

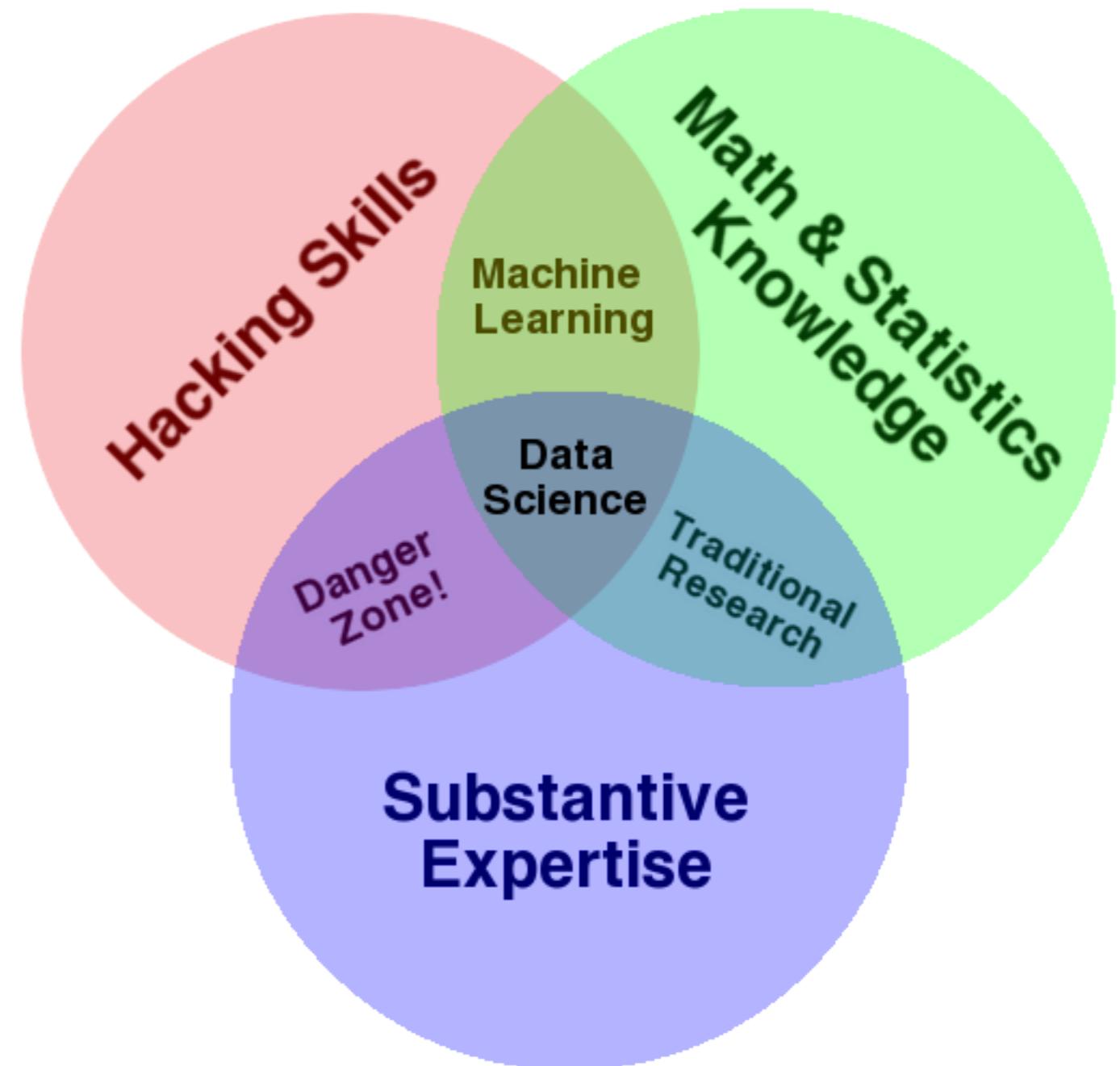
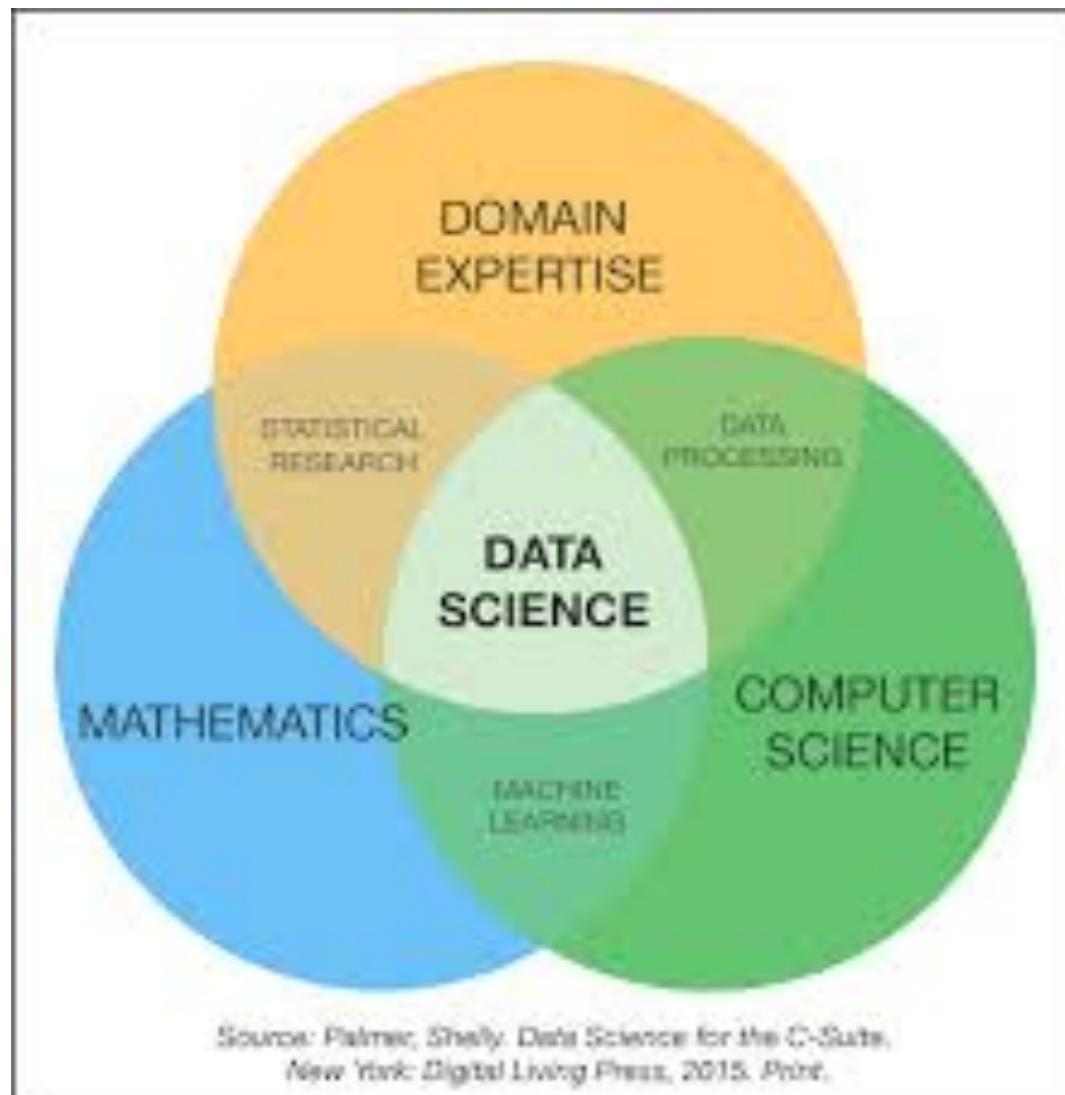
Introduction to Scientific Computing I

