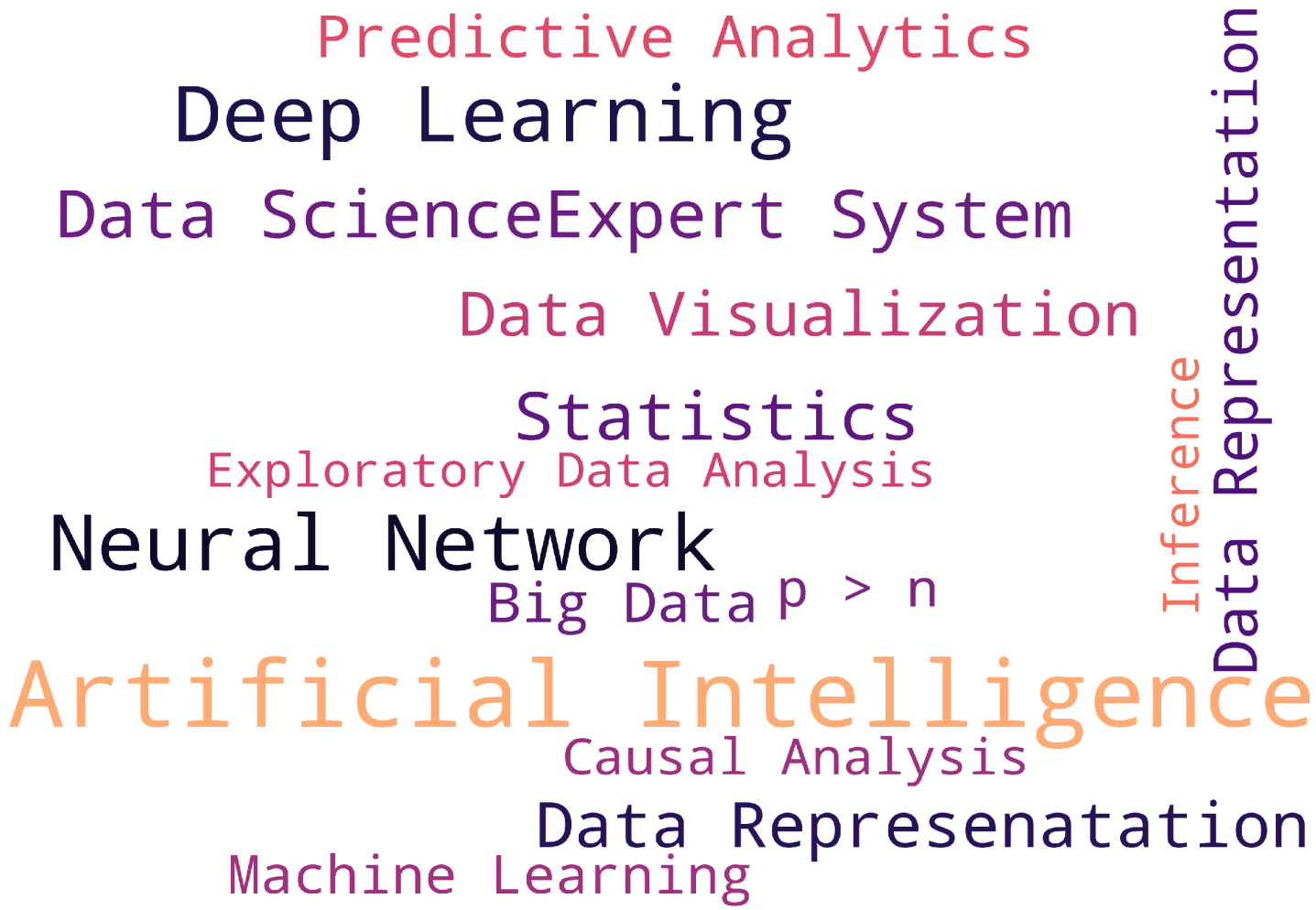
The Promise, Pitfalls and Process of AI

Stephen J. Mildenhall

May 20-21, 2019

ST. JOHN'S UNIVERSITY
Tobin College of Business
School of Risk Management


Predictive Analytics
Deep Learning
Data ScienceExpert System
Data Visualization
Statistics
Exploratory Data Analysis
Neural Network
Big Data
p > n
Artificial Intelligence
Causal Analysis
Data Represenatation
Machine Learning
Inference
Data Representation


