Big Data Science in Astroparticle Research: sharing and exploiting data & knowledge

Martin Erdmann & Andreas Haungs

21-Feb-2018

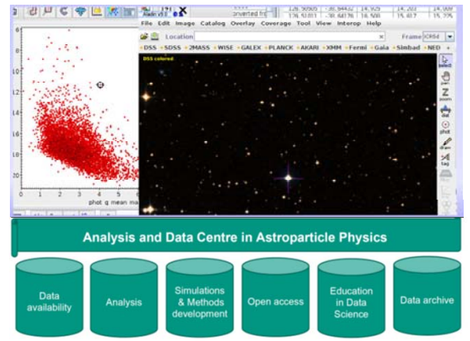
Today: 2nd Workshop 19.-21. Feb-2018

Deep Learning

- Supervised & unsupervised machine learning
- Causality & stability of networks
- Network training with simulations ≠ data

Tutorial: 14:15 basics & convolutional, 16:15 adversarial

Open Data



- National & international initiatives:
  research data alliance, data centers,
  application for funds
- Perspective United Kingdom

Evaluation: Tuesday 16:00

Discussion on future: Wednesday

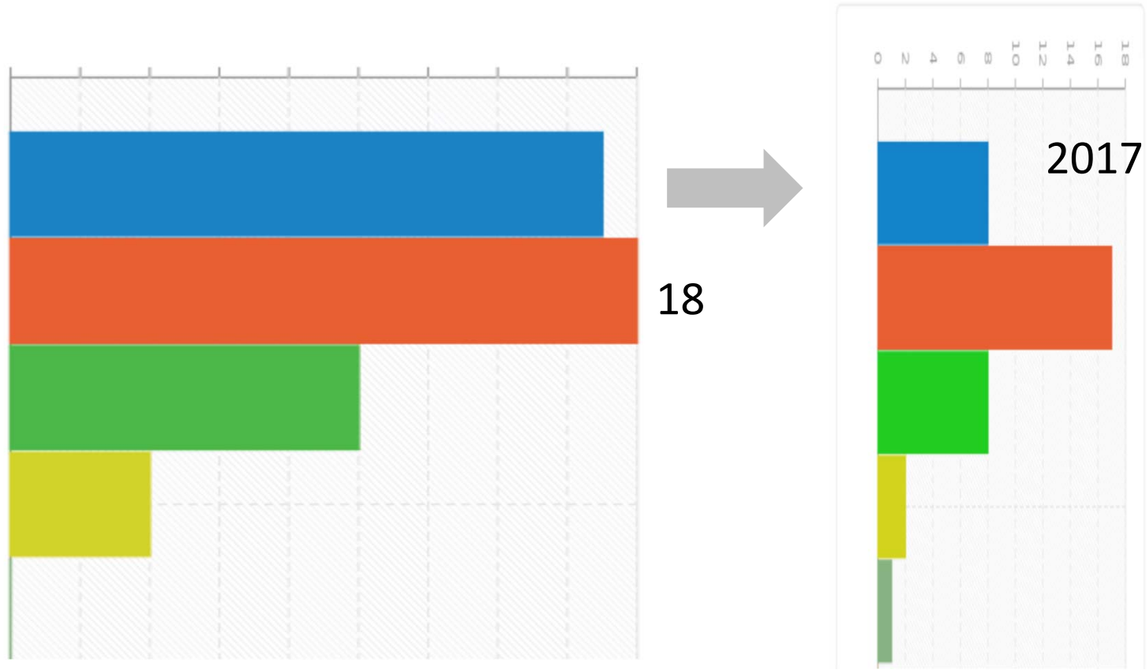
lots of discussions & initiatives happend in 2017

Participants

snapshot 20-Feb-2018, 56 answers of 95 participants

I participate in this workshop as

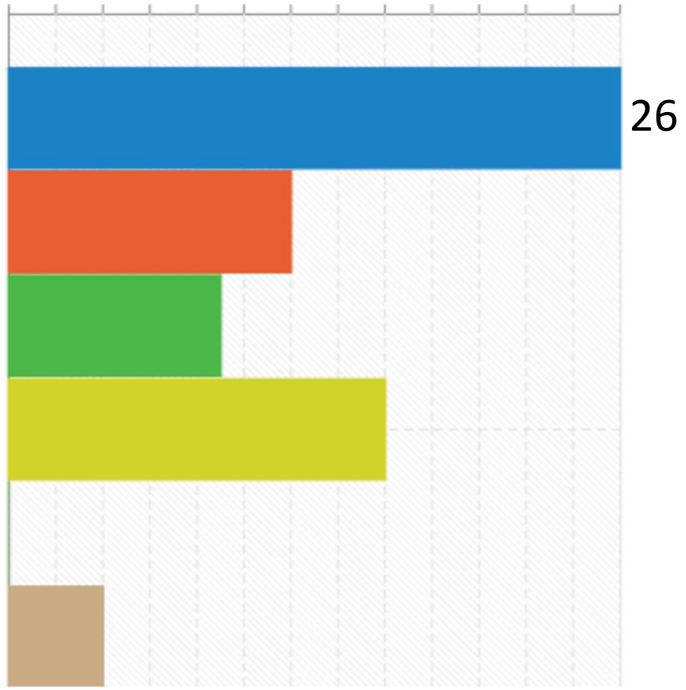
- Student
- PhD student
- Postdoctoral researcher
- Senior physicist (>5 years PhD)
- Other

