

# The Promise, Pitfalls and Process of AI

---

Stephen J. Mildenhall

November 14, 2019

Predictive Analytics  
Deep Learning  
Data Science Expert System  
Data Visualization  
Statistics  
Exploratory Data Analysis  
Neural Network  
Big Data  $p > n$   
Artificial Intelligence  
Causal Analysis  
Data Representation  
Machine Learning

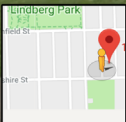
Inference  
Data Representation

1006 Woodbine Ave

Oak Park, Illinois

Google

Street View - Oct 2018



Google

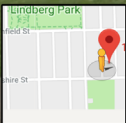
1006 Woodbine Ave

Oak Park, Illinois

Google

Street View - Oct 2018

# Buying a House: How Much to Offer?



Google

## Facts and Features



**Type**  
Single Family



**Year Built**  
1928



**Heating**  
Radiant



**Cooling**  
Central



**Parking**  
2 spaces



**Lot**  
7,400 sqft

### INTERIOR FEATURES

#### Bedrooms

Beds: 3

#### Bathrooms

Baths: 1 full, 1 half

#### Heating and Cooling

Heating: Radiant

Heating: Gas

Cooling: Central

#### Basement

#### Appliances

Appliances included:

Dishwasher, Dryer, Garbage disposal,  
Microwave, Range / Oven, Refrigerator,  
Trash compactor, Washer

#### Flooring

Floor size: 2,346 sqft

Flooring: Hardwood

#### Other Interior Features

*View more*

[See More Facts and Features](#)

## Home Value

Zestimate   
**\$565,863**

1010 Woodbine Ave, Oak Park,

Contact Agent

Or call 773-974-9996 for more info

## Nearby Similar Sales

**SOLD: \$565,000**

Sold on 05/23/18

3 bds, 2 ba, 1,932 sqft

[846 Linden Ave, Oak Park, IL 60302](#)

**SOLD: \$560,000**

Sold on 05/29/18

3 bds, 2.5 ba, 2,660 sqft

[830 Belleforte Ave, Oak Park, IL 60302](#)

**SOLD: \$570,000**

Sold on 04/26/19

4 bds, 2.5 ba, 2,332 sqft

[913 Linden Ave, Oak Park, IL 60302](#)

**SOLD: \$550,000**

Sold on 11/13/18

3 bds, 2.5 ba, 1,696 sqft

[1047 N Grove Ave, Oak Park, IL 60302](#)

**SOLD: \$540,000**

Sold on 11/30/18

4 bds, 2.5 ba, 2,204 sqft

[1015 Woodbine Ave, Oak Park, IL 60302](#)

[See sales similar to 1010 Woodbine Ave](#)

# Comparables

- Statistics: learning from data
- **Descriptive** statistics: summarize with no interpretation
- **Exploratory** data analysis (EDA): claims about current sample

# Generalizing Comparables

- Relevant comparables may not be available
- Extend limited data
- **Predictive** modeling: estimate value for an **out of sample** unit
- Buyer's perspective

# Understanding Comparables

- Seller's perspective: does an investment to
  - remodel the kitchen or
  - paint or
  - landscape or
  - add a bathroom or
  - make life easier for my broker

make sense?

- **Inferential** statistics



# Machine Learning (ML)

- **Machine learning**: an algorithm allowing computer to build and improve a model from more **data**

# Machine Learning (ML)

- **Machine learning**: an algorithm allowing computer to build and improve a model from more **data**
- GLMs
- Logistic regression
- Decision trees
- iCAS CSPA material

# Machine Learning (ML)

- **Machine learning**: an algorithm allowing computer to build and improve a model from more **data**
- GLMs
- Logistic regression
- Decision trees
- iCAS CSPA material
- **Not ML**: expert system, knowledge base

# Machine Learning (ML)

- **Machine learning:** an algorithm allowing computer to build and improve a model from more **data**
- GLMs
- Logistic regression
- Decision trees
- iCAS CSPA material
- **Not ML:** expert system, knowledge base

## ML Problems

- Classification
- Regression
- Transcription
- Machine translation
- Anomaly detection
- Synthesis
- Impute missing values
- De-noise
- Density estimation

Comparables. . .

