Essentials of Deep Learning for NLP

Concepts & Examples

Dipanjan (DJ) Sarkar

Session Agenda

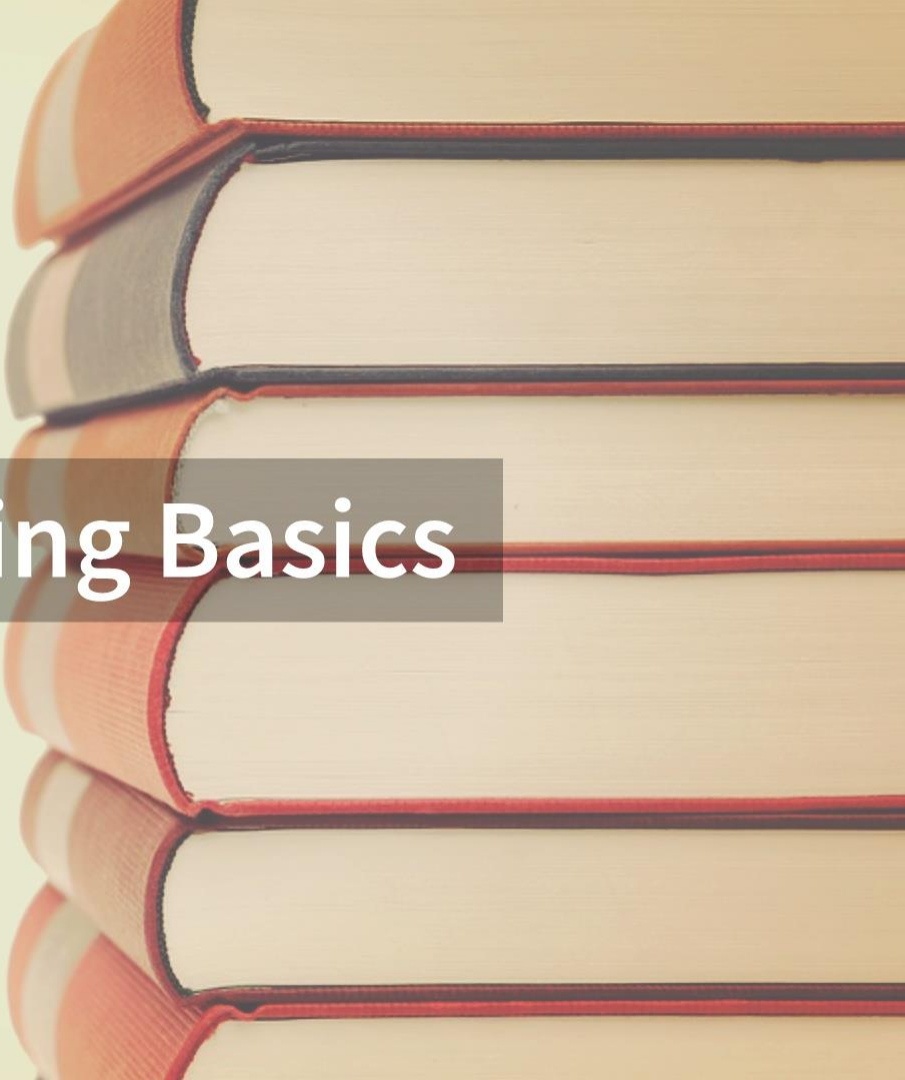
Deep Learning Basics

Deep Learning Effectiveness

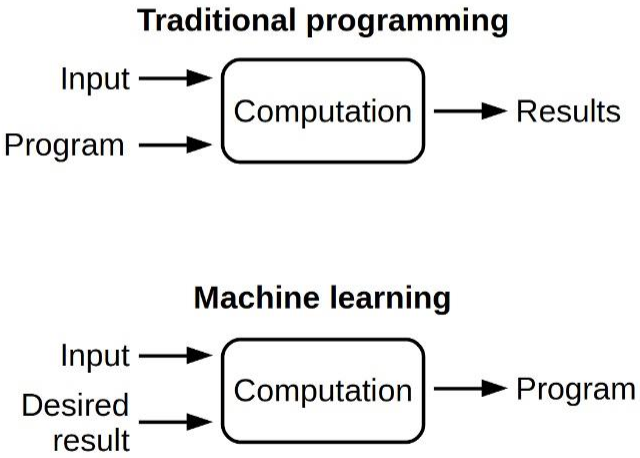
Deep Learning Frameworks

Deep Learning Model Architectures & Applications for NLP

Deep Learning Basics

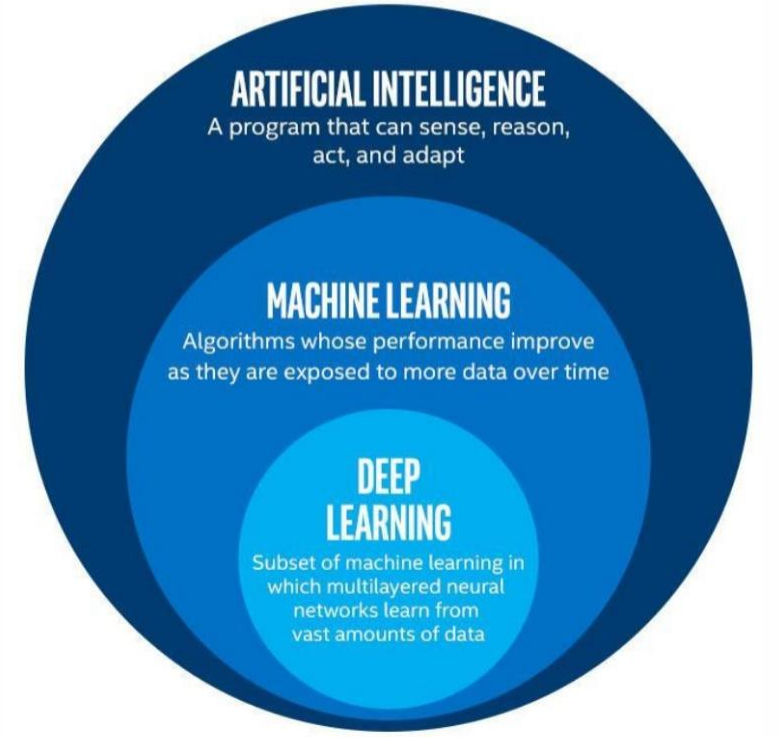


Why Learning?