Amir Farbin

What is Data Science?

- From *Wikipedia*: Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms...
- There are concepts that unify statistics, data analysis, machine learning,
...
- Take techniques from mathematics, statistics, information science, and computer science
- Apply it to a domain:
 - science: biology, physics, psychology, ...
 - marketing, business analytics, ...

Data Science



Data Science @ UTA

- College of Science is in the process of getting approval for undergraduate BA and BS degrees in Data Science.
- Minor in Data Science possible now.
- Tentative New Courses:
 - DATA 1301: Introduction to Data Science
 - DATA 1401/2: Introduction to Scientific Computing 1 and 2
 - DATA 34xx: Data Mining, Management, and Curation
 - DATA 34xx: Statistical Modeling
 - DATA 34xx: Simulation
 - DATA 34xx: Machine Learning
 - DATA 43xx: Data Problems
 - DATA 43xx: Data Capstone Project 1 and 2

Logistics

- **Lectures:** Monday/Wednesday
 - No official Textbook. Unique class. Not following an existing paradigm.
 - Lecture and labs are essential.
 - Welcome to use any python book as a reference.
 - We will learn to use online docs as resource.
 - **Quizzes** (10%): Short and Easy. “Pop”: Randomly scheduled. Meant mostly as attendance tool.
 - Drop 2 lowest grades (including being sick, unless previously made arrangements).
- **Lab** (50%) Friday
 - Hands-on: I'll walk through something.
 - **Lab**: (15%) has to be accomplished in during the session (maybe end of day).
 - **Homework** (35%): Remainder is general is “Homework”. Due by following lab session.
- **Midterm** (20%) / **Final** (20%): Comprehensive... similar to homework. Take home. 24-48 hours. No makeups.

Logistics (2)

- Homework Policy
 - You can work with others, but do not copy/paste code from another student.
 - Submitted via git (Version Control System).
- Help
 - **Clinic:** Joe Cloud is the TA. Intend to have three, 3-hour sessions a week (TBA), where you can ask for help.
 - **Office Hours:** Open door policy. Generally at UTA MW before class 10 am - 2:30 pm. Best to let me know after class or e-mail.
 - Google group... <https://groups.google.com/forum/#!forum/uta-data-1401-spring-2020>
- **Laptop** (with a physical keyboard).
 - Highly recommended to have a laptop for this class
 - Doesn't matter what OS you run... all you need is a browser.
 - If you don't have a laptop, you can rent one. Details ...
- First lab will be NEXT Friday (1/31), will be to get setup.

Style

- Really a Python Programming Class taught from a Data Science perspective.
- Introduce and then reinforce data science concepts through successively more advanced implementations each introducing new programming techniques and culminating in use of advanced libraries.
- For example,
 - Learn about distributions and histogram
 - Example: Construct histograms
 - by hand
 - procedurally
 - write histogramming functions
 - write histogramming object oriented classes
 - use standard libraries that provide histogramming

Course Plan

- Class Intro
- What is Data Science?
- Conceptual Understanding of Computing
 - From Transistor to the iPhone (Hardware)
 - From Machine Language to Python (Programming)
 - From Operating Systems to Apps (Software)
- Fundamentals
 - Git
 - Basic Unix
 - Jupyter
- Python Introduction → Build your first game: Checkers / Tic-Tac-Toe
 - Data types / Syntax
 - Loops / Control Flow
 - Abstraction: Functions
 - Basic I/O
 - Modules / Libraries
 - → Build your first game: Checkers / Tic-Tac-Toe
- Linear Algebra
 - Basic algorithms
 - → Matrix Operations
- Object Oriented Programming (take 1)
 - → Matrix Class
- Probability Theory
 - Understanding Distributions
 - → Histograms and Monte Carlo / Functional Programming
- Object Oriented Programming (take 2)
 - Scripting vs Building Software
 - OO Design Patterns / UML
 - → Gradebook Software
- Data Analysis
 - Introduction to: numpy, Pandas, matplotlib
 - Representing Data
 - Visualization
 - → Basic Data Analysis
- Machine Learning
 - scikit
 - → LHC Data Analysis

What do I do ?

Was the Universe an Accident?

