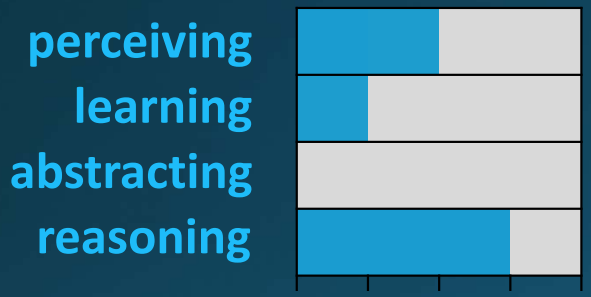
A DARPA Perspective on Artificial Intelligence

John Launchbury

Director I2O, DARPA

Ability to process information

Notional intelligence scale



perceive
rich, complex and subtle information
learn
within an environment
abstract
to create new meanings
reason
to plan and to decide

Artificial intelligence is a *programmed* ability to process information

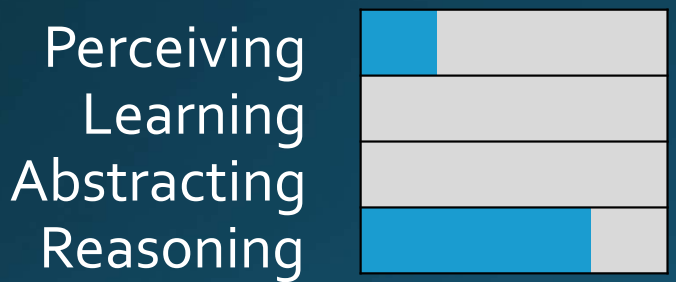
Three waves of AI

Handcrafted Knowledge
Statistical Learning
Contextual Adaptation

The first wave of AI

Handcrafted Knowledge

The first wave of AI



Enables reasoning over narrowly defined problems

No learning capability and poor handling of uncertainty

The first wave of AI

Engineers create sets of rules
to represent knowledge in
well-defined domains

The **structure** of the knowledge is defined by humans
The **specifics** are explored by the machine

First wave stumbles

2004
completed: 0

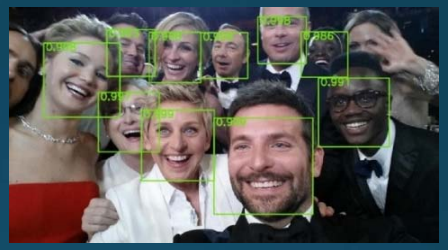
Source: DARPA

2005
completed: 5

DARPA Autonomous Vehicle Grand Challenge
140 miles of dirt tracks in California and Nevada

The second wave of AI

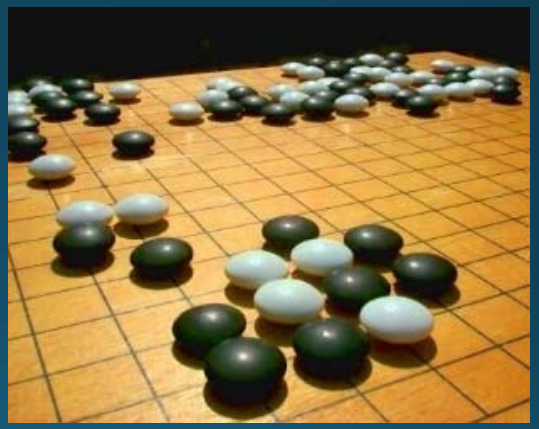
Source: thrillist.com



Statistical Learning

The second wave of AI

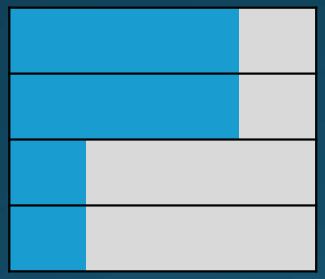
Engineers create statistical models for specific problem domains and train them on big data



