i.am.ai
AI Expert Roadmap

Roadmap to becoming an Artificial Intelligence Expert in 2022

Below you find a set of charts demonstrating the paths that you can take and the technologies that you would want to adopt in order to become a data scientist, machine learning or an AI expert. We made these charts for our new employees to make them AI Experts but we wanted to share them here to help the community.

If you are interested to become an AI EXPERT at AMAI in Germany, or you want to hire an AI Expert, please say hi@am.ai.

Note

👉 An **interactive version with links to follow** about each bullet of the list can be found at i.am.ai/roadmap 👈

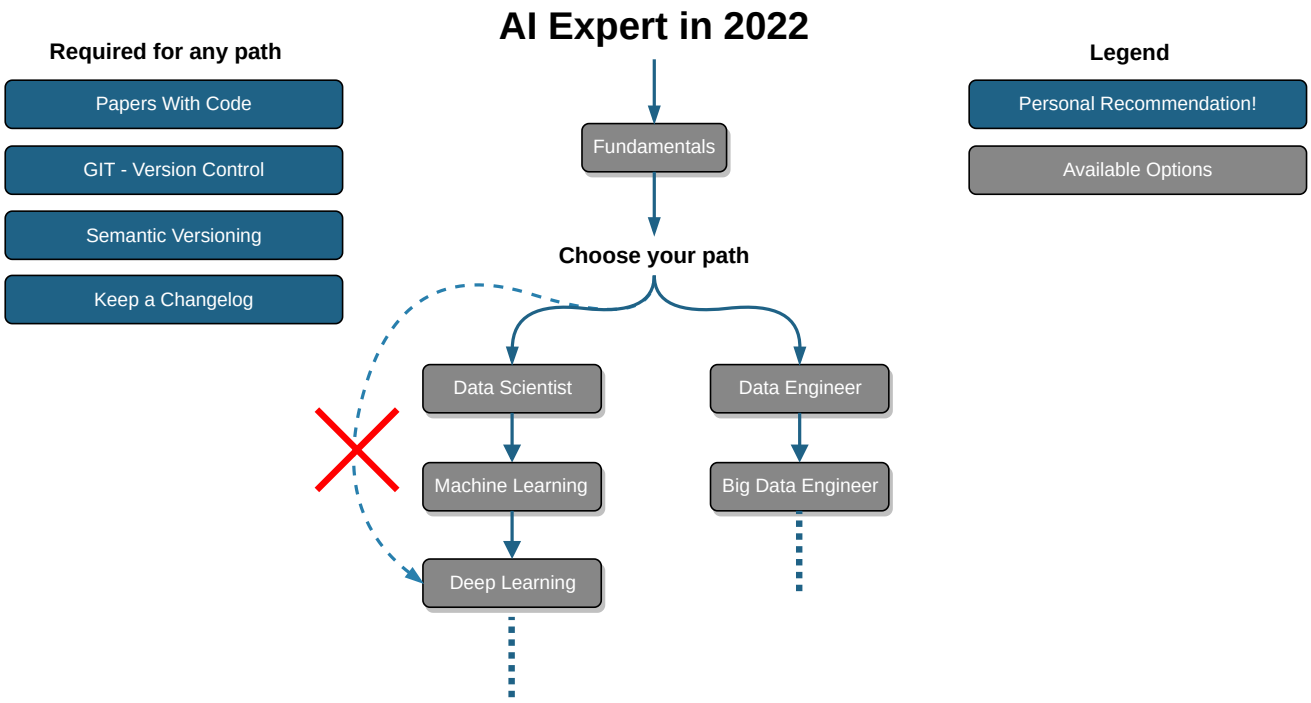
To receive updates star ⭐ and watch 👀 the GitHub Repo to get notified, when we add new content to stay on the top of the most recent research.

Follow our AI Newsletter to stay up to date with the latest developments in AI. We cover new use cases and research topics.

