

# Nikhef AI group

Sascha Caron (Radboud University and Nikhef)

Wouter Verkerke (Nikhef and UvA)



# 2026:AI/Deep Learning in Fundamental Physics

AI – unfolding

AI – inference

AI – event selection

AI – reconstruction

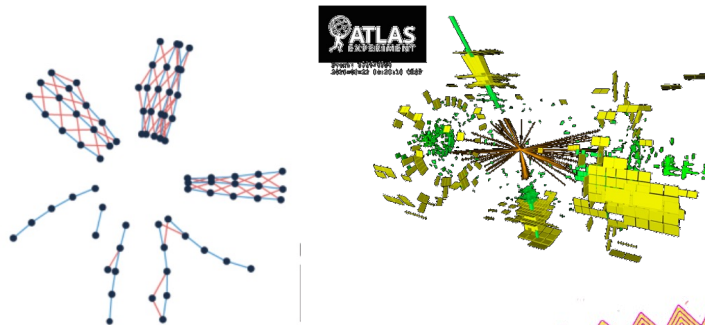
AI - tracking

AI – detector simulation

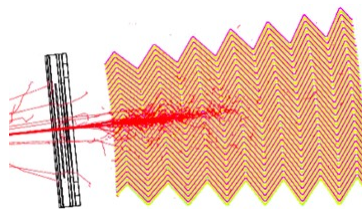
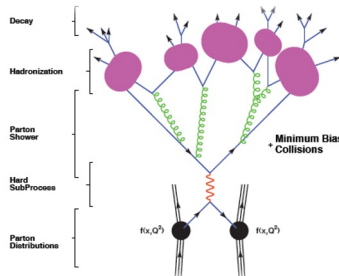
AI – event generation



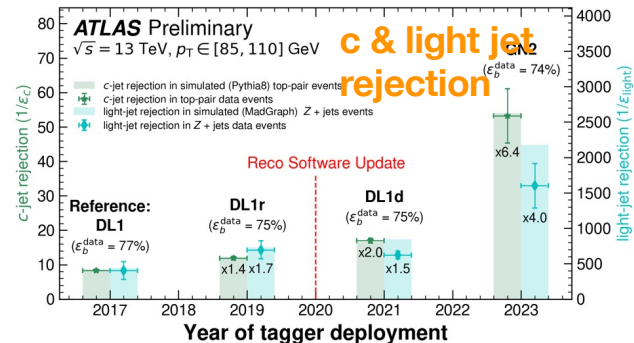
## Energy and angles of reconstructed particles



## Detector Simulator



## AI - trigger



# Goals of the AI of group

AI – unfolding

AI – inference

AI

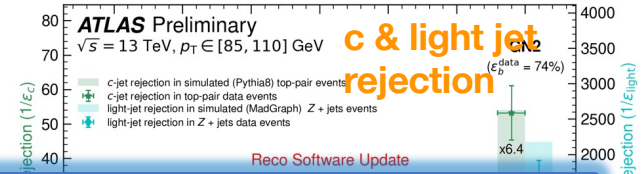
AI

AI

AI – detector simulation

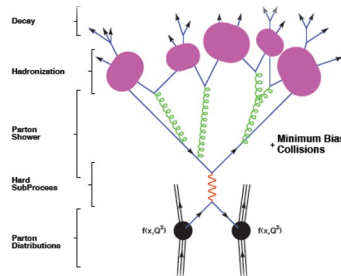
AI – event generation

Energy and angles of reconstructed particles



**2026: Start of the Nikhef AI/ML group, with 3 objectives:**

- 1) **Expertise on AI/ML methods** between programs, training, and joint R+D.
- 2) **Liaison function for the national landscape** and point of contact for AI in fundamental physics in the Netherlands
- 3) Platform for interaction and influence in the **European AI landscape**



AI - trigger

# AI group members



# AI group members



## 2026 Staffing plan

- **now:** 2 current Nikhef staff members (W. Verkerke, S. Caron) as **leadership**
- **spring:** recruitment of 2 **new staff positions**
  - dedicated staff position with focus on algorithm development for (A)PP
  - AI infrastructure staff position (joint with PDP group)
- **beyond:**
  - recruit **2 postdocs** dedicated to AI group research program
  - *Around the core members we aim for **associate members** from the physics programs at Nikhef (staff, PDs, PhD students)*

