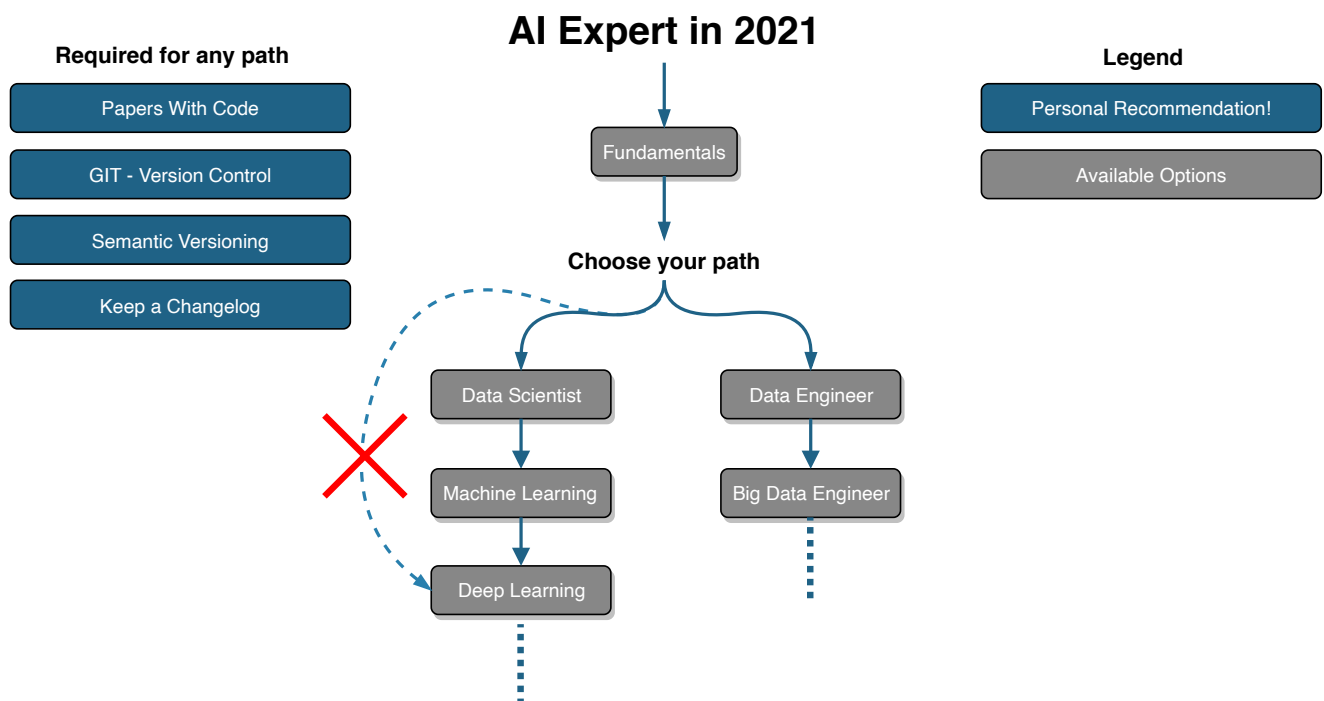


# i.am.ai

## AI Expert Roadmap

Roadmap to becoming an Artificial Intelligence Expert in 2021

## Introduction



## Fundamentals

# Fundamentals

## Basics

Matrices & Linear Algebra Fundamentals

Database Basics

Tabular Data

Data Frames & Series

Extract, Transform, Load (ETL)

Reporting vs BI vs Analytics

Data Formats

Regular Expressions (Regex)

Relational vs. non-relational databases

SQL + Joins (Inner, Outer, Cross, Theta Joins)

NoSQL

Some boxes link to additional resources  
Interactive version on  
[i.am.ai/roadmap](https://i.am.ai/roadmap)

JSON

XML

CSV

## Python Programming

Python Basics

Important libraries

Virtual Environments

Jupyter Notebooks / Lab

Numpy

Pandas

Expressions

Variables

Data Structures

Functions

Install packages (via pip, conda or similar)

