

facebook

Data Science In Industry

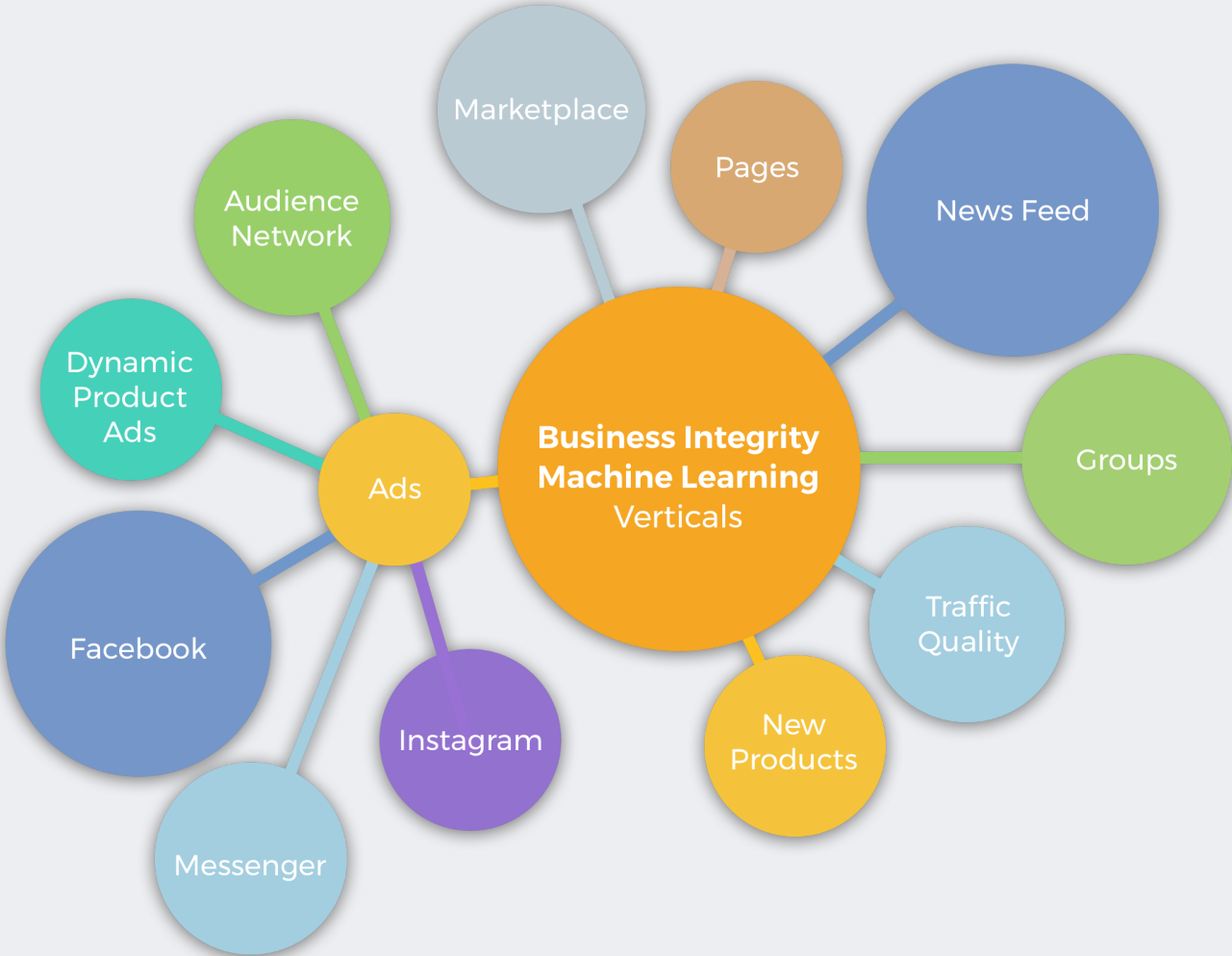
Emanuel Strauss

Engineering Manager, Business Integrity ML

eman@fb.com

Facebook Business Integrity

Ensure safe and fair connections between people and businesses



Culture

<https://medium.com/@jeremybirnholtz/what-is-hacker-culture-795459649a05>

What is Hacker Culture? Comparing the World's Largest Social Network and the World's Largest Particle Collider



JEREMY BIRNHOLTZ · THURSDAY, JUNE 25, 2015

“basic foundation of understanding the data analysis portion of ML I developed in HEP”

“My data analysis skills come in to play in the more general sense of problem solving, analytical thinking, that kind of thing”