# Goals of the AI of group

## 1 - Connect and coordinate AI/ML activities of PP/APP scientists within the Nikhef portfolio

- Support Nikhef science programs **over the full range** (LHC, APP, GW, Theory)
- Maintain **in-house expertise** on 'AI for Science' techniques that are of importance for Nikhef science
- **Work with scientists** on physics groups to work on concrete AI issues in their physics application
- Define and organize **AI education/training programs** for PhD/PD as well as staff
- Pursue **own research program on AI techniques** with potential impact on PP/APP, in collaboration with experts in the field (national & international / physics & computer science & math)

✓ **Theme 1: How can we use AI's power for fundamental physics?**

✓ **Theme 2: How do we make AI scientifically trustworthy?**

# Group plans - before summer

- **Connect to the Nikhef programs**

- **Visit all programs** (e.g. in their weekly meeting) for a brief presentation followed by an **interactive discussion** on areas of interest in AI/ML

- **Define focus areas of AI/ML research & organization of the group**

- **Staffing:** Ongoing staff **recruitment** (closing date May 20<sup>th</sup>), upcoming Postdoc recruitment (summer)
- **Organization:** **Associate membership** & mailing lists for all interested, education & scientific events
- **Focus areas:** Based on input of groups **define a few key research themes** for core group

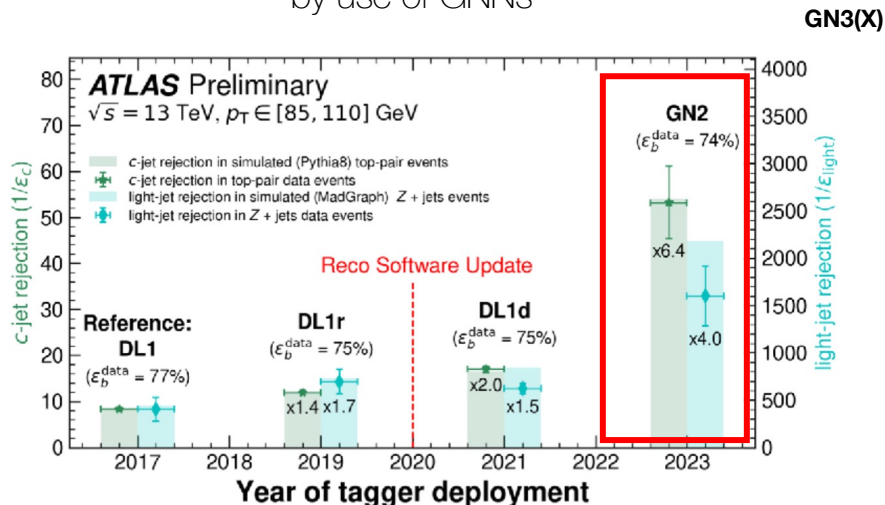
- **Submit full ENW-XL proposal on AI tracking**

- Collaborative effort between ATLAS, LHCb, Alice, PDP and AI group
- Goal – demonstrator of AI tracking framework for LHC experiments

# AI impact – already here

## Object ID with Graph Neural Nets

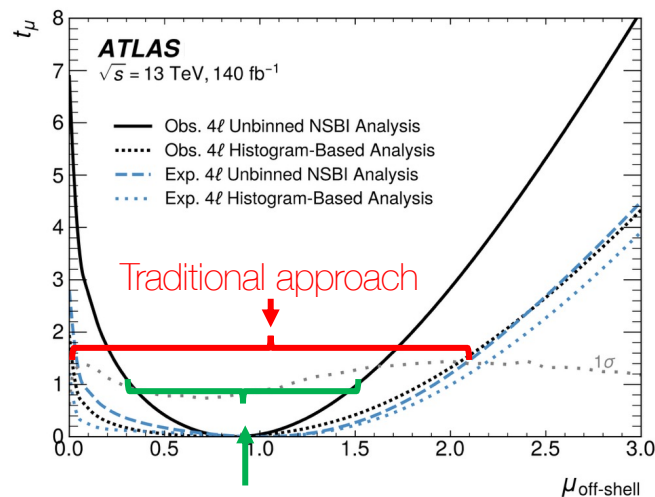
Identification of b-quark jets strongly enhanced by use of GNNs



Factor 6 reduction in background

## Statistical inference with Neural Simulation Based Inference

ATLAS Cross section  $H \rightarrow ZZ$  offshell ('Higgs lifetime')