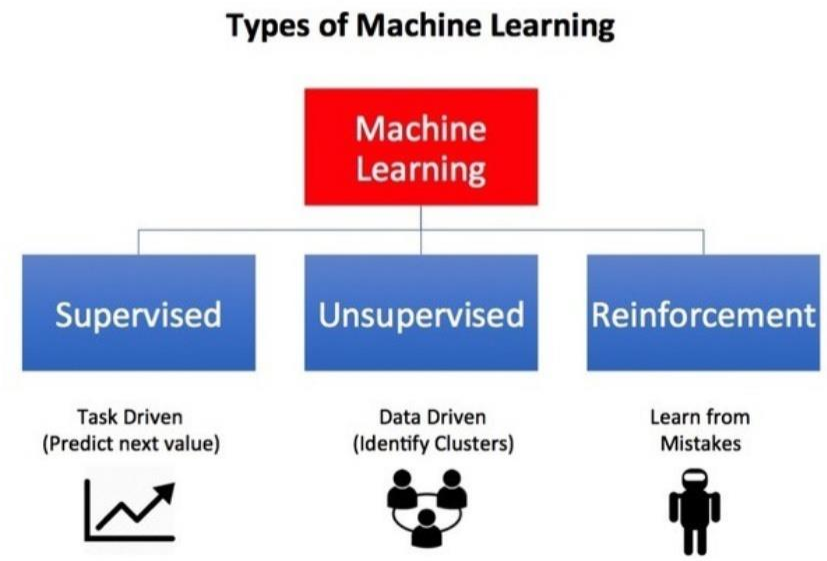
- We humans are wired to discover, inspect, comprehend and learn automatically thanks to our brain
- The automatic extraction of semantic information from raw data is the heart of most useful applications
  - Image Recognition
  - Object Detection
  - Natural Language Processing
  - Audio Processing
  - Robotics and so on...
- Making machines learn by extracting useful semantic information is hard!
- We need models to learn from data (a lot of it!)

AI vs. ML vs. DL



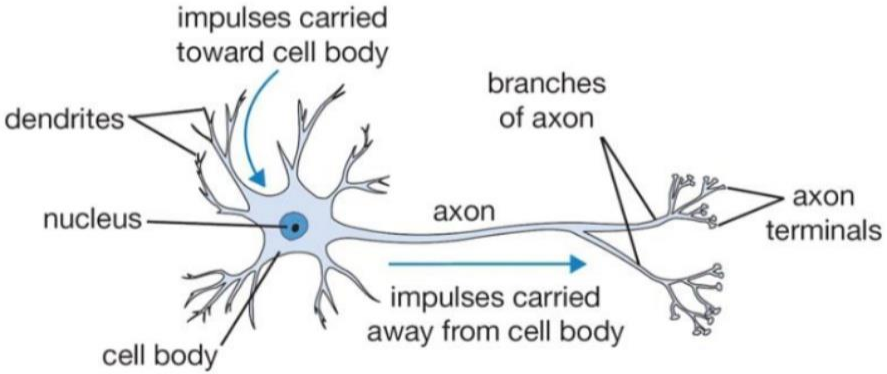
- AI can be any model or technique which can mimic human intelligence and perform complex tasks
- Machine Learning is a way towards building AI by leveraging models which learn from patterns in data
- Deep Learning is a specialized area of ML dealing with complex multi-layered neural network models which scale and perform better with big and complex datasets

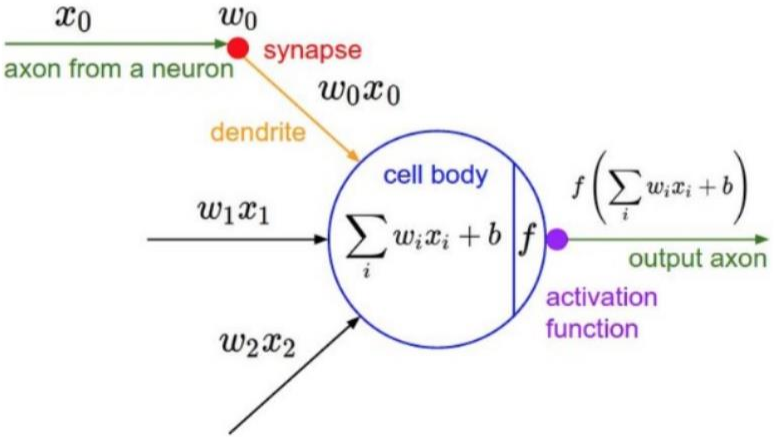
Learning Types



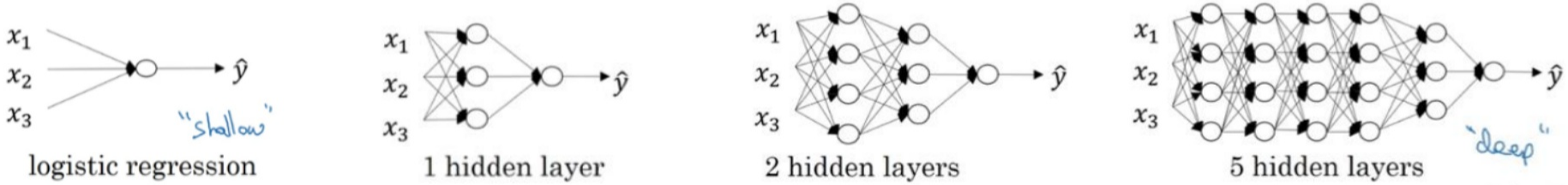
- Supervised Learning requires historical labeled data based on outcomes to predict
- Unsupervised Learning is about finding out frequent patterns or similar groups in data
- Reinforcement Learning is about trying to achieve a goal based on learning from rewards and mistakes