Source: gobelluno.it

The second wave of AI

Perceiving
Learning
Abstracting
Reasoning



Nuanced classification and prediction capabilities

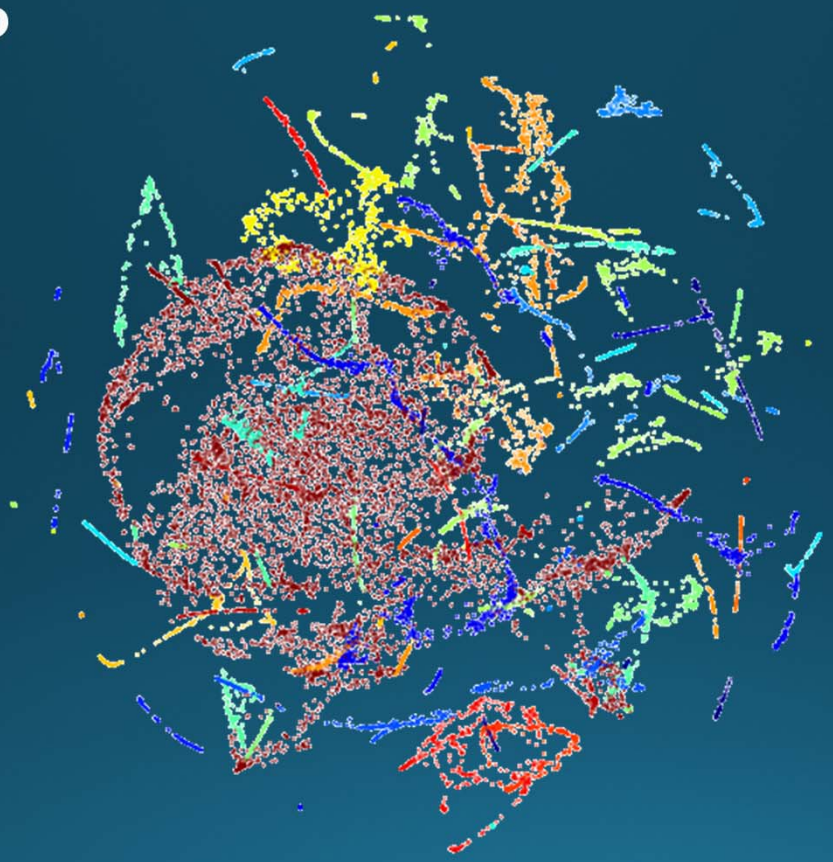
No contextual capability and minimal reasoning ability

Second wave and natural data

Manifold hypothesis

Natural data forms lower dimensional structures (manifolds) in the embedding space

