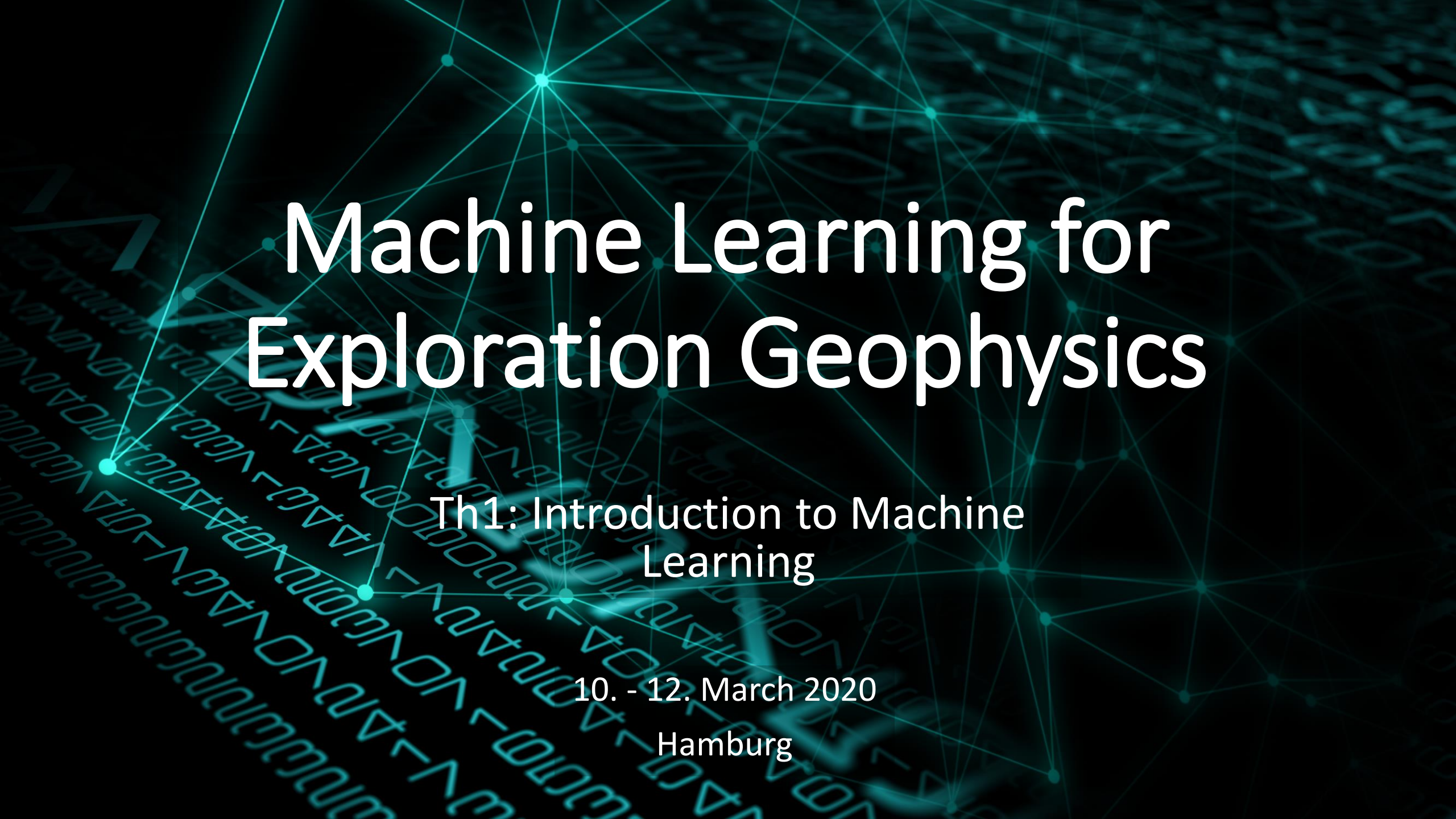


Machine Learning for Exploration Geophysics

Th1: Introduction to Machine Learning

10. - 12. March 2020

Hamburg

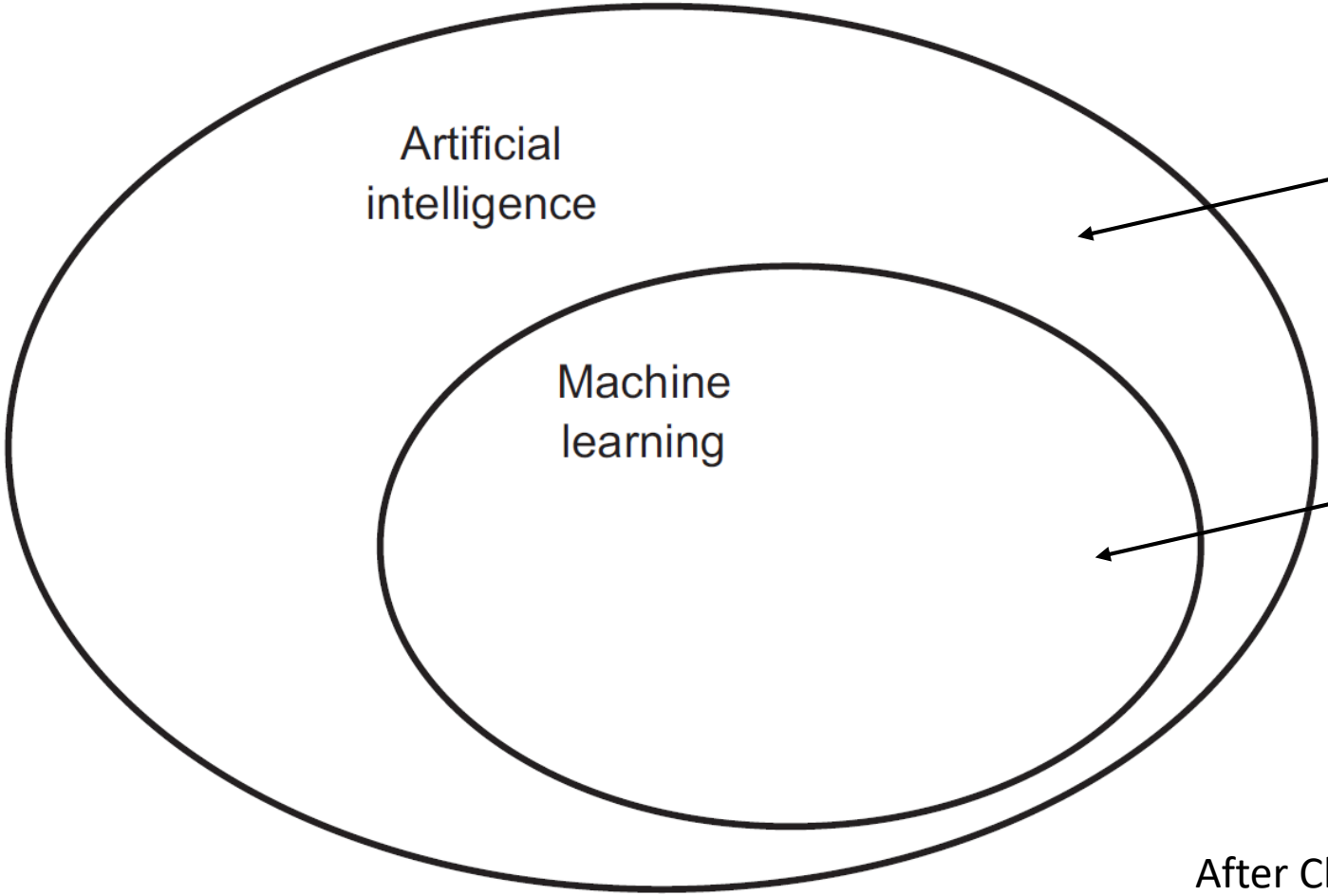
Outline

- Aim of the course
- AI, ML & DL – what is the difference?
- What is ML?
- Map of the ML world
- Main categories of ML algorithms
- Why is there so much hype around ML?
- Why ML is a Game Changer?
- Courses, Tutorials, Books
- Online resources
- ML Packages

Aim of the course

- Introduce how ML is used in geophysical applications
- Give an understanding of the workflows used in ML
- Learn the high-level principles
- Give references for the used algorithms and software
- Practice the methods on Python-related codes

What is the difference between Artificial Intelligence, Machine Learning, and Deep Learning?



- Automated planning and scheduling
- Natural language processing
- Perception
- Robotics
- Social intelligence
- General intelligence

Dozens of different ML methods!

After Chollet, 2017

Chollet, F. (2017) Deep learning with Python.

What is Machine Learning?

"Field of study that gives computers the ability to learn without being explicitly programmed"

