Machine Learning and AI at IBM

Lennart Frantzell PhD, Developer Advocate, IBM

4/24/2018

http://ibm.biz/mlatibm

IBM and AI

- **1996, 1997. Deep Blue** The first match was played in Philadelphia in 1996 and won by Kasparov. The second was played in New York City in 1997 and won by **Deep Blue**.
- **Watson Jeopardy 2011**: The IBM Challenge aired February 14–16, 2011, and **featured IBM's Watson computer** facing off against Ken Jennings and Brad Rutter in a two-game match played over three shows. This was the first man-vs.-machine competition in Jeopardy!'s history. Watson won both the first game and the overall match

Watson uses IBM's DeepQA software and the Apache UIMA, Unstructured Information Management Architecture, is an OASIS standard for content analytics.

IBM and AI

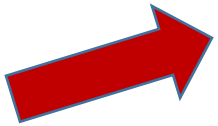
- **1996, 1997. Deep Blue** The first mat… a in 1996 and won by Kasparov. The se… 1997 and won by **Deep Blue**.
- **Watson Jeop… featu…** …, 2011, and …nings and Brad …s. This was the first man-vs… …y. Watson won both the first ga…

Ke Jie, the world's best player of what might be humankind's most complicated board game was defeated on Tuesday by a Google computer program. Adding insult to potentially deep existential injury, he was defeated at Go — a game that claims centuries of play by humans — in China, where the game was invented.
May 23 2017

Watso… DeepQA software and the Apache UIMA, Unstructured Informa… Management Architecture, is an OASIS standard for content analytics,

From an analog to a digital world

Watson For Jeopardy

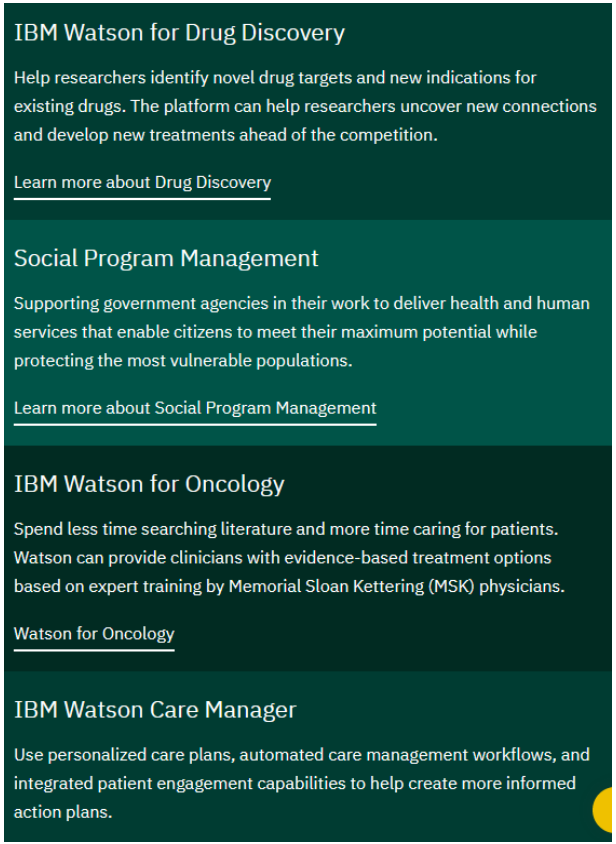


Watson for Healthcare etc deployed in Private Clouds

Watson services in the IBM Public Cloud

Watson for Healthcare

https://www.ibm.com/watson/health/oncology-and-genomics/genomics/



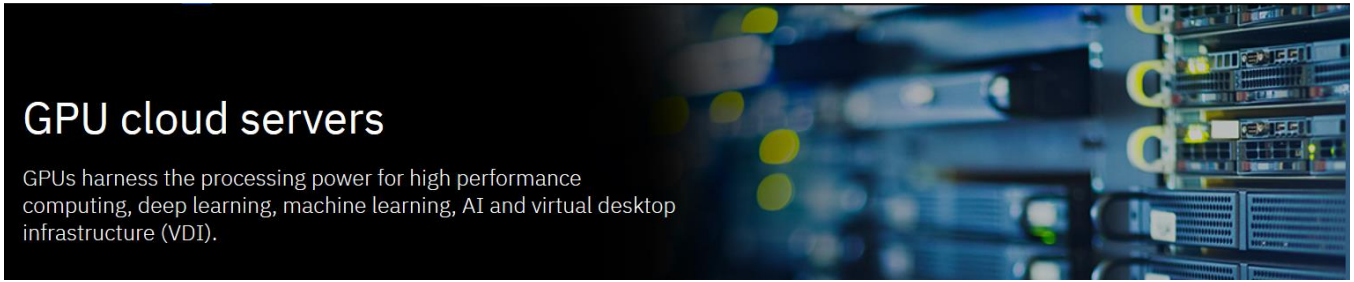
https://www.ibm.com/watson/health/

Data in Healthcare

- With IBM's planned $2.6 billion acquisition of Truven Health, the company will add "200 million lives" to its data trove. "Lives" is a term typically used in the healthcare business for a data asset or record.
- And when it comes to big data analytics, the more data, the better, said IBM (ibm, -0.84%) Watson Health general manager Deborah DiSanzo. Truven brings still more data into IBM, which has already assembled quite the data pool, both on its own and via acquisition.

http://fortune.com/2016/02/18/ibm-truven-health-acquisition/

The GPU and Machine Learning

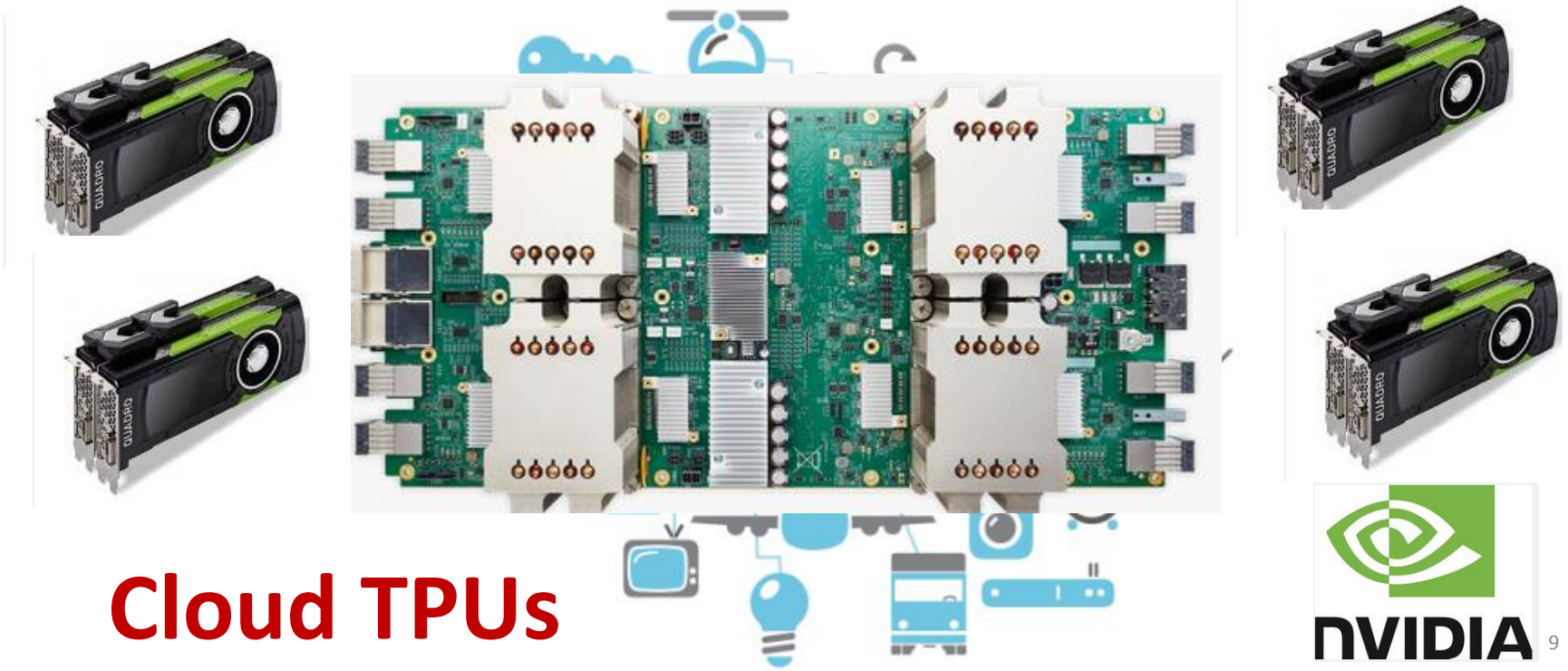


…The answer lies in the rise of deep learning, an advanced machine learning technique that is heavily used in AI and Cognitive Computing. Deep learning powers many scenarios including autonomous cars, cancer diagnosis, computer vision, speech recognition and many other intelligent use cases.

Like most of the ML algorithms, deep learning relies on sophisticated mathematical and statistical computations. Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are some of the modern implementations of deep learning. …. Each type of neural net is aligned with a complex use case like classification, clustering and prediction.

https://www.forbes.com/sites/janakirammsv/2017/08/07/in-the-era-of-artificial-intelligence-gpus-are-the-new-cpus/#16d91afc5d16

Big Data, Hardware and the GPU



Cloud TPUs

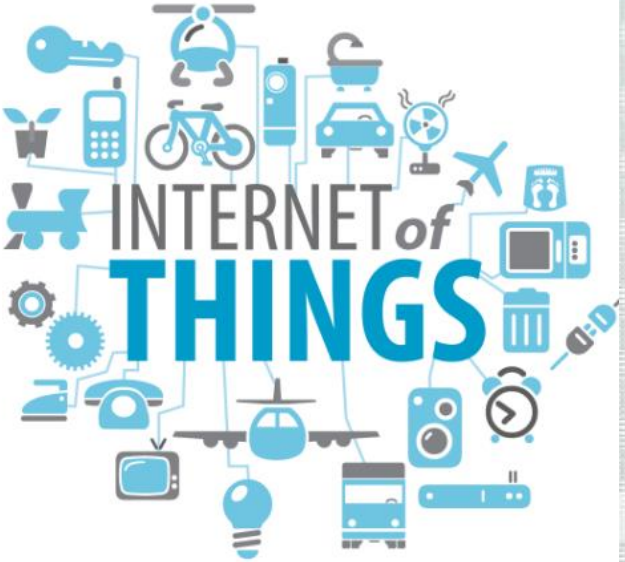
AI, Machine Learning and hardware

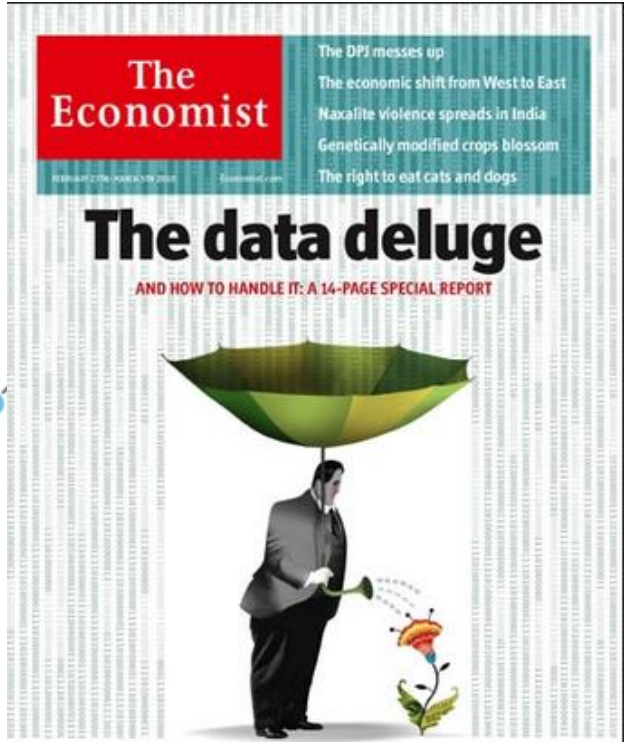
IBM Bets on Artificial Intelligence With Its POWER9 Chip

