ARTIFICIAL INTELLIGENCE AND NEURAL NETWORKS

SETTING THE STAGE

IDLEWYLD Sysabee DAVHILL

data-action-lab.com

OUTLINE

1. A Brief History of A.I.
2. Basics of Neural Networks
3. Case Study: Japanese Pharmacy Chain
4. Deep Learning

[Yes, Androids Do Dream of Electric Sheep, The Guardian UK]

HEADLINES

"AlphaGo vanquishes world's top Go player, marking A.I.'s superiority over human mind" [*South China Morning Post*, May 27, 2017]

"A Japanese A.I. program just wrote a short novel, and it almost won a literary prize" [*Digital Trends*, March 23, 2016]

"Elon Musk: Artificial intelligence may spark World War III" [*CNET*, September 4, 2017]

"A.I. hype has peaked so what's next?" [*TechCrunch*, September 30, 2017]

Q: How many legs does a cat have if you call the tail a leg?
A: Four. Calling the tail a leg doesn't make it a leg.

(old riddle, attributed to Abraham Lincoln)

WHAT IS ARTIFICIAL INTELLIGENCE (A.I.)?

What are the **essential qualities and skills** of an intelligence?

- provides flexible responses in various scenarios
- takes advantage of lucky circumstances
- makes sense out of contradictory messages
- recognizes the relative importance of a situation's elements
- finds similarities between different situations
- draws distinctions between similar situations
- comes up with new ideas from scratch or by re-arranging previous known concepts

WHAT IS ARTIFICIAL INTELLIGENCE?

A.I. research is defined as the study of **intelligent agents**: any device that perceives its environment and takes actions that maximize its chance of success at some goal.

Examples

- Expert Systems

  TurboTax, WebMD, technical support, insurance claim processing, air traffic control, etc.
- Decision-Making

  Deep Blue, auto-pilot systems, ”smart” meters, etc.
- Natural Language Proc.

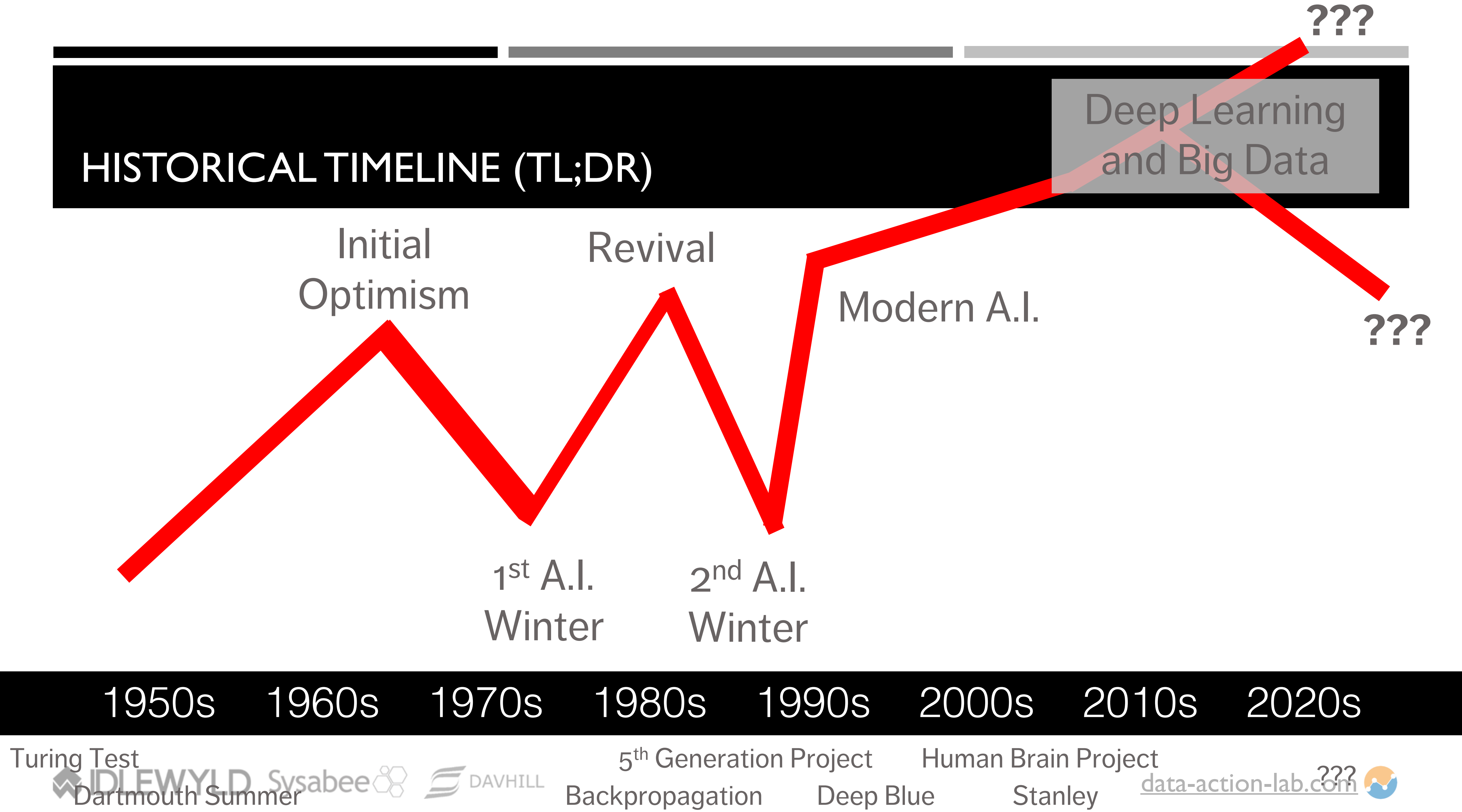
  machine translation, Siri, named-entity recognition, etc.
- Recommenders

