

Introduction to Scientific Computing II

Amir Farbin

Lecture 1

A Data Science BS Degree

- The Degree Proposal was in progress for 3 years.
- **Approved by Texas Higher Education Coordinating Board (October 22, 2020)**
 - Full program launch Fall 2021.
- Courses available since Fall 2018.
- *Minor* defined Fall 2020.
- Unique Degree
 - Undergraduate
 - Most programs are professional masters or PhD.
 - Within *College of Science*
 - Most are in Computer Science or Business
 - Requires concentration and Capstone Project
- Aim to prepare students:
 - Entry-level Data Science jobs
 - Better Science Research
 - Undergraduate
 - Better positioned for Graduate School
 - Stronger Application
 - Start on research quicker

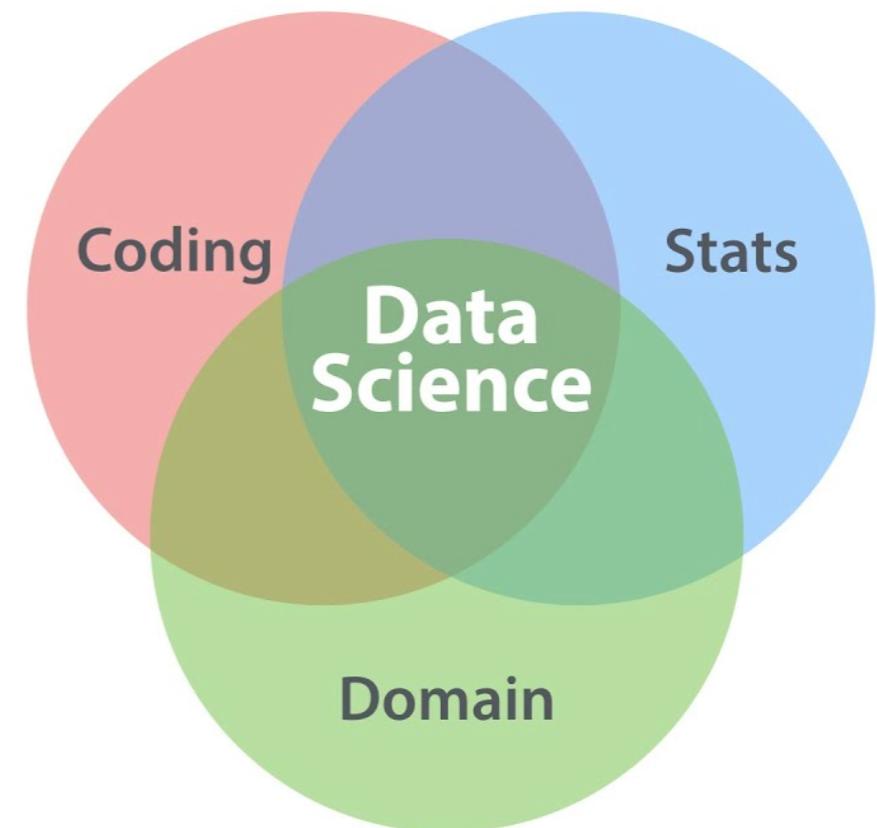


The UTA Data Science faculty

<https://www.uta.edu/science/data-science/>

The Challenge

- Practicing Data Science requires
 - Coding
 - Math (e.g. Statistics)
 - Domain expertise (e.g. physics, biology, ...)
- Each of these areas can be a degree onto itself...
 - Usually people come into data science from one of these areas.
- *Challenge:* Start with a University Freshman with no preparation.



The Degree

- Choose a concentration: Biology, Physics, Math, Psychology, Chemistry, Earth and Environmental Science, Geology
 - Note courses are being renumbered/rename for Fall 2021. Pending UCC approval:
 - **DATA 1301** – Introduction to Data Science
 - **DATA 3311** (new course) – Mathematics for Data Science
 - **DATA 3401** (formerly 1401) -- Python for Data Science 1
 - **DATA 3402** (formerly 1402) -- Python for Data Science 2
 - **DATA 3421** (formerly 3401) -- Data Mining, Management, and Curation
 - **DATA 3441** (formerly 3402) -- Statistical Methods for Data Science 1
 - **DATA 3442** (formerly 3403) -- Statistical Methods for Data Science 2
 - **DATA 3461** (formerly 3404) -- Machine Learning
 - **DATA 4380** (formerly 4301) -- Data Problems
 - **DATA 4381** (formerly 4302) -- Data Capstone Project 1
 - **DATA 4382** (formerly 4303) -- Data Capstone Project 2
 - Major: core + these courses + 2 math courses (and calculus I)
 - Minor: ~ 5 courses
- The DS BS Degree requirements are as follows:
- UTA Core curriculum (46 hours)
 - Data Science Foundations (18 hours)
 - Student Success
 - Intro Data Science
 - Python 1 and 2
 - Statistical Inference
 - Linear Algebra/Probability
 - Core Data Science courses (16 hours)
 - Statistical Methods 1 and 2
 - Data management
 - Machine Learning
 - Data Science Capstone courses (9 hours)
 - Data Problems
 - Capstone 1 and 2
 - Domain specific courses (23 hours)
 - Domain Concentration Specific Requirements
 - Science elective with lab (8 hours)
 - Total = 120 hours

Introduce Yourself

- Your UTA Degree
 - What is your major?
 - What year?
 - When will you graduate?
- Interests
 - Is there a specific scientific or professional field?
 - Have you done any research?
 - Any hobbies, etc, that you can apply DS to ...?
- Your goals
 - What's next (job, grad school)?
 - How can this course help?
- Your setup
 - What kind of computer? (Windows, Mac, Linux)
- Anything else...