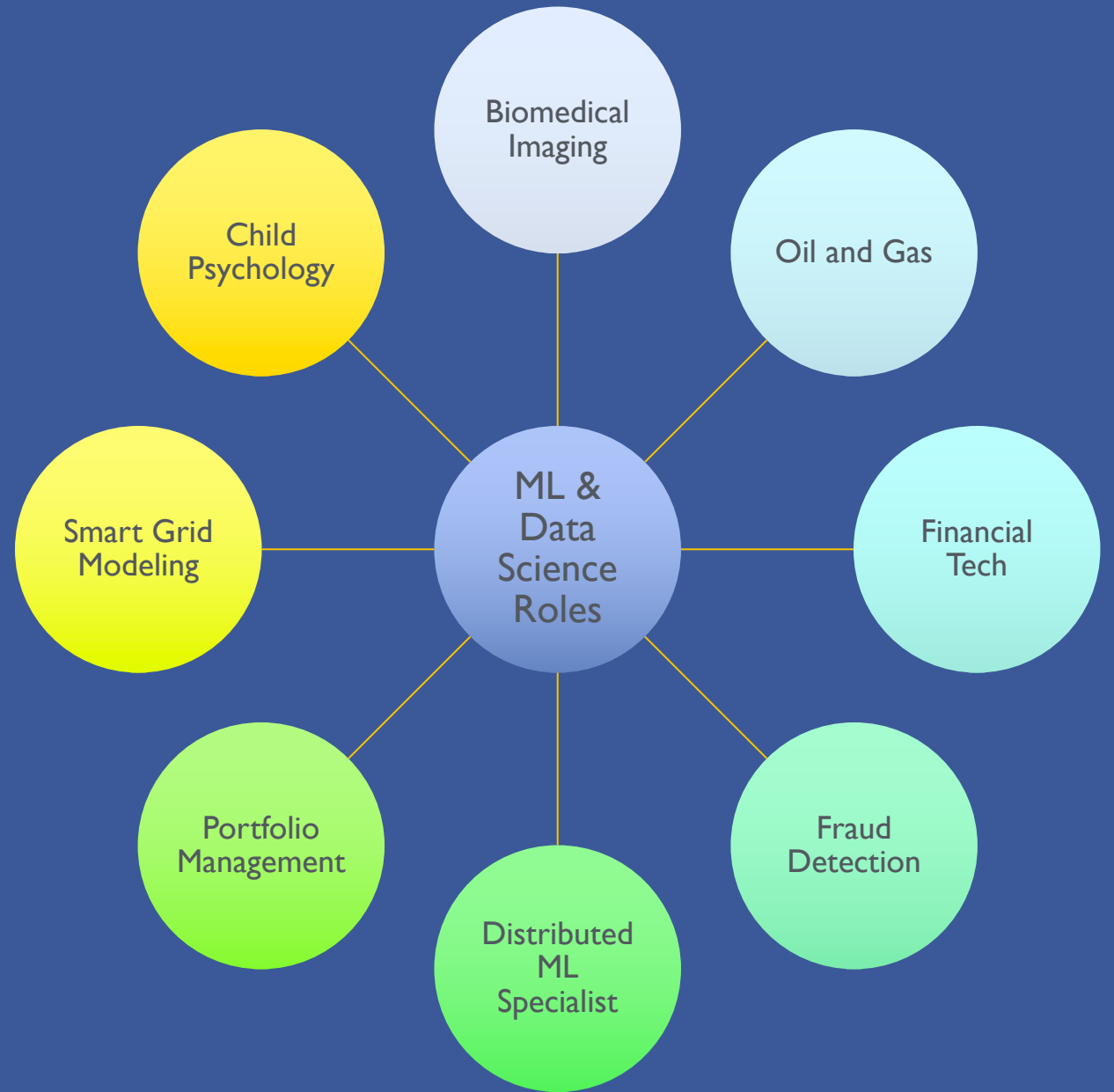
“I have learned most of my deep learning skill since I joined my company”

“I had some basic coding skills from HEP, some data vis stuff, and the ability to learn new things. That last one is the most important.”

“I coded in C++ but was never taught or shown best practices”

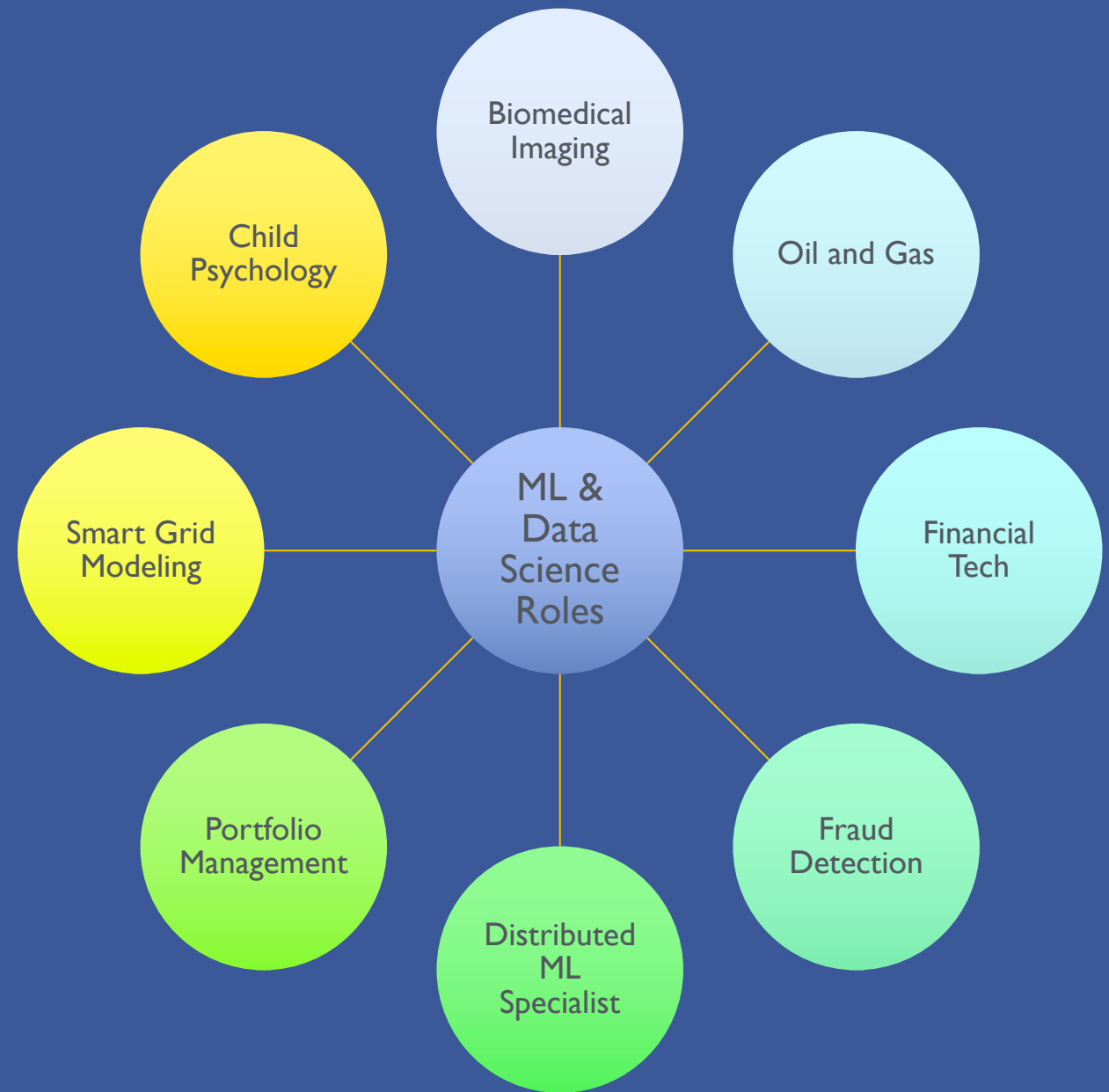
“being able to read and implement stuff off of arxiv is really useful”