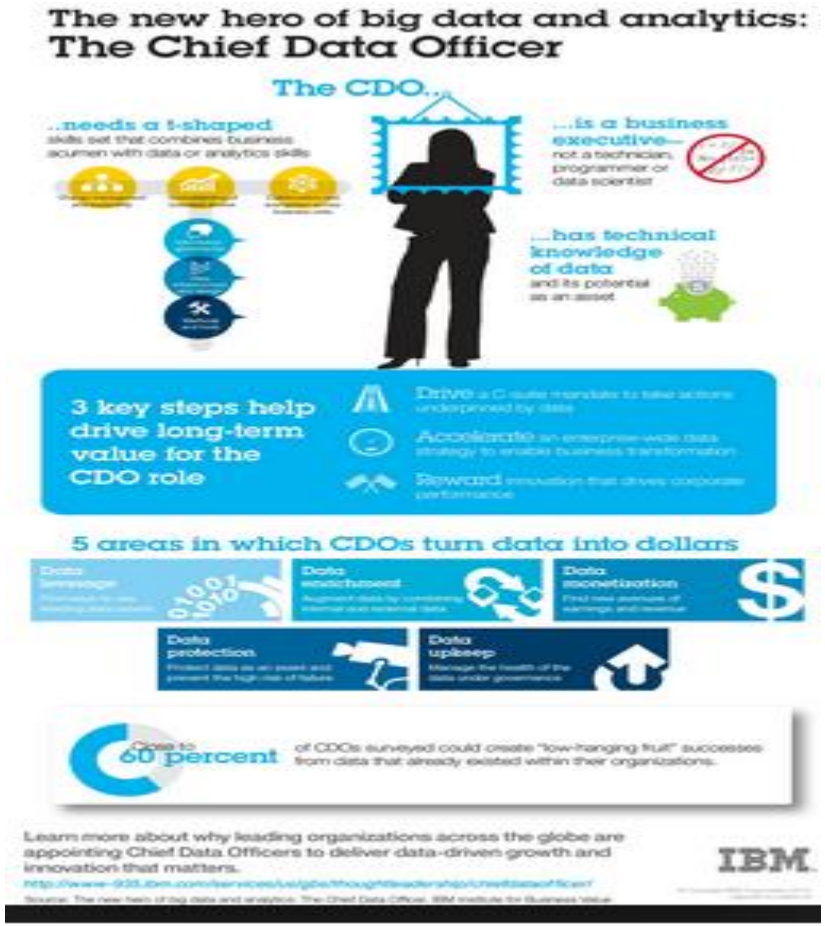
The company is hoping to carve out a presence in the hyperscale data center market with its AI-optimized processor.

Machine Learning, the Software story

The Internet of Things and Big Data



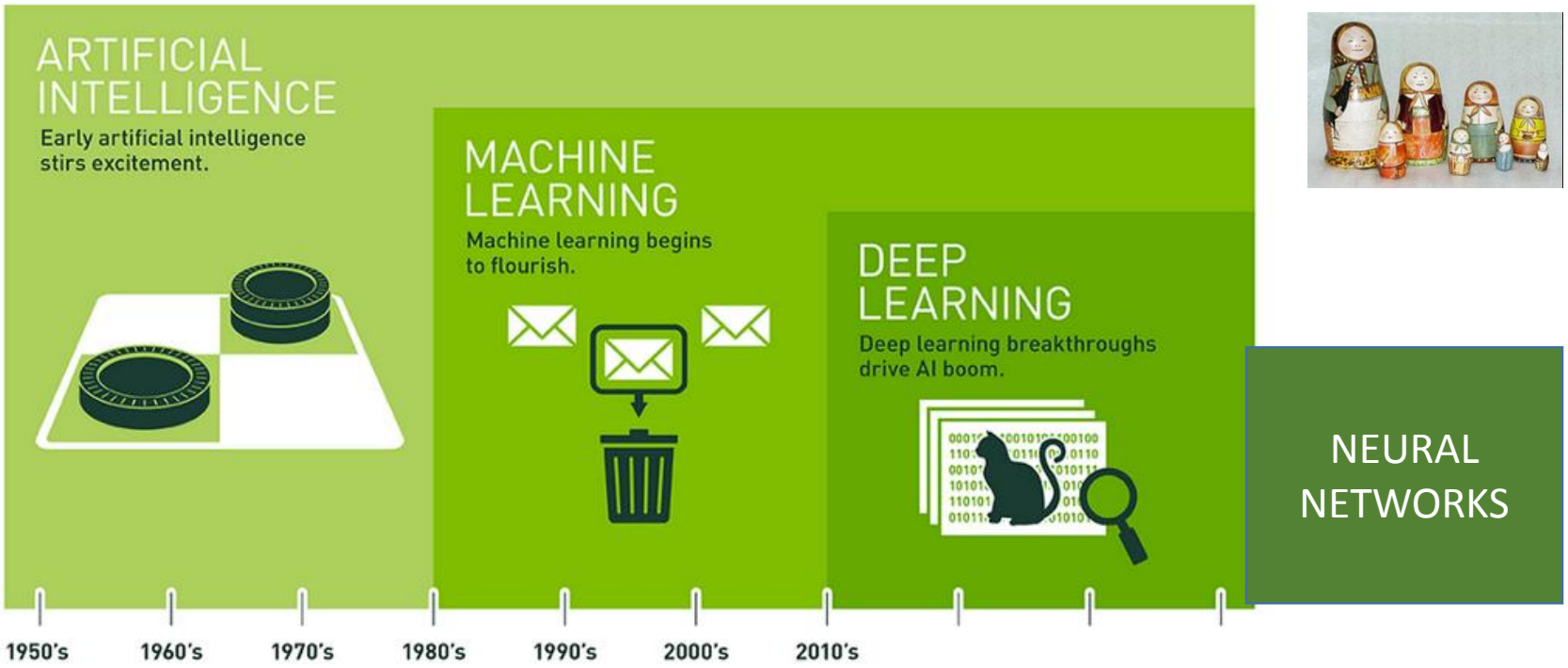



The new hero of big data and analytics:
The Chief Data Officer
The CDO...
...needs a t-shaped
skills set that combines business acumen with data or analytics skills
...is a business executive—
not a technician, programmer or data scientist
...has technical knowledge of data
and its potential as an asset
3 key steps help drive long-term value for the CDO role
Drive a C-suite mandate to take actions underpinned by data
Accelerate an enterprise-wide data strategy to enable business transformation
Reward innovation that drives corporate performance
5 areas in which CDOs turn data into dollars
Data leverage
Data enrichment
Augment data by combining internal and external data
Data monetization
Find new avenues of earnings and revenue
Data protection
Protect data as an asset and prevent the high risk of failure
Data upkeep
Manage the health of the data under governance
Close to 60 percent
of CDOs surveyed could create "low-hanging fruit" successes from data that already existed within their organizations.
Learn more about why leading organizations across the globe are appointing Chief Data Officers to deliver data-driven growth and innovation that matters.
IBM

Just in case you didn't know

But, ~ 80% of all ML projects are taken up with finding, transferring, washing and structuring the data.