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 ?

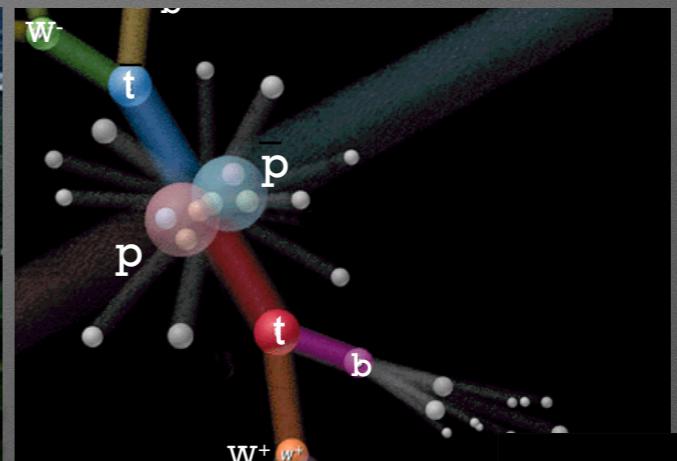
Large Hadron Collider (LHC)



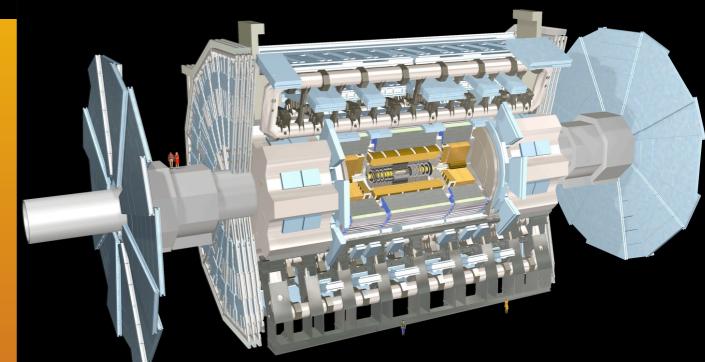
Largest Machine Ever Built



10^{11} Protons Collide 40 Million Times per Second