“I didn't do any machine learning (even old style pre deep learning stuff) while I was in HEP”



“Physicists are usually my top candidate because I know there’s gold behind there”

“I wasn't able to get my boss interested in seriously considering any of them because he wanted to see significant prior experience doing ML development”



What Makes Data Science Work?

Data

Feature
Engineering

Modeling
Algorithms

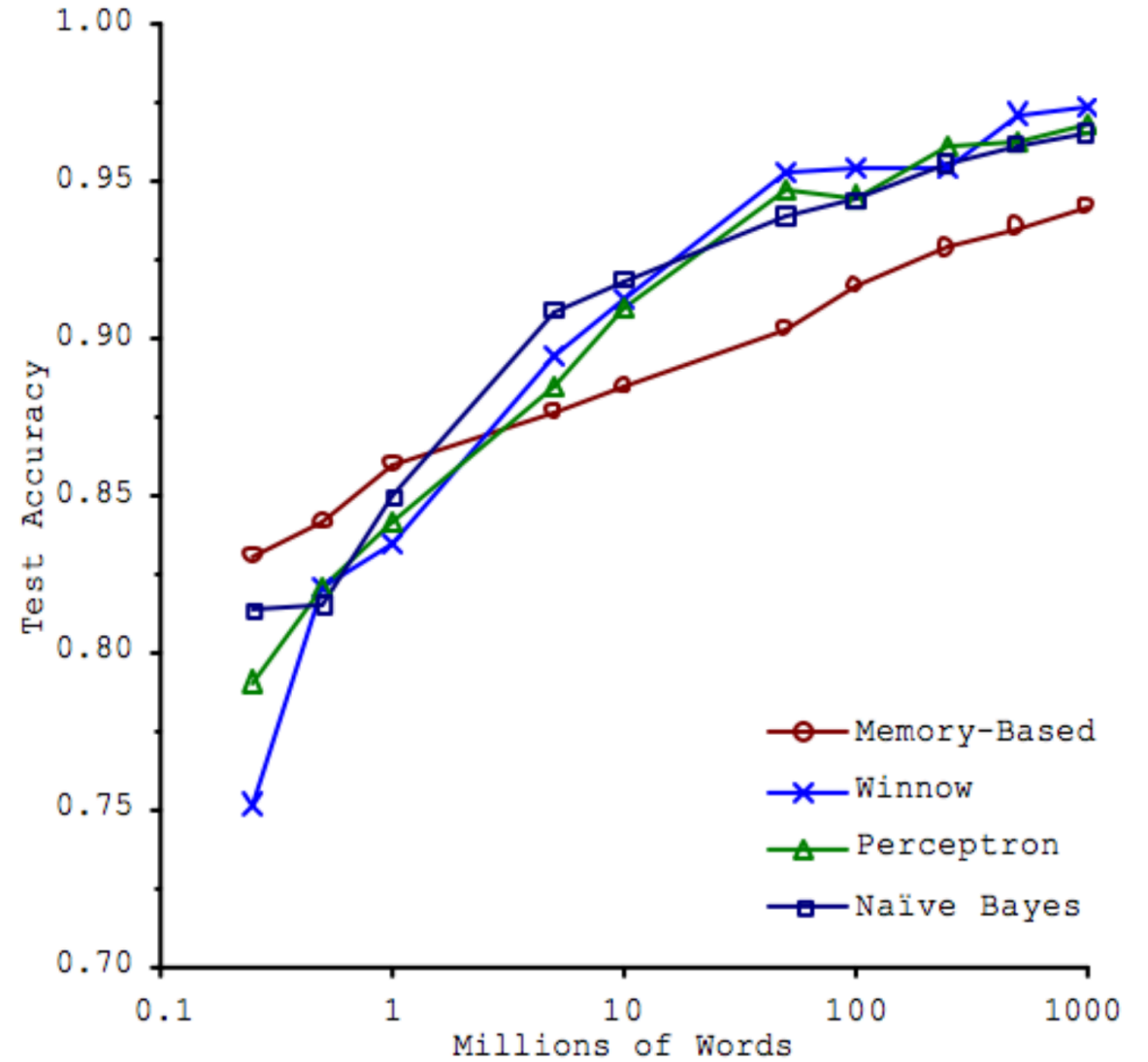
Statistics

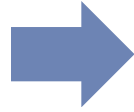
Product
Integration

Communication

Experience

It's All About The Data

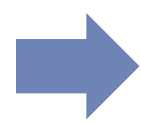