# Comparables. . .

- Encapsulate considerable domain expertise
- “**Raw data** is both an oxymoron and a bad idea; to the contrary, data should be cooked with care.” (Geoffrey Bowker)

...and are a powerful data representation

# Three Houses in Same Town: Rank by Market Value



# Three Houses in Same Town: Rank by Market Value





# Three Houses in Same Town: Rank by Market Value



# Three Houses in Same Town: Rank by Market Value



- 1)  $A < B < C$
- 2)  $B < A < C$
- 3)  $B < C < A$
- 4)  $C < B < A$
- 5) None of the above

FYI, values are  $x : 3x : 9x$

Home Facts by Data Source ALL SOURCES COMBINED COUNTY RECORDS County website Beds: 6 - Baths: 7.5 4.5 SqFt: 12,000 7,914 Lot: 39,622 sq ft / 0.91 acres 39,622 sq ft / 0.91 acres Type: Single Family Single Family Year Built: 1929 1929 Last Sold: Aug 2011 for \$2,100,000 - Parking: Attached Garage Garage - Attached Cooling: Central Central Heating: Gas - Fireplace: Yes Yes Days on Zillow: 47 - MLS #: 1032977 - Other facts Additional Features: Air Conditioning: Zoned, Space Pac. Appliances: All Stainless Steel Kitchen Appliances, 2nd Bedroom Level: 2nd Level, Built Before 1978 (Y/N): Yes. Electricity: Circuit Breakers, 400 Amp Service or Greater, Family Room Level: Not Applicable, Heat/Fuel: Hot Water/Steam, 2+ Sep Heating Systems: Zoned, Sewer: Sewer-Public, Listing Type: Exclusive Right To Sell, Master Bedroom Level: 2nd Level, Parking Type: Garage, Water: Lake Michigan, Kitchen Level: Main Level, Living Room Level: Main Level, Exposure: W (West), S (South), E (East), N (North), Sale Terms: Conventional, 3rd Bedroom Level: 2nd Level, Equipment: Sump Pump, Sprinkler Lawn, Power Generator, Kitchen Type: Pantry, Closet: Eating Area, Table Space, Island, Basement Bathrooms (Y/N): Yes, Addtl Room 2 Level: 2nd Level, Addtl Bath Amenities: Double Sink, Steam Shower, Lot Description: Corner, Addtl Room 4 Level: Main Level, Addtl Room 1 Level: 3rd Level, Addtl Room 3 Level: 3rd Level, 4th Bedroom Level: 2nd Level, Addtl Room 3 Name: Office, Addtl Room 5 Level: 3rd Level, Addtl Room 1 Name: 5th Bdrm, Addtl Room 2 Name: 6th Bdrm, Frequency: Not Applicable, Attic: Full, Interior Stair, Finished, Addtl Room 6 Level: Basement, Addtl Room 7 Level: 2nd Level, Addtl Room 9 Level: Main Level, Addtl Room 8 Level: 2nd Level, Status: New, Addtl Room 4 Name: Library, Addtl Room 5 Name: Attic, Addtl Room 9 Name: Gallery, Square Feet Source: Estimated, Master Bedroom Bath (Y/N): Full, Age: 81-90 Years, Area Amenities: Street Lights, Street Paved, Sidewalk, Pool, Garage On-Site: Yes, Additional Rooms: Recreation Room, 5th Bedroom, Sitting Room, 6th Bedroom, Recent Rehab (Y/N): Yes, Additional Sales Information: List Agent Must accompany, 2nd Floor Laundry, In-Law Arrangement, Heated Floors, Breezeway, Dog Run, Fireplace Location: Other, Is Parking Included in Price: Yes, Type of House 2: 2 Stories, Garage Ownership: Owned, Laundry Level: Not Applicable, Additional Rooms: Terrace, Attic, Office, Gallery, Library, Addtl Room 6 Name: Recreation Rm, Addtl Room 7 Name: Sitting, Addtl Room 10 Level: Main Level, Addtl Room 8 Name: Sitting, Addtl Room 10 Name: Terrace, Aprox. Total Finished Sq Ft: 0, Total Sq Ft: 0, Lot Dimensions: 215 X 184, Tax Year: 2017, Parcel Identification Number: 15013170210000 - Appliances Included: Dishwasher, Dryer, Garbage disposal, Microwave, Range / Oven, Refrigerator, Washer - County Name: Cook County Cook Covered Parking Spaces: 4 4 Exterior Material: Brick, Stone Stone Floor Covering: Carpet, Hardwood - Heating Type: Forced air - Laundry: In Unit - Parcel #: 1501317021 1501317021 Pets: Contact manager - Roof Type: Slate Slate Room Count: 0 - Room Types: Dining room - Stories: 0 0 2 Structure Type: Tudor Split level Unit Count: 0 0 Zillow Home ID: 123561615 -

From [https://www.zillow.com/homes/for\\_sale/Oak-Park-IL-60302/123561615\\_zpid/84453\\_rid/globalrelvancesort\\_41.897302\\_-87.808603\\_41.886408\\_-87.826799\\_react/15\\_zm/](https://www.zillow.com/homes/for_sale/Oak-Park-IL-60302/123561615_zpid/84453_rid/globalrelvancesort_41.897302_-87.808603_41.886408_-87.826799_react/15_zm/)

