

JetGPT — Wie versteht man den perfekten Datensatz?

Tilman Plehn

Universität Heidelberg

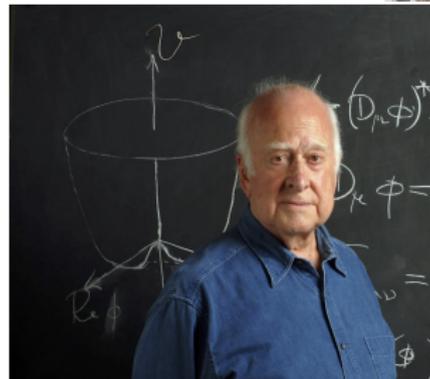
70 Jahre CERN



KI in Teilchenphysik

Suchen

- dunkle Materie
- Materie vs Antimaterie
- Ursprung des Higgs-Bosons



KI in Teilchenphysik

Suchen

- dunkle Materie
- Materie vs Antimaterie
- Ursprung des Higgs-Bosons

LHC-Forschung

- riesige Datensätze
 - kontrollierte Unsicherheiten
 - präzise Simulationen
- [Fall für KI?](#)



KI in Teilchenphysik

Suchen

- dunkle Materie
- Materie vs Antimaterie
- Ursprung des Higgs-Bosons

LHC-Forschung

- riesige Datensätze
 - kontrollierte Unsicherheiten
 - präzise Simulationen
- [Fall für KI?](#)

Früher

- Teilchen-Zählraten
 - Theorie-Motivation [Higgs-Entdeckung]
 - Suche nach Entdeckungen
- [Fall für KI?](#)



KI in Teilchenphysik

Suchen

- dunkle Materie
- Materie vs Antimaterie
- Ursprung des Higgs-Bosons

LHC-Forschung

- riesige Datensätze
- kontrollierte Unsicherheiten
- präzise Simulationen

→ Fall für KI?

Früher

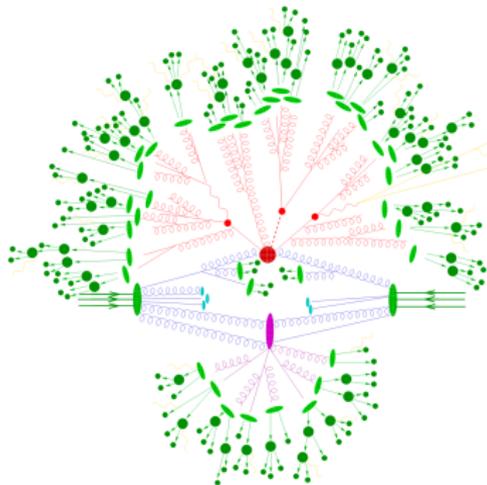
- Teilchen-Zählraten
- Theorie-Motivation [Higgs-Entdeckung]
- Suche nach Entdeckungen

→ Fall für KI?

Virtuellen Welten

- Ausgangspunkt: Elementarteilchen
- Rechnungen: Quantenfeldtheorie
- Simulation: Kollisionen
- Simulation: Detektoren

→ Fall für KI!



KI in Teilchenphysik

Suchen

- dunkle Materie
- Materie vs Antimaterie
- Ursprung des Higgs-Bosons

LHC-Forschung

- riesige Datensätze
 - kontrollierte Unsicherheiten
 - präzise Simulationen
- Fall für KI?

Früher

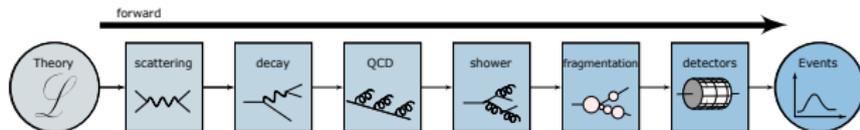
- Teilchen-Zählraten
 - Theorie-Motivation [Higgs-Entdeckung]
 - Suche nach Entdeckungen
- Fall für KI?

Virtuellen Welten

- Ausgangspunkt: Elementarteilchen
 - Rechnungen: Quantenfeldtheorie
 - Simulation: Kollisionen
 - Simulation: Detektoren
- Fall für KI!