  Google, Expedia, Facebook, LinkedIn, Netflix, Amazon, etc.
- Content generators

  music composer, novel writer, animation creator, etc.
- Classifiers

  facial recognition, object identification, fraud detection, etc.

HISTORICAL TIMELINE (TL;DR)



NEURAL NETWORKS IN A NUTSHELL

A trained **Artificial Neural Network** (ANN) is a function that maps inputs to outputs in a useful way:

- receives input(s)
- computes values
- provides output(s)

ANNs use a Swiss-army-knife approach to things (**plenty of options, but it's not always clear which one should be used**).

The user does not need to decide much about the function or know much about the problem space in advance (**quiet model**).

NEURAL NETWORKS IN A NUTSHELL

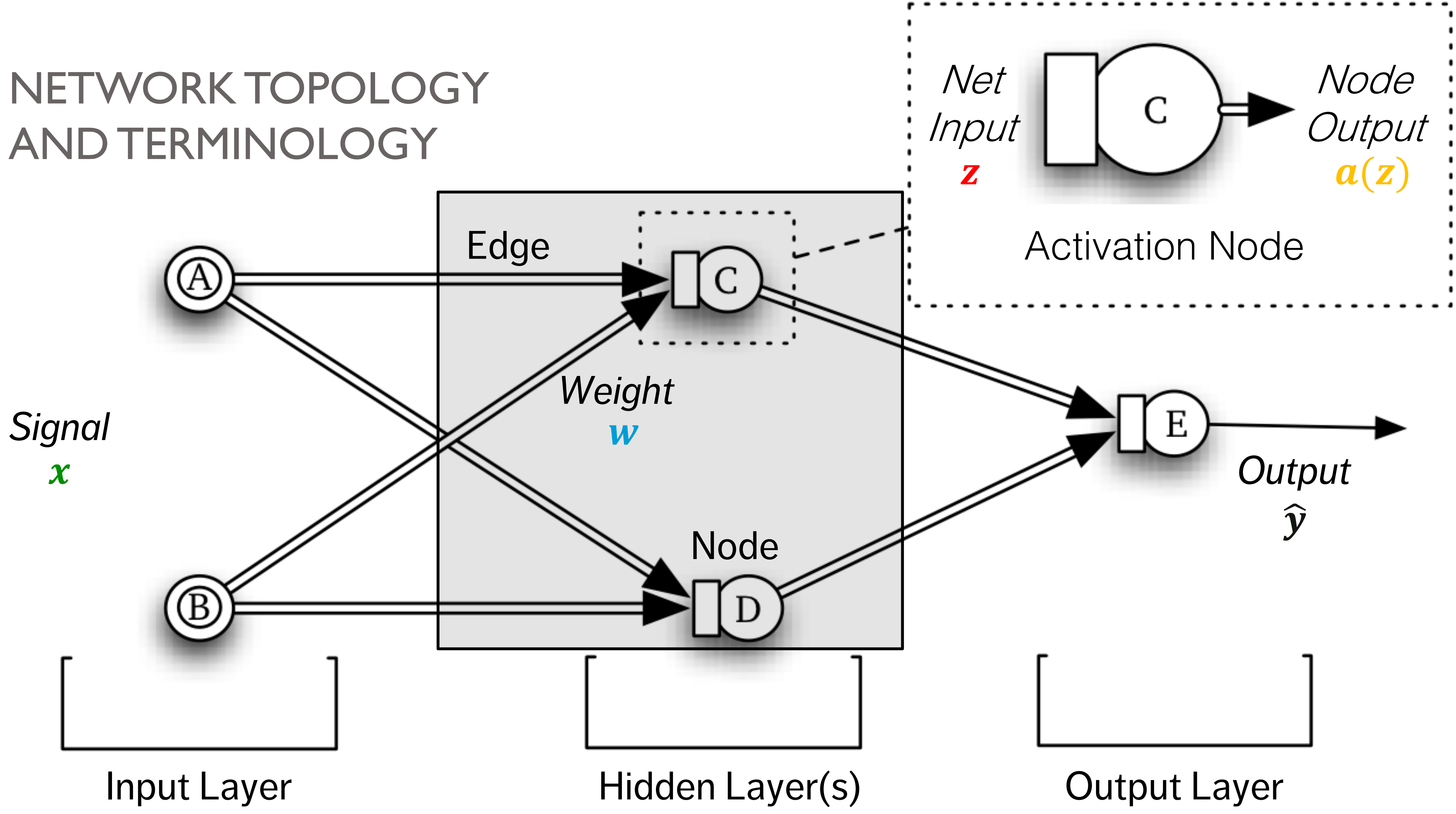
Algorithms allow ANNs to **learn** (i.e. generate the function and its internal values) **automatically**.