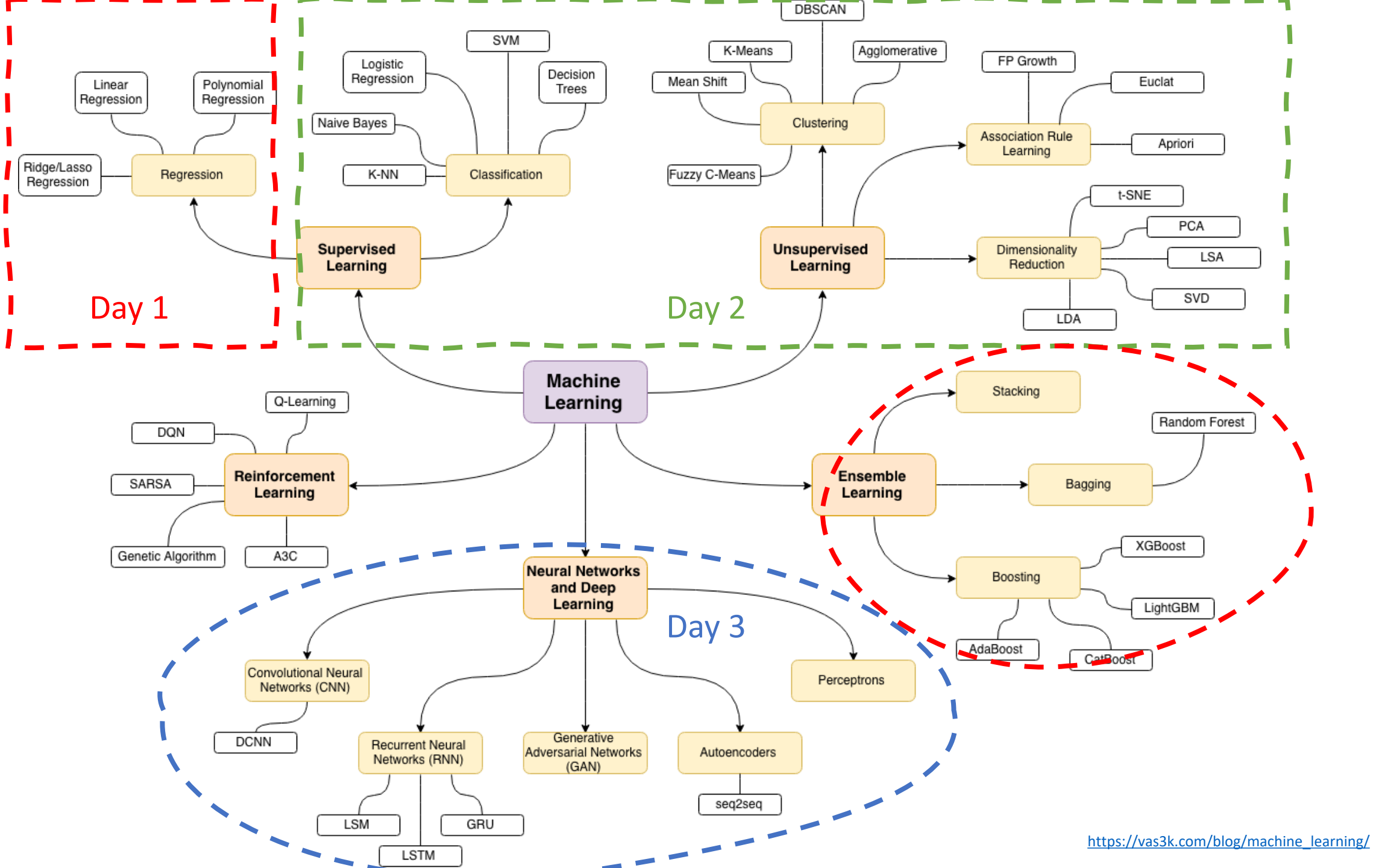
(Arthur Samuel, 1959)

"A computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program"

"Programming computers to learn from experience should eventually eliminate the need for much of this detailed programming effort"

Professor Arthur Lee Samuel is known for his groundbreaking work in computer checkers in 1959

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. IBM Journal of research and development, 3(3), 210-229.
Figure source: https://en.wikipedia.org/wiki/Arthur_Samuel



https://vas3k.com/blog/machine_learning/

Tuesday, March 10th			Wednesday, March 11th	Thursday, March 12th
8:30	9:00	Question Time (optional)	Question Time (optional)	Question Time (optional)
9:00	10:00	Th1: Introduction to Machine Learning	Th3: Classification	Th5: Deep Neural Networks
10:00	11:00	Th2: Basic Principles of Supervised Learning	Th4: Unsupervised Learning	Th6: Convolutional Neural Networks
11:00	11:30	*Coffee Break*	*Coffee Break*	*Coffee Break*
11:30	13:00	CC1: Regression	CC3: Classification examples	CC5: DNN example
13:00	14:00	*Lunch Break*	*Lunch Break*	*Lunch Break*
14:00	15:30	CC2: Ensemble Learning	CC4: Clustering	CC6: CNN example
15:30	16:00	*Coffee Break*	*Coffee Break*	*Coffee Break*
16:00	17:30	Kaggle Competition 1	Kaggle Competition 2	Kaggle Competition 3
17:30	18:00	Question Time (optional)	Question Time (optional)	Question Time (optional)

Three main categories of ML algorithms

- Supervised Learning
  - Labeled data
  - Make predictions about the output and then compare it with the actual answer
- Unsupervised learning
  - Unlabeled data
  - Objective: create some structure from the data
    - Clustering
    - Association
- Reinforcement learning
  - No data, but we have an environment to interact with
  - Learn by trial and error

Why is there so much hype around ML?

- Machine learning has been around for decades (centuries)
- Traditional neural networks are used in the industry during the last 30 years

- So, what's new? Why ML is a Game Changer?

Why ML is a Game Changer?