Neural Networks

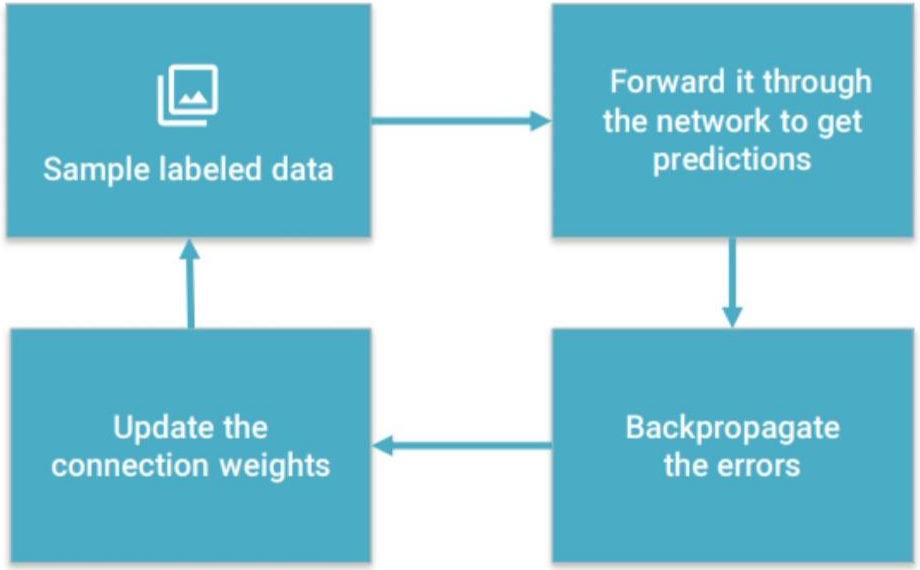




Deep Learning Models - Multi-layered Neural Networks

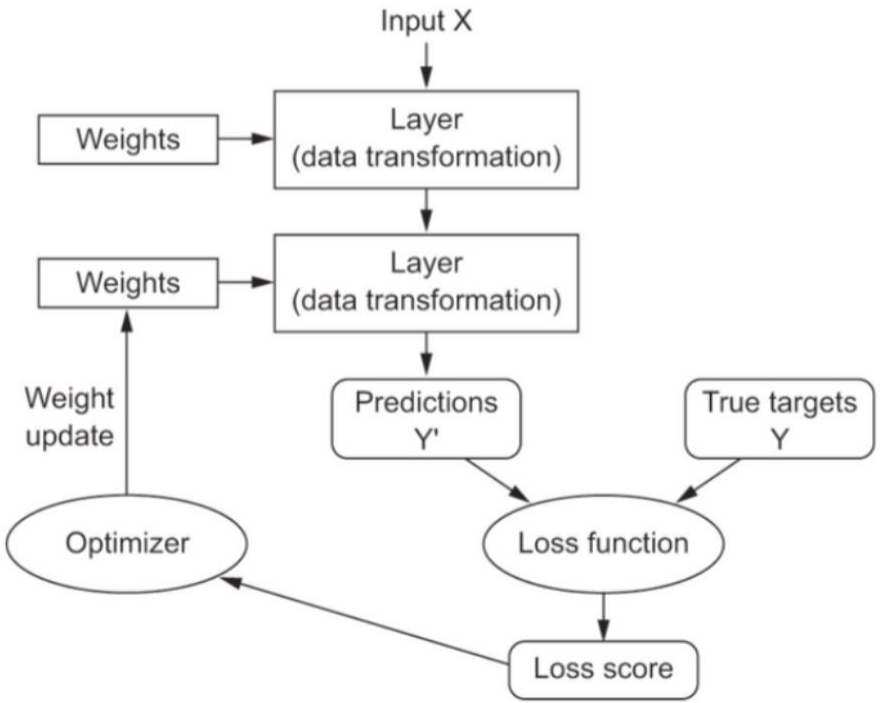


Neural Networks - Training



Learns by generating an error signal that measures the difference between the predictions of the network and the desired values and then using this error signal to change the weights (or parameters) so that predictions get more accurate.