ANNs can be used for:

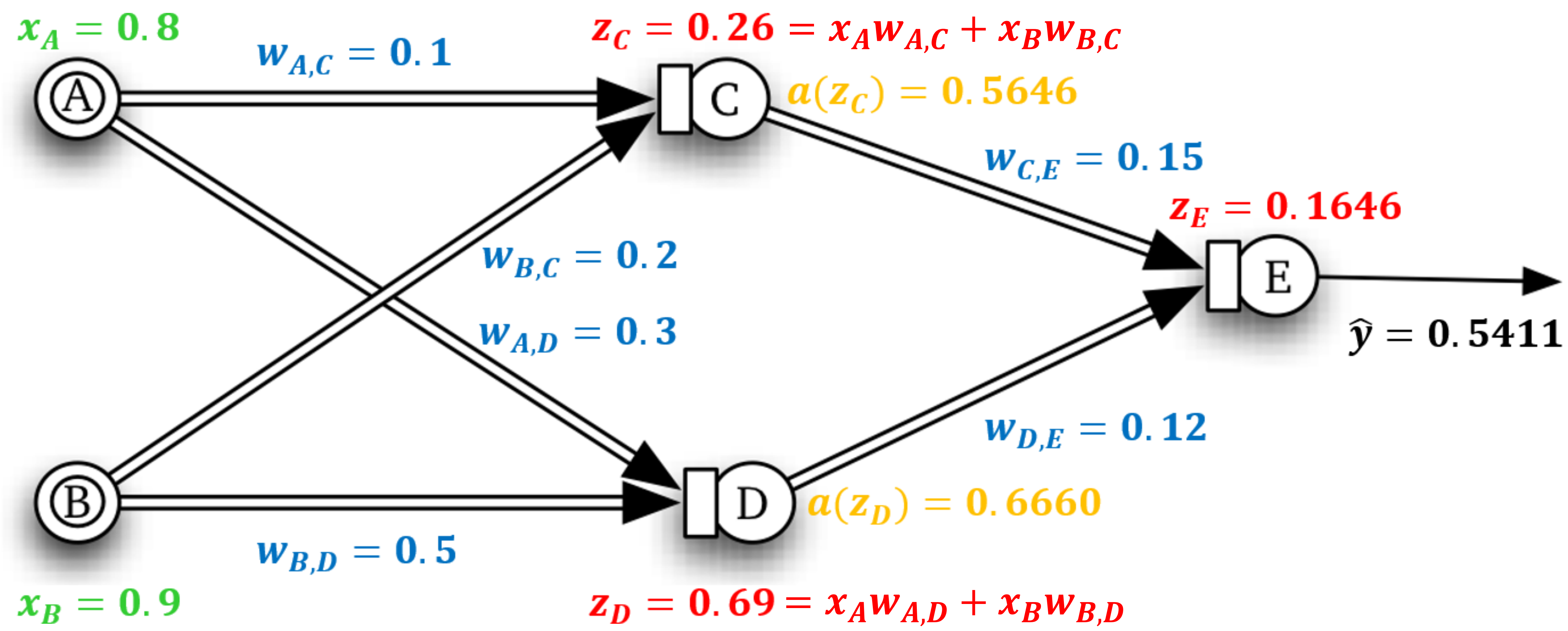
- supervised learning (**multi-layered feedforward neural networks**)
- unsupervised learning (**self-organizing maps**)
- reinforcement learning.

