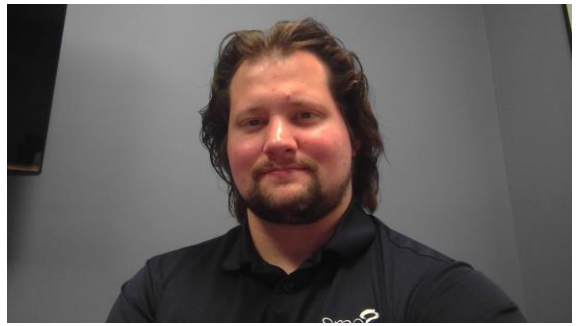
Applied Python for Data Analytics and Machine Learning

Adam J. Cook, Chair of SME Chapter 112

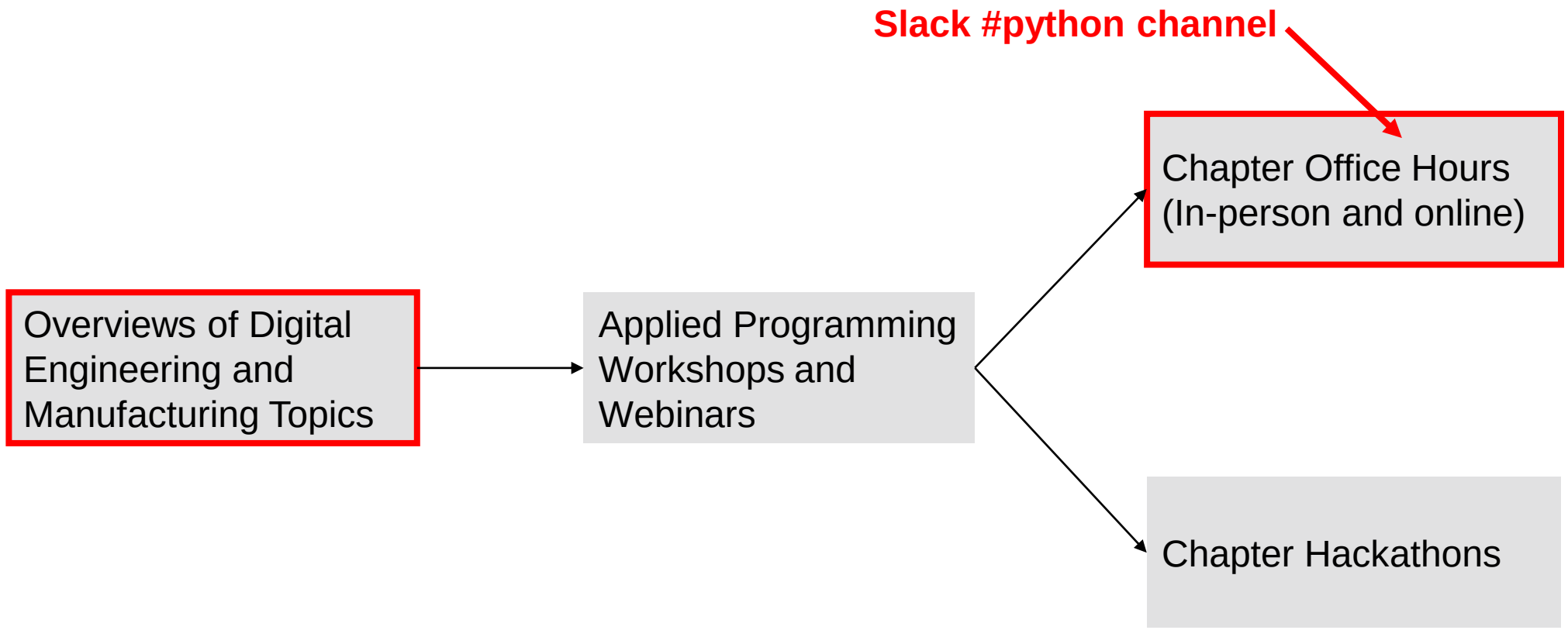
About the Presenter



- Adam Cook
- B.S. in Mechanical Engineering from Purdue University West Lafayette.
- Chief Technical Officer of Alliedstrand in Hammond, Indiana.
- Chair of SME Chapter 112 (Northwest Indiana and South Chicago).
- Embedded systems engineering, custom automation systems, industrial software.
- Lives in Chicago (come on by).
- Contact me at adam.j.cook@alliedstrand.com.

Chapter “Digital Initiative”

Slides and code:
http://bit.ly/2uzCQqR



What is Python?

Slides and code:
http://bit.ly/2uzCQqR

- High-level programming language.
- Free and open-source.
- Cross-platform.
- Extensive standard library.
- Designed to be highly readable, explicit and **productive**.
- Proven to be quite versatile (and popular).

Today's Agenda

Slides and code:
http://bit.ly/2uzCQqR

- Recap of some questions received about the last event.
- Look at a package called pandas for data analytics.
- **Super** high-level overview of artificial intelligence.
- The future paths of these Python webinars and announcements.

Caveats and Warnings

Slides and code:
http://bit.ly/2uzCQqR

- This event assumes you are a novice. If you have keep in mind that we will be watering down a bunch.
- Programming and data analytics is challenging – the following presentation will not make you into an expert. **Practice and read code.**
- For data analytics and machine learning applications, in particular, knowing Python is **not enough**.
- We are starting to get **advanced** now. The industry is **extremely** fluid. **Do not try to memorize everything!**
- We are going to talk today at a **high-level**. Let us know if you want to break down things into separate webinars.
- **Think about what kind of actual applications you want to build and let us know.** After a couple of projects, things will start clicking together.