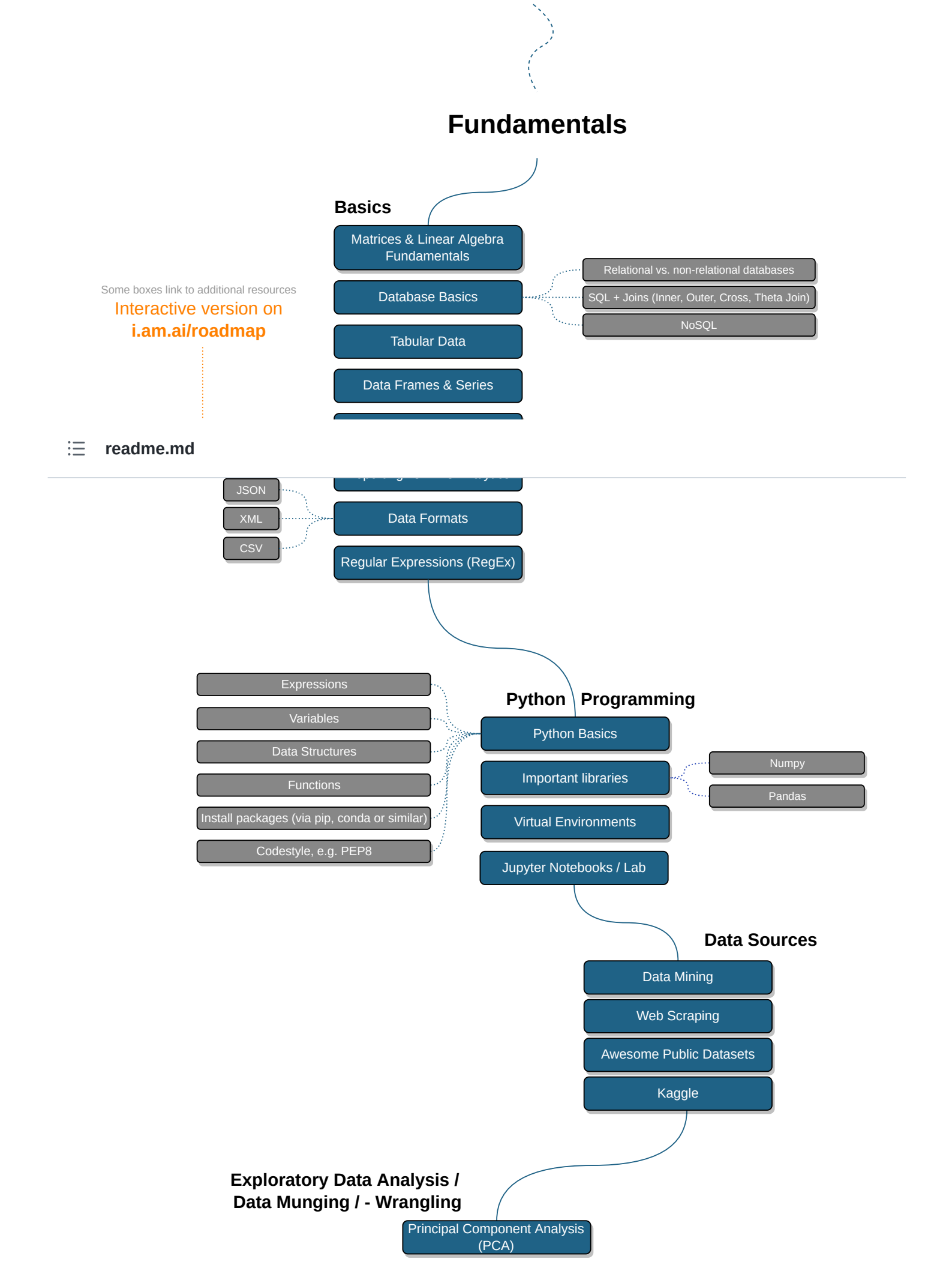
Disclaimer

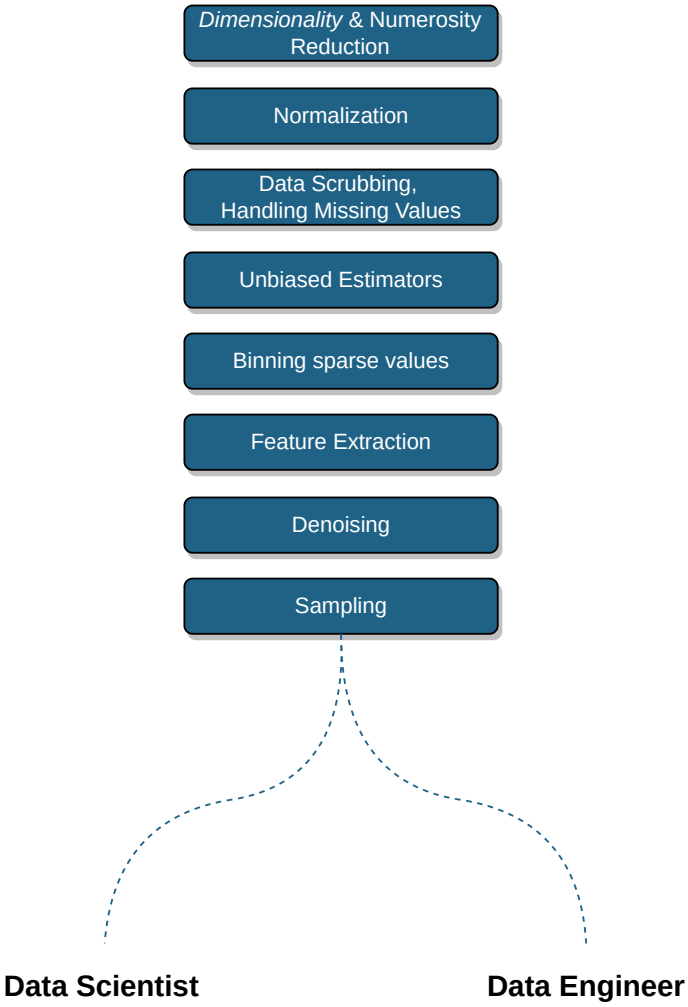
The purpose of these roadmaps is to give you an idea about the landscape and to guide you if you are confused about what to learn next and not to encourage you to pick what is hip and trendy. You should grow some understanding of why one tool would be better suited for some cases than the other and remember hip and trendy never means best suited for the job.