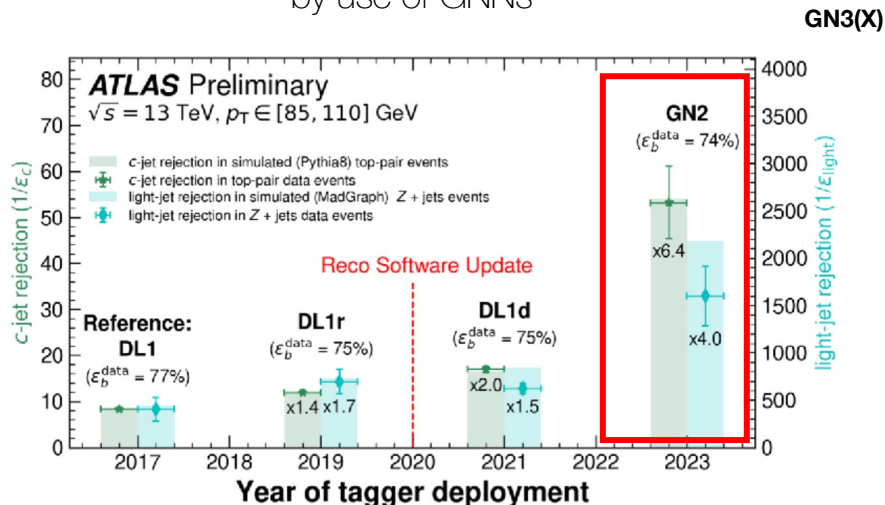


NSBI approach on the same dataset  
(30% smaller uncertainty)

# AI impact – already here

## Object ID with Graph Neural Nets

Identification of b-quark jets strongly enhanced by use of GNNs

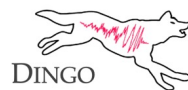


**Factor 6 reduction in background**

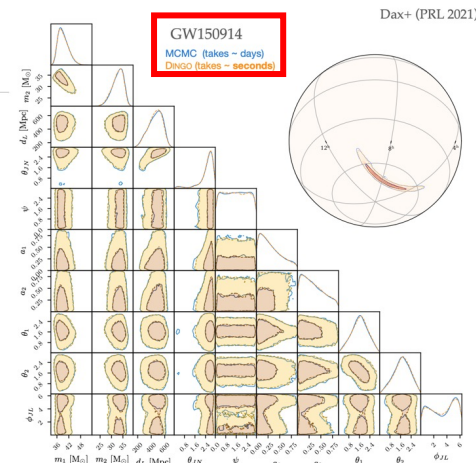
## Statistical inference with Neural Simulation Based Inference

**GW: Inference on binary black holes**

Binary black holes



- Inference in seconds to minutes using pre-trained networks (1000x speed up)
- Extremely good agreement with standard samplers
- Likelihood-free



Maximilian Dax

10

**Speedup From days (MCMC) to second (SBI based)**

# AI impact – data analysis pipeline

## End-to-end optimization: differentiable models of analysis chains

Traditional analysis pipelines consists of many stages, many of which are tuned independently (or not really at all)

What if you could simultaneously optimize your *entire analysis pipeline*?



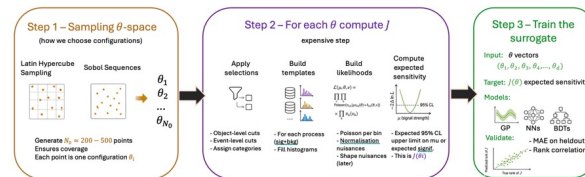
Potential applications in all particle and astro-particle physics domains

Part of ENW-XL diHiggs (figures by E. Shalugin)

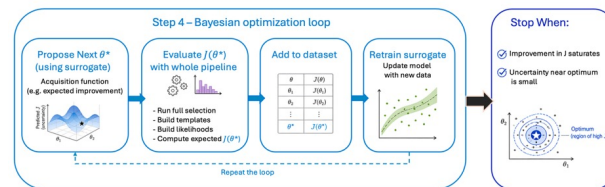
### 1) Identify full set of Tunable Parameters $\theta$