Recap - Data Sanity

Slides and code:
http://bit.ly/2uzCQqR

- Data can (and it will, at times) lie to you.
- Think about data delivery – particularly if it is arriving from human sources (i.e. manual data collection on clipboard).
- Data anomalies will occur (i.e. sensor failures). How do you address them?
- Are you collecting the right data and, more importantly, enough **relevant** data?
- Careful of biases (i.e. confirmation bias, selection bias and others) and anecdotal evidence. **Be scientific!**
- When facts are weapons.
- Good example of selection bias. **Why?**

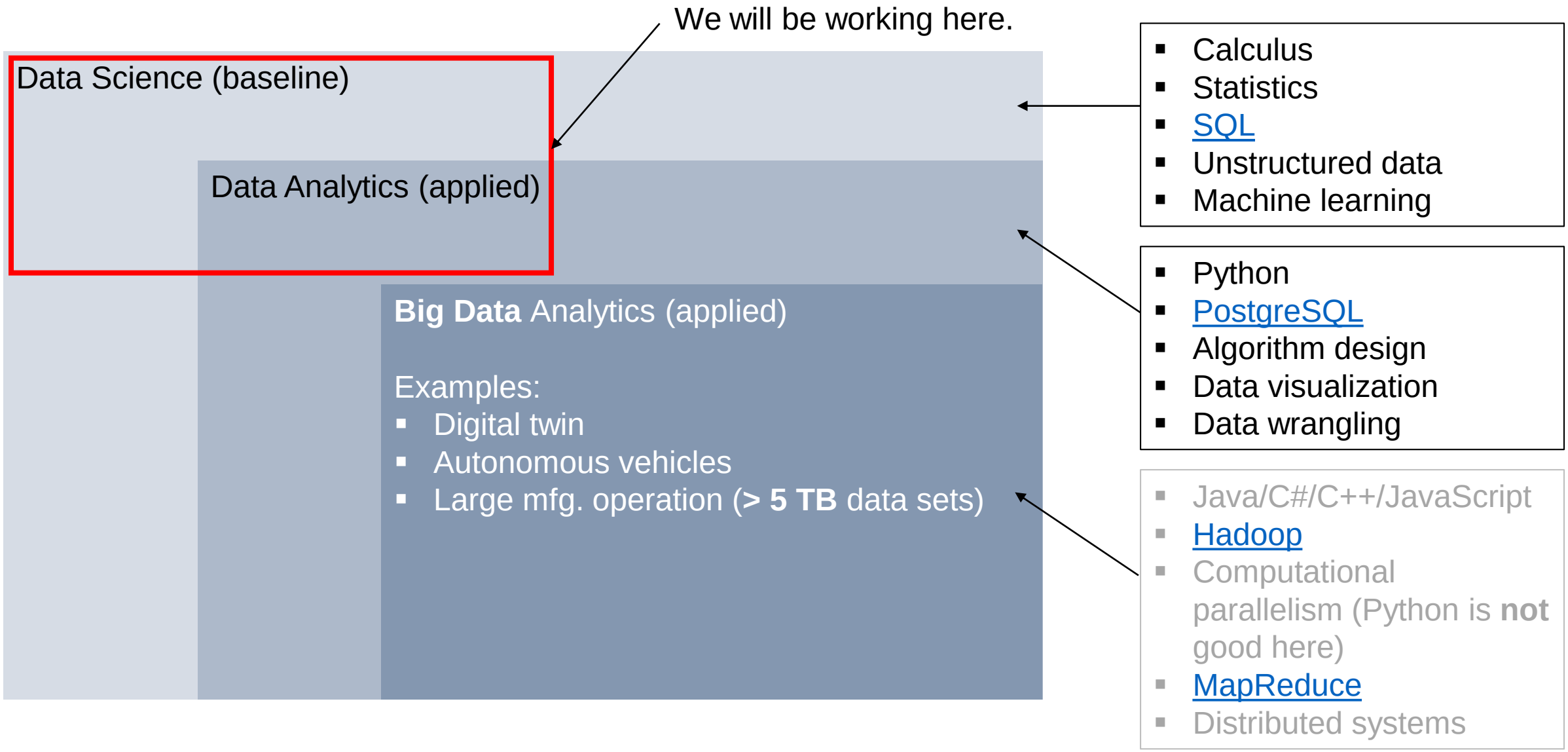
Funny, but true…or is it?

Slides and code:
http://bit.ly/2uzCQqR

Source: https://twitter.com/bigdataborat/status/306596352991830016?lang=en

Recap - Word Soup

Slides and code:
http://bit.ly/2uzCQqR

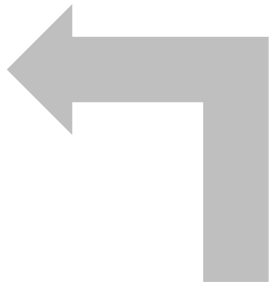


Recap – Data Architecture

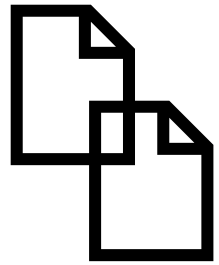
Slides and code:
http://bit.ly/2uzCQqR

This architecture is suitable when the data is small and static.