Introduction

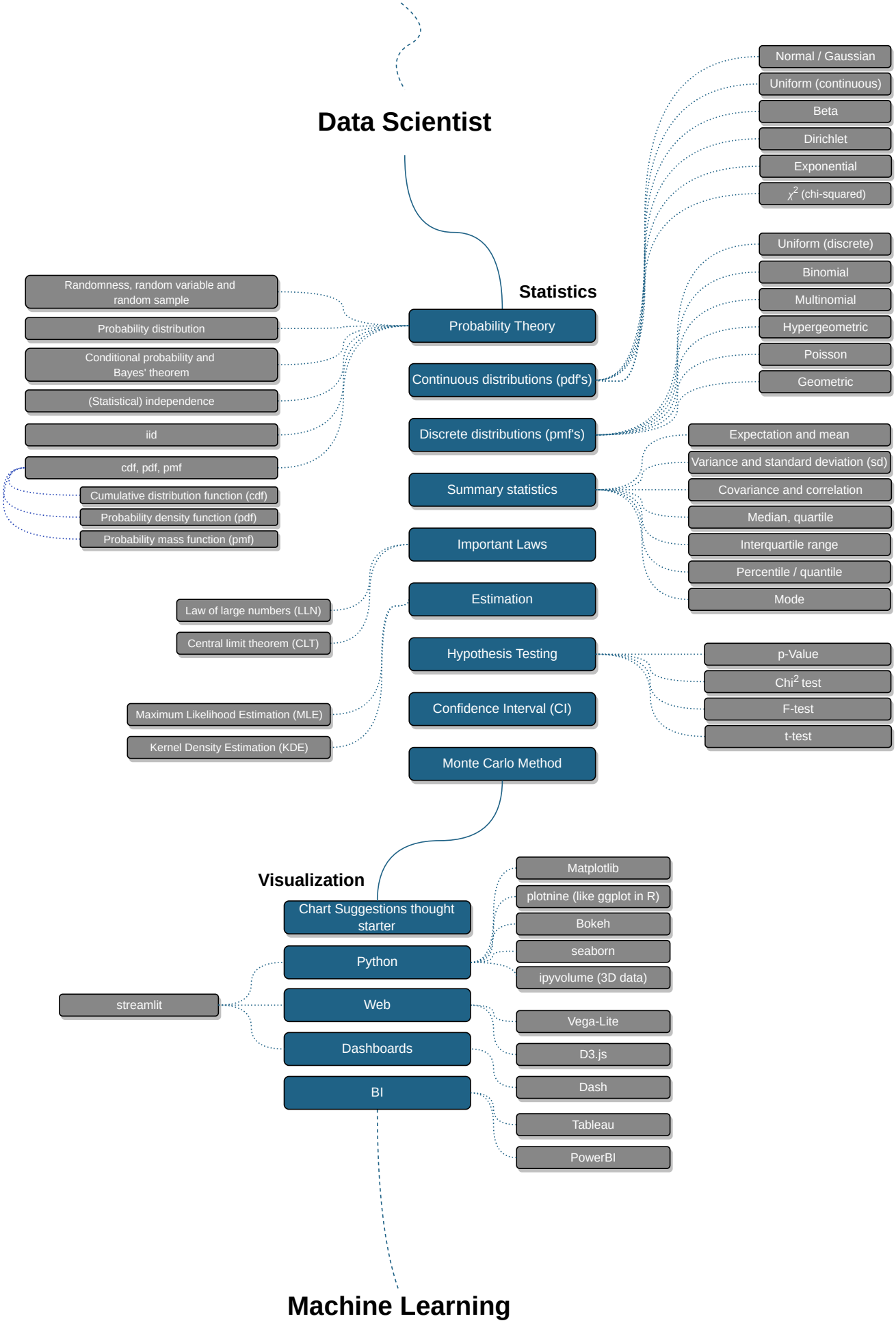


Fundamentals

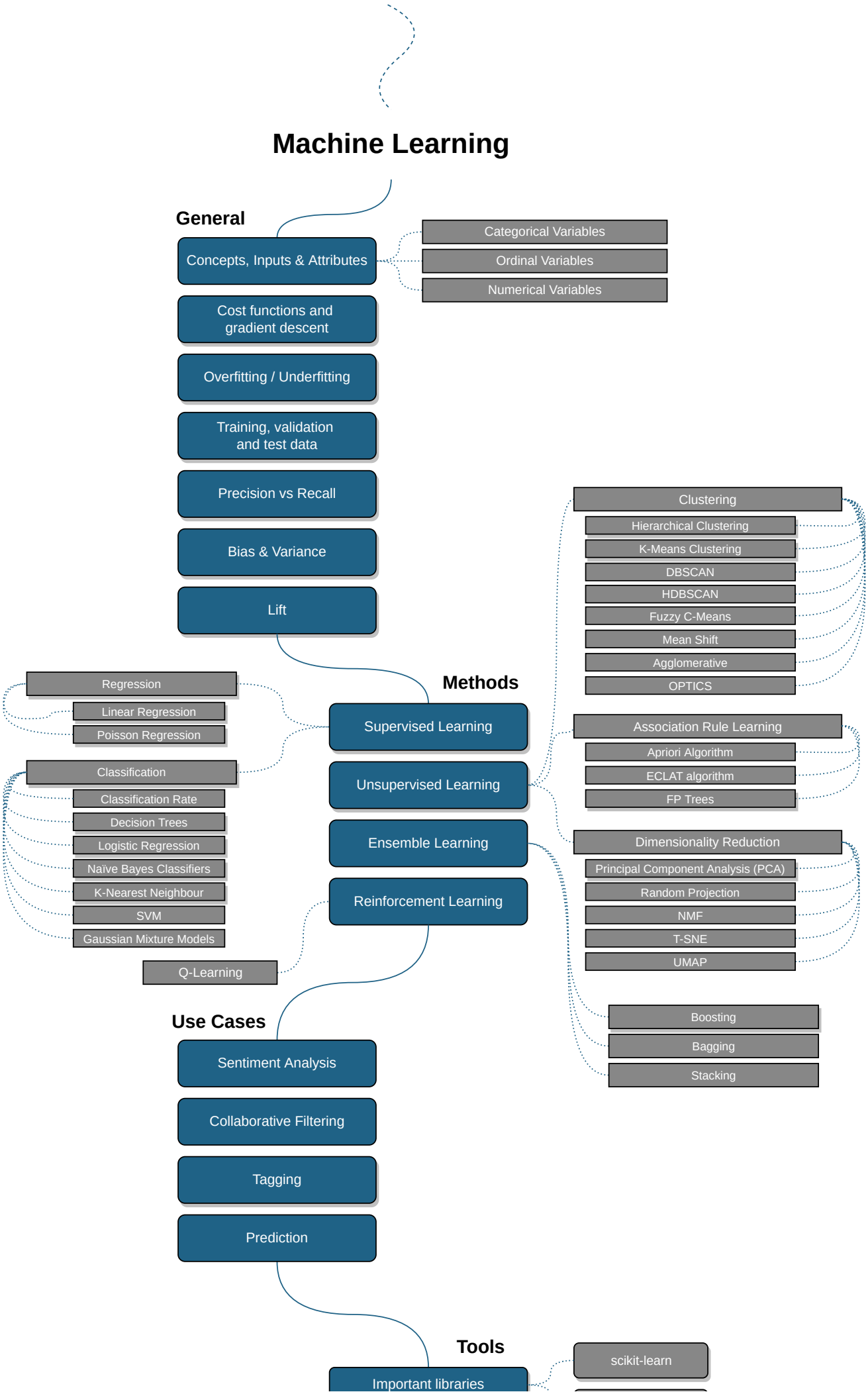

Fundamentals
Basics
Matrices & Linear Algebra Fundamentals
Database Basics
Relational vs. non-relational databases
SQL + Joins (Inner, Outer, Cross, Theta Join)
NoSQL
Tabular Data
Data Frames & Series
Some boxes link to additional resources
Interactive version on
i.am.ai/roadmap
readme.md
JSON
XML
CSV
Data Formats
Regular Expressions (RegEx)
Expressions
Variables
Data Structures
Functions
Install packages (via pip, conda or similar)
Codestyle, e.g. PEP8
Python Programming
Python Basics
Important libraries
Numpy
Pandas
Virtual Environments
Jupyter Notebooks / Lab
Data Sources
Data Mining
Web Scraping
Awesome Public Datasets
Kaggle
Exploratory Data Analysis /
Data Munging / - Wrangling
Principal Component Analysis (PCA)

