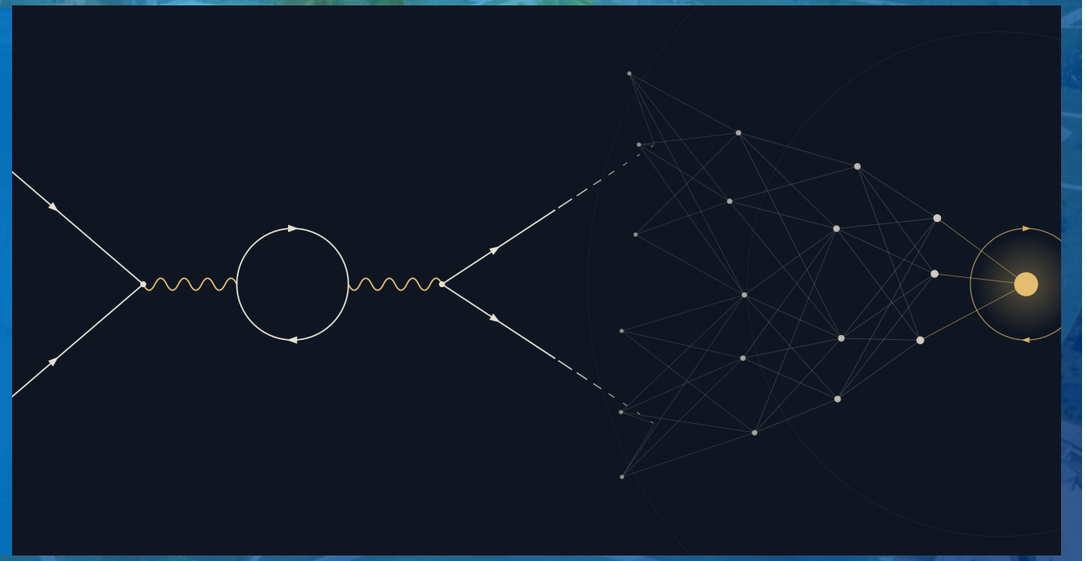
Theory Foundation Models as Inference Engines at the Discovery Frontier



unifying results across the model ladder

T. J. Hobbs — Argonne HEP Theory Group

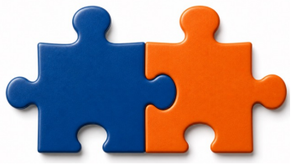
CLARIPHY, UW Madison — 15 Sept 2026



Supratim Das Bakshi, Sam Foreman

Special thanks:
Brandon Kriesten

AI Revolution: theory building and discovery



Experimental reach requires deep AI-theory integration beyond simple application of ML tools

- ❑ AI-native theory representation(s)? Autonomous discovery loop? Theory implications of rapid AI advances? …

Bespoke models probe specific physics mechanisms; agentic systems, foundation models scale to discover emergent patterns.

Paradigm 1 · PRECISION PREDICTION

Paradigm 2 · ROBUST INFERENCE WITH UQ

1 · BESPOKE MODELS

VAIM — gluon PDF directly decoded from lattice QCD [2507.17810]

Evidential UQ — distinguishing BSM scenarios in neutrino DIS [2412.16286]

2 · AGENTIC SYSTEMS

ArgoLOOM — agents steer the HEP + nuclear toolchain [2510.02426]

Knowledge Extraction — theory agents guide EW analyses [2606.00224 (SLD)]

3 · FOUNDATION MODELS (FMs)

SMEFT demonstrator — reusable theory representations for colliders [2512.15862]

MORPH — PDF foundation model, developed on ALCF resources [ongoing]

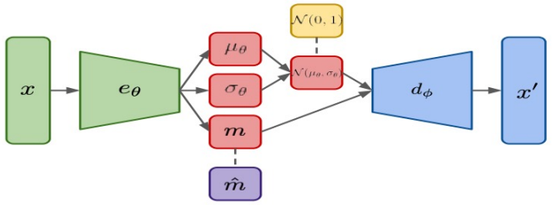
→ foundation models are an ideal setting for AI-theory induction

Decoding hadronic structure for LHC precision

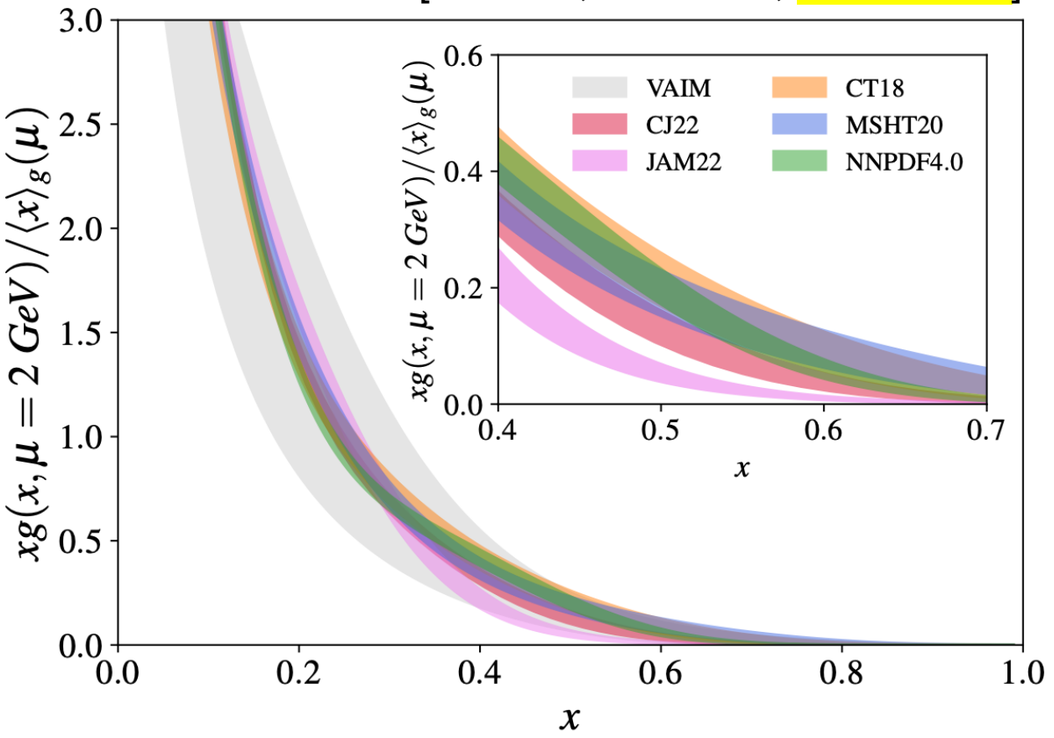
Bespoke models (eg, VAIM) may exploit problem-specific symmetries, properties

Example: PDF unfolding from lattice

- PDFs underlie every LHC cross section (*e.g.*, Higgs production) but never directly measured: an ill-posed inverse problem --- parametrization bias limits discovery
- First-ever ML decoding [right] of the gluon PDF from lattice data (VAIM); VAIM structured around flavor symmetries of underlying PDFs; consistent with phenomenology.
- Systematically improvable, but **generalizable to FM scale?**

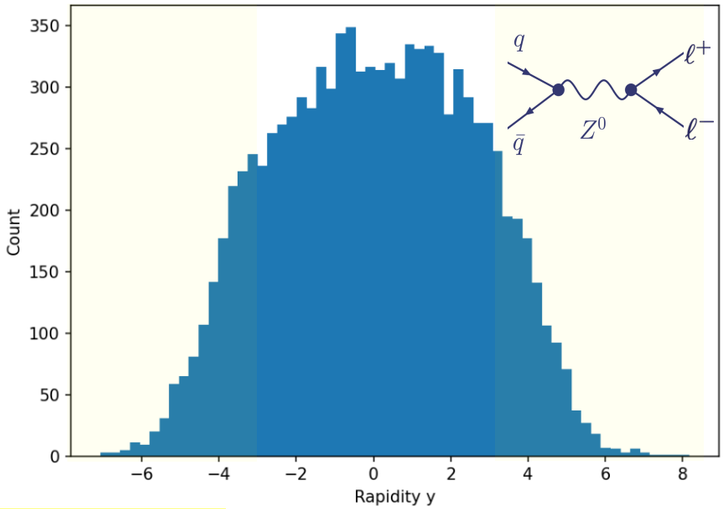


[Kriesten, TJH et al., 2507.17810]