Codestyle, e.g. PEP8

## Data Sources

Data Mining

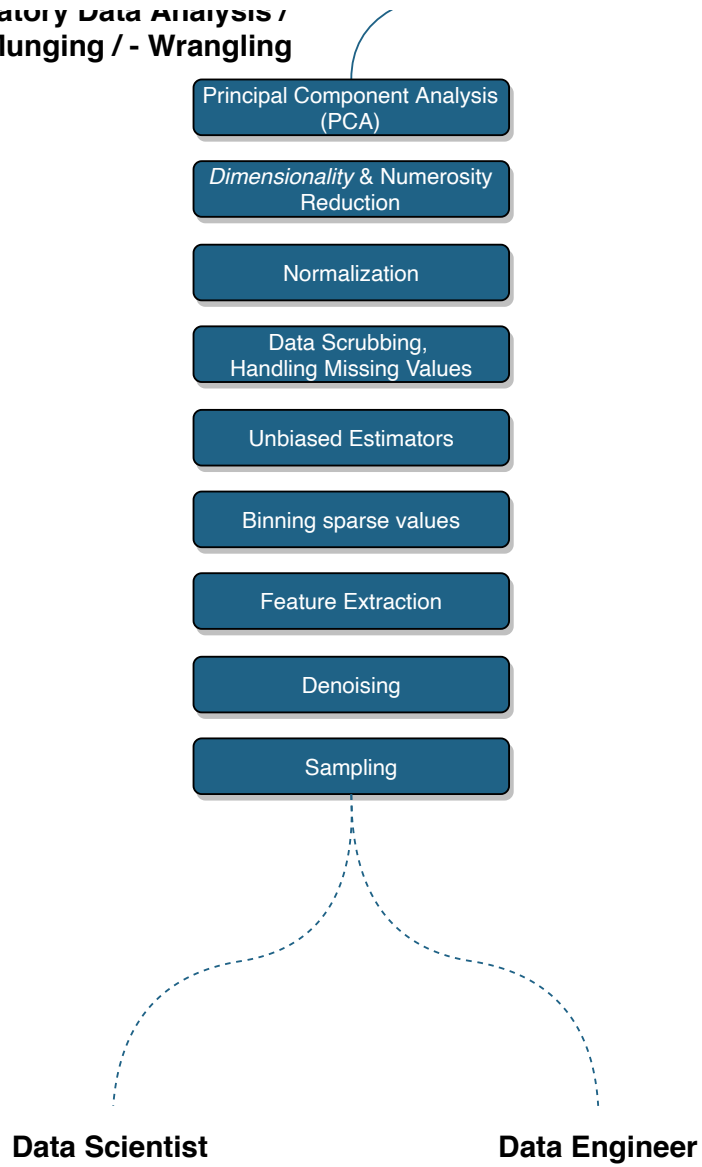
Web Scraping

Awesome Public Datasets

Kaggle

Exploratory Data Analysis /

## Exploratory Data Analysis / Data Munging / - Wrangling

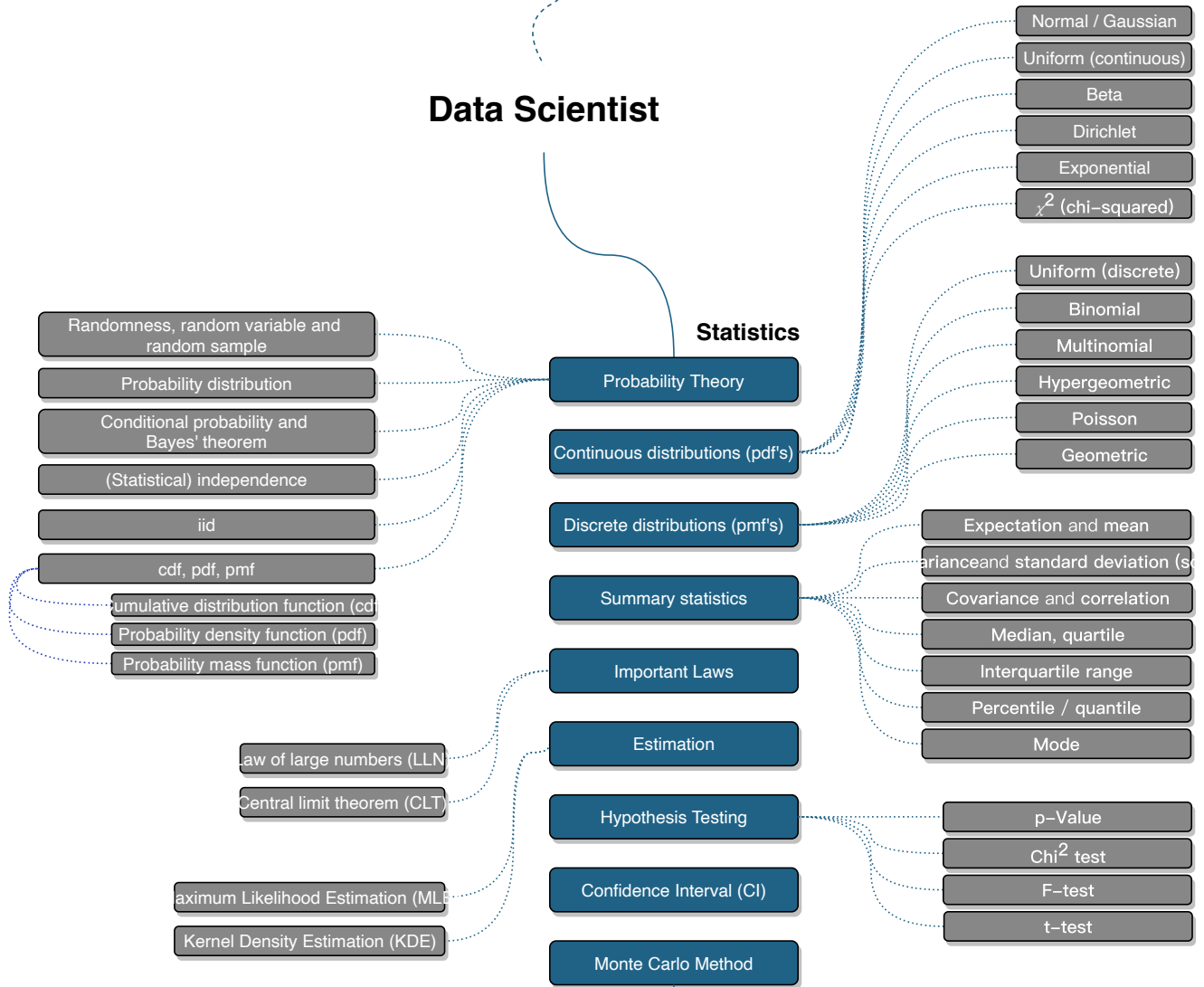


## Data Science Roadmap

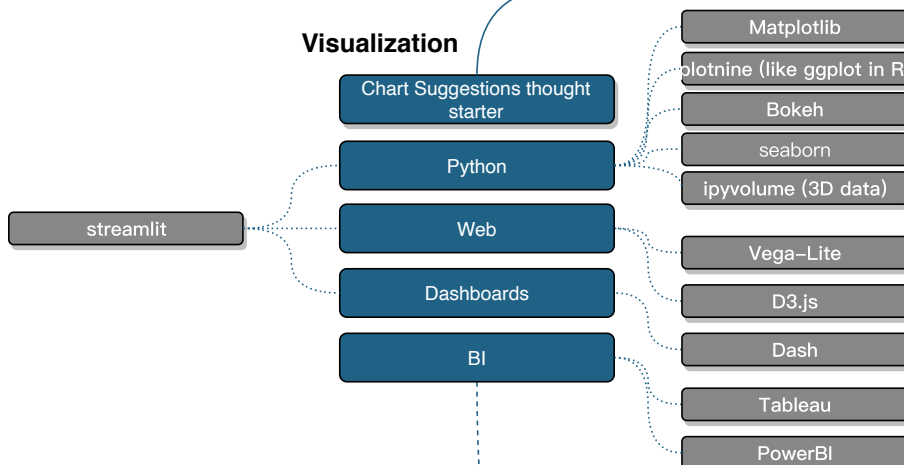
---

# Data Scientist

## Statistics



## Visualization





# Machine Learning

## Machine Learning Roadmap

---

# Machine Learning

## General

Concepts, Inputs & Attributes

Categorical Variables

Ordinal Variables

Numerical Variables

Cost functions and  
gradient descent

Overfitting / Underfitting

Training, validation  
and test data

Precision vs Recall

Bias & Variance

Lift

## Methods

Supervised Learning

Unsupervised Learning

Ensemble Learning

Reinforcement Learning

Regression

Linear Regression

Poisson Regression

Classification

Classification Rate

Decision Trees

Logistic Regression

Naïve Bayes Classifiers

K-Nearest Neighbour

SVM

Gaussian Mixture Models

Q-Learning

Clustering

Hierarchical Clustering

K-Means Clustering

DBSCAN

HDBSCAN

Fuzzy C-Means

Mean Shift

Agglomerative

OPTICS

Association Rule Learning

Apriori Algorithm

ECLAT algorithm

FP Trees

Dimensionality Reduction

Principal Component Analysis (PCA)

