

Agentic AI Applications to particle theory and collider phenomenology

Konstantin Matchev

University of Alabama

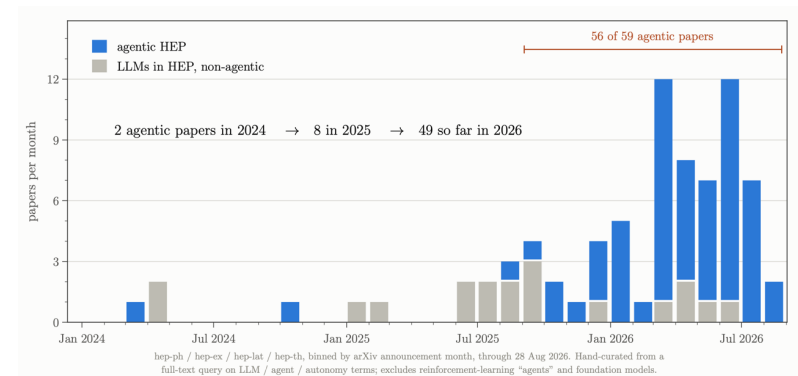
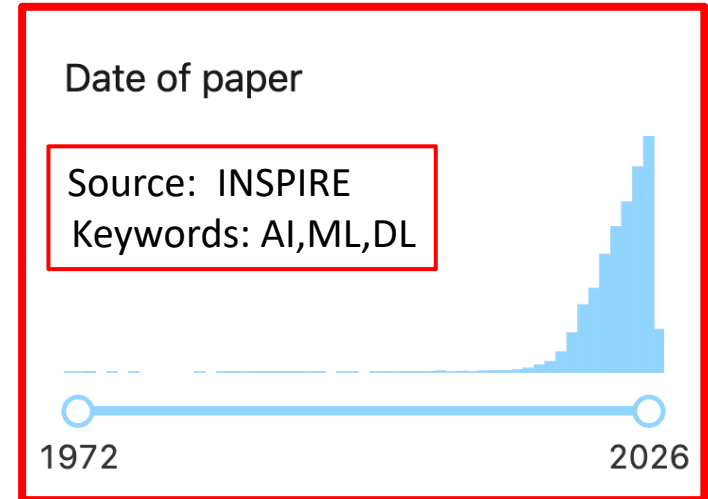


In collaboration with: G. Fleming, S. Gleyzer, S. Hoeche, M. Knipfer, K. Matcheva,
T. Menzo, S. Mrenna, A. Roman, P. Shyamsundar
2512.15867, 2602.19387, 2603.26953, 2603.26990, 2609.00107, 2609.xxxxx



Outline

- Types of Machine Learning
 - ‘narrow AI’ (regression, classification...)
 - ‘proto-AGI’ (agentic stuff)
- Agentic scientific workflows
 - **HEPTAPOD** (HEP Toolkit for Agentic Planning, Orchestration & Deployment)
 - Agentic **Diagrammatica**
 - Other examples in:
 - E. Moreno’s plenary talk this morning
 - Talks in this and the previous session
- Pressing tasks in agentic AI
 - Benchmarking
 - Harnessing/Safety
- Public Service Announcements
 - Agentic AI BSM workshop, Fermilab
 - CLARIPHY Meeting, Madison WI
 - **US AI Theory** **whitepaper**



Compiled by T. Menzo (last week)