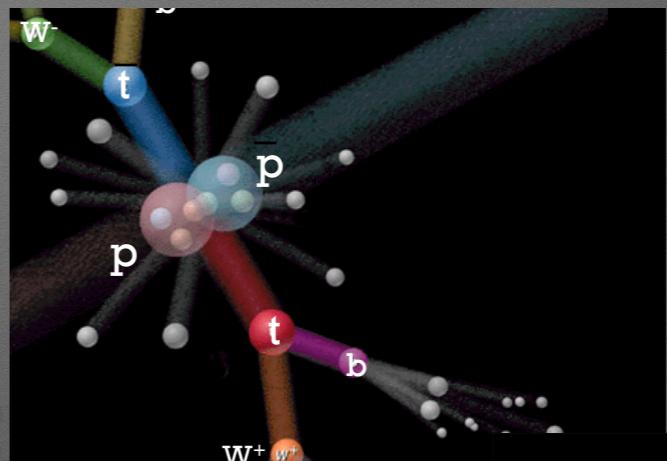
*Artificial Intelligence may find the answer in
data from the Large Hadron Collider*

Amir Farbin

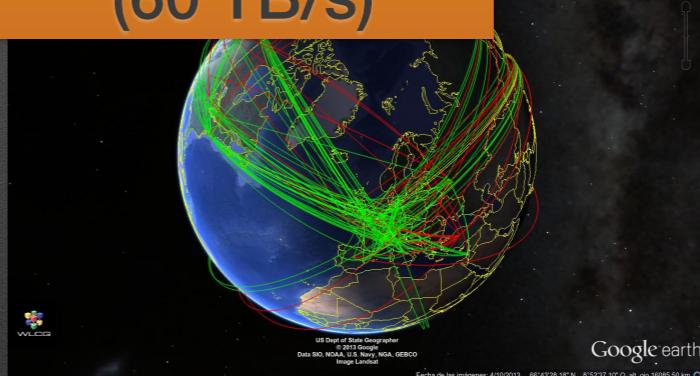


What is HEP ?

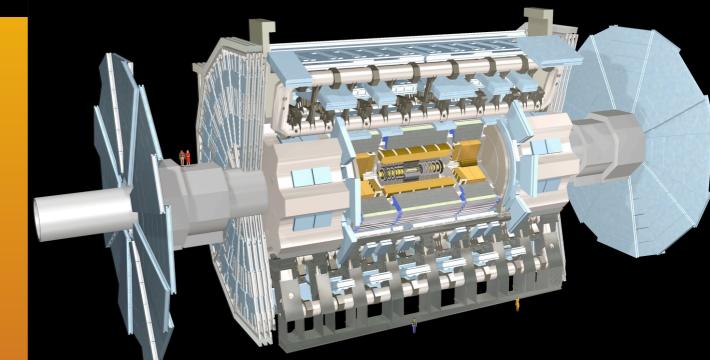
Large Hadron Collider (LHC)



Record with 5
Story
100M Channel
“Camera”
(60 TB/s)



Create New
Particles
Only 1 in 10^{11}
are “interesting”



Processed by
300k Cores
Around the
World

Higgs Discovery - Nobel Prize Physics 2013

Physics Letters B 716 (2012) 1–29

Contents lists available at SciVerse ScienceDirect

Physics Letters B

www.elsevier.com/locate/physletb

 ELSEVIER

Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC[☆]

ATLAS Collaboration*
This paper is dedicated to the memory of our ATLAS colleagues who did not live to see the full impact and significance of their contributions to the experiment.

ARTICLE INFO

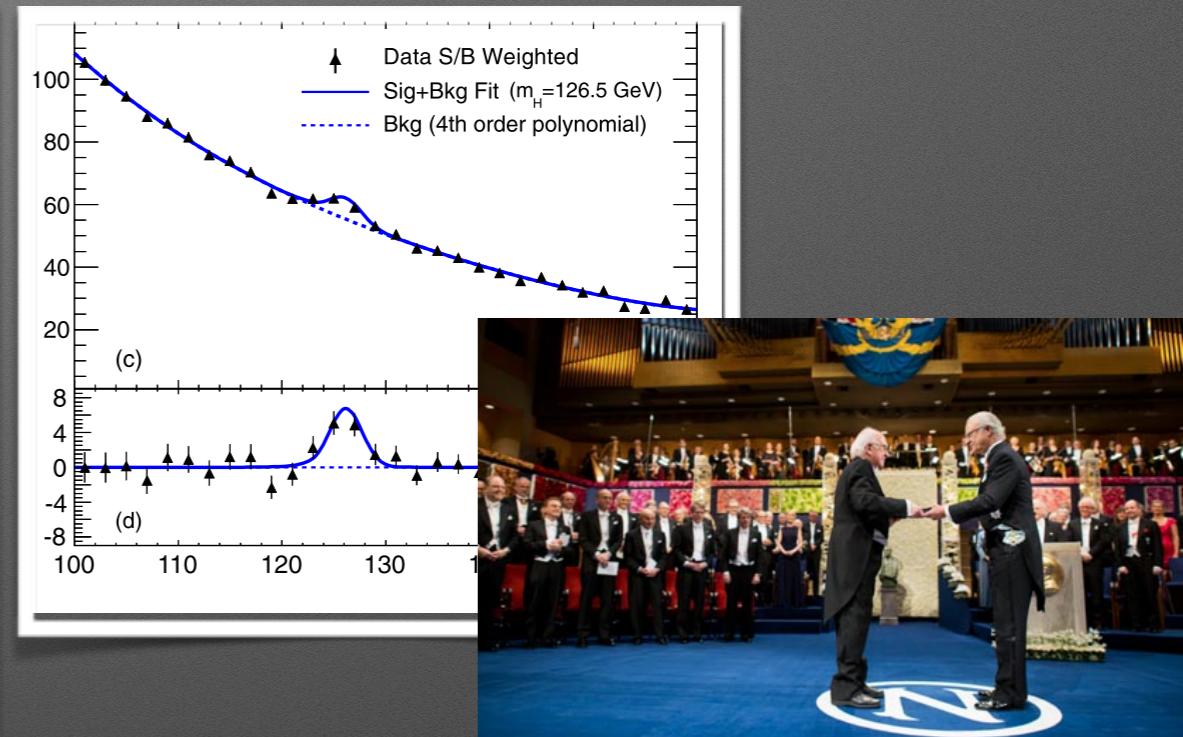
Article history:
Received 31 July 2012
Received in revised form 8 August 2012
Accepted 11 August 2012
Available online 14 August 2012
Editor: W.-D. Schlatter

ABSTRACT

A search for the Standard Model Higgs boson in proton–proton collisions with the ATLAS detector at the LHC is presented. The datasets used correspond to integrated luminosities of approximately 4.8 fb^{-1} collected at $\sqrt{s} = 7 \text{ TeV}$ in 2011 and 5.8 fb^{-1} at $\sqrt{s} = 8 \text{ TeV}$ in 2012. Individual searches in the channels $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$, $H \rightarrow \gamma\gamma$ and $H \rightarrow WW^{(*)} \rightarrow e\nu\mu\nu$ in the 8 TeV data are combined with previously published results of searches for $H \rightarrow ZZ^{(*)}$, $WW^{(*)}$, $b\bar{b}$ and $\tau^+\tau^-$ in the 7 TeV data and results from improved analyses of the $H \rightarrow ZZ^{(*)} \rightarrow 4\ell$ and $H \rightarrow \gamma\gamma$ channels in the 7 TeV data. Clear evidence for the production of a neutral boson with a measured mass of $126.0 \pm 0.4 \text{ (stat)} \pm 0.4 \text{ (sys)} \text{ GeV}$ is presented. This observation, which has a significance of 5.9 standard deviations, corresponding to a background fluctuation probability of 1.7×10^{-9} , is compatible with the production and decay of the Standard Model Higgs boson.