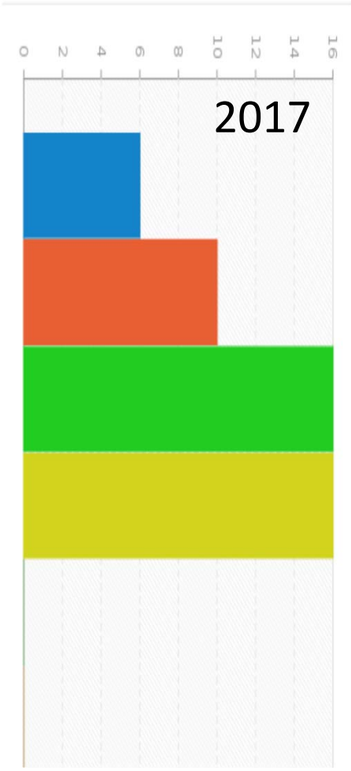


University: courses, seminars, research

I have used machine learning techniques before

- Deep learning networks
- Neural networks (shallow)
- Boosted decision trees
- None
- Don't know
- Other



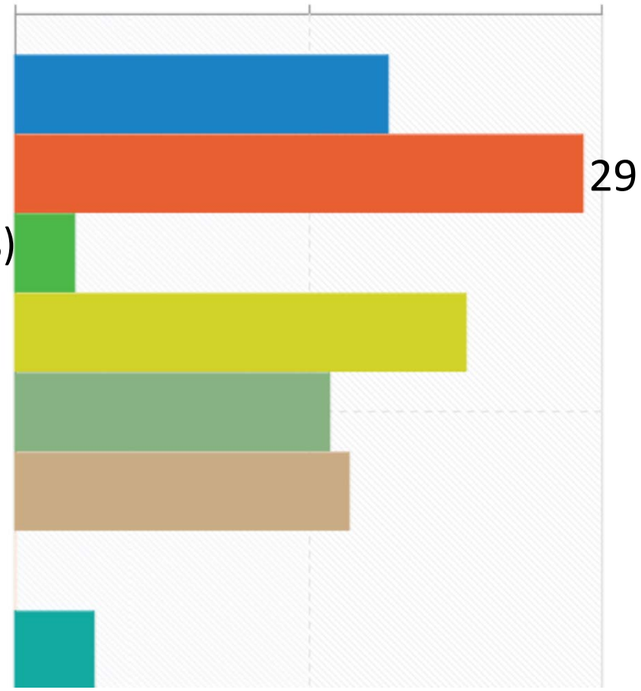


Broadcast new possibilities arising from deep learning to physics institutes

Physicists take advantage of new opportunities arising with deep networks

My research interests concerning deep learning networks are

- Image pattern recognition (e.g. event reconstruction)
- Classification (e.g. particle identification, physical processes)
- Assignments (e.g. solving ambiguities with >2 identical particles)
- Regression (calculation of variables, e.g. energy corrections)
- Adversarial training
- Curiosity
- Applications not obvious
- Other applications