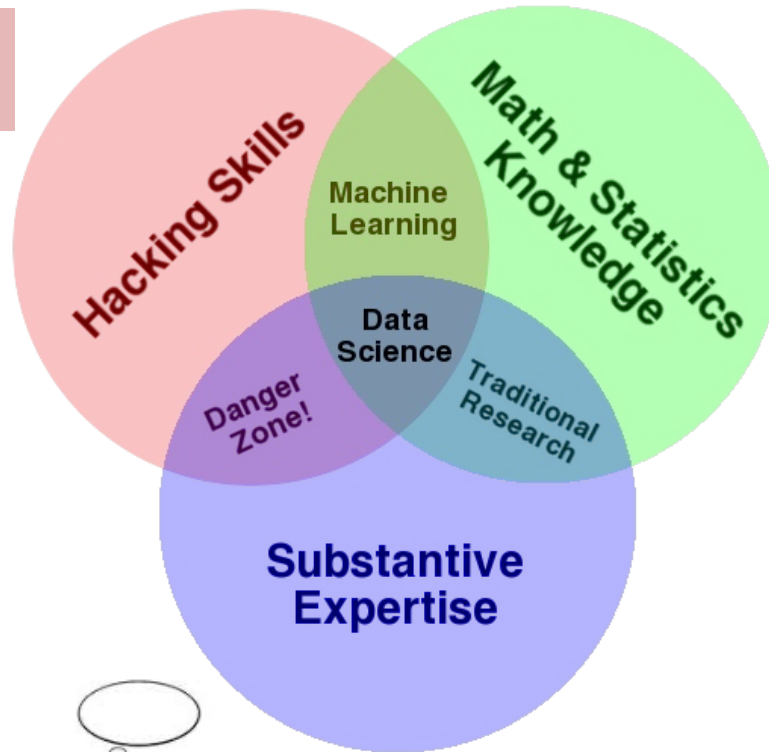
Data Science in STEM

<http://drewconway.com/>

Automation



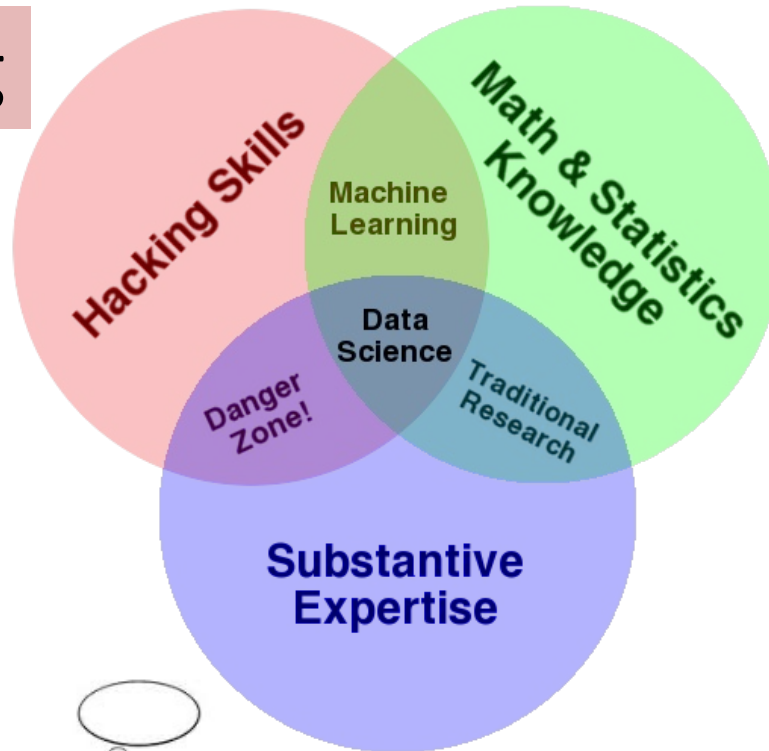
Validation



Intuition

The Tasks of a High-Energy Theorist

Programming



Calculating

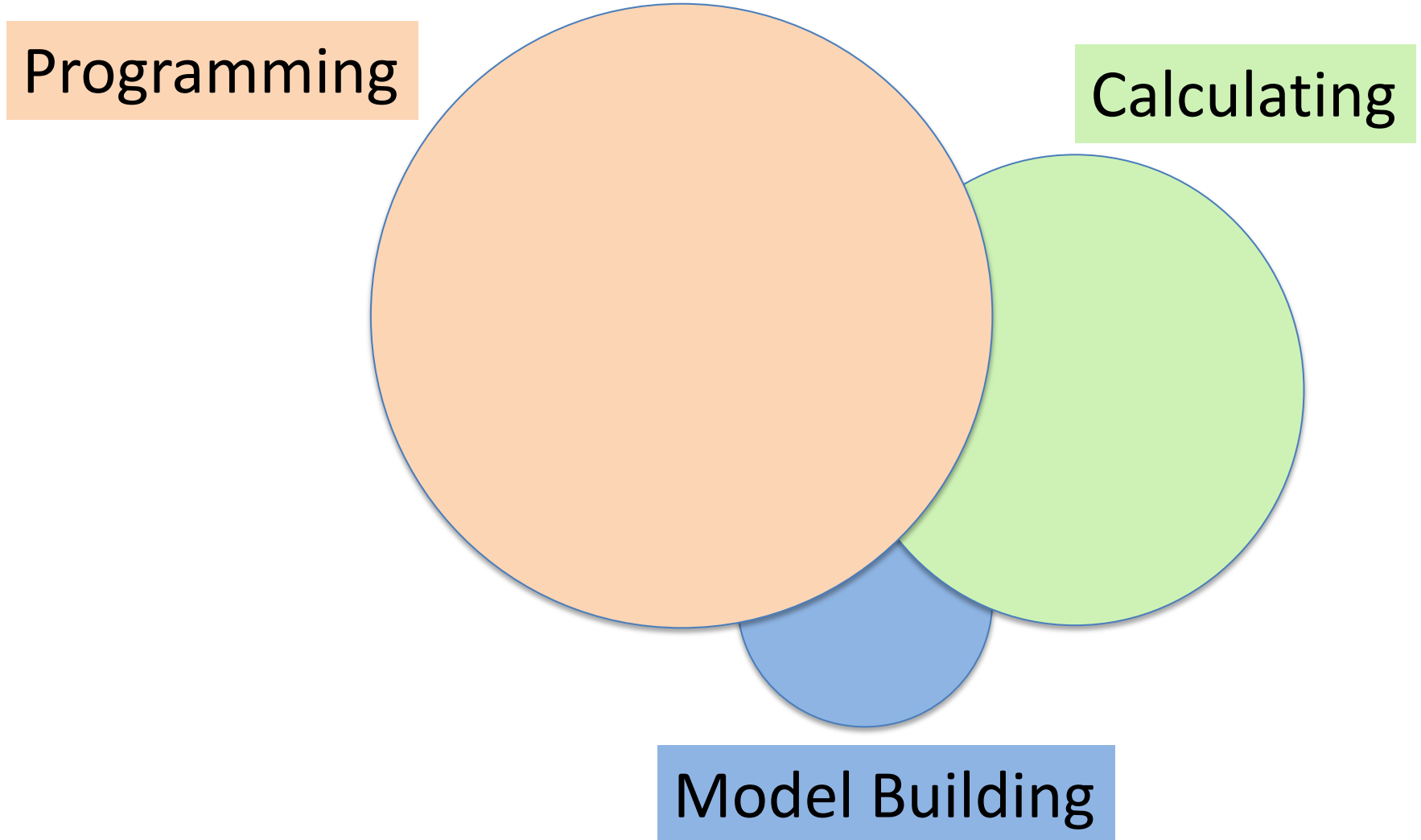


designed by freepik

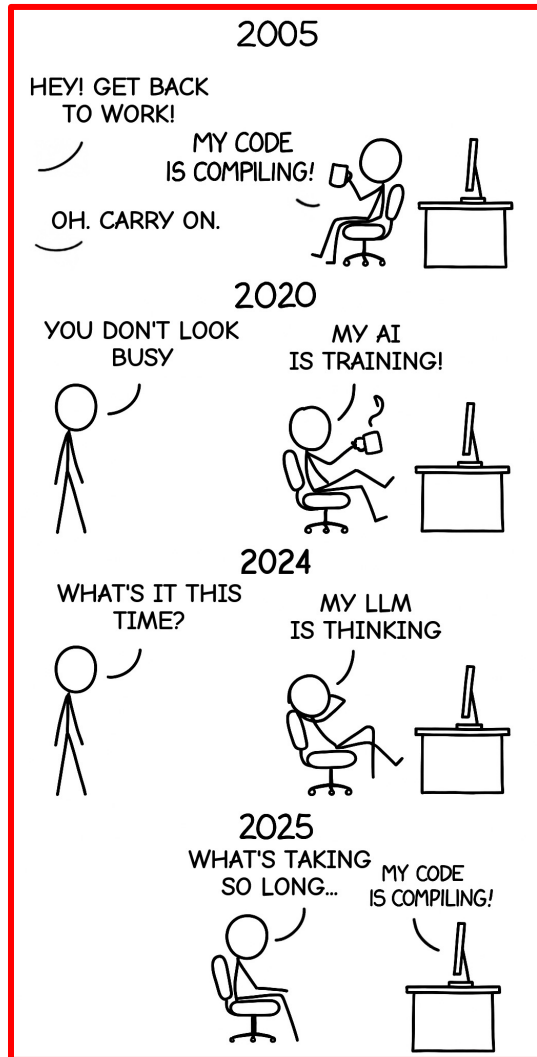


Model Building

Time Spent by a High-Energy Theorist



Enter LLMs



Activity Type

	I am an expert	I'm not sure how to do
Enjoyable	X	X
Boring	Yes!	X

Agentic programming

- Modern LLMs are
 - Pretty **reliable** compressions of human knowledge
 - Becoming very **powerful**
 - **Evolving** rapidly
- Agent: Given **context** C_t at a decision step t , and an **action** A_t selected from an action space, an agent is a system that implements a **conditional distribution** over actions, $p(A_t | C_t)$ and is embedded in a **feedback loop** such that executed actions influence future context.

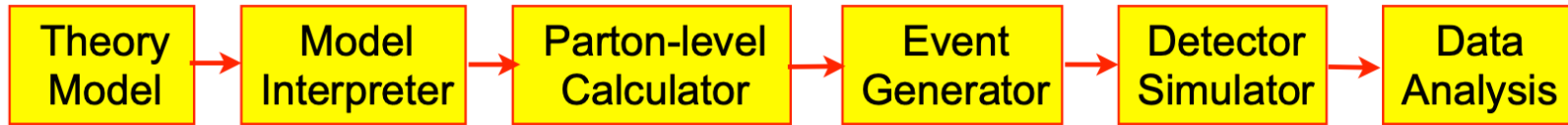
Traditional programming

- Specify how to do something
- Scripting
- Control flow is explicit and tractable.

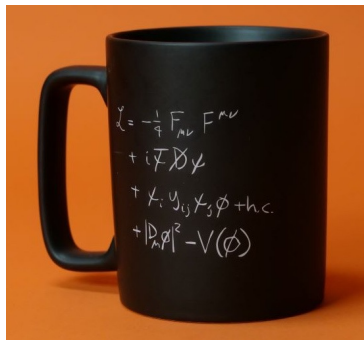
Agentic programming

- Specify what to do
- Prompt engineering
- Control flow is non-deterministic

The simulation pipeline in HEP

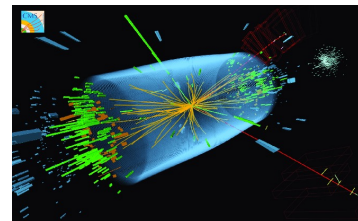
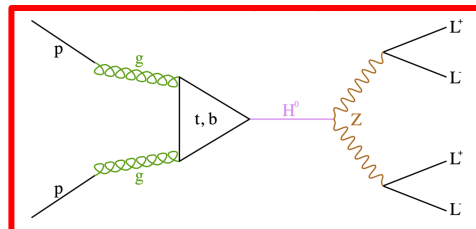


- Symmetries
- Representations
- Fields (basis)
- Lagrangian
- Aesthetics

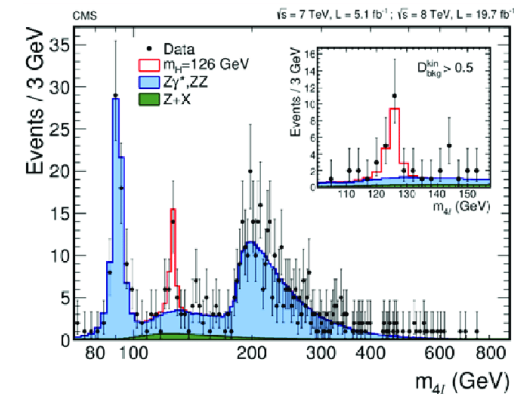


See talk by
Theo Heimel

- RGEs
- Vacuum structure
- Mass spectrum
- Feynman rules
- Feynman diagrams
- Cross sections (rates)
- Branching fractions (rates)
- MC event generation
- Fragmentation
- Hadronization
- Higher Order Corrections