Technically, the only requirement is the ability to minimize a cost function (**optimization**).

NETWORK TOPOLOGY AND TERMINOLOGY



FEED FORWARD NETWORK



ANN IN MATRIX NOTATION

This *vanilla* neural net example can be expressed as:

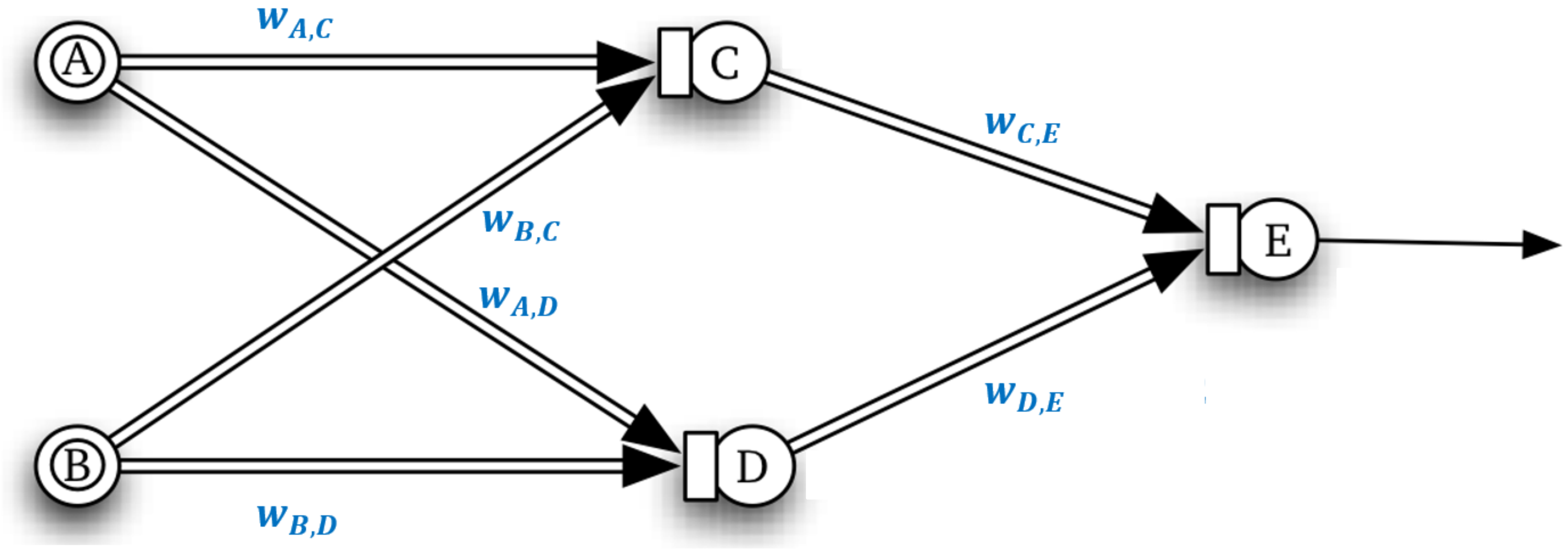
$$\hat{\boldsymbol{y}} = \boldsymbol{a}^{(3)} = g\left(\boldsymbol{Z}^{(3)}\right) = g\left[\boldsymbol{a}^{(2)}\boldsymbol{W}^{(2)}\right] = g\left[g\left(\boldsymbol{X}\boldsymbol{W}^{(1)}\right)\boldsymbol{W}^{(2)}\right]$$

In a nutshell, at each node, the neural net

1. computes a **weighted** **sum** of **inputs**
2. applies **activation** functions, and
3. sends a **signal**,

until the signal reaches the final **output** node.