© 2012 CERN. Published by Elsevier B.V. Open access under CC BY-NC-ND license.

**Last Piece of the Standard Model
Best Tested Theory... Ever.**



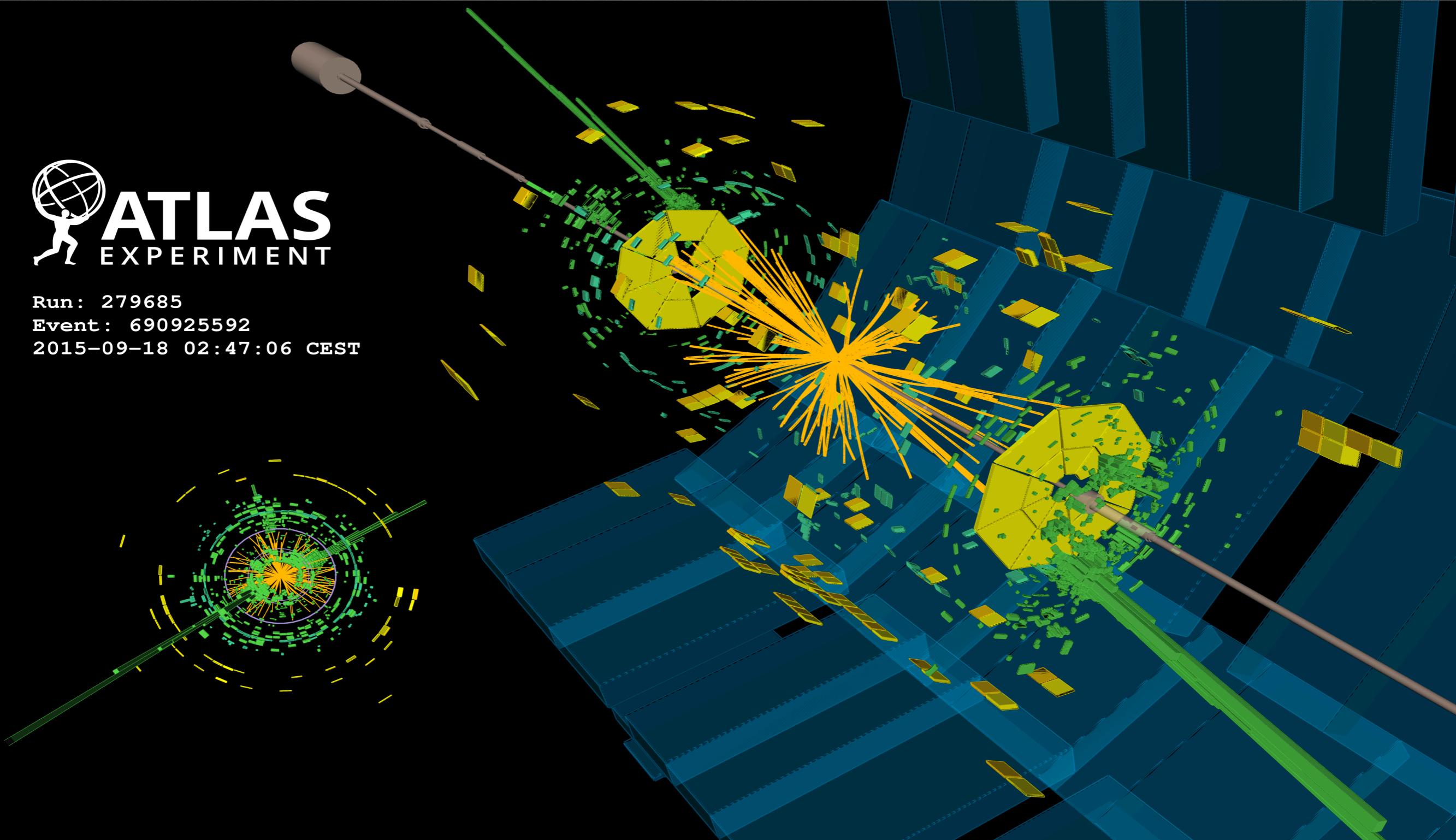
Not Done... Higgs is Light! Possibilities:

- *Fine-tuned Theory: Accident or Multiverse + Anthropic Principle*
- *Mechanism: Supersymmetry, Extra-Dimensions, Sub-structure*
 - Focus of LHC
- *Design?*

Deep Learning in High Energy Physics



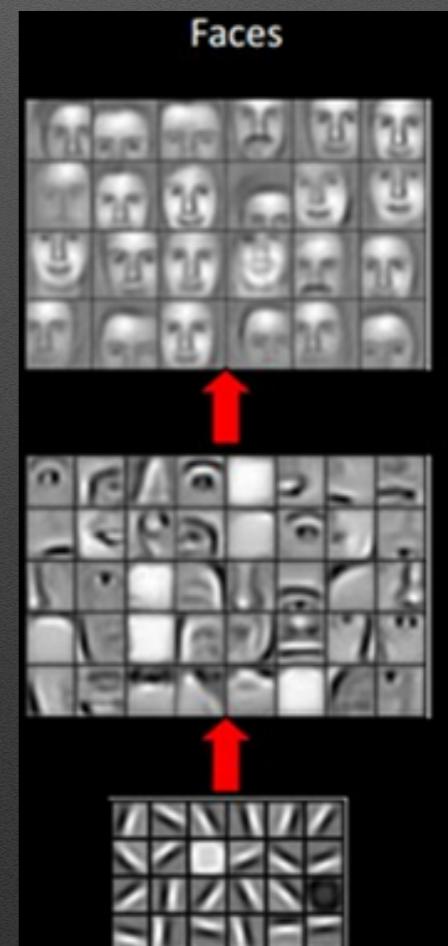
Run: 279685
Event: 690925592
2015-09-18 02:47:06 CEST



- Requires lots of computing
- Upgrade to LHC will give us 100x the data.
 - We won't have 100x the computing power or storage.
- Use Artificial Intelligence and newest processors...