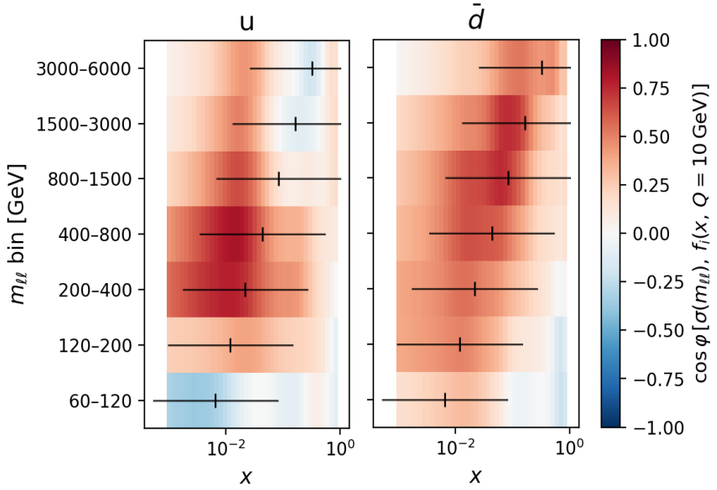


Agentic AI from Quarks to Cosmos: ArgoLOOM

AI-orchestrated package for Cosmo-Nucl-HEP predictions



[2510.02426]



AI for cross-frontier phenomenology

- One solution: while no single bespoke model does all physics (analyses chain many theory tools) – **agents can make more intelligent and adaptive**
- Example: ArgoLOOM, first agentic platform to compute collider [left-top], cosmology observables; and nuclear (QCD) metrics [left-bottom] to explore correlations for BSM searches
- Cf. ongoing effort: agentic determination of fundamental physics information [AmSC Knowledge Extraction]

(see talk, N. Ramachandra)

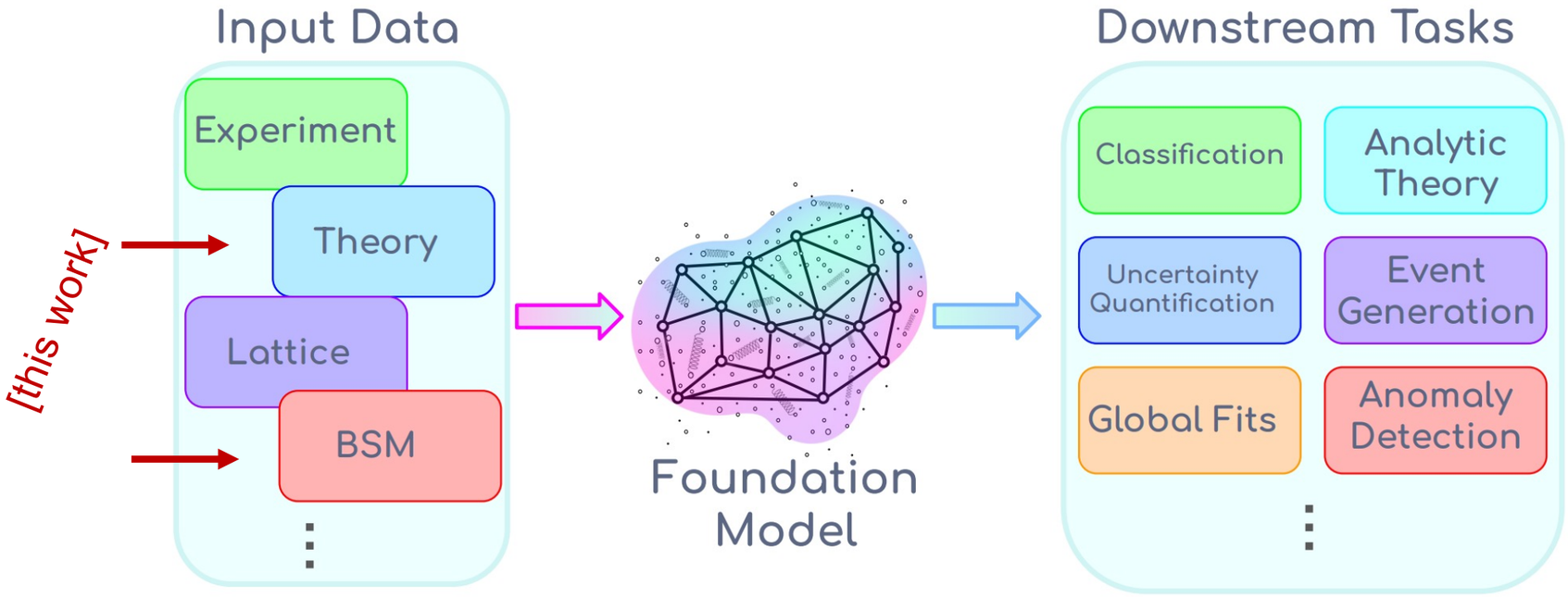
- While agents may correlate across frontiers, **might FMs learn subtle correlations via large model scaling (?)**

foundation (or world) models for HEP theory

❑ *idea*: pretrained foundation model **(FM) learns essential theory** properties (*e.g.*, of SMEFT, PDFs, or BSM sensitivities) through simulated data; build by SMEs to pioneer AI-theory representations

→ collider observables form 'language' through which model interrogates underlying theory

$$O_{\ell q}^{(1)}, O_{\ell q}^{(3)}$$
$$O_{eu}, O_{ed}$$
$$O_{\ell u}, O_{\ell d}, O_{qe}$$



2512.15862: Das Bakshi, TJH, Kriesten

➢ pretraining *not the goal*: embedding supports downstream tasks unseen in training (product is **reusable representation**)

Conceptual similarity to global analysis: solve inverse problem, optimize large model for downstream phenomenological calculations

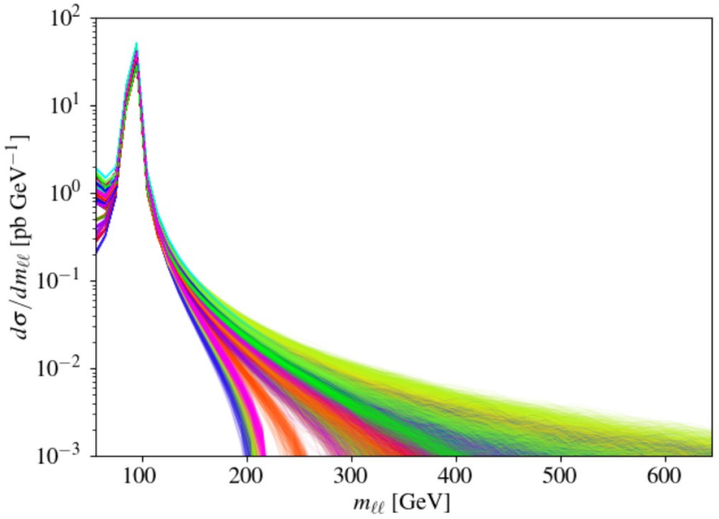
demonstration

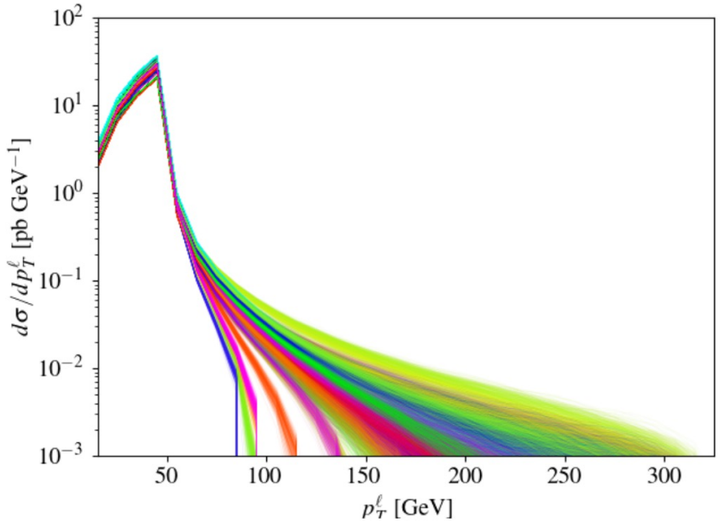
teaching FMs (SMEFT) theory: pre-training

2512.15862

- training instances are $O(100)$ **SMEFT *universes*: samplings of dim-6 Warsaw** coefficients at $O(\Lambda^{-2})$, $\Lambda = 1$ TeV (log-uniform, controlled sparsity), fed to *contrastive encoder model*
  - → base simulations: MadGraph5 + SMEFTsim, concatenated NC Drell-Yan $m_{\ell\ell}$ and p_T spectra; CT18LO PDFs

- $O(10^4)$ MC replicas sampled with physics-motivated uncertainties: effective-luminosity Poisson statistics; Gaussian-kernel bin correlations
  - → **supervised contrastive loss**: pull same-universe replicas together, push distinct universes apart