Home Facts by Data Source ALL SOURCES COMBINED COUNTY RECORDS County website Beds: 5 4 Baths: 4.0 3.5 SqFt: 3,360 3,360 Lot: 13,560 sq ft / 0.31 acres 13,560 sq ft / 0.31 acres Type: Single Family Single Family Year Built: 1899 1898 Last Sold: -- - Parking: Detached Garage Garage - Detached Cooling: Central - Heating: Gas - Fireplace: Yes Yes Days on Zillow: 117 - MLS #: 10344749 - Other facts Additional Features: Addtl Room 5 Level: Not Applicable, Appliances: Oven-Double, All Stainless Steel Kitchen Appliances, 2nd Bedroom Level: 2nd Level, Built Before 1978 (Y/N): Yes, Sewer: Sewer-Public, Listing Type: Exclusive Right To Sell, Master Bedroom Level: 2nd Level, Parking Type: Garage, Space/s, Tax Exemptions: Homeowner, Equipment: Ceiling Fan, CO Detectors, Water: Lake Michigan, Kitchen Type: Eating Area-Breakfast Bar, Eating Area-Table Space, Island, Kitchen Level: Main Level, Living Room Level: Main Level, 3rd Bedroom Level: 2nd Level, Other Information: School Bus Service, Commuter Train, Historical District, Exposure: E (East), Electricity: 200+ Amp Service, Exterior Building Type: Frame, Addtl Room 2 Level: Main Level, Addtl Room 2 Name: Foyer, Addtl Room 3 Level: 2nd Level, Addtl Room 1 Level: 3rd Level, Addtl Room 4 Level: 2nd Level, 4th Bedroom Level: 2nd Level, Driveway: Side Drive, Concrete, Addtl Room 4 Name: Balcony, Addtl Room 1 Name: 5th Bdrm, Family Room Level: 3rd Level, Age: 100+ Years, Addtl Room 10 Level: Not Applicable, Addtl Room 6 Level: Not Applicable, Addtl Room 7 Level: Not Applicable, Addtl Room 8 Level: Not Applicable, Addtl Room 9 Level: Not Applicable, Frequency: Not Applicable, Attic: Full, Interior Stair, Finished, Foundation: Stone, Status: New, Square Feet Source: Assessor, Master Bedroom Bath (Y/N): Full, Bath Amenities: Soaking Tub, Area Amenities: Curbs/Gutters, Street Lights, Street Paved, Sidewalks, Garage On-Site: Yes, Parking On-Site: Yes, Additional Rooms: Balcony, 5th Bedroom, Sitting Room, 6th Bedroom, List Agent Must accompany, Heated Floors, Balcony, Parking Ownership: Owned, Fireplace Location: Other, Is Parking Included in Price: Yes, Parking: Driveway, Type of House 2: 3 Stories, Lot Dimensions: 78X175, Additional Rooms: Foyer, Garage Ownership: Owned, Laundry Level: Not Applicable, Aprox. Total Finished Sq Ft: 0, Total Sq Ft: 0, Basement Sq Ft: 0, Tax Year: 2017, Parcel Identification Number: 16071030160000 - Appliances Included: Dishwasher, Dryer, Garbage disposal, Microwave, Range / Oven, Refrigerator, Trash compactor, Washer - Assisted Living Community: No - Attic: Yes - Barbecue Area: No - Basement Type: Partial - Basketball Court: No - Cable Ready: No - Ceiling Fan: Yes - Controlled Access: No - County Name: Cook County Cook Covered Parking Spaces: ? 2 Deck: No - Disability Access: No - Dock: No - Double Pane Glass Windows: No - Elementary School: Oliver W Holmes Elementary School - Elevator: No - Exterior Material: Stone, Wood Wood Fenced Yard: Yes - Floor Covering: Hardwood, Slate - Garden: Yes - Gated Entry: No - Greenhouse: No - Fitness Center: No - Heating Type: Baseboard, Radiant - High School: Oak Park & River Forest High Sch - Hot Tub/Spa: No - Intercom: No - Jetted Tub: No - Laundry: In Unit - Lawn: No - Legal Description: - (WORRWICKS) SUB OF LTI E N7T4 & ALL OF LT2 IN BLK12 IN KETTLETRINGS ADD TO HARLEM SUB OF PT OF NW SEC 07.39.133 Middle School: GWENDOLYN BROOKS MIDDLE SCHOOL - Mother-in-Law Apartment: No - Near Transportation: No - Over 55 Active Community: No - Parcel #: 1607103016 1607103016 Petals: Yes - Pets: Contact manager - Pond: No - Pool: Yes - Porch: Yes - Roof Type: Asphalt Shale/Shingle Room Count: 9 0 Room Types: Dining room, Family room, Laundry room, Master bath, Walk-in closet - RV Parking: No - Sauna: No - Security System: No - Skylight: Yes - Sports Court: No - Sprinkler System: No - Storage: No - Stories: 3 0 2 Structure Type: Victorian Other Tennis Court: No - Unit Count: 0 0 Vaulted Ceiling: No - Waterfront: No - Wet Bar: No - Wired: No - Zillow Home ID: 3803093 -