Inferenz

- Vergleich Simulationen und Daten
 - Beschreibung aller LHC-Daten
- korrekte Theorie
- Fall für KI!



LHC-Community

- High-luminosity LHC: 2029-2040
- ATLAS & CMS bigger is better
LHCb, ALICE, FASER... spezialisiert
- Top-Unis **weltweit**

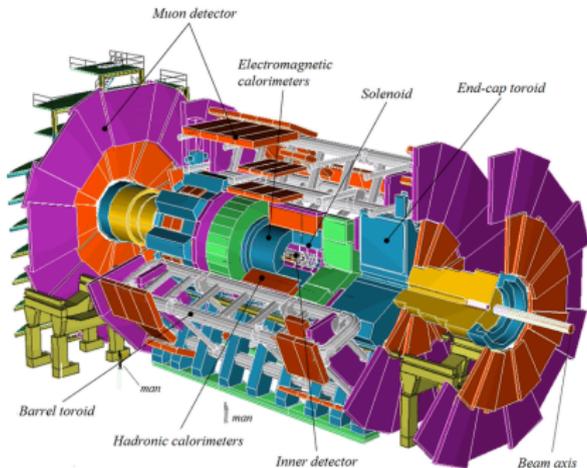


LHC-Community

- High-luminosity LHC: 2029-2040
- ATLAS & CMS bigger is better
LHCb, ALICE, FASER... spezialisiert
- Top-Unis **weltweit**

LHC-Detektoren

- Messung aller produzierter Teilchen
- Riesig und komplex [ohne Theorie]



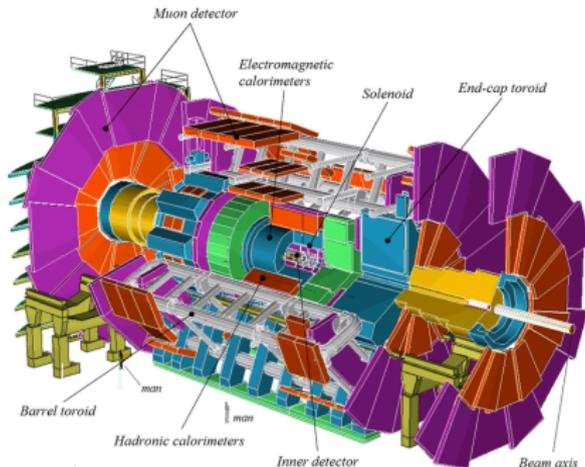
CERN und LHC

LHC-Community

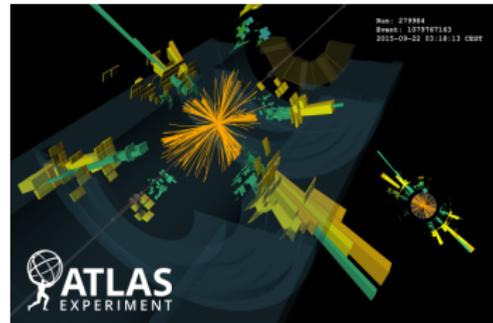
- High-luminosity LHC: 2029-2040
- ATLAS & CMS bigger is better
LHCb, ALICE, FASER... spezialisiert
- Top-Unis **weltweit**

LHC-Detektoren

- Messung aller produzierter Teilchen
- **Riesig und komplex** [ohne Theorie]



LHC-Daten



- Proton-Stöße mit 40 MHz
- Produktion von (Anti)-Teilchen
dann Teilchen-Zerfälle
- Messungen: Energie, Ladung,...
- Elektronen, Muonen einfach
Quarks, Gluons herausfordernd
- Event: 100+ Vektoren (E, \vec{p}, Q)

