

DeepJetCore

Swapneel Mehta, Mauro Verzetti, Jan Kieseler, Markus Stoye, Maurizio Pierini
CMS Experiment, EP-CMG-PS
CERN

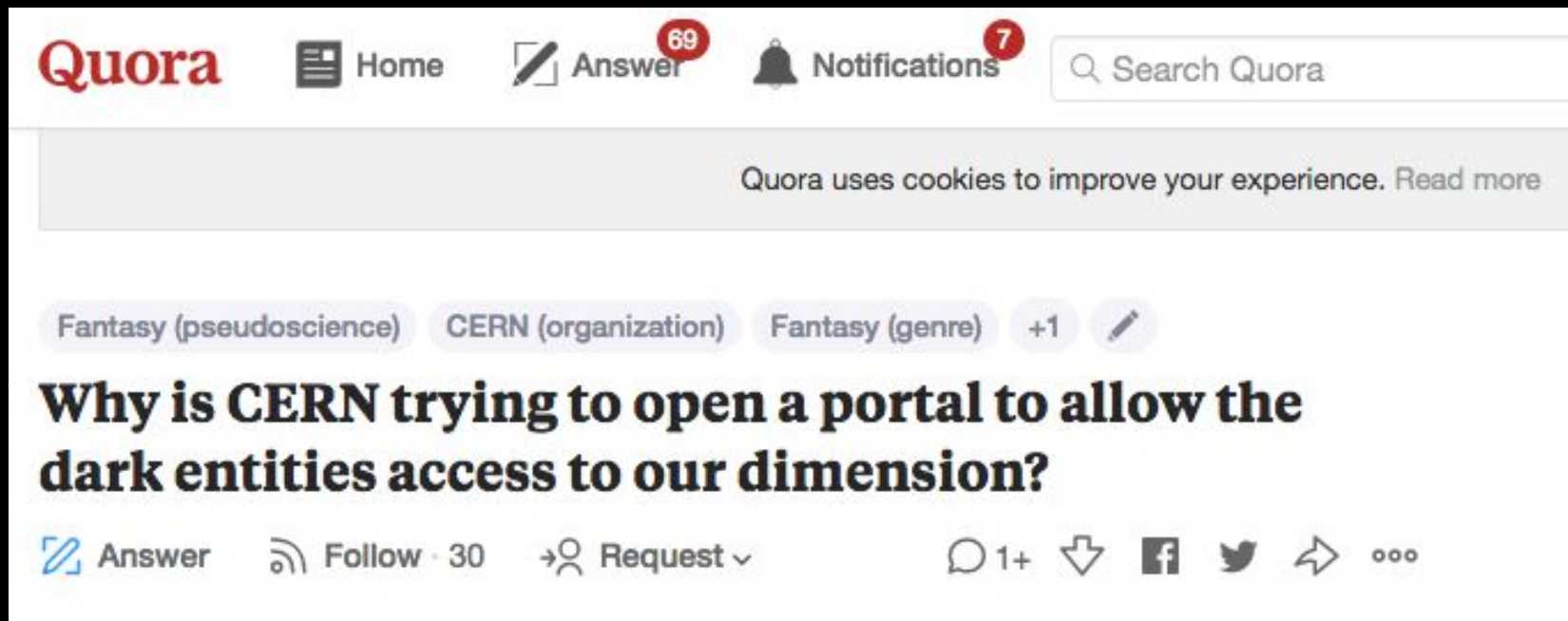
Machine Learning

1. Comprehensive libraries
2. Fantastic documentation
3. Interactive Tutorials
4. Developer Community Support



Why build a
library designed
for high-energy
physics?

Computer Scientists don't always understand requirements for particle physics...