BACKPROPAGATION OBJECTIVE



BACKPROPAGATION – ANN TRAINING

Given a signal, an ANN can produce an output, as long as the weights are specified.

For **supervised** learning tasks (i.e. when an ANN attempts to emulate the results of training examples), simply picking weights at random is a failing proposition.

Backpropagation is a method to optimize the choice of the weights against an error function $R(\mathbf{W})$ (usually done with numerical methods: gradient descent).

STRENGTHS

ANNs can be quite **accurate** when making predictions – better than other algorithms with a proper set up.

ANNs often work when other things fail:

- when the relationship between attributes is **complex**
- when there are a lot of dependencies/**nonlinear relationships**
- **messy**, highly connected inputs (images, text and speech)
- non-linear classification

ANNs are relatively easy to set up (with available packages).

ANNs degrade gracefully (important in robotics).

LIMITATIONS

ANNs are relatively slow (creating and using) and prone to overfitting (may require **large/diverse** training set).

ANNs usually do not provide good interpretation (unlike decision trees or logistic regression, say). Can you live with that?

No algorithms for selecting the optimal network topology.

Even when they do perform better than other options, ANNs may not perform that much better due to **No Free-Lunch Theorems**; and they're susceptible to various forms of **adversarial attacks**.

DISCUSSION

The biggest challenge (to our minds) is to overcome the black box nature of ANNs. How important is it to you and your organization to be able to explain data-driven decisions?

NEURAL NETWORKS VIDEOS (BRILLIANT!)

1. Neural Networks Demystified, Welch Labs
   https://www.youtube.com/watch?v=bxe2T-V8XRs (first in the series)
2. Learning to See, Welch Labs
   https://www.youtube.com/watch?v=i8D90DkCLhI (first in the series)
3. Neural Networks, 3 Blue 1 Brown
   https://www.3blue1brown.com/videos/2017/10/9/neural-network