→ **Daten-Rate 3 PB/s**



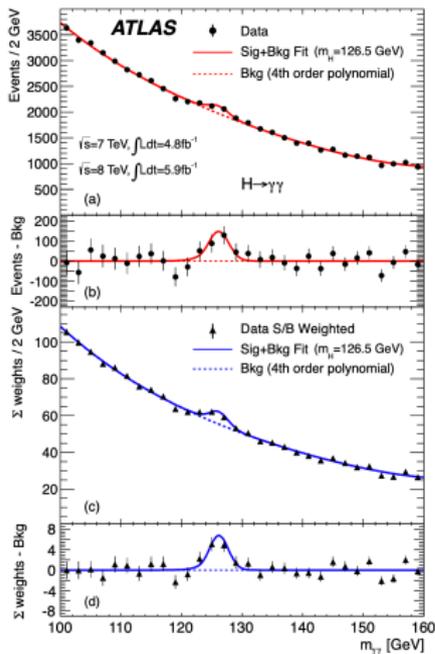
Entdeckungen und KI

LHC-Entdeckungen

- Higgs-Entdeckung, 4. Juli 2012
Nobel-Preis 2013



CERN-PH-EP-2012-218
Accepted by: Physics Letters B



√2 [hep-ex] 31 Aug 2012

Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC

The ATLAS Collaboration

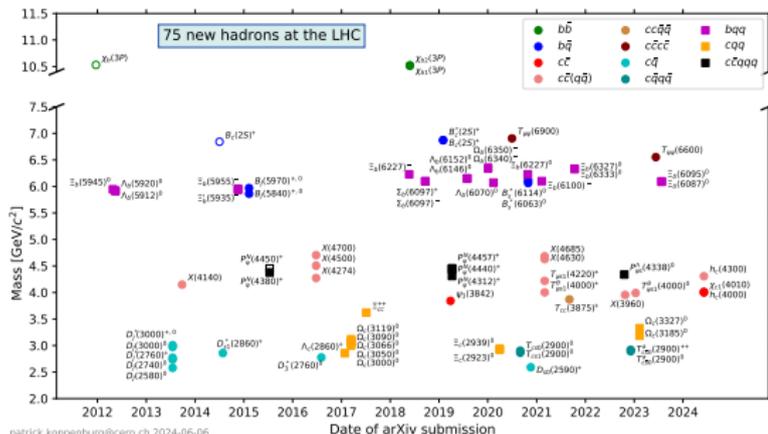
This paper is dedicated to the memory of our ATLAS colleagues who did not live to see the full impact and significance of their contributions to the experiment.



Entdeckungen und KI

LHC-Entdeckungen

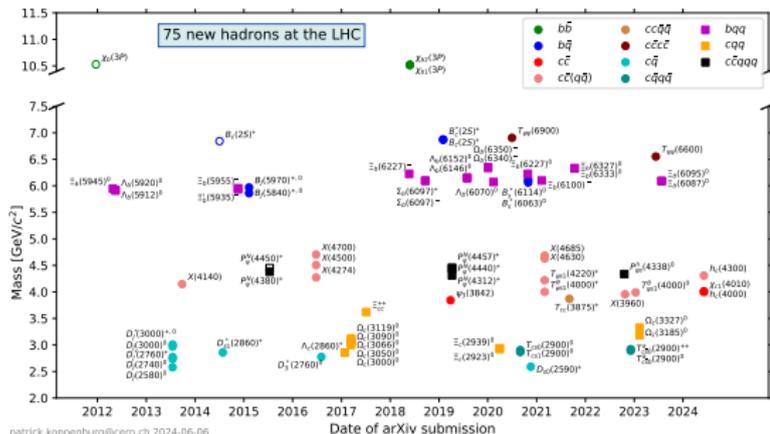
- Higgs-Entdeckung, 4. Juli 2012
Nobel-Preis 2013
- seither: 75 entdeckte Teilchen
- Teilchen \leftrightarrow Elementarteilchen?
wie Proton vs Elektron



Entdeckungen und KI

LHC-Entdeckungen

- Higgs-Entdeckung, 4. Juli 2012
Nobel-Preis 2013
- seither: 75 entdeckte Teilchen
- Teilchen ↔ Elementarteilchen?
wie Proton vs Elektron



→ verpasste Entdeckungen?
Garantie, in PB/s nichts zu übersehen?