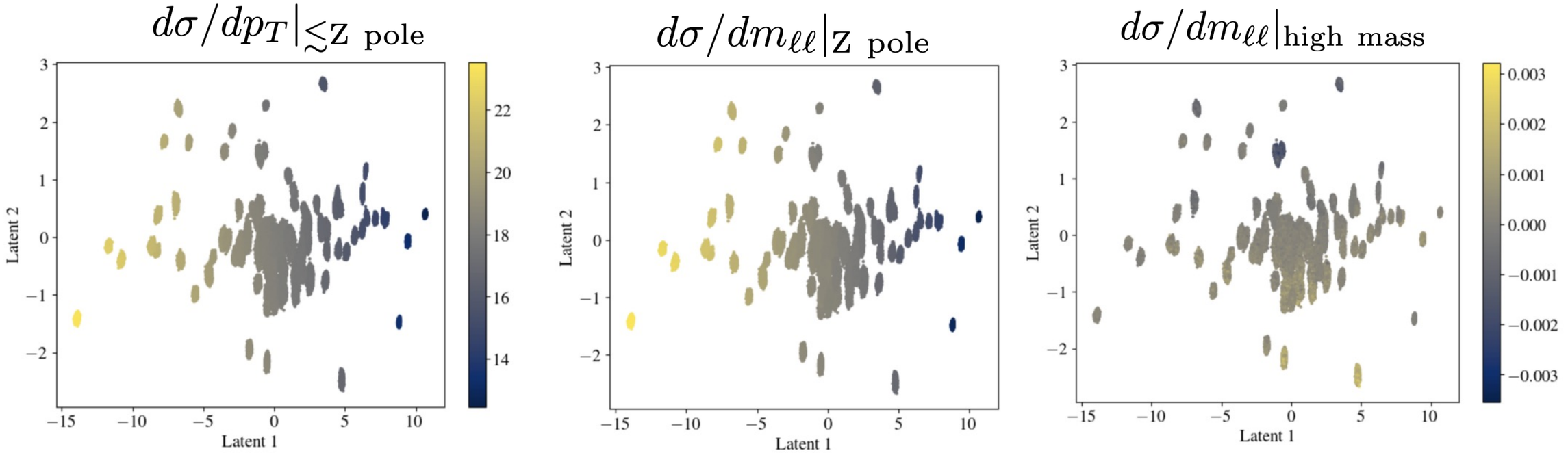




$$\mathcal{L}_\theta(x, x', m) = y\, D_\theta(x, x')^2 + (1-y)\, \max\left[0,\, m - D_\theta(x, x')\right]^2$$

interpretability: interrogating the learned latent representation

- **latent space organizes physics without explicit direction**: kinematic bins (Z-pole, high-mass tail[s]) vary contiguously and distinctly across the embedding

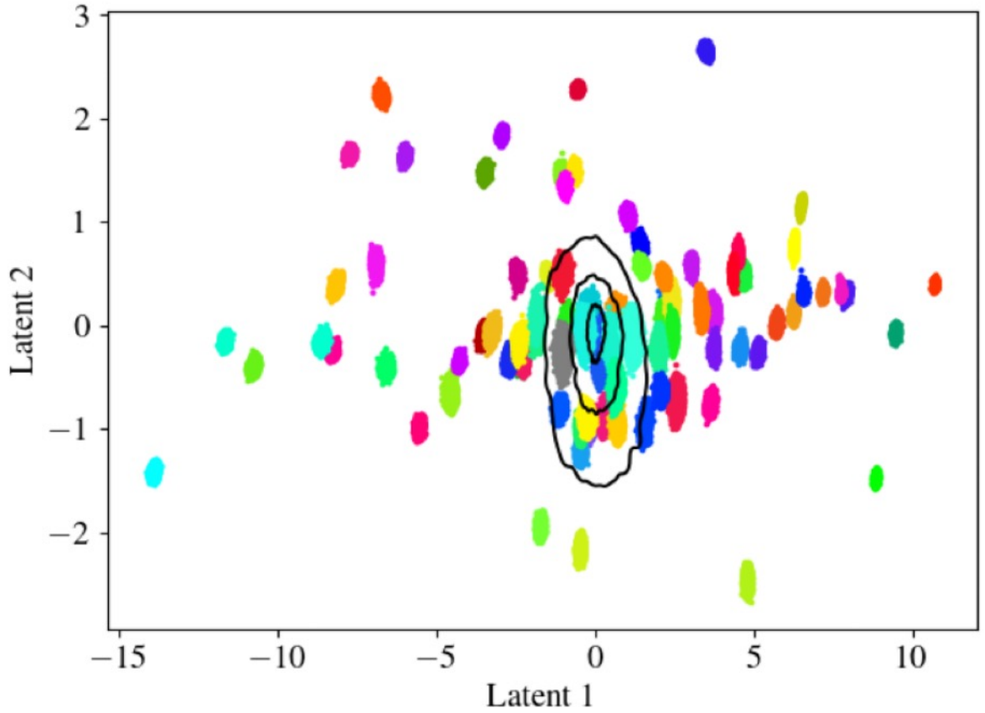


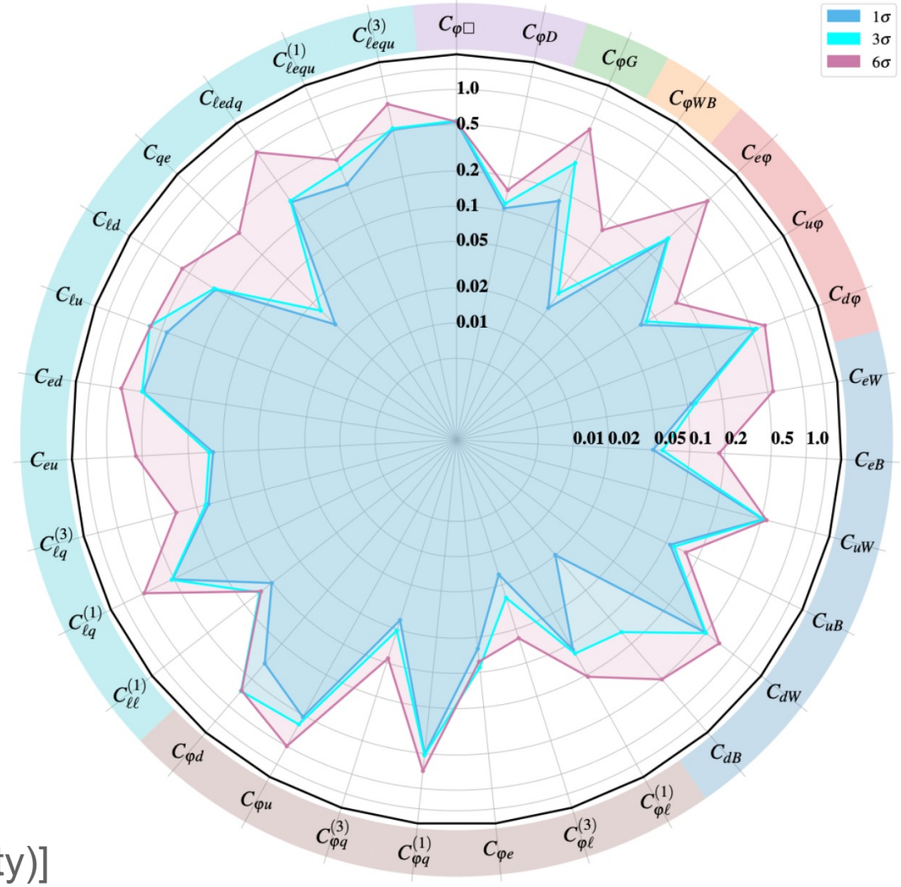
→ model learns correlated properties; for instance, between $m_{\ell\ell}$ and p_T peaks, organized along same latent axis (not a training objective)

query pretrained model for rapid BSM inferencing

❑ FM acts as *fast analyzer*: *e.g.*, identifies sensitivity to specific Wilson coefficients (above, NC Drell-Yan with dim-6 sampling)

2512.15862





[retrieval head connects SM-consistency contours (1/3/6σ) to allowed families of Wilson coefficients (global-fit complementarity)]

➢ pattern for inverse problems across HEP theory: **amortize the forward model once, invert on demand**

foundation model: downstream tasks with uncertainty

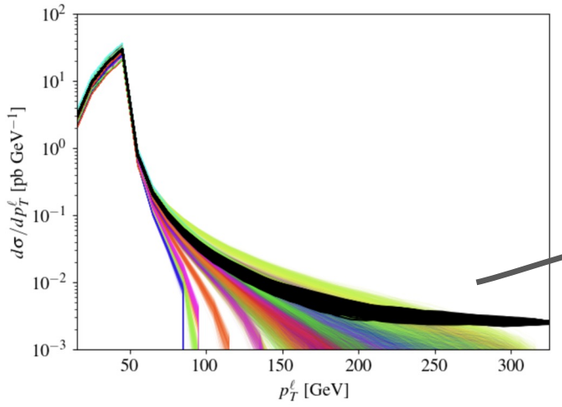
- ❑ universal latent representation of SMEFT theory space obtained through contrastive learning; this space can be repurposed for downstream tasks like anomaly detection or model classification

