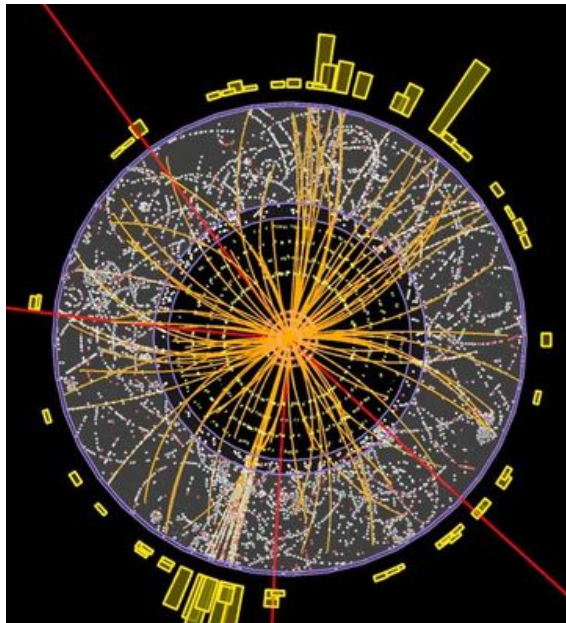# Wrapping up and next steps

Juan Rojo
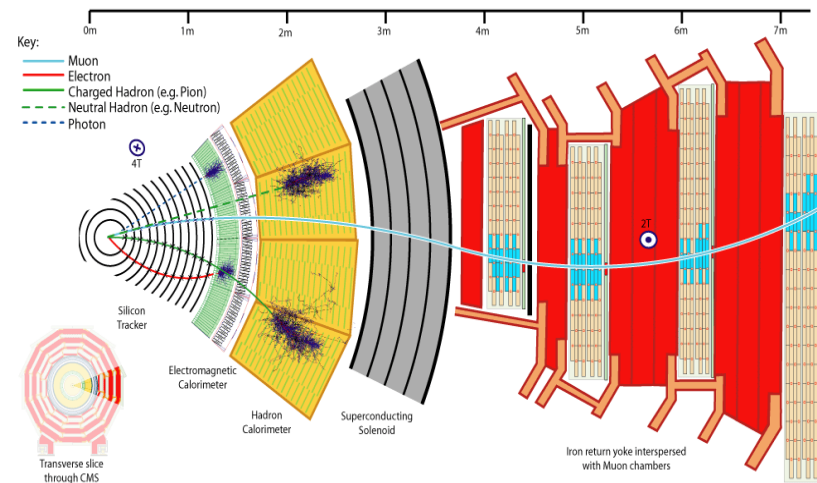VU Amsterdam & Theory group, Nikhef
(on behalf of the workshop organisers)

Big Data Tools in Physics and Astronomy
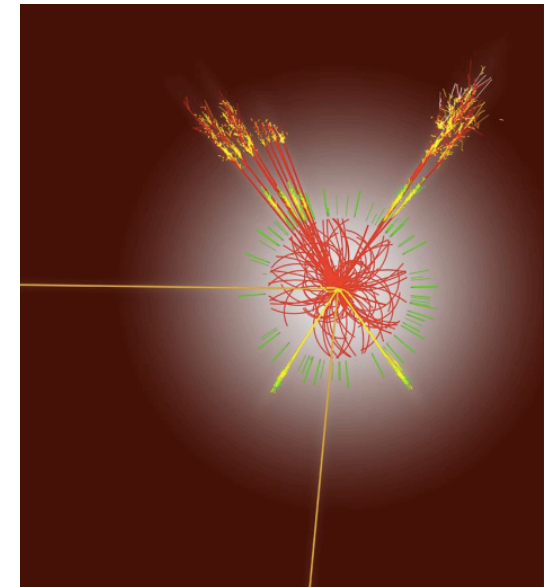**Amsterdam, 12/06/2017**

Machine Learning tools are everywhere!

Deep Kalman RNNs

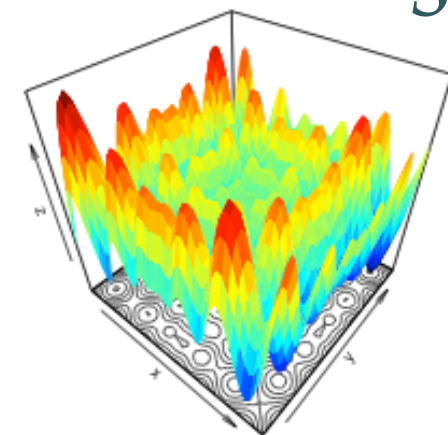**Generative Models, Adversarial Networks**