- Big data
- Cloud computing
- Advanced Algorithms
- Enthusiasm
- Openness of the AI community

THE 4 V'S OF BIG DATA

Volume

SCALE OF DATA

40 ZETTABYTES
of data will be created by 2020, an increase of 300 times from 2005

6 BILLION PEOPLE
have cell phones
WORLD POPULATION: 7 BILLION

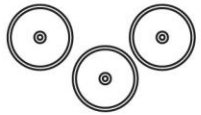
2.5 QUINTILLION BYTES
of data are created each day

Most companies in the U.S. have at least
100 TERABYTES
of data stored

Variety

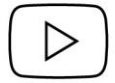
DIFFERENT FORMS OF DATA

As of 2011, the global size of data in healthcare was estimated to be
150 EXABYTES



30 BILLION PIECES OF CONTENT
are shared on facebook every month

4 BILLION + HOURS OF VIDEO
are watched on You Tube each month



4 MILLION TWEETS
are sent per day by about 200 million monthly active users

Velocity

ANALYSIS OF STREAMING DATA

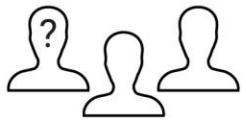
The New York Stock Exchange captures
1TB OF TRADE INFORMATION
during each trading session

Modern cars have close to
100 SENSORS
that monitor items such as fuel level and tire pressure

Veracity

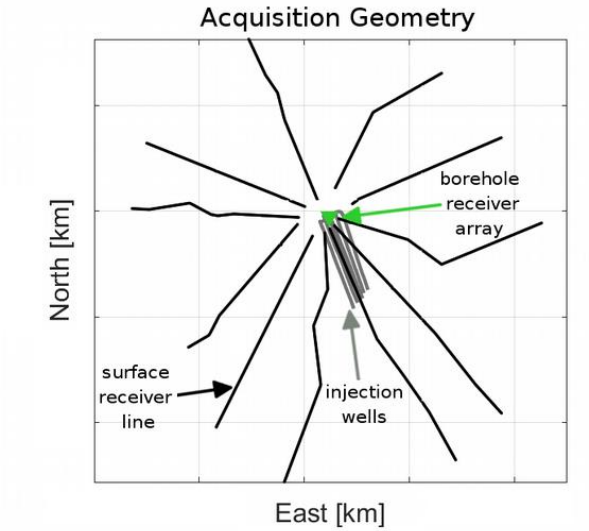
UNCERTAINITY OF DATA

1 IN 3 BUSINESS LEADERS
don't trust the information they use to make decisions



27% OF RESPONDENTS
in one survey were unsure of how much of data was inaccurate

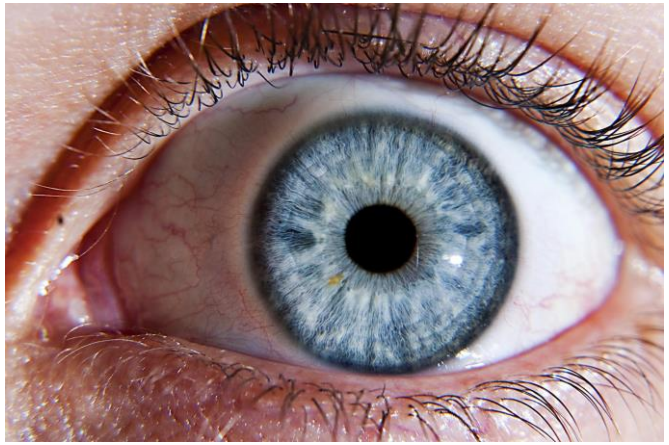
Figure sourse: https://www.ibmbigdatahub.com/infographics/four-vs-big-data



1500 stations x 500 samples/second =
0.75 Megapixel/second

1280 x 720 x 3 x 30 fps =
83 Megapixel/second

3840 x 2160 x 3 x 60 fps =
1.5 Gigapixel/second



576 Megapixel/frame x 1000 fps =
500 Gigapixel/second

Machine Learning Cloud Service Providers

- AWS Google Cloud
  - https://aws.amazon.com/machine-learning/amis
- Google Cloud
  - https://cloud.google.com/products/ai
- Google Colab
  - https://colab.research.google.com
- IBM Watson
  - https://www.ibm.com/watson
- Kaggle
  - https://www.kaggle.com
- Microsoft Azure
  - https://azure.microsoft.com