### 2) Use AI/ML to build surrogate model as function of $\theta$



### 3) Automated optimization of analysis w.r.t $\theta$

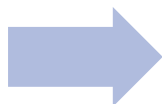
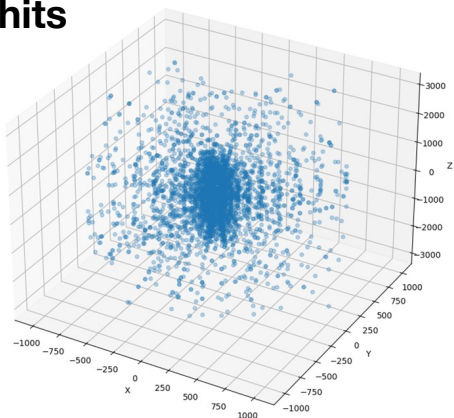


# AI impact – data analysis pipeline

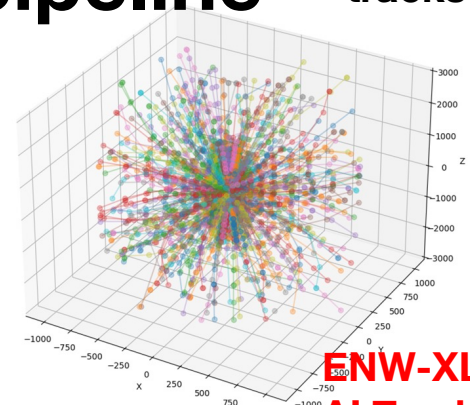
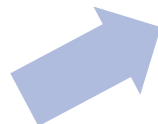
tracks

End-to-end optimization:  
direct inference from raw data

hits



Trackformer  
/  
Higgsformer

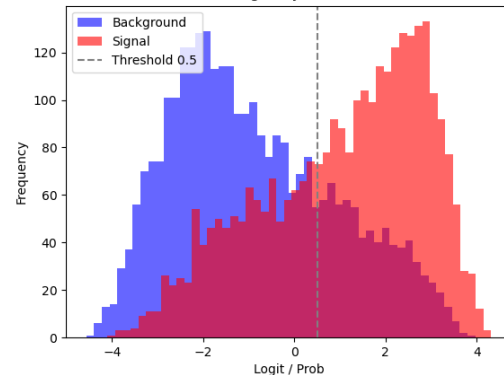


ENW-XL  
AI Tracking

See presentation by Eugene (next talk)

ENW-XL  
diHiggs

Higgs classification



# AI impact – experimental design

## differential models of entire detector systems

### On the Co-Design of Scientific Experiments and Industrial Systems

<https://arxiv.org/pdf/2603.26613>

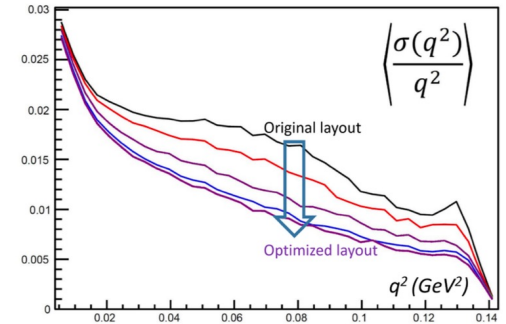
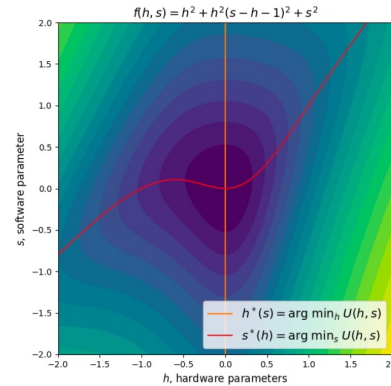
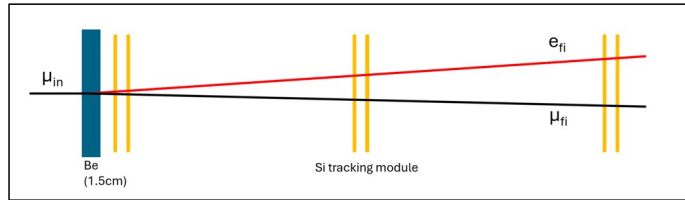


Figure 6: Progressive improvement of the resolution in muon-electron scattering  $q^2$  as a function of  $q^2$ , obtained by the staged optimization discussed in Ref. (77).