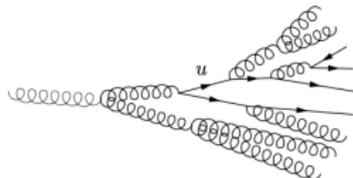
→ Vertrauenswürdige KI



LHC und Jets

Komplexität der LHC-Daten

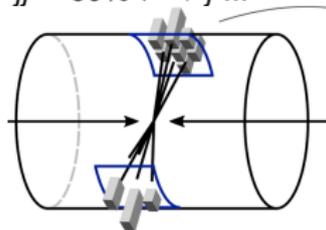
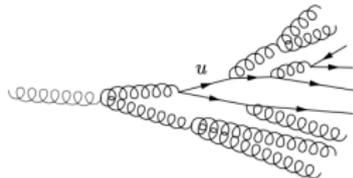
- LHC-Produktion vor allem $q\bar{q}, gg \rightarrow q\bar{q}, gg$
 - Quark-Gluon-Splittings aus Quantenfeldtheorie
Hadronisierung und Hadron-Zerfälle
beobachtet als sogenannte **Jets**
 - Jets als Zerfallsprodukte
67% $t \rightarrow jjj$ 60% $H \rightarrow jj$ 70% $Z \rightarrow jj$ 67% $W \rightarrow jj$ 60% $\tau \rightarrow j \dots$
 - **Dark-Jets** mit neuen Teilchen
- **LHC heißt Präzisions-Jetphysik**



LHC und Jets

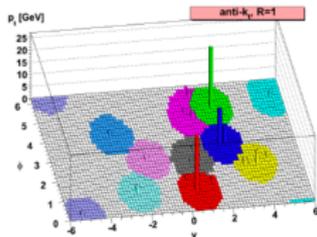
Komplexität der LHC-Daten

- LHC-Produktion vor allem $q\bar{q}, gg \rightarrow q\bar{q}, gg$
 - Quark-Gluon-Splittings aus Quantenfeldtheorie
Hadronisierung und Hadron-Zerfälle
beobachtet als sogenannte **Jets**
 - Jets als Zerfallsprodukte
67% $t \rightarrow jjj$ 60% $H \rightarrow jj$ 70% $Z \rightarrow jj$ 67% $W \rightarrow jj$ 60% $\tau \rightarrow j \dots$
 - **Dark-Jets** mit neuen Teilchen
- **LHC heißt Präzisions-Jetphysik**



Jet-Verständnis

- 50-200 Konstituenten pro Jet
Dutzende simultaner Kollisionen
 - Cluster-Algorithmen für Jet-Energie
 - Klassifikations-Algorithmen für Jet-Inhalt
- **Basis und Grenze jeder LHC-Analyse**



November-Revolution

Stanford, 16.11.2015

- Jet-Signale sind Bilder
- Jet-Analyse heißt Bild-Analyse



Jet-Images – Deep Learning Edition

Loic de Oliveira,^a Michael Kagan,^b Lester Mackey,^c Benjamin Nachman,^b and Ariel Schwartzman^d

^a*Institute for Computational and Mathematical Engineering, Stanford University, Stanford, CA 94305, USA*
^b*SLAC National Accelerator Laboratory, Stanford University, 2575 Sand Hill Rd, Menlo Park, CA 94025, U.S.A.*

^c*Department of Statistics, Stanford University, Stanford, CA 94305, USA*

E-mail: lschwartz@stanford.edu, mkagan@cern.ch, lmackey@stanford.edu, bnachman@cern.ch, schwartz@stanford.edu

ABSTRACT: Building on the notion of a particle physics detector as a camera and the collimated streams of high energy particles, or jets, it measures as an image, we investigate the potential of machine learning techniques based on deep learning architectures to identify highly boosted W bosons. Modern deep learning algorithms trained on jet images can out-perform standard physically-motivated feature driven approaches to jet tagging. We develop techniques for visualizing how these features are learned by the network and what additional information is used to improve performance. This interplay between physically-motivated feature driven tools and supervised learning algorithms is general and can be used to significantly increase the sensitivity to discover new particles and new forces, and gain a deeper understanding of the physics within jets.