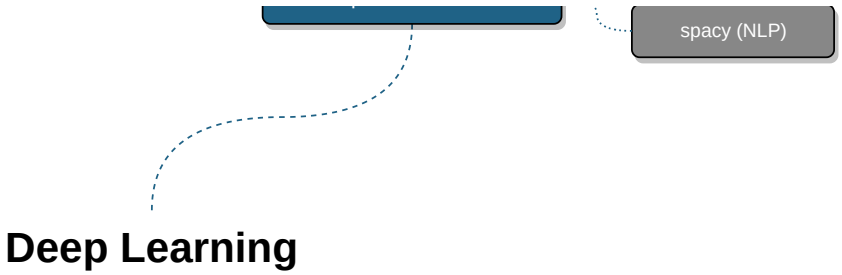


Data Science Roadmap

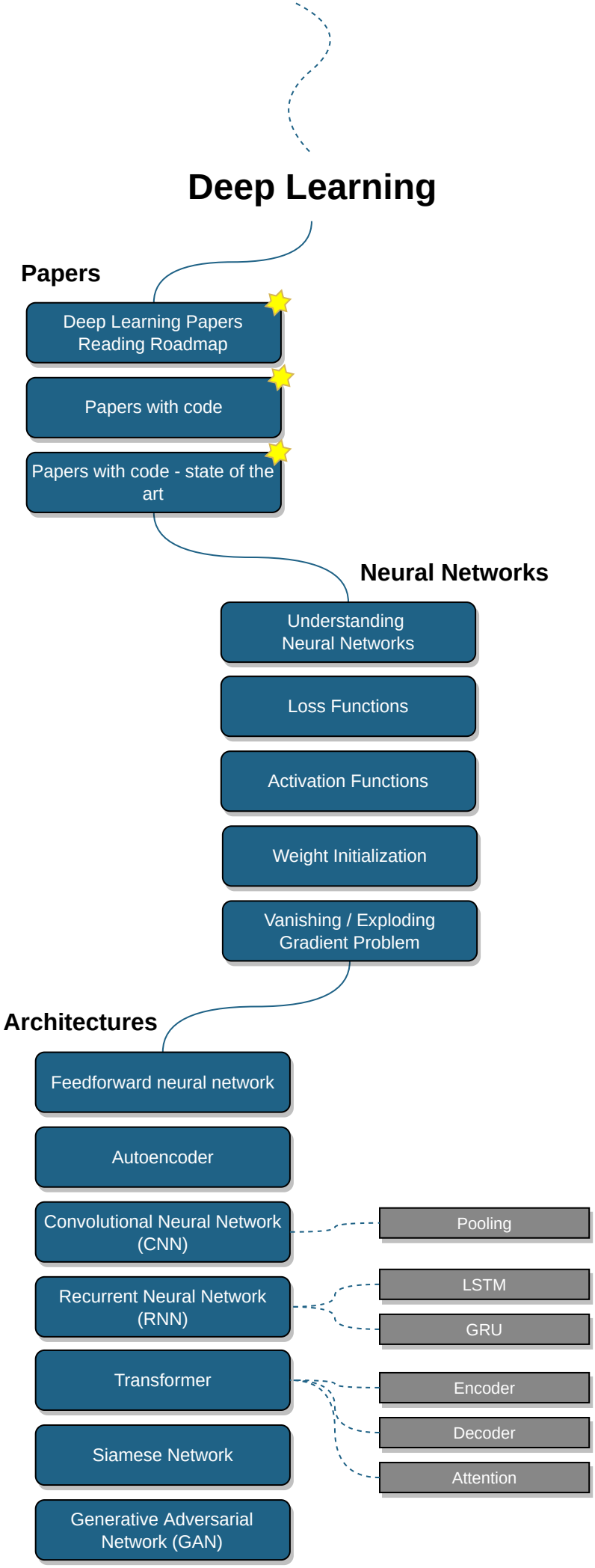


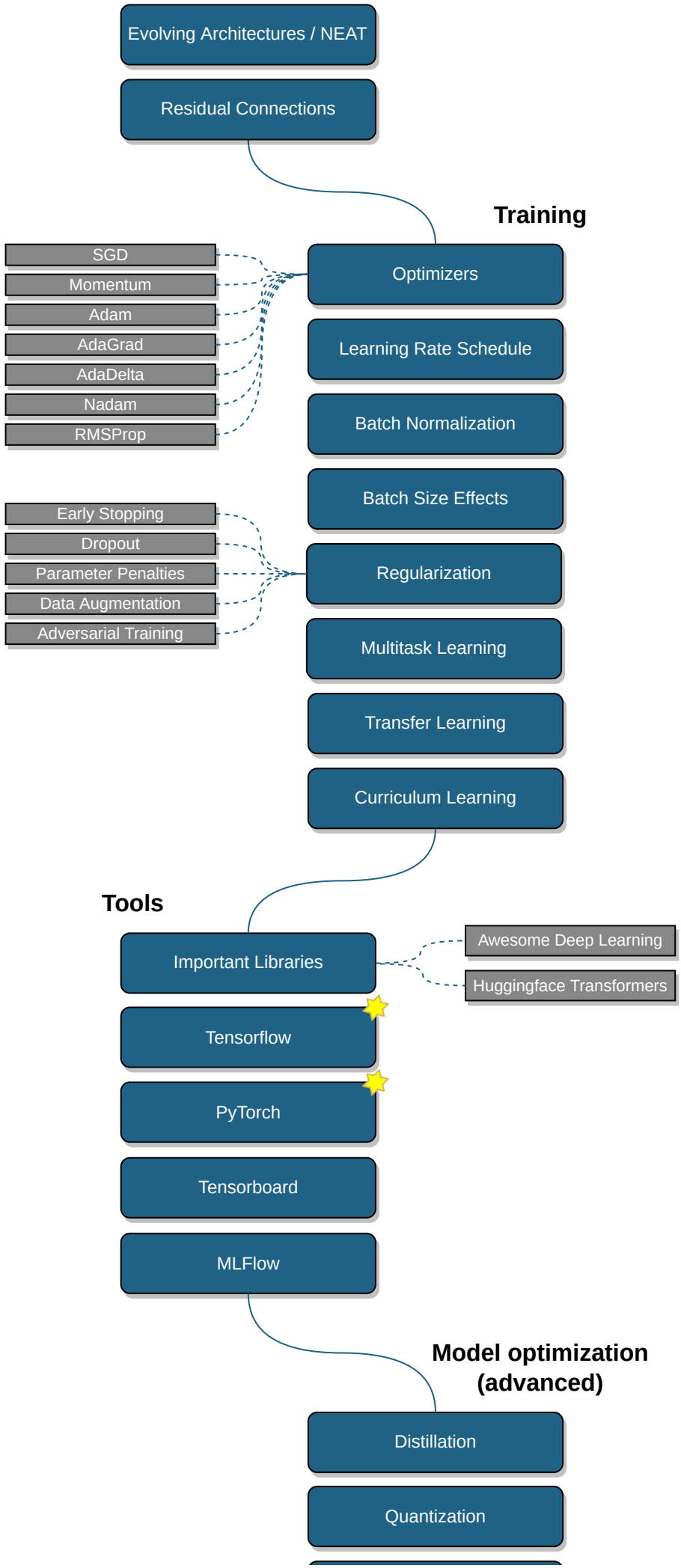
Machine Learning Roadmap

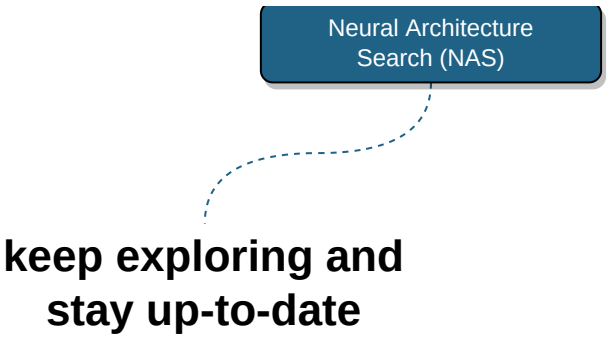

Machine Learning
General
Concepts, Inputs & Attributes
Categorical Variables
Ordinal Variables
Numerical Variables
Cost functions and gradient descent
Overfitting / Underfitting