CASE STUDY: JAPANESE PHARMACY – CONTEXT

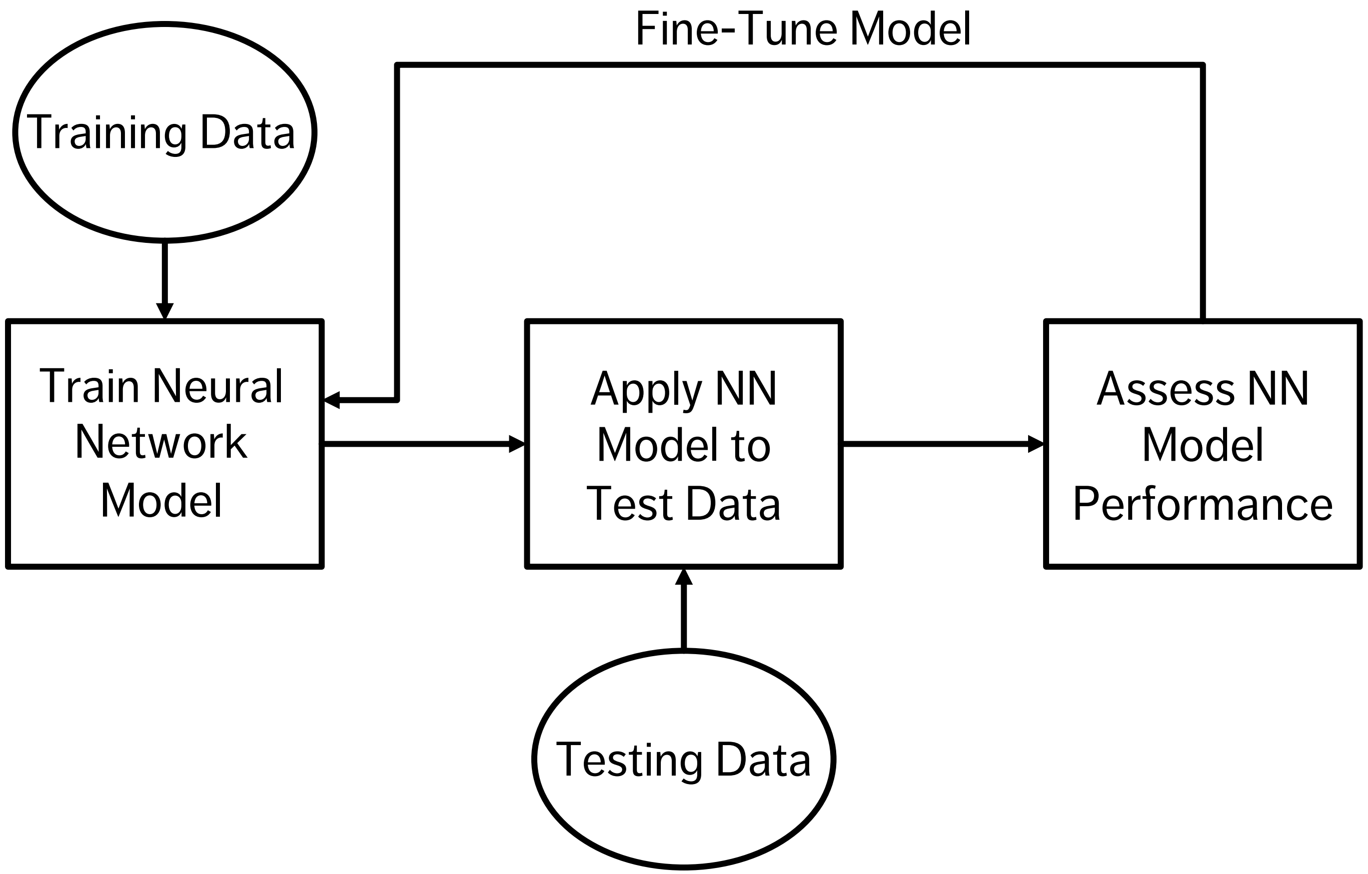
6 × as costly to sell to a new customer than to an existing one
(Kalakota, Robinson, Tapscott, 1999)

Annual customer retention up by 5% can lead to an 85% increase in profits
(Kalakota, Robinson, Tapscott, 1999)

Retaining the "right" customers plays a role in long-term profitability (Reicheld, 1993)

How can **loyal** and **profitable** customers be identified **early on**?

CASE STUDY: JAPANESE PHARMACY – LEARNING FLOW



CASE STUDY: JAPANESE PHARMACY – DATA

114,069 customers who made purchases during a 1-yr period

Customer value measured using

- frequency of visit (1 – 5 scale)
- profitability per visit (1 – 5 scale)

High-value customers (HVC): (4,5), (5,4), (5,5), making up 10.6% of observations

HVCs generate 52.5% of profits, 38.4% of revenues

CASE STUDY: JAPANESE PHARMACY – DATA

Target variable: customer value

Input variables: total number of categories purchased, profit per visit, # of units purchased per visit, # visits, purchase of:

All variables were scaled from 0 – 1:

- paper product
- detergent
- eye drops
- kitchen cleaner
- bottled supplement
- hair care products
- fabric softener
- household cleaner
- toothpaste
- cold medicine

CASE STUDY: JAPANESE PHARMACY – DATA

Learning set: 104,069 observations (selected randomly)

Training/Testing set ratio: 70-to-30

Validation set: remaining 10,000 observations

Calibration metrics: predictive accuracy, overall accuracy

predictive accuracy = # correctly predicted HVCs / # HVCs

overall accuracy = # correctly predicted class / # customers

CASE STUDY: JAPANESE PHARMACY – RESULTS

Using a **Multilayered Feed Forward Neural Network** (MFFN), researchers were able to capture 80% of the HVCs by targeting (a model-specified) 25% of new customers.

At a threshold parameter value of 30%, the model performs 5 × better than randomly classifying customers.