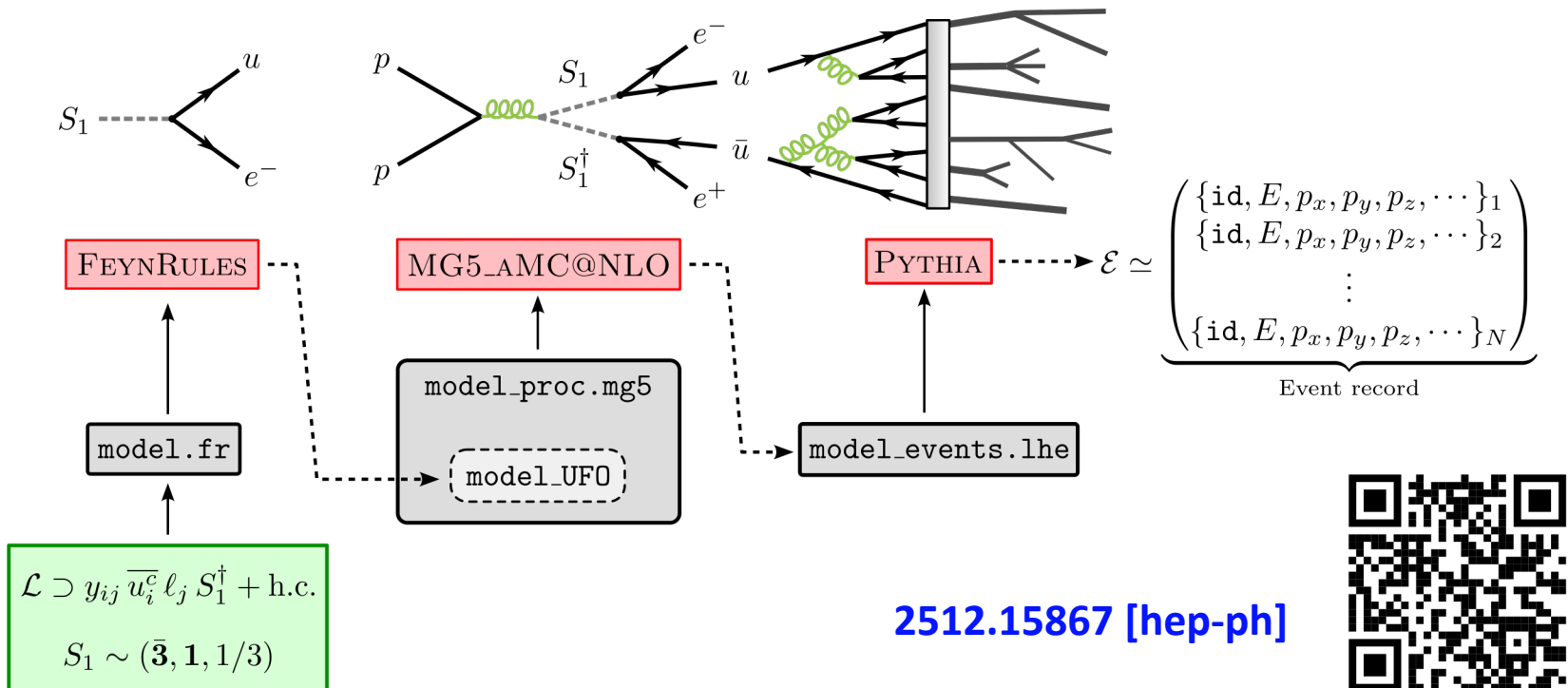
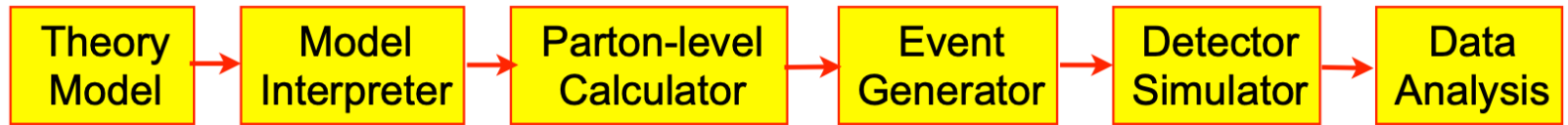
- Resolution
- RecoObjects
- Histograms
- Backgrounds
- Selection cuts
- ROC curve
- Error analysis



Agents to the rescue

- HEPTAPOD** (HEP Toolkit for Agentic Planning, Orchestration & Deployment)

Menzo, Roman, KM, Gleyzer, Fleming, Hoche, Mrenna, Shyamsundar 2025



2512.15867 [hep-ph]

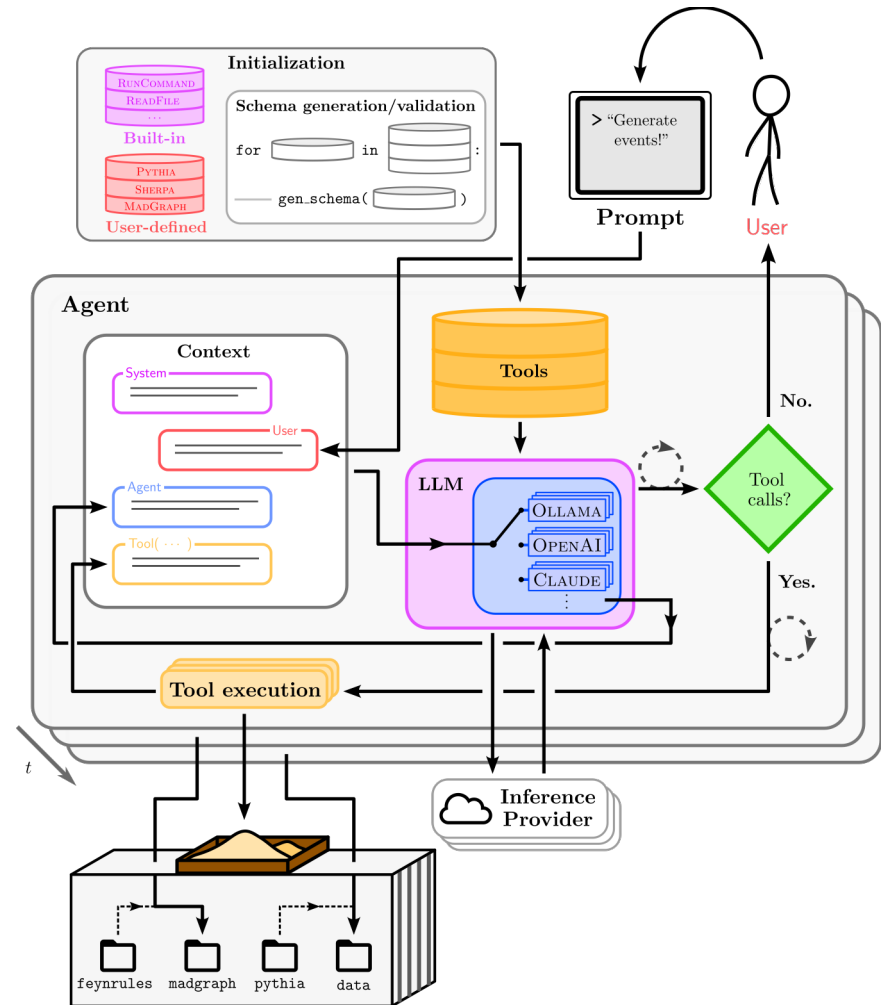
See also D. Schiller's talk on MadAgents

How does the agent work?

Menzo, Roman, KM, Gleyzer, Fleming, Hoche, Mrenna, Shyamsundar 2025

HEPTAPOD: A HEP-focused toolkit for agentic programming that integrates LLMs into HEP workflows spanning theoretical calculations, simulation, and data analysis.

- Define workflows at the level of **physics intent**, not scripts
- Query particle properties, literature, and unit conversions
- Execute multi-stage pipelines using run-cards and filesystem tools
- Equipped with tools like to-do lists and terminal access
- Approval hooks preventing misbehavior
- Fully reproducible, auditable logs
- Built on the **Orchestral AI** engine [Roman&Roman 2601.02577]



See also A. Upadhyay's talk on Ariadne

HEPTAPOD

HEP Toolkit for Agentic Programming, Orchestration, and Deployment

Emphasis on toolkit: 3+ harnesses, 60+ tools, prompts, skills

#	Tool	Reliability	Auditable	Object	External	Repeated	Diagrammatica / exact symbolic tools					
Event generation and simulation							20	ComputeSymbolicAmplitudeTool	*	✓	✓	✓
1	FeynRulesToUFOTool	✓		✓	*	✓	21	RunWolframTool	✓		*	✓
2	MadGraphFromRunCardTool	✓	✓	✓	*	✓	22	RunWolframScriptBatch	✓		*	✓
3	PythiaFromRunCardTool	✓		✓	*	✓	23	SimplifyResultTool	*			✓
4	SherpaFromRunCardTool	✓		✓	*	✓	24	ConvertToPythonTool	✓	✓	✓	*
5	JetClusterSlowJetTool	✓	✓	*	✓	✓	NDA and FeynGraph					
Event-format conversion and array bridges							25	EstimateDecayWidthNDATool	✓	✓	*	✓
6	LHETOJSONLTool	*	✓	✓		✓	26	EstimateBranchingRatioNDATool	✓	✓	*	✓
7	EventJSONLToNumpyTool	✓		✓		*	27	EnumerateDiagramsTool	✓	✓	*	✓
8	JetsJSONLToNumpyTool	✓	✓	✓		*	28	DiagramVisualizationTool		*	✓	✓
Kinematic analysis and event selection							PDG and INSPIRE tools					
9	CalculateInvariantMassTool	*	✓	✓		✓	29	PDGDatabaseTool	✓		✓	*
10	CalculateTransverseMomentumTool	✓		✓		*	30	PDGSearchTool	✓	✓	*	✓
11	CalculateDeltaRTool	*	✓	✓		✓	31	PDGPropertyTool	✓		✓	*
12	ApplyCutsTool	✓	*	✓		✓	32	InspireSearchTool		✓	*	✓
13	GetHardestNTool	✓	*	✓		✓	33	InspirePaperTool		✓	✓	*
14	GetHardestNJetsTool	✓	*	✓		✓	34	InspireCitationTool			✓	*
15	FilterByPDGIDTool	✓	*	✓		✓	35	InspireBibTeXTool	✓		✓	*
16	SortByPtTool	✓	*			✓	36	InspireAuthorTool			✓	*
17	MergeObjectCollectionsTool	✓	✓	*		✓	37	InspireInstitutionTool			✓	*
18	FilterByDeltaRTool	*	✓	✓		✓	38	InspireConferenceTool			✓	*
Resonance reconstruction							39	InspireJournalTool			✓	*
19	ResonanceReconstructionTool	✓	✓	*		✓	40	InspireExperimentTool			✓	*
							41	InspireReadingListTool		✓	✓	*
							42	InspireNotesTool		✓	✓	*