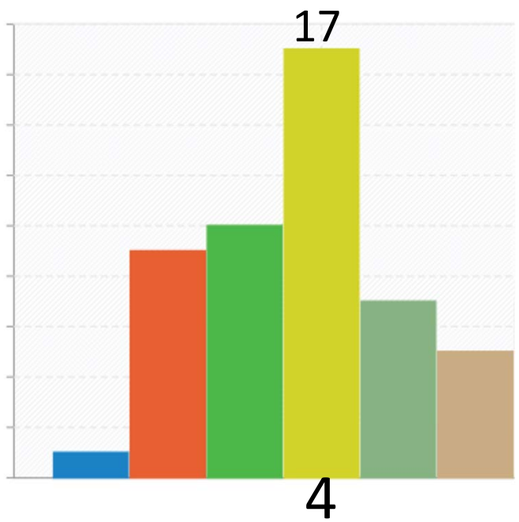
Similar distribution in 2017

Strong demands, expect exponentially increasing applications

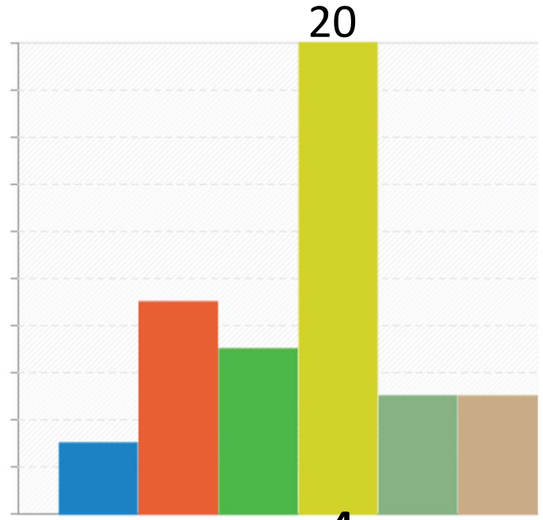
Workshop tutorial

- Introduction to machine learning with deep neural networks
- Generative adversarial networks

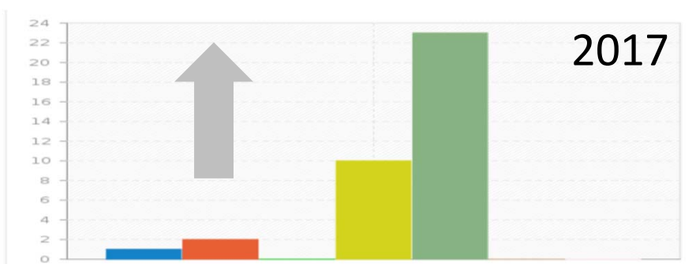
The deep learning introductory tutorial was helpful

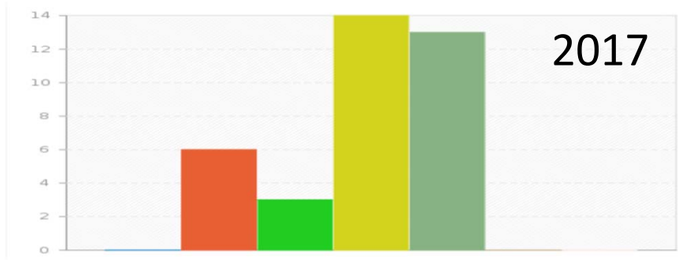


The example applications deepened my understanding of deep networks



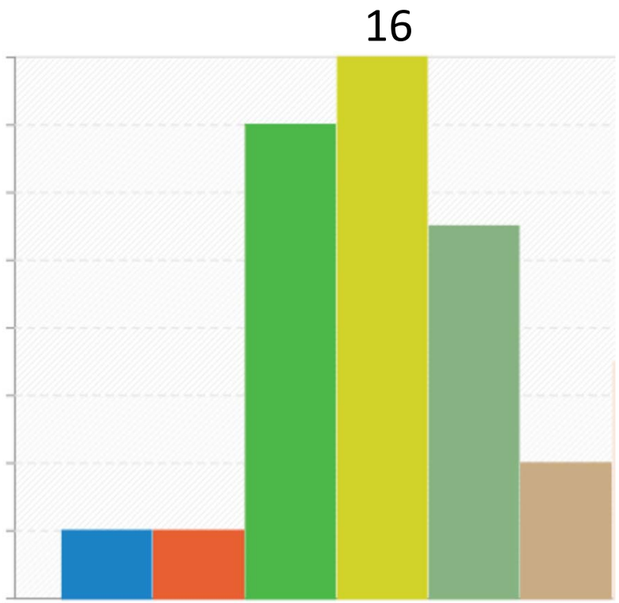
1=strongly disagree, 2=disagree, 3=undecided, 4=agree, 5=strongly agree, 6=no answer