Dataset	Training	Validation
Predictive Accuracy	55.6%	57.4%
Overall Accuracy	90.6%	91.2%
% Classified as HVC	10.6%	10.3%

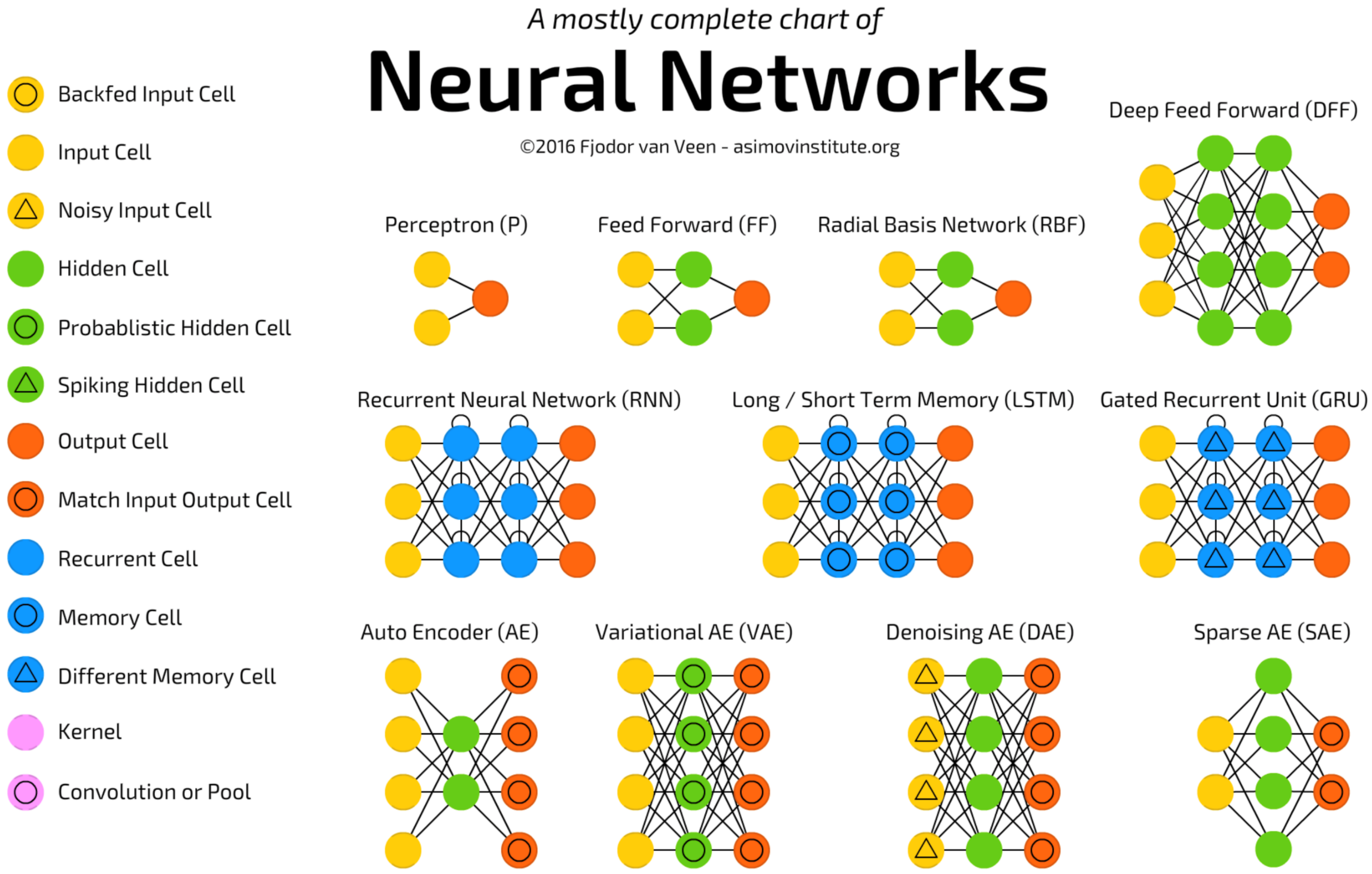
DEEP LEARNING NETWORKS

Deep Learning networks are simply ANNs with **a large number of hidden layers** (and various types of nodes)