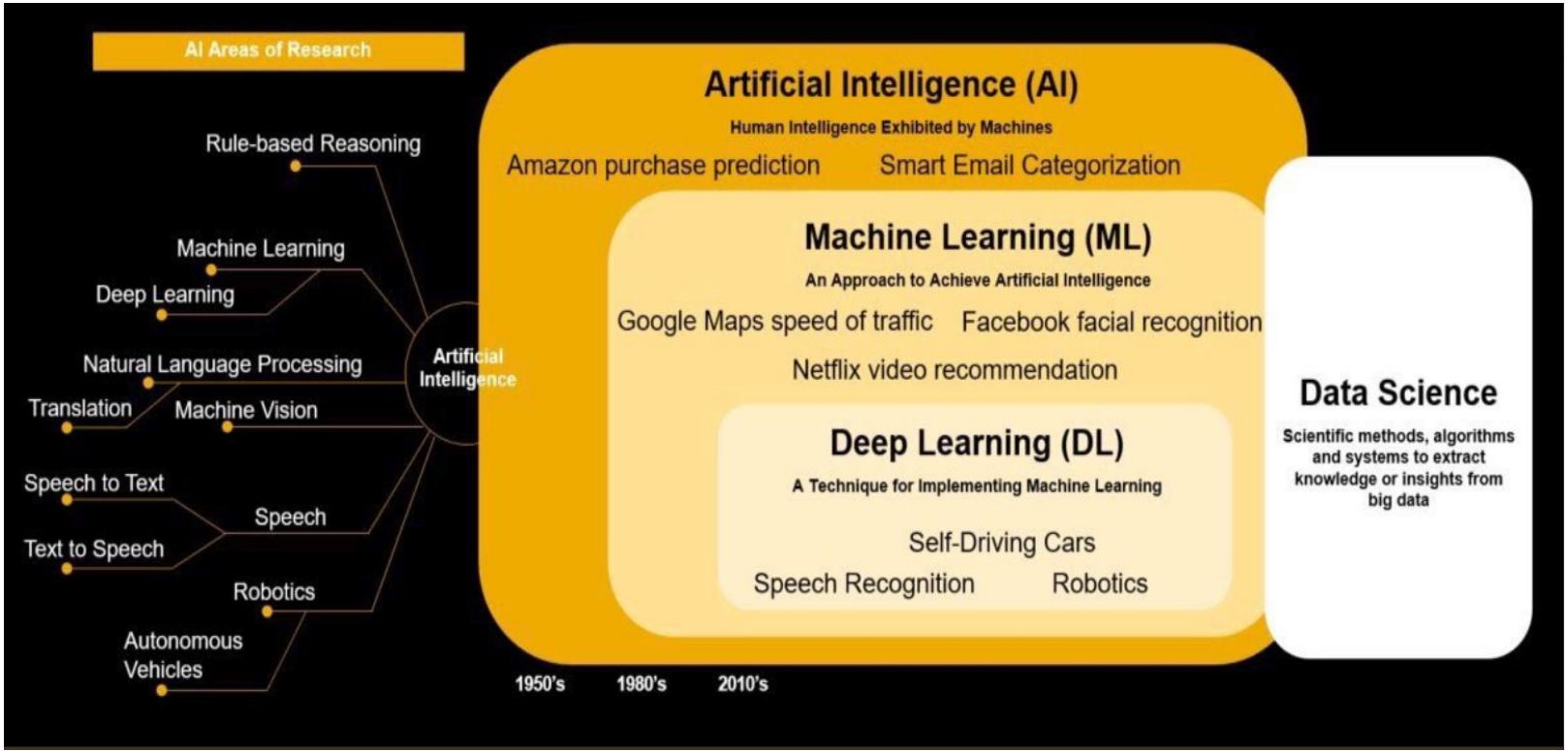
Terminology: From AI to ML to DL



NEURAL NETWORKS

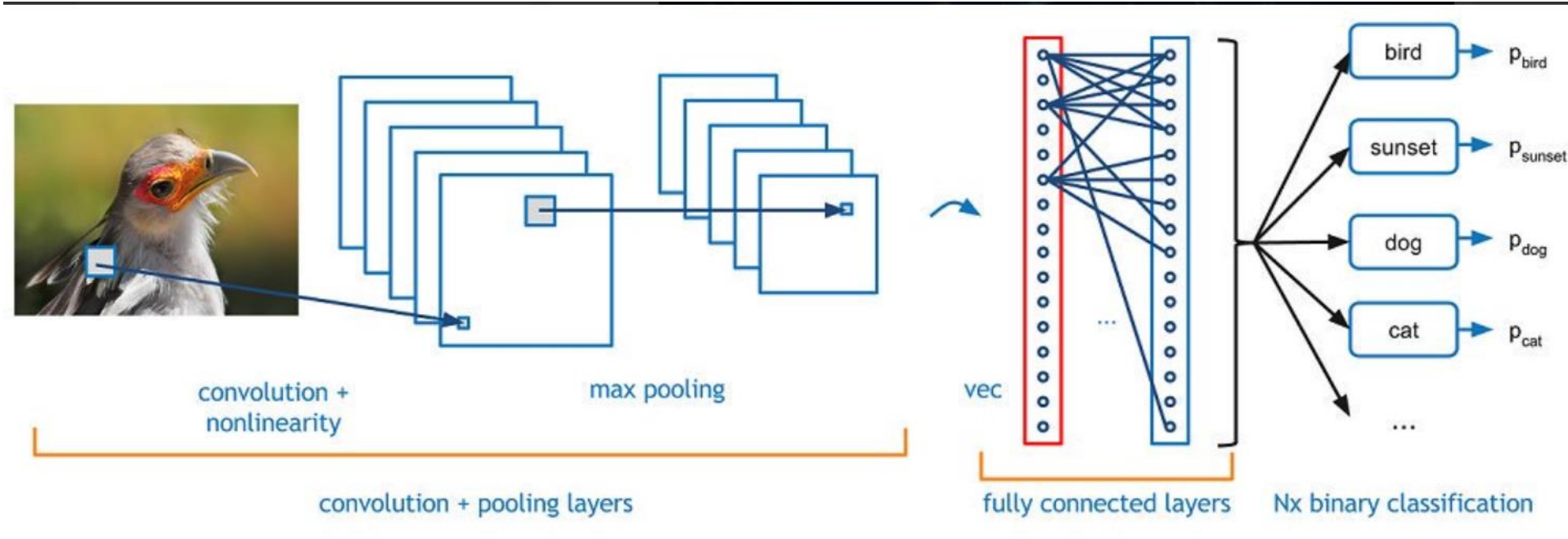
Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

AI, Machine Learning, Deep Learning



https://twitter.com/mikedelgado/status/982340054927331328

Convolutional Neural Networks



- https://dzone.com/articles/a-beginners-guide-to-understanding-convolutional-n

CheXNet: Deep Learning in Radiology

Our algorithm, CheXNet, is a 121-layer **Convolutional neural network** trained on ChestX-ray14, currently the largest publicly available chest X-ray dataset, containing over 100,000 Frontal view X-ray images with 14 diseases.

1. **Four practicing academic radiologists annotate a test set, on which we compare the performance of CheXNet to that of radiologists**.
2. **We find that CheXNet exceeds average radiologist performance** on the F1 metric. (In statistical analysis of binary classification, the F1 score (also F-score or F-measure) is a measure of a test's accuracy.)

https://stanfordmlgroup.github.io/projects/chexnet/

CheXNet: Deep Learning in Radiology

Our algorithm, CheXNet, is a 121-layer **Convolutional neural network** trained on ChestX-ray14, currently the l… publicly available chest … containing over … Frontal … 14 di…

Test Set with 420 images

1. **Four practicing academic radiologists annotate a test set, on which we compare the performance of CheXNet to that of radiologists**.
2. **We find that CheXNet exceeds average radiologist performance** on the F1 metric. (In statistical analysis of binary classification, the F1 score (also F-score or F-measure) is a measure of a test's accuracy.)

https://stanfordmlgroup.github.io/projects/chexnet/

But wait, is it the whole story?

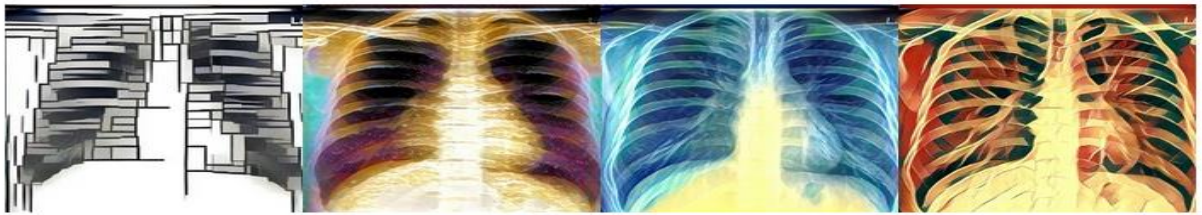
Medium

Bálint Botz Follow

MD/PhD, Diagnostic radiology resident, Researcher mainly focusing on neuroinflammation

Nov 20, 2017 · 4 min read

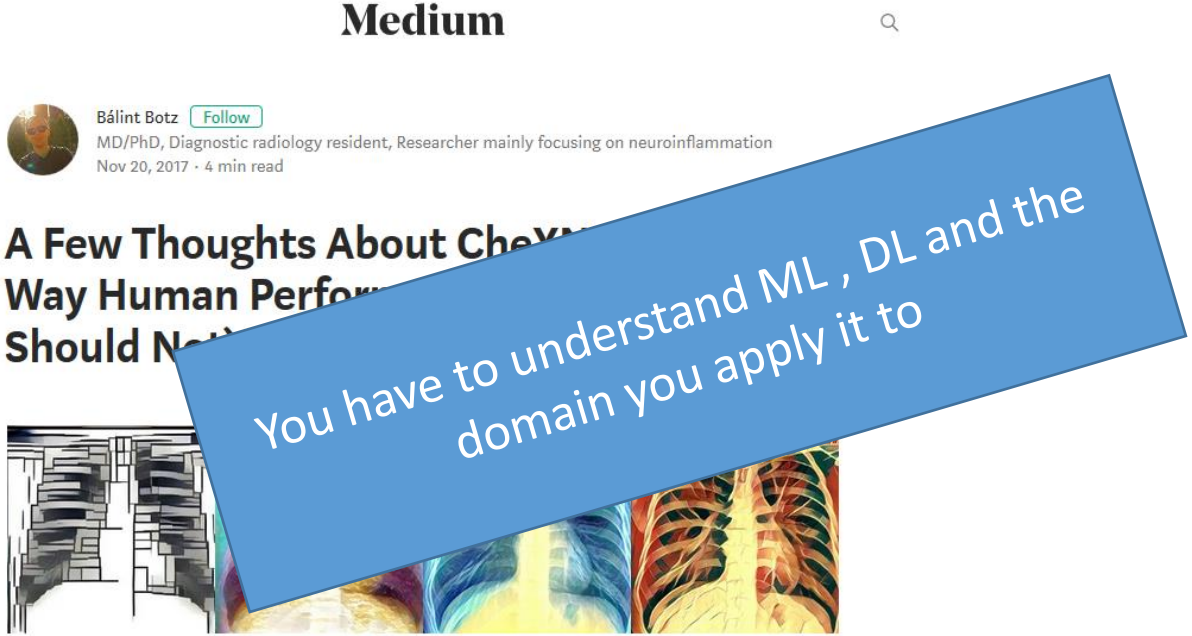
A Few Thoughts About CheXNet—And The Way Human Performance Should (And Should Not) Be Measured



https://medium.com/@BalintBotz/a-few-thoughts-about-chexnet-and-the-way-human-performance-should-and-should-not-be-measured-68031dca7bf

https://lukeoakdenrayner.wordpress.com/2018/01/24/chexnet-an-in-depth-review/

But wait, is it the whole story?