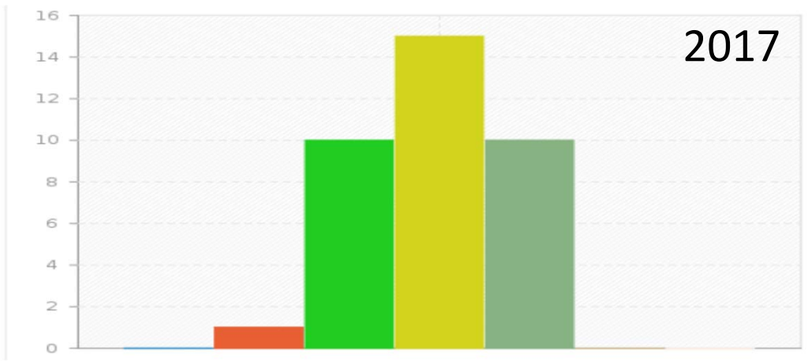


VISPA platform

My overall score for the VISPA platform



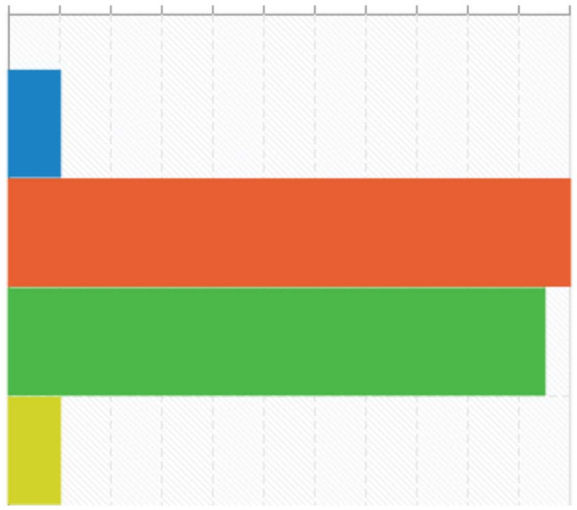
1=insufficient, 2=sufficient, 3=ok, 4=good, 5=very good, 6=no answer

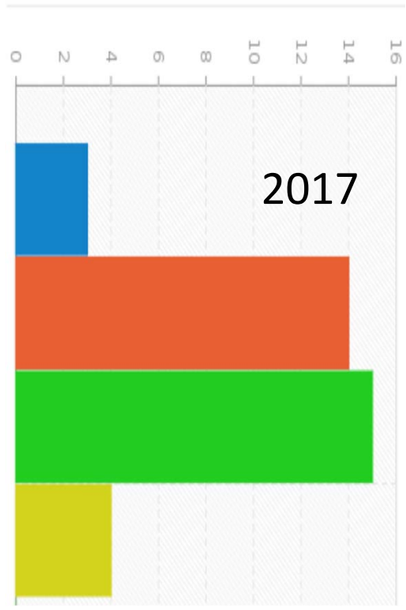


GPU resources

I have access to GPU resources for network training at my research institute

- None
- Some
- Sufficient
- Don‘t know



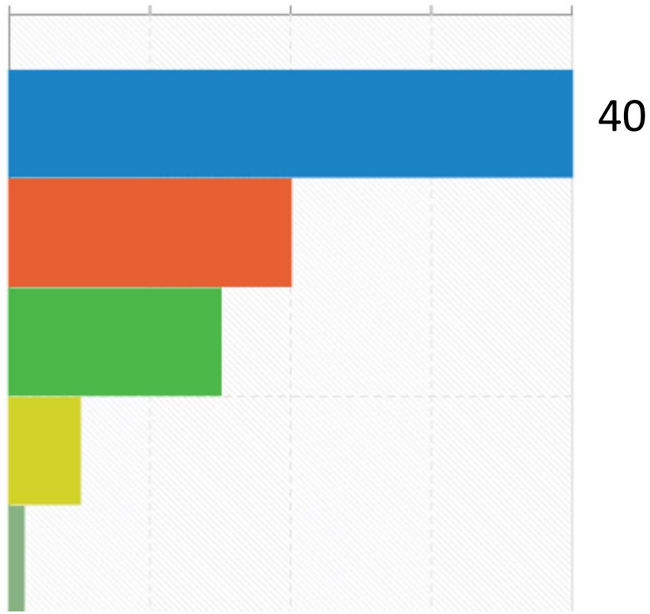


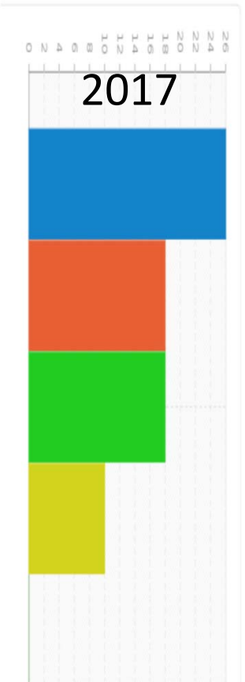
Need for resources

Experiment data, analyses, publications

I get the most important analysis support from

- My research group/institute
- Colleagues of my collaboration
- Information on websites
- Previous analyses
- Other





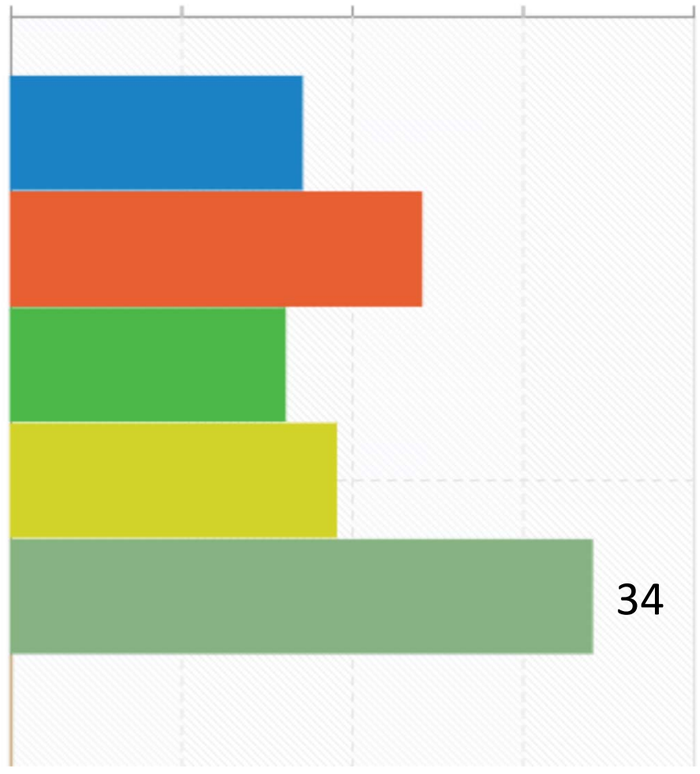
Support is an obvious prerequisite for successful exploitation of data