Perfect
Modeling
Technology



It's All About The Features



Perfect
Modeling
Technology



It's All About The Features



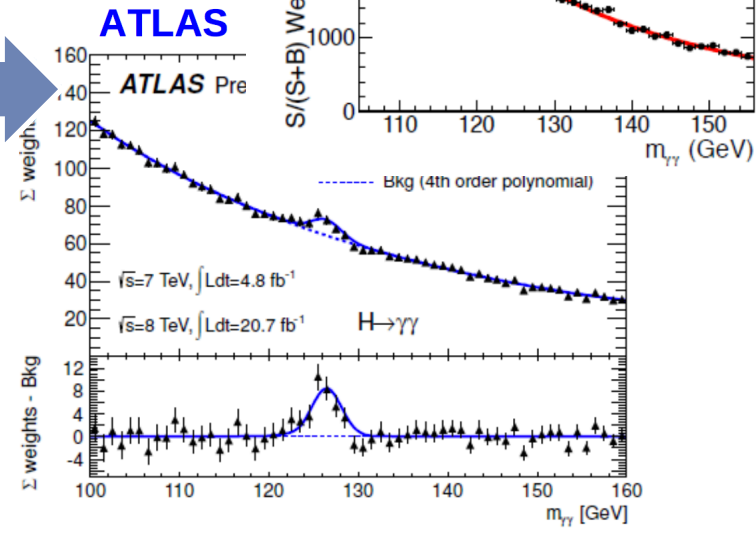
$$m_{jj}, \Delta\eta_{jj}, \eta_{j1}, \eta_{j2}, p_{Tt}$$

$$\Delta\phi_{\gamma\gamma;jj}, \eta^* = \eta_{\gamma\gamma} - \frac{\eta_{j1} + \eta_{j2}}{2}, \Delta R_{\min}^{\gamma j}$$

$$p_T^{\gamma}(1)/m_{\gamma\gamma}, p_T^{\gamma}(2)/m_{\gamma\gamma}, p_T^j(1), p_T^j(2)$$

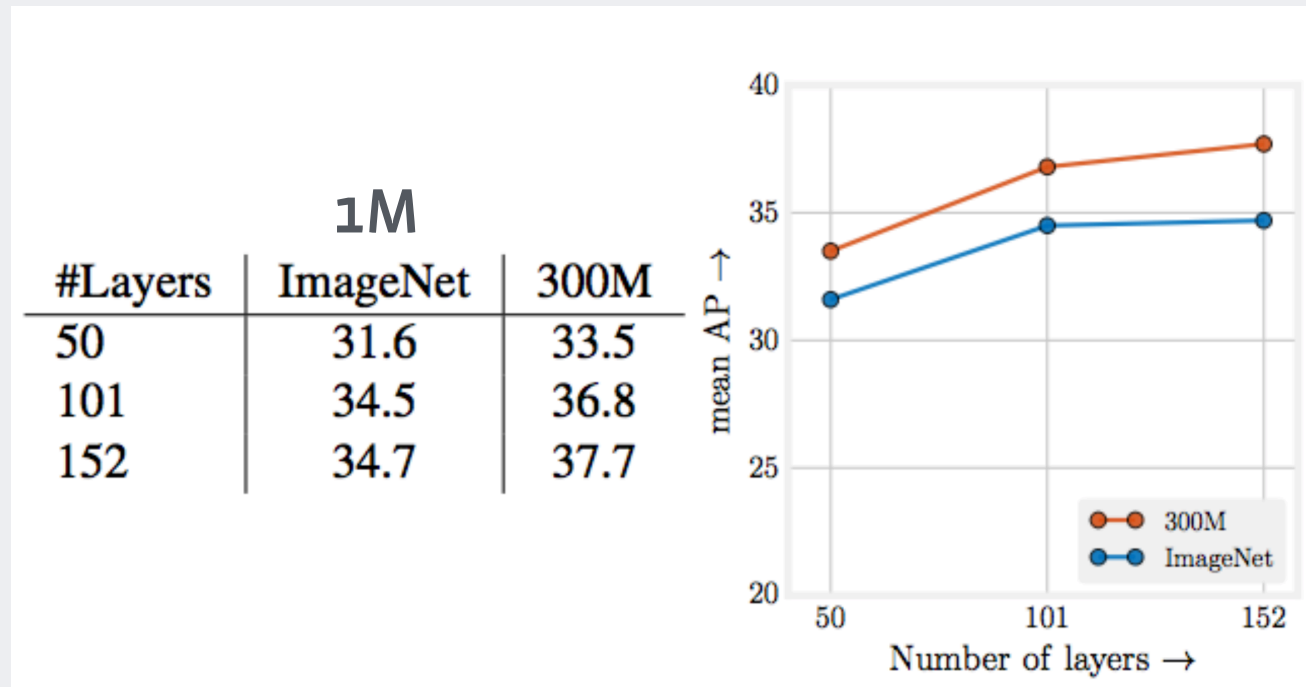
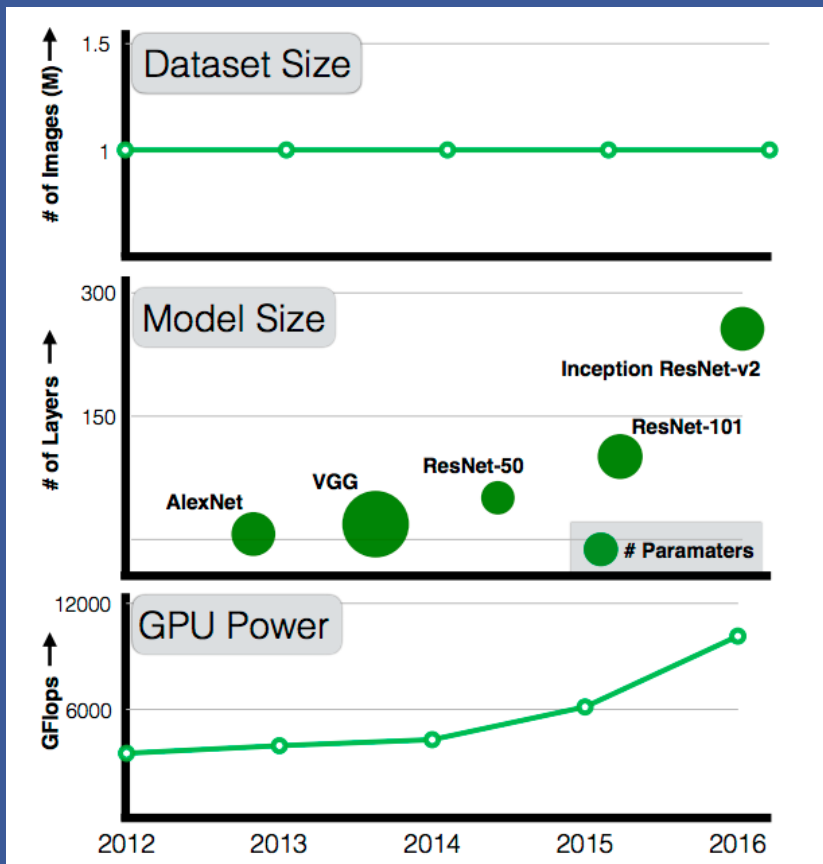
$$m_{jj}, \Delta\varphi_{jj,\gamma\gamma}, \Delta\eta_{jj,\gamma\gamma}$$