https://medium.com/@BalintBotz/a-few-thoughts-about-chexnet-and-the-way-human-performance-should-and-should-not-be-measured-68031dca7bf

https://lukeoakdenrayner.wordpress.com/2018/01/24/chexnet-an-in-depth-review/

The work horses: Deep (Machine) Learning Frameworks

Caffe theano TensorFlow

Spark MLlib Keras

- http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture8.pdf

The Open Source Caffe Deep Learning Framework

Caffe

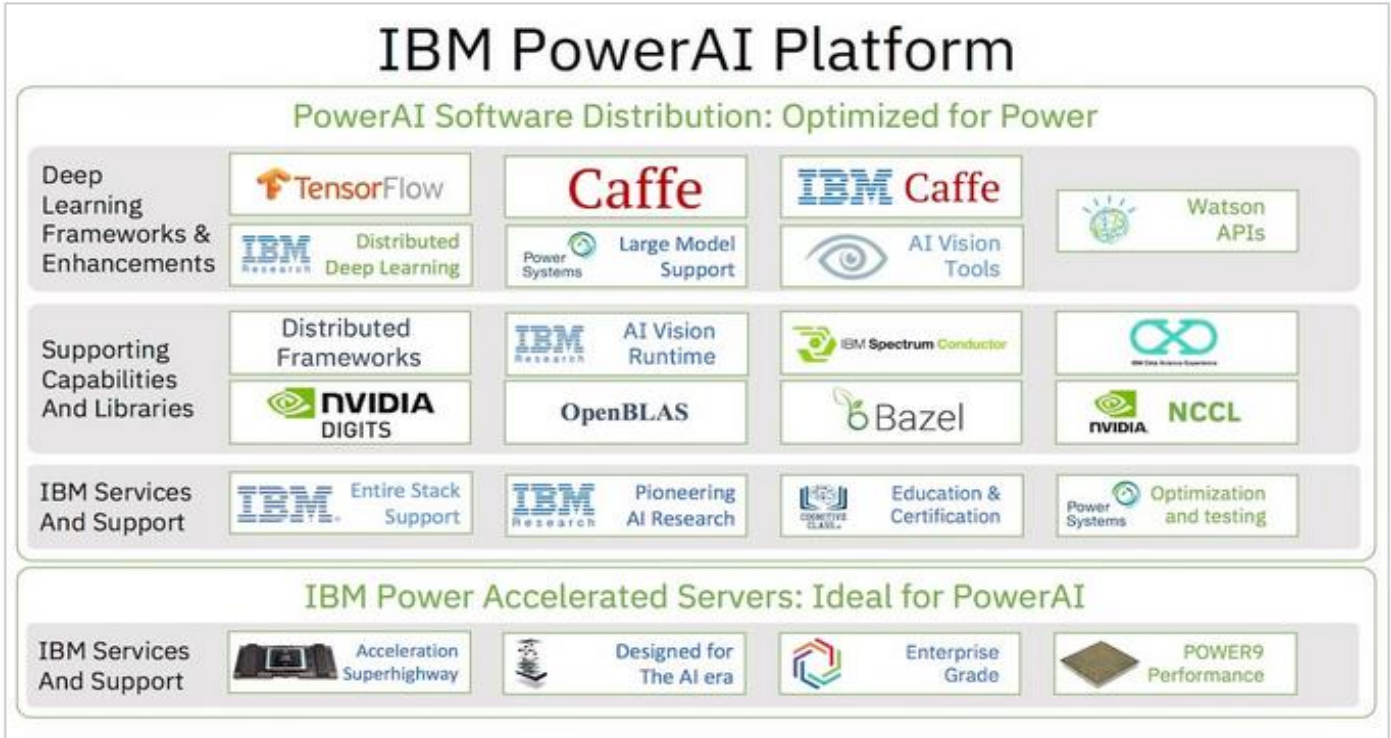
- **Developed at Berkeley**
- **Expressive architecture** encourages application and innovation. Models and optimization are defined by configuration without hard-coding. Switch between CPU and GPU by setting a single flag to train on a GPU machine then deploy to commodity clusters or mobile devices.
- **Extensible code** fosters active development. **In Caffe's first year, it has been forked by over 1,000 developers and had many significant changes contributed back.** Thanks to these contributors the framework tracks the state-of-the-art in both code and models.
- **Speed** makes Caffe perfect for research experiments and industry deployment**. Caffe can process over 60M images per day with a single NVIDIA K40 GPU*.** That's 1 ms/image for inference and 4 ms/image for learning and more recent library versions and hardware are faster still. We believe that Caffe is among the fastest convnet implementations available.

http://caffe.berkeleyvision.org/

Note on Keras

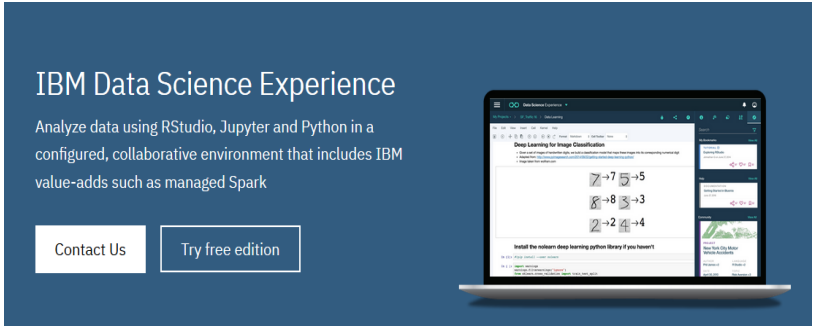
Keras is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation.

IBM PowerAI Platform



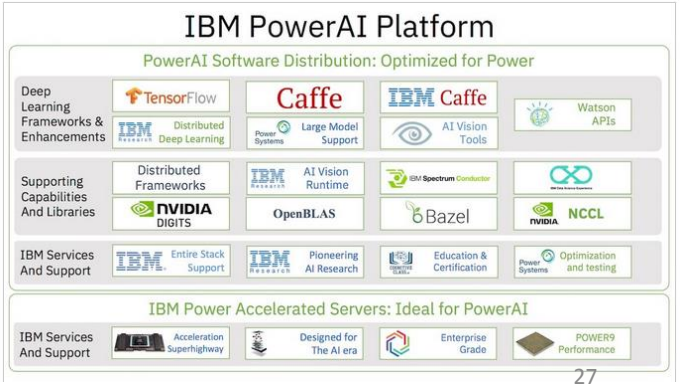
https://www.ibm.com/us-en/marketplace/deep-learning-platform

Watson Studio and PowerAI



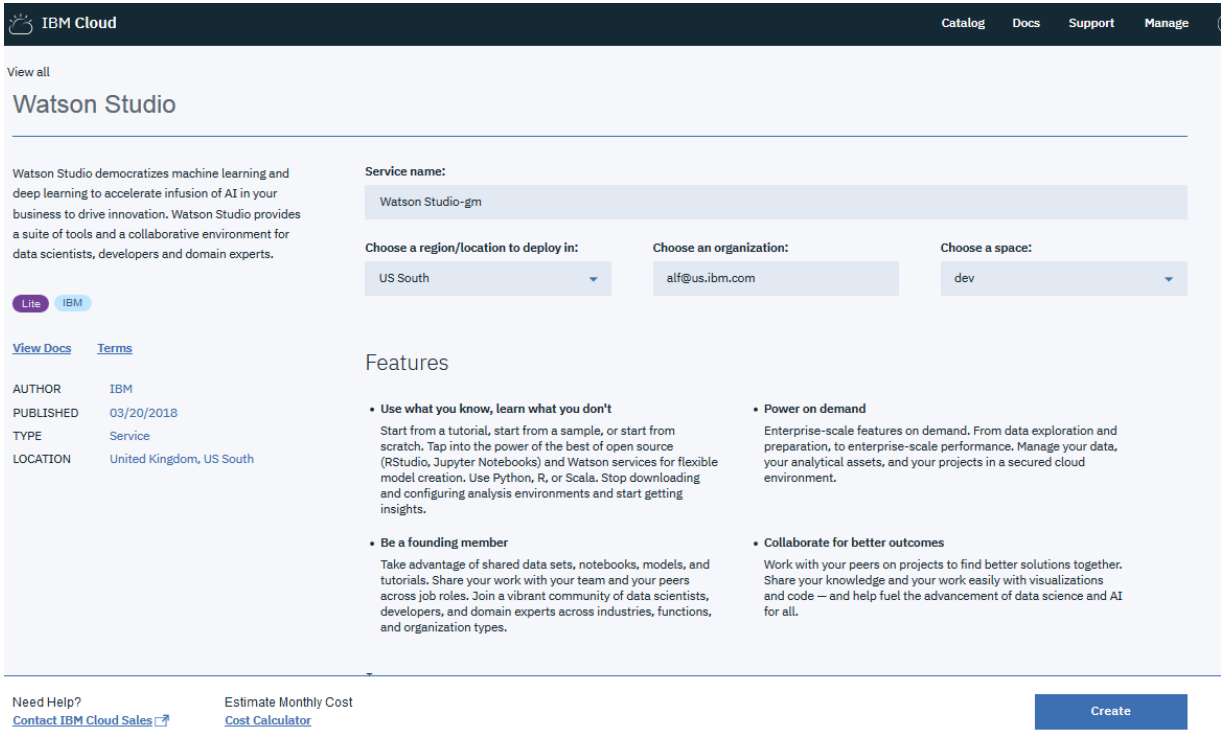
The Data Science Experience is an interactive and collaborate cloud-based environment designed to be **a place where data scientists can use such tools as RStudio, Jupyter, Python, Scala, Spark and IBM's Watson Machine Learning technology** to drive insights into their data and derive information useful to their businesses. It was rolled out last year first for the public cloud, and later was optimized for private clouds.

IBM is bringing the two together by integrating the PowerAI deep learning enterprise software distribution into the Data Science Experience.



Watson in the IBM Public Cloud

Watson Studio in the IBM Cloud



https://console.bluemix.net/catalog/services/watson-studio

Watson Studio

Welcome Lennart!

Watson Studio is part of IBM Watson.

Try out other IBM Watson apps.

Get started with key tasks

Refine data

New notebook

Deep learning

New Modeler flow

New model

Deep Learning in Watson Studio

What is Deep Learning?

IBM's experiment-centric deep learning service within Watson Studio allows data scientists to visually design their neural networks and scale out their training runs while auto-allocation means paying only for the resources utilized. Optimized for production environments, scale up your training using the NVIDIA® Tesla® V100 GPU with your preferred deep learning framework then easily deploy to the cloud or at the edge.

Deep Learning features

Experiment Assistant

Initiate and monitor batch training experiments then compare cross-model performance in real-time without worrying about log transfers and scripts to visualize results. You focus on designing your neural networks. We'll manage and track your assets.

Open and flexible

Use your preferred deep learning framework: Tensorflow, Keras, PyTorch, Caffe and more. Manage your deep learning experiments with the tools you prefer: command-line interface (CLI), Python library or an interactive user interface.

Elastic GPU compute

Train neural networks in parallel using market-leading NVIDIA® Tesla® GPUs - K80, P100, and V100. Pay only for what you use. Auto-allocation means no more remembering to shutdown your cloud training instances. No clusters or containers to manage.