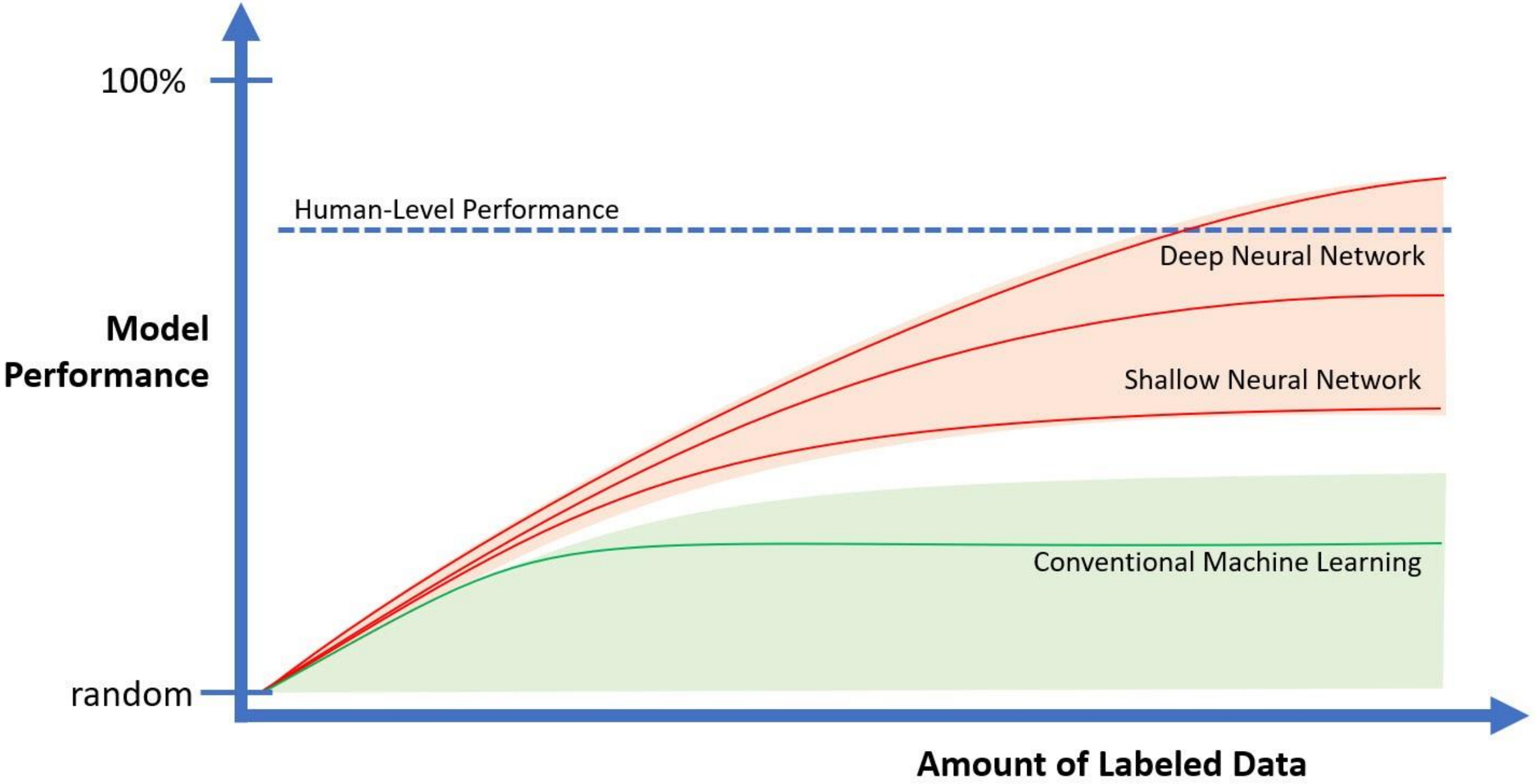
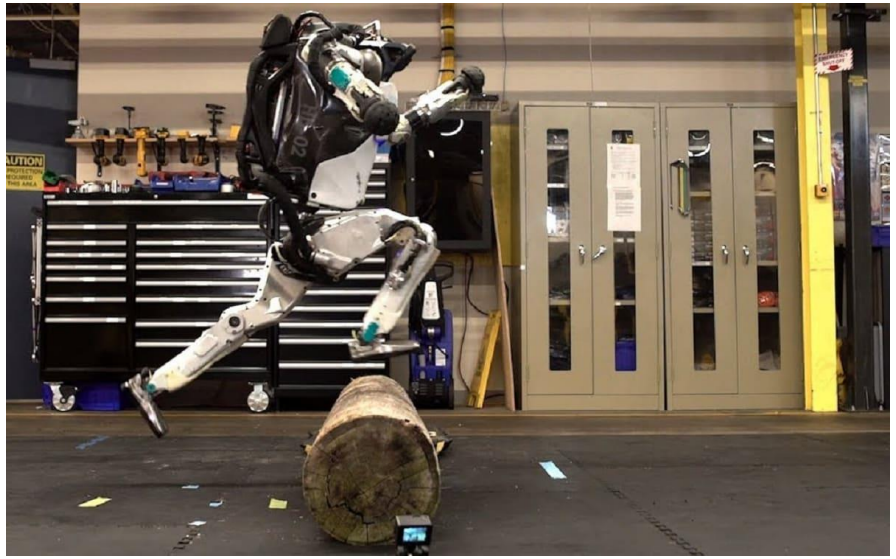


Figure sourse: Hackathorn, R. (2018) How Managers Should Prepare for Deep Learning: New Values, http://bit.ly/3cvfIO1

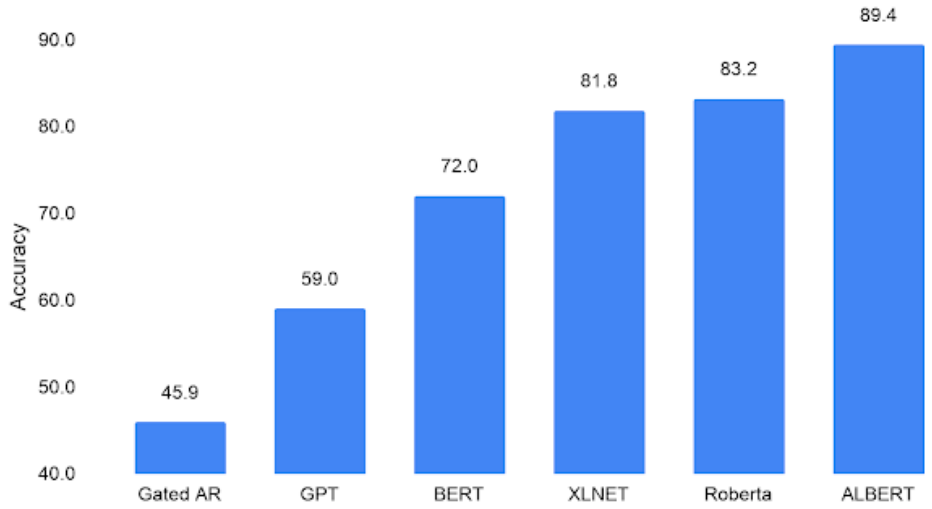
Why ML is a Game Changer?