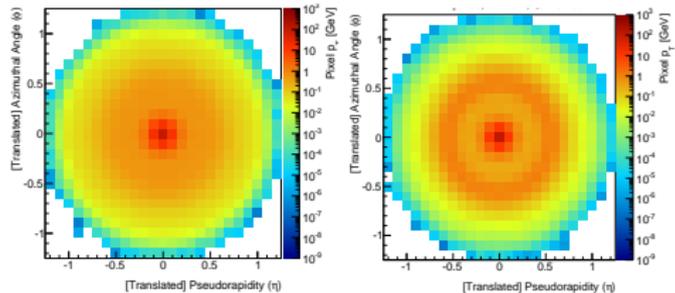
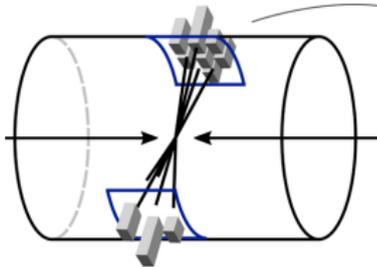
w1 [hep-ph] 16 Nov 2015



November-Revolution

Stanford, 16.11.2015

- Jet-Signale sind Bilder
 - Jet-Analyse heißt Bild-Analyse
- Revolution?



November-Revolution

Stanford, 16.11.2015

- Jet-Signale sind Bilder
 - Jet-Analyse heißt Bild-Analyse
- Revolution!



Einmal angeben...

- Boston 2016
Heidelberg 2017
 - Wissenschaft:
überzeugende Anwendung
Vergleich mit Standard-Methoden
Ohne Einstein geht nichts...
- Überzeugende KI



Hello World in LHC-KI

Jet-Klassifikation (Tagging)

- 2017: Welche Netzwerk-Architektur?
- 2018: Bild, Text, Physik alle funktioniert

SciPost Physics

Submission

The Machine Learning Landscape of Top Taggers

G. Kasieczka (ed)¹, T. Plehn (ed)², A. Butter², K. Cranmer³, D. Debnath⁴, M. Fairbairn⁵, W. Fedorilo⁶, C. Gay⁶, L. Gouskos⁷, P. T. Komiske⁸, S. Leiss¹, A. Lister⁶, S. Macaluso^{3,4}, E. M. Metodiev⁸, L. Moore⁹, B. Nachman^{10,11}, K. Nordström^{12,13}, J. Pearce⁶, H. Qu⁷, Y. Rath¹⁴, M. Rieger¹⁴, D. Shih⁴, J. M. Thompson², and S. Varma⁵

¹ Institut für Experimentalphysik, Universität Hamburg, Germany

² Institut für Theoretische Physik, Universität Heidelberg, Germany

³ Center for Cosmology and Particle Physics and Center for Data Science, NYU, USA

⁴ NHECT, Dept. of Physics and Astronomy, Rutgers, The State University of NJ, USA

⁵ Theoretical Particle Physics and Cosmology, King's College London, United Kingdom

⁶ Department of Physics and Astronomy, The University of British Columbia, Canada

⁷ Department of Physics, University of California, Santa Barbara, USA

⁸ Center for Theoretical Physics, MIT, Cambridge, USA

⁹ CP3, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

¹⁰ Physics Division, Lawrence Berkeley National Laboratory, Berkeley, USA

¹¹ Simons Inst. for the Theory of Computing, University of California, Berkeley, USA

¹² National Institute for Subatomic Physics (NIKHEF), Amsterdam, Netherlands

¹³ LP'THE, CNRS & Sorbonne Université, Paris, France

¹⁴ III. Physics Institute A, RWTH Aachen University, Germany

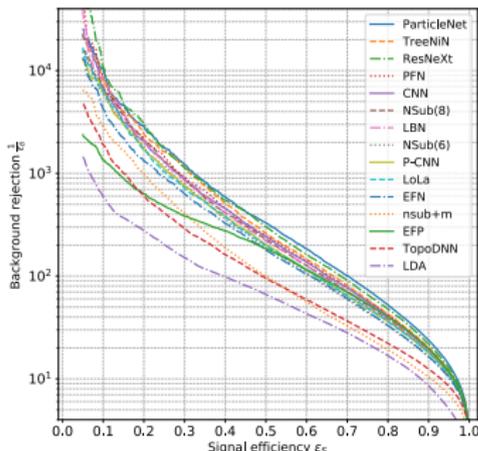
gregor.kasieczka@uni-hamburg.de

plehn@uni-heidelberg.de

April 12, 2019

Abstract

Based on the established task of identifying boosted, hadronically decaying top quarks, we compare a wide range of modern machine learning approaches. We find that they are extremely powerful and great fun.