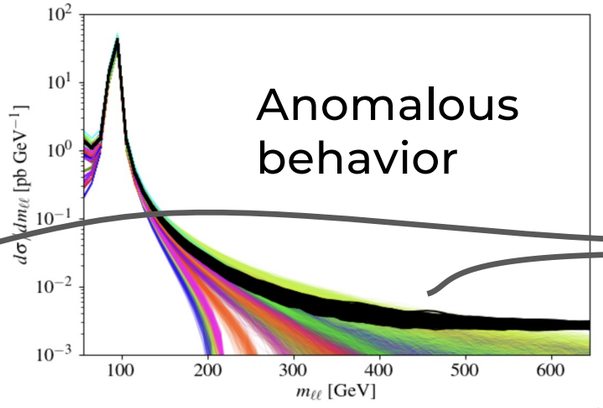
  → embedding encodes predictive uncertainty, allowing the model to possess a semantic notion of "I don't know"

[see DPNs for BSM, 2412.16286]

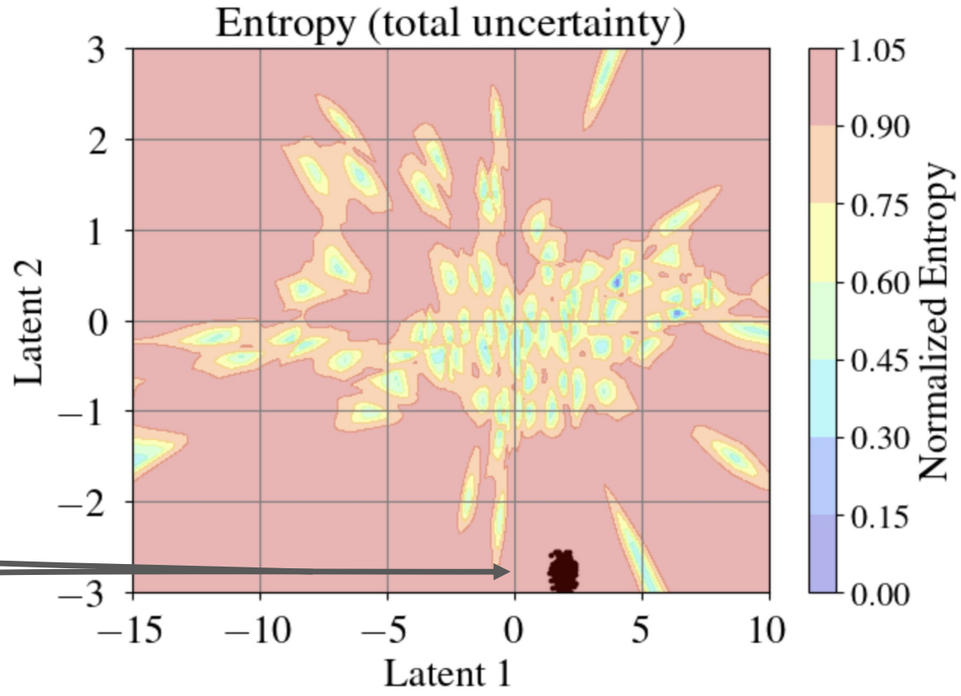
- ❑ classifier identifies known SMEFT scenarios; entropy highlights uncertain, OOD, and **anomalous signatures**

proof of principle established – learned embedding can encode SMEFT's internal structure and put it to work





Anomalous behavior

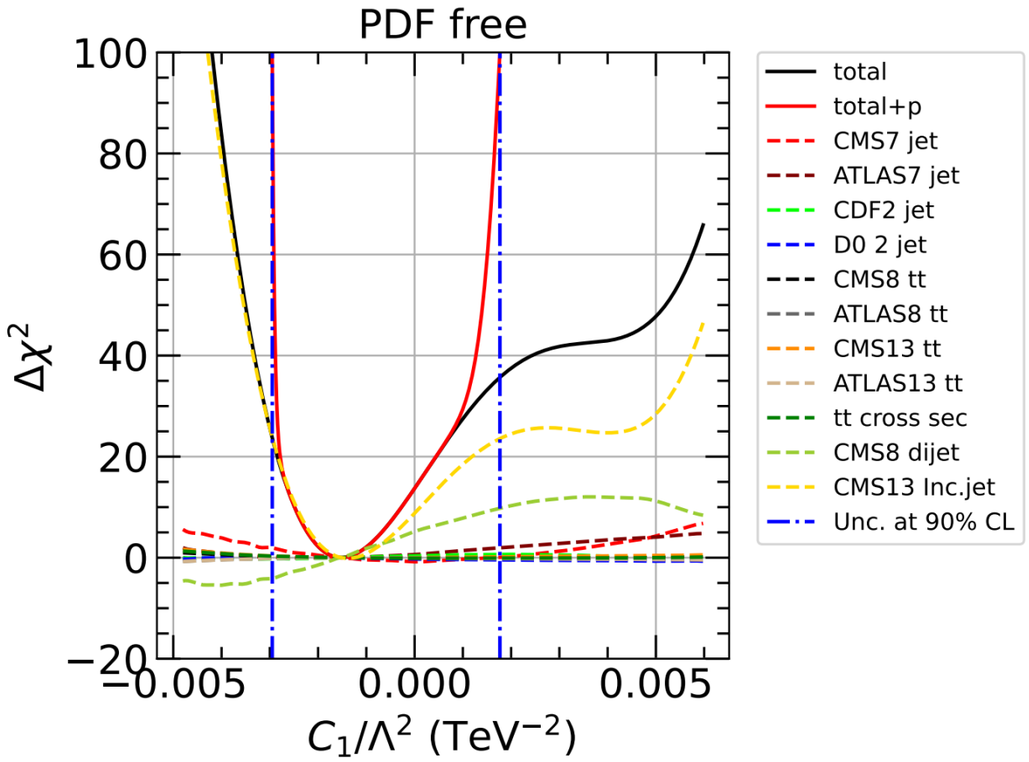


FM extensibility to related inverse problems (PDFs)

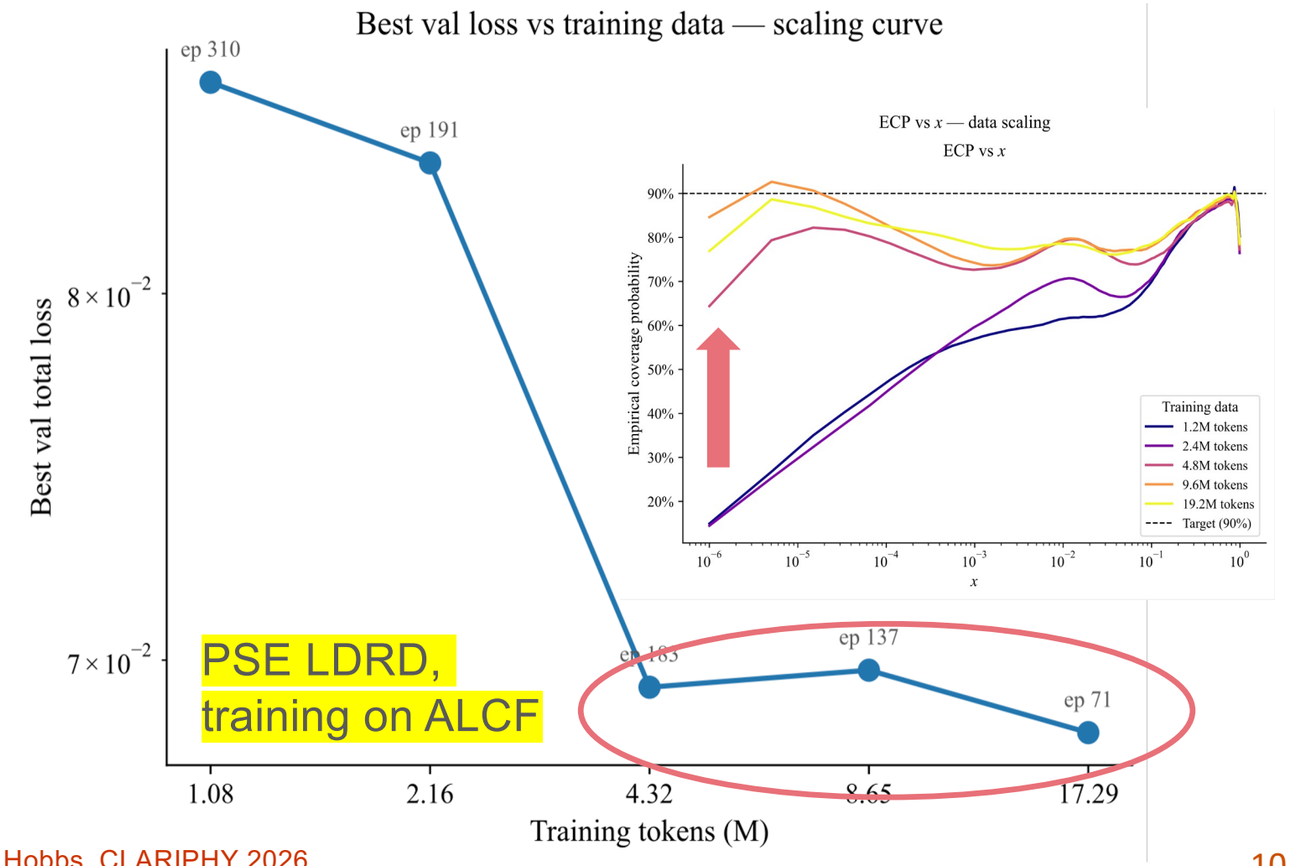
- theory FMs applicable to larger class of inverse problems --- *e.g.*, modeling of nucleon structure, which may be entangled with BSM signatures [left]; exploring presently in **MORPH** FM [right]