Perfect Modeling Technology

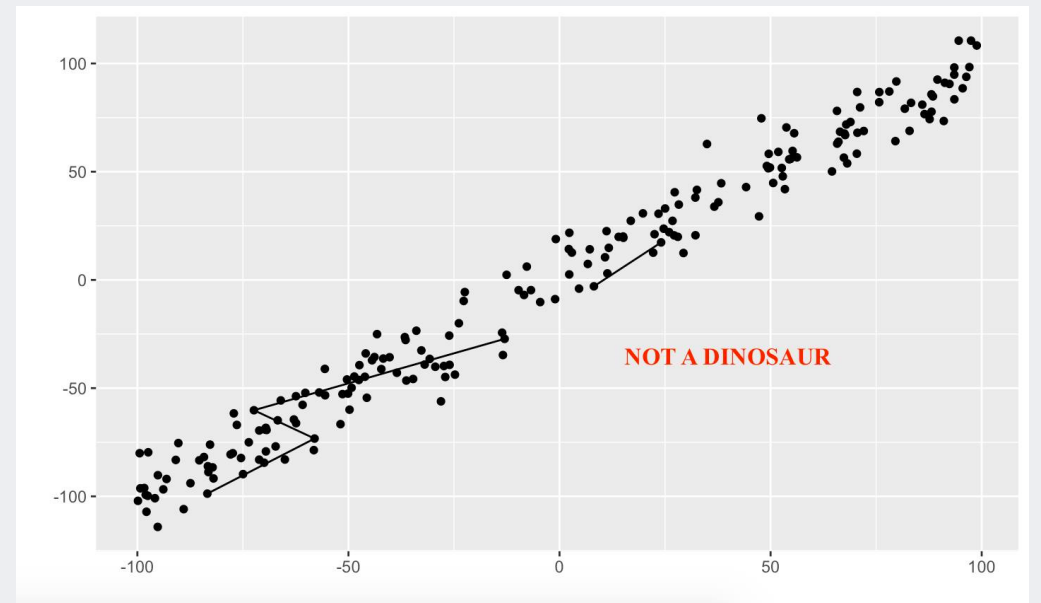
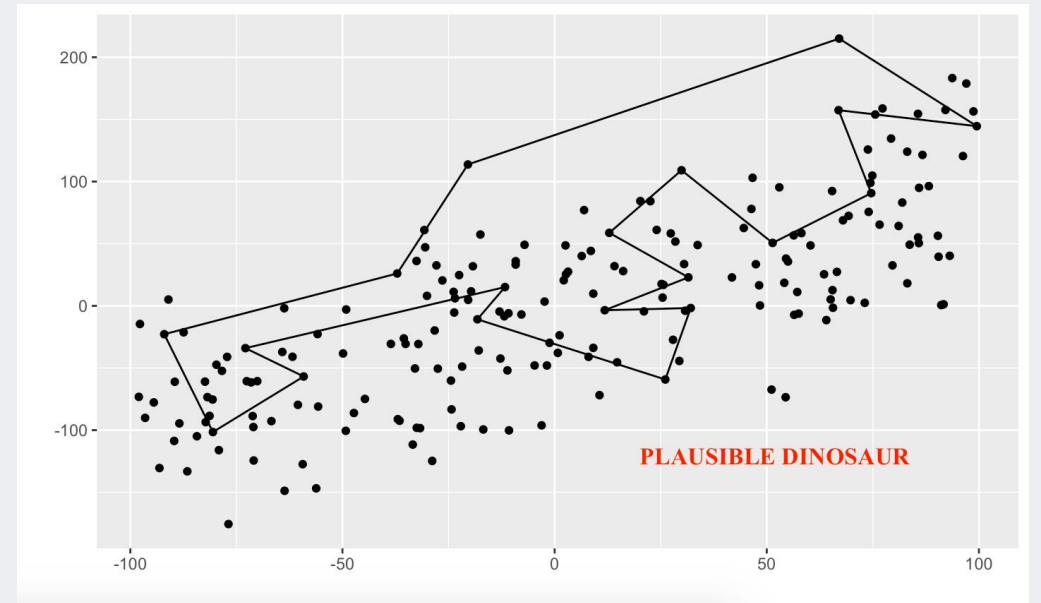
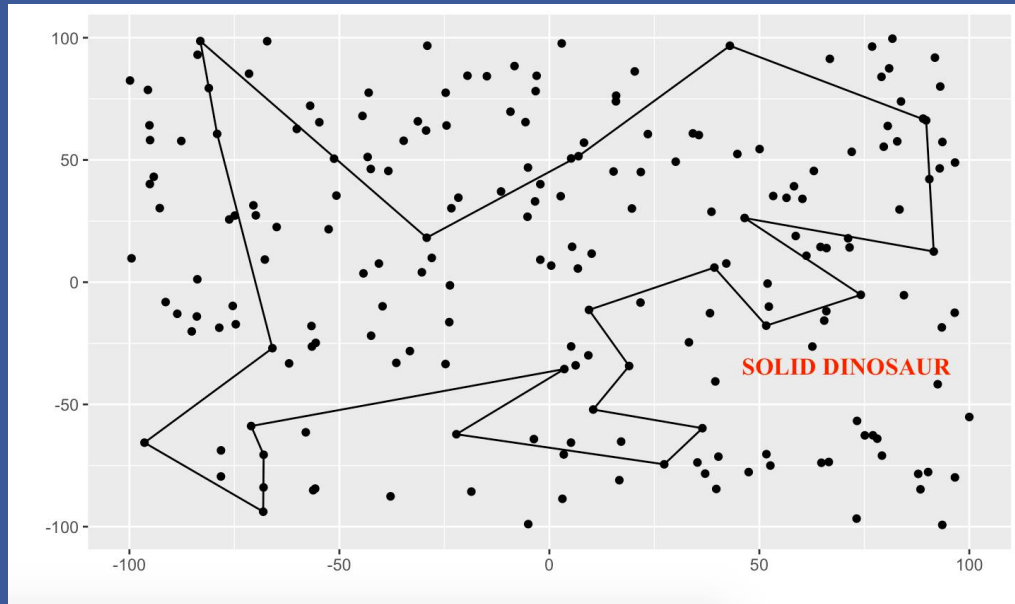