Random Projection

NMF

T-SNE

UMAP

Boosting

Bagging

Stacking

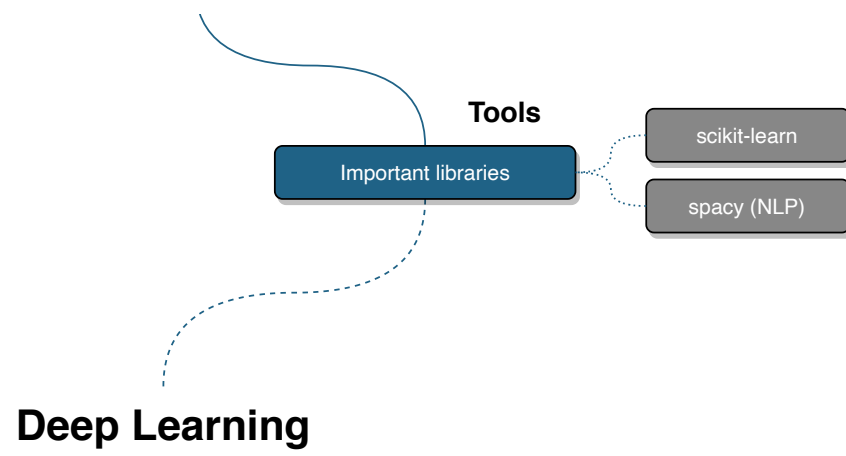
## Use Cases

Sentiment Analysis

Collaborative Filtering

Tagging

Prediction



## Deep Learning Roadmap

---

# Deep Learning

## Papers

Deep Learning Papers  
Reading Roadmap

Papers with code

Papers with code - state of the  
art

## Neural Networks

Understanding  
Neural Networks

Loss Functions

Activation Functions

Weight Initialization

Vanishing / Exploding  
Gradient Problem

## Architectures

Feedforward neural network

Autoencoder

Convolutional Neural Network  
(CNN)

Recurrent Neural Network  
(RNN)

Transformer

Siamese Network

Generative Adversarial  
Network (GAN)