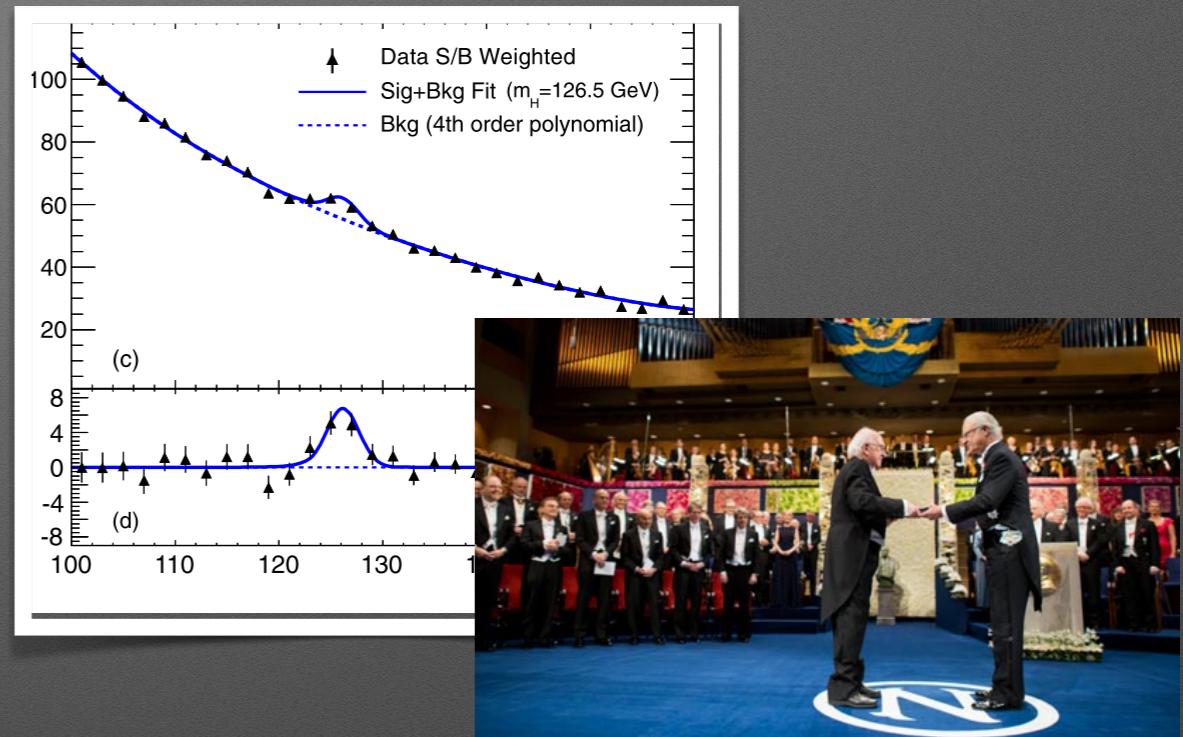
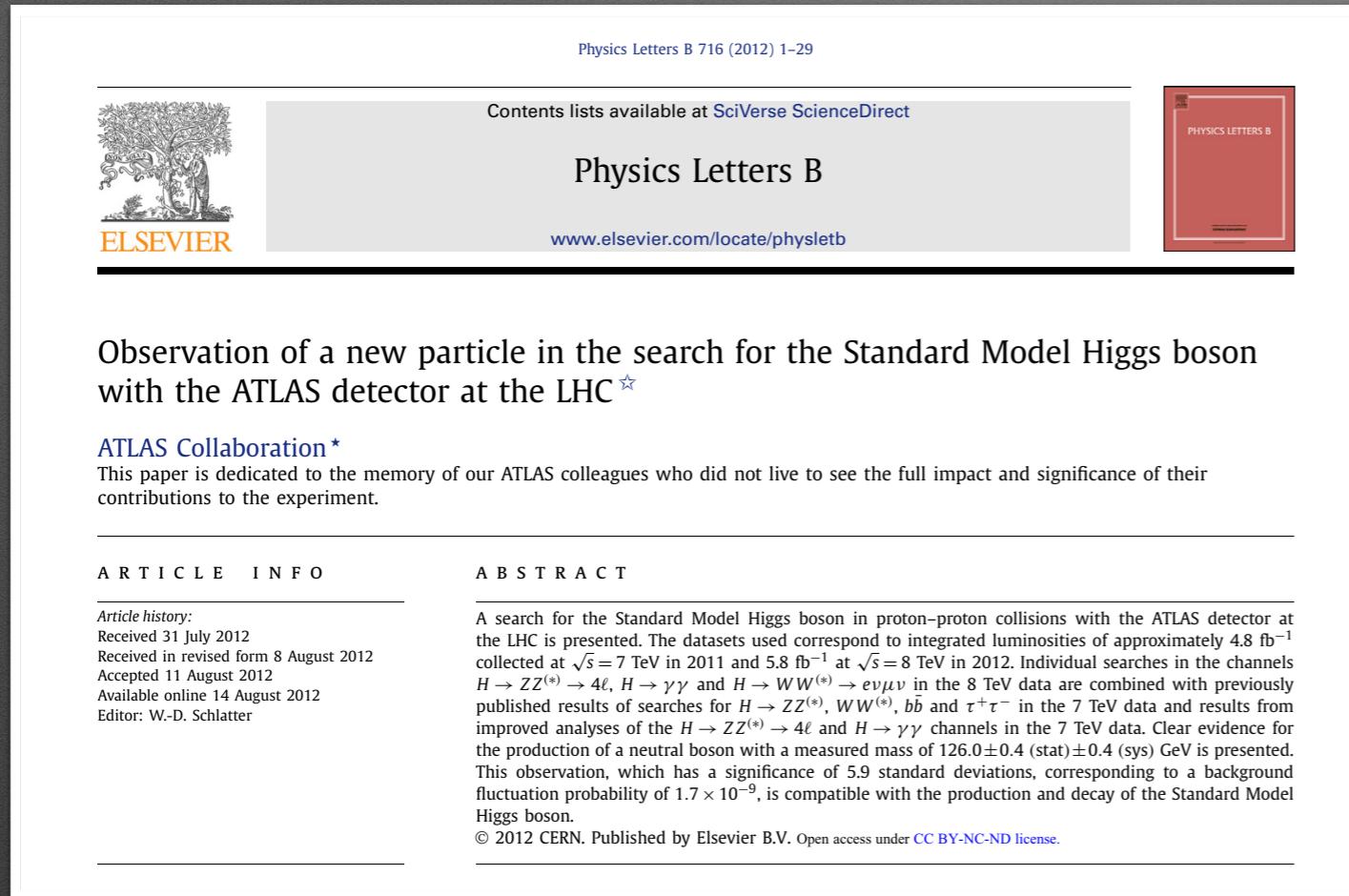


Record with 5
Story
100M Channel
“Camera”
(60 TB/s)



Processed by
300k Cores
Around the
World

Higgs Discovery - Nobel Prize Physics 2013



Not Done... Higgs is Light! Possibilities:

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

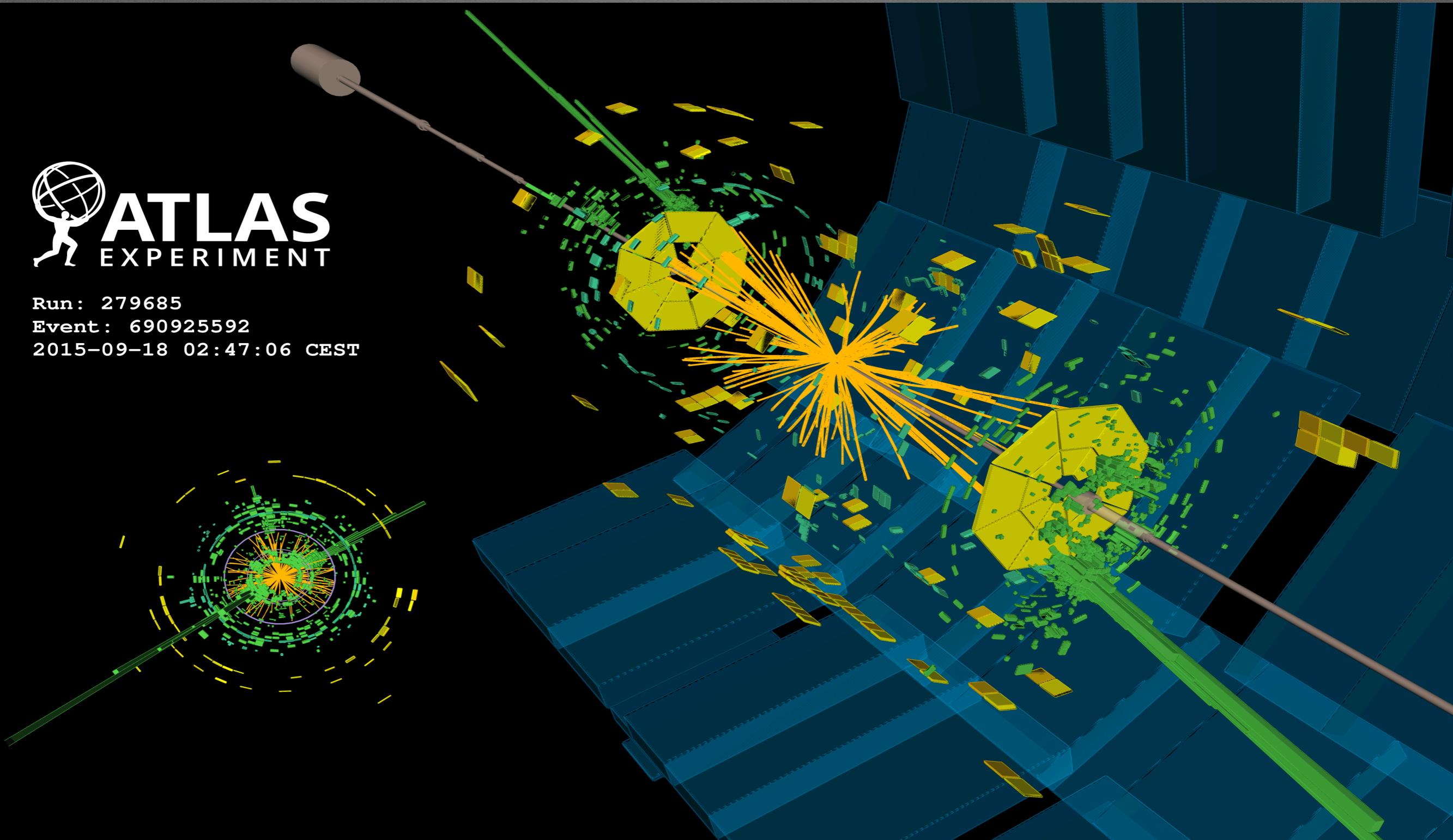
Last Piece of the Standard Model

Best Tested Theory... Ever.

Deep Learning in High Energy Physics



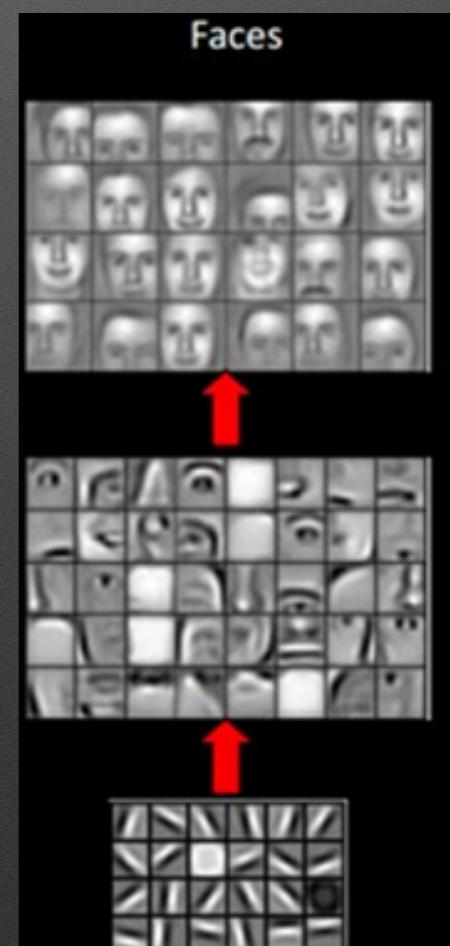
Run: 279685
Event: 690925592
2015-09-18 02:47:06 CEST



- 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



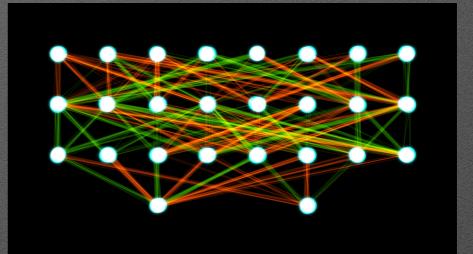
Brief History of AI

Artificial Intelligence

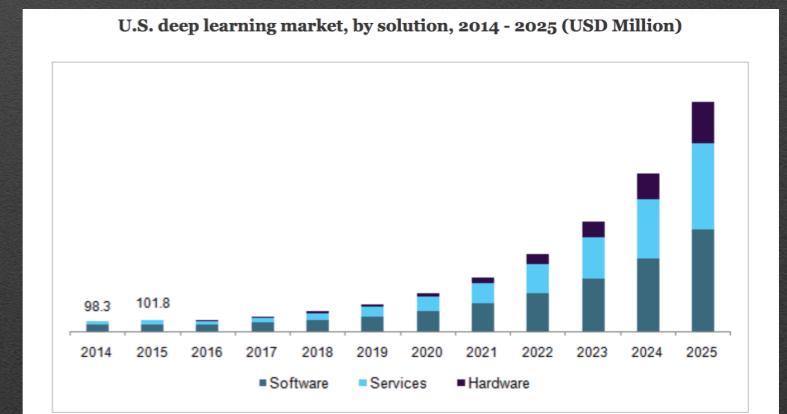
- Goal: Systems that reason and act as well as or better than humans
- Heuristic AI (1990's)
- Machine Learning AI
 - Knowledge learned from data
 - Neural Networks ~ Brain inspired computing (1943)
 - Universal Computation Theorem (1989)
 - Multi-layer hidden networks (a.k.a. Deep) (1965)
 - Vanishing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