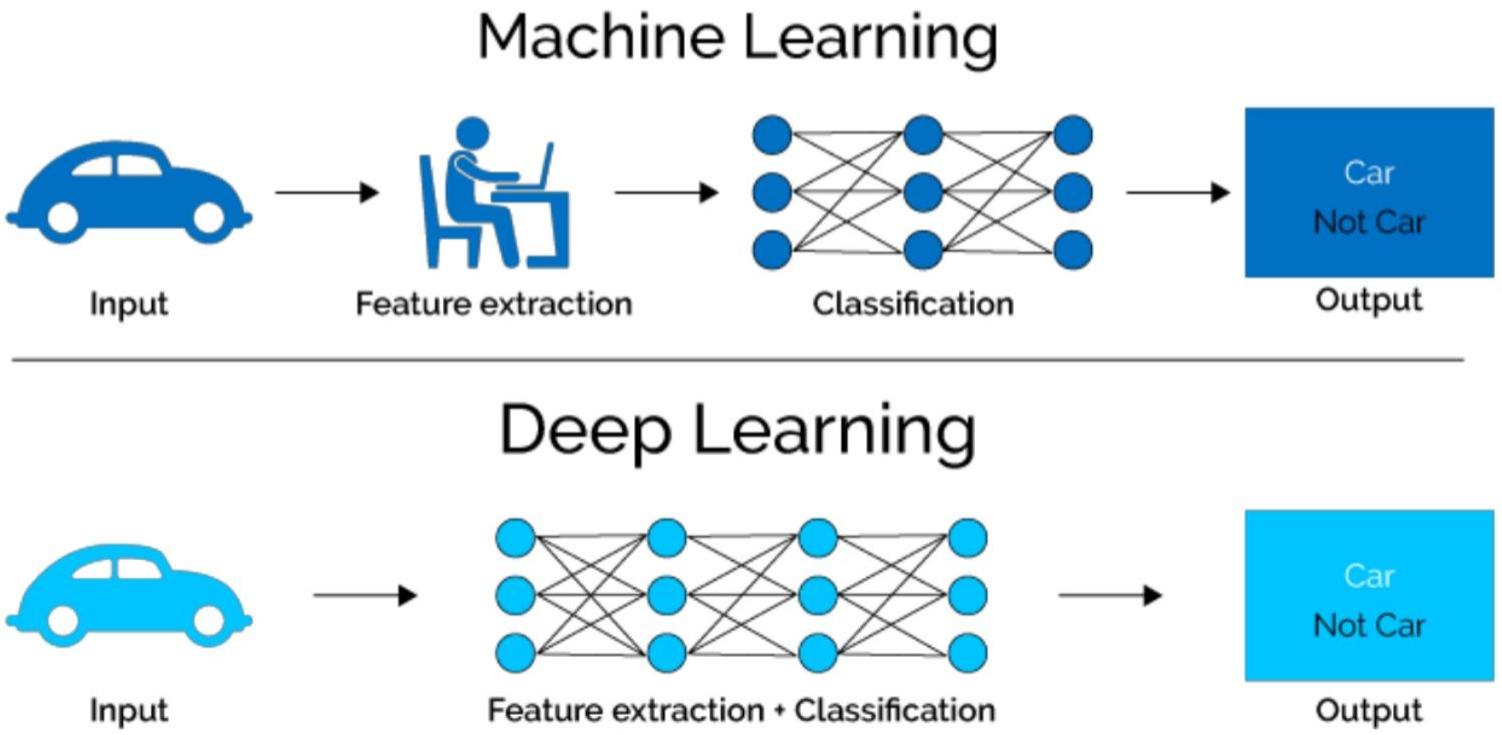
Training a Neural Network



Deep Learning with Python, Francois Chollet

- **Perform parameter updates to minimize the loss (training objective)**
- **Typical flow involves:**
  - Forward pass with the input going through various transformations
  - Compute the loss based on predictions and actuals
  - Compute gradients
  - Backpropagate gradients to update layer weights
- **TensorFlow / PyTorch enables easy Automatic Differentiation**

Deep Learning - ML on Steroids




Deep Learning Effectiveness

Deep Learning - No longer a Hype!

I have worked all my life in Machine Learning, and I've never seen one algorithm knock over benchmarks like Deep Learning
– Andrew Ng (Stanford & Baidu)

Deep Learning is an algorithm which has no theoretical limitations of what it can learn; the more data you give and the more computational time you provide, the better it is *– Geoffrey Hinton (Google)*

Human-level artificial intelligence has the potential to help humanity thrive more than any invention that has come before it *– Dileep George (Co-Founder Vicarious)*

For a very long time it will be a complementary tool that human scientists and human experts can use to help them with the things that humans are not naturally good *– Demis Hassabis (Co-Founder* DeepMind*)*

Adoption of Deep Learning is massive

CONSUMER

- Smart Assistants
- Chatbots
- Search
- Personalization
- Augmented Reality
- Robots

HEALTH

- Enhanced Diagnostics
- Drug Discovery
- Patient Care
- Research
- Sensory Aids

FINANCE

- Algorithmic Trading
- Fraud Detection
- Research
- Personal Finance
- Risk Mitigation

RETAIL

- Support
- Experience
- Marketing
- Merchandising
- Loyalty
- Supply Chain
- Security

GOVERNMENT

- Defense
- Data Insights
- Safety & Security
- Resident Engagement
- Smarter Cities