Tony Menzo (U. Alabama + Fermilab)

24

Agentic phenomenology

- **“Agentic Diagrammatica”**: additional tools for
 - Naïve dimensional analysis of any number of tree-level Feynman diagrams
 - Exact diagrammatic analysis for tree-level symbolic calculations

Pictures from the agent in action

User

Plan a Monte Carlo signal-validation analysis of the BSM scalar leptoquark defined by the supplied run cards. Generate 10,000 events, select objects, and create lepton-jet invariant-mass distributions for each parameter point. Attempt to isolate and reconstruct the leptoquark resonance by applying reasonable lepton-jet separation and experimental cuts. Compute and plot the resulting $m_{LQ}^{\min}$ distribution. Write a todo list first and then execute the todo list until completed.

```
## Monte Carlo Signal-Validation Analysis for S1 Leptoquark

- [*] Generate UFO model from FeynRules file
    'feynrules/models/S1_LQ_RR.fr'
- [] Generate parton-level events with MadGraph using card
    'madgraph/cards/S1_LQ_RR_pp_lqlq_scan.mg5'
    - set 'nevents = 10000'
    - keep all parameter points
- [] Shower and hadronize events with Pythia using card
    'pythia/cards/S1_LQ_RR_pp_ljlj.cmd'
    - shower_lhe = True
- [] Convert Pythia JSDML events to NumPy array
- [] Apply kinematic cuts to leptons and jets
    - leptons: pt>20 GeV, leta<2.5, PDG IDs 11,13 (and charge
      conjugates)
    - jets: pt>30 GeV, leta<2.5
- [] Select hardest 2 leptons and hardest 2 jets per event
- [] Compute lepton-jet invariant masses for all pairings and retain
    the minimum per event ('m_min_LQ')
- [] Plot the 'm_min_LQ' distribution (histogram, 50 bins,
    appropriate range) and save as 'mmin_LQ.png'
- [] Summarize results (number of events, cross-section, plot
    location)
```

GetHardestNJets(...)
Same as GetHardestNJets above with run_01 → run_02

ResonanceReconstruction(min_delta_r=0.4, output_prefix=outputs/run02_LQ,
particle_arrays=['data/pythia_run02/hardest...',
template=two_body_symmetric

Error: Processing Error
- Reason: Array 1 not found: data/pythia_run02/hardest2_jets.jsonl

GetHardestNJets(input_path=data/pythia_run03/jets.jsonl, ,
n_hardest=2,
output_path=data/pythia_run03/hardest2...

Error: File Not Found
- Reason: Input file not found: data/pythia_run03/jets.jsonl

JetClusterSlowJet(...)
Same as above with run_02 → run_03

GetHardestN(...)
Same as GetHardestN above with run_02 → run_03

Approval Required

This command will delete the file 'example_module.py'

Technical Details

Tool:

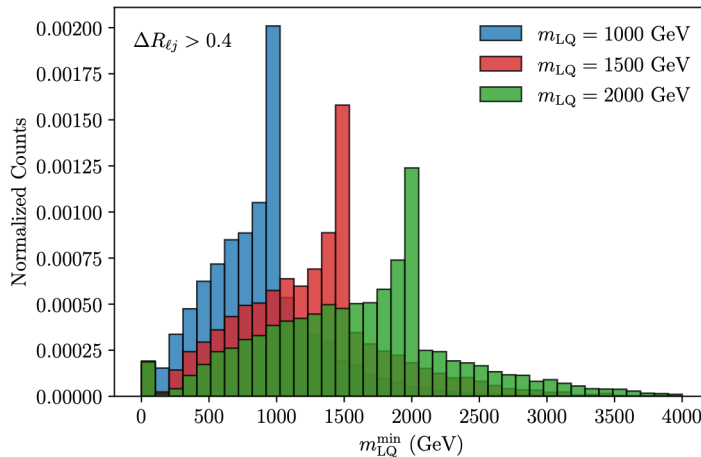
runcommand

command:

rm example_module.py

Deny

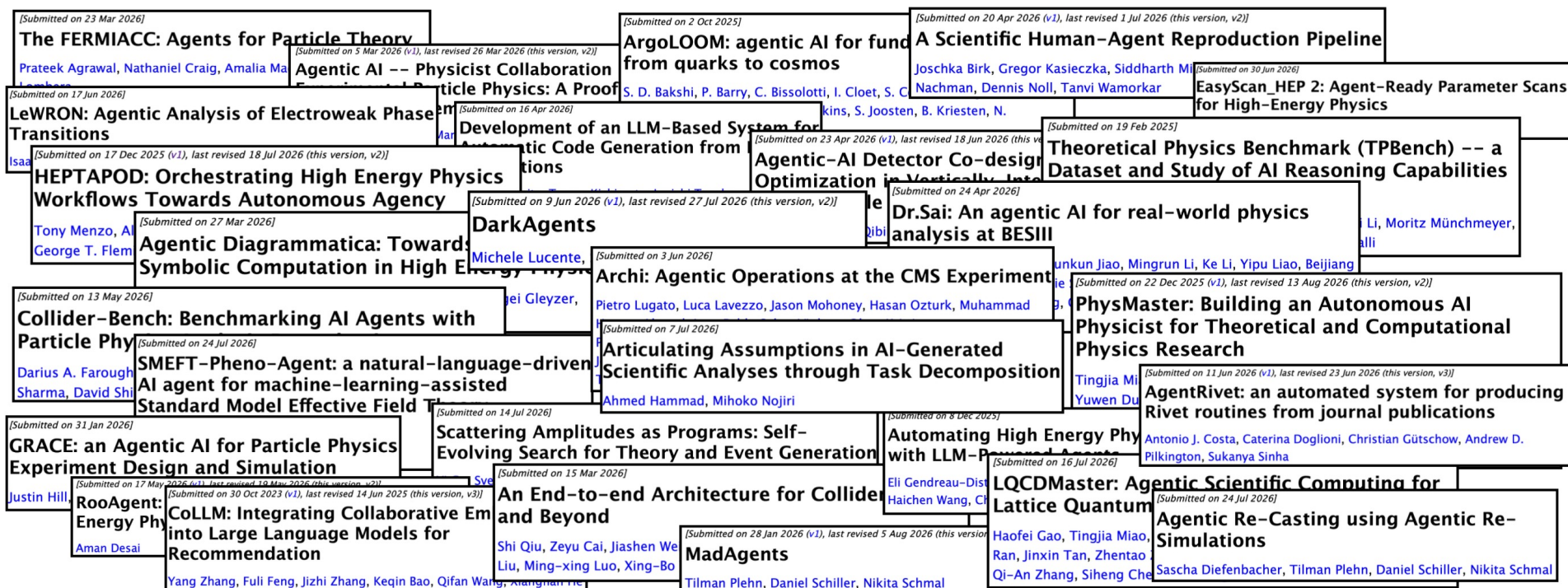
Approve



Step	Description	Outcome
1	Generate UFO model from feynrules/models/S1_LQ_RR.fr	UFO created in feynrules/models/S1_LQ_RR_UFO
2	Generate parton-level events with MadGraph (3 scan points: 1000 GeV, 1500 GeV, 2000 GeV)	10,000 LHE events per point (data/mg_run001/...)
3	Shower & hadronize with Pythia8 (pythia/cards/S1_LQ_RR_pp_ljlj.cmd)	10,000 final-state events per point (data/pythia_run0X/events.jsonl)
4	Jet clustering (anti- k_t $R = 0.4$)	Jets stored in data/pythia_run0X/jets.jsonl
5	Lepton selection (electrons ± 11 , muons ± 13)	leptons_all.jsonl for each run
6	Keep hardest 2 leptons & hardest 2 jets per event	hardest2_leptons.jsonl, hardest2_jets.jsonl
7	Reconstruct leptoquark candidates (two-body symmetric template, $\Delta R > 0.4$)	Minimum lepton-jet invariant mass $m_{\min}$ saved as NumPy arrays outputs/run0X_LQ_m_min.npy
8	Plot the distribution of $m_{LQ}^{\min}$ for the three mass points	Histogram saved as mmin_LQ.png
9	Summarize results	See below

The “Cambrian agent explosion”

Since ~November/December of last year (2025) there has been a “Cambrian explosion” of papers demonstrating the use of agents in HEP.