  → transformer-based PDF model shows signs of **scaling and emergent learning**; capturing low-x PDFs requires sufficiently large training corpus of PDF samples

TJH, **Foreman, Kriesten**: 2610.xxxx



Gao, Gao, TJH, Liu, Shen: 2211.01094



toward practical theory world models: progress and outlook

- AI revolution has deep implications for HEP theory: not only how theorists work, but an emerging science of fundamental AI for/from particle theory; **necessitates coordinated effort(s)**
  - AI-Theory Induction Loop --- pure theory development for optimal representations for AI
  - Synergy with HEP experiments: data must be AI- *and* theory inference-ready
  - **Theory Foundation Models are valuable to both challenges**
  - Natural problem for dedicated working groups to establish conventions, best practices

- This talk: representative examples drawn from a recurring template, **inverse problems in HE(N)P theory and phenomenology (BSM/SMEFT, PDFs, ….)**
  - Pay-off: scalable, reusable models which amortize forward sim once; invertible on demand
  - ongoing applications to the BSM-PDF entanglement problem

THANKS!!

Backup

the challenge of model-independent New Physics searches

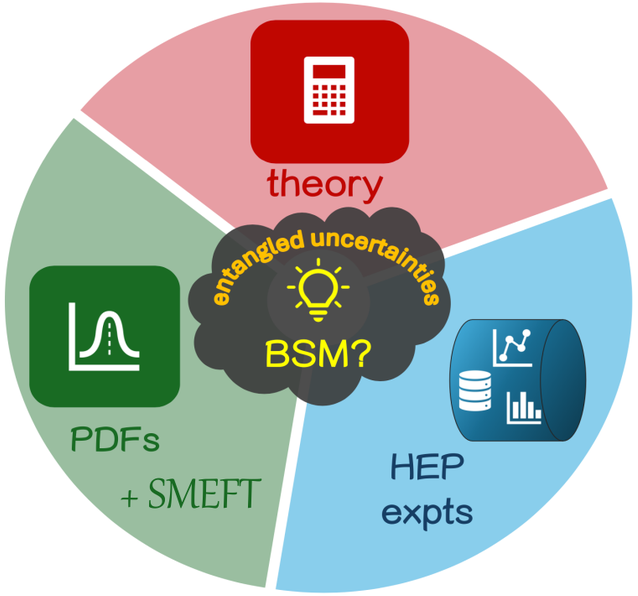
- definitive **signatures of BSM remain elusive**, despite extensive searches

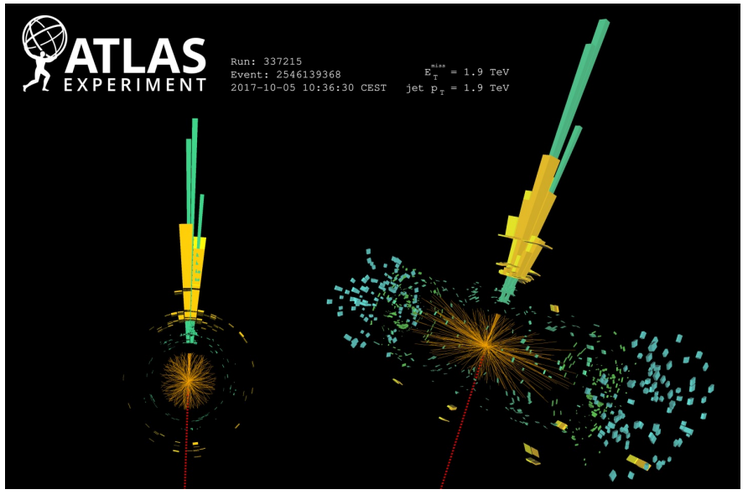
→ New Physics may lie above the kinematic reach of colliders — entering only virtually (non-resonant)

→ collider signatures may be very subtle; *e.g.*, appearing in tails of kinematical distributions

---

→ reach enhanced by model-independent BSM parametrizations, and novel analysis strategies





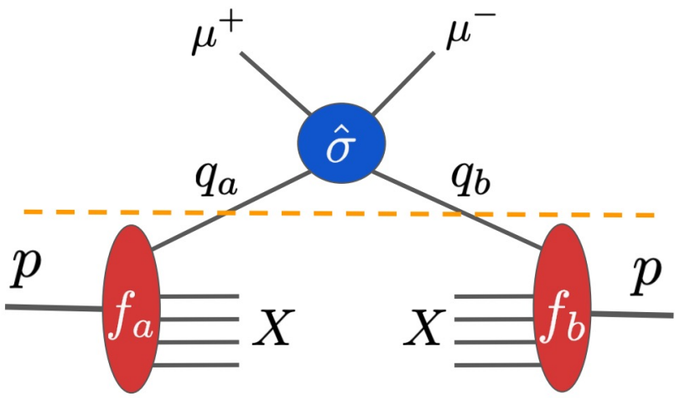
→ widely adopted framework: **standard model effective field theory (SMEFT)**

high theory accuracy; modeling (*e.g.*, SMEFT, PDF) precision; faithful uncertainties all essential to theory interpretation high-energy collider data

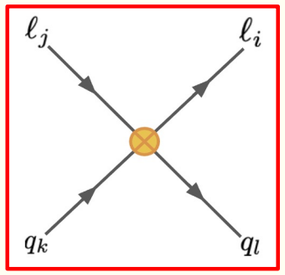
constraining massive BSM in SMEFT

❑ **SMEFT provides a systematic expansion for possible BSM interactions** mediated by New Physics at $\Lambda \gg$ EW scale; integrate these away to leave higher-dimensional operators involving SM fields

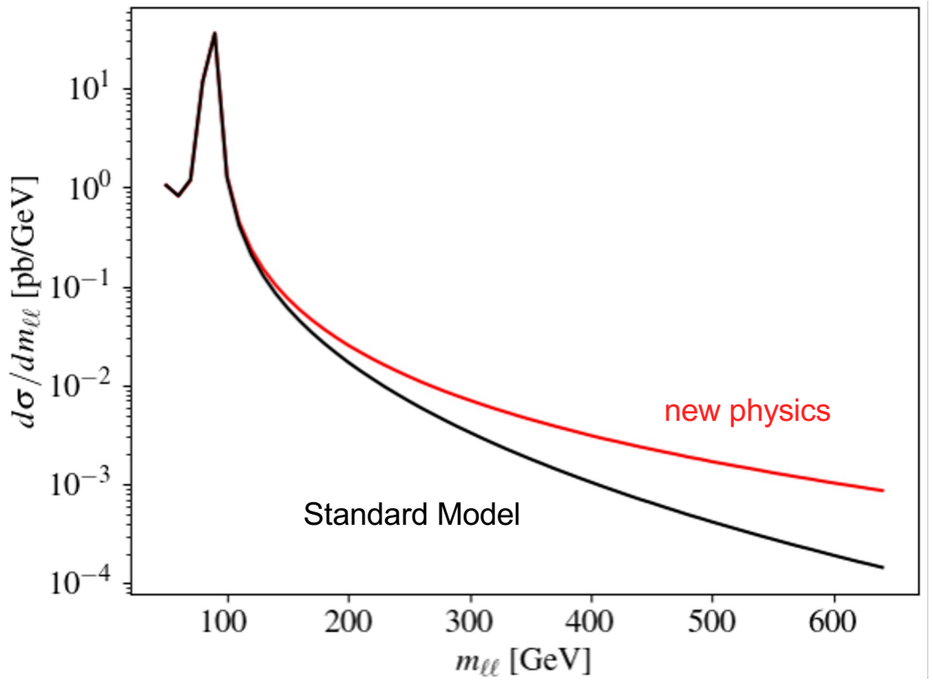
$$\mathcal{L}_{\mathrm{SMEFT}} = \mathcal{L}_{\mathrm{SM}} + \sum_i \frac{C_i O_i^{(6)}}{\Lambda^2} + \cdots$$



