU)；单点点评估： $\sim 10^{-6}$ 秒 (GPU)	10 个边界点
	传统数值法 (DiffExp)	$\sim 2\%$	网格生成：4 小时 (CPU)；单点点评估： $\sim 10^{-3}$ 秒 (CPU)	无 (网格生成)
	本文神经网络 (PIDL)	3.9×10^{-3}	训练：32 分钟 (GPU)；单点点评估： $\sim 10^{-6}$ 秒 (GPU)	20 个边界点

2. AI for HEP: from data to theory

7. Symbolic Aggression

2412.07839

Symbolic Tree
(数学表达式用树结构表示)



均方误差 (Accuracy) + 复杂度 (complexity)

Rediscovery from QED

$$e^+e^- \rightarrow \gamma^* \rightarrow \mu^+\mu^-$$

$$\frac{d\sigma}{d\cos\theta} = \frac{\pi\alpha^2}{2s}(1 + \cos^2\theta)$$

使用MadGraph生成模拟数据

SR恢复表达式

Bins	Accuracy	Best	Score
10	$c_\theta(c_\theta + 0.00798)(0.00111 \cdot c_\theta^3 + 0.03459) + 0.03503$	$c_\theta^2(0.00111 \cdot c_\theta + 0.03459) + 0.03503$	$0.03459 \cdot c_\theta^2 + 0.03503$
20	$c_\theta(c_\theta + 0.01825)(-0.00155 \cdot c_\theta(c_\theta - 0.05138) + 0.03579) + 0.03485$	$c_\theta(0.03447 \cdot c_\theta + 0.00064) + 0.03498$	$0.03447 \cdot c_\theta^2 + 0.03498$
200	$c_\theta^2(-0.64647 \cdot c_\theta(0.00119 \cdot c_\theta - 0.00151) + 0.03495) + 0.03495$	$0.03447 \cdot c_\theta^2 + 0.03495$	$0.03447 \cdot c_\theta^2 + 0.03495$

2. AI for HEP: from data to theory

8. Large Language Model (LLM) 2501.09729

双向自回归 Transformer (3.57亿参数)

符号 token 化方案

拉格朗日量

自然语言结构

符号类型	处理方式	示例
场	按 “FIELD + 自旋 + 对称性 + helicity+ dagger+ ID” 编码，ID 用于跟踪指标收缩	$\phi(\bar{3}, 2, -1/3)$ → [FIELD, SPIN, 0, SU3, -, 3, SU2, 2, U1, -, 1, /, 3, ID3]
导数	按 “DERIVATIVE + 对称性 + ID” 编码，协变导数隐含规范场	$D_{(SU3, SU2, U1)\mu}$ → [DERIVATIVE, SU3, SU2, U1, ID1]
对易子	用 COMMUTATOR_A/B 标记位置	$[X, Y]$ → [COMMUTATOR_A, X, COMMUTATOR_B, Y]
收缩信息	单独标记 “CONTRACTION + 对称群 + ID”，明确指标收缩规则	洛伦兹收缩 → [CONTRACTION, LORENTZ, ID6, ID4]

场

单词

相互作用



句子

对称性

语法

对称性约束：群表示组合、量子数守恒

2501.09729

共轭操作学习: “国王 - 男人 ~ 女王 - 女人” 的向量偏移

拉格朗日量:

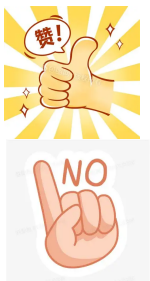
“规则驱动的复杂语言表达式”

分布内性能:

$S > 0.92$

分布外泛化能力:

missing term
anomaly free

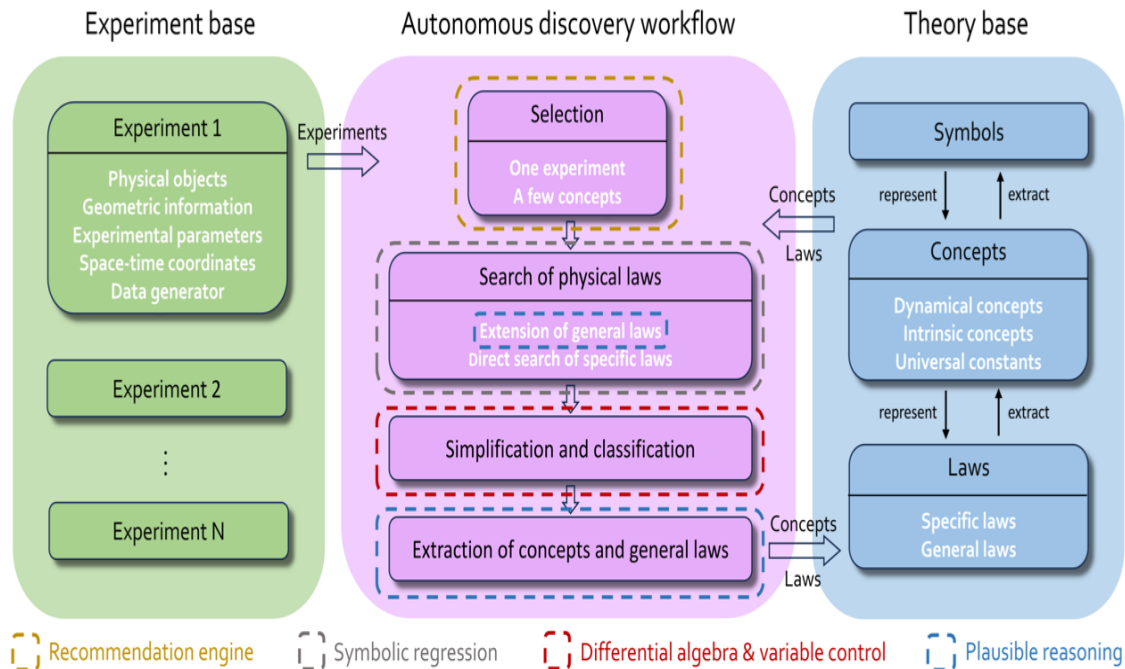


2. AI for HEP: from data to theory

9. SR/LLM for HEP in China

Y.-Q. Ma + students 2504.01538

AI-Newton 知识库+DSL+无先验



数学、物理如漆似胶？

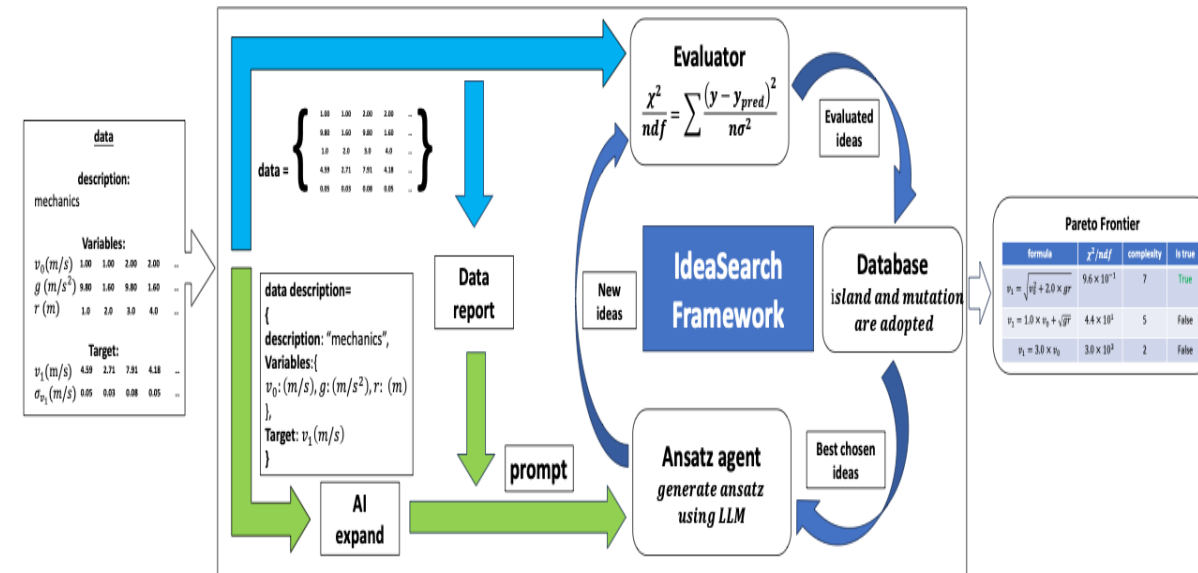
Cao, Liu, Luo, Zhu + students 2510.08317



将搜索转向语义空间：先解释，后表达

LLM作为语义变异和交叉操作符

IdeaSearch Collaboration



3. Prospects and Challenges

现有AI模型主要**依赖数据中的统计相关性**，
而**非基于物理定律的因果理解**。

物理感知：从感知数据中提取物理属性

物理推理：用符号、数学解决物理问题

虚拟建模：构建能预测环境动态的模型

具身交互：在真实物理环境中进行交互



第十四届 TeV 物理工作组学术研讨会 (14th Workshop on TeV Physics)