- Big data
- Cloud computing
- Advanced Algorithms
- Enthusiasm
- Openness of the AI community

Self-Driving cars



Robots on two legs



ALBERT

OpenAI Five first beats the world champions of Dota 2

Figure sourses: https://www.tesla.com/videos/autopilot-self-driving-hardware-neighborhood-short; https://www.bostondynamics.com/; https://ai.googleblog.com/2019/12/albert-lite-bert-for-self-supervised.html; https://openai.com/blog/openai-five-defeats-dota-2-world-champions/

Online courses

- Machine Learning Course
  - https://www.coursera.org/learn/machine-learning
- Machine Learning Crash Course
  - https://developers.google.com/machine-learning/crash-course
- Machine Learning A-Z: Hands-On Python & R In Data Science
  - https://www.udemy.com/course/machinelearning/
- Deep Learning Specialization
  - https://www.coursera.org/specializations/deep-learning
- TensorFlow: Data and Deployment Specialization
  - https://www.coursera.org/specializations/tensorflow-data-and-deployment
- Reinforcement Learning Course by David Silver
  - http://www0.cs.ucl.ac.uk/staff/d.silver/web/Teaching

Over 200 of the Best Machine Learning, NLP, and Python Tutorials (by Robbie Allen)

Link: http://bit.ly/2TrodT0
Topics:

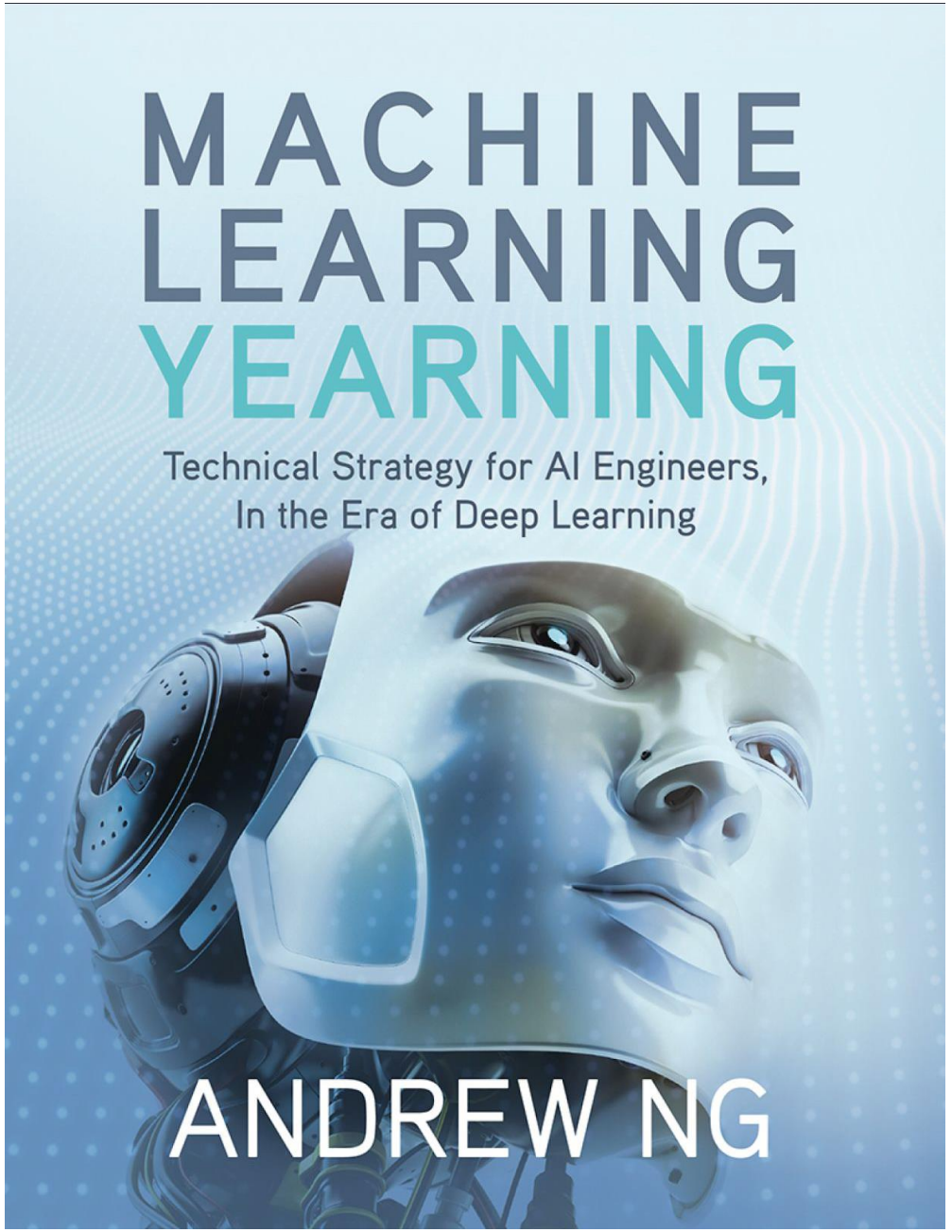
- Machine Learning
- Activation and Loss Functions
- Bias
- Perceptron
- Regression
- Gradient Descent
- Generative Learning
- Support Vector Machines
- Backpropagation
- Deep Learning
- Optimization and Dimensionality Reduction
- Long Short Term Memory (LSTM)
- Convolutional Neural Networks (CNNs)
- Recurrent Neural Nets (RNNs)
- Reinforcement Learning
- Generative Adversarial Networks (GANs)
- Multi-task Learning
- NLP
- Deep Learning and NLP
- Word Vectors
- Encoder-Decoder
- Python
- Scipy and numpy
- scikit-learn
- Tensorflow
- PyTorch
- Math
- Linear algebra
- Probability
- Calculus