ENERGY

- Oil & Gas Exploration
- Smart Grid
- Operational Improvement
- Conservation

TRANSPORT

- Autonomous Cars
- Automated Trucking
- Aerospace
- Shipping
- Search & Rescue

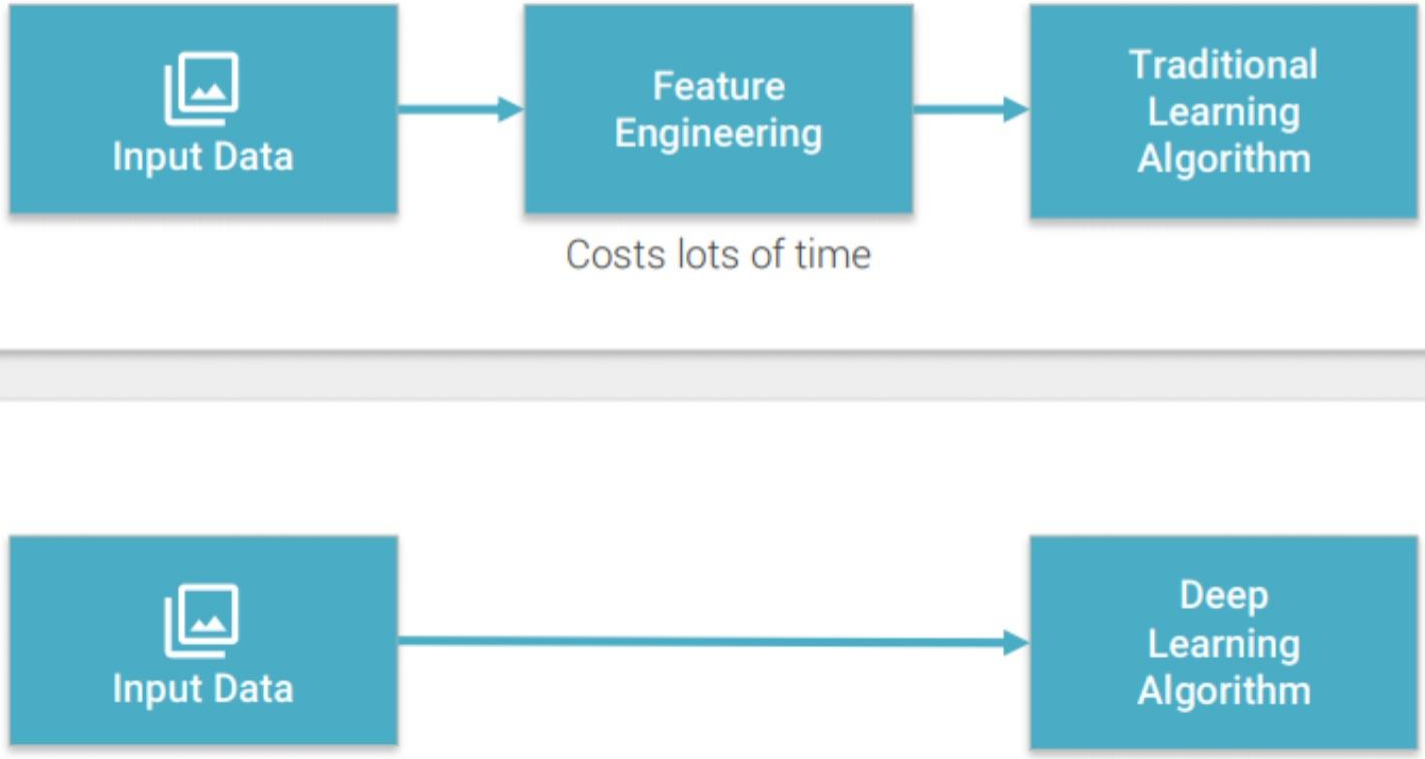
INDUSTRIAL

- Factory Automation
- Predictive Maintenance
- Precision Agriculture
- Field Automation

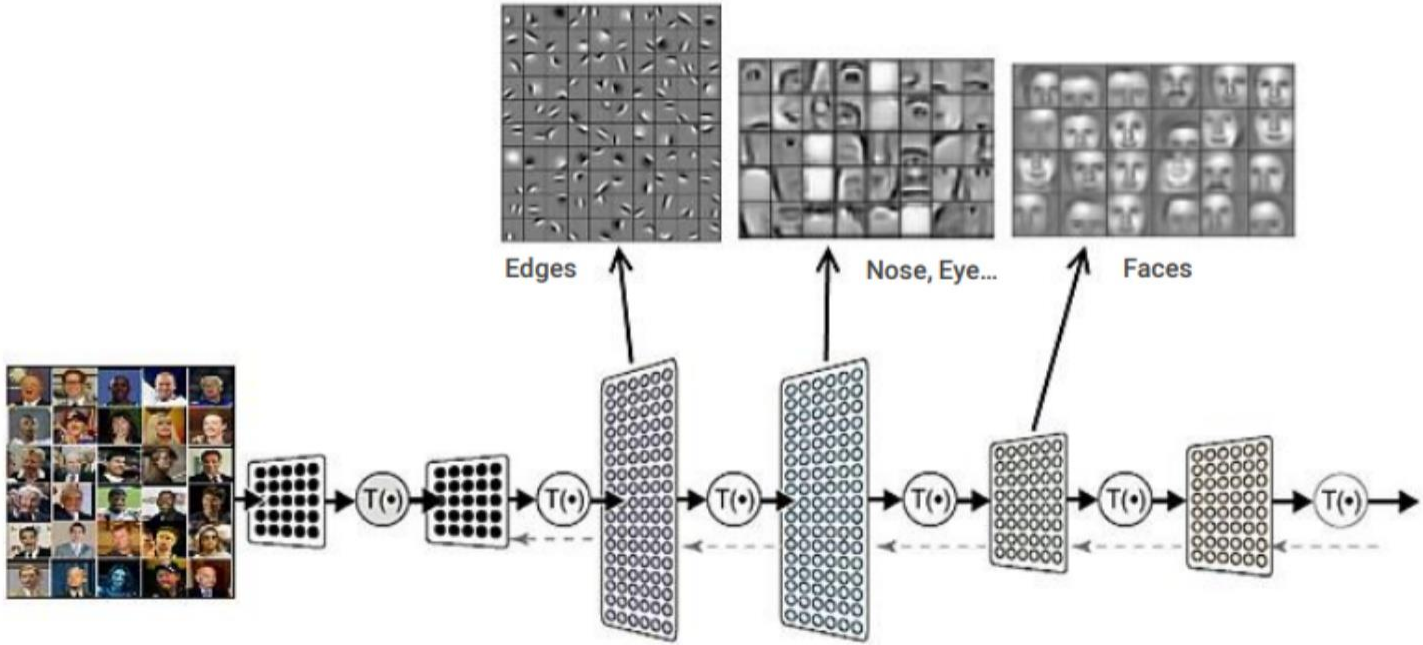
OTHER

- Advertising
- Education
- Gaming
- Professional & IT Services
- Telco/Media
- Space Exploration
- Sports

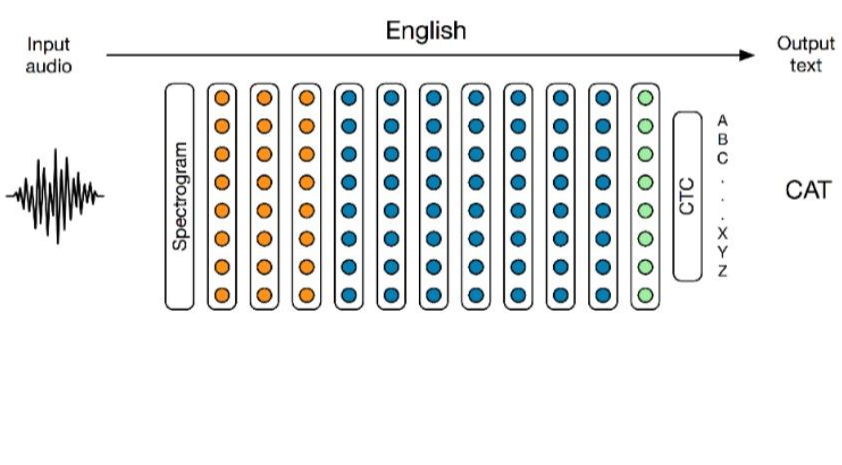
How is Deep Learning effective?