**Potential applications in all particle and astro-particle physics domains**

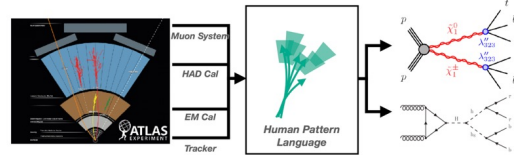
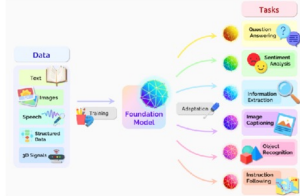
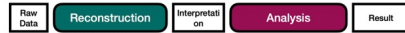
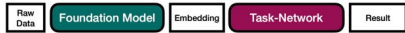
# AI impact – foundation models

differential models of entire detector systems

## Foundation Models

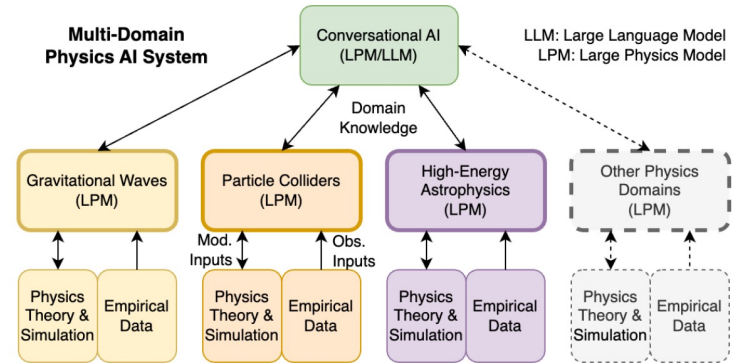
Train AI to learn its **own, multi-purpose, pattern vocabulary**

→ **Goal:** make it useful for not one, but many tasks



*Actually very similar to the human two-step strategy we know so well  
→ but combine with precision and autonomy of AI*

## Large Physics Models



Eur.Phys.J.C 85 (2025) 9, 1066  
(Barman, Caron, Sullivan, de Regt, et al)

Potential applications in all particle and astro-particle physics domains

# AI impact – coding assistance

Hands-on experience in Monday's session – by Niels & Samuel



W O R D L E

## Solver Showdown

*Build a bot. Beat the field. Pick your AI workflow.*

60 participants!

Winning solution:  
using Claude Code

# AI impact – analysis design & execution

## Designing & executing analysis code at PhD level

Goal: Let LLM model implement **ATLAS search for  $pp \rightarrow t\bar{t}t$  process**

### Instructions

**\*\* Instructions \*\*** You are an expert at programming in Python, machine learning, particle and high energy physics. You will help me answer a question in a machine learning challenge format where you strive to maximise a scalar metric in order to learn more about your scientific creativity and scientific understanding. You will follow all of the instructions to your best capabilities. Your first priority is to produce a correct solution in terms of runnable python code. Your second priority is to maximise the scoring metric defined below.

**\*\* Problem Description \*\*** A major task in particle physics is the measurement of rare signal processes with very small cross-sections. With the unprecedented amount of data provided by the upcoming runs of the Large Hadron Collider (LHC), one can start to measure these processes. An example is the recent observation of four top quarks originating from a single proton-proton collision event. Accurate classification of these events is crucial, as even a small reduction in background noise on the order of a few tens of percent while maintaining the same signal detection efficiency can lead to a profound increase in sensitivity.

**\*\* Evaluation Metric \*\*** The evaluation metric for this classification task is the area under the curve (AUC), specified by the area under the receiver operating characteristic (ROC) curve. The AUC summarizes a model's ability to distinguish between positive and negative classes. The higher the score the better.

**\*\* Dataset Description \*\*** The dataset used for this problem consists of simulated proton-proton collision at a center of mass energy of 13 TeV. The signal process is defined as  $pp \rightarrow t\bar{t}t$ . The relevant production processes of the backgrounds are  $t\bar{t} + X$  where  $X = Z, W+, W+W-$ . The dataset includes 302072 events, of which roughly 50% is signal and 50% are background processes. All background processes have an equal number of events. There is no cut on the maximum number of objects and there is no order

**\*\* IMPORTANT: Your Challenge \*\*** Write Python code for a binary classification model focusing on maximising the AUC using the code template above. You may freely choose any pre-processing methods and techniques as well as model architecture and training conventions. Do absolutely everything in your power to achieve the highest possible AUC.