FCN, Recurrent, LSTM NN
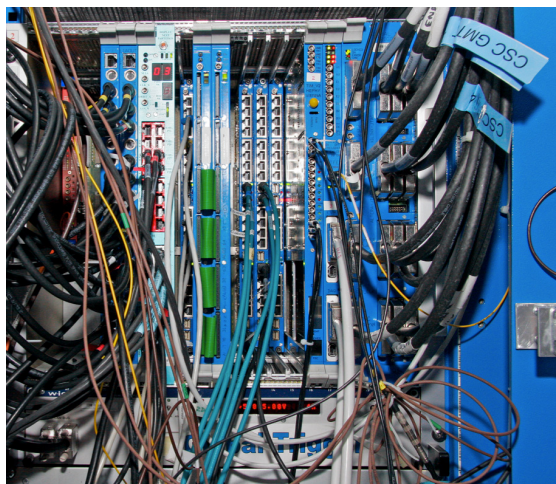
S. Glayzer

Deep ML +FPGA
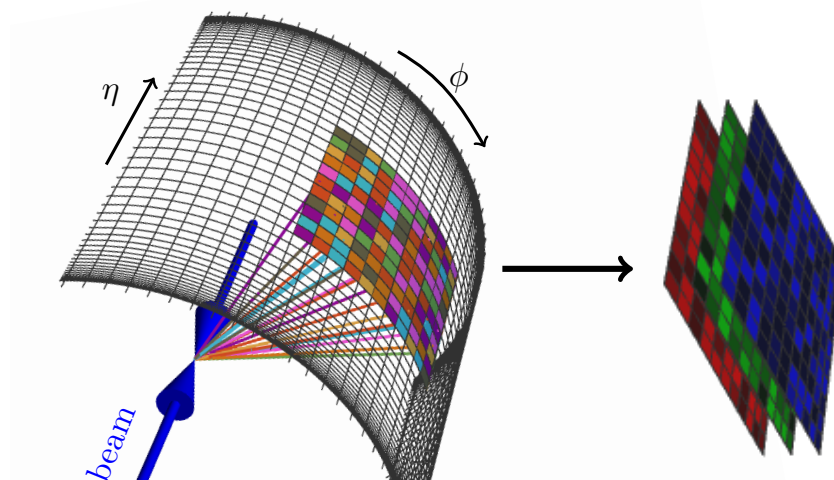
Convolutional DNN

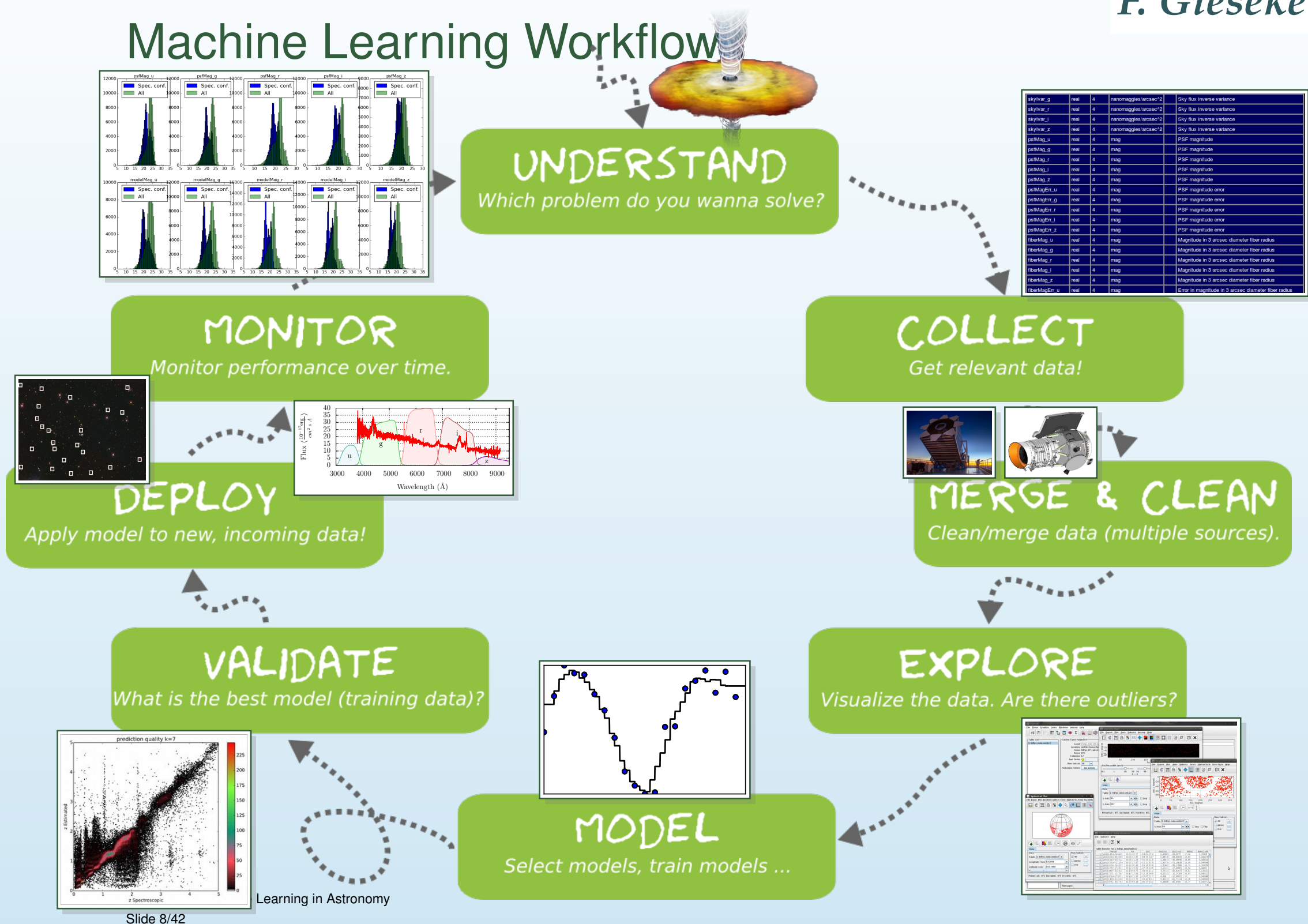**Multiobjective Regression**

06/19/2017   15

For many crucial applications, ML tools not just one option, but **the only option**

ML cheat sheet

F. Gieseke

Machine Learning Workflow

UNDERSTAND
Which problem do you wanna solve?

COLLECT
Get relevant data!