Deep Learning Renaissance (> 2007 - now)

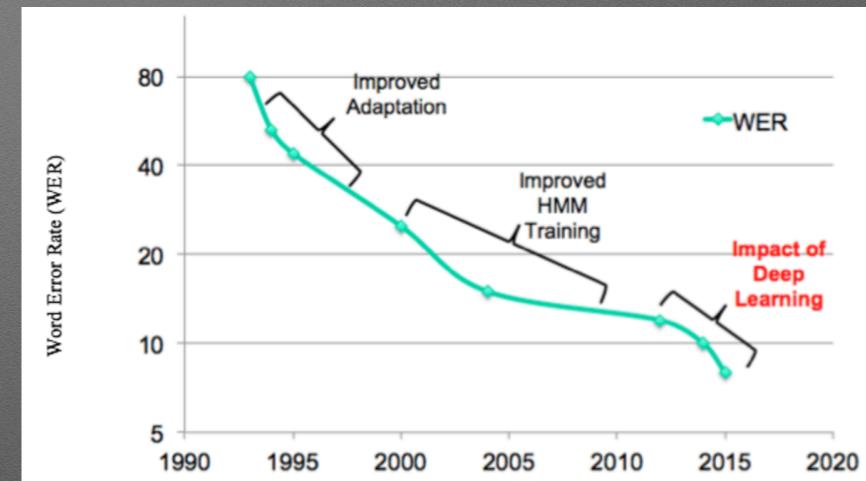
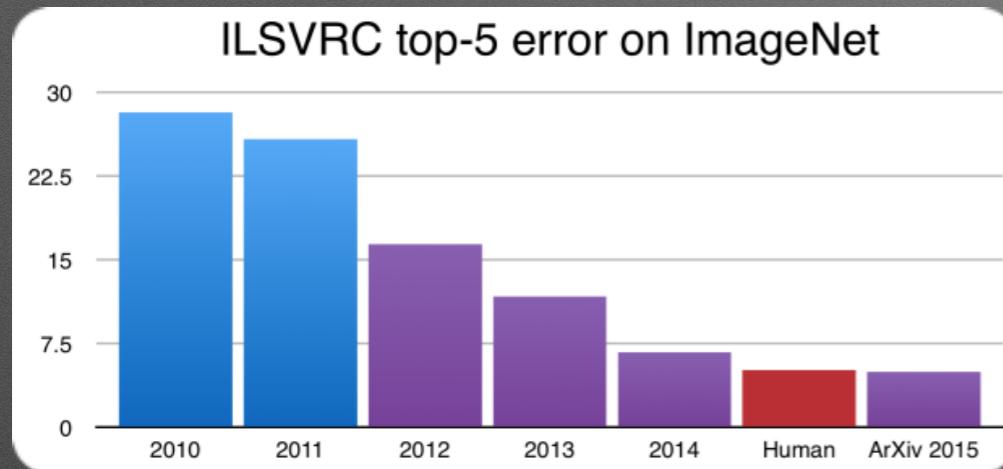
- Driven by:
 - New NN Innovation
 - Big Data
 - Graphical Processing Units
- Amazing Feats