From [https://www.zillow.com/homes/for\\_sale/Oak-Park-IL-60302/3803093\\_zpid/84453\\_rid/globalrelvancesort\\_41.898596\\_-87.787571\\_41.887702\\_-87.805953\\_react/15\\_zm/](https://www.zillow.com/homes/for_sale/Oak-Park-IL-60302/3803093_zpid/84453_rid/globalrelvancesort_41.898596_-87.787571_41.887702_-87.805953_react/15_zm/)

Home Facts by Data Source ALL SOURCES COMBINED COUNTY RECORDS County website Beds: 3 3 Baths: 1.0 1.0 Sqft: 2,080 1,376 Lot: 4,159 sq ft / 0.10 acres 4,158 sq ft / 0.10 acres Type: Single Family Single Family Year Built: 1914 1913 Last Sold: -- - Parking: Detached Garage Garage - Detached Cooling: Central Central Heating: Gas - Fireplace: -- Days on Zillow: 104 - MLS #: 10260750 - Other facts Additional Features: Addtl Room 4 Level: Not Applicable, Addtl Room 5 Level: Not Applicable, 4th Bedroom Level: Not Applicable, Built Before 1978 (Y/N): Yes, Sewer: Sewer-Public, Listing Type: Exclusive Right To Sell, Parking Type: Garage, Tax Exemptions: Homeowner, Senior, Senior Freeze, Addtl Room 1 Level: Main Level, 2nd Bedroom Level: Main Level, Water: Lake Michigan, Kitchen Level: Main Level, Living Room Level: Main Level, Master Bedroom Level: Main Level, Other Information: School Bus Service, Commuter Bus, Addtl Room 3 Level: Main Level, 3rd Bedroom Level: Main Level, Addtl Room 2 Name: Recreation Rm, Addtl Room 1 Name: Breakfast Rm, Family Room Level: Basement, Addtl Room 9 Level: Basement, Age: 100+ Years, Addtl Room 10 Level: Not Applicable, Addtl Room 6 Level: Not Applicable, Addtl Room 7 Level: Not Applicable, Addtl Room 8 Level: Not Applicable, Addtl Room 9 Level: Not Applicable, Frequency: Not Applicable, Status: New, Square Feet Source: Plans, Master Bedroom Window Treatments (Y/N): Curtains/Drapes, 3rd Bedroom Window Treatments (Y/N): Shades, Exterior Building Type: Aluminum Siding, Addtl Room 2 Window Treatments (Y/N): Blinds, Family Room Window Treatments (Y/N): Blinds, Area Amenities: Curbs/Gutters, Street Lights, Street Paved, Sidewalks, Park/Playground, Addtl Room 3 Window Treatments (Y/N): Curtains/Drapes, Garage On-Site: Yes, 1st Floor Bedroom, 1st Floor Full Bath, Additional Rooms: Recreation Room, Breakfast Room, Sun/Florida Room Heated, Is Parking Included in Price: Yes, Addtl Room 3 Name: Sun/Florida Room Heated, Type of House 2: 1 Story, Lot Dimensions: 125 X 33, Garage Ownership: Owned, Laundry Level: Not Applicable, Aprox. Total Finished Sq Ft: 2080, Main Sq Ft: 1376, Total Sq Ft: 1376, Tax Year: 2016, Parcel Identification Number: 16053110300000 - Appliances Included: Dishwasher, Dryer, Range / Oven, Refrigerator, Washer - County Name: Cook County Cook Covered Parking Spaces: 2 2 Elementary School: Whittier Elementary School - Exterior Material: Wood Wood Floor Covering: Carpet, Hardwood, Linoleum / Vinyl - Heating Type: Forced air - High School: Oak Park & River Forest High Sch - Laundry: In Unit - Legal Description: - (HOOKERS) SUB OF NE SW SEC 05-39-13 Middle School: GWENDOLYN BROOKS MIDDLE SCHOOL - Parcel #: 1605311030 1605311030 Pets: Contact manager - Roof Type: Tile Shale/Shingle Room Count: 0 0 Room Types: Dining room - Stories: 0 0 1.0 Structure Type: Bungalow Bungalow Unit Count: 0 0 Zillow Home ID: 3799641 -