Machine Learning Yearning

Andrew Ng

https://www.deeplearning.ai/machine-learning-yearning/

The book is focused not on teaching you ML algorithms, but on how to make ML algorithms work.



Deep Learning

Ian Goodfellow, Yoshua Bengio,
and Aaron Courville

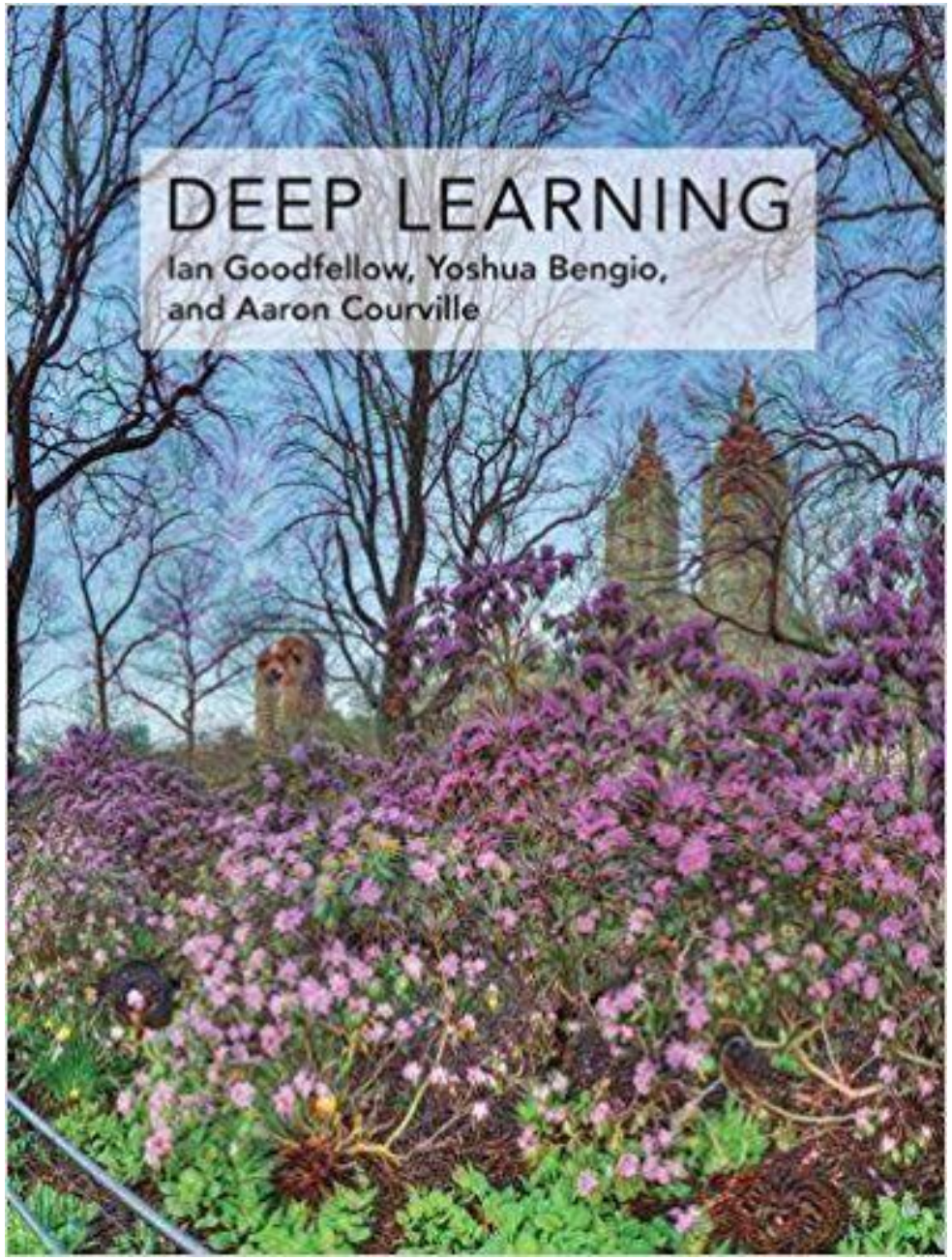
- http://www.deeplearningbook.org

“Written by three experts in the field, Deep Learning is the only comprehensive book on the subject.“