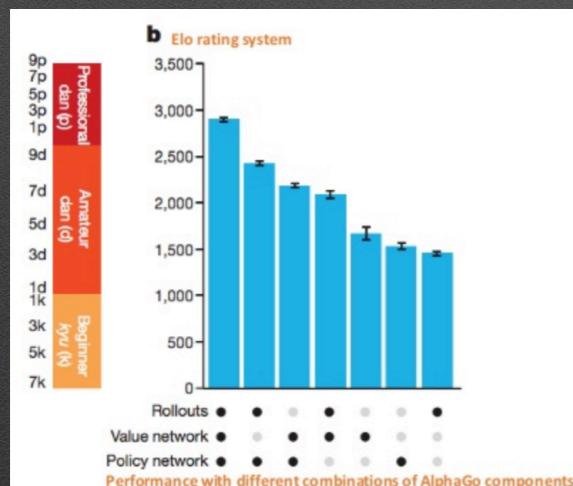
- Market Growth
- Industry Adoption



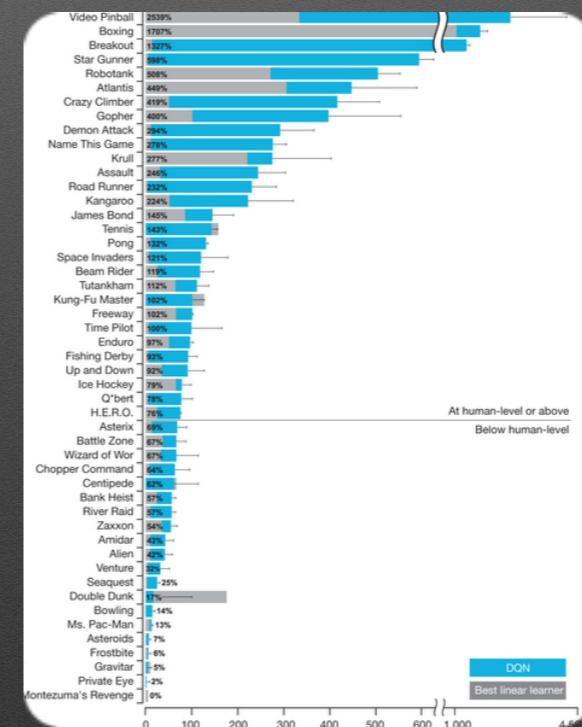
Amazing Feats : Some Examples