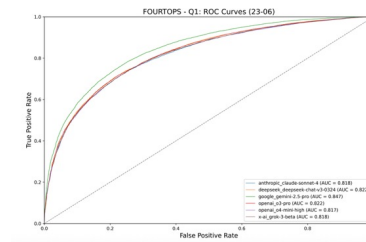
**\*\* Response Format \*\*** Your response must strictly be python code. If you must wrap it, put it in a "python fenced block and nothing else. Your response must follow these rules: 1. Do not add any formatting, such as markdown, to the response. 2. Replace each "# <LLM : ... >" comment, in the code template, with the required code. No placeholder should remain. 3. Before finalizing your answer, double-check that your code runs without errors and meets all requirements (all functions implemented, correct tensor shapes, etc.). 4. To prevent dimensional mismatches, make sure to annotate tensor sizes as comments. 5.

**IMPORTANT:** Remember, your first, and most important priority is to produce (syntactically) correct code. Prioritise what you can implement reliably above all else. Then prioritise maximising the metric



### Results separation of $pp \rightarrow t\bar{t}t$ vs $pp \rightarrow t\bar{t}W$

**Performance (AUC of ROC) already on par with specialized physics models!**



(a) LLM results (23-06)

LLM	AUC
ChatGPT 4o-mini-high	0.8175
ChatGPT o3 Pro	0.8221
Claude Sonnet 4.0	0.8179
Gemini 2.5 Pro	0.8469
X-AI Grok	0.8183
Deepseek Chat v3	0.8224

(b) Specialized physics models

Model Type	AUC
PN	0.8471(1)
PN <sub>int.SM</sub>	0.8725(0)
ParT	0.8404(0)
ParT <sub>int.SM</sub>	0.8732(0)

# AI impact – analysis design & execution

## Agentic AI - Complete analysis workflow execution

AI Agents Can Already Autonomously Perform  
Experimental High Energy Physics

Eric A. Moreno<sup>\*†1,2</sup>, Samuel Bright-Thonney<sup>\*†1,2</sup>, Andrzej Novak<sup>\*§1,2</sup>, Dolores Garcia<sup>#3</sup>,  
and Philip Harris<sup>‡1,2</sup>

<sup>1</sup>Department of Physics, Massachusetts Institute of Technology

<sup>2</sup>NSF AI Institute for Artificial Intelligence and Fundamental Interactions

<sup>3</sup>CERN

April 2, 2026

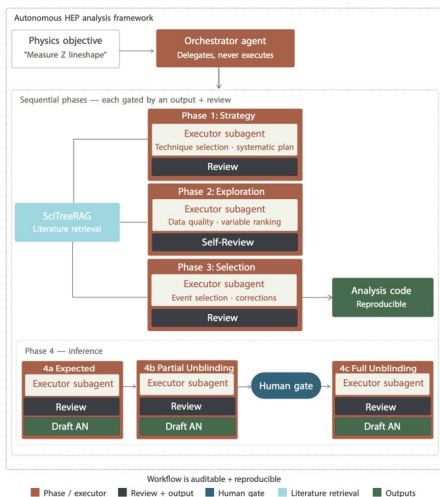
“Large language model-based AI agents are now able to **autonomously execute substantial portions of a high energy physics (HEP) analysis pipeline** with minimal expert-curated input. Given access to a HEP dataset, an execution framework, and a corpus of prior experimental literature, we find that Claude Code succeeds in **automating all stages of a typical analysis**: **event selection, background estimation, uncertainty quantification, statistical inference, and paper drafting.**”

<https://arxiv.org/abs/2603.20179>

# AI impact – analysis design & execution

## Example: CMS $h \rightarrow \tau\tau$ cross-section

Output: 73-page paper  
with 53 figures and 47 tables



## A $H \rightarrow \tau\tau$ Signal Strength Measurement with CMS Open Data

### A.1 Introduction

The discovery of the Higgs boson by the ATLAS and CMS collaborations in 2012 ([1]; [2]) opened a new era of precision measurements of the properties of this particle. Among the most fundamental properties to be established is the Yukawa coupling of the Higgs boson to fermions, which is a direct prediction of the Standard Model (SM) mechanism of electroweak symmetry breaking. The  $H \rightarrow \tau\tau$  decay channel provides the most sensitive probe of the Higgs boson coupling to leptons at the Large Hadron Collider (LHC), as the tau lepton is the heaviest lepton and this has the largest Yukawa coupling among the leptons.