The image shows a screenshot of a Quora question page. At the top, the Quora logo is on the left, followed by navigation links for Home, Answer (with a red badge showing 69), and Notifications (with a red badge showing 7). A search bar is on the right. Below the navigation is a grey banner stating "Quora uses cookies to improve your experience. Read more". Underneath the banner are several topic tags: "Fantasy (pseudoscience)", "CERN (organization)", "Fantasy (genre)", and "+1" with a pencil icon. The main question is "Why is CERN trying to open a portal to allow the dark entities access to our dimension?". Below the question are interaction options: "Answer", "Follow · 30", "Request" (with a dropdown arrow), and social sharing icons for comments (1+), a download icon, Facebook, Twitter, a share icon, and a menu icon (three dots).

Quora

Home Answer ⁶⁹ Notifications ⁷ Search Quora

Quora uses cookies to improve your experience. Read more

Fantasy (pseudoscience) CERN (organization) Fantasy (genre) +1

Why is CERN trying to open a portal to allow the dark entities access to our dimension?

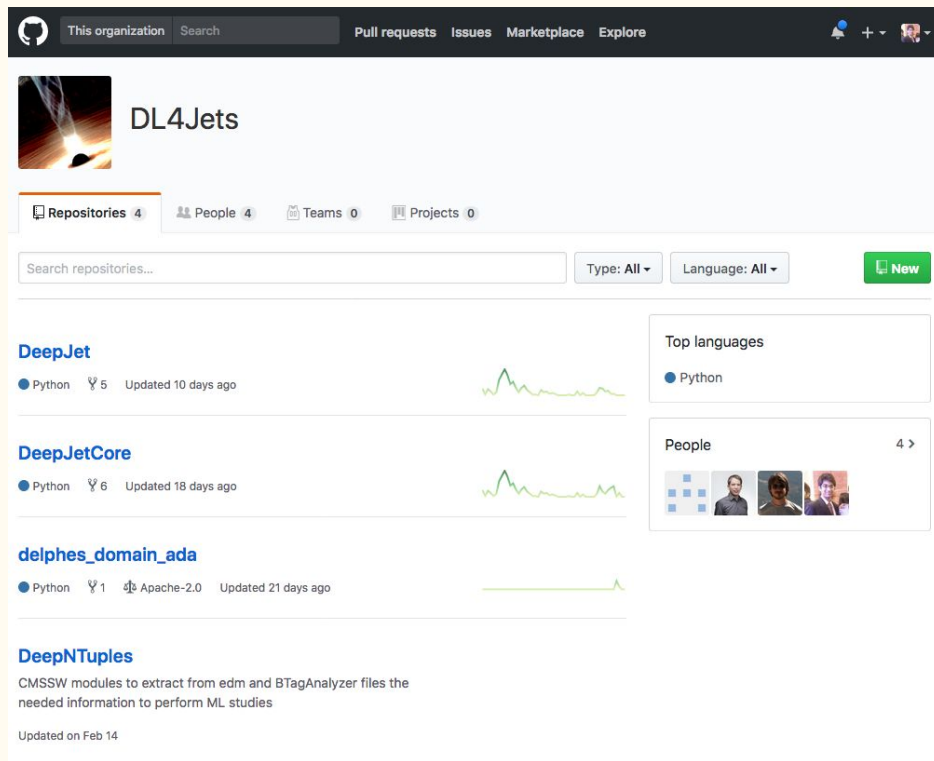
Answer Follow · 30 Request ¹⁺ [Social Sharing Icons]

Physicists don't always write great code...

```
public Date getNextDay() {  
    try {  
        Thread.sleep(TimeUnit.DAYS.toMillis( duration: 1));  
        return new Date(); //success  
    } catch (InterruptedException e) {  
        e.printStackTrace();  
        return null; //failure  
    }  
}
```

Best of Both Worlds

1. Implement fast, efficient machine learning algorithms for physics
2. Provide high-level functions/wrappers for low-level tasks
3. Handle common bottlenecks - esp. memory -related issues
4. Create an extensible, easy-to-use framework