Animal Brains

- The brain takes in sensory data... *builds hierarchical models of the world.*
- So effectively, a *representation* of the input is assembled in the brain.
 - Eyes see a limited window... but...
 - Location Cells
 - Imagining locations



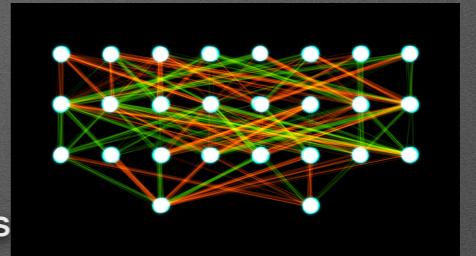
Brief History of AI

Artificial Intelligence

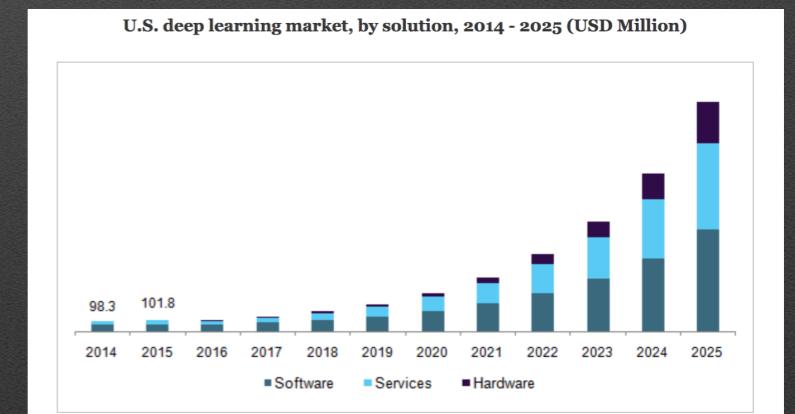
- Goal: Systems that reason and act as well as or better than humans
- Heuristic AI (1990's)
- Machine Learning AI
 - Knowledge learned from data
 - Neural Networks ~ Brain inspired computing (1943)
 - Universal Computation Theorem (1989)
 - Multi-layer hidden networks (a.k.a. Deep) (1965)
 - Vanishing Gradient Problem (1991)

Deep Learning Renaissance (> 2007 - now)

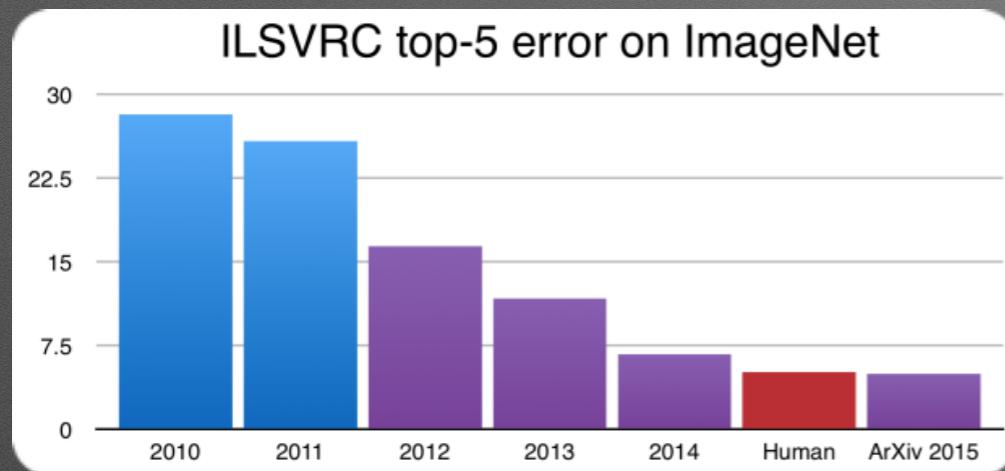
- Driven by:
 - New NN Innovation
 - Big Data
 - Graphical Processing Units
- Amazing Feats