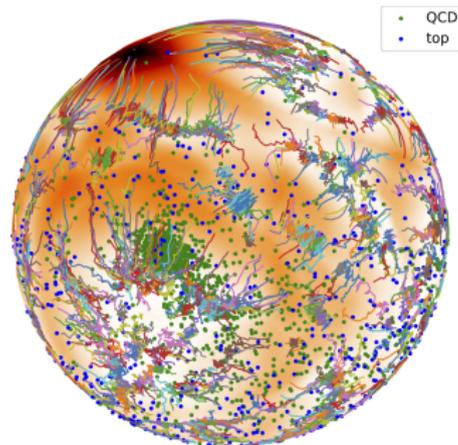
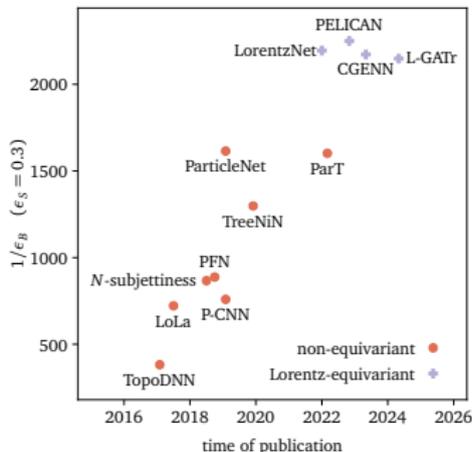


Hello World in LHC-KI

Jet-Klassifikation (Tagging)

- 2017: Welche Netzwerk-Architektur?
- 2018: Bild, Text, Physik alle funktioniert
- 2020: ML-Klassifikation etabliert
- 2024: latente Jet-Darstellung mit Symmetrien

→ [Erklärbare AI](#)



Detektor-Kalibration

Neu: Kalibration mit Unsicherheiten

- Jet mit E^{wahr} gemessen als E^{mess}
- KI-Korrektur aus Simulationen

$$\mathcal{R}^{\text{BNN}}(x) = \mathcal{R}(x) = \frac{E^{\text{wahr}}(x)}{E^{\text{mess}}(x)}$$

- Physik heißt Fehlerbalken: $\mathcal{R}(x) \pm \Delta\mathcal{R}(x)$



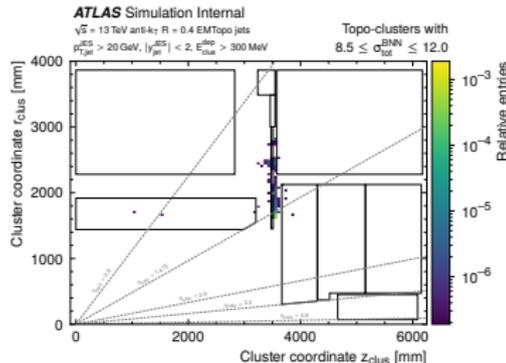
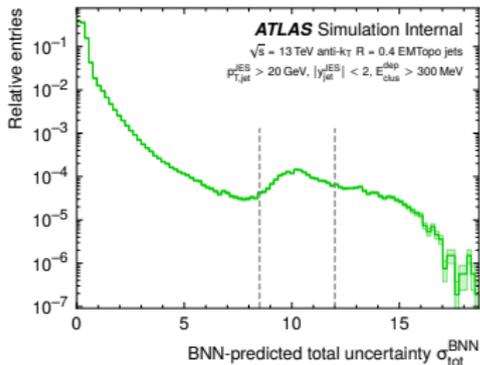
Detektor-Kalibration