Hyperparameter optimization

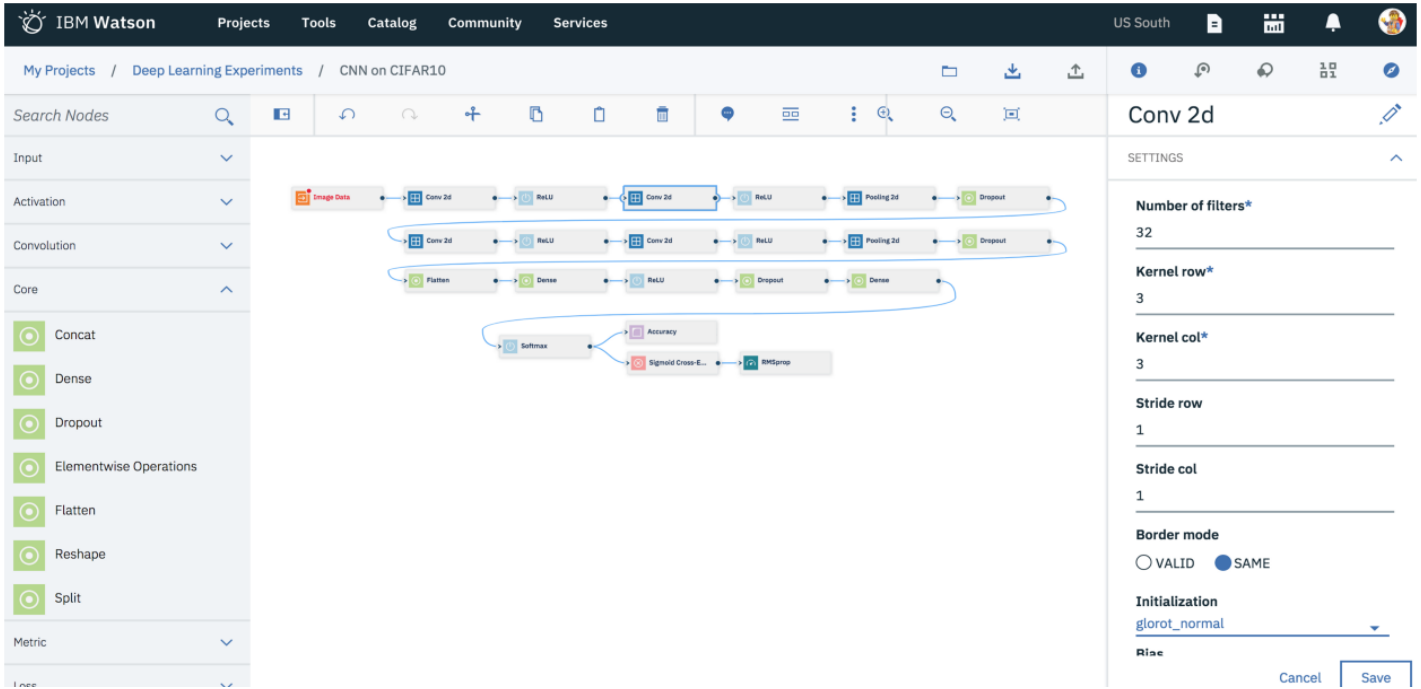
Efficiently automate searching your network's hyperparameter space to ensure the best model performance with the fewest training runs.

Neural Network Modeler (beta)

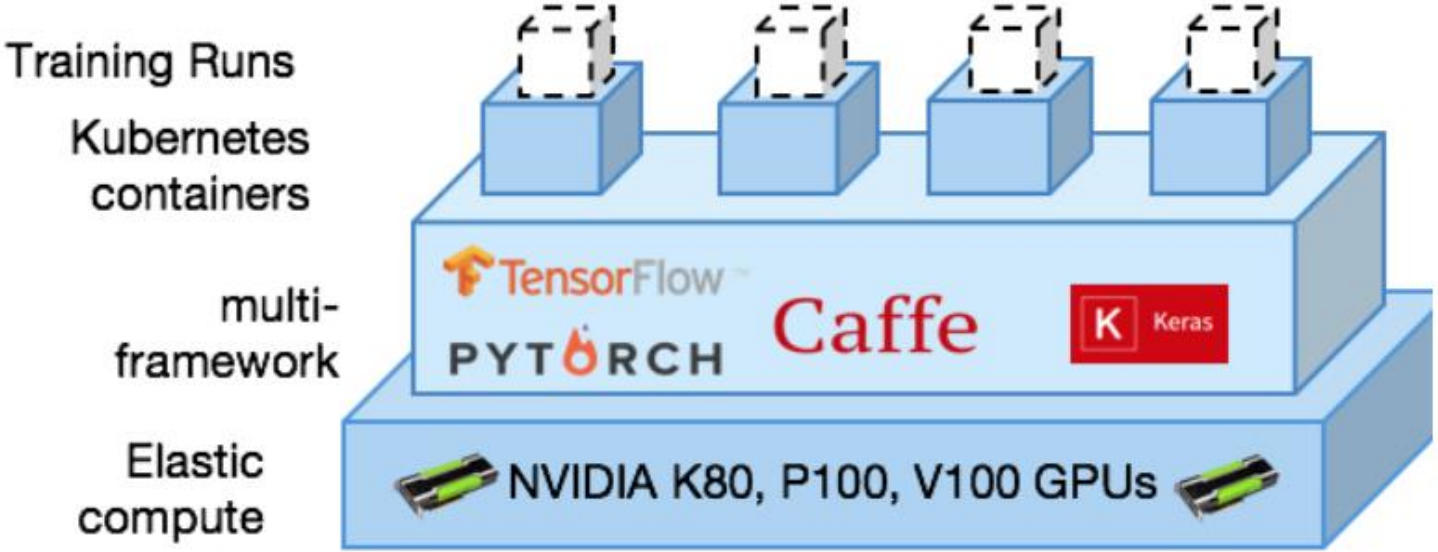
Visually design your neural networks. Drag-and-drop layers of your neural architecture then configure and deploy using the most popular deep learning frameworks.

Let's talk

IBM Watson Studio and Neural Network Modeler

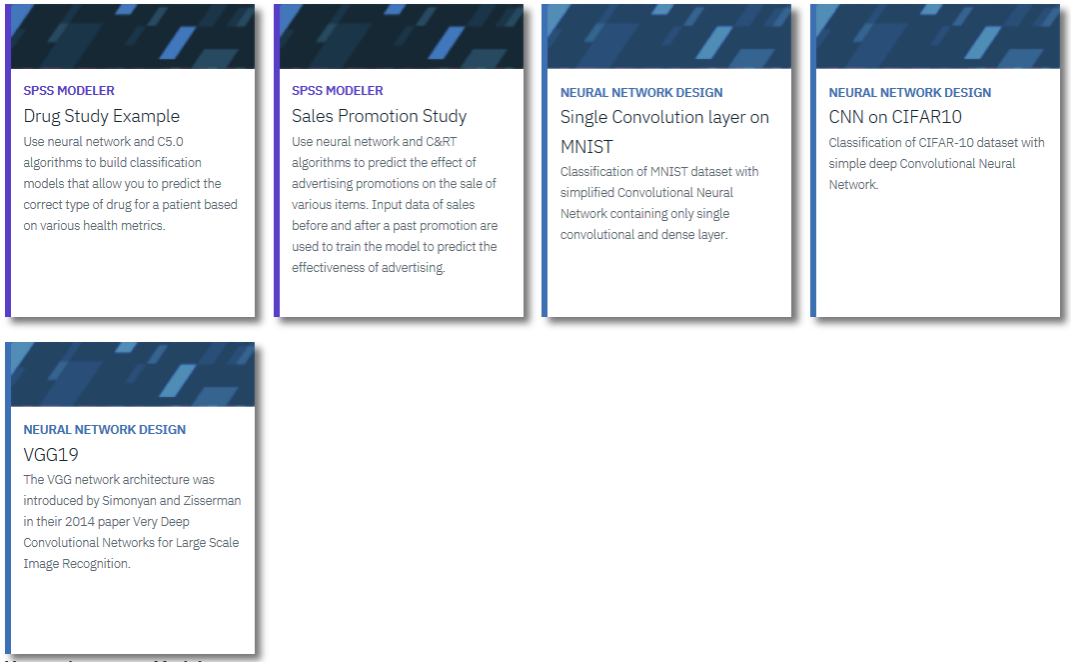


Deep Learning as a Service within Watson Studio



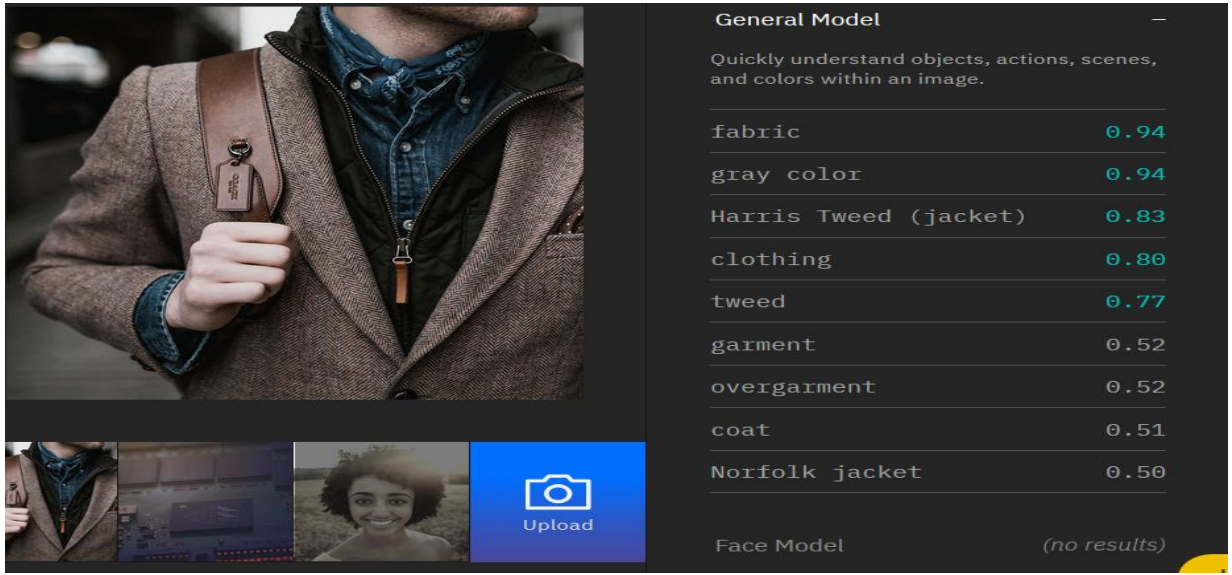
https://www.ibm.com/blogs/research/2018/03/deep-learning-advances/

Watson Studio Modeler, From Example



You can't create a Modeler yet.