Python application or Jupyter
(you can either write a custom Python application or just write your code using Jupyter)



Read-only



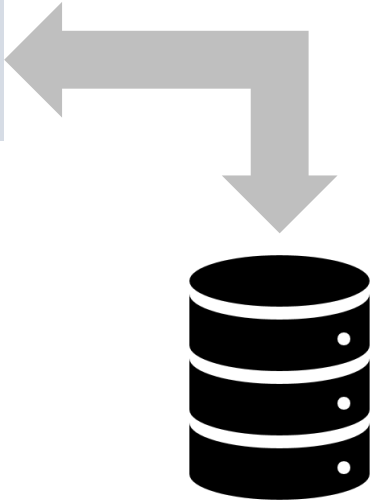
Raw data in one or multiple files

Recap – Data Architecture

Slides and code:
http://bit.ly/2uzCQqR

Python application or Jupyter
(you can either write a custom Python application or just write your code using Jupyter)

This architecture is suitable when the data is too large for files (100s of MBs) and the data might be real-time.



Database (PostreSQL)

Broker written in Python, perhaps
(you can write this or you can potentially use someone else's, like Predix)

Read-only

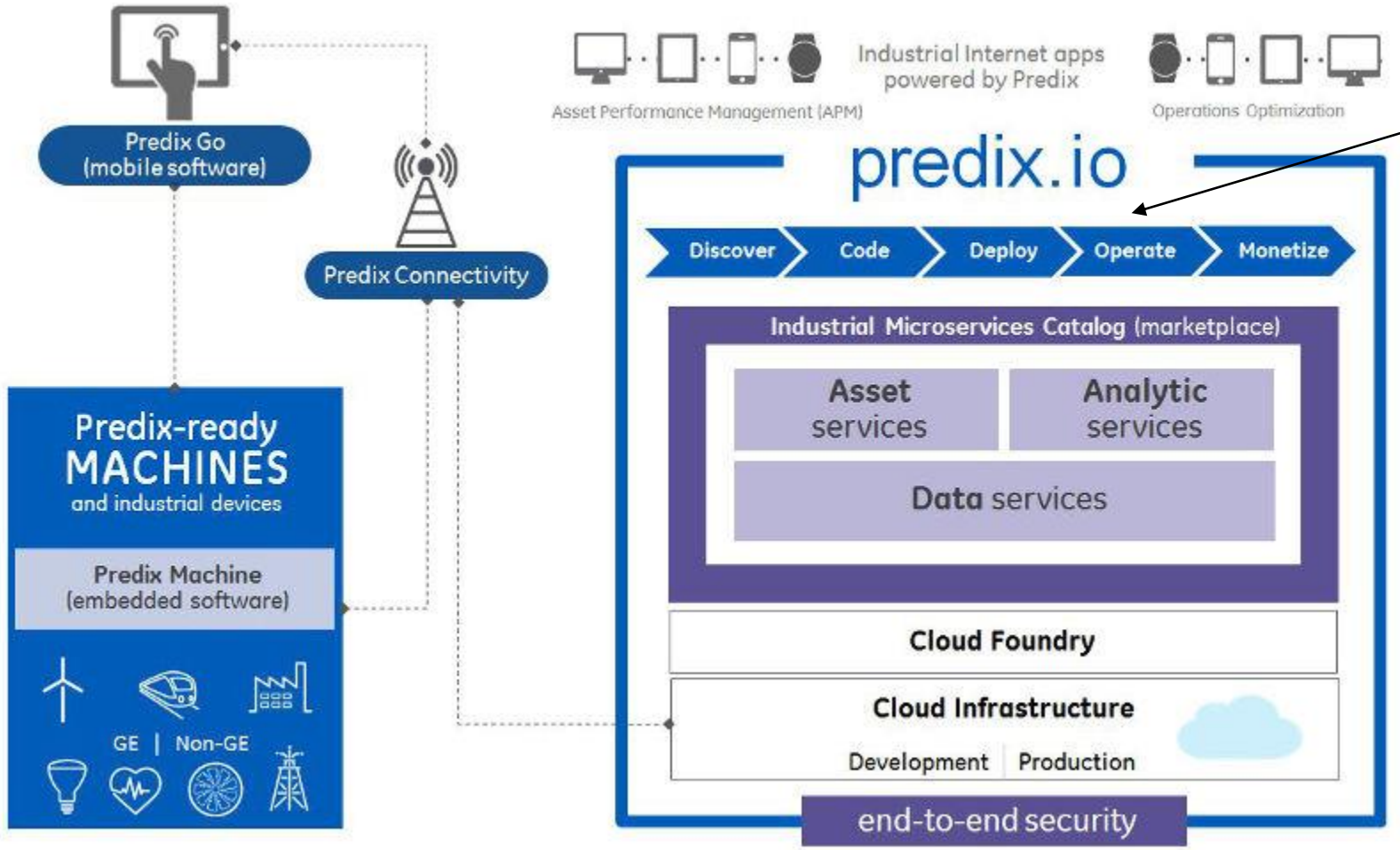
Recap – Data Architecture

Slides and code:
http://bit.ly/2uzCQqR

If you are questioning whether you need Big Data (Hadoop), then you do not need Big Data (something like GE Predix is also an option)!

Recap – Data Architecture

Slides and code:
http://bit.ly/2uzCQqR



This is where you can write your own Python application.

Source: http://bit.ly/2w7duQ8

What is the point of data analytics?

Simple. To extract value from data.