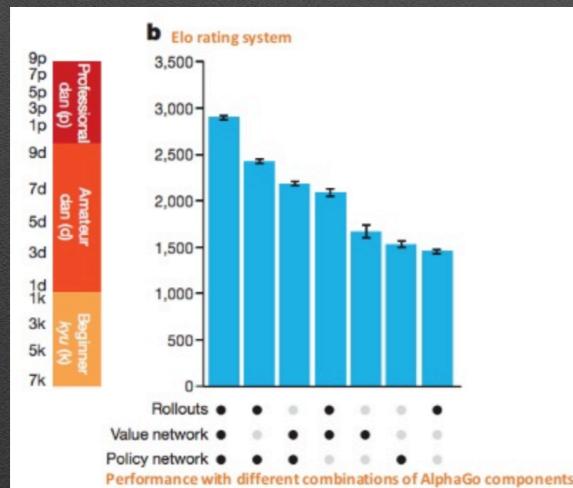
- Market Growth
- Industry Adoption



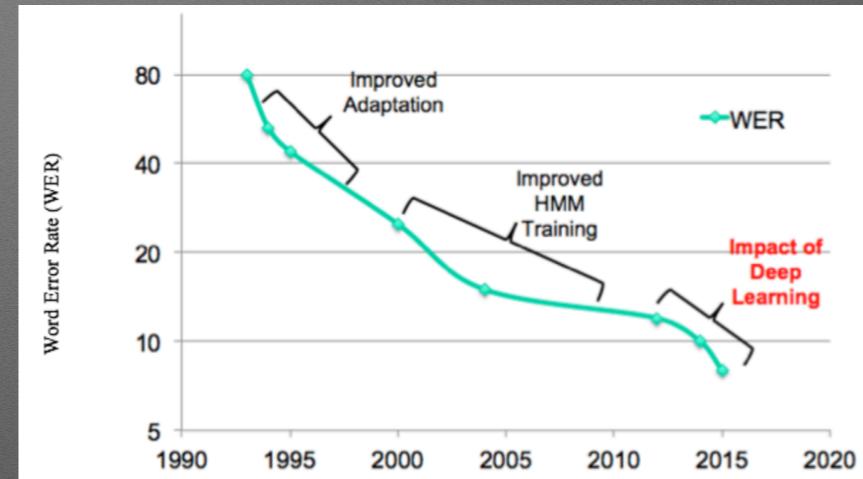
Amazing Feats : Some Examples



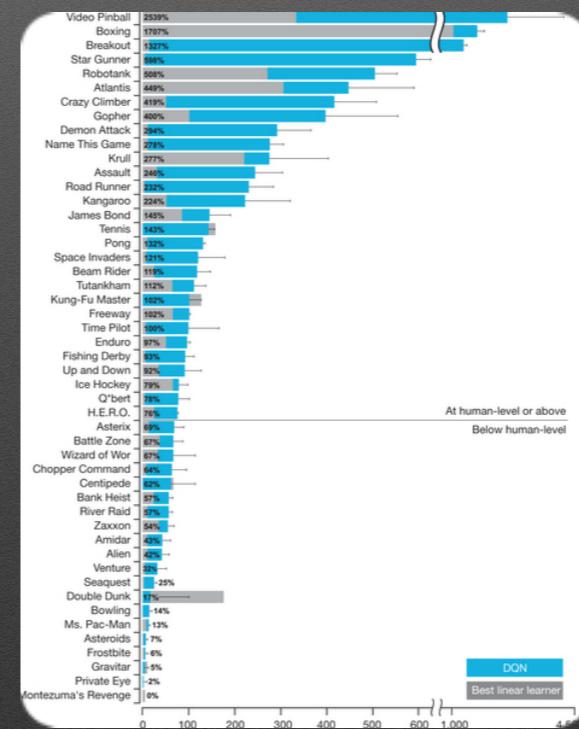