Tony Menzo (Fermilab + U. Alabama)

3

Benchmarking w/ HEPTAPOD

A Giant Table

Menzo, Fleming, KM, Mrenna, Roman 2609.00107

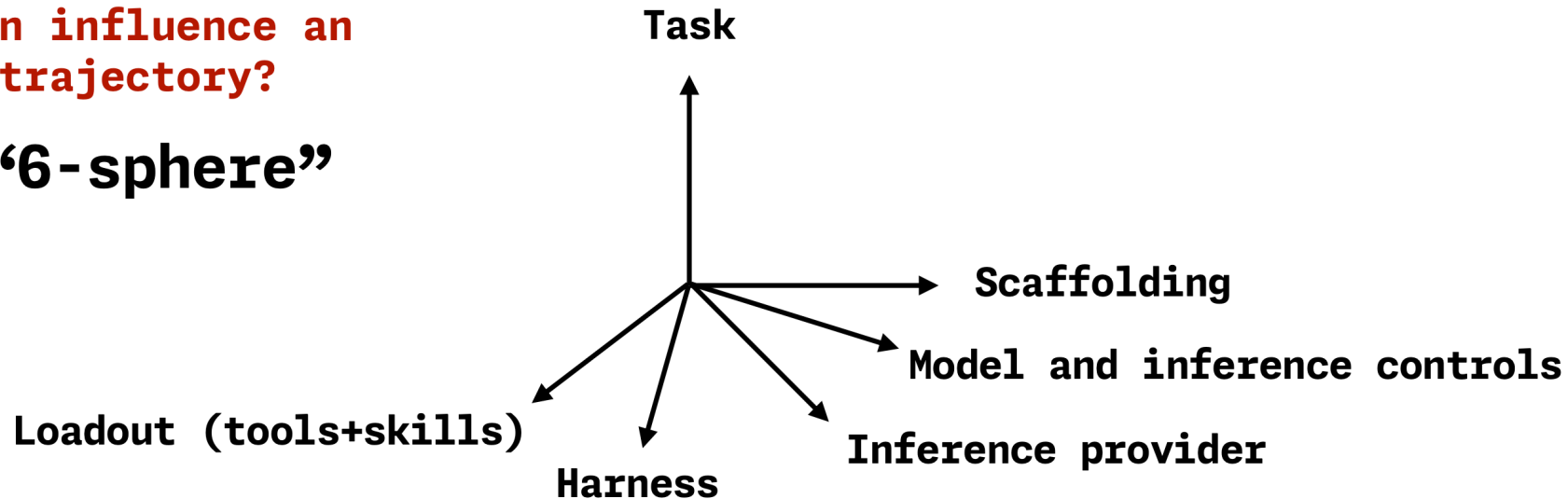
	Name	Architecture	MCP	Multi-agent	Domain / role
Free-form	MadAgents [4]	Free-form code generation	–	✓	Collider phenomenology
	HEP Data Analysis Agents [7]	Free-form generation in fixed workflow DAG	–	✓	Experimental analysis
	JFC [1]	Free-form generation, literature-grounded	–	✓	Experimental analysis
	CoLLM [8]	Free-form generation, structured prompt	–	✗	Collider DL analysis
	PhysMaster [9, 10]	Free-form generation, validated-trace memory	–	✓	Theory / lattice QCD
	Agents of Discovery [5]	Free-form code generation	–	✓	New-physics analysis
	ArgoLOOM [2]	Orchestrated module wrappers, RAG-grounded	–	✓	Cosmology, collider, nuclear
	ALEPH thrust [3]	Free-form generation, physicist-directed	–	✗	Experimental analysis
	SFitter Agents [6]	Orchestrator with consultant agents	–	✓	Global SMEFT re-casting
	CLVisc Agent [23]	Free-form execution, agent-authored skill	–	✗	Heavy-ion hydrodynamics
	Amplitude program search [35]	Coding agent in generate-evaluate-select loop	–	✗	Scattering amplitudes
	Detector co-design [36]	Agentic loop over bi-level optimizer	–	✗	Detector design
Structured handoffs	Dr. Sai [11]	Role agents over experiment software	–	✓	Experimental analysis
	DarkAgents [12]	Sub-agents with JSON handoff contracts	–	✓	Astroparticle / dark sectors
	SHARP [13]	Machine-readable project spec, subagents	–	✓	Analysis reproduction
	AgentRivet [14]	Provider-agnostic agents, output schemas	–	✓	Analysis preservation
	Semantic differencing [15]	Multi-agent trace and critique of generated code	–	✓	Analysis audit
	Selection-list codegen [16]	Structured selection list between stages	–	✓	Analysis reproduction
Fixed backend	LQCDMaster [17]	Hybrid: skill-guided generation, deterministic Wick tool	–	✓	Lattice QCD
	ColliderAgent [18]	CLI/template-mediated	✓	✓	Collider phenomenology
	HepScript [19]	Constrained-grammar DSL	–	✗	Experimental analysis
	EasyScan_HEP 2 [20]	Agent-written configuration, fixed backend	–	✗	Parameter scans
Valid-by-rejection	SMEFT-Pheno-Agent [21]	Locked configuration, phase manifests	–	✗	SMEFT phenomenology
	RooAgent [25]	Typed tool registry over LangGraph and MCP	✓	✗	Experimental analysis
	FERMIACC [24]	Typed schemas over a guarded pipeline	–	✓	Particle theory
	GRACE [26]	Tool-invocation DAG; value-level verification against physics constraints	–	✗	Detector / experiment design
	bsm_agent [27]	Tool calling over a symbolic backend	–	✗	BSM model building
	LeWRON [33]	Audited toolbox with auditor agents	–	✓	Particle cosmology
	Archi [34]	ReAct agent over corpus and live MCP tools	✓	✗	Experimental operations
Valid-by-construction	HEPTAPOD (event generation) [31]	Typed schema constrains decoding	✓	✗	Collider phenomenology
	HEPTAPOD (Diagrammatica) [32]	Typed schema constrains decoding	✓	✗	Symbolic computation

Testing the harness

“Is this thing on?”

What can influence an agents trajectory?

The “6-sphere”

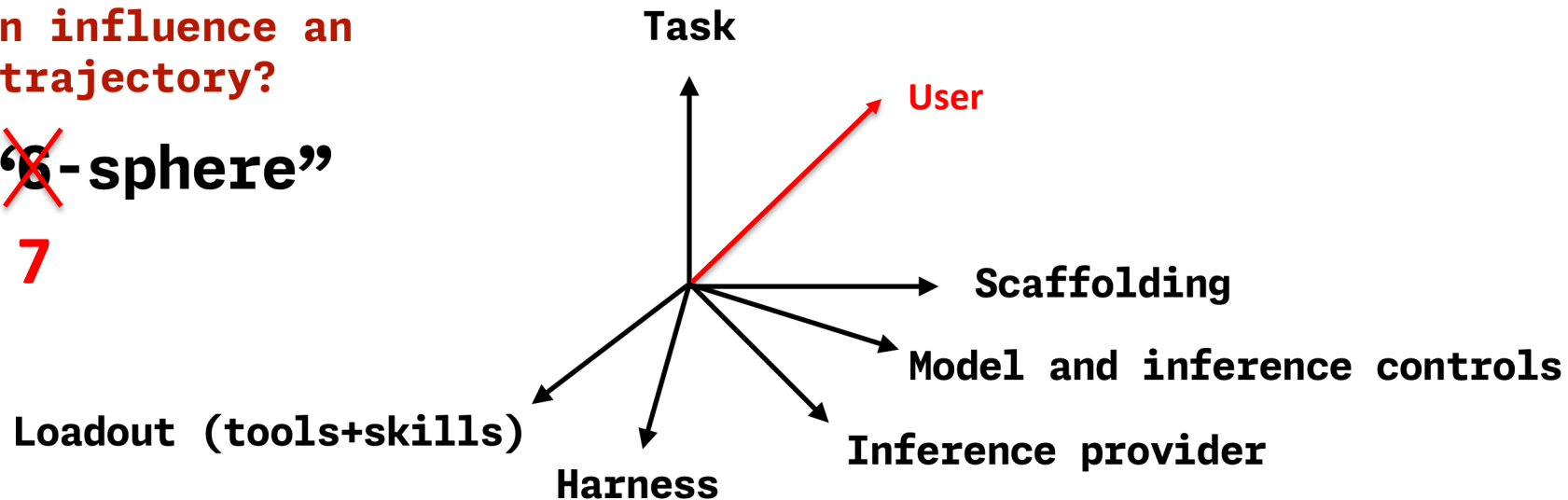


Testing the harness

“Is this thing on?”