1006 Woodbine Ave
Oak Park, Illinois
Google
Street View - Oct 2018
Lindberg Park
Google


1006 Woodbine Ave
Oak Park, Illinois
Google
Street View - Oct 2018
Lindberg Park
Google

House Buying: How Much Should You Offer?

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

Fireplace

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

- **Descriptive** statistics summarize with no interpretation
- **Exploratory** data analysis (EDA): only makes claims about current sample

Extending Comparables

- **Predictive** modeling or analytics: estimating value for a **new** unit
- Buyer's perspective
- Leverage limited data

Understanding Comparables

- Seller's perspective: should I invest to remodel the kitchen or paint or landscape or make life easier for my broker?
- **Inferential** statistics

Machine Learning (ML)

- Statistics is learning from data
- **Machine learning** is an algorithm allowing computer to extract and improve model from raw data

Machine Learning (ML)

- Statistics is learning from data
- **Machine learning** is an algorithm allowing computer to extract and improve model from raw data
- Predictive and inferential statistics are machine learning
- Logistic regression, decision trees, GLMs. . . what you use everyday
- Generally an expert system or knowledge base is **not** ML
- Big question: form of input data?
- Exam MAS material

Machine Learning (ML)

- Statistics is learning from data
- **Machine learning** is an algorithm allowing computer to extract and improve model from raw data
- Predictive and inferential statistics are machine learning
- Logistic regression, decision trees, GLMs. . . what you use everyday
- Generally an expert system or knowledge base is **not** ML
- Big question: form of input data?
- Exam MAS material

Examples of ML Problems

- Classification (with missing inputs)
- Regression, prediction
- Transciption (image to text)
- Machine translation
- Anomaly detection (fraud, spam)
- Synthesis (speech, synthetic landscape)
- Impute missing values
- Denoise
- Density estimation

Rank Three Houses in Same Town by Market Value

Rank Three Houses in Same Town by Market Value

Rank Three Houses in Same Town by Market Value

Rank Three Houses in Same Town 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 39,621 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 #: 10322977 – 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, 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, Sidewalks, 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 2.0 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/globalrelevanceex_sort/41.897302,-87.808603,41.886408,-87.826799_rect/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 13,562 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, Additional Sales Information: 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: 7 2 Deck: No – Disability Access: No – Dock: No – Doorman: No – Double Pane/Storm Windows: No – Elementary School: Oliver W Holmes Elementary Schoo – 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: – (WORWICKS) SUB OF LT1 EX N7FT & ALL OF LT2 IN BLK2 IN KETTLESTRINGS ADD TO HARLEM SUB OF PT OF NW SEC 07-39-13 Middle School: GWENDOLYN BROOKS MIDDLE SCHOOL – Mother-in-Law Apartment: No – Near Transportation: No – Over 55 Active Community: No – Parcel #: 1607103016 1607103016 Patio: Yes – Pets: Contact manager – Pond: No – Pool: Yes – Porch: Yes – Roof Type: Asphalt Shake/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 2.0 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/globalrelevanceex_sort/41.898596,-87.787757,41.887702,-87.805953_rect/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 2 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 Shake/Shingle Room Count: – 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/globalrelevanceex_sort/41.902253,-87.774485,41.89136,-87.792681_rect/15_zm/? "'

Humans Excel at Image Recognition...