Neu: Kalibration mit Unsicherheiten

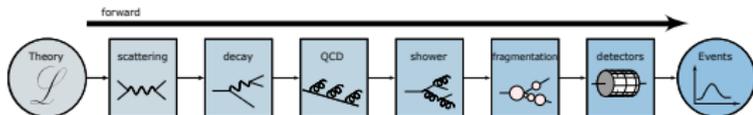
- Jet mit E^{wahr} gemessen als E^{mess}
- KI-Korrektur aus Simulationen

$$\mathcal{R}^{\text{BNN}}(x) = \mathcal{R}(x) = \frac{E^{\text{wahr}}(x)}{E^{\text{mess}}(x)}$$

- Physik heißt Fehlerbalken: $\mathcal{R}(x) \pm \Delta\mathcal{R}(x)$
- Fehlerbalken bedeutet Verständnis



Generative KI

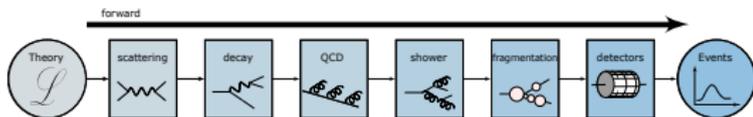


Theorie-Simulationen

- Startpunkt:
Elementarteilchen und Wechselwirkungen
 - Quantenfeldtheorie:
Produktion & Zerfall
und immer wieder Jets
 - überwiegend Theorie:
Hadronisierung und Hadron-Zerfälle
 - Detektor-Simulationen [Experimentalphysik!?!]
- Präzisions-Theorie-Simulation



Generative KI



Theorie-Simulationen

- Startpunkt:
Elementarteilchen und Wechselwirkungen
 - Quantenfeldtheorie:
Produktion & Zerfall
und immer wieder Jets
 - überwiegend Theorie:
Hadronisierung und Hadron-Zerfälle
 - Detektor-Simulationen [Experimentalphysik!?!]
- Präzisions-Theorie-Simulation

Beschleunigter KI-Fortschritt [Bilder & Text]

- Generative Adversarial Networks [2019]
- Normalizing Flow [2020]
- Diffusion [2023]
- Diffusion mit Attention [2023]
- Autoregressive Transformer [2023]
- symmetrische Diffusion [2024]
- Niemals ohne Fehlerbalken!

SciPost Physics

Submission

Jet Diffusion versus JetGPT — Modern Networks for the LHC

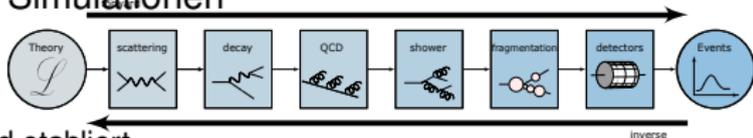
Anja Butter^{1,2}, Nathan Huetsch¹, Sofia Palacios Schweitzer¹,
Tilman Plehn¹, Peter Sorrenson³, and Jonas Spinner¹

¹ Institute for Theoretical Physics, Universität Heidelberg, Germany
² LPNHE, Sorbonne Université, Université Paris Cité, CNRS/IN2P3, Paris, France
³ Heidelberg Collaboratory for Image Processing, Universität Heidelberg, Germany



Transformation: Inverse Simulationen

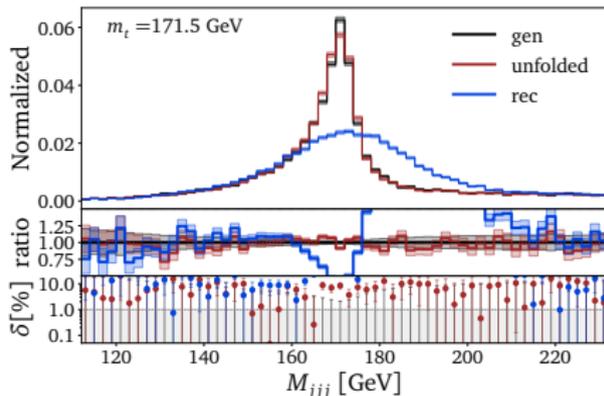
Zahl von LHC-Analysen



- Untergrund bekannt und etabliert
 - Potenzielle Signale im Fluss
- **Test ohne Detektor-Effekte**