What can influence an agent's trajectory?

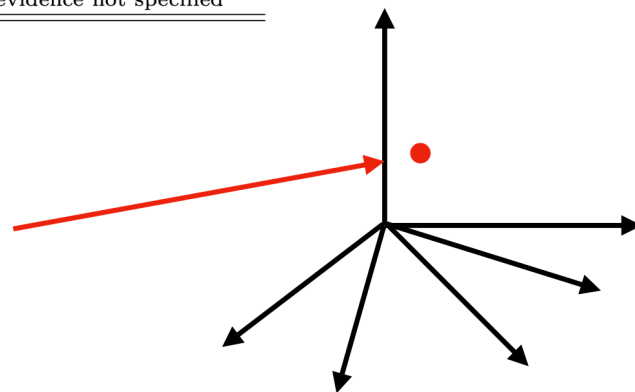
The “~~S~~-sphere”
7



Benchmarking in HEP

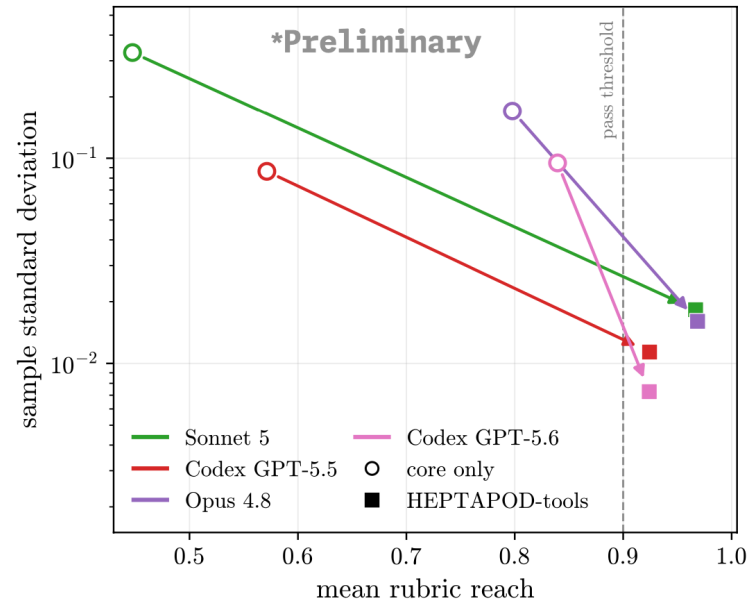
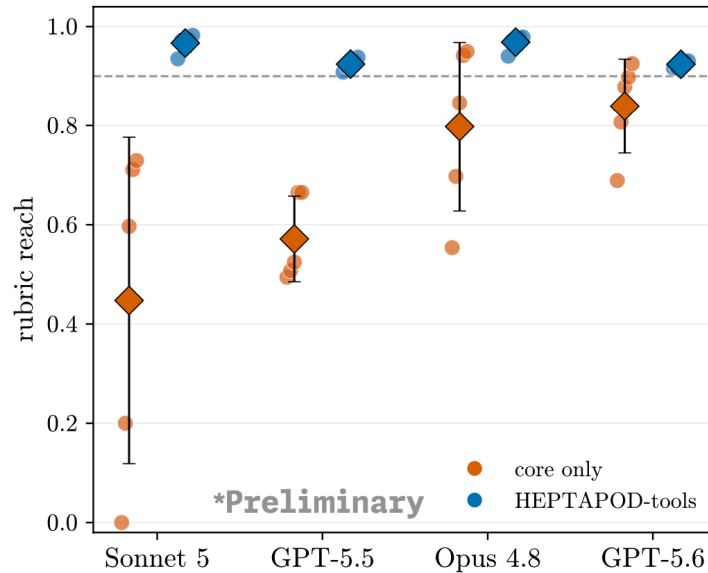
Name	Scope	Grading	Integrity control / retained evidence
Collider-Bench [22]	LHC analysis reproduction; 10 tasks from 4 papers	Histogram fidelity, no hand-written rubric; 3 repeats; cost reported	LLM judge audits workspace and session trace for fabrication
TPBench [37]	HEP theory and cosmology; 57 problems	Automated numeric verification; holistic grading of reasoning	Public/private split maintained against leakage
CritPt [38]	Frontier physics research incl. HEP; 71 challenges, 190 checkpoints	Physics-format-aware auto-grading; reliability metric; cost reported	Unpublished problems, guess-resistant by construction
PRL-Bench [39]	End-to-end physics research; 100 papers across 5 subfields	Objective verifiability; expert-validated tasks	Published-source tasks; retained execution evidence not specified

Each of these benchmarks only operate at a single point in the 6-sphere

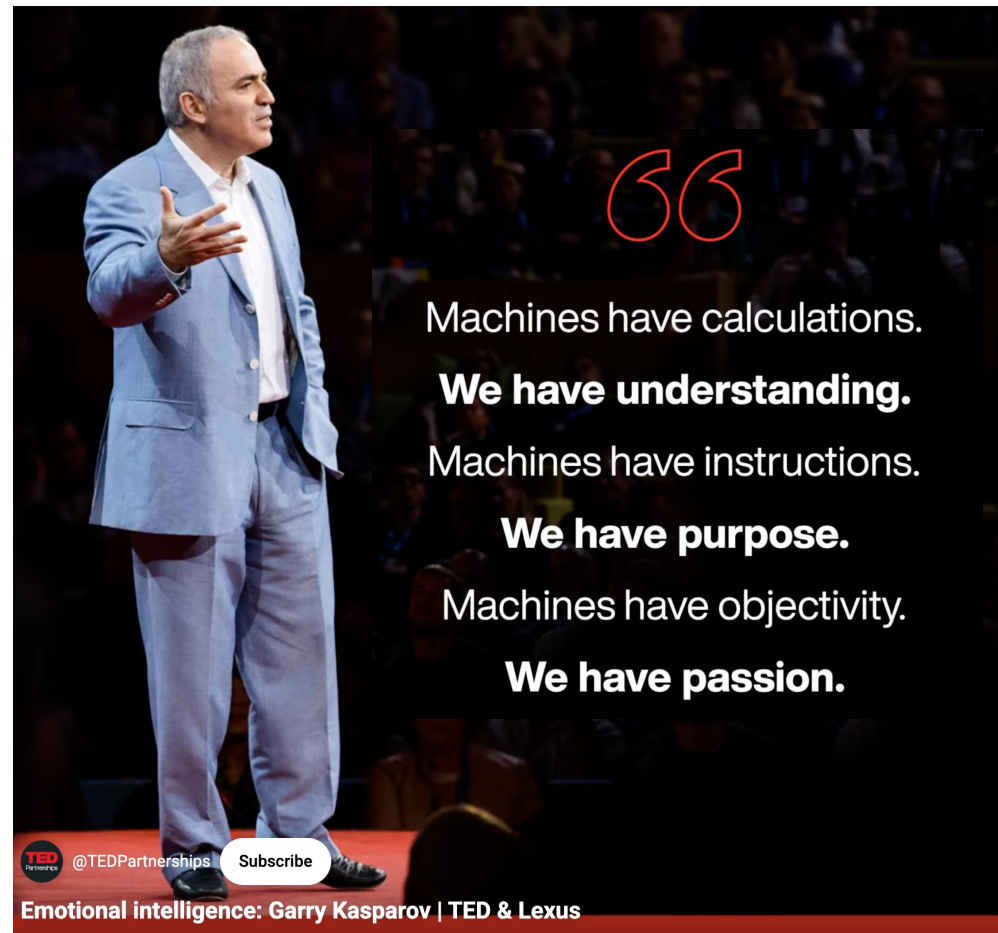


Results

II. Forward LLP reach estimate



Domain tools improve performance and reliability, even in frontier systems.



Kasparov's law: “A weaker player with an ordinary machine can beat a stronger player with a faster machine if they leverage synergy effectively. This principle extends beyond chess.”