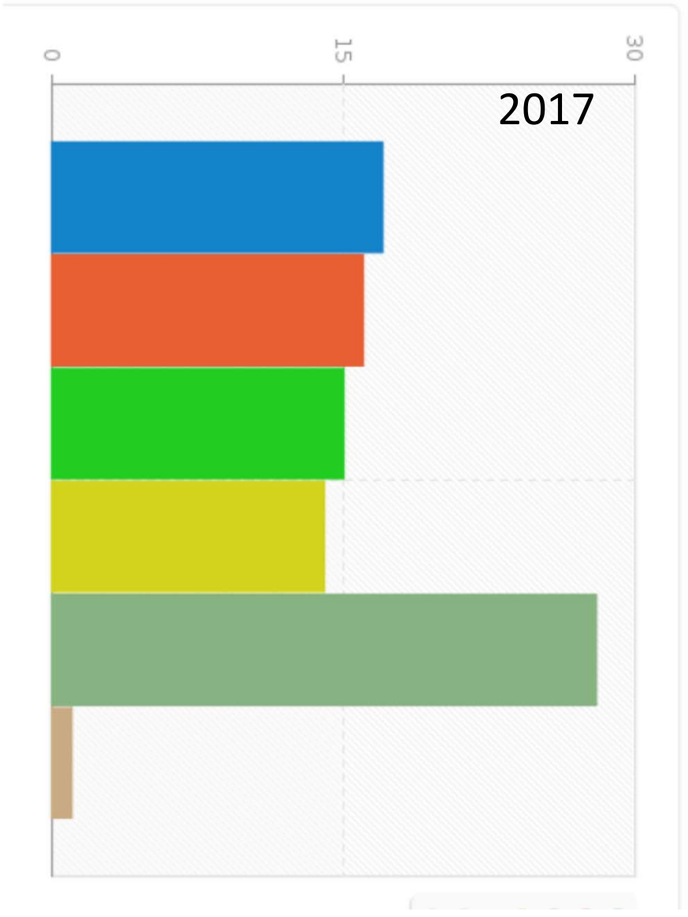
Martin Erdmann, Andreas Haungs

Experiment data, analyses, publications

For performing my research I need access to

- Raw data
- Calibrated data
- Reconstructed data
- High level data
- Simulated data
- Other



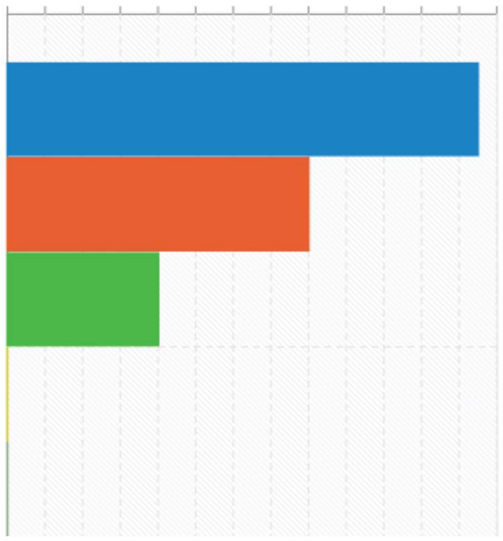


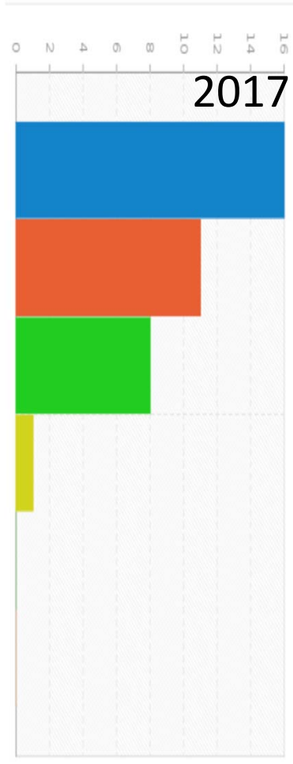
Resource intensive, aim to maximally exploite old & new data