From [https://www.zillow.com/homes/for\\_sale/Oak-Park-IL-60302/3799641\\_zpid/84453\\_rid/globalrelvancesort\\_41.902253\\_-87.774485\\_41.89136\\_-87.792681\\_react/15\\_zm/](https://www.zillow.com/homes/for_sale/Oak-Park-IL-60302/3799641_zpid/84453_rid/globalrelvancesort_41.902253_-87.774485_41.89136_-87.792681_react/15_zm/)

# Humans excel at image recognition. . .

- Humans can process images very quickly
  - The three images contain several **megabytes** of data,  $p \gg n$
- Computers can process structured text very quickly
  - The raw text contains a few **kilobytes** of information
- Which system would you rather try to program?

...and many other problems computers find hard

# Artificial Intelligence (AI)

- **Artificial Intelligence**: solving problems **people find easy but computers find hard**
  - Image interpretation
  - Conversation
  - Reading emotions
  - Catching frisbees
- Skills shared with other animals having neuron-based brains
- AI is a **problem domain**



# The Premise of AI...

Appears people have gotten a long way with a **single** trick

- Recent studies suggests the mammalian brain uses a **single algorithm** to solve most tasks
- Previously machine learning research fragmented: distinct approaches for natural language processing, vision, motion planning, speech recognition

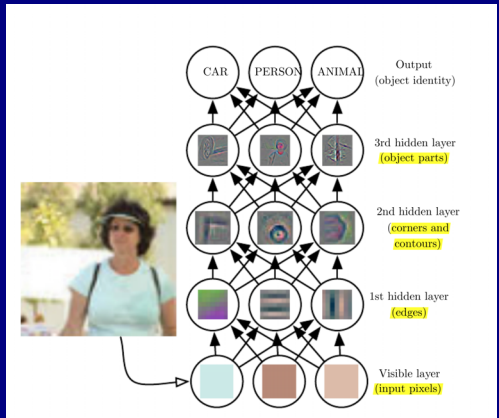
...today AI leverages a single general purpose algorithm inspired by, but not simulating, brain function to solve a wide range of problems

# AI General Purpose Algorithm

- Learn **model** from experience
  - Use machine learning
  - *Show don't tell* background information
- Learn what is **important**
  - Data **representation** created simultaneously with model
- Understand the world as a **hierarchy of concepts**
  - Complicated concepts built out of simpler ones
- Hierarchy of concepts has many layers
  - **Deep** learning

# AI General Purpose Algorithm. . .

- Hierarchy layers built using simple *neuron* nodes
- **Neural networks** build new covariates from non-linear **hinge** functions of existing covariates and the output of other layers
- Optimize overall fit with least squares



...algorithm inspired by our understanding of brain function



# AI Has a Long History...

---

## Old Approach

---

### **Knowledge base:**

hard-coded knowledge about world

**Expert system:** humans decide important features

### **Neural networks**

## New Approach

---

**Machine Learning:** extract model from training data

**Representation Learning:** determine important factors of variation

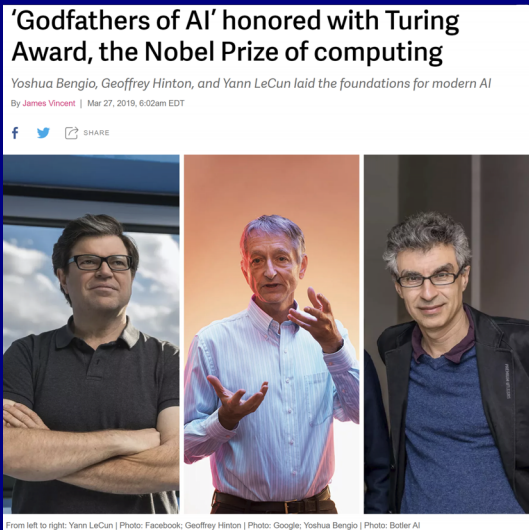
**Deep learning** neural networks

---

...but hasn't always performed as expected

# Why Is AI Working Now?

- Better algorithms
- More data
- Bigger models



# Why AI Is Working Now: **Better Algorithms**

Pre-2006 deep networks were believed to be very **difficult to train**

- Better algorithms devised by Hinton in 2006 allowed deeper models by making them faster to train
- Beginning of the deep learning era
- Tens of thousands of researchers working in AI/ML

## A fast learning algorithm for deep belief nets

[GE Hinton](#), [S Osindero](#), [YW Teh](#) - Neural computation, 2006 - MIT Press

We show how to use “complementary priors” to eliminate the explaining-away effects that make inference difficult in densely connected belief nets that have many hidden layers.