MONITOR
Monitor performance over time.

MERGE & CLEAN
Clean/merge data (multiple sources).

DEPLOY
Apply model to new, incoming data!

EXPLORE
Visualize the data. Are there outliers?

VALIDATE
What is the best model (training data)?

MODEL
Select models, train models ...

Learning in Astronomy
Slide 8/42

Endless possibilities - but also many non-trivial hurdles to overcome
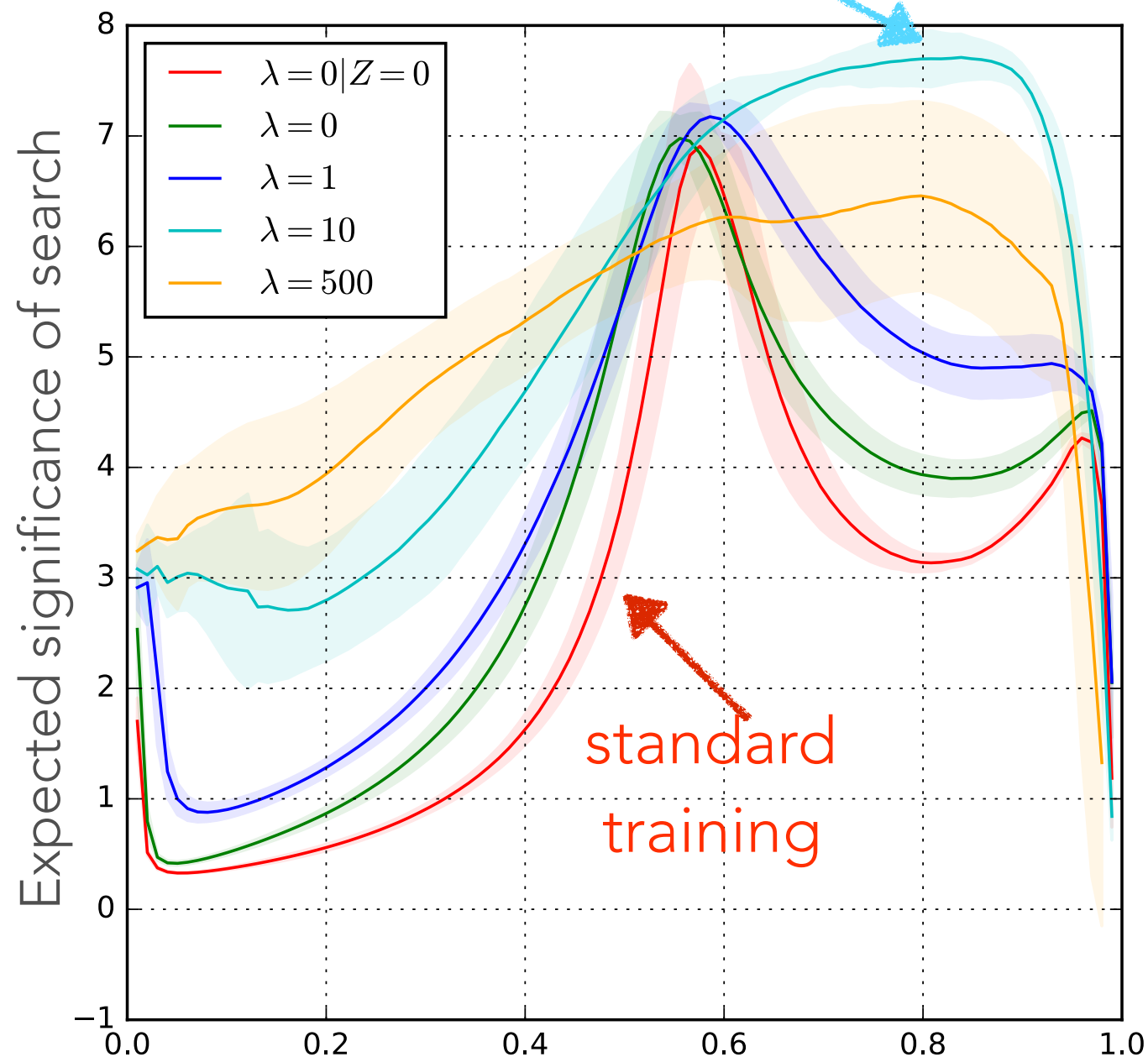
Science drivers of ML tools