It's All About The Features

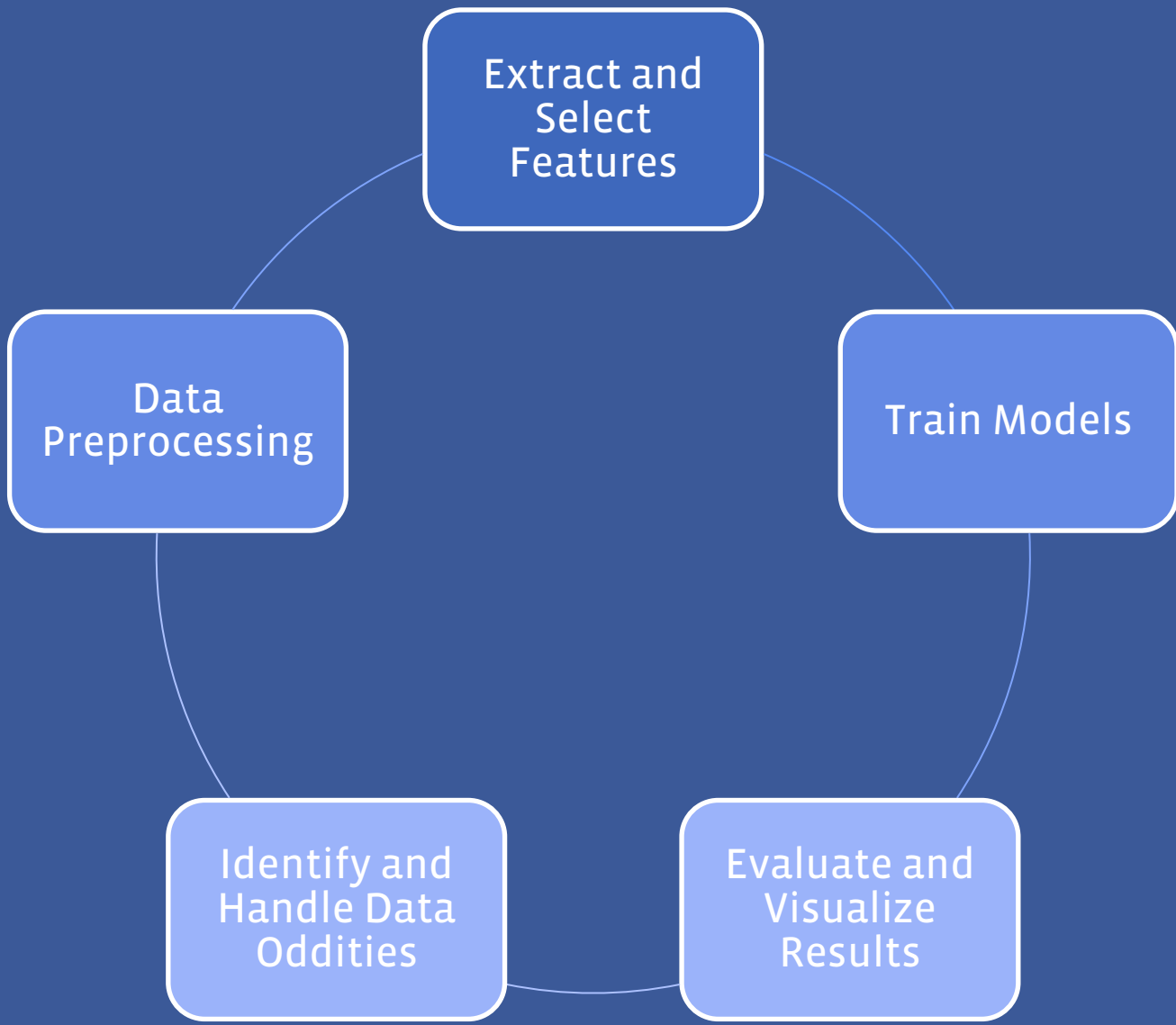
It's All About The Algorithms



It's All About Statistics



Put It All Together



1

Process

2

Visualize

3

Communicate

4

Drive Action

Data Scientists Tell Stories

What's your advice for undergraduate student who aspires to be a research scientist in deep learning or related field one day?