Experiment data, analyses, publications

I have accessed open (*public*) data for my research purposes before

- Never
- Sometimes
- Frequently





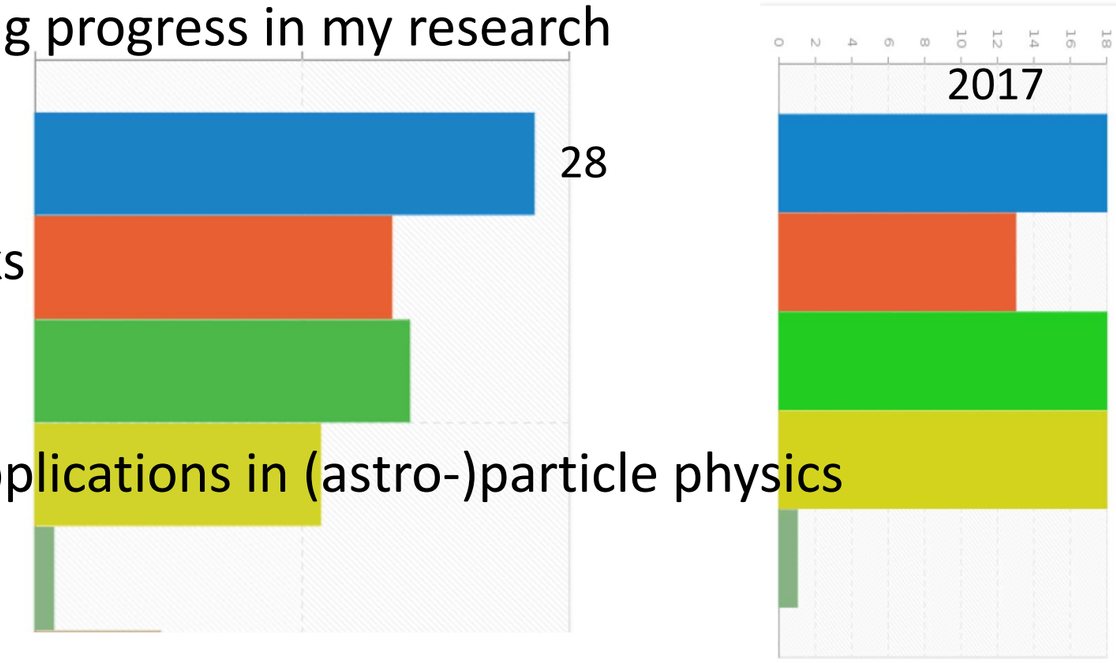
Requires scientific experience to judge what to extract from which data

Martin Erdmann, Andreas Haungs

Improve exploitation of new and old data

The following **measures** are important for achieving progress in my research

- User support deep learning
- Common platform to collect and re-use networks
- National GPU resources for basic research
- Common national structure for deep learning applications in (astro-)particle physics
- others