Training, validation and test data
Precision vs Recall
Bias & Variance
Lift
Methods
Supervised Learning
Regression
Linear Regression
Poisson Regression
Classification
Classification Rate
Decision Trees
Logistic Regression
Naïve Bayes Classifiers
K-Nearest Neighbour
SVM
Gaussian Mixture Models
Unsupervised 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
Ensemble Learning
Boosting
Bagging
Stacking
Reinforcement Learning
Q-Learning
Use Cases
Sentiment Analysis
Collaborative Filtering
Tagging
Prediction
Tools
Important libraries
scikit-learn



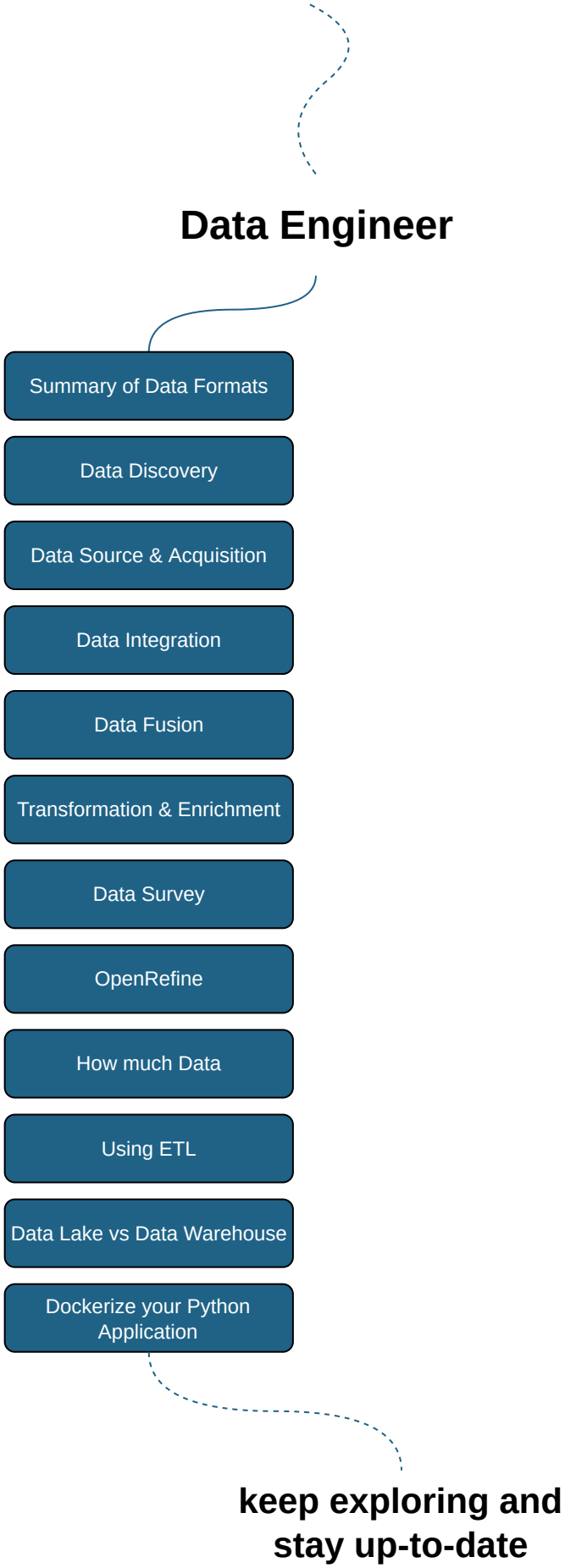
Deep Learning Roadmap


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)
Pooling
Recurrent Neural Network (RNN)
LSTM
GRU
Transformer
Encoder
Decoder
Attention
Siamese Network
Generative Adversarial Network (GAN)