- Humans: very fast image processing
  - The three images contain several **megabytes** of data, $n \ll p$
- Computers: very fast text processing
  - The raw text contains a few **thousand** bytes 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 recognition
  - Conversation, spoken words
  - Driving
  - Reading emotions
- Skills shared with other animals having neuron-based brains

- AI is a **problem domain**

The Premise of AI

People have gotten a long way with a single algorithm

- Studies suggests that much of the mammalian brain uses a **single algorithm** to solve most tasks
- Before this hypothesis, machine learning research was more fragmented, with different communities of researchers studying natural language processing, vision, motion planning and speech recognition
- Deep learning taps this single, powerful general problem solving paradigm
- Deep learning is inspired by brain function, not simulating brain function

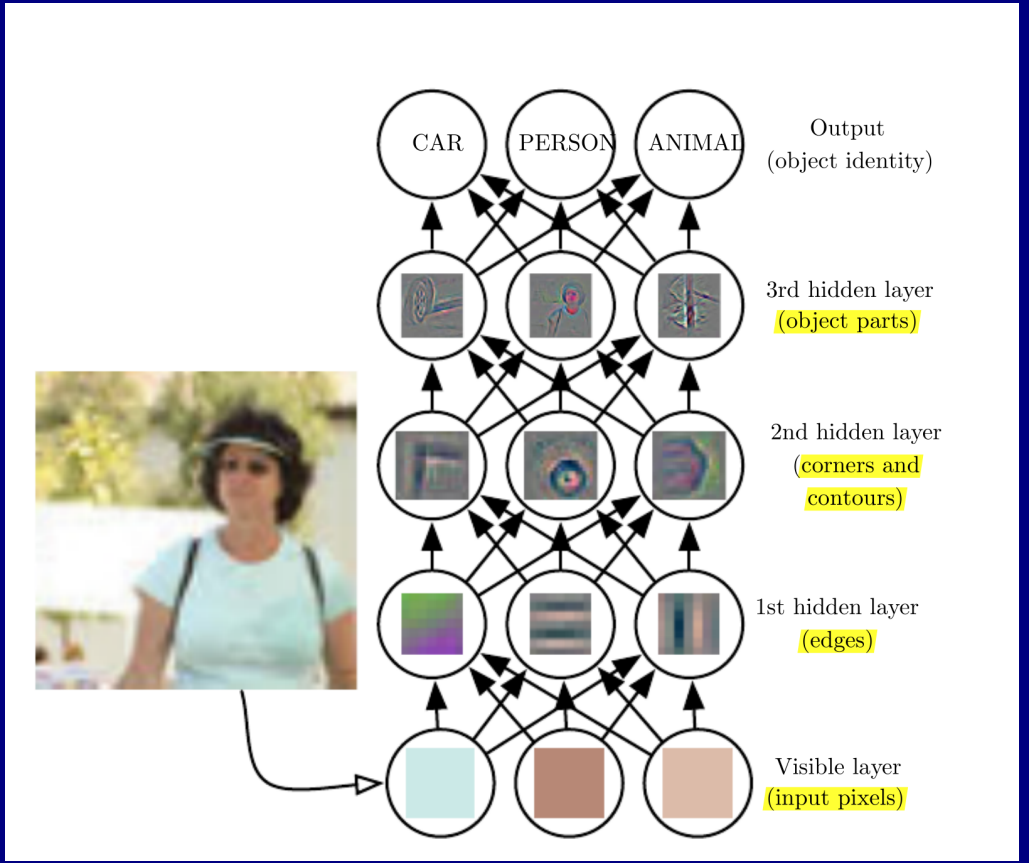
Current AI Assumptions

- Learn **model** from experience
  - Use machine learning
  - Show don't tell the background needed for everyday life
- Understand the world as a **hierarchy of concepts**
  - Learn complicated concepts by building out of simpler ones
- Learn what is important
  - Learn the **data representation** and the model
- Graph of concepts generally has many layers
  - Hence **deep** learning

Miracle of AI: the same general purpose tool works for a wide range of problems

Current AI Tools. . .