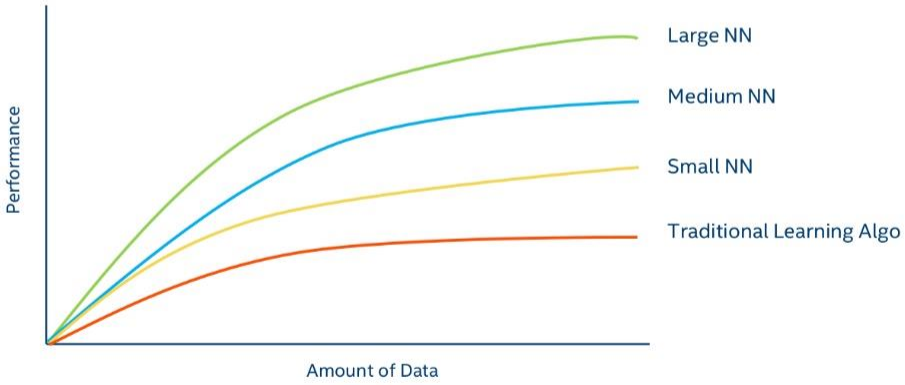
How is Deep Learning effective?



How is Deep Learning effective?



Scale drives Deep Learning Models



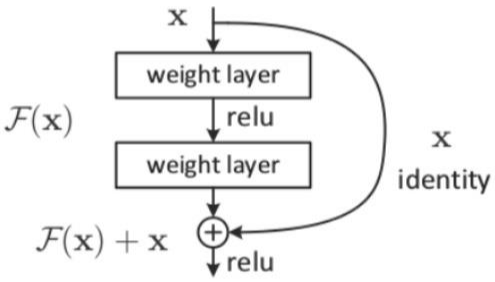
Bigger and more complex architectures based on various interchangeable modules/techniques

Deeper models that can learn from much fewer training cases

Harder problems such as video understanding and natural language processing will be successfully tackled by deep learning algorithms

Why is Deep Learning effective now?