### A.13 Conclusions

This analysis presents a measurement of the Higgs boson signal strength modifier  $\mu$  in the  $H \rightarrow \tau\tau$  decay channel, performed in the  $\mu\tau_\tau$  final state using  $11.5 \text{ fb}^{-1}$  of CMS Open Data at  $\sqrt{s} = 8 \text{ TeV}$ . Three fitting approaches are compared using a simultaneous binned profile likelihood fit across Baseline and VBF event categories.

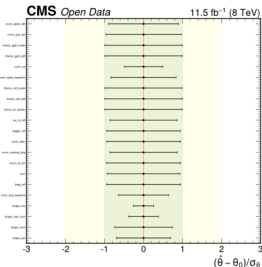
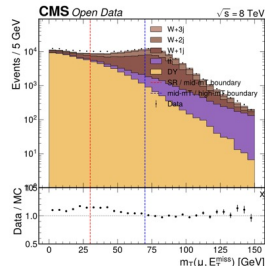
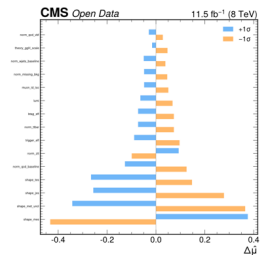
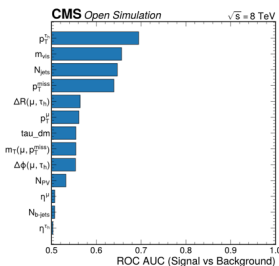
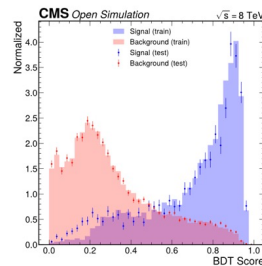
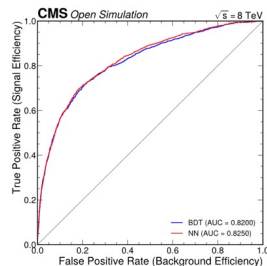
The primary result, using the neural network discriminant on the full dataset, is:

$$\mu(H \rightarrow \tau\tau) = 0.63 \pm 1.08$$

### A.15 Future Directions

Several concrete improvements could enhance this measurement:

- Higher-order MC samples.** NLO-matched DY and  $W$ +jets MC (e.g., from MadGraph5\_aMC@NLO) would improve the VBF category modeling and likely resolve the 31% normalization deficit. This is the single most impactful improvement.
- Pileup reweighting.** Implementing PV-based pileup reweighting using the data pileup profile would reduce the pileup-related modeling uncertainty and potentially improve the VBF category agreement.
- Per-decay-mode tau energy scale.** Evaluating TES variations separately for 1-prong ( $\pm 1\%$ ), 1-prong+ $\pi^0$  ( $\pm 2\%$ ), and 3-prong ( $\pm 3\%$ ) decay modes would provide a more precise systematic evaluation.
- Additional final states.** Including the  $e\tau_\tau$ ,  $e\mu$ , and  $\tau_\tau\tau_\tau$  final states from CMS Open Data would improve the expected sensitivity by a factor of approximately  $1/\sqrt{0.23} \approx 2.1$ , bringing the expected uncertainty to  $\sigma(\mu) \approx 0.5$ .
- SVfit mass reconstruction.** Implementing the SVfit algorithm ([3]) would provide 15-20% better mass resolution for the visible mass and collinear mass approaches, potentially making them competitive with the NN discriminant.



all 6h30 using Claude Opus!

But do we really want to give everything to Claude & pay for the privilege?

LHC

DM

Neutrinos

GW

Cosmic

Theory

R&D

PDP



Nikhef

## 2026: Start of the Nikhef AI/ML group, with 3 objectives:

- 1) **Expertise on AI/ML methods** between programs, training, and joint R+D.
- 2) **Liaison function for the national landscape**  
and point of contact for AI in fundamental physics in the Netherlands
- 3) Platform for interaction and influence in the **European AI landscape**



# Agenda for the AI session

## **Introduction & Overview – 10 minutes**

Wouter Verkerke



## **End-to-end optimization – 12 minutes**

Eugene Shalugin