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



...follow our understanding of brain function

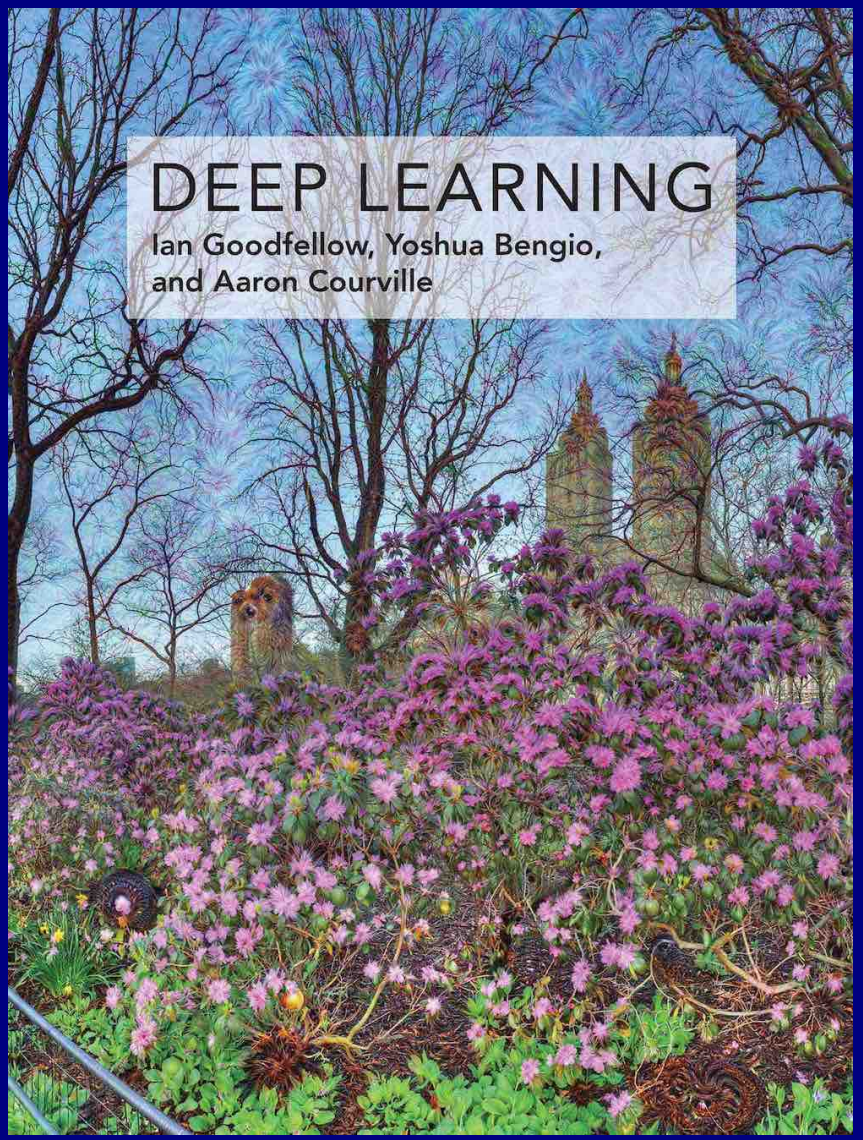
Current AI Tools: The Naïve View

- Neural networks and deep learning are fancy least squares and gradient descent = walking downhill
  - Stochastic gradient descent = probably walking downhill
- Build complex functions from **compositions of simple ones**
- Built specialized problem solving widgets
  - recognize color
  - recognize car vs truck vs bike vs motorcycle vs person
  - can combine to recognize red car, blue truck, ...
- Pool, share and feedback information between widgets
- Widgets learn model and appropriate data representation

Deep Learning

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**



AI has a Long History...

Old Approach	New Approach
Knowledge base: hard code knowledge about world	**Machine Learning**: extract model from
Expert system: humans decide important features	**Representation Learning**: determine important factors of variation
Neural networks	Deeper **neural networks** trained with more data

...but hasn't always performed as expected

Why Is AI Working Now?