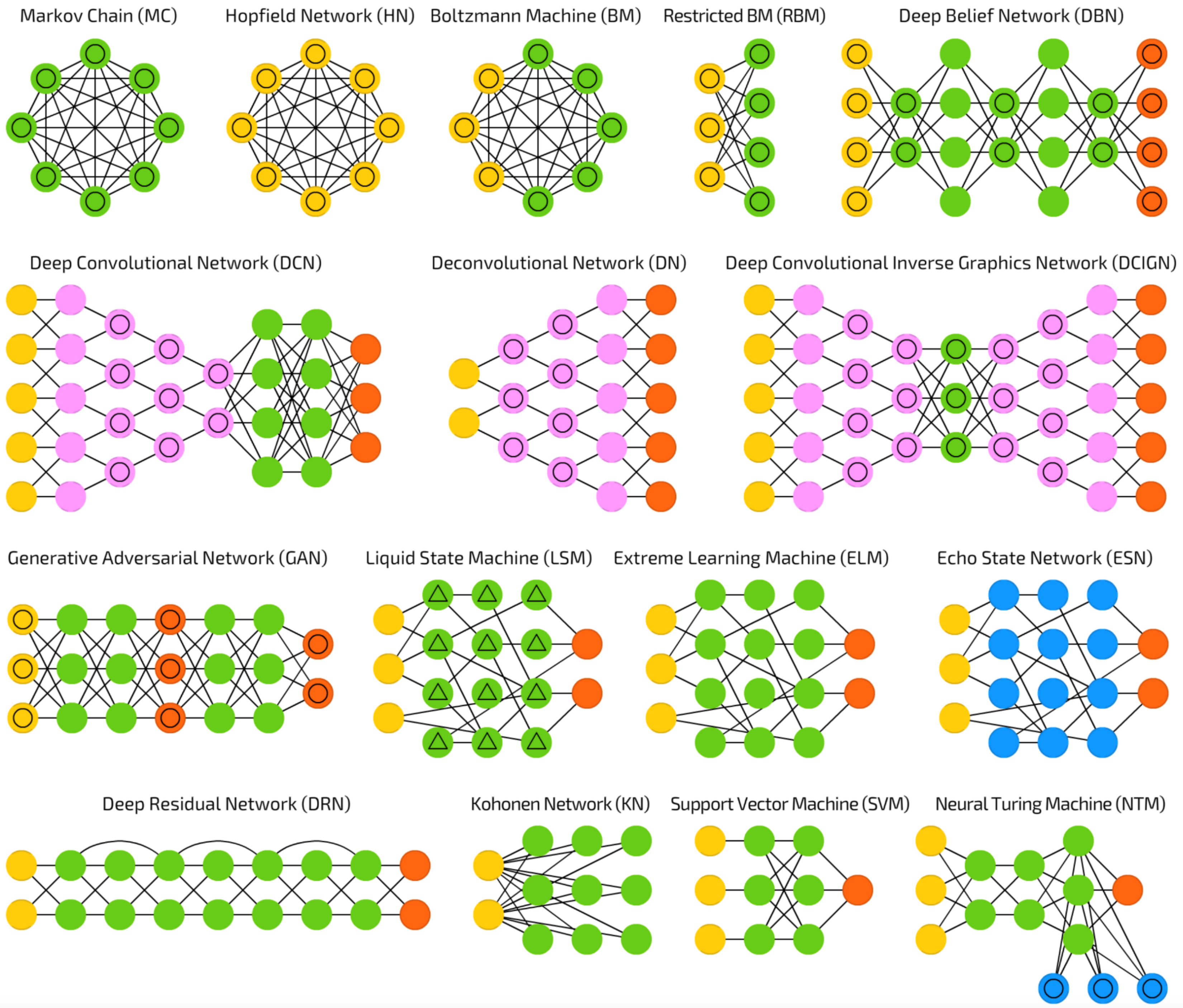
Types:

- *Convolution Neural Networks*
  Handwritten digit recognition, 99.7% accuracy in 2013, Self-driving cars
- *Recurrent Neural Networks*
  Natural language processing (speech recognition, machine translation, etc.)
- *Autoencoders*
- *Restricted Boltzmann Machines*
  BellKor's Pragmatic Chaos, Netflix Prize, 2009

[Fjodor van Veen, Asimov Institute, 2016]



[Fjodor van Veen, Asimov Institute, 2016]



LIMITATIONS

Require **large**, **diverse**, and **correctly labeled** training sets.

Accurate on average, but they can still be **spectacularly** wrong.

They can be “hacked” (NFL).

Humans don’t need that much labeled data to make decisions: so **what’s really going on under the hood**? (3rd AI Winter?)