An example:

background: 1000 QCD jets
signal: 100 boosted W's

Train W vs. QCD classifier

Simple cut-and-count
analysis with background
uncertainty.

optimal tradeoff of classification vs. & robustness

standard
training

Legend:
- $\lambda = 0 | Z = 0$
- $\lambda = 0$
- $\lambda = 1$
- $\lambda = 10$
- $\lambda = 500$

y-axis: Expected significance of search

Focus on **scientific requirements** for ML development

ie, **robustness** wrt experimental systematic uncertainties

Science drivers of ML tools

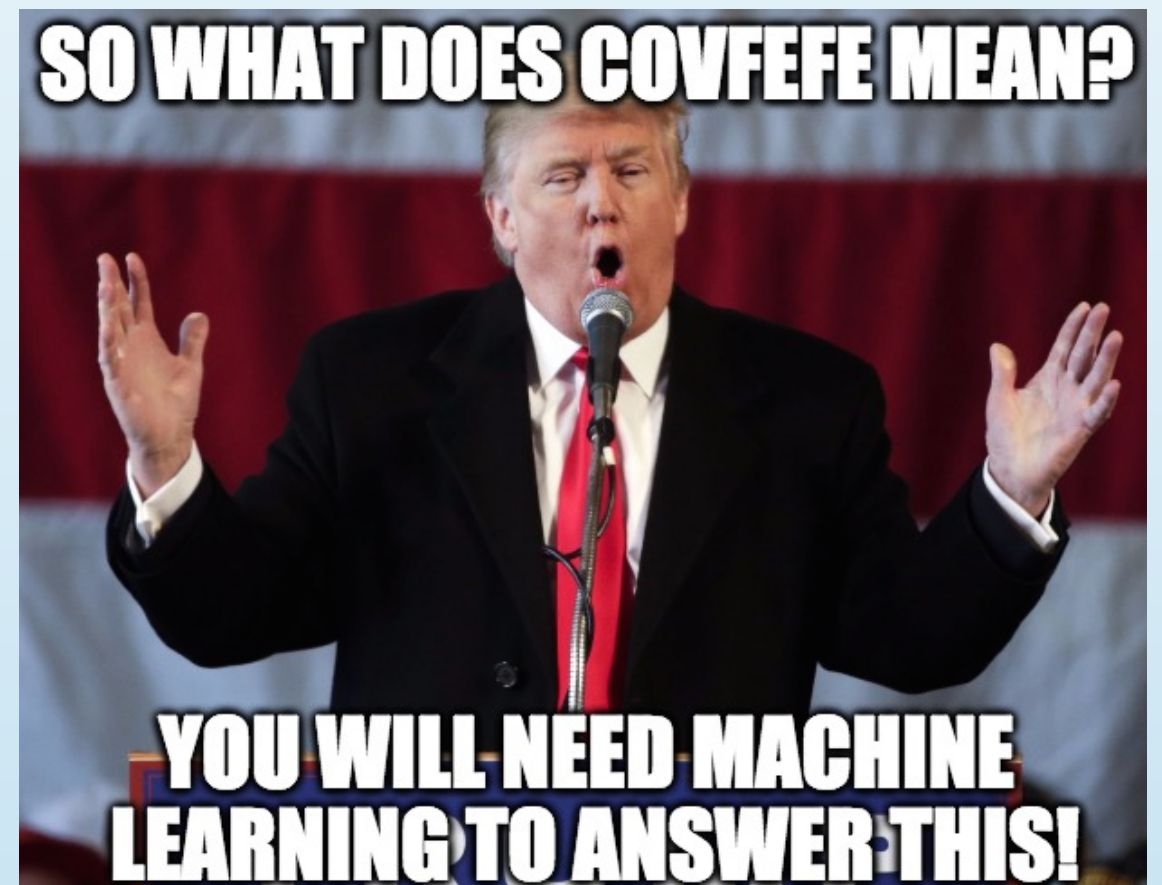though of course we all love **fun applications of ML tools!**
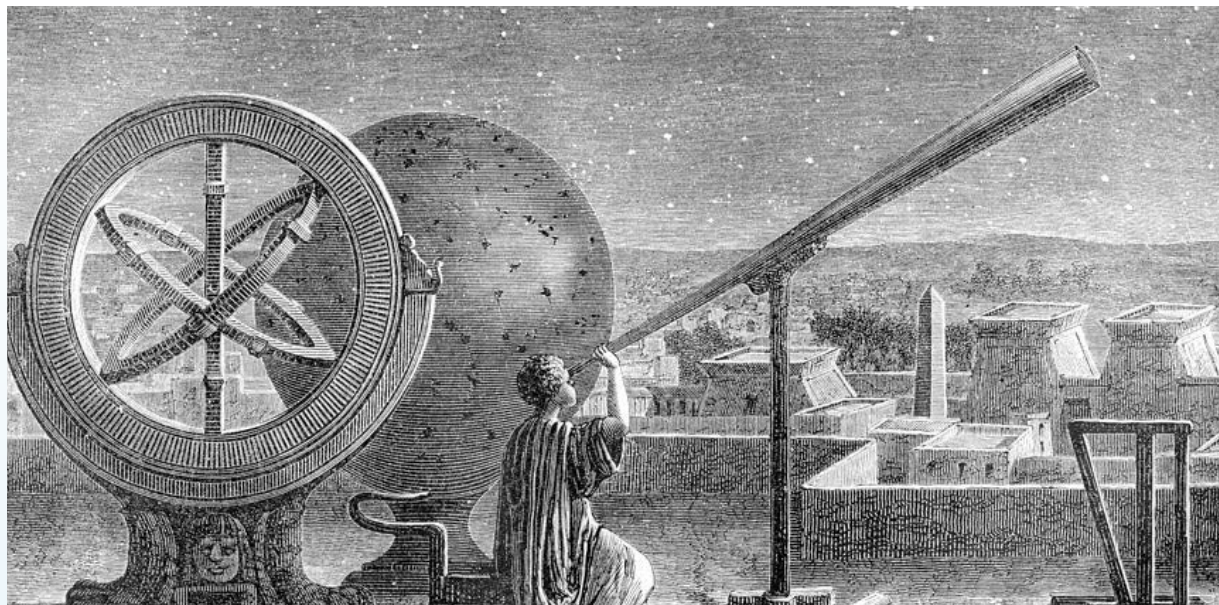
Some ideas for the next steps

- This workshop has nicely illustrated the lively activities of the **Machine Learning / Big Data Physics & Astronomy** community in Amsterdam

- Many other groups in the NL: **Broaden scope to the whole Dutch community**

- Is there momentum to create a **Dutch ML P&A Working Group**? With a webpage with the list of ongoing activities and interested groups, code resources, literature, maybe also a forum for technical discussions?

- What about educational activities? Maybe a **ML school** for advanced PhD students and postdocs?

- Please let us know your suggestions - these ideas will only move forward of there is **genuine interest in the community!**

Stronger Together

- Exploiting **synergies within ongoing ML efforts** can boost their scientific output, as well as to avoid reinventing the wheel

- Machine Learning tools have enormous potential, but one should also make an effort to be **pedagogical and minimise jargon,** to facilitate discussions within the community and strengthen collaboration

- Joint efforts could also help in fostering **ML applications to "real world problems"** beyond academic research

Thank you!

Celeste: Variational inference for a generative model of
astronomical images