- Better algorithms
- More data
- Bigger models

‘Godfathers of AI’ honored with Turing Award, the Nobel Prize of computing

Yoshua Bengio, Geoffrey Hinton, and Yann LeCun laid the foundations for modern AI

By James Vincent | Mar 27, 2019, 6:02am EDT

SHARE

From left to right: Yann LeCun | Photo: Facebook; Geoffrey Hinton | Photo: Google; Yoshua Bengio | Photo: Botler AI

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 …

☆ 99 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 …
☆ 99 Cited by 10337 Related articles All 61 versions

Premium calculation by transforming the layer premium density
S **Wang** - ASTIN Bulletin: The Journal of the IAA, 1996 - cambridge.org
This paper examines a class of premium functionals which are (i) comonotonic additive and (ii) stochastic dominance preservative. The representation for this class is a transformation of the decumulative distribution function. It has close connections with the recent developments …
★ 99 Cited by 706 Related articles All 8 versions

Why AI Is Working Now: More Data

- More data: IOT, Big Data, digitization of society
- **More data = less skill** required to train model
- Goodfellow 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 models enabled by greater computational power and better algorithms
- Model size/depth: **connections** and **number** of neurons or nodes
- Biological neurons not densely connected
  - Models within order of magnitude of mammal brains connections
- Number of neurons: current models are very small
  - Doubling every 2.4 years
  - Will not match human brain until **2050**

Can't expect much from a leech-brain sized network

When does AI Work Well?

Stunning results when

- Static, rules based environment
- Clean, direct observations
- Essentially limitless data
- Definitive right answer
- Classification problems
- Simple dynamic control

For example

- Image, speech recognition
- Pedestrian detection
- Traffic sign classification
- Reinforcement learning, robotics
- GANs

When does AI Work Well?

Stunning results when

- Static, rules based environment
- Clean, direct observations
- Essentially limitless data
- Definitive right answer
- Classification problems
- Simple dynamic control

For example

- Image, speech recognition
- Pedestrian detection
- Traffic sign classification
- Reinforcement learning, robotics
- GANs

Characteristics of insurance

- Behavioral feedbacks
- Dynamic: reacts as we learn
- Proxy data, not direct
- Granularity drives small classes
- Uncertain information: claim development, trend
- Latency: asbestos and environmental
- Most of the time not good enough