Pooling

LSTM

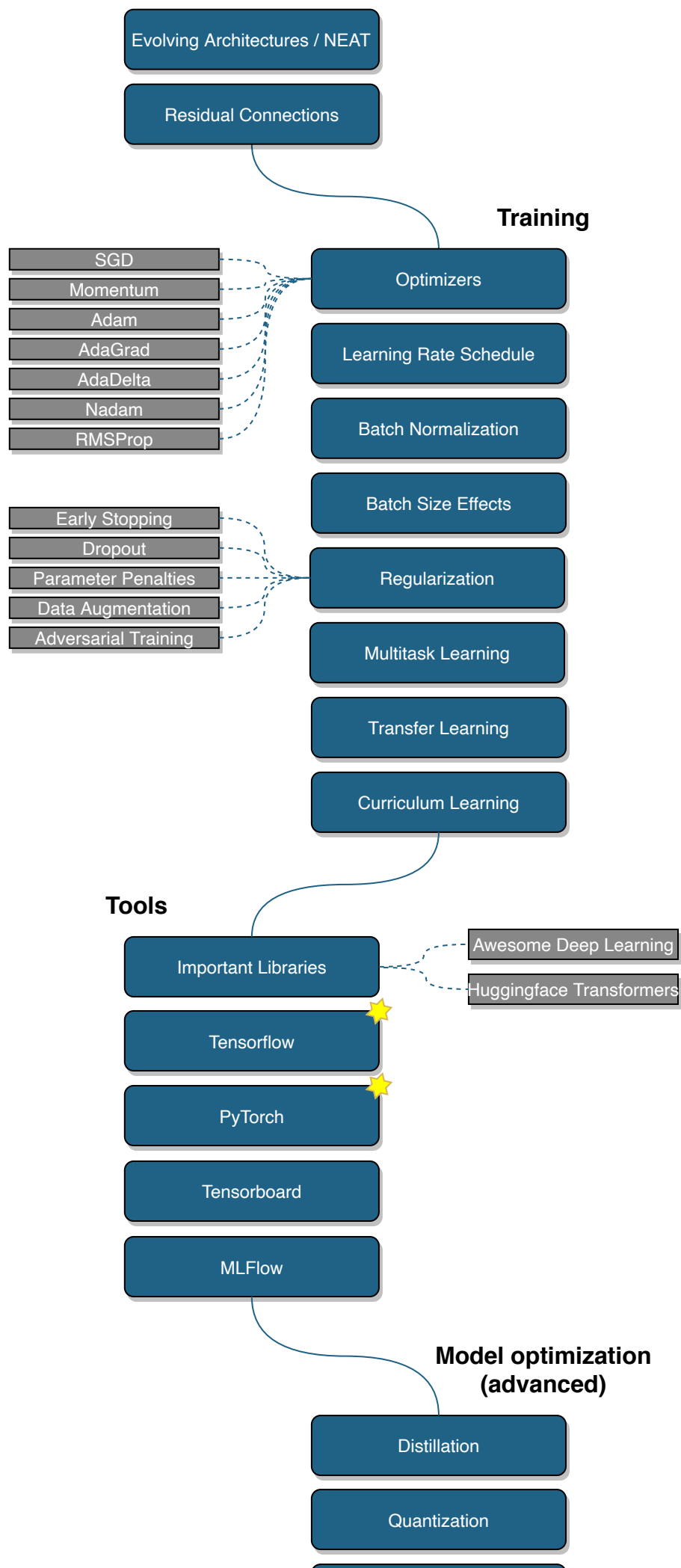
GRU

Encoder

Decoder

Attention





Neural Architecture  
Search (NAS)

**keep exploring and  
stay up-to-date**

## Data Engineer Roadmap

---



## Data Engineer

Summary of Data Formats

Data Discovery

Data Source & Acquisition

Data Integration

Data Fusion

Transformation & Enrichment

Data Survey

OpenRefine

How much Data

Using ETL

Data Lake vs Data Warehouse

Dockerize your Python  
Application

**keep exploring and  
stay up-to-date**

# Big Data Engineer

## Big Data Architectures

Architectural Patterns & Best Practices (video)

## Principles

Horizontal vs vertical scaling

Map Reduce

Data Replication

Name & Data Nodes

Job & Task Tracker

## Tools

Check the Awesome Big Data List

Hadoop (large data)

Spark (in memory)

RAPIDS (on GPU)

Flume, Scribe: For Unstruct Data

Data Warehouse with Hive

Elastic (EKL) Stack

Avro

Flink

Dask

Numba

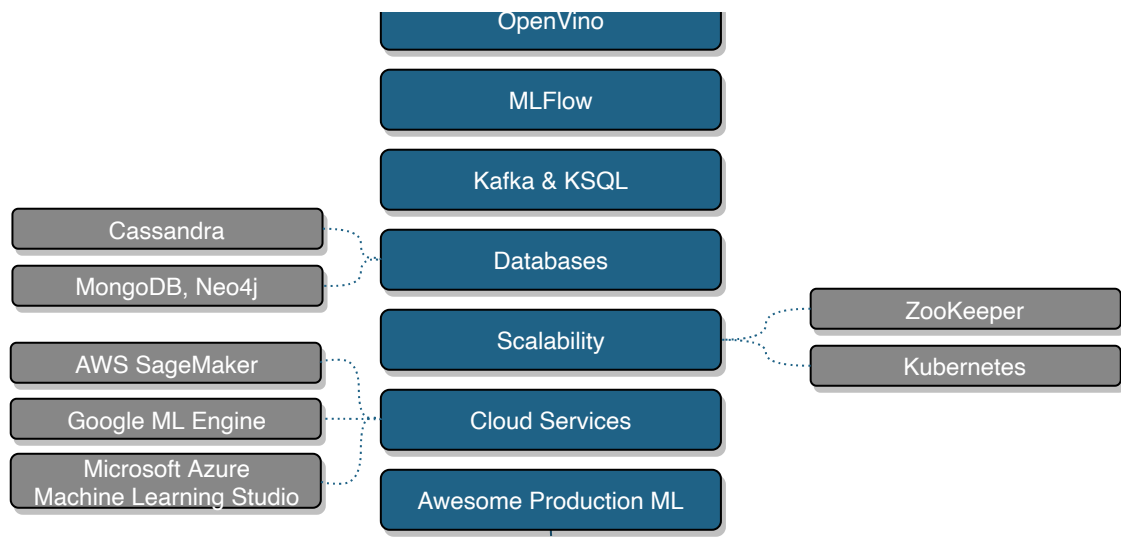
Onnx

HDFS

Loading data with Sqoop and Pig

Storm: Hadoop Realtime

to get data (e.g. logging), search, analyze and visualize it in realtime



**keep exploring and  
stay up-to-date**