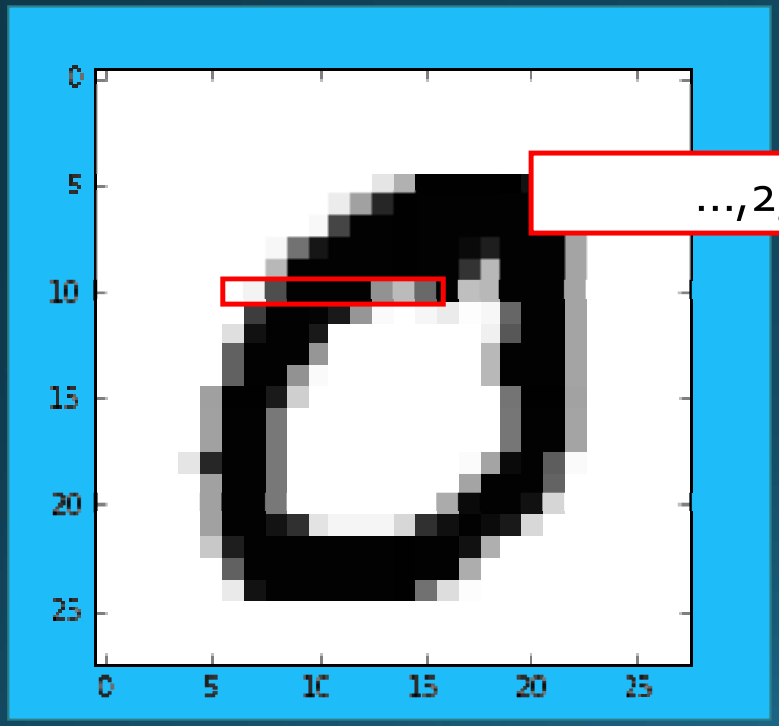
Manifolds



Each manifold represents a different entity

Understanding data comes by separating the manifolds

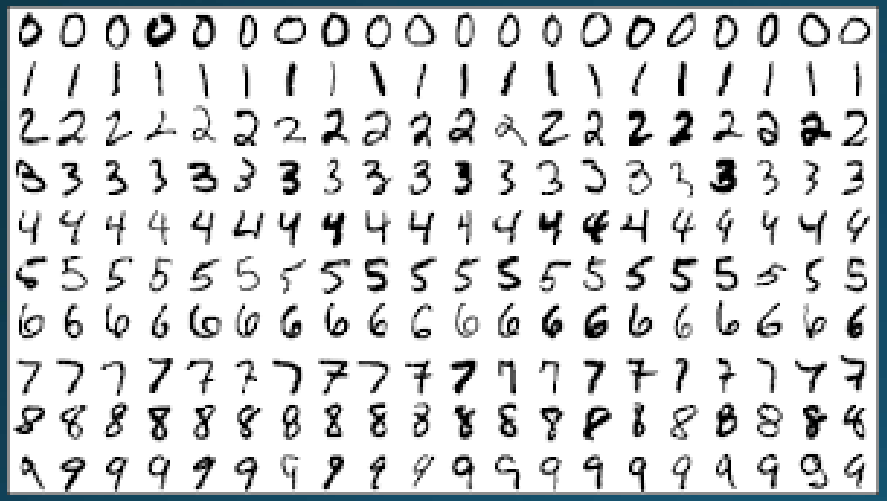
Handwritten digits



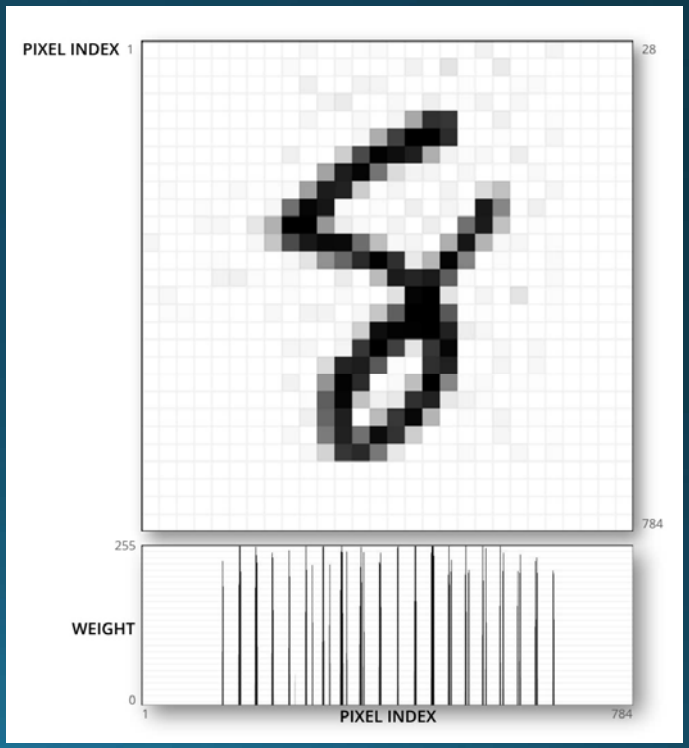
…,253,247,228,118,41,38,34,39,147,198,…

Each image is an element living within a data space of 784 "dimensions"