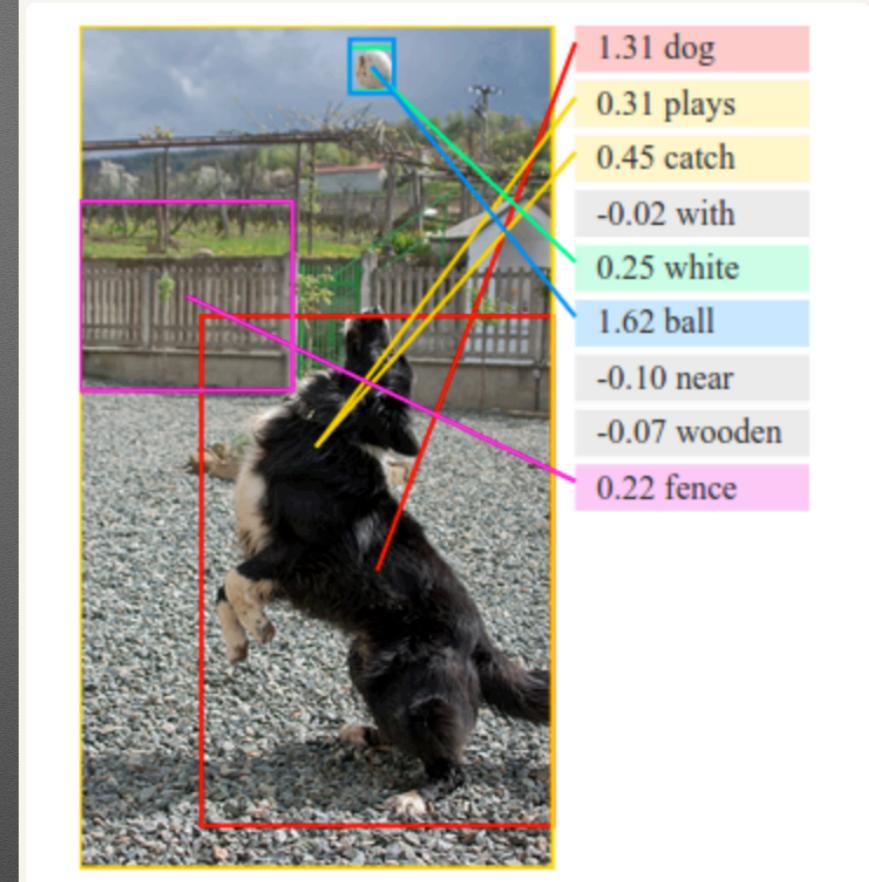
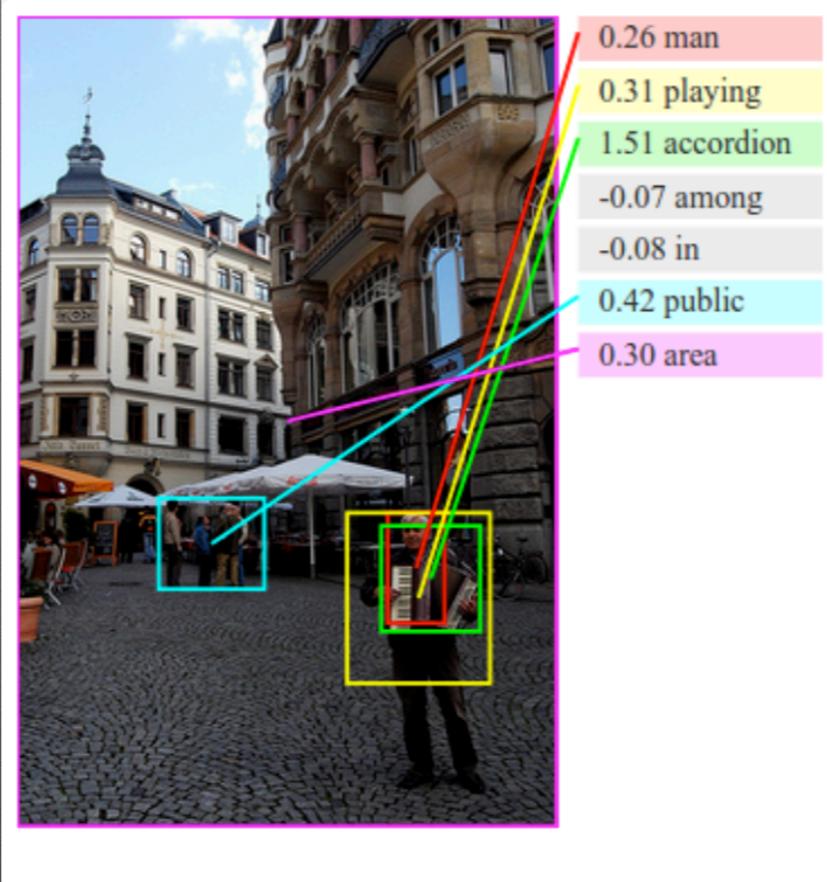
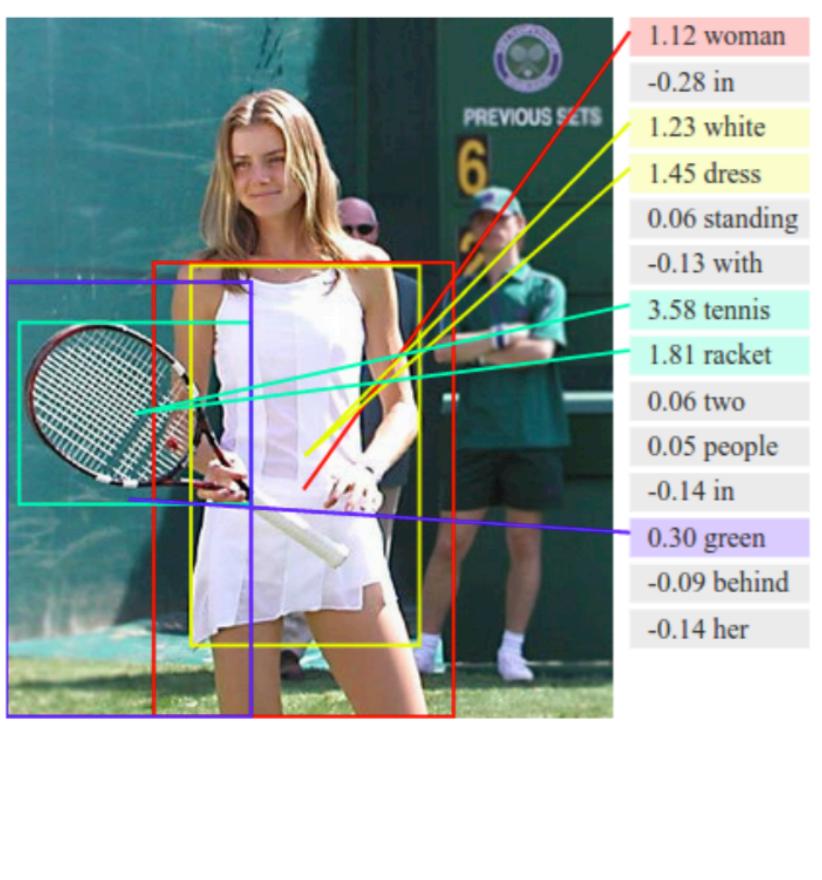
ImageNet Outperforms humans