Pandas

Slides and code:
http://bit.ly/2uzCQqR

Essential tool for data analytics in Python. It's efficient and fast.

pandas – provides data structures and data analysis tools for Python

Demonstration

Slides and code:
http://bit.ly/2uzCQqR

Let’s take a quick look at what pandas can do!

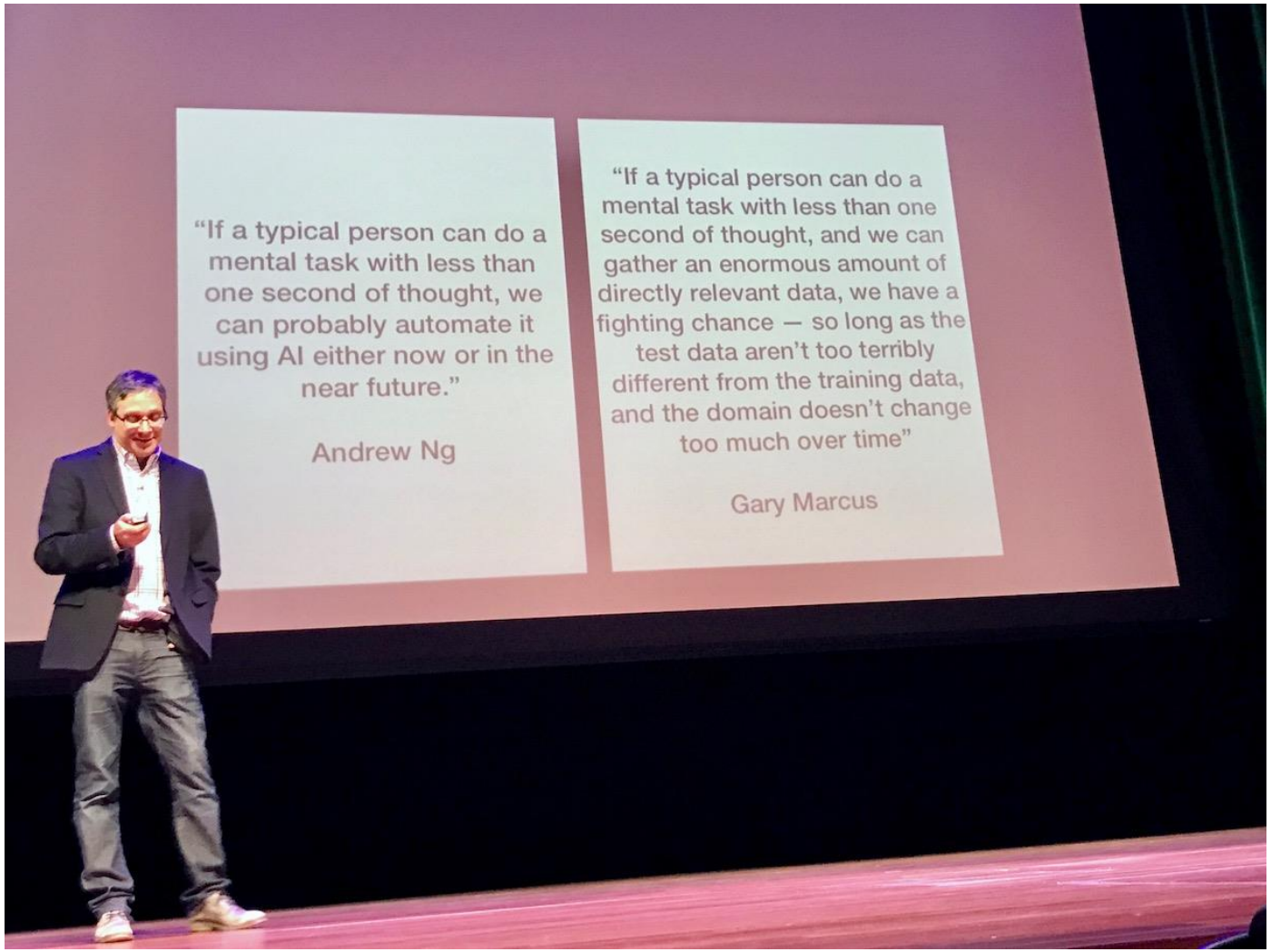
Artificial Intelligence

Slides and code:
http://bit.ly/2uzCQqR

- Narrow AI – non-sentient form of intelligence that focuses on one specific task.
- Strong AI (Artificial General Intelligence) – the intelligence of a machine can conceivably perform any task a human can.
- There is some debate on if or when we will achieve strong AI.
- There is also a fair amount of hype around artificial intelligence currently in the press. At this point, it is all speculation to some degree. Some AI technologies are showing promise for manufacturing, however (i.e. predictive maintenance and defect detection).
- We will focus on narrow AI.