ImageNet Outperforms humans



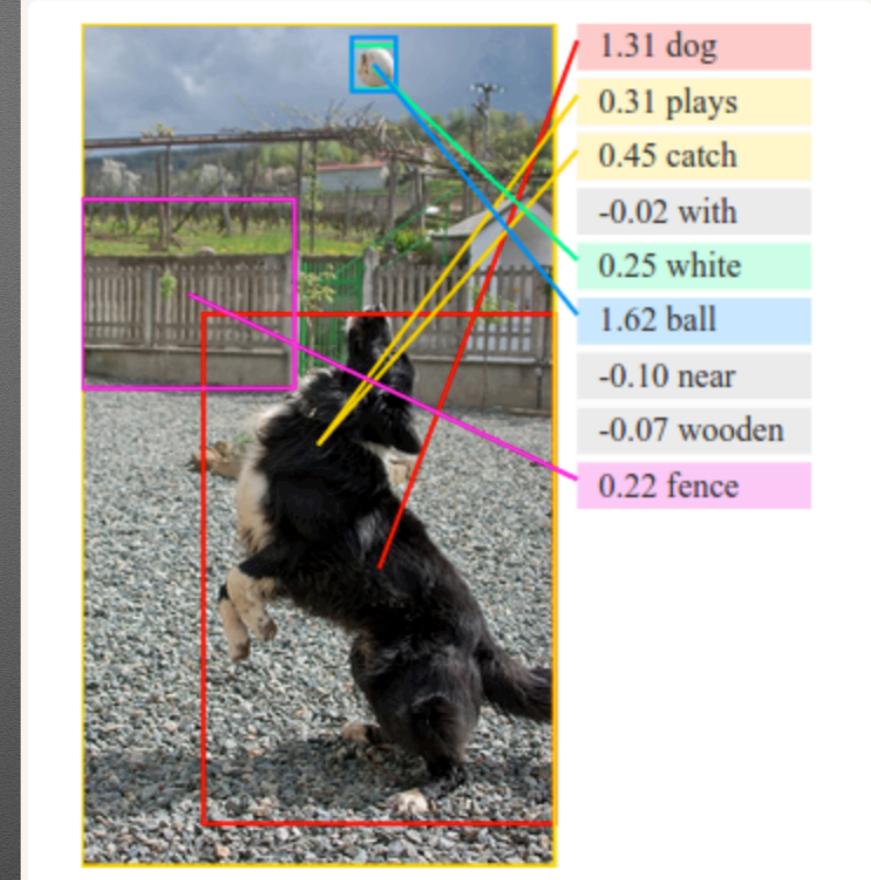
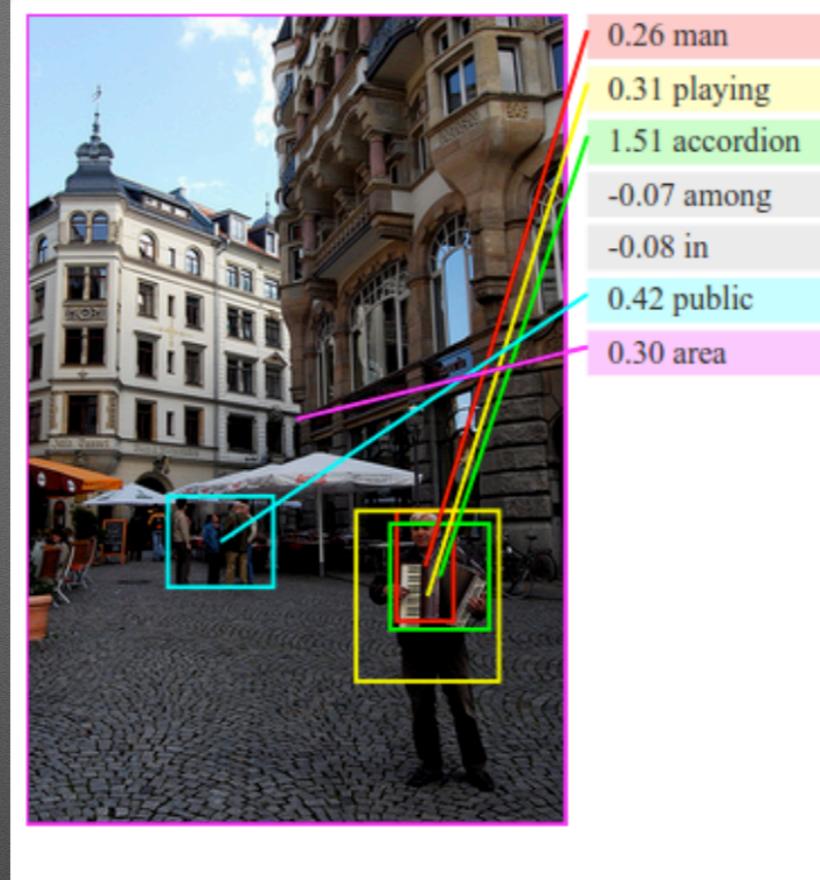
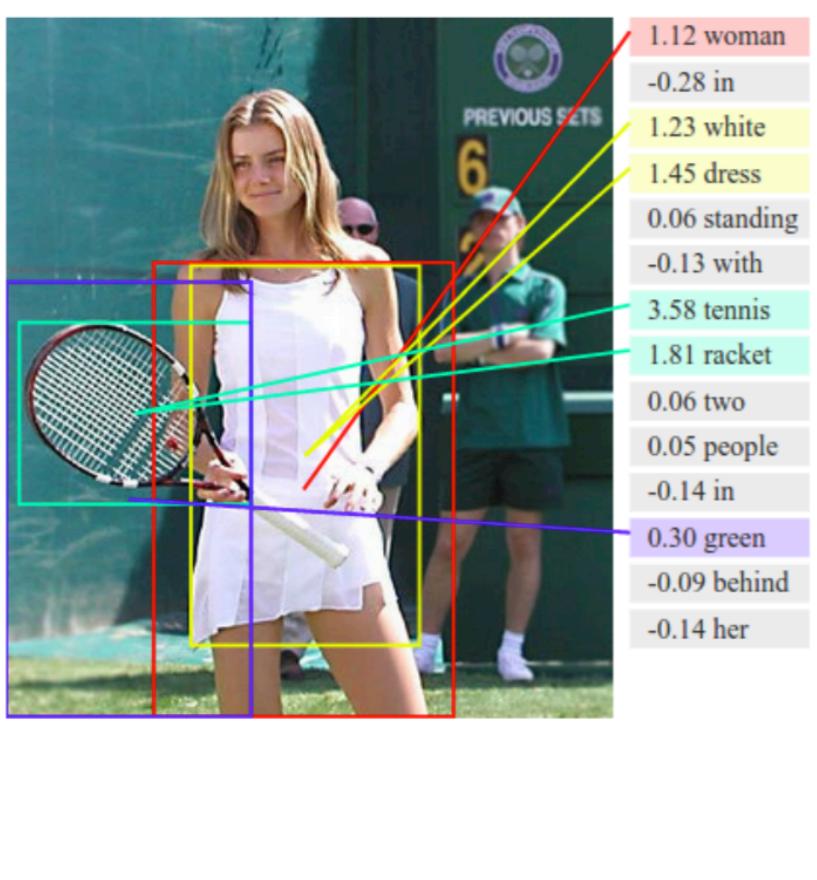
AlphaGo beats
Lee Sedol