Öffentliche LHC-Daten

- Standard-Argument:
LHC-Daten zu kompliziert für Amateure
 - Wirklichkeit:
Theorie-Berechnungen öffentlich
- **Analysen auf Amateur-Level**



Teilchenphysik & KI

LHC perfekter KI-Spielplatz

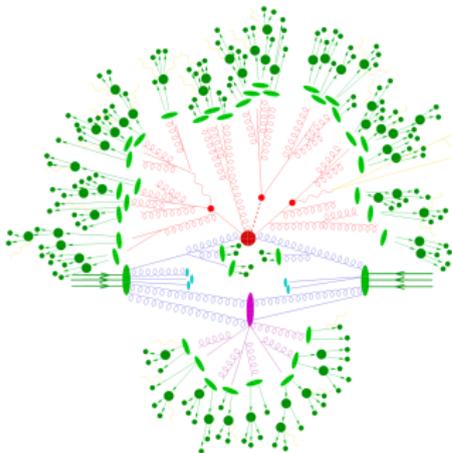
- **Vertrauenswürdige KI** für quantitative Analysen
- **Überzeugende KI** für komplexe Probleme
- **Erklärbare KI** für neue Einsichten
- **Transformative KI** für echten Fortschritt
- Gesellschaft schwierig? Lieber Hunderte Profs?



Teilchenphysik & KI

LHC perfekter KI-Spielplatz

- **Vertrauenswürdige KI** für quantitative Analysen
 - **Überzeugende KI** für komplexe Probleme
 - **Erklärbare KI** für neue Einsichten
 - **Transformative KI** für echten Fortschritt
 - Gesellschaft schwierig? Lieber Hunderte Profs?
- Relevanz: Von LHC zu Covid



JUNE: open-source individual-based epidemiology simulation

Joseph Bullock^{1,2*}, Carolina Cuesta-Lazaro^{1,3,4}, Arnau Quer-Bofarril^{1,3,4}, Miguel Icaza-Lizola^{1,5,6*}, Aidan Sedgwick^{1,7,8}, Henry Truong^{1,9}, Aisde Curran^{1,3}, Edward Elliott^{1,3}, Tristan Casfield¹, Kevin Fong^{6,8}, Ian Vernon^{1,6}, Julian Williams², Richard Bower^{1,3}, and Frank Krauss^{1,2*}

¹Institute for Data Science, Durham University, Durham DH1 1TA, UK

²Institute for Particle Physics Phenomenology, Durham University, Durham DH1 1TA, UK

³Institute for Computational Cosmology, Durham University, Durham DH1 1TA, UK

⁴Centre for Extragalactic Astronomy, Durham University, Durham DH1 1TA, UK

⁵Institute for Health, Risk & Behaviour, Durham University, Durham DH1 1TA, UK

⁶Department of Mathematical Sciences, Durham University, Durham DH1 1TA, UK

⁷Department of Computer Science, University College London, London WC1E 6BT, UK

⁸Department of Science, Technology, Engineering and Public Policy, University College London, London WC1E 6BT, UK

⁹Department of Anesthesia, University College London Hospital, London NW1 2BU, UK

*Equal contribution

**Signal contribution

†Corresponding author: frank.krauss@durham.ac.uk

Abstract: We introduce JUNE, an open-source framework for the detailed simulation of epidemics on the basis of social interactions in a virtual population constructed from geographically granular census data, reflecting age, sex, ethnicity, and socio-economic indicators. Interactions between individuals are modelled in groups of various sizes and properties, such as households, schools and workplaces, and other social activities using social mixing matrices. JUNE provides a suite of flexible parameterisations that describe infectious diseases, how they are transmitted and affect contaminated individuals. In this paper we apply JUNE to the specific case of modelling the spread of COVID-19 in England. We discuss the quality of initial model outputs which reproduce reported hospital admission and mortality statistics at national and regional levels as well as by age strata.