Artificial Intelligence

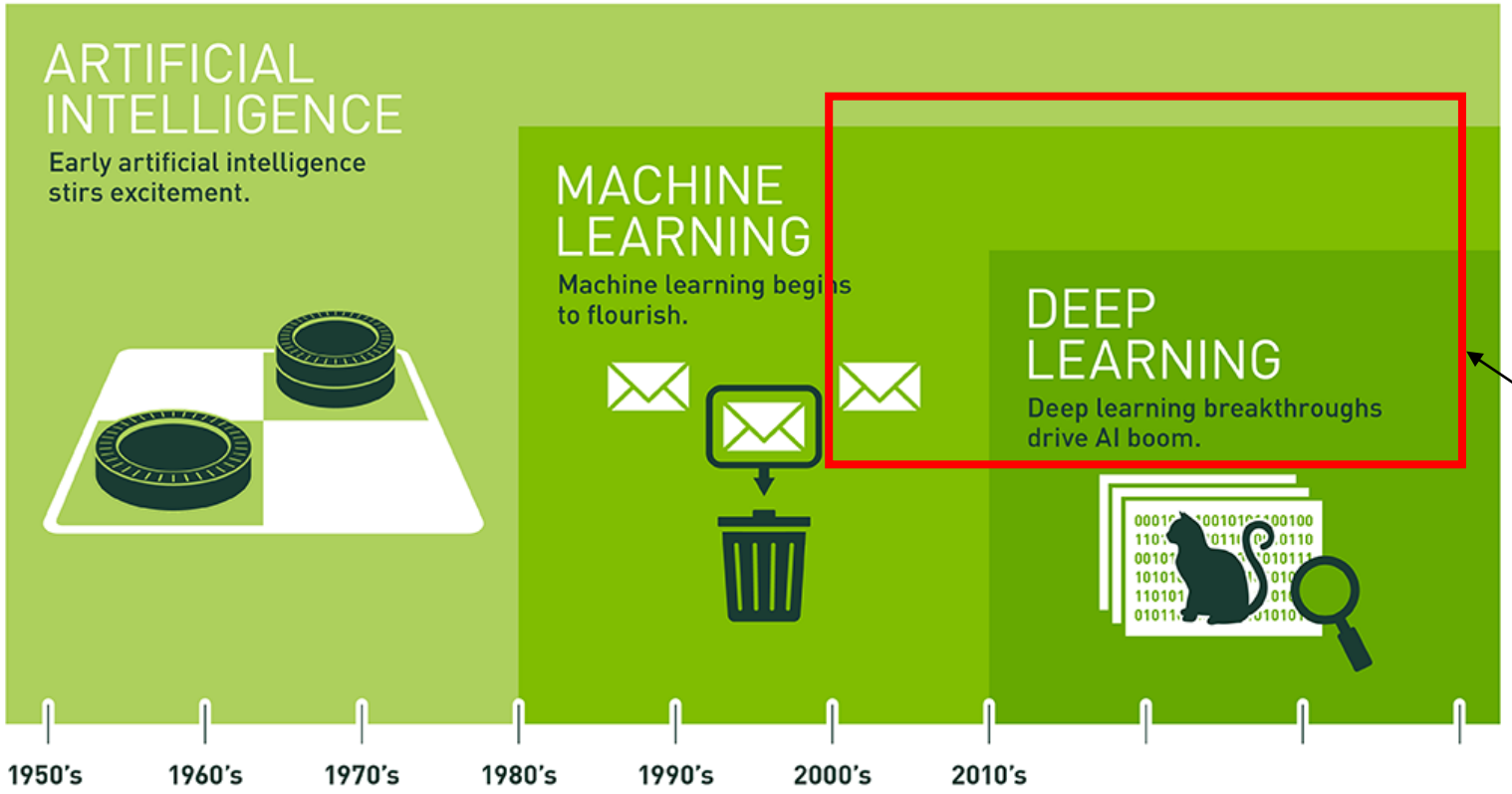
Slides and code:
http://bit.ly/2uzCQqR



Source: https://technical.ly/brooklyn/2017/04/07/ai-boom-legit-just-hype/