Using complementary priors, we derive a fast, greedy algorithm that can learn deep, directed ...



Cited by 10337

[Related articles](#)

[All 61 versions](#)

# Why AI Is Working Now: **Better Algorithms**

Pre-2006 deep networks were believed to be very **difficult to train**

- Better algorithms devised by Hinton in 2006 allowed deeper models by making them faster to train
- Beginning of the deep learning era
- Tens of thousands of researchers working in AI/ML

## A fast learning algorithm for deep belief nets

[GE Hinton](#), [S Osindero](#), [YW Teh](#) - Neural computation, 2006 - MIT Press

We show how to use “complementary priors” to eliminate the explaining-away effects that make inference difficult in densely connected belief nets that have many hidden layers.

Using complementary priors, we derive a fast, greedy algorithm that can learn deep, directed ...

☆ 📄 Cited by 10337 Related articles All 61 versions

## [PDF] A systematic relationship between **minimum bias** and generalized linear models

[SJ Milderhall](#) - Proceedings of the Casualty Actuarial Society, 1999 - casact.org

The minimum bias method is a natural tool to use in parameterizing classification ratemaking plans. Such plans build rates for a large, heterogeneous group of insureds using arithmetic operations to combine a small set of parameters in many different ways. Since the ...

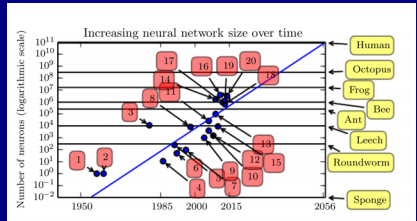
☆ 📄 Cited by 77 Related articles All 5 versions 🔗

# Why AI Is Working Now: **More Data**

- More data: IOT, Big Data, digitization of society
- **More data = less skill** required to train model
- Data rules of thumb: a supervised deep learning algorithm will generally
  - achieve **acceptable** performance with around **5,000** labeled examples per category
  - will match or **exceed human** performance when trained with a dataset containing at least **10 million** labeled examples

# Why AI Is Working Now: **Bigger Models**

- Bigger = **deeper** models
  - More computer power
  - Better algorithms
- Model size: **connections** and **number** of neurons or nodes
- Biological neurons not densely connected
  - Models within order of magnitude of mammal brains
- Number of neurons: current models are very small
  - Doubling every 2.4 years
  - **Match human brain by 2050**



...can't expect much from a leech-brain sized network

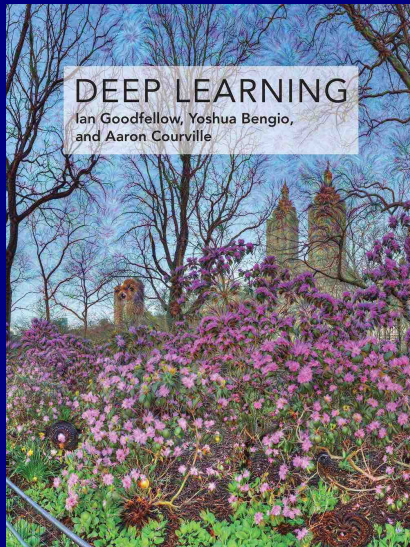
# Deep Learning Summary

**Deep learning** is an approach to AI. Specifically, it is a type of machine learning, a technique that allows computer systems to improve with experience and data.

- ... machine learning is the **only viable approach** to building AI systems that can operate in complicated, real-world environments
- Deep learning... achieves great power and flexibility by learning to represent the world as a **nested hierarchy of concepts**

Book available online for free!

11,110 GSCs.



Central Park Dreamscape CGI

# When Does AI Work Well?

## Stunning results

- Static, rules based environment
- Clean, direct observations
- Essentially limitless data
- Definitive right answer

## Examples

- Classification problems
- Image, speech recognition
- Pedestrian detection
- **Traffic sign classification**
- Simple dynamic control
- Reinforcement learning, robotics
- Synthetic data, alphaZero
- GANs



# When Does AI Work Well?

## Stunning results

- Static, rules based environment
- Clean, direct observations
- Essentially limitless data
- Definitive right answer

## Examples

- Classification problems
- Image, speech recognition
- Pedestrian detection
- **Traffic sign classification**
- Simple dynamic control
- Reinforcement learning, robotics
- Synthetic data, alphaZero
- GANs