Watson Visual Recognition

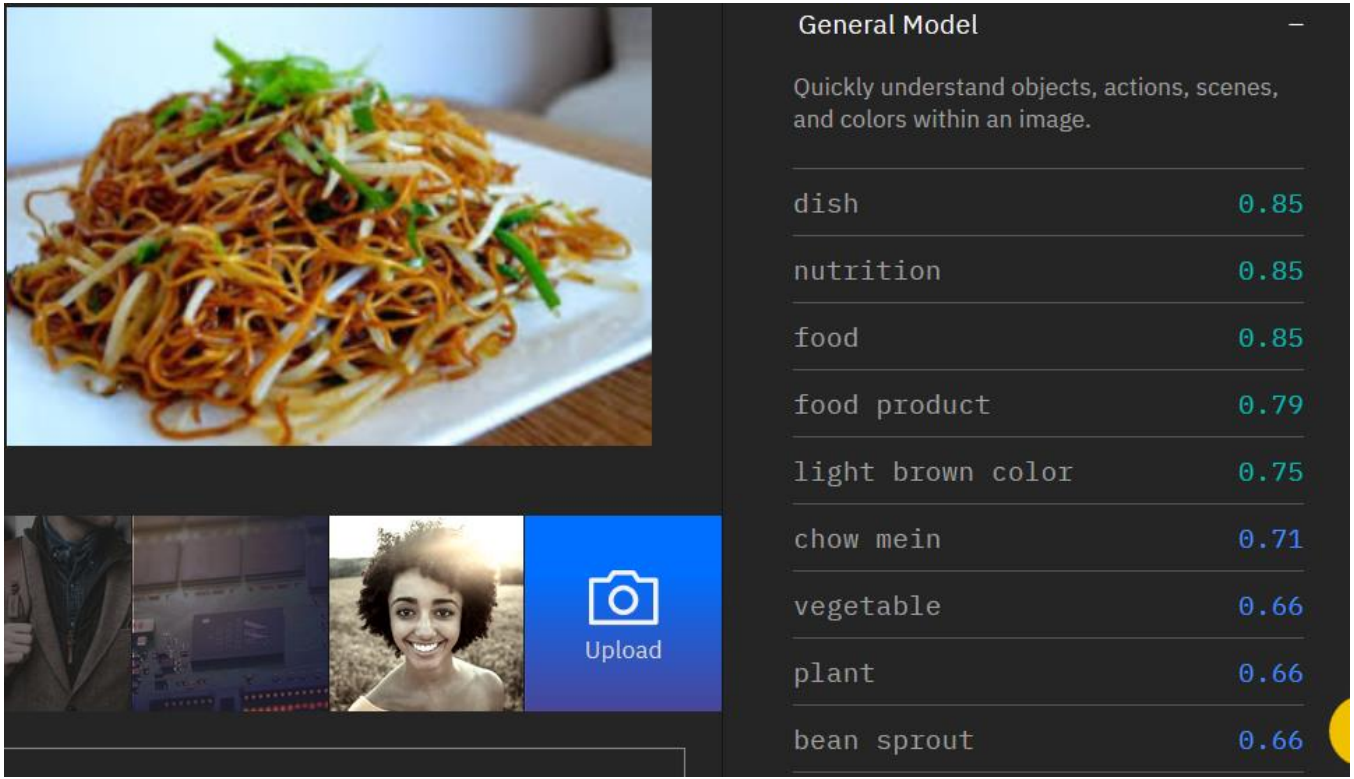


Quickly and accurately tag, classify and train visual content using machine learning.

Train models effortlessly with Watson Studio

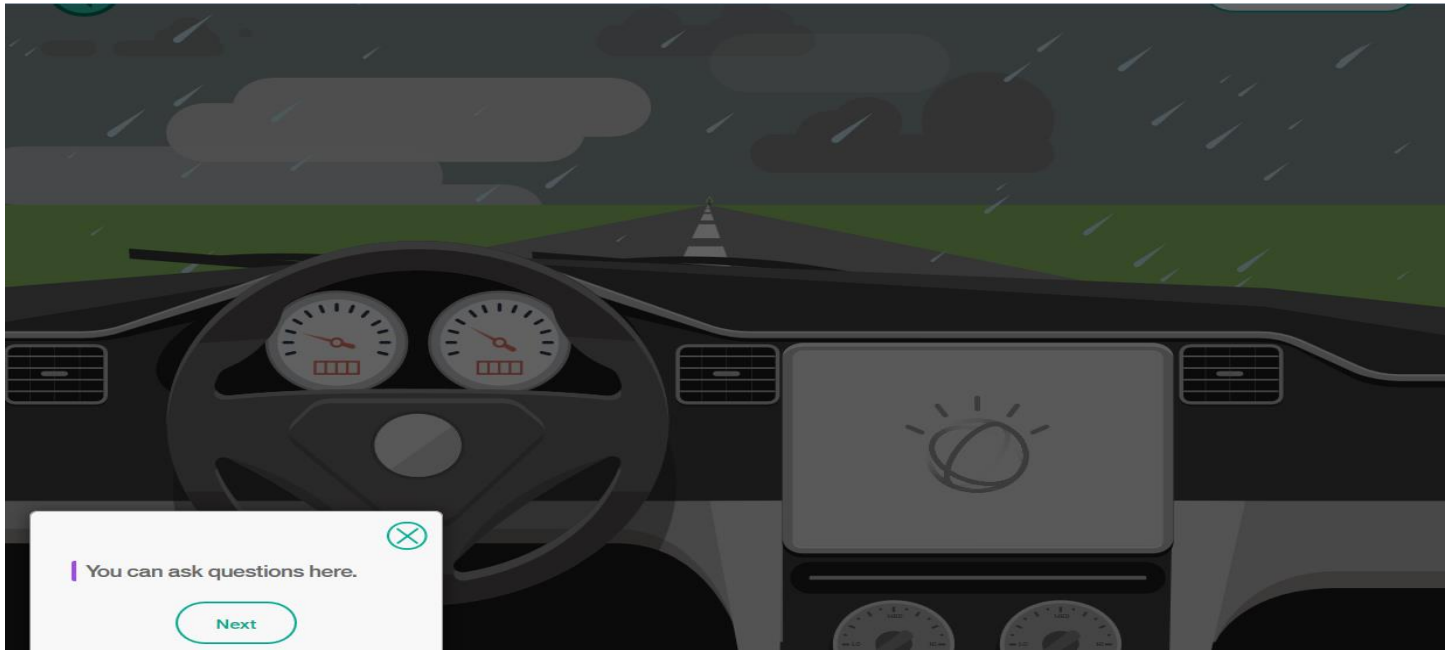
https://www.ibm.com/watson/services/visual-recognition/demo/index.html#watson-demo

How about ML and some Chow Mien?



https://www.ibm.com/watson/services/visual-recognition/demo/

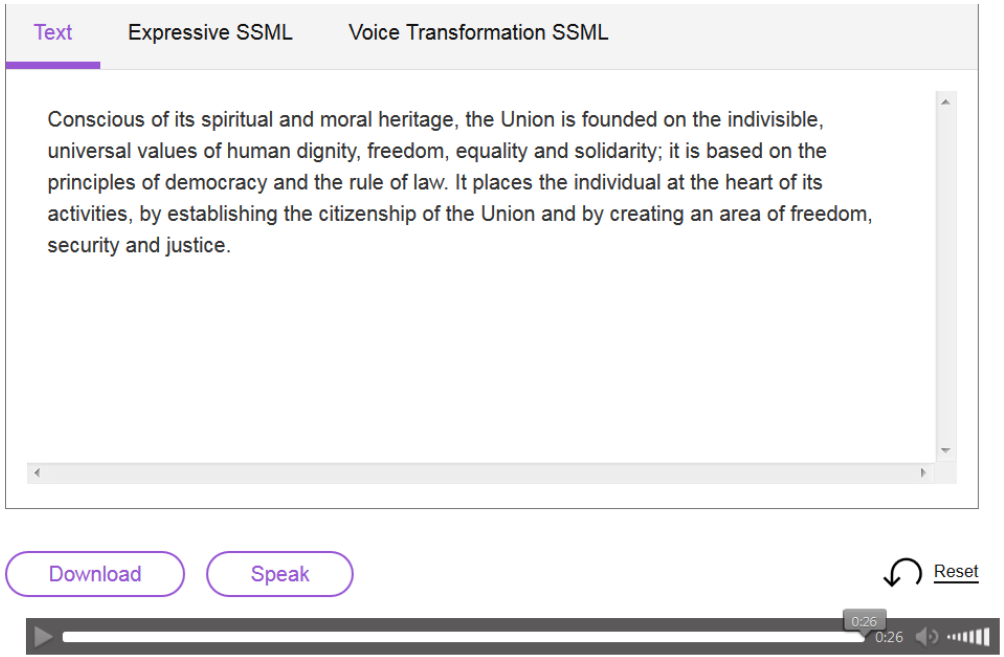
Let’s go for a drive with a Chatbot



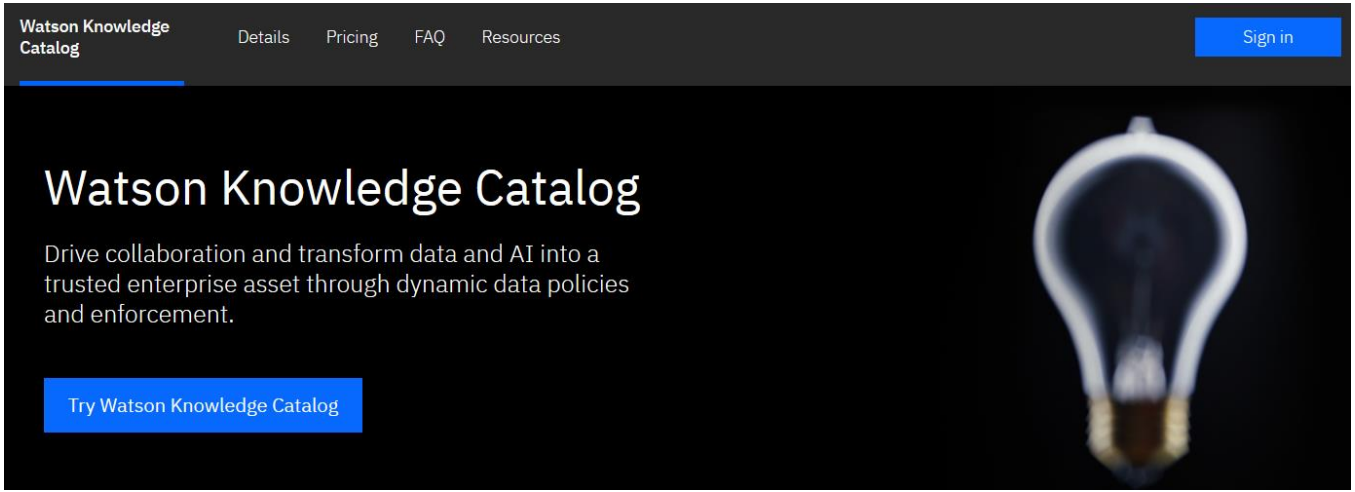
- https://conversation-demo.ng.bluemix.net/