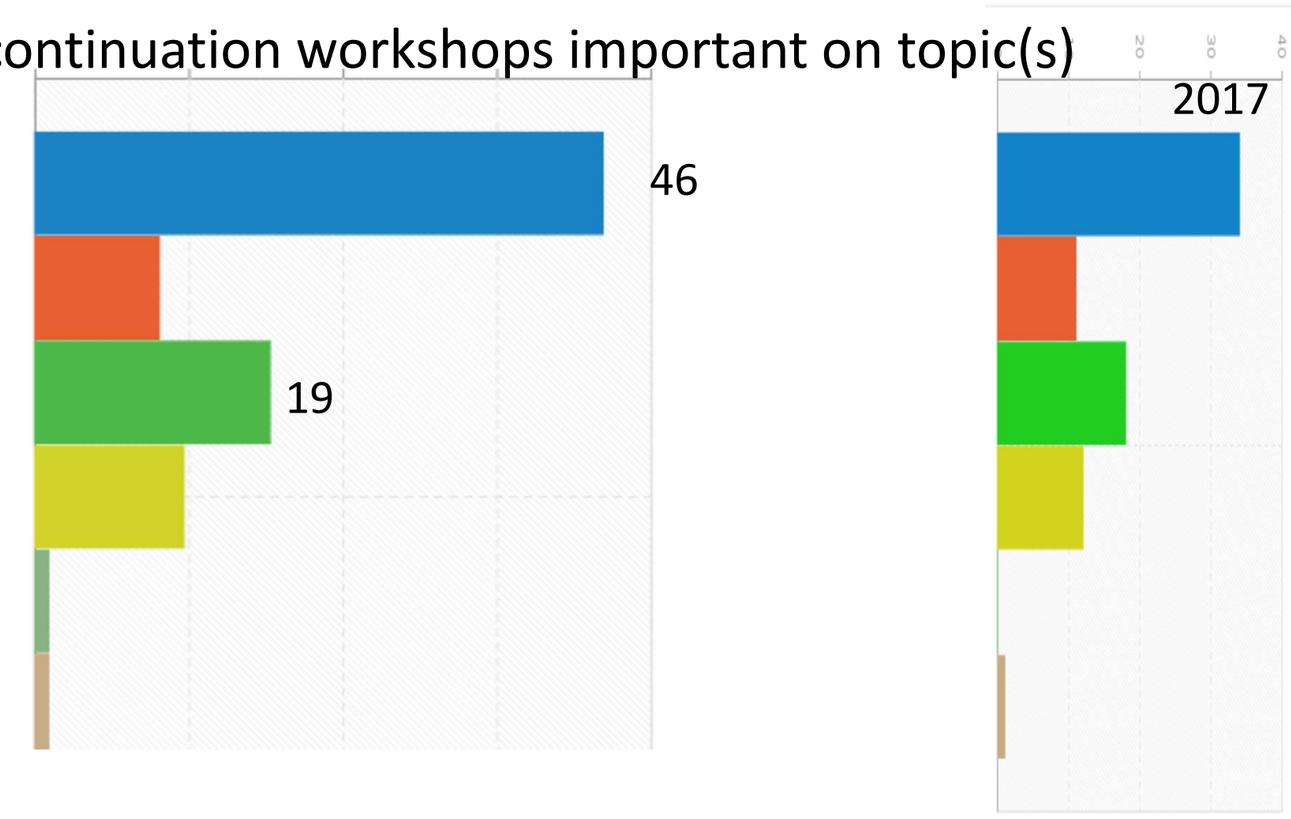


Need sustainable structures for accelerating research & development

Workshops perspectives

For the future, I consider continuation workshops important on topic(s)

- Deep learning
- Open data
- Open analysis software
- Analysis preservation
- Not important
- Other



Strong wish to benefit from each other‘s progress

Martin Erdmann, Andreas Haungs

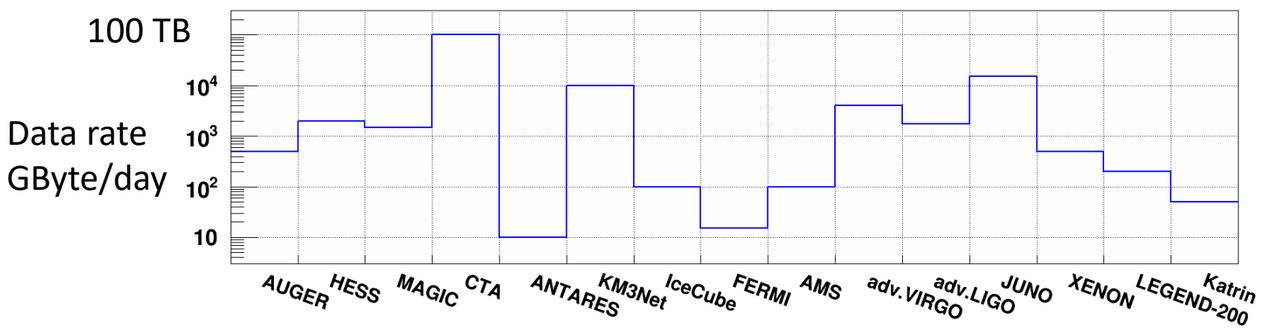
KAT. Komitee für Astro.Teilchen.Physik

Working group
13.4.2017

White paper

Astroparticle physics related to the Digital Agenda

1. Digital primary care for scientists (data storage, data processing, analysis)
2. Receipt & Publication of Measurement Data: Cross-Experimental Data Center (Multimessenger)
3. Modern techniques for physics data analysis: interdisciplinary platform & user support
4. Dedicated education for dealing with big data as a socially relevant challenge
5. Participate interested public: Outreach



extended from APPEC arXiv:1512.00988

RECOMMENDATIONS

Empfehlungen des KAT
Das KAT unterstreicht mit Nachdruck die Bedeutung der Einrichtung bzw. den Ausbau von Zentren zur Datenspeicherung, Zurverfügungstellung der Daten und der erforderlichen Rechenressourcen als digitale Grundversorgung der deutschen Wissenschaftlerinnen und Wissenschaftler und darüber hinaus für die öffentliche Teilhabe an den wissenschaftlichen Daten.
Das KAT unterstützt den Aufbau einer Netzwerkstruktur, die die Kommunikation zwischen Wissenschaftlerinnen und Wissenschaftlern als Nutzer wissenschaftlicher Daten und moderner Datenanalysemethoden einerseits ermöglicht und weiterhin eine Expertenberatung im Rahmen von User Support umsetzt.
Das KAT weist auf die zentrale Bedeutung drittmittelfinanzierter sowie nachhaltig angelegter Personalstellen hin, die für die Unterstützung der Nutzer zwingend erforderlich sind.