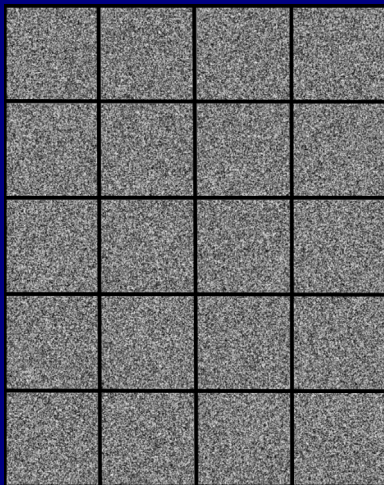
## Characteristics of insurance

- Behavioral feedbacks
- Dynamic: reacts as we learn
- Proxy data, not direct
- **Inherently small** classes
- Uncertain data: development
- Latent data

...AI works well on  
certain types of  
real world data

# How is *Real Data* Special?

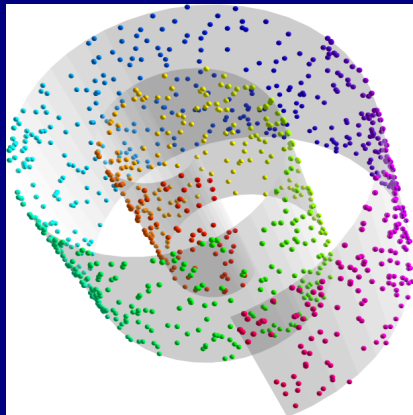
- Real data often appears very high dimensional ( $p \gg n$ ) but really contains **hidden structure**
- Photographs do not look like random images
  - Continuous: image near a photograph recognizable
  - No jumps or cliffs
  - Not fractal-like



Real images occupy a **negligible proportion** of the total volume of image space

# How is *Real Data* Special?

- Real data often appears very high dimensional ( $p \gg n$ ) but really contains **hidden structure**
- Photographs do not look like random images
  - Continuous: image near a photograph recognizable
  - No jumps or cliffs
  - Not fractal-like



Real images occupy a **negligible proportion** of the total volume of image space

# How is *Real Data* Special?

## Helpful AI data **rules of thumb**

### Representation learning: A review and new perspectives

[Y Bengio](#), [A Courville](#), [P Vincent](#) - IEEE transactions on pattern ... , 2013 - [ieeexplore.ieee.org](#)

The success of machine learning algorithms generally depends on data representation, and we hypothesize that this is because different representations can entangle and hide more or less the different explanatory factors of variation behind the data. Although specific domain knowledge can be used to help design representations, learning with generic priors can also be used, and the quest for AI is motivating the design of more powerful representation-learning algorithms implementing such priors. This paper reviews recent work in the area of ...

☆ 77 Cited by 4662 Related articles All 40 versions

- Smooth
- Common material drivers
- Hierarchical organization
- Low dimensional
- Sparse
- Temporal and causal coherence

...properties often fail for insurance data!

# Insurance-Related AI Applications

**Good fit** general business

- Marketing
- Customer service
- Underwriting  
homogeneous classes
- Fraud detection

# Insurance-Related AI Applications

## **Good fit** general business

- Marketing
- Customer service
- Underwriting  
homogeneous classes
- Fraud detection

## **Potential?**

- Claims process  
management, reserves
- Data evaluation,  
anomalies, completion
- Risk management in  
high frequency lines
- Underwriting accounts:  
pricing or ranking?

# Insurance-Related AI Applications

## **Good fit** general business

- Marketing
- Customer service
- Underwriting  
homogeneous classes
- Fraud detection

## **Potential?**

- Claims process  
management, reserves
- Data evaluation,  
anomalies, completion
- Risk management in  
high frequency lines
- Underwriting accounts:  
pricing or ranking?

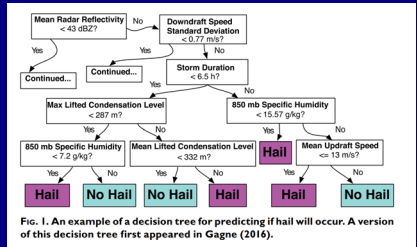
## **More exposure to see rarer events**

- Good performance = **5000** observations
- 50,000 exposures at 10% claim rate
- 500,000 exposures at 1% claim rate
- 5 million exposures at 0.1% claim rate
  
- For approx. 1% claim rate see examples  
on [r/CatastrophicFailure](#), [r/IdiotsInCars](#),  
[r/nononono](#), [r/OSHA](#), [r/Roadcam](#)

**AI needs lots of data...**  
**...can you collect enough?**

# AI Compared to Catastrophe and Weather Models

- Is enough data **practically possible**?
- Weather and catastrophe models are physics-based models
- Physical models better than a learned phenomenological models
- Re-learning physics inefficient
- AI can work effectively in **conjunction** with physics-based models