Yann LeCun, Director of AI Research at Facebook and Professor at NYU

Answered Jul 28, 2016 · Featured on Forbes and Quora Sessions's Twitter · Upvoted by Joaquin Quiñero Candela, [studied Machine Learning](#) and Nikhil Badugu, [M.S Computer Science & Machine Learning, Northeastern University](#)

- (0) take all the continuous math and physics class you can possibly take. If you have the choice between “iOS programming” and “quantum mechanics”, take “quantum mechanics”. In any case, take Calc I, Calc II, Calc III, Linear Algebra, Probability and Statistics, and **as many physics courses as you can**
But make sure you learn to program

↑ [-] [ylecun](#) [S] 22 points 3 years ago

↓ **Physics is about modeling actual systems and processes. It's grounded in the real world.** You have to figure out what's important, know what to ignore, and know how to approximate. These are skills you need to conceptualize, model, and analyze ML models.

Another set of courses that are relevant is signal processing, optimization, and control/system theory.

That said, taking math and statistics courses is good too.

Repetition, Repetition, Repetition

“basic foundation of understanding the data analysis portion of ML I developed in HEP”

“My data analysis skills come in to play in the more general sense of problem solving, analytical thinking, that kind of thing”

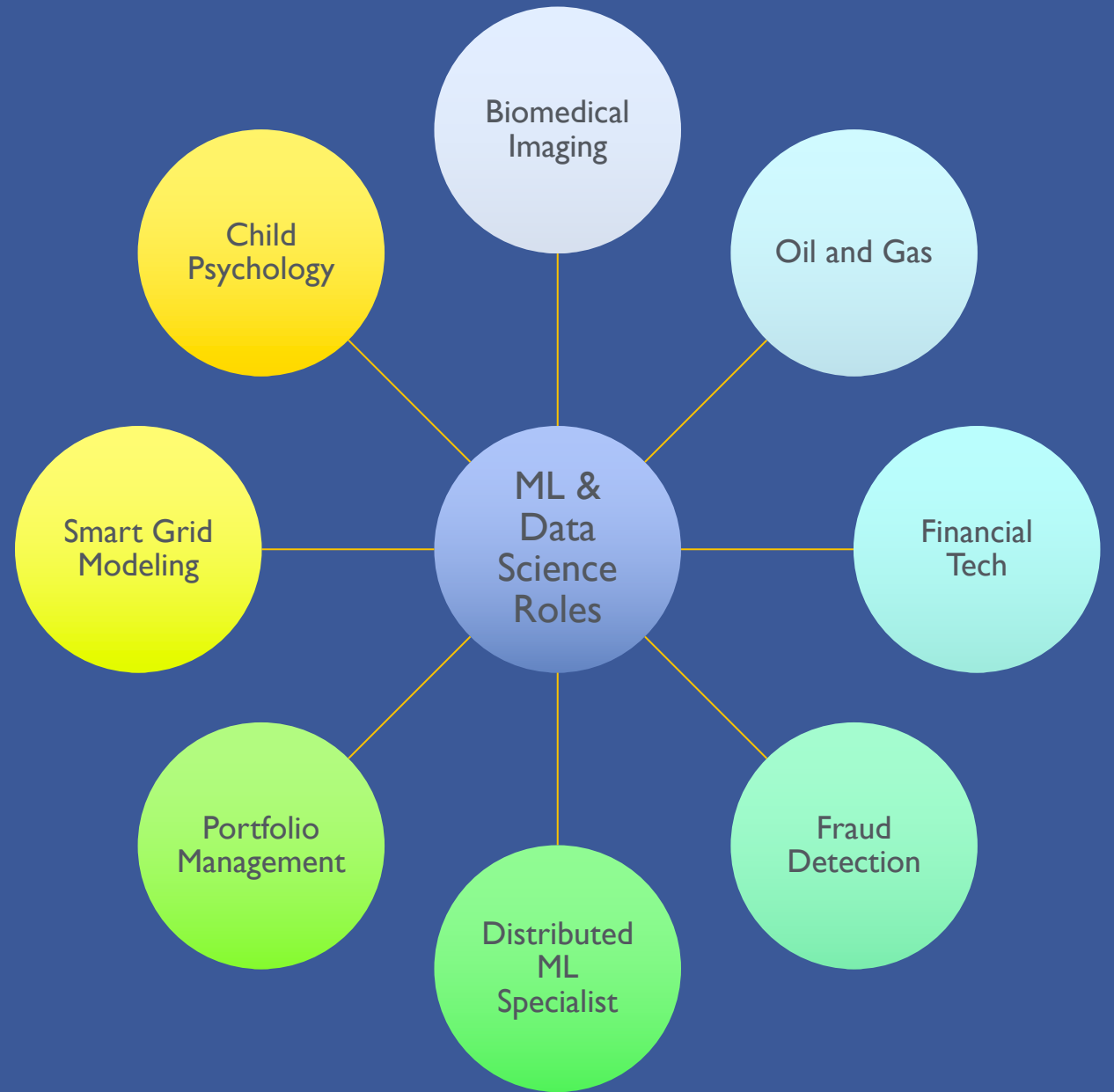
“I have learned most of my deep learning skill since I joined my company”

“I had some basic coding skills from HEP, some data vis stuff, and the ability to learn new things. That last one is the most important.”

“I coded in C++ but was never taught or shown best practices”

“being able to read and implement stuff off of arxiv is really useful”

“I didn't do any machine learning (even old style pre deep learning stuff) while I was in HEP”



What Makes Data Science Work?