Evolving Architectures / NEAT
Residual Connections
Training
SGD
Momentum
Adam
AdaGrad
AdaDelta
Nadam
RMSProp
Optimizers
Learning Rate Schedule
Batch Normalization
Batch Size Effects
Early Stopping
Dropout
Parameter Penalties
Data Augmentation
Adversarial Training
Regularization
Multitask Learning
Transfer Learning
Curriculum Learning
Tools
Important Libraries
Awesome Deep Learning
Huggingface Transformers
Tensorflow
PyTorch
Tensorboard
MLFlow
Model optimization (advanced)
Distillation
Quantization



Data Engineer Roadmap



Big Data Engineer Roadmap

Big Data Engineer

Big Data Architectures

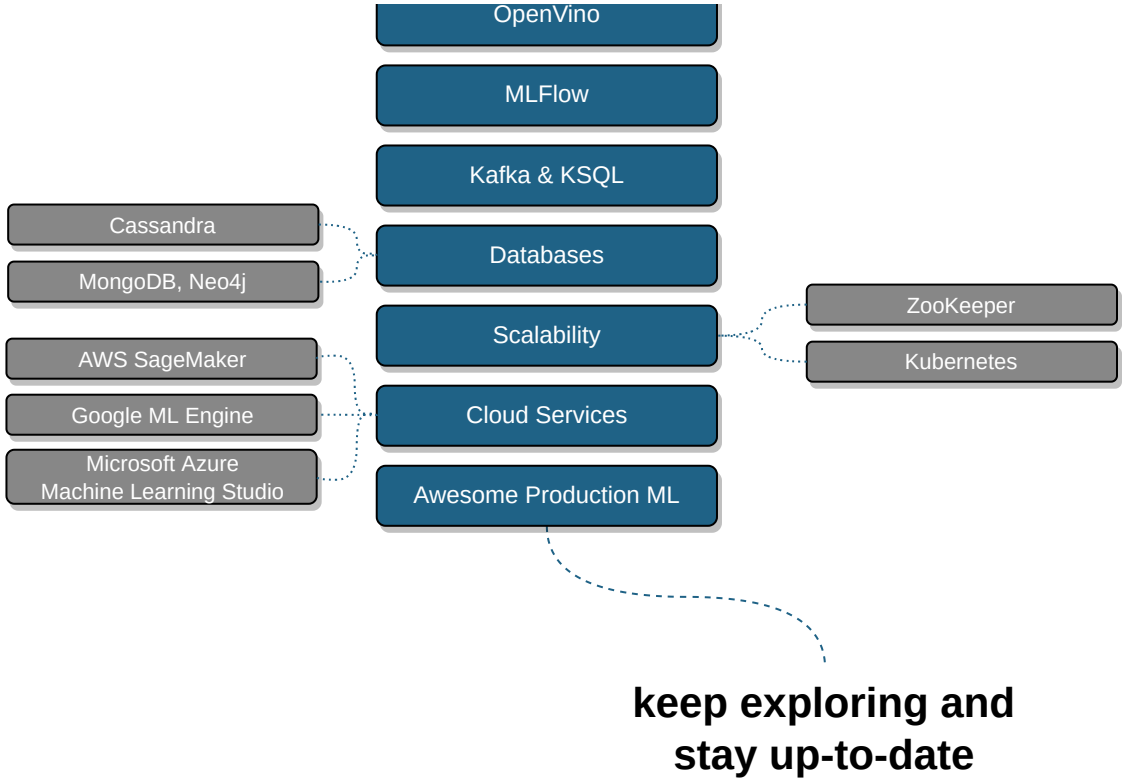
- Architectural Patterns & Best Practices (video)

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)
  - HDFS
  - Loading data with Sqoop and Pig
  - Storm: Hadoop Realtime
- Spark (in memory)
- RAPIDS (on GPU)
- Flume, Scribe: For Unstruct Data
- Data Warehouse with Hive
- Elastic (EKL) Stack
  - to get data (e.g. logging), search, analyze and visualize it in realtime
- Avro
- Flink
- Dask
- Numba
- Onnx



🚦 Wrap Up

If you think any of the roadmaps can be improved, please do open a PR with any updates and submit any issues. Also, we will continue to improve this, so you might want to watch/star this repository to revisit.

🙌 Contribution

> Have a look at the contribution docs for how to update any of the roadmaps

- Open pull request with improvements
- Discuss ideas in issues
- Spread the word
- Reach out with any feedback

Supported By

AMAI AI EXPERTS

de:hub KARLSRUHE Artificial Intelligence

Contributors 9

Languages

JavaScript 63.5% **Vue** 18.7% **Stylus** 17.8%