Artificial Intelligence

Slides and code:
http://bit.ly/2uzCQqR



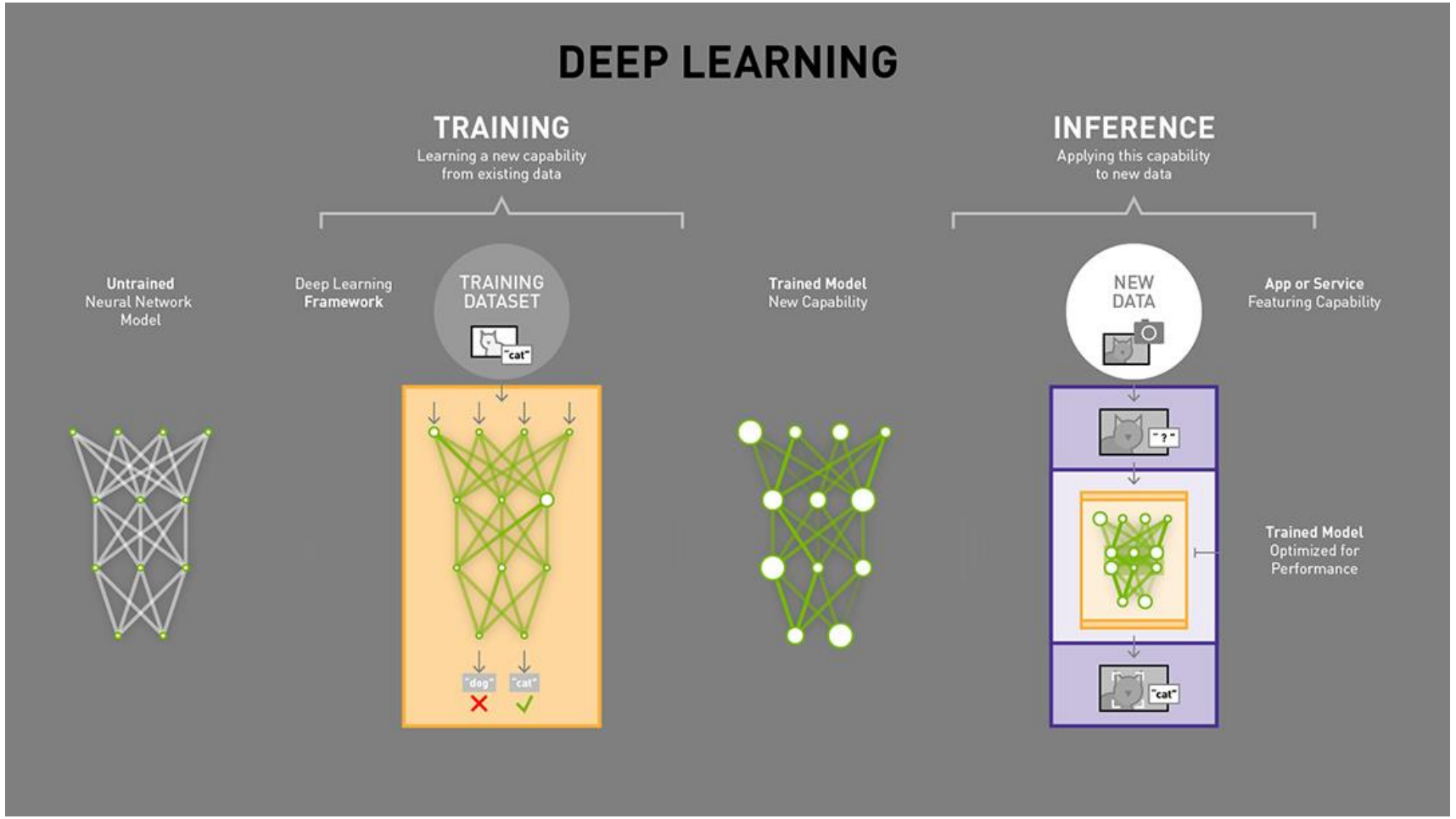
Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