Watson Text to Speech

- https://text-to-speech-demo.ng.bluemix.net/

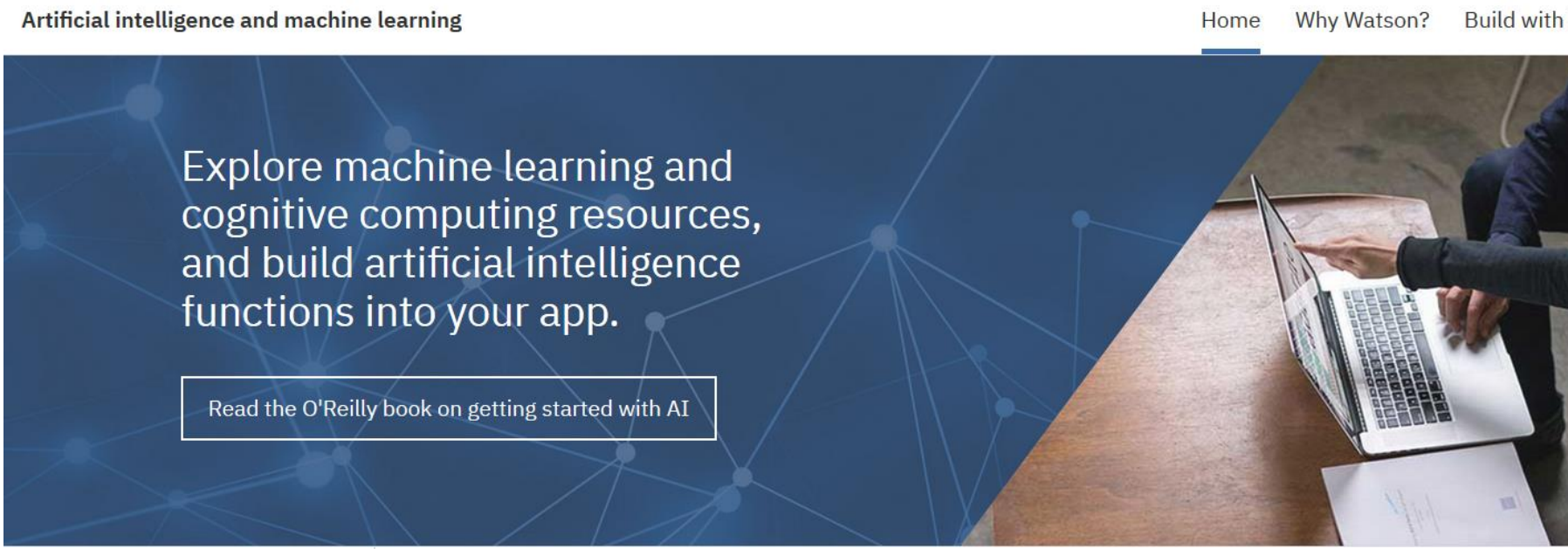


Watson Knowledge Catalog



https://www.ibm.com/blogs/bluemix/2018/03/ibm-watson-knowledge-catalog-find-prepare-data-analytic-assets-power-ai/

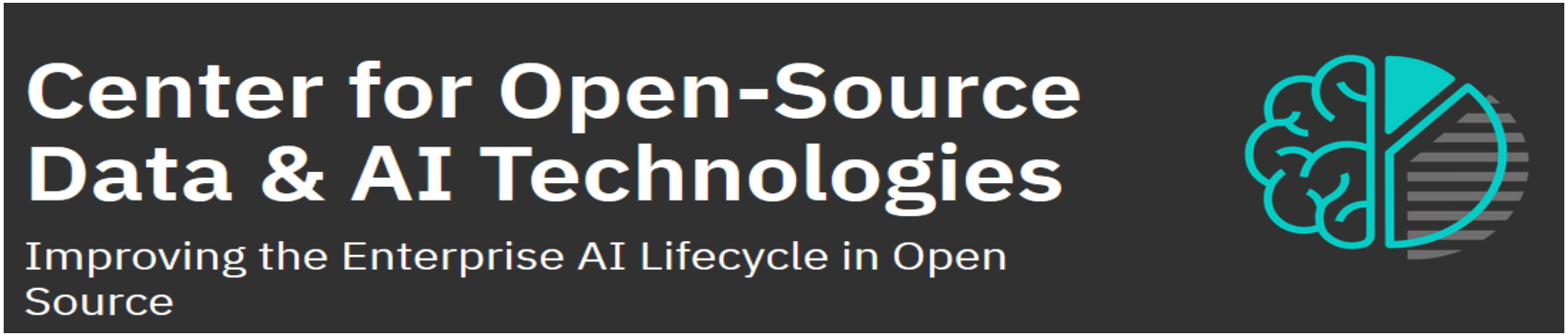
Watson Machine Learning



https://developer.ibm.com/ai/

IBM’s CodaIT

IBM CodaIT



https://developer.ibm.com/code/open/centers/codait/

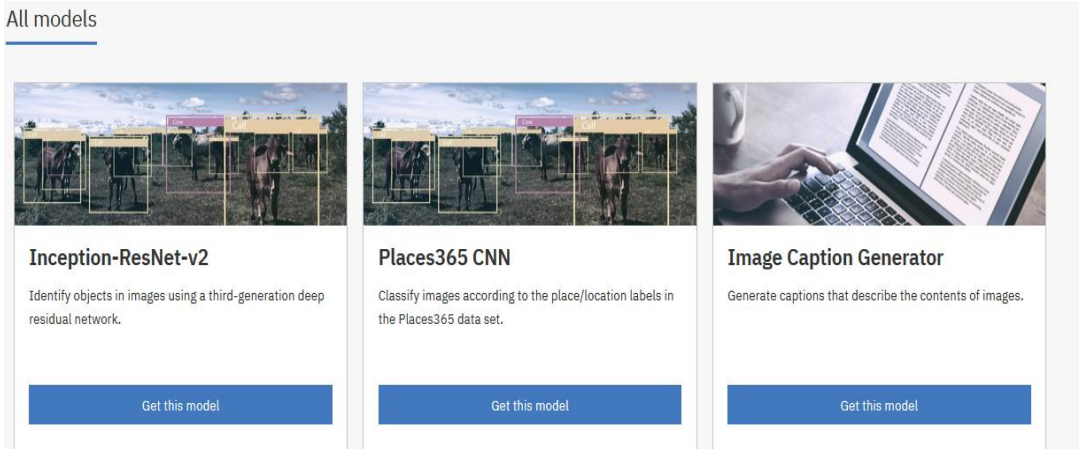
Model Asset eXchange

- MAX is a one-stop exchange for data scientists and AI developers to consume models created using their favorite machine learning engines, like TensorFlow, PyTorch, and Caffe2
- provides a standardized approach to classify, annotate, and deploy these models for prediction and inferencing,
- including models that can be deployed and customized in IBM's AI application development platform, **Watson Studio**.

IBM Code Model Asset Exchange

A place for developers to find and use free and open source deep learning models.

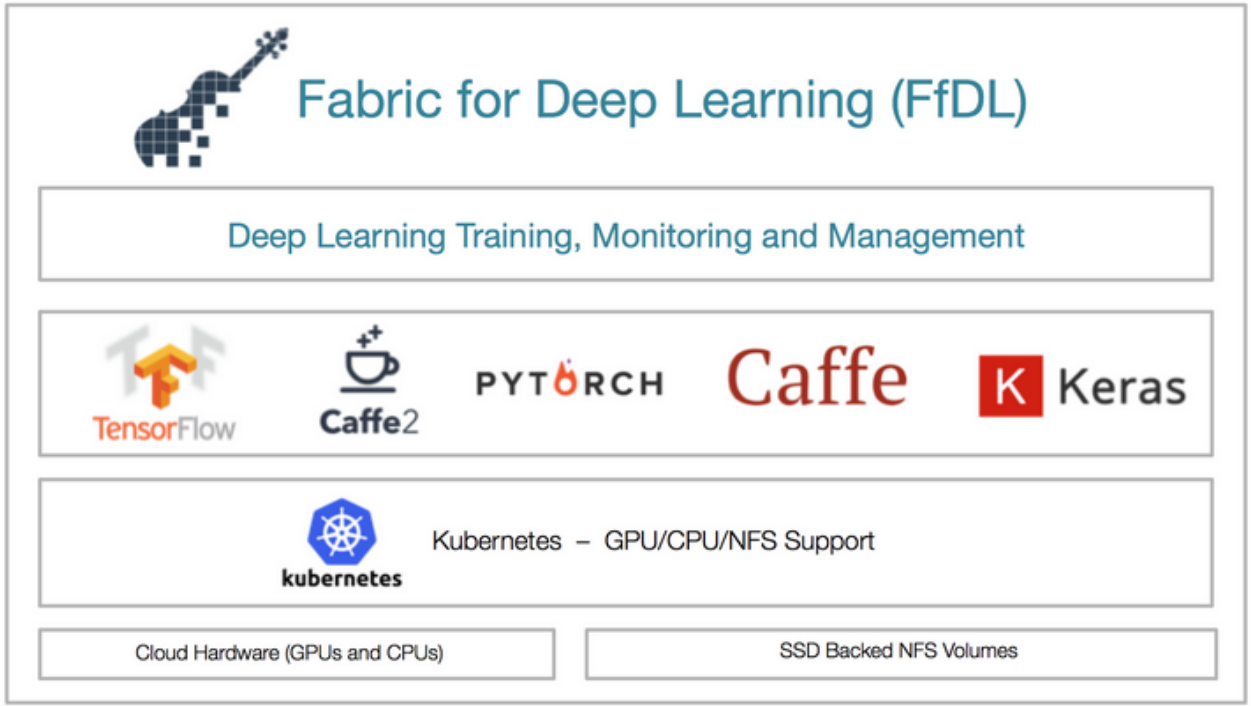
https://developer.ibm.com/code/exchanges/models/



a one-stop exchange for data scientists and AI developers to consume models created using their favorite machine learning engines, like TensorFlow, PyTorch, and Caffe2

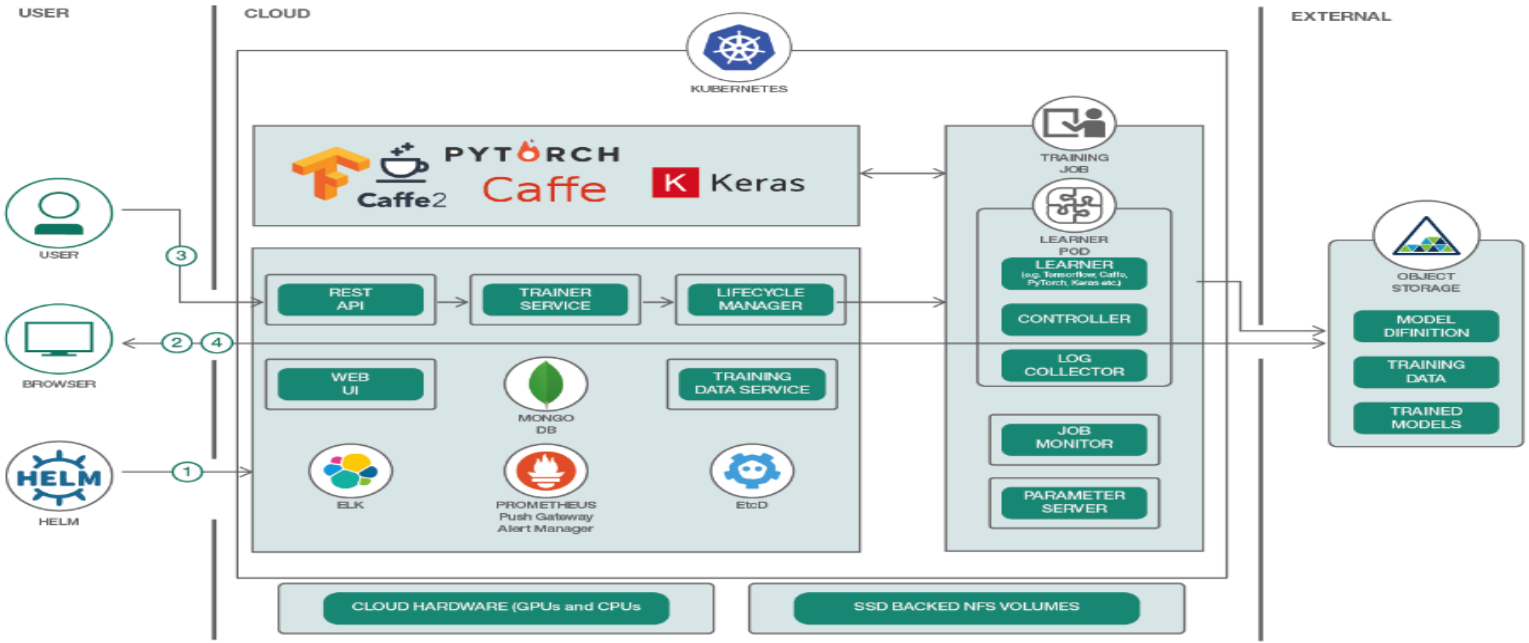
including models that can be deployed and customized in IBM's AI application development platform, **Watson Studio**.

