DeepJet Framework

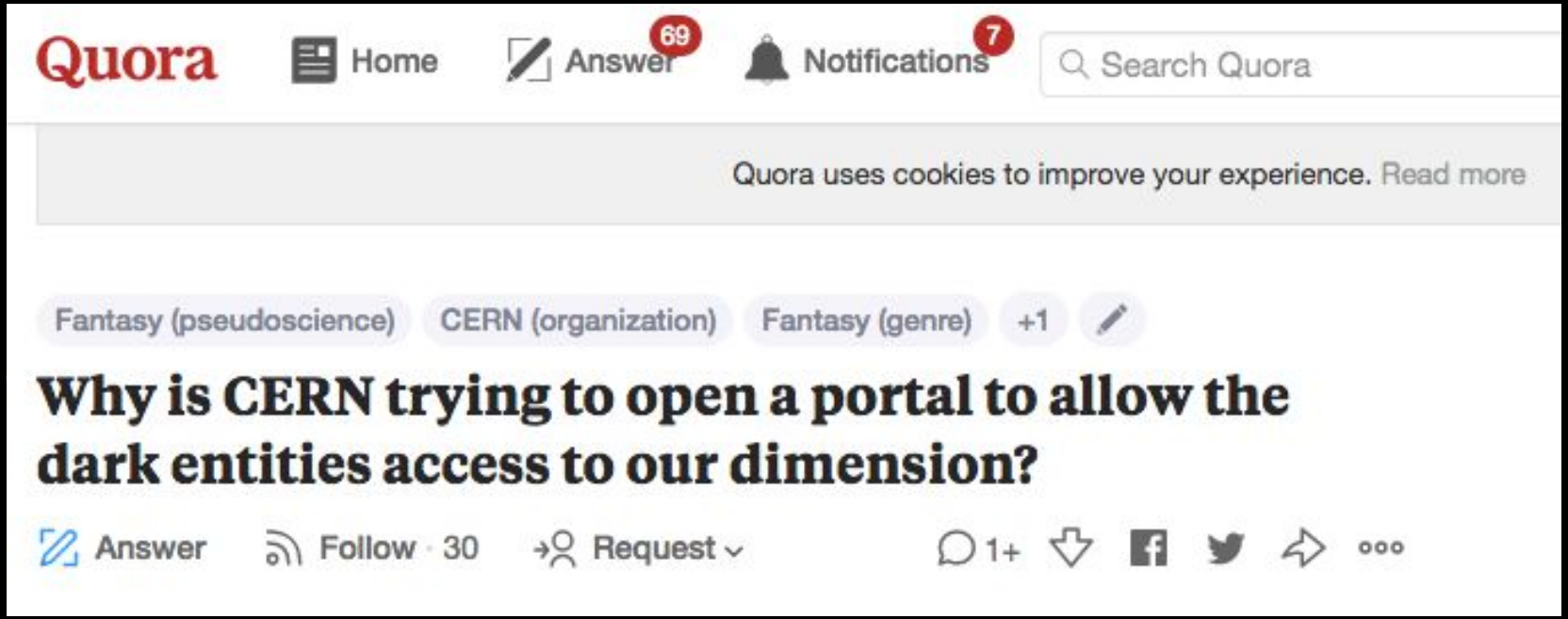
Swapneel Mehta, Mauro Verzetti, Jan Kieseler, Markus Stoye, Maurizio Pierini
CMS Experiment, EP-CMG-PS
CERN

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...

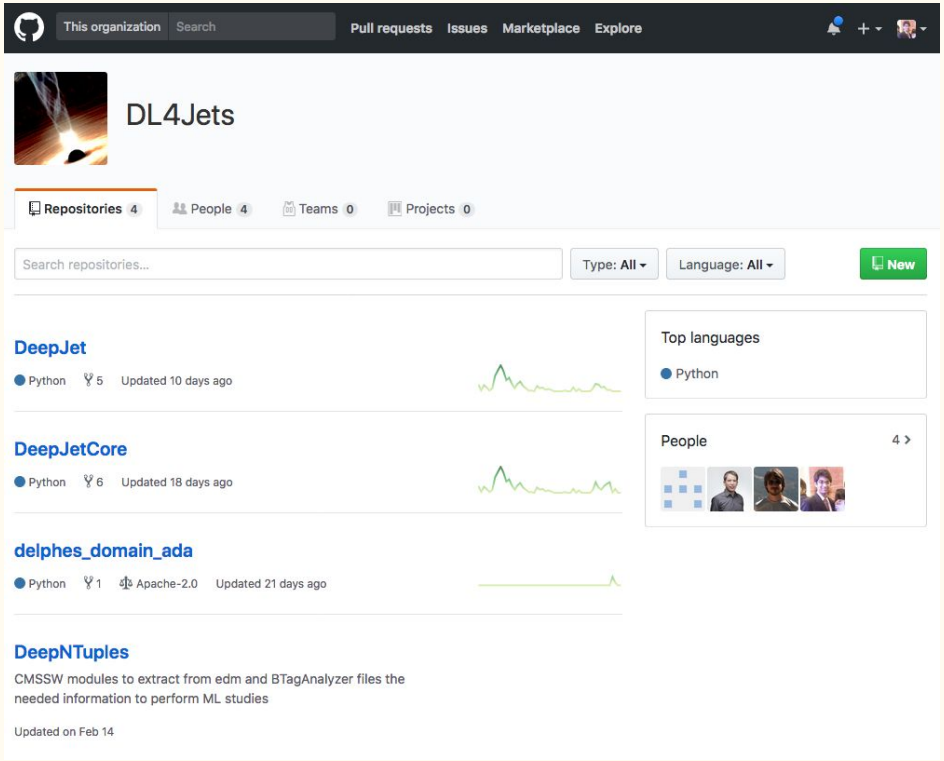


Physicists don't always write great code...

```
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



What does this library do?

Features of DeepJet

- Data Conversion
- Model Training
- Prediction
- Model Evaluation

Conversion

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

Training

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

Prediction and Evaluation

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

Yeah, but why should I use it?

Simplicity

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

Support

- Custom CPP Extensions improve efficiency for Python
- Automation of specific tasks
- Anaconda Environment

Upgrades

- Available as a pip package for Python 3.6
- Tensorflow 1.8 supported
- Integrating support for TFRecords
- Docker Image Distribution

Interesting! Tell me more about this library

DeepJet Demo

Conclusion

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