We will be working here.

Source: http://bit.ly/2aoQwIx

Artificial Intelligence

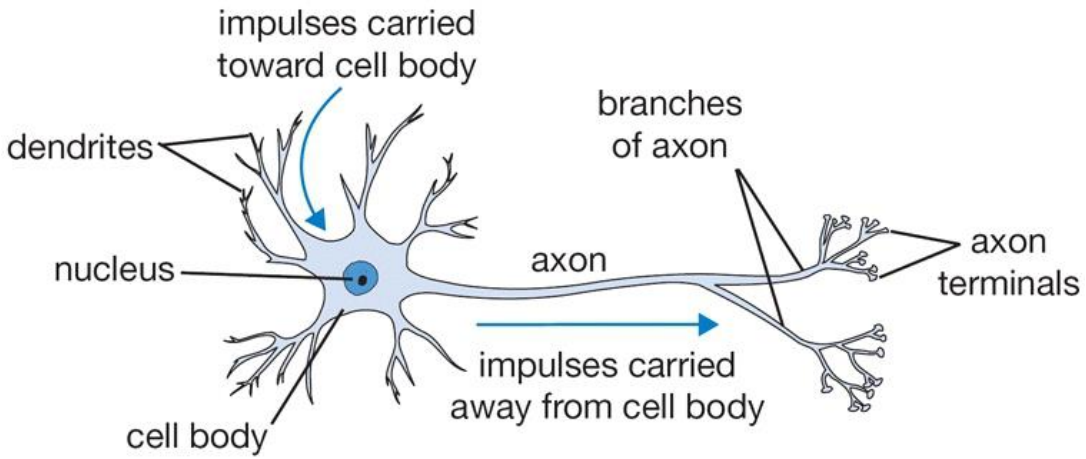
Slides and code:
http://bit.ly/2uzCQqR

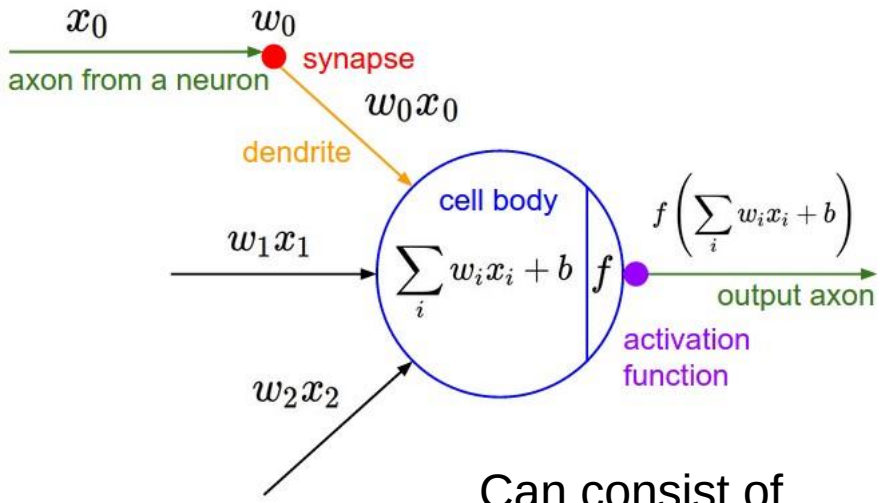


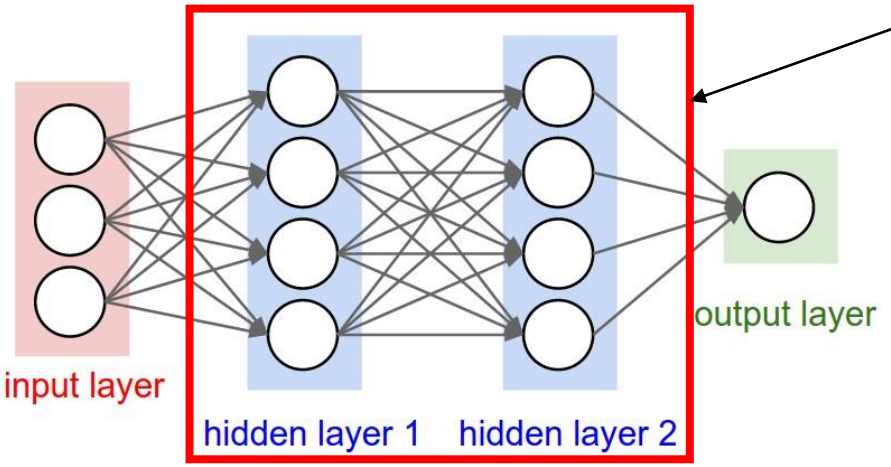
Source: http://bit.ly/2bvtm3Z

Artificial Intelligence

Slides and code:
http://bit.ly/2uzCQqR







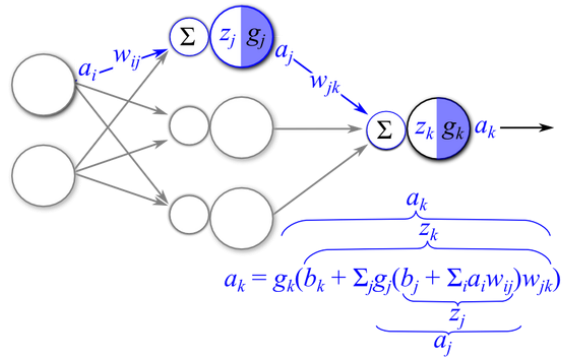
Can consist of infinitely many nodes and layers.

Source: http://cs231n.github.io/neural-networks-1/

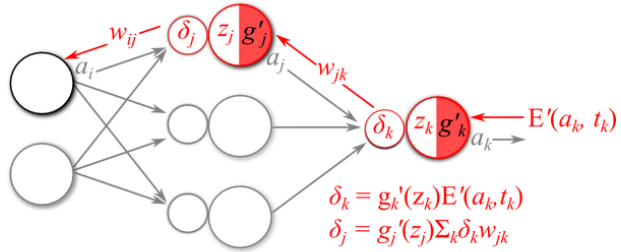
ANN Mechanics

Slides and code:
http://bit.ly/2uzCQqR

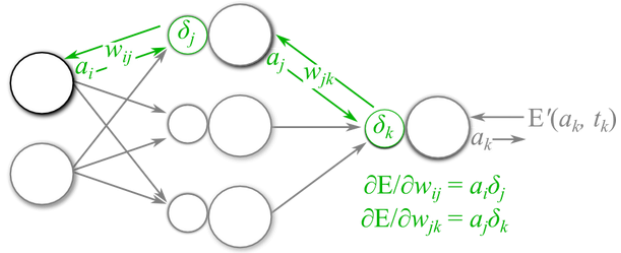
I. Forward-propagate Input Signal



II. Back-propagate Error Signals



III. Calculate Parameter Gradients



IV. Update Parameters

$w_{ij} = w_{ij} - \eta(\partial E/\partial w_{ij})$
$w_{jk} = w_{jk} - \eta(\partial E/\partial w_{jk})$
for learning rate η