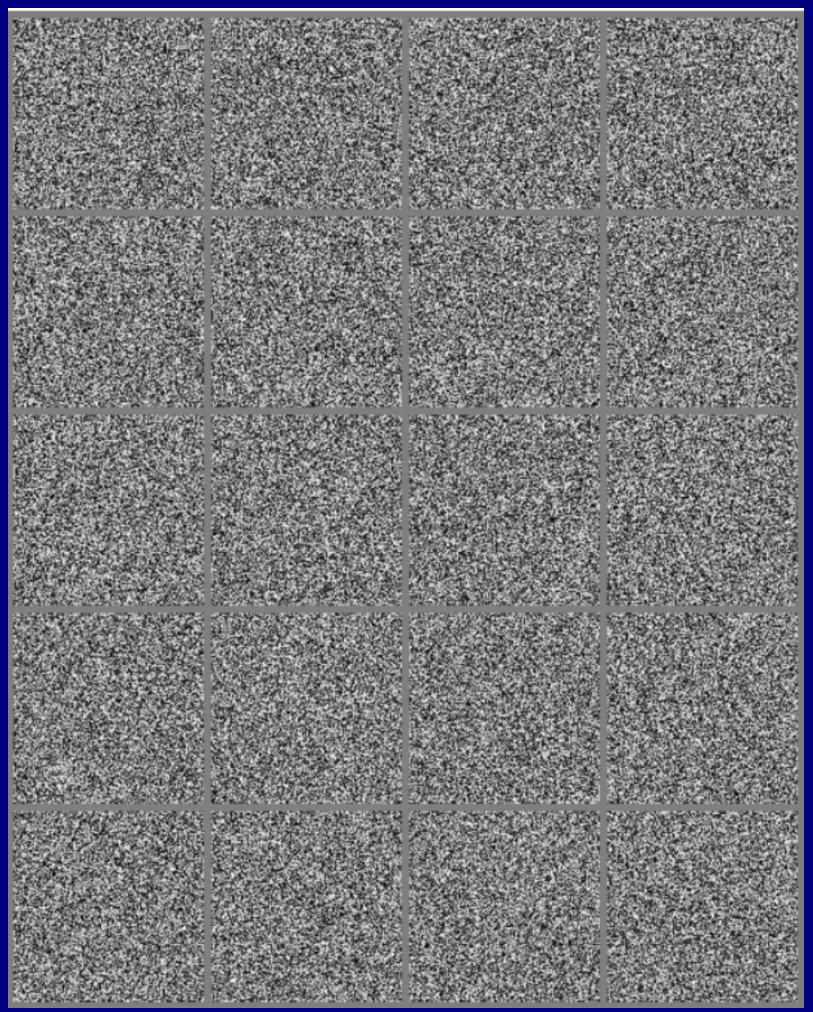
Real World Data

- **No Free Lunch** Theorem for ML
  *Averaged over **all** possible data generating distributions,* **every** *classification algorithm has the* **same error rate** *on new data*
- No machine learning algorithm is universally any better than any other
- The most sophisticated algorithm has the same average performance as predicting every point belongs to the same class
- Despite grim prognosis ML finds rules that are probably correct most of the time

ML builds tools that work well on real data

AI Assumptions About *Real Data*

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



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

AI Assumptions About *Real Data*

Helpful **general purpose priors** or 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 …

☆ 99 Cited by 4662 Related articles All 40 versions

AI Assumptions About *Real Data*

Assumption	**Concerns**
Smoothness: functions are smooth	Weather, butterfly effect; singularities
Multiple explanatory factors	Sudden change in relationship
Common material drivers	Market crash
Shared factors across tasks	
Hierarchical organization of factors	
Simple factor dependencies, linear	Log-linear: earthquakes
Manifolds: real data is low d	
Natural clustering	
Sparsity, 80/20 explanatory rule	
Temporal and spatial coherence	
Causal factors	State dependent causality
Most of the time	. . . there are no claims

AI Applications: Everything

Insurance-Related Applications

Good Fit—Already Live

- Marketing
- Customer service
- Underwriting homogeneous classes = predictive modeling
- Fraud detection
- Claims processing: identify complex claims
- Healthcare

Potentially Good Fit

- Claims reserving, see Brian...
- Loss reserving
- Data quality
- Anomaly detection
- Data completion
- Underwriting **accounts** = detecting risk-aware management
- Risk management in high frequency lines

More Challenging for AI

Challenge Data Rules of Thumb

- Discontinuities
- State change, sentiment change
- Singularities: tornados, black holes
- Weather model
- Cat model
- Identify next mass tort

The Limits of Predictability

- Frequency = **Trials** × **Probability**
- Materiality race between Trials and Probability
- More trials lowers observable probability threshold = Big Data helping
- r/CatastrophicFailure, r/IdiotsInCars, r/nononono, r/OSHA, r/Roadcam

AI Compared to Catastrophe and Weather Models

- Frequency = **Trials** × **Probability** again. . . given enough data. . .
- Weather and catastrophe models are physics-based models
- AI works most effectively **in conjunction** with physics-based models to avoid re-learning physics: a physical model is better than a learned phenomenological model
- AI already being applied to high impact weather forecasting