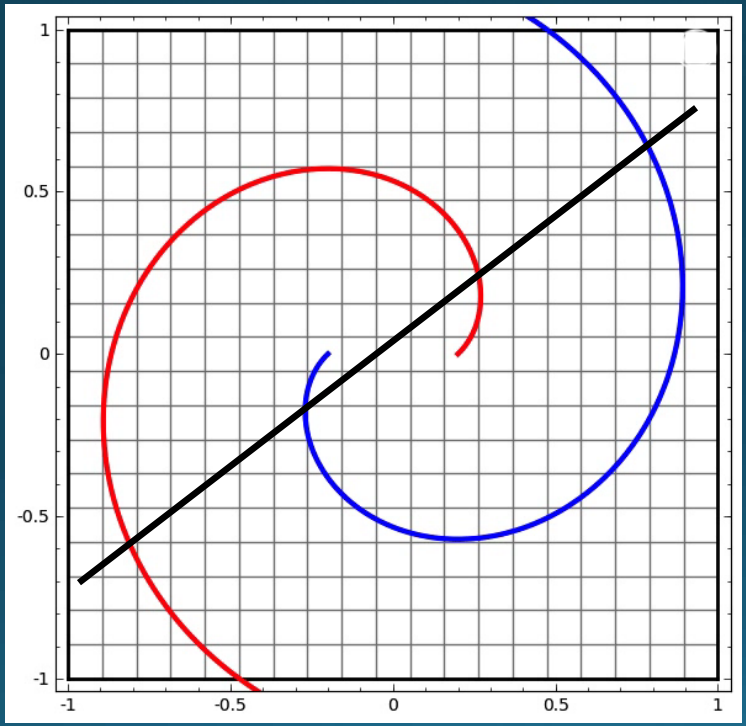
Manifolds of handwriting



Variation in handwritten digits
form 10 distinct manifolds within the
28x28 dimensional space of pixel values



Separating manifolds



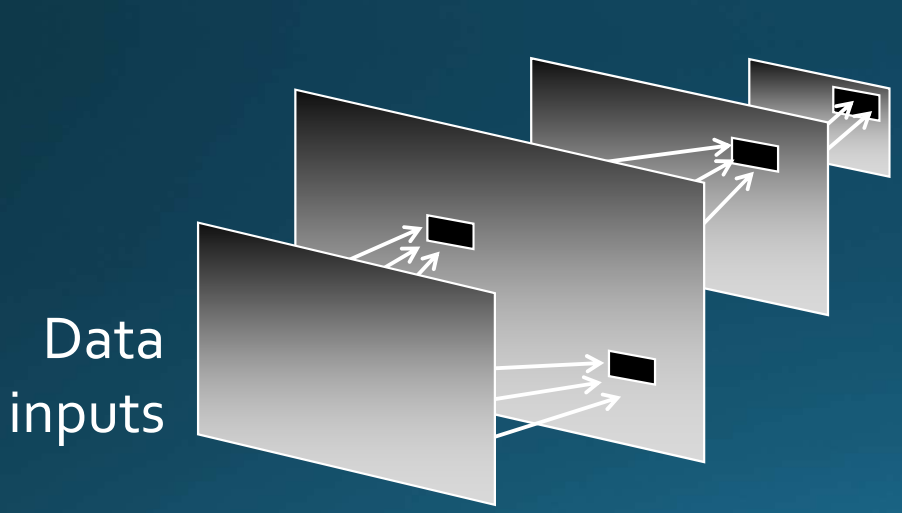
Imagine the spiral arms are each clusters of data

Stretching and squashing the data space separates them cleanly

Separating manifolds

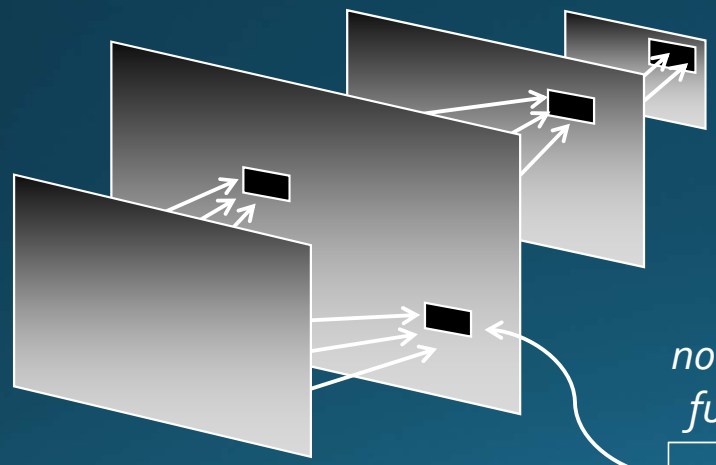
Stretching in a new dimension enables enclosed manifolds to be isolated