Source: http://bit.ly/2w7RpB6

AI Tools

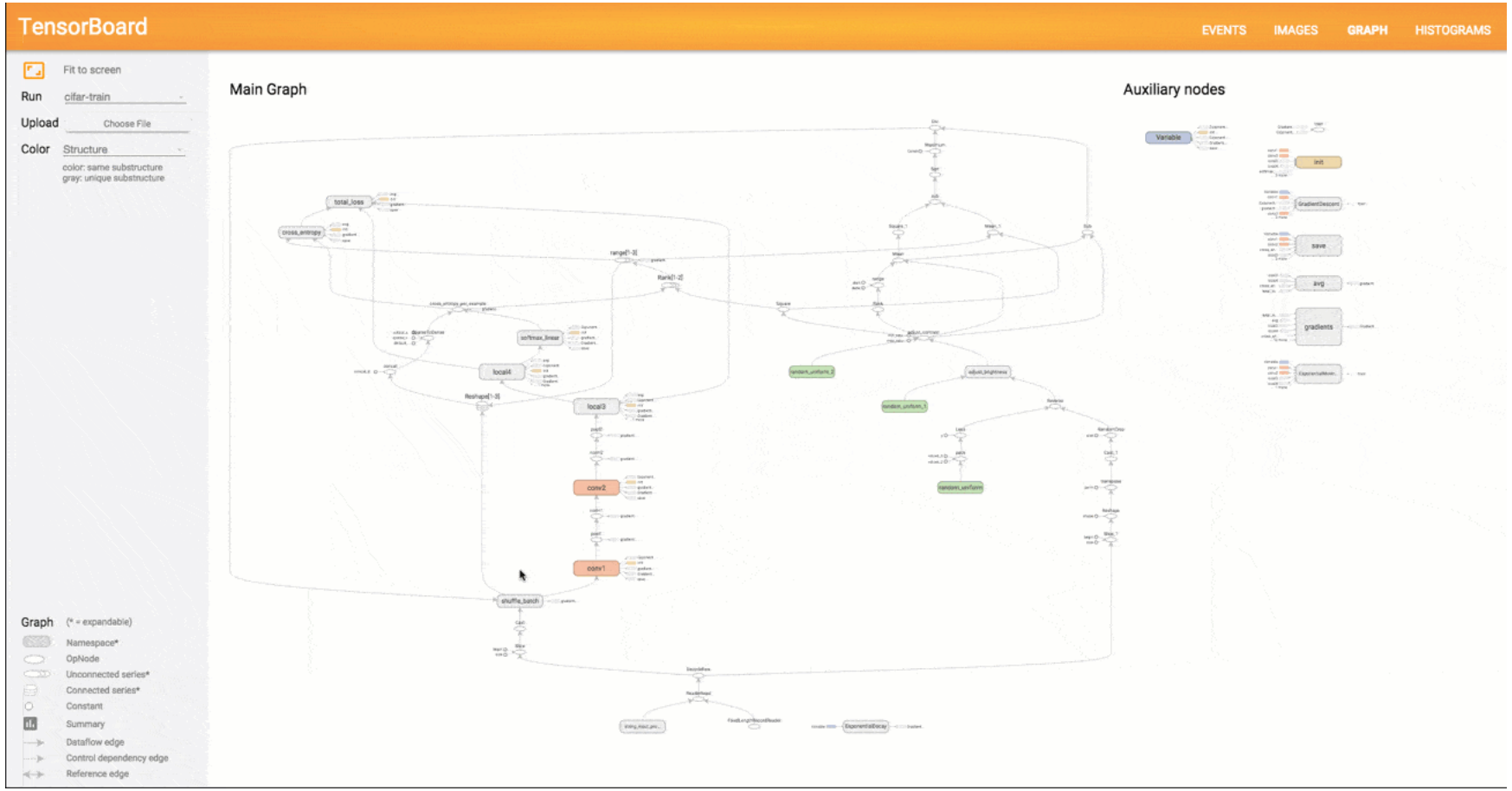
Slides and code:
http://bit.ly/2uzCQqR

- scikit-learn – high-level tools for machine learning (out-of-the-box algorithms such as support vector machines or logistic regressions).
- Tensorflow – low-level tools for building deep learning systems (**GPU support**).

Both of these have Python APIs

AI Visualization

Slides and code:
http://bit.ly/2uzCQqR



[Tensorboard]

Tensorflow

Slides and code:
http://bit.ly/2uzCQqR

- Developed internally by Google and is now open-source.
- Powerful library written mostly in C++ which allows you to generate computation graphs (neural network).
- In my cases, it handles the complexities of interacting with GPUs for neural network training.
- What can you use Tensorflow for? Image recognition (defect detection), OCR, **manufacturing asset insights**, engineering analysis models and much more.
- Try to do the Tensorflow OCR tutorial here: http://bit.ly/2v8iali

Relationship between data and AI?

Slides and code:
http://bit.ly/2uzCQqR

- **Sane** data is used for training the neural network.
- For some problems, large amounts of data is needed to train the network. Getting this data can be **difficult**. Think about the amount of data that Google has about you and everyone else.
- For some problems, however, there are success stories for producing highly accurate neural networks with small data sets.
- For other problems, there is a technique called transfer learning.

Important Online Resources for AI

Slides and code:
http://bit.ly/2uzCQqR

- http://bit.ly/2danP4n - Applied Data Science with Python Specialization – University of Michigan (free to audit)
- http://bit.ly/1IXp8Lg – Machine Learning Specialization – Sanford University (free)
- http://bit.ly/2tT9aSC - Deep Learning Specialization (free to audit)
- **http://www.fast.ai – Applied Deep Learning for Coders (free)**

Where can I get this slide deck and code?

Slides and code:
http://bit.ly/2uzCQqR

http://bit.ly/2uzCQqR
(actually, go ahead and bookmark this link – this web page will be updated constantly with new content)

Call for Feedback and Help

Slides and code:
http://bit.ly/2uzCQqR

Please provide us with feedback (and/or your data/ML problems)!

Want to Keep the Conversation Going?

Slides and code:
http://bit.ly/2uzCQqR

We have a Slack channel! Send me an invite request at my e-mail address. (we are working on an automatic invite link)

Fall/Spring Webinars

Slides and code:
http://bit.ly/2uzCQqR

- Regular Expressions with Python – September 2017
- Computational Geometry in Python (might be in multiple parts) – October 2017
- MTConnect Fundamentals – November 2017
- **Industrial Internet of Things – September 2017 through December 2017**
- **Robotics and Artificial Intelligence – January 2018 through May 2018**
- **Subscribe to our chapter's YouTube channel: http://bit.ly/2iv32iu**

Thank you!

Slides and code:
http://bit.ly/2uzCQqR

Thanks for attending!

Special thanks to our hosting partner – GreenCow Coworking. Check them out at greencow.space!

Suggestions? Feedback? Comments? Complaints? Contact us below!

Adam J. Cook

Chief Technical Officer of Alliedstrand

Chair of SME Chapter 112

adam.j.cook@alliedstrand.com

https://linkedin.com/in/adam-j-cook

https://github.com/adamjcook

SME

www.sme.org

https://facebook.com/SMEmfg

https://twitter.com/SME_MFG

https://linkedin.com/company/sme

SME Chapter 112

Serving Northwest Indiana and Chicagoland

https://facebook.com/sme112

https://linkedin.com/company/sme112

https://github.com/sme112