Martin Erdmann, Andreas Haungs

White paper

Draft: common strategy of national comittees KAT, KET, KfB, KFN, KFS, KHuK, RDS

1. Access to high-performance information and communication technology (ICT) and ensuring basic digital primary care for scientists
2. Federation of Data Centers in National Physics Science Cloud, Network Connectivity (> 100 Gb/s), Long-term Archiving (Data, Metadata, Software, Data Analysis)
3. Algorithms and Software Development Skills: interdisciplinary development of novel algorithms and tools, metadata systems, user friendly scientific concepts for large-scale devices
4. Governance, policies: involvement of communities, development of data policies (blocking periods, access rights), copyrights, active participation in European science cloud strategy
5. Education and User Support: Utilization of state-of-the-art IT technology for junior scientists, curricula for data scientists, better reputation and career opportunities for ICT close people

Martin Erdmann, Andreas Haungs

White paper

A Roadmap for HEP Software and Computing R&D for the 2020s

The HEP Software Foundation

Conclusions (shortend version)

If there was any lingering notion that software or computing could be done cheaply by a few junior people for modern experimental programmes, that should now be thoroughly dispelled.

Software and Computing needs investment in people who understand problems & solutions employed today & have skills to provide innovative solutions for the future. There needs to be recognition from community with a recognized career path for these experts.

Martin Erdmann, Andreas Haungs

Deutsche Physikalische Gesellschaft DPG: New working group „Physics, modern Information Technologie and Articifial Intelligenz“

1. BIG DATA: archiving, processing, management, analysis and simulation of complex data streams, HPC, information theory, statistical methods
2. IT: High-performance Data Readout Systems and Mass Storage, Visualization, Smart Sensors, Bridge Technologies for the Next Level of Big Data
3. AI & ROBOTICS: Data driven algorithms & software, autonomous devices, remote control, innovative applications, algorithms for quantum computers
4. HIGHER EDUCATION: Curricula and multi-disciplinary research centers, cooperation with the GI Data Scientist Task Force, IT Infrastructure
5. INDUSTRY and SOCIETY: ethics, technology assessment, sustainability, economics, law, start-ups, public relations

DPG conference Würzburg 19.-23.3.2018
Arbeitskreis Physik, IT & KI (AKPIK)
Monday 16:30-18:45
Tuesday 16:50-18:50
Mitgliederversammlung des Arbeitskreises
Election: Wahl des Vorstands

Martin Erdmann, Andreas Haungs

Summer term 2017 & 2018

Education:
Deep Learning in Physics Research

12 lectures and exercises

1. Fundamentals of deep learning
2. Regularization & generalization
3. Optimization & hyperparameter tuning
4. Convolutional neural networks
5. Classification of magnetic phases
6. Advanced computer vision methods
7. Application in astroparticle physics
8. Autoencoders & application in solid state physics
9. Generative adversarial networks
10. Restricted Boltzmann machines
11. Recurrent networks
12. Summary & a bit more on recurrent networks

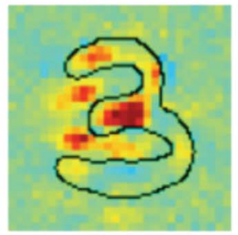
Martin Erdmann, Andreas Haungs

Why Interpretability ?

Different dimensions of "interpretability"

prediction

"Explain why a certain pattern x has been classified in a certain way f(x)."



model

"What would a pattern belonging to a certain category typically look like according to the model."

data

"Which dimensions of the data are most relevant for the task."

Initiative for a
National Data Centre for Astroparticle Physics

Initiative for a Data and Analysis Centre for Astroparticle Physics
Karlsruhe, 2-Nov-2017
2 November 2017
Karlsruhe Institute of Technology (KIT)
Europe/Berlin timezone

- Overview
- Scientific Programme
- Timetable
- Contribution List
- Author List
- Registration
  - Registration Form
- Participant List

Campus Plan

for the slides, please click left on "Contribution List"!!

preliminary Agenda: click left on 'Timetable' or below on 'Poster'

Organizing Committee: Andreas Haungs (KIT), Christian Stegmann (DESY), Achim Streit (KIT), Sabine Bucher (KIT)

Discuss with you
future prospects:

Goals

- Data catalogues & computing resources
- Analysis & simulations
- User interface to data & analysis
- User support & user platforms

Martin Erdmann, Andreas Haungs

New paradigm in physics

- Up to now: **physics laws & mathematics** are basis for algorithm development → machine learning with physicist‘s favorite observables

- New: **machines exploit data deeper** than physicist‘s algorithms so far

- **Together we ought to keep working on the fundamental change to include machines in our daily work**

it is your generation – you – who will drive the future

Thanks very much!

Workshop administration:
Gerlinde Schmidt, Melanie Roder

Local RWTH workshop team:
Florian Briechle, Jonas Glombitza, Yannik Rath, Marcus Wirtz

and the VISPA team:
Benjamin Fischer, Erik Geiser, Marcel Rieger, Felix Schlüter, Martin Urban, Florian von Cube

Martin Erdmann, Andreas Haungs