Neural nets



Each layer stretches and squashes the data space until the data manifolds are cleanly separated

Neural nets learn from data



Data inputs

Computed outputs

Compute outputs from inputs

Adjust weights by error propagation

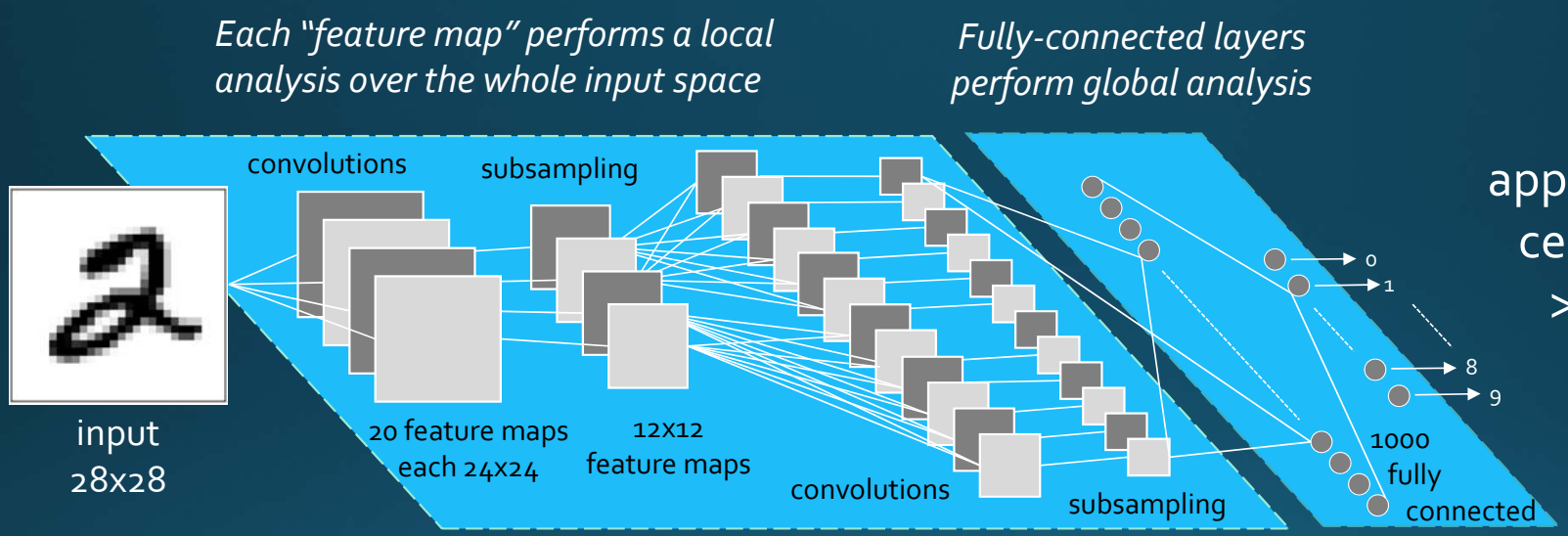
non-linear function

cell weights (learned)

cell inputs (from prev layer)

= POS (SUMPRODUCT(W1:W16, V1:V16))