AlphaGo beats
Lee Sedol

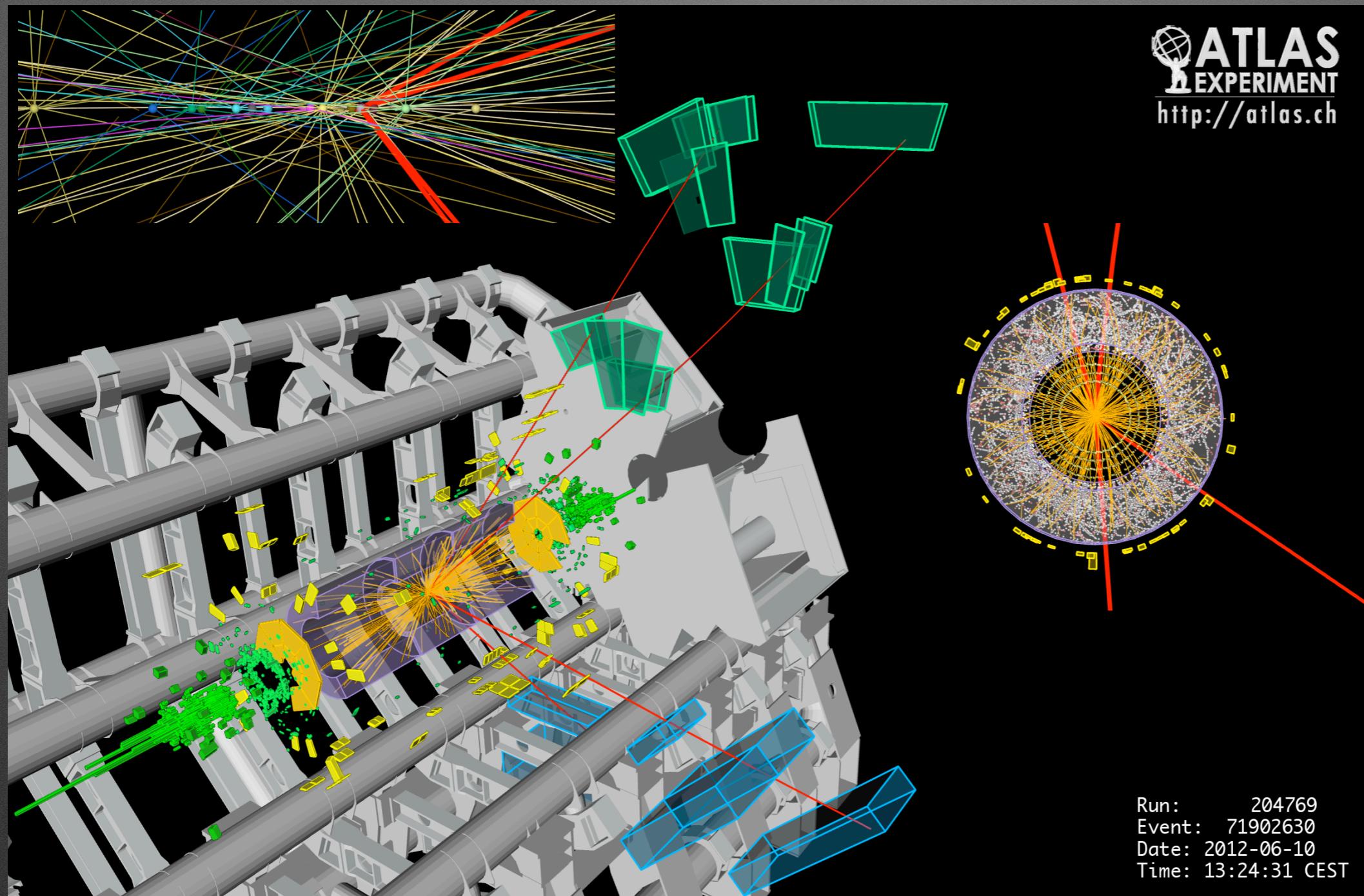


Human Level
control in playing
Atari games



Deep Learning can be used to tell a story with context through data

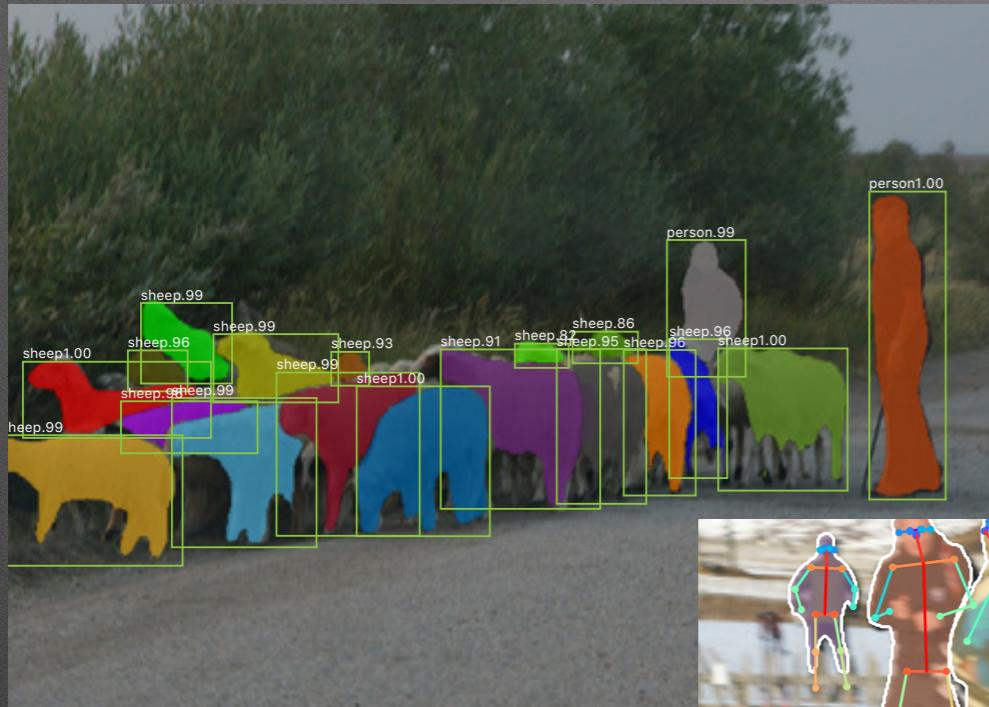
Used in HEP to understand what physics is happening



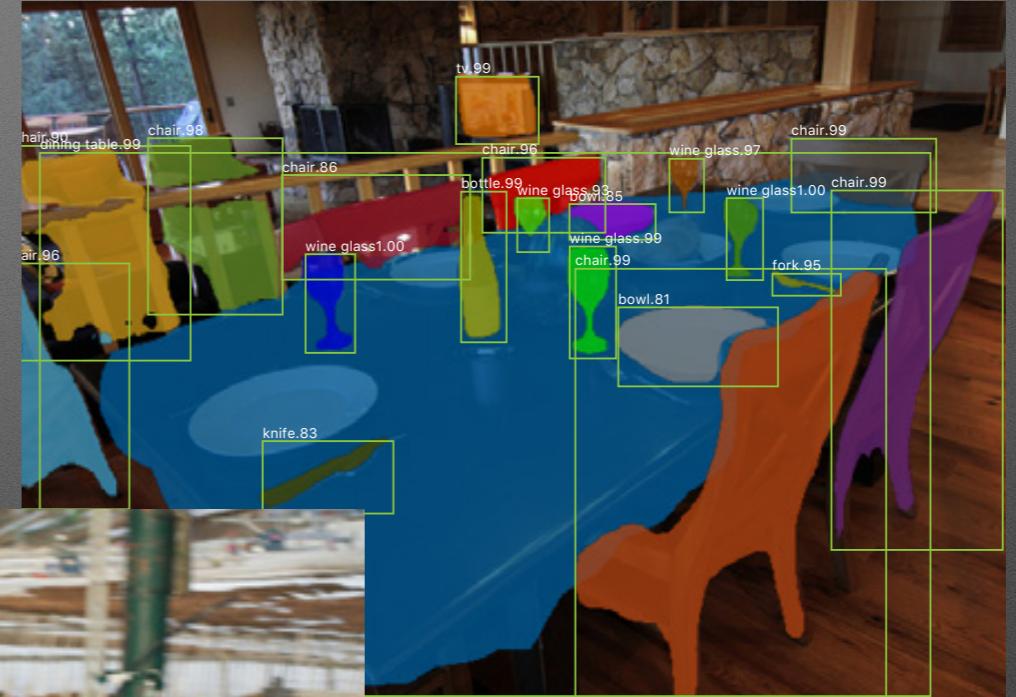
$H \rightarrow ZZ \rightarrow 4\ell$

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.