Fabric for Deep Learning(FfDL)



https://developer.ibm.com/code/2018/03/20/fabric-for-deep-learning/

Install and consume a deep learning platform on Kubernetes with TensorFlow, Caffe, PyTorch, and more



https://developer.ibm.com/code/patterns/deploy-and-use-a-multi-framework-deep-learning-platform-on-kubernetes/

Adversarial Robustness Toolbox (ART v0.1)

This is a library dedicated to **adversarial machine learning**. Its purpose is to allow rapid crafting and analysis of attacks and defense methods for machine learning models. The Adversarial Robustness Toolbox provides an implementation for many state-of-the-art methods for attacking and defending classifiers.

The library is still under development. Feedback, bug reports and extension requests are highly appreciated.

Supported attack and defense methods

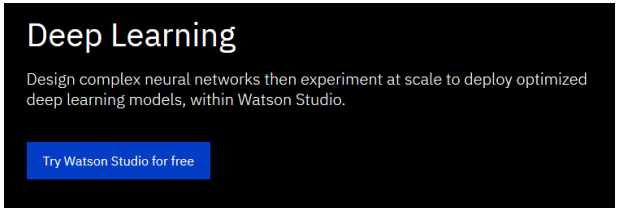
The Adversarial Robustness Toolbox contains implementations of the following attacks:

- Deep Fool (Moosavi-Dezfooli et al., 2015)
- Fast Gradient Method (Goodfellow et al., 2014)
- Jacobian Saliency Map (Papernot et al., 2016)
- Universal Perturbation (Moosavi-Dezfooli et al., 2016)
- Virtual Adversarial Method (Moosavi-Dezfooli et al., 2015)
- C&W Attack (Carlini and Wagner, 2016)
- NewtonFool (Jang et al., 2017)

https://github.com/IBM/adversarial-robustness-toolbox

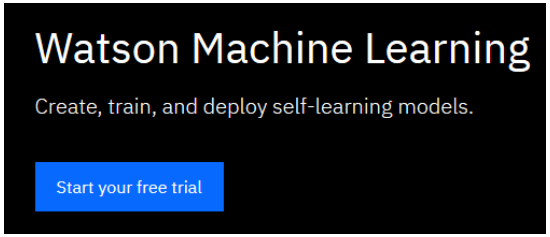
Where do we go from here?

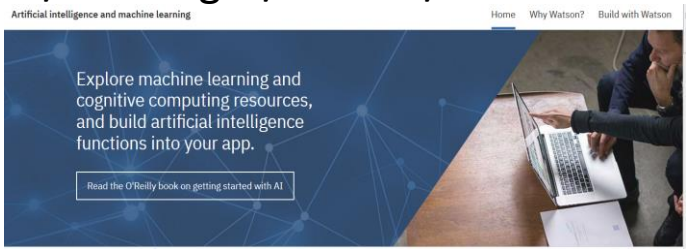
- kaggle.com Open Source Datasets



https://www.ibm.com/cloud/machine-learning

https://developer.ibm.com/code/exchanges/models/





https://www.ibm.com/cloud/machine-learning

https://developer.ibm.com/ai/

IBM Machine Learning for z/OS

A private cloud (on-premises) deployment of IBM Machine Learning

Cost effective, low latency, high security

Benefit from your private cloud (on-premises) z Systems investments

- Gain advantage from z Systems infrastructure, people and processes
- Leverage z Systems data in place while combining structured and unstructured data from z and non-z data sources
- Access live transactional data

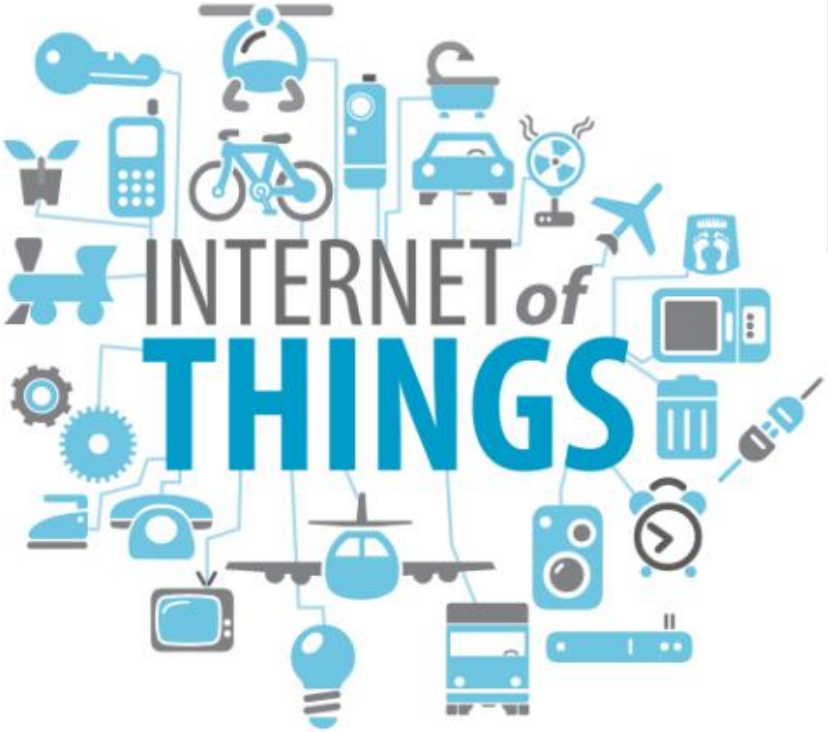
Tell me about the terminology

SAN FRANCISCO — Apple has hired Google's chief of search and artificial intelligence, John Giannandrea, a major coup in its bid to catch up to the artificial intelligence technology of its rivals.

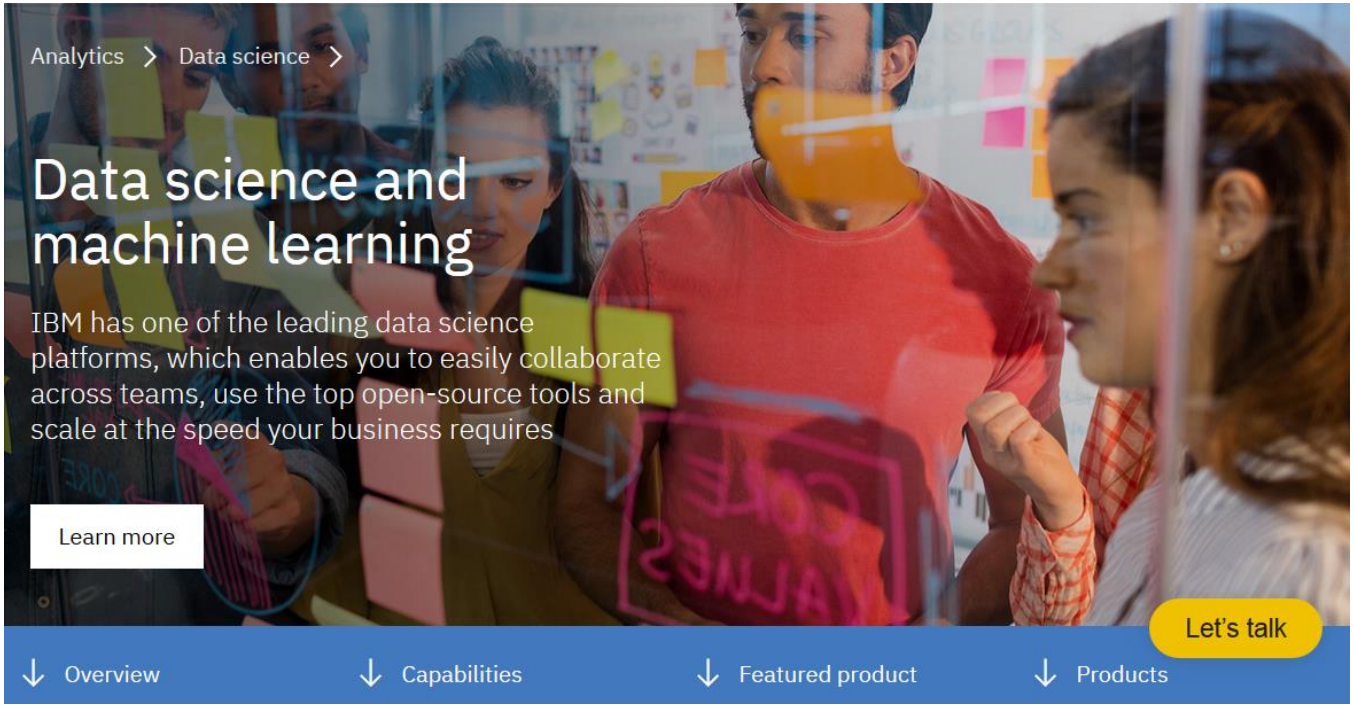
Apple said on Tuesday that Mr. Giannandrea will run Apple's "**machine learning and A.I. strategy,**" and become one of 16 executives who report directly to Apple's chief executive, Timothy D. Cook.

https://www.nytimes.com/2018/04/03/business/apple-hires-googles-ai-chief.html

Big Data, Hardware and the GPU



Data science and machine learning at IBM



Runtimes Supported by Watson Machine Learning*

Machine Learning(ML)

More frameworks including deep learning are on the Roadmap

*IBM ML in DSX Local is currently in Beta and only supports Spark ML models (scikit-learn and other deep learning frameworks are on the roadmap).