The screenshot shows the GitHub organization page for DL4Jets. The header includes navigation links for Pull requests, Issues, Marketplace, and Explore. The organization's profile picture is a stylized image of a particle detector. Below the profile, there are statistics for Repositories (4), People (4), Teams (0), and Projects (0). A search bar for repositories is present, along with filters for Type and Language, and a 'New' button. The main content area lists several repositories:

- DeepJet**: Python, 5 forks, updated 10 days ago. Includes a green line graph showing activity.
- DeepJetCore**: Python, 6 forks, updated 18 days ago. Includes a green line graph showing activity.
- delphes_domain_ada**: Python, 1 fork, Apache-2.0 license, updated 21 days ago. Includes a green line graph showing activity.
- DeepNTuples**: CMSSW modules to extract from edm and BTagAnalyzer files the needed information to perform ML studies. Updated on Feb 14.

On the right side, there are sections for 'Top languages' (Python) and 'People' (4 members).

What does this
library do?

Features of DeepJet

- Data Conversion
 - Model Training
 - Prediction
 - Model Evaluation
-

- File-by-File
- Avoids memory threshold crossed (EOS)
- Handles user-defined data structures
- Preprocessing support
- Parallelized operation

Conversion

- Keras-wrapped Tensorflow backend
- Additional callbacks
- Monitor validity of tokens
- Bookkeeping support

Training

- Create compatible prediction data structures
- Support for Plots
- Export of models and data structures

Prediction and Evaluation

Yeah, but why
should I use it?

- Modularised code, easy to understand
- Templates for quick-start
- Step-by-step documentation
- Elaborate examples and use-cases

Simplicity

- C++ Extensions improve efficiency for Python
- Anaconda support
- Available as a Python Package
- Docker support

Support

Interesting! Tell
me more about
this library

DeepJetCore

📁 DeepJetCore	override build_ext for custom make
📁 bin	modified README, added Apache License
📄 .gitignore	modified cmdclass adding custom build in setup
📄 CHANGELOG.rst	breaking changes for setup and compilation of scripts as extensions
📄 MANIFEST.in	modified MANIFEST to add swig wrappers in package
📄 README.md	restructuring and modifications to setup
📄 README.rst	modified README, added Apache License
📄 environment.yml	modified setup for manual build
📄 setup.cfg	modified setup for manual build
📄 setup.py	override build_ext for custom make

bin	little tool to add prediction labels to a data collection - only need...
compiled	OSX setup
environment	OSX setup
evaluation	fix
preprocessing	more checks and fixes. fixed other global treename per traindata
training	better handling of prediction labels, more options, better bookkeepin...
.gitignore	updates for domain adaptation
DataCollection.py	better handling of prediction labels, more options, better bookkeepin...
LICENSE.txt	adding license
MultiDataCollection.py	not working restructuring commit
README.md	Update README.md
TrainData.py	better handling of prediction labels, more options, better bookkeepin...
Weighter.py	not working restructuring commit
__init__.py	not working restructuring commit
gpu_env.sh	more checks and fixes. fixed other global treename per traindata
lxplus_env.sh	more checks and fixes. fixed other global treename per traindata
modeltools.py	not working restructuring commit
stopwatch.py	not working restructuring commit

DeepJet

📁 Analysis	fixed makefile
📁 Train	better label bookkeeping - connected to Core, and bugfixes for dom ad...
📁 independence/example	decorrelation and independence
📁 modules	better label bookkeeping - connected to Core, and bugfixes for dom ad...
📁 scripts	removed core part
📄 .gitignore	yet another update
📄 LICENSE.txt	added license
📄 README.md	Update README.md
📄 env.sh	one-size-fit-all environment
📄 gpu_env.sh	more restructuring
📄 lxplus_env.sh	more restructuring

- Install not entirely automated yet
- Platform supported: CentOS 7
- Breaking changes in recent releases of Tensorflow

Limitations

Conclusion

- Easy-to-use Framework
 - Faster conversion and training
 - Diverse use-cases
 - Scalable to large datasets
-

Thank You!