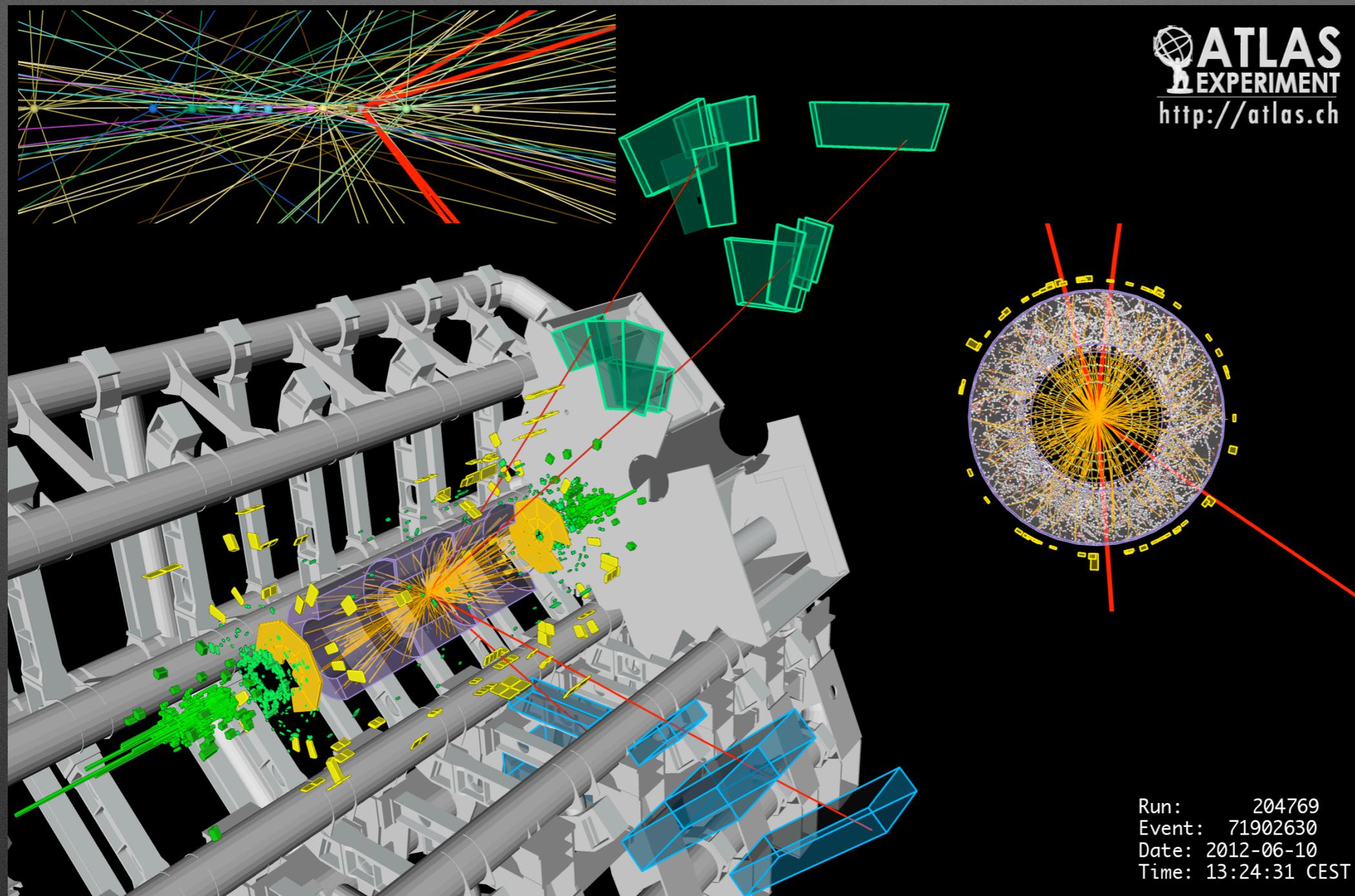
Rapid
Advances in
Speech
Recognition



Human Level
control in playing
Atari games



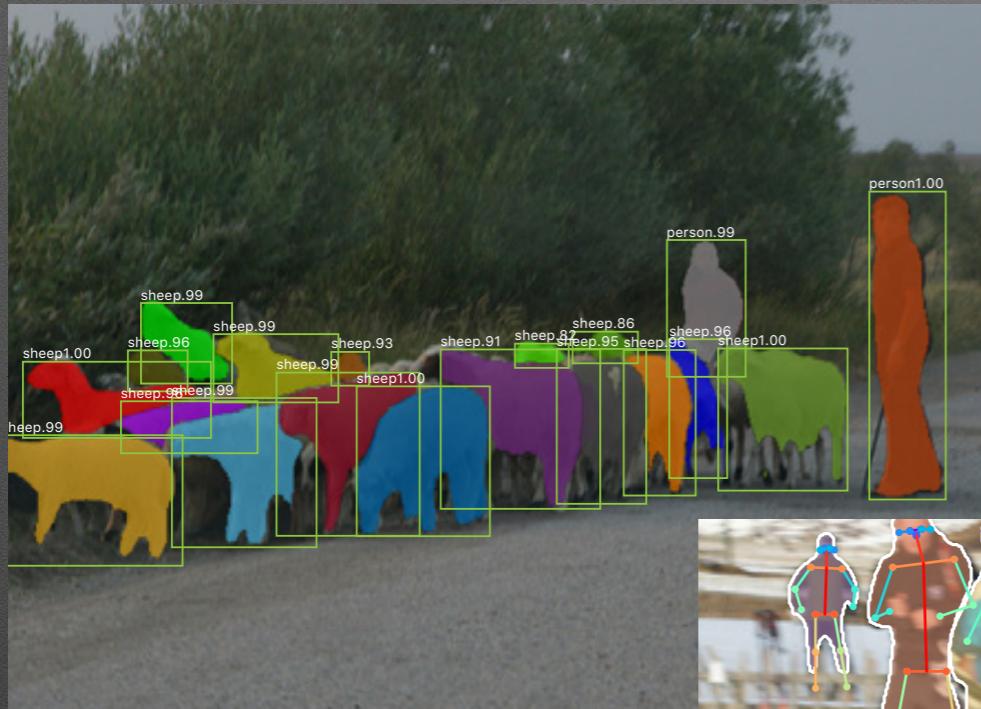
Deep Learning can be used to tell a story with context
through data
Used in HEP to understand what physics is happening



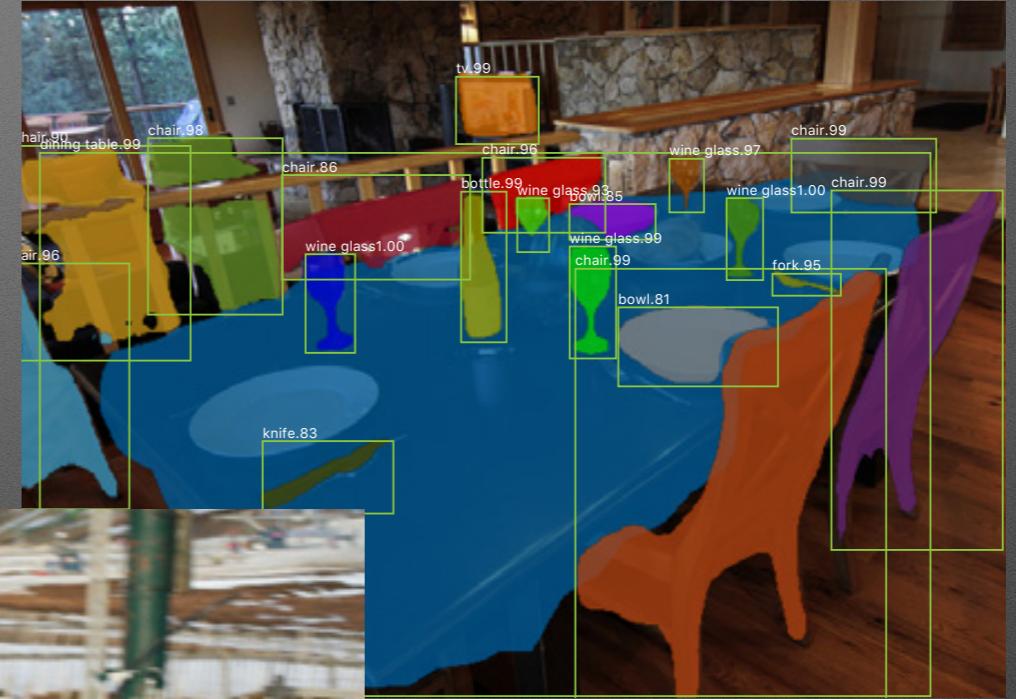
$H \rightarrow ZZ \rightarrow 4\ell$

We can use deep learning object tagging techniques in the data to find the decaying particle

Amazing Feats : Some More Examples



<https://arxiv.org/pdf/1703.06870.pdf>



Why go Deep?

Better Algorithms

- Better results
- Solution where there is none
- Make sense of complicated data

Easier Development

- Feature Learning, not Feature Engineering
- Save time and cost

Faster Algorithms

- DNNs Faster than traditional Algs
- Neuromorphic processors

Why Physicists ?

High Energy Physicists (HEP) ideally suited

- HEP Systems and Machine Learning and Deep Learning Systems confront similar challenges
- Decades of Experience at the Data Frontier
- Bridge between science and industry
- HEP scientists are also engineers by training

MOVE OVER, CODERS— PHYSICISTS WILL SOON RULE SILICON VALLEY

... it's happening across Silicon Valley., *the things that just about every internet company needs to do are more and more suited to the skill set of a physicist.*

new wave of data science and AI is something that suits physicists right down to their socks.

"There is something very natural about a physicist going into machine learning ... more natural than a computer scientist."

Physicists know how to handle data ... building these enormously complex systems requires its own breed of abstract thought.

Abstraction

- Definitions:
 - the quality of dealing with ideas rather than events
 - the process of considering something independently of its associations, attributes, or concrete accompaniments.
- Usage: (wikipedia)
 - *Computer scientists* use abstraction to make models that can be used and re-used without having to re-write all the program code for each new application on every different type of computer.
 - Abstraction in *mathematics* is the process of extracting the underlying essence of a mathematical concept, removing any dependence on real world objects with which it might originally have been connected, and generalizing it so that it has wider applications or matching among other abstract descriptions of equivalent phenomena.
- Imagine zooming out from a picture, start from individual spec of paint ... all the way to the full picture.
- Imagine a cooking recipe...