AI in US HEP **theory** and **experiment** - I



Building an AI-native Research Ecosystem for Experimental Particle Physics: A Community Vision

February 20, 2026 - Version 1.00

U.S. DEPARTMENT OF ENERGY

About ▾ Our Mission ▾ Topics ▾ Funding Opportunities ▾ Consumer Savings ▾ Services & Opportunities ▾

Office of Science ▸ Department of Energy Seeks Input on Advancing AI for Science and Engineering Workforce Development and Genesis Mission Challenges

Department of Energy Seeks Input on Advancing AI for Science and Engineering Workforce Development and Genesis Mission Challenges



A National Research Collaboration in Particle Physics as a Strategic Platform for AI Workforce Development

*A response to the RFI on Mobilizing Talent for the Genesis Mission and Developing
an American Workforce to Advance Artificial Intelligence (AI) for Science and Engineering
Submitted on behalf of the authors of arXiv 2602.17582.*

AI-Accelerated Discoveries Through Particle Theory*

George T. Fleming¹, Daniel C. Hackett¹, James Halverson², Timothy Hobbs³, Stefan Hoeche¹,
Konstantin T. Matchev^{4†}, Stephen Mrenna¹, Benjamin Nachman⁵, Prasanth Shyamsundar¹,
David Shih⁶, Jesse Thaler⁷, Jure Zupan⁸

¹Fermi National Accelerator Laboratory, ²Northeastern University, ³Argonne National Laboratory,
⁴University of Alabama, ⁵SLAC National Laboratory, ⁶Rutgers University,
⁷Massachusetts Institute of Technology, ⁸University of Cincinnati

March 4, 2026

Extended version for the arXiv
coming soon...

GENESIS MISSION (PHASE 1 proposals)



2026 timeline

AI in US HEP **theory** and **experiment** - II

Blueprint workshop - Towards a National-Scale
AI Collaboration in HEP, May 18-19, DC

APS DPF Meeting,
Fermilab, July 20-24 2026

GENESIS MISSION: PHASE 1 awards announced

Agentic AI for BSM Event Generation
and Analysis Workshop,
Fermilab LPC, September 9-11 2026



ML4Jets Conference,
University of Vienna,
Vienna, September 14-18 2026

CLARIPHY
AI Collaboration Meeting,
Madison, September 14-16 2026

Action item: endorse the US AI Theory blueprint white paper

2026 timeline

Agentic AI for BSM Event Generation and Analysis Workshop at the LPC

Sep 9 – 11, 2026
Fermilab
America/Chicago timezone

Enter your search term



Overview

Timetable

Contribution List

My Conference

My Contributions

Registration

Participant List

Surveys

Mattermost chat

Fermilab Site Access

Arrival at Fermilab

Ride share/carpool

Accommodation/Transport

For more info

✉ mrenna@fnal.gov

✉ lpc-support@fnal.gov

☎ 2556

With the ubiquity of LLMs and agentic platforms, there is a pressing need to coordinate ongoing efforts across experiments and tools, and to build a shared community. This workshop focuses on the real-world problem of automating BSM event generation and experimental data analysis. Topics include:

- What harness-independent tools (usable across different agent frameworks) are needed to complete event generation and analysis tasks?
- How do we build a repository of validated tools?
- How do we build a knowledge base of user-agent interactions?
- What is the best feedback mechanism for improving agents using that knowledge base?

While this workshop is organized around CMS use cases, the underlying tools and challenges are largely experiment-agnostic – members of the broader experimental and theory community are invited.

This workshop aligns with Fermilab science mission and its leadership in the development and use of AI to accelerate HEP research and discovery.

We gratefully acknowledge University of Alabama for sponsoring this event.

Organizing Committee:

- Konstantin Matchev, University of Alabama
- Tony Menzo, University of Alabama/Fermilab
- Steve Mrenna, Fermilab

Fermilab LPC Events Committee:

- Gabriele Benelli, Brown University
- Murtaza Safdari, Fermilab

Fermilab LPC Coordinators:

- James Hirschauer, Fermilab
- Isobel Ojalvo, Princeton University

64 participants, mostly remote
<https://indico.fnal.gov/event/75639/>

Where to start

1. National Lab theory groups serve as an anchor: organize a meeting ASAP focused on the four assets.
2. A pilot artifact repo on the GMP, open to all registering theorists. Artifacts == human-agent interactions, skills, tools. *
3. Connect to the broader national effort through the GMP, using common formats.

Education, training & workforce development.
Important to get this right.^a

^aLet's stop comparing models to "undergrad", "first year graduate student", etc.

*This one first: it is the cheapest to build, and its value grows fastest with the number of groups contributing.

The bottom line

Theorists can keep what is valuable and unique in how they work and still share resources

- ☞ Organizing is how we **survive** the shift in funding, and how we **thrive** on what the tools now make possible.
- ☞ Event generator groups and USQCD show it already works.
- ☞ Theory needs a voice in a field-wide AI effort, and that requires a community.

Theory is just forming.
The best time to connect the two wings is now.

Blueprint White Paper

Open for endorsements

Blueprint for an AI-Enabled Particle Theory Community

George T. Fleming¹, Daniel C. Hackett¹, James Halverson², T. J. Hobbs³, Stefan Hoeche¹, Brandon Kriesten³, Konstantin T. Matchev⁴, Tony Menzo^{1,4}, Stephen Mrenna¹, Benjamin Nachman⁵, Prasanth Shyamsundar¹, David Shih⁶, Jesse Thaler⁷, Jure Zupan⁸

¹ Fermi National Accelerator Laboratory, ² Northeastern University, ³ Argonne National Laboratory,

⁴ University of Alabama, ⁵ SLAC National Laboratory, ⁶ Rutgers University,

⁷ Massachusetts Institute of Technology, ⁸ University of Cincinnati

September 14, 2026

ABSTRACT

Rapid advances in the capabilities and reach of artificial intelligence (AI) are changing research in particle physics. In this new era, AI is an innovative tool for theoretical physicists, while the interaction of AI and particle theory is itself a developing field. We argue that particle theory requires a community-scale effort to realize its full potential. We show how particle theory occupies a unique position in the AI for physics landscape, and describe the complementarity of AI and theory developments. We conclude with a series of recommendations on how to build an AI-enabled theory community for the Genesis Mission.

<http://smrenna.github.io/share/whitepaper.pdf>

<https://indico.fnal.gov/event/75639/surveys/>

AI-Accelerated Discoveries Through Particle Theory

Fleming, Hackett, Halverson, Hobbs, Hoeche, KM, Mrenna, Nachman, Shyamsundar, Shih, Thaler, Zupan 2026

- Science tasks for AI agents
 - Forward problem: theory building from first principles, predictions
 - Inverse problem: theory discovery from experimental data
- Roles for AI agents
 - Supervised research assistant: proofs, simplifications, optimizations, ...
 - Autonomous scientific assistant: calculations, simulations, data analysis
 - Autonomous scientist: hypothesis generation and testing, orchestration
- Application areas
 - Formal theory: hidden mathematical relations
 - Monte Carlo simulations and sampling
 - Hypothesis testing and inference
 - Symbolic regression
 - Quantum circuit optimization and algorithm construction
 - ...

Summary and Outlook

- **AI (LLMs)** are now ubiquitous in science and in everyday life.
 - There is still (lots of) room for narrow AI applications in HEP
- **Bold prediction:** all talks at ML4Jets 2028 will use agentic AI.
- **Where do we go next?** Three questions:
 - Can we teach AI routine research tasks? **Community tool repository?**
 - Can we teach AI high-level thinking? What is the  for HEP theory?
 - How do we integrate AI and quantum? **Knipfer,Roman,Gleyzer,KM,Matcheva (2026)**
- Mark your calendars: **ML4Jets 2028**

ML4Jets2028 @ The University of Alabama

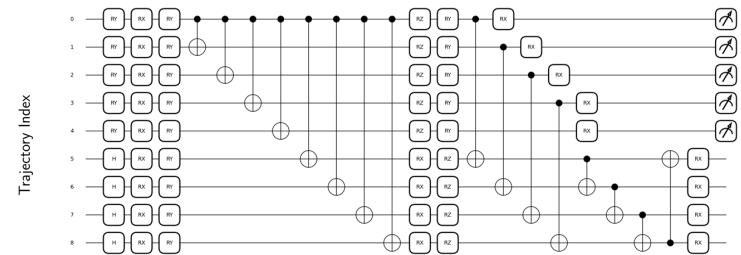
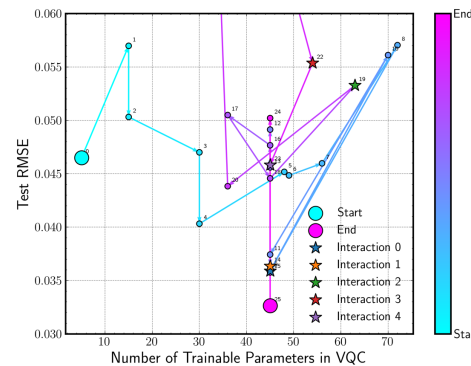
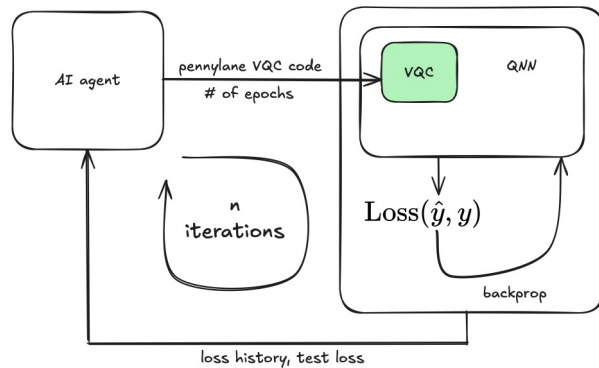