+



SMEFT contact interaction

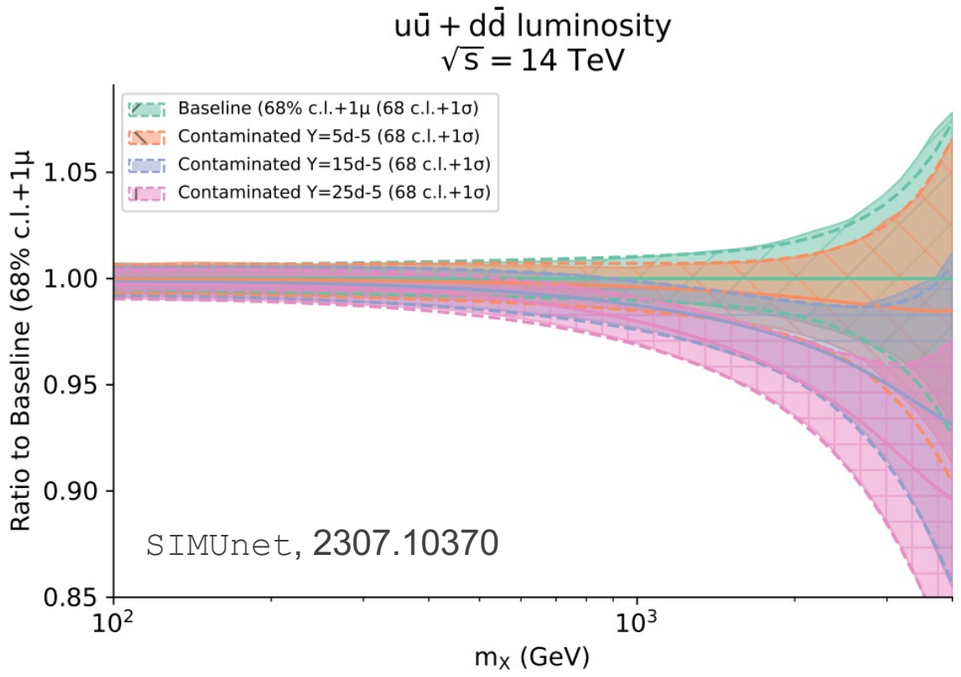


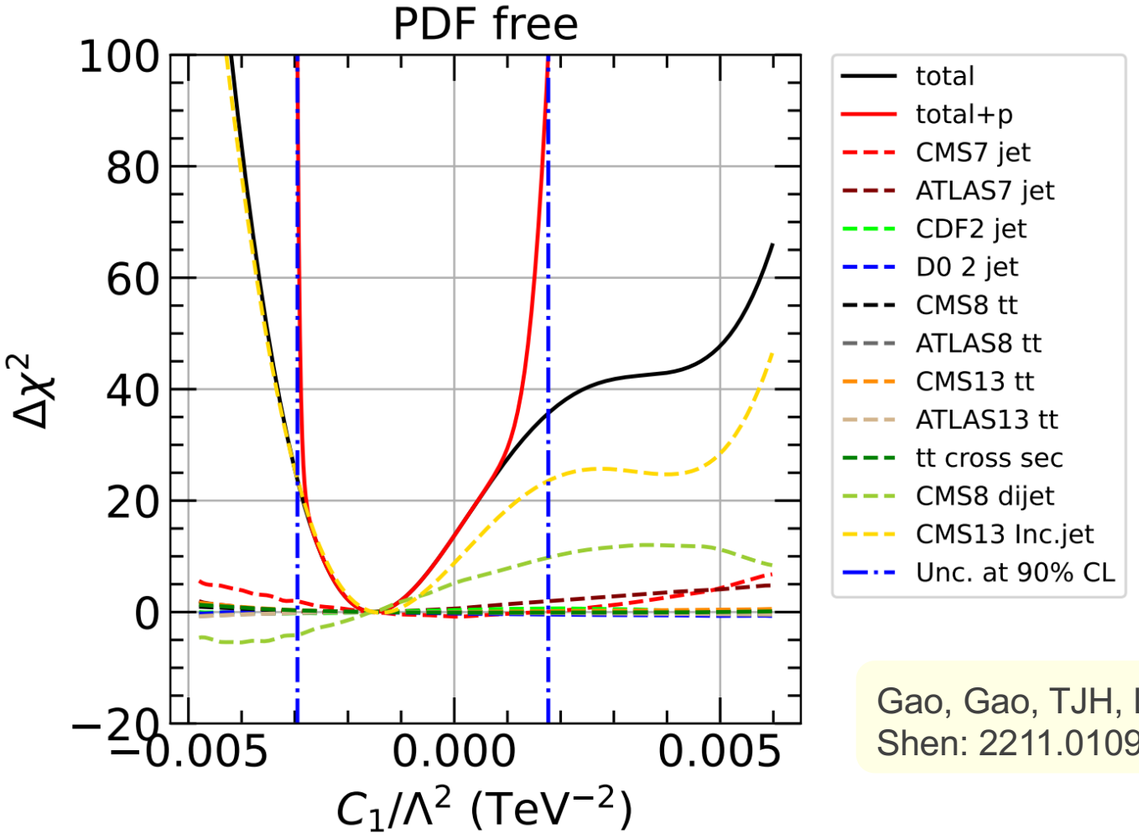
❑ inclusion of SMEFT contributions [here, dim-6] produce smooth excesses (deformations) in tails of cross sections (e.g., neutral-current Drell-Yan)

→ challenge: the SMEFT parameter space is large – **difficult inverse problem**

the PDF-BSM (SMEFT) entanglement problem

- **BSM entangled with PDF uncertainties**: hadronic data used in New Physics searches also constrain PDF shape parameters!





Gao, Gao, TJH, Liu, Shen: 2211.01094

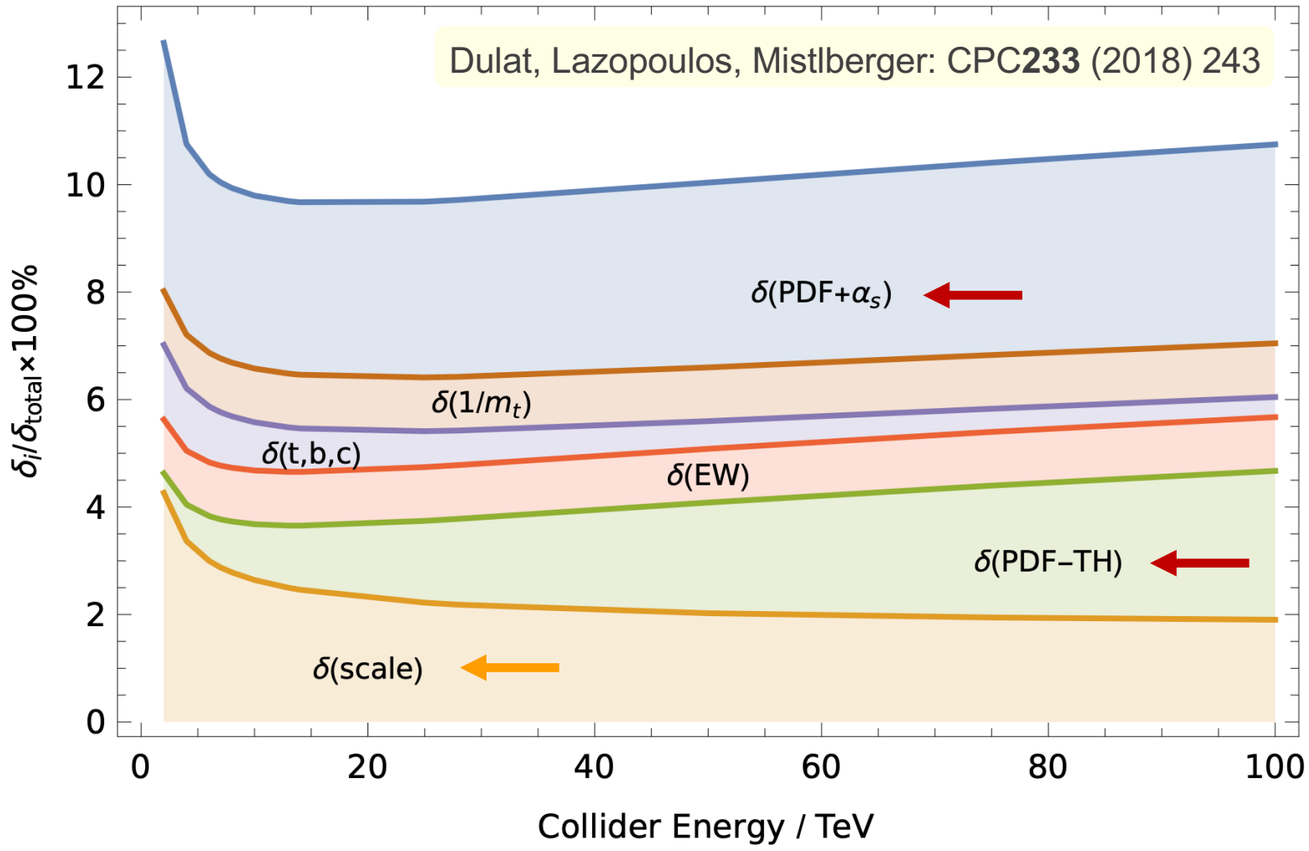
→ need to unravel SMEFT/PDF modeling uncertainties – implies **larger scope for theory inferencing** with foundation model