...AI already being applied to forecast high impact weather events



# AI is Not a Universal Panacea—No Singularity

---

Problems	Computers find Easy	Computers find Hard
<b>People find Easy</b>	Clustering Consistency Inconsistency Outliers	Reading emotion Conversation Unstructured interaction Physical interactions
<b>People find Hard</b>	Big data Complex computations Simulations Dispassionate	<b>Underwriting</b> Investing Middle East Peace Plan Identifying fake news Gödelization, <i>thinking outside the box</i>

---

# AI Bias and Sources of Modeling Failure

Statisticians know training data must be a **random** cross section of population

- **Extend** tools to include new characteristic picked up by ML
  - Computers do not see images with semantic understanding
  - Computers see things the eye cannot detect: lighting, type of camera
  - Computers don't filter irrelevant information
- **Enhance** skills for the AI modeling era
  - Enhanced methodological rigor **managing training data**
  - Technical tools to analyze and diagnose the **behavior** of the model
  - Training, education and caution in the **deployment** of ML in products

# AI Bias and Sources of Modeling Failure

- Model-in-use can fail for three reasons
  - Algorithm
  - Implementation
  - Bureaucracy
- Understanding decisions in big organizations is very complex
  - Like ML models, people complex and opaque

...AI is just a machine—think washing machine  
not HAL9000—powerful but limited

Benedict Evans

# Three Things to Remember...

- AI = **solving problems people find easy but computers find hard**
- Spectacular progress: bigger models; better algorithms; more data
  - Show don't tell
  - Derive data **representation** and model **form**
- AI relies on **special characteristics of data**: applied to non-AI data it can disappoint

Thank you!

# References

---

Slide(s)	Reference
3,4,10,11	<a href="https://www.zillow.com">https://www.zillow.com</a>
13-16, 21-23,	Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, (2016).
25-26	<a href="https://deeplearningbook.org">https://deeplearningbook.org</a>
18	Godfathers of AI, The Verge 2019, <a href="https://www.theverge.com/2019/3/27/18280665/ai-godfathers-turing-award-2018-yoshua-bengio-geoffrey-hinton-yann-lecun">https://www.theverge.com/2019/3/27/18280665/ai-godfathers-turing-award-2018-yoshua-bengio-geoffrey-hinton-yann-lecun</a>
19, 26	Google Scholar
25	Swiss Roll: <a href="http://www1.osu.cz/~zusmanovich/talks/habil-pres/src/test-swiss-roll.png">http://www1.osu.cz/~zusmanovich/talks/habil-pres/src/test-swiss-roll.png</a>
26	Bengio, Yoshua, Aaron Courville, and Pascal Vincent. "Representation learning: A review and new perspectives." IEEE transactions on pattern analysis and machine intelligence 35.8 (2013): 1798-1828. <a href="https://arxiv.org/pdf/1206.5538.pdf">https://arxiv.org/pdf/1206.5538.pdf</a>
28	McGovern, Amy, et al. "Using artificial intelligence to improve real-time decision-making for high-impact weather." Bulletin of the American Meteorological Society 98.10 (2017): 2073-2090. <a href="https://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-16-0123.1">https://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-16-0123.1</a>
30-31	Benedict Evans, Notes on AI Bias (2019) <a href="https://www.ben-evans.com/benedictevans/2019/4/15/notes-on-ai-bias">https://www.ben-evans.com/benedictevans/2019/4/15/notes-on-ai-bias</a>

---

```
## AI talk buzzword bingo image, Python Code
from imageio import imread
import matplotlib.pyplot as plt
from wordcloud import WordCloud
import numpy as np
```

```
args = {'width':3600, 'height':2400, 'max_words': 200,
        'max_font_size':240, 'relative_scaling' : 0.4,
        'background_color' : 'white', 'prefer_horizontal':0.75}
```

```
words=['Artificial Intelligence', 'Big Data',
        'Causal Analysis', 'Data Representation', 'Data Visualization',
        'Deep Learning', 'Data Representation', 'Data Science'
        'Expert System', 'Exploratory Data Analysis', 'Inference',
        'Machine Learning', 'Neural Network', 'Predictive Analytics',
        'p > n', 'Statistics']
```

```
word_freqs = { w: (0.4 + 0.6 * np.random.rand()) for w in words}
word_freqs['Artificial Intelligence'] = 2.5
word_freqs['Deep Learning'] = 1.5
word_freqs['Neural Network'] = 1.5
```

```
wordcloud = WordCloud(colormap=plt.cm.magma, **args)
```

```
for i in range(30):
    wordcloud.generate_from_frequencies(word_freqs)
    plt.figure(figsize=(11,8.5))
    plt.imshow(wordcloud)
    plt.axis("off")
    plt.savefig(f'word_cloud_{i}.png', dpi=300)
    plt.show()
```