-- Elon Musk, cochair of OpenAI; cofounder and CEO of Tesla and SpaceX

Some useful links for this learning:

- Exercises
- Lecture Slides
- External links



Deep Learning with Python

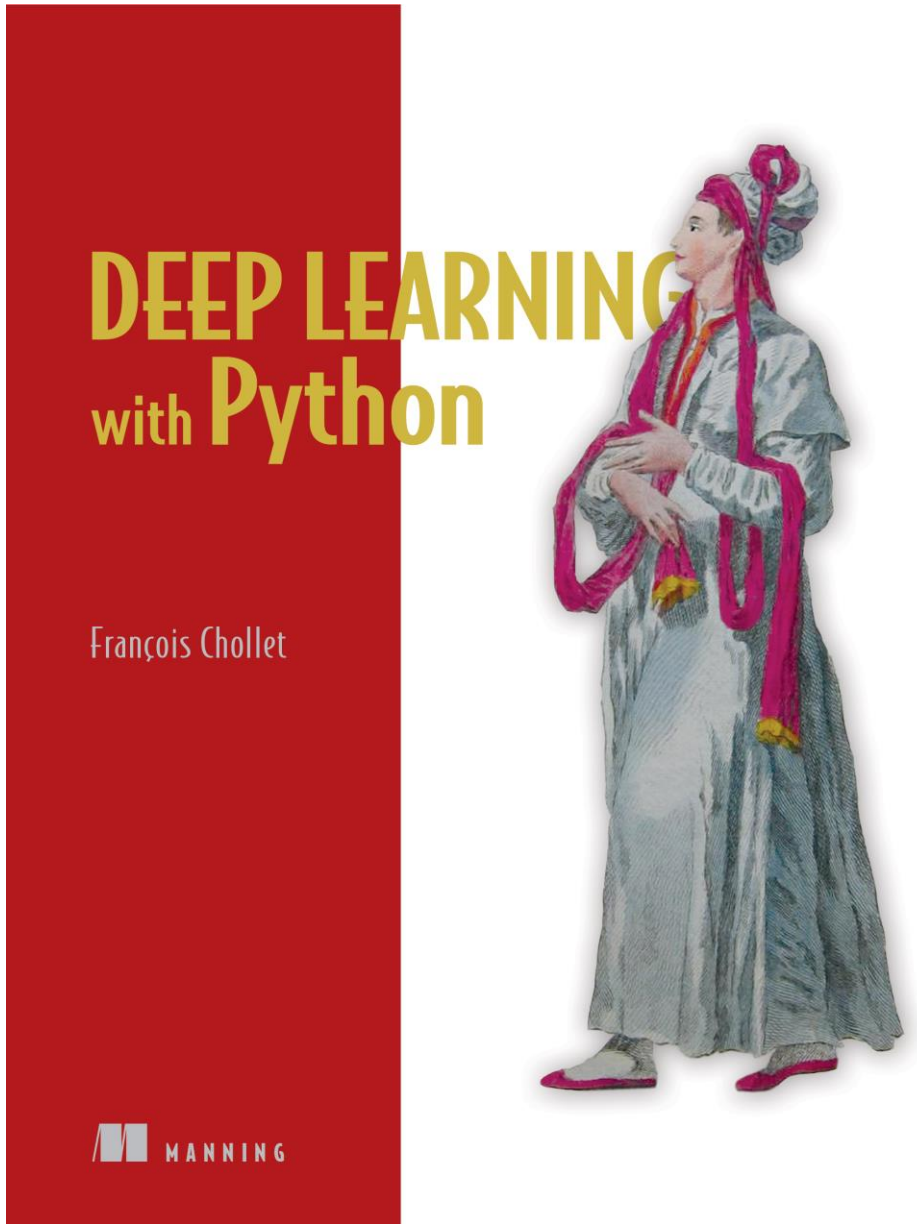
François Chollet

“The clearest explanation of deep learning I have come across... it was a joy to read.“

-- Richard Tobias, Cephasonics

Companion Jupyter notebooks for the book:

- https://github.com/fchollet/deep-learning-with-python-notebooks



Online resources

- https://www.kaggle.com/
- https://github.com
- https://mybinder.org
- https://jupyter.org/try
- https://arxiv.org/
- https://ai.google/tools/
- https://www.docker.com/
- https://cloud.google.com/automl

Online resources

- https://medium.com/
- https://towardsdatascience.com/
- https://www.quora.com/
- https://datascience.stackexchange.com/
- https://blog.dominodatalab.com

ML Packages

- NumPy
  - https://numpy.org/
- Scikit-learn
  - https://scikit-learn.org/
- XGBoost
  - https://xgboost.readthedocs.io/
- TensorFlow
  - https://www.tensorflow.org/
- Keras
  - https://keras.io/
- PyTorch
  - https://pytorch.org/