2019.04.20 南京



2018.04



2025/11/2



第五届量子场论及其应用研讨会



2024.05

22

恭祝李老师

松柏长青

桃李永芳



网络模块	层数	核心参数	功能
输入层	1	1 通道、40×40 像素	接收预处理后的喷注图像（经平移、旋转、归一化处理，消除位置 / 角度依赖）
卷积层（Conv）	3	第 1-2 层：32 个 5×5 卷积核，步长 = 1，零填充（Padding=2）；第 3 层：32 个 1×1 卷积核	提取图像局部特征（如像素分布密度、边缘结构），零填充确保卷积后尺寸保持 40×40；1×1 卷积压缩通道维度以减少参数
激活函数	3	Leaky ReLU（斜率 0.01）	引入非线性，解决 Sigmoid/tanh 的梯度饱和问题和 ReLU 的“死亡神经元”问题，提升梯度收敛效率
池化层（MaxPool）	2	2×2 池化核，步长 = 2	对卷积特征图下采样，保留关键特征的同时降低维度 (40×40→20×20→5×5)，减少计算量并抑制过拟合
展平层（Flatten）	1	将 5×5×128 特征图展平为 1 维向量（长度 = 5×5×128=3200）	连接卷积层与全连接层，将空间特征转化为向量特征
全连接层（FC）	2	第 1 层：128 个神经元；第 2 层：32 个神经元	对展平后的特征向量进行高阶组合，学习信号与背景的全局差异
输出层	1	Softmax 激活函数，2 个输出节点（信号 / 背景）	将全连接层输出转化为概率分布（满足 $\sum p=1$ ），公式为 $p_i = \frac{e^{S_i}}{\sum_j e^{S_j}}$ (S_i 为全连接层输出， $j=2$ 为类别数)

1. 事件图的构建

- **节点 (Node)**: 每个节点代表碰撞事件的末态粒子 (光子、轻子、b 喷注、轻喷注、缺失横向能量 MET), 节点特征向量 x_i 为 7 维:
 - a. 光子标识 (1 = 光子, 0 = 非光子);
 - b. 轻子电荷 (轻子为实际电荷, 非轻子为 0);
 - c. 喷注类型 (1=b 喷注, -1 = 轻喷注, 0 = 非喷注);
 - d. MET 标识 (1=MET, 0 = 非 MET);

5-7. 粒子的横向动量 p_T 、能量 E 、质量 m 。

- **边 (Edge)**: 为无向加权边, 权重 $d_{ij} = \sqrt{\Delta y_{ij}^2 + \Delta \phi_{ij}^2}$ (Δy 为快度差, $\Delta \phi$ 为方位角差), 编码粒子间的几何距离; 且通过径向基函数 (RBF) 将 d_{ij} 映射到高维向量, 提升模型学习能力。

3. 训练与对比设置

- 损失函数: 二元交叉熵; 优化器: Adam (学习率 0.001); 训练框架: PyTorch (GPU 加速);
- 对比模型: DNN (固定输入长度, 需零填充)、MPNN6/DNN6 (仅用 6 个核心粒子: 1 个轻子、2 个 b 喷注、2 个领先轻喷注、MET, 模拟传统截断输入)。

2. MPNN 模型架构

MPNN 通过 “消息传递 - 状态更新” 迭代, 将整个事件图的信息编码到每个节点的状态向量中, 实现端到端分类, 流程如下:

1. **节点嵌入 (Node Embedding)**: 通过单层感知机 f_e 将 7 维节点特征 x_i 映射为 30 维初始状态向量 $s_i^{(0)}$;
2. **消息传递 (Message Passing)**: 第 t 次迭代中, 节点 i 收集所有其他节点 j 的消息 $m_{i \leftarrow j}^{(t)} = f_m^{(t)}(s_j^{(t-1)}, d_{ij})$, 并求和得到总消息 $m_i^{(t)} = \sum_{j \neq i} m_{i \leftarrow j}^{(t)}$;
3. **状态更新 (State Update)**: 节点 i 通过 $f_u^{(t)}(s_i^{(t-1)}, m_i^{(t)})$ 更新自身状态, 迭代 $T = 2$ 次后, 每个节点状态已包含全图的粒子特征与几何关联;
4. **投票与分类 (Voting & Prediction)**: 每个节点通过投票函数 f_v 输出信号相似度 y_i , 最终取所有节点的平均值 $y = \frac{1}{|\mathcal{V}|} \sum_i y_i$ 作为事件的 discrimination score (分数越高越可能为信号)。

研究里用了个叫 PySR 的软件，它像 “进化” 一样找公式 —— 先生成一堆简单公式当 “祖先”，然后让它们不断 “变异”（比如把 “+” 改成 “-”）、“结合”（把两个公式拼一起），再按 “算得准不准” “公式简不简单” 这两个标准筛选，最后留下最好的。

大语言模型（Large Language Model, LLM）：是一种基于深度学习的自然语言处理技术，通过海量文本数据训练，具备理解、生成和推理文本的能力。

架构基础：采用Transformer神经网络，依赖自注意力机制处理长文本序列，显著提升并行计算效率

训练流程：分为预训练（无监督学习）、微调（有监督学习）和RLHF（基于人类反馈的强化学习）三个阶段

参数规模：通常包含数十亿至数千亿参数（如GPT-3的1750亿参数）