Data

Feature
Engineering

Modeling
Algorithms

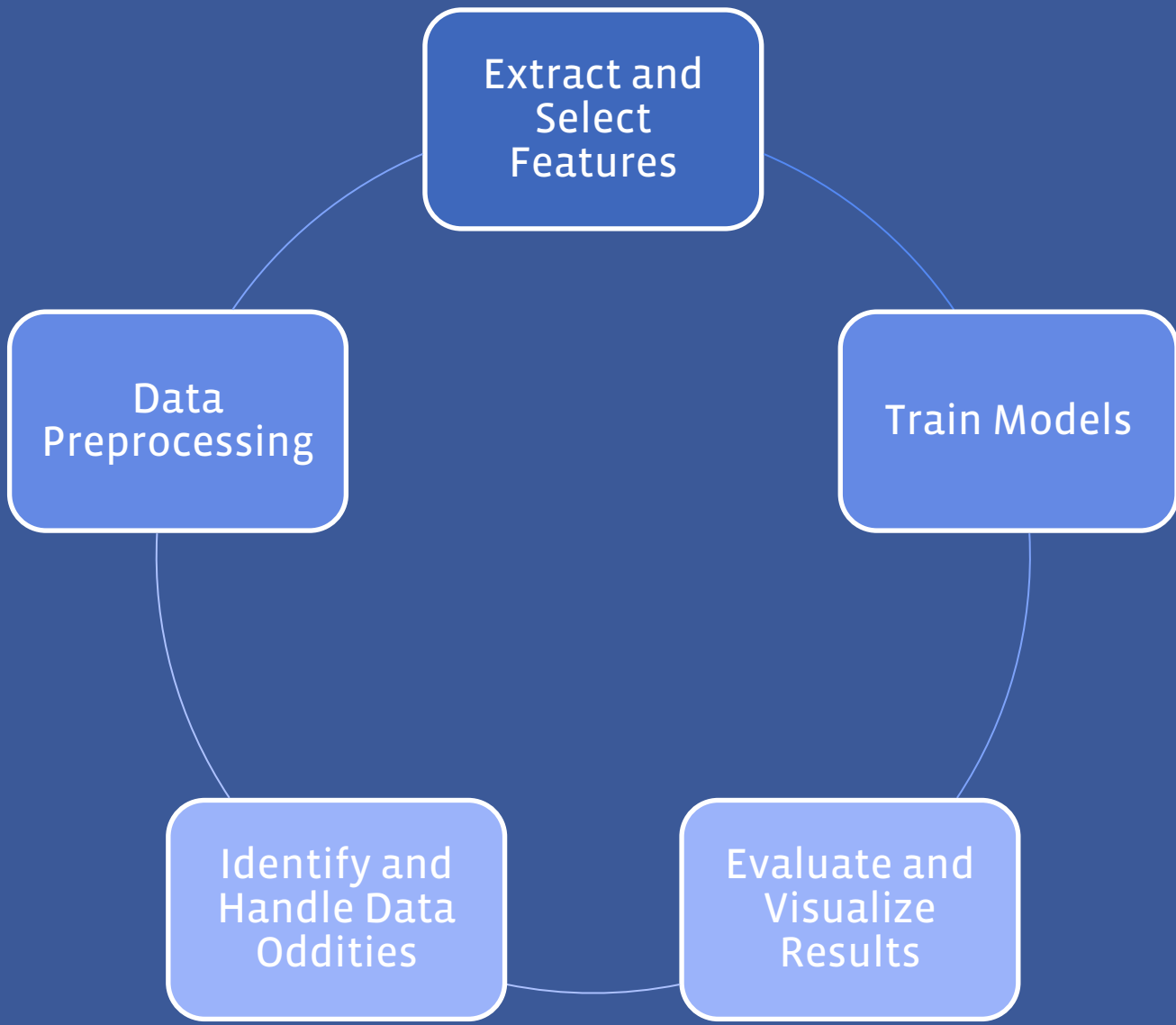
Statistics

Product
Integration

Communication

Experience

Put It All Together





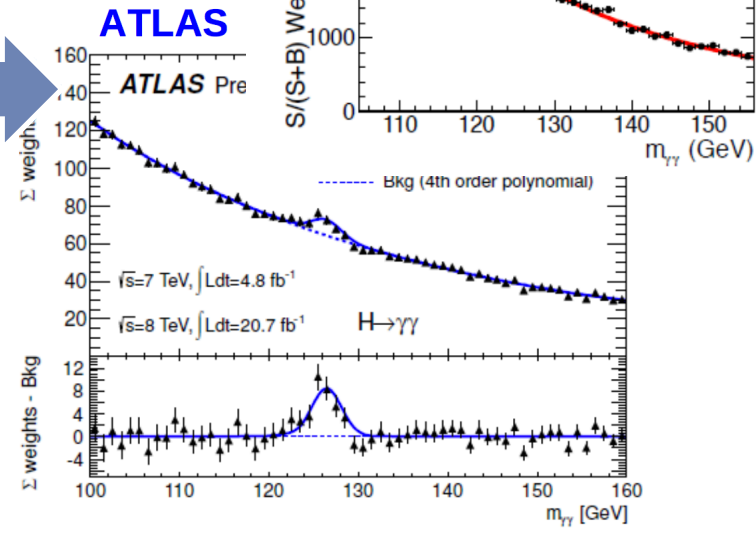
$$m_{jj}, \Delta\eta_{jj}, \eta_{j1}, \eta_{j2}, p_{Tt}$$

$$\Delta\phi_{\gamma\gamma;jj}, \eta^* = \eta_{\gamma\gamma} - \frac{\eta_{j1} + \eta_{j2}}{2}, \Delta R_{\min}^{\gamma j}$$

$$p_T^{\gamma}(1)/m_{\gamma\gamma}, p_T^{\gamma}(2)/m_{\gamma\gamma}, p_T^j(1), p_T^j(2)$$

$$m_{jj}, \Delta\varphi_{jj,\gamma\gamma}, \Delta\eta_{jj,\gamma\gamma}$$

Perfect Modeling Technology



It's All About The Features

Thank You

Emanuel Strauss
eman@fb.com