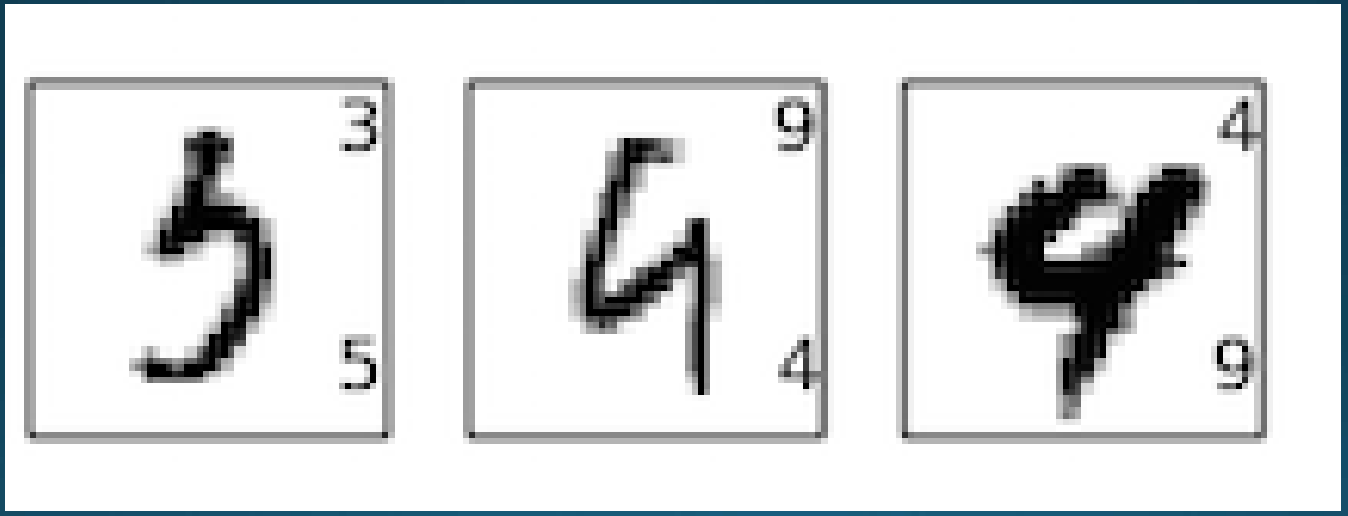
Structured neural net



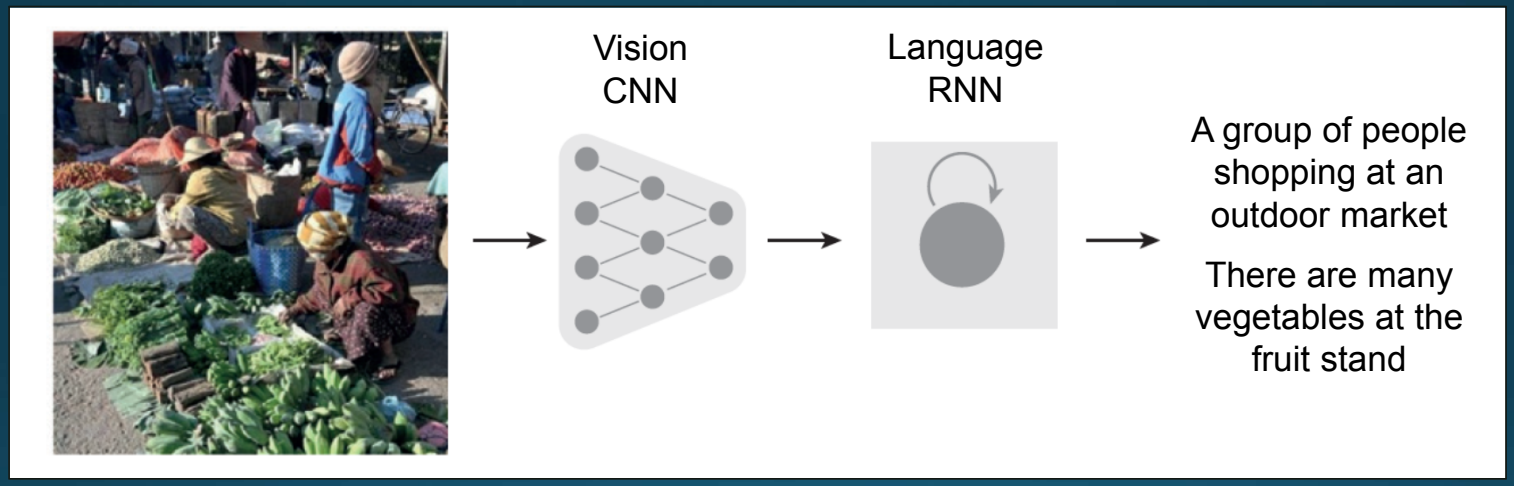
approx. 30,000 cells in total for >99.5% accuracy