New algorithms



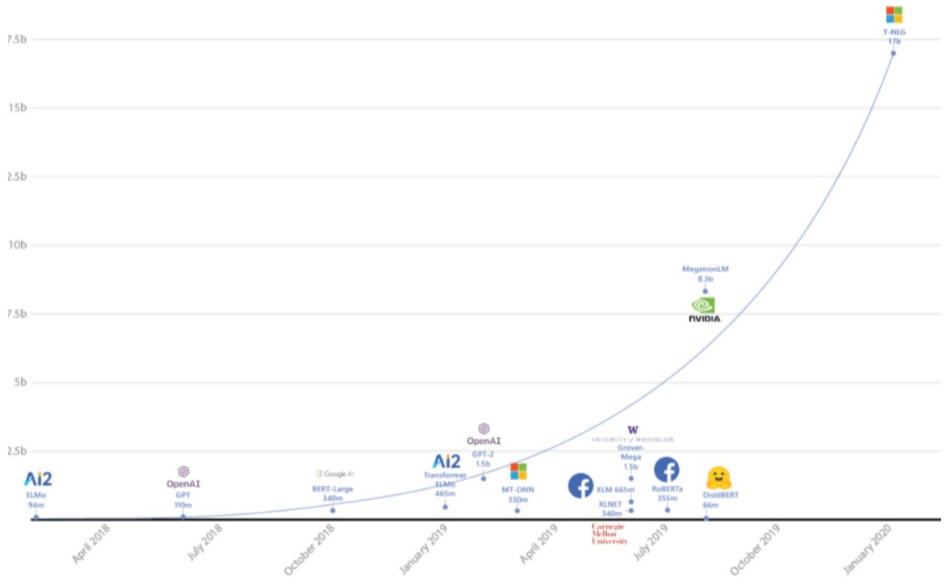
More data

Software

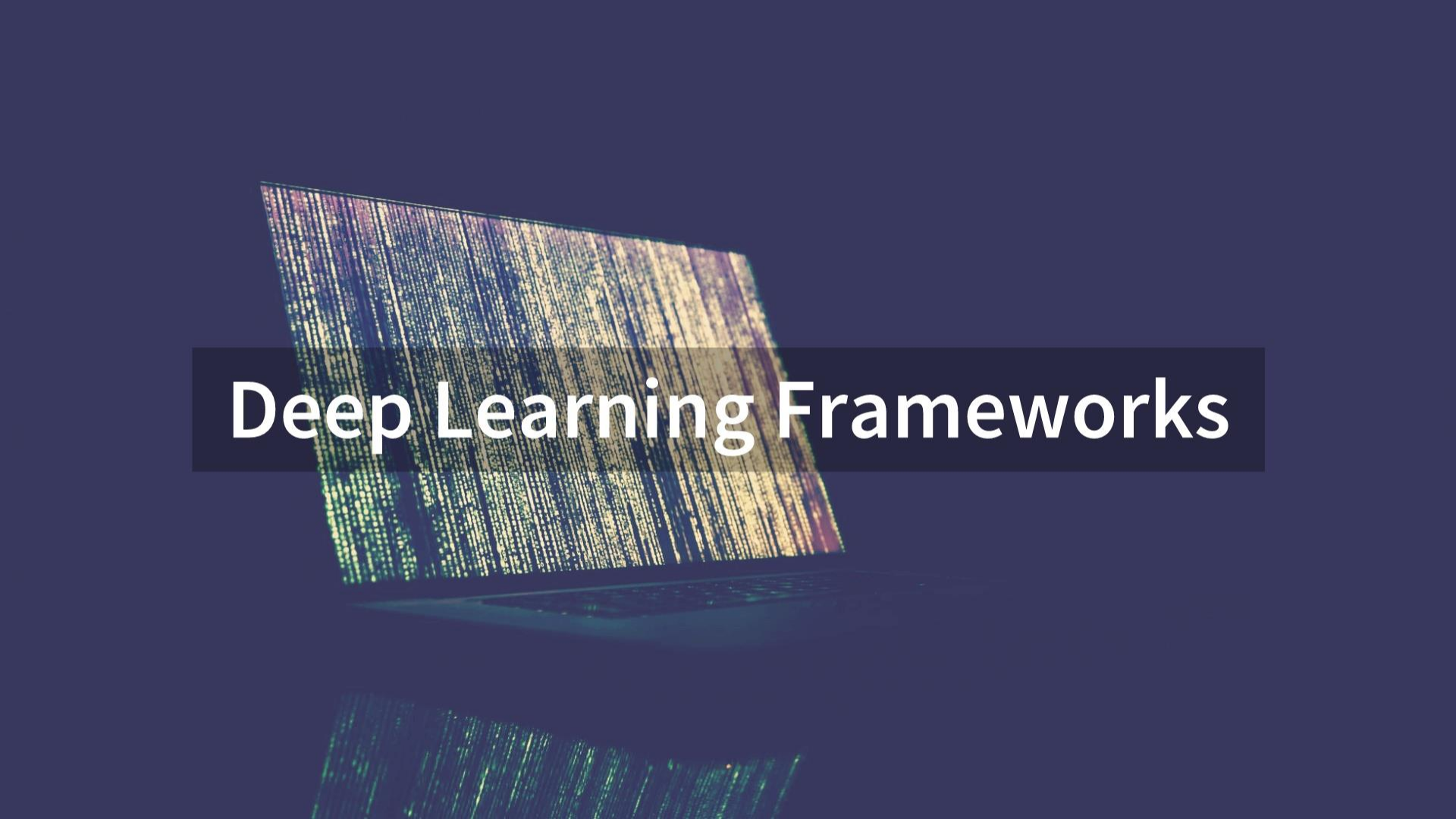
Caffe2 · Microsoft CNTK · TensorFlow · theano · Chainer · PYTORCH · dmlc mxnet · K

Faster compute engines

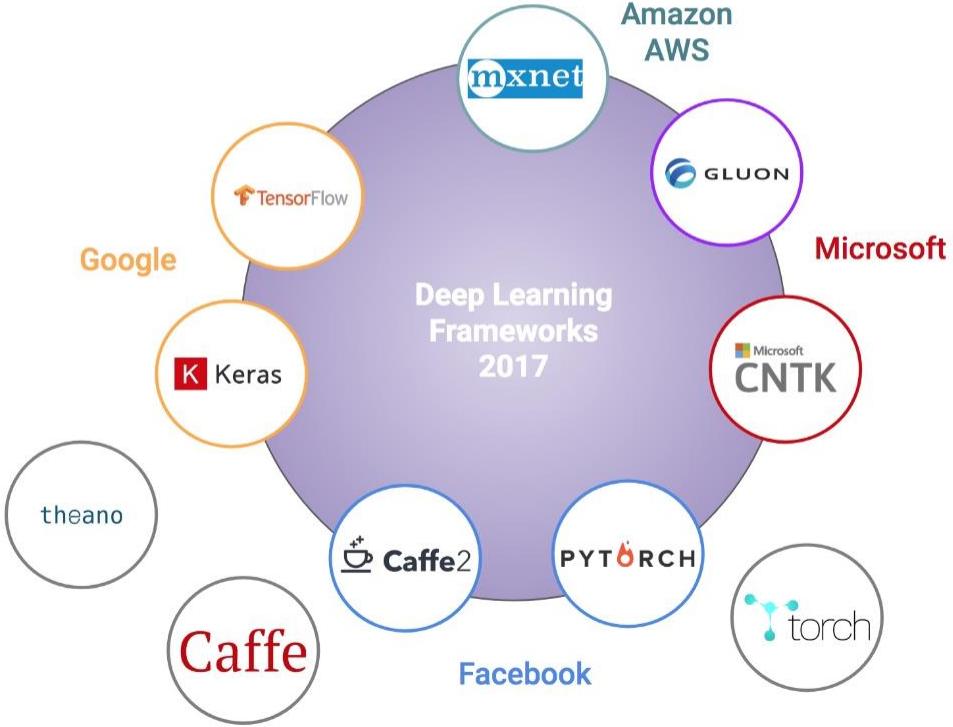
Transfer Learning - Advantage of Pre-trained Models