BACKUPS

Agentic AI for quantum algorithm discovery

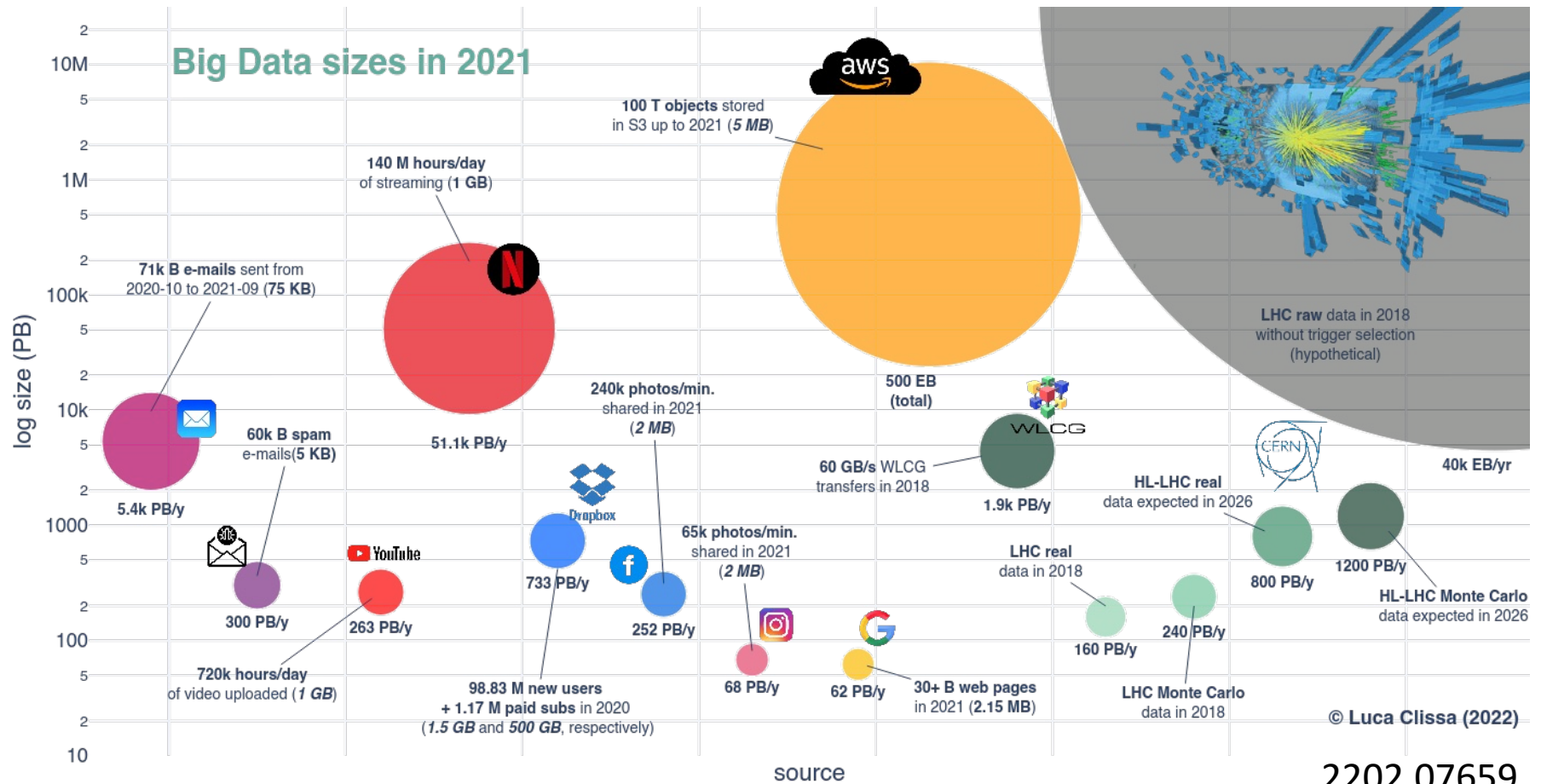
Knipfer, Roman, KM, Matcheva, Gleyzer 2026

- Let us set up a similar agentic framework to design, simulate, and evaluate variational quantum circuits
 - Tasks: Gaussian regression, EQGNN for quark-gluon tagging



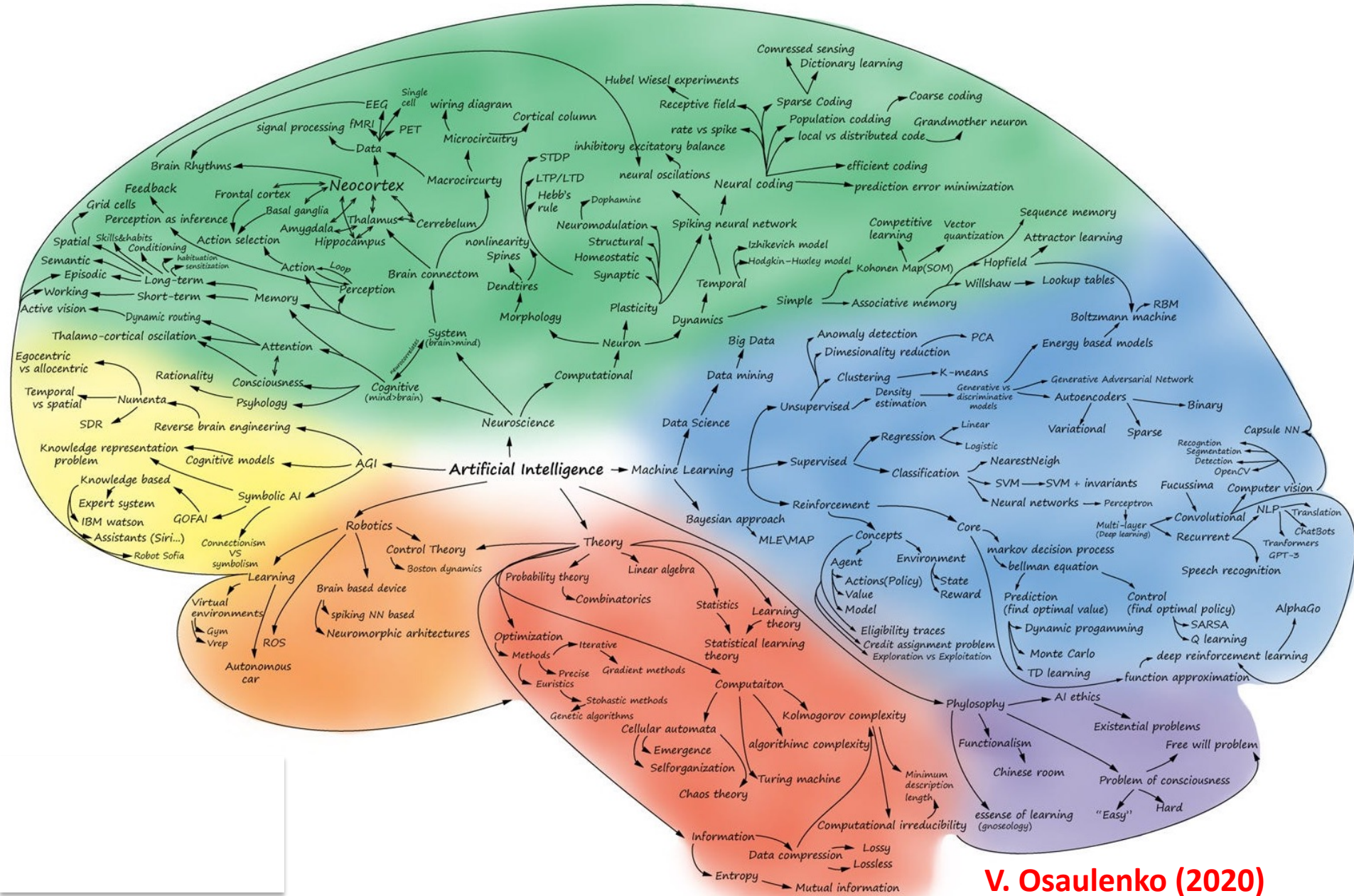
- Benefits of agentic AI for quantum algorithm discovery:
 - automation of the trial-and-error process of VQC design
 - exploration of unconventional architectures without human bias
 - ability to quickly iterate and refine designs based on performance feedback.

High Energy Physics has (Lots of) Data



2202.07659

The AI Map



V. Osaulenko (2020)