Machine-learning "programmers" design the network structure with experience and by trial and error

Example failures



Layering neural networks



Yann LeCun, Yoshua Bengio, & Geoffrey Hinton (2015). Deep Learning, Nature, Vol. 521, (pp. 436-444)

A deep convolution neural net (CNN) produces a set of outputs (abstract "words")

A language-generating recurrent neural net (RNN) "translates" the abstract "words" into captions

AI technology is powerful

Code and network flows

Observe real-time cyber attacks at scale

Electromagnetic spectrum

Overcome spectrum scarcity to meet wireless data demand

Autonomous platforms

Reshape defense missions

Challenges with second wave

Statistically impressive,
but individually unreliable

Challenges with second wave

"Panda"

+

< 1%
targeted distortion

=

"Gibbon"
(99.3% confidence)

Inherent flaws can be exploited

Challenges with second wave

Internet trolls cause the AI bot, Tay, to act offensively

Skewed training data creates maladaptation

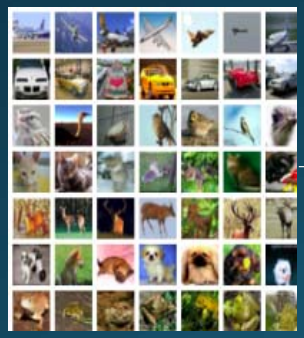
The (future) third wave of AI

Contextual adaptation

Systems construct contextual explanatory models
for classes of real world phenomena

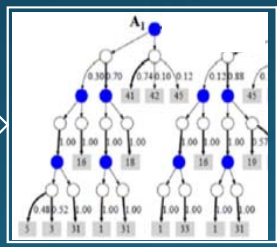
Models to explain decisions

Source: SPIN South West



Training Data

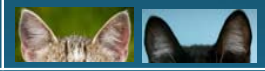
Learning Process



Explainable Model

This is a cat:

- It has fur, whiskers, and claws.
- It has this feature:



Explanation Interface

Decision

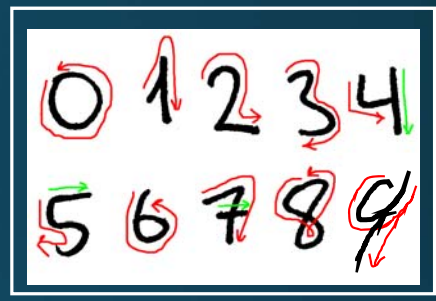
Explanation

I understand why
I understand why not
I know when you'll succeed
I know when you'll fail
I know when to trust you
I know why you made that mistake

Models to drive decisions


```
Probable number of strokes: 1 – 4
Each stroke: probable trajectory
Each trajectory: probable shift in
                 shape and location
```

Seed model



Training data

Generative model

Generates explanations of how a test character might have been created

The third wave of AI

perceive

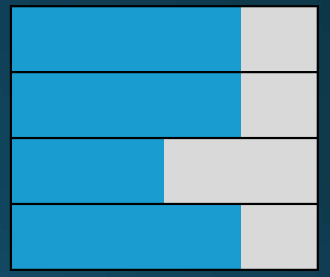
abstract

contextual model

learn

reason

Perceiving
Learning
Abstracting
Reasoning



Three waves of AI

Handcrafted Knowledge
Statistical Learning
Contextual Adaptation