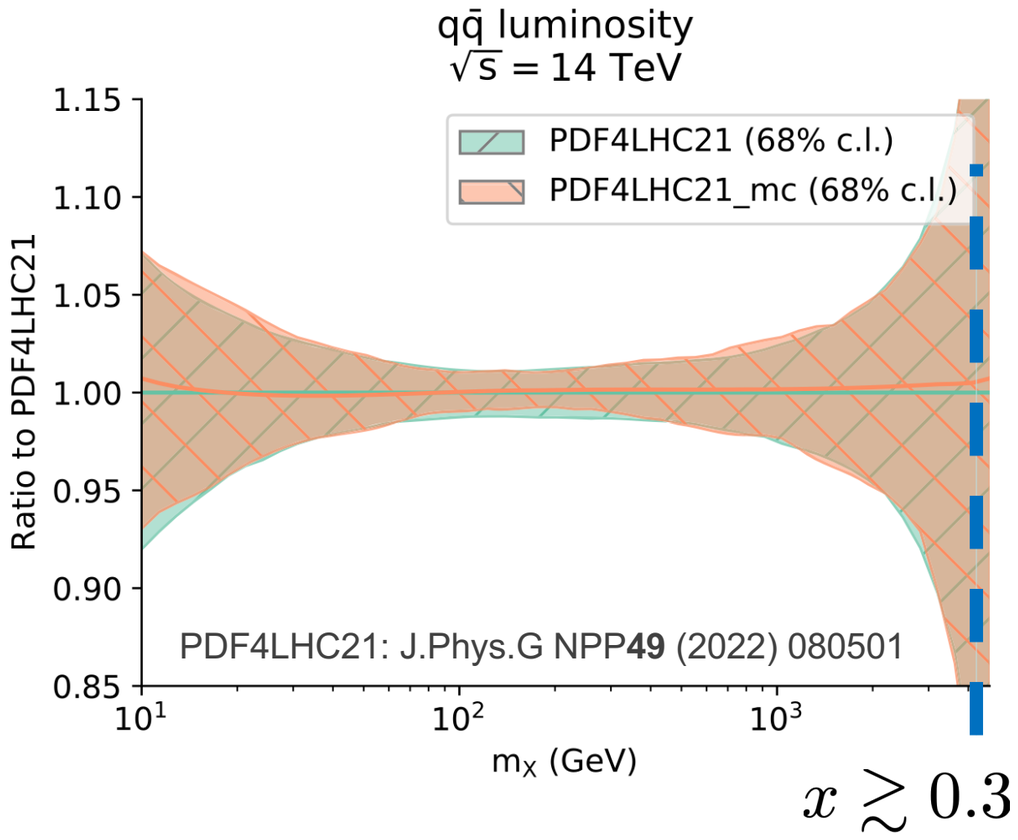
- joint fits alongside PDFs: **jet data** accommodate wider contact interaction ranges (C_1/Λ^2) relative to fixed-PDF extractions

PDFs as precision bottleneck at LHC

- Higgs; Z', W' production (as exemplar cases): uncertainties receive substantial PDF contribution

→ significant role from related effects (perturbative mismatch, scale, quark masses, …)



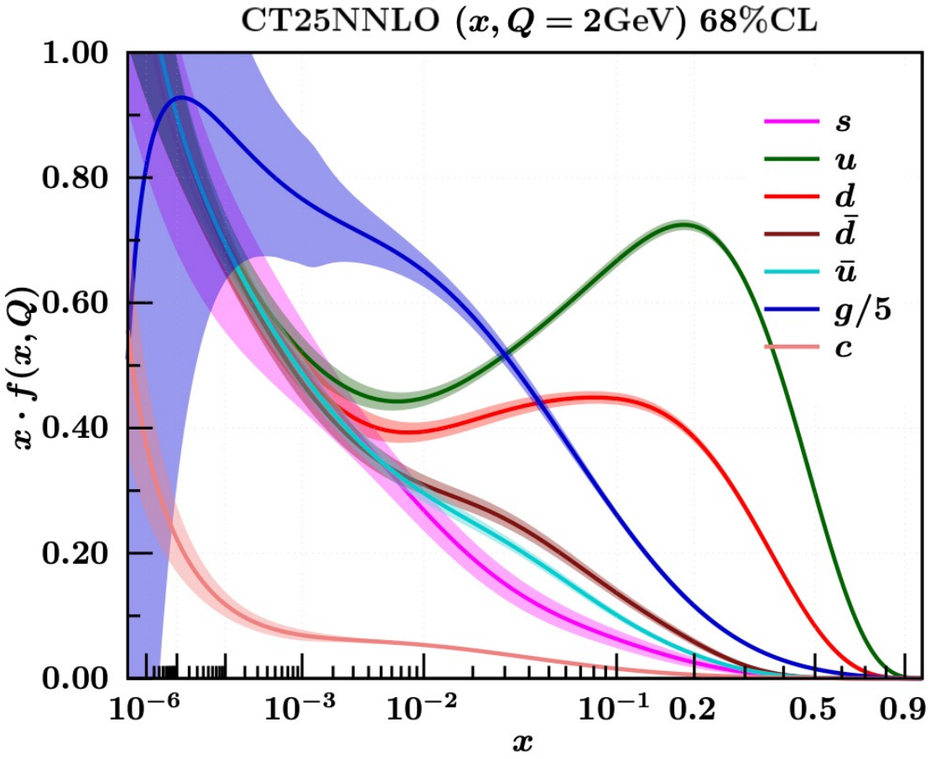


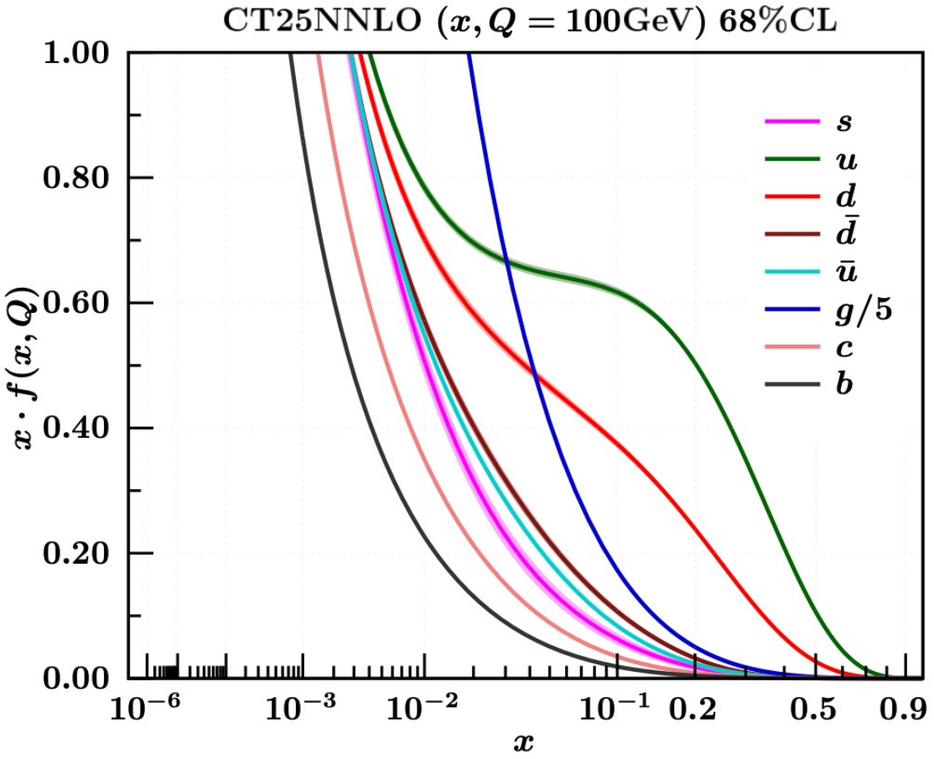
$x \gtrsim 0.3$

→ (high-x) PDF uncertainties limit searches for TeV-scale particles at the LHC

newest CTEQ-TEA global fit: **CT25** analysis

- CT25 global analysis PDF grids now public, 2512.19779 (includes corresponding series in α_s and n_F)
- forthcoming 'long paper': systematic exploration of fit variations; phenomenological implications



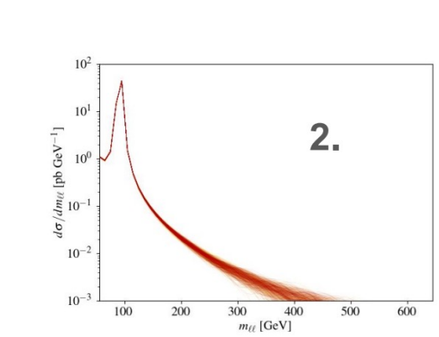


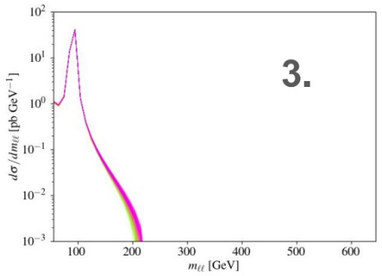
→ includes adoption of updated CT dynamical tolerance prescription for PDF uncertainties