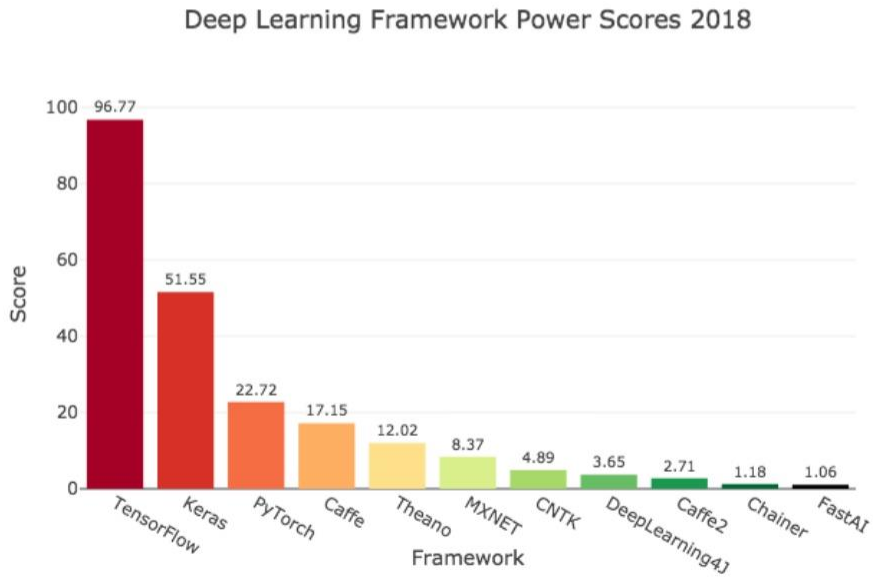


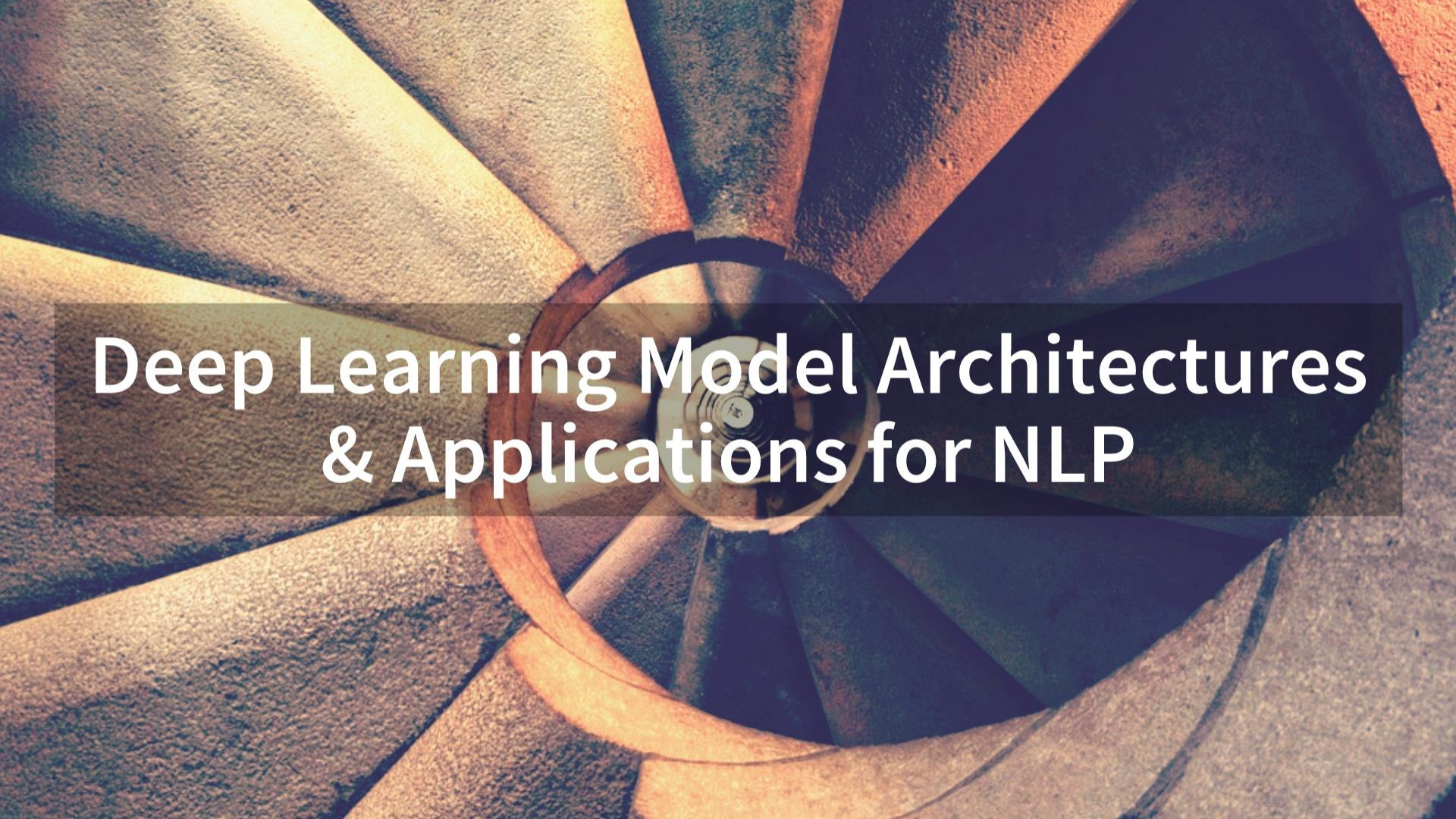
- Training DL models from scratch can take a lot of time
- Pre-trained models have been trained on a lot of data and specialized for performing specific tasks
- Can be used for both feature extraction and inference
- Models are available for computer vision and natural language processing tasks


Deep Learning Frameworks

Popular Deep Learning Frameworks - Open Source






Deep Learning Model Architectures
& Applications for NLP

Deep Learning Model Architectures

1. **Convolutional Neural Networks (CNNs)**

   Used extensively in computer vision problems with image, video. Can also be used for audio and text

2. **Recurrent Neural Networks**

   Good for sequential data, used for time series forecasting and NLP problems

3. **Long Short Term Memory Networks (LSTMs)**

   Can remember longer sequences of data and better than RNNs

4. **Gated Recurrent Units (GRUs)**

   Can remember longer sequences of data and faster than LSTMs

5. **Bi-directional Models**

   Processes sequences of data in both directions for capturing better contextual information

6. **Encoder-Decoder Models**

   Takes in a sequence of data and generates a sequence of data as output

7. **Transformer Models**

   Stack of encoder-decoder models used for language modeling and can be tuned for different NLP tasks

Deep Learning Applications for NLP

- **Text Classification**

  Support Ticket Classification, News Article Categorization

- **Text Clustering & Similarity**

  Recommender systems, Duplicate Detection with Fuzzy Matching

- **Search and Information Retrieval**

  Search Engines, Document Ranking

- **Parsing and Named Entity Recognition**

  Entities from health records, legal documents

- **Text Summarization**

  Topic models, summarizing entire documents

- **Machine Translation**

  Speech to Text, Language Translation

- **Conversational Interfaces**

  Chatbots, Personal Assistants, Q&A Systems

- **Sentiment Analysis**

  Survey result analysis, NPI analysis