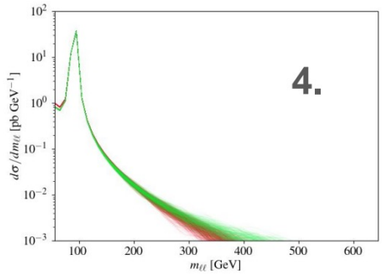
→ analysis variants: CT25 (baseline); CT25FlatP (proton-only prior); CT25m (parametrization ensemble); CT25pd (deuteron-corrected data)

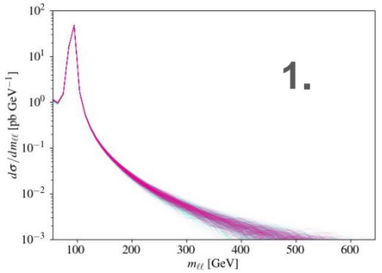
interpretability: interrogating the learned latent representation

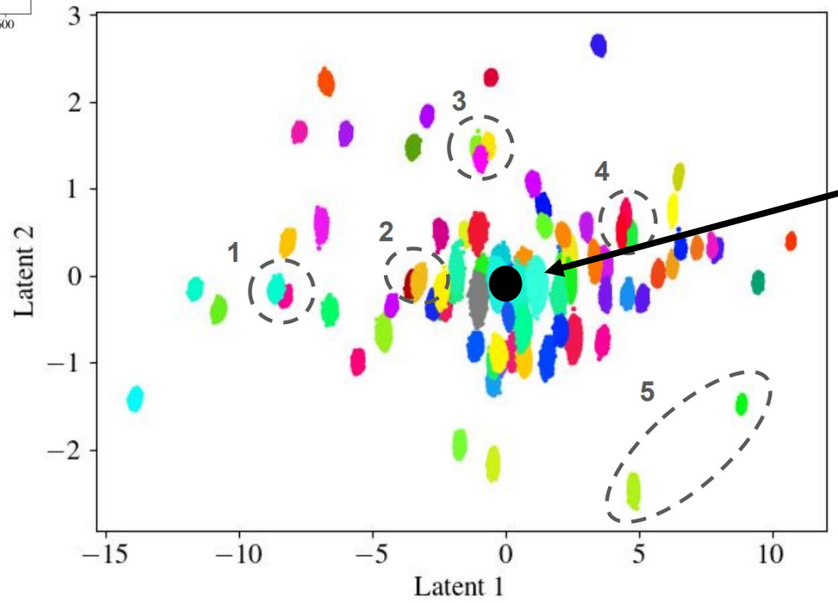
- traversals across latent space correspond to **controlled morphing of high-mass tails**; clusters = families of similar SMEFT deformations

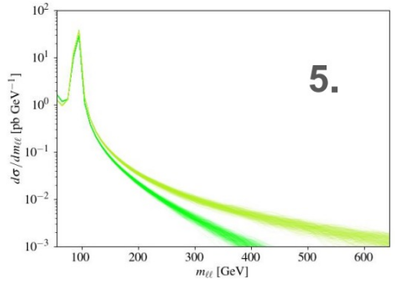










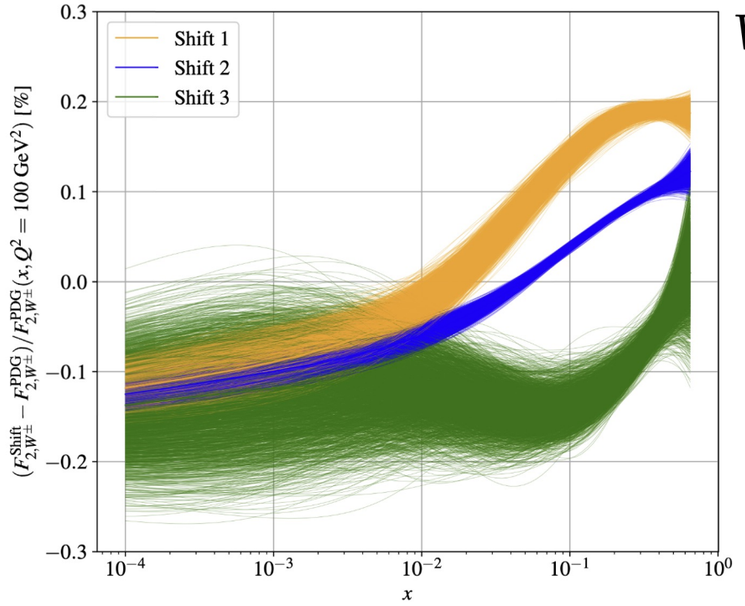


SM-like behavior

→ small-deformation universes concentrate around this central locus → SM sits at (0,0)

(evidential) learning BSM theory differences alongside PDF variations

$$F_2^{W^\pm}(x,Q^2) \equiv \frac{1}{2}\left(F_2^{W^+}(x,Q^2) + F_2^{W^-}(x,Q^2)\right)$$

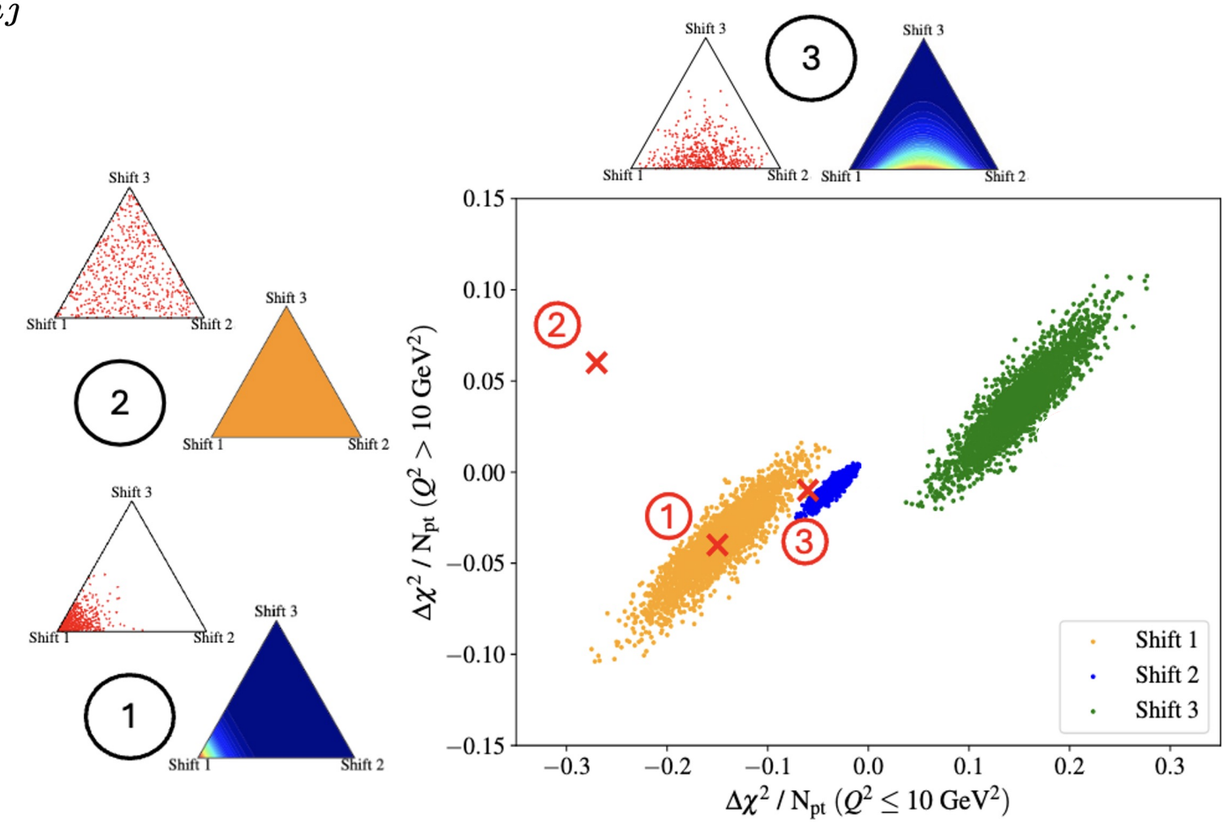


$$V_{ij} \to V'_{ij} = \Delta_{ij} \cdot V_{ij}$$

$$\mathcal{L}_{\text{WEFT}} \supset -\frac{2V_{ij}}{v^2}\left\{\left[1+\epsilon_L^{ij}\right]_{ab}\left(\bar{u}^i\gamma^\mu P_L d^j\right)\left(\bar{\ell}_a\gamma_\mu P_L \nu_b\right) + \text{h.c.}\right\}$$

Kriesten, TJH: 2412.16286

- non-standard neutrino interactions can shift high-x DIS structure functions (*e.g.*, CDHSW predictions)



- EDL techniques: distinguish BSM scenarios up to PDF uncertainties; quantify confidence in BSM separation