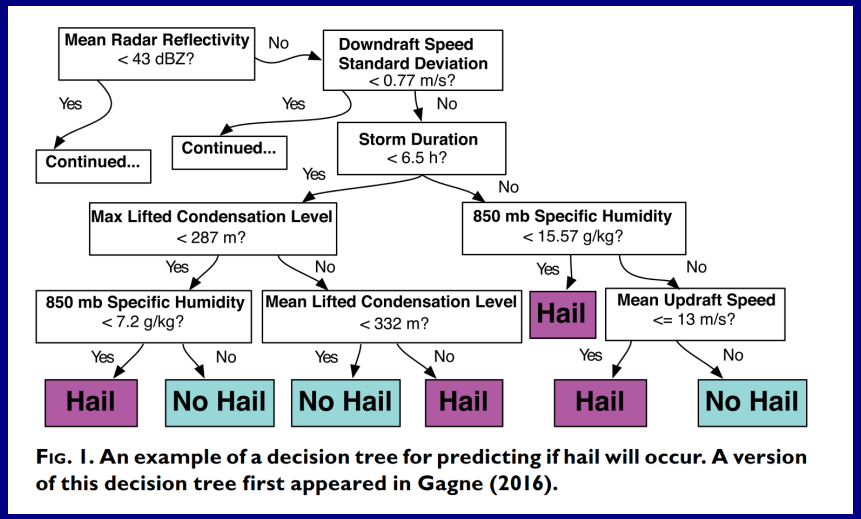
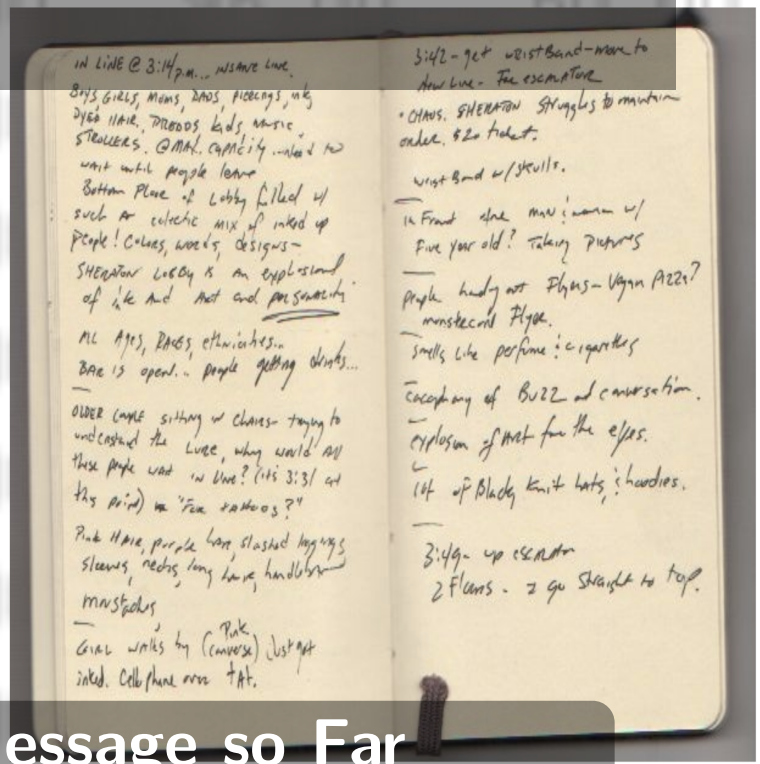


FIG. 1. An example of a decision tree for predicting if hail will occur. A version of this decision tree first appeared in Gagne (2016).

“Raw data is both an oxymoron and a bad idea; to the contrary, data should be cooked with care.”

Geoffrey Bowker

```
0000000 0000 0001 0001 1010 0010 0001 0004 0128
0000010 0000 0016 0000 0028 0000 0010 0000 0020
0000020 0000 0001 0004 0000 0000 0000 0000 0000
0000030 0000 0000 0000 0010 0000 0000 0000 0204
0000040 0004 8384 0084 c7c8 00c8 4748 0048 e8e9
0000050 00e9 6a69 0069 a8a9 00a9 2828 0028 fdfc
0000060 00fc 1819 0019 9898 0098 d9d8 00d8 5857
0000070 0057 7b7a 007a bab9 00b9 3a3c 003c 8888
0000080 8888 8888 8888 288e be88 8888 8888
0000090 3b83 5788 8888 8888 7667 778e 8828 8888
00000a0 d61f 7abd 8818 8888 467c 585f 8814 8188
00000b0 8b06 e8f7 88aa 8388 8b3b 88f3 88bd e988
00000c0 8a18 880c e841 c988 b328 6871 688e 958b
00000d0 a948 5862 5884 7e81 3788 1ab4 5a84 3eec
00000e0 3d86 dcb8 5cbb 8888 8888 8888 8888 8888
00000f0 8888 8888 8888 8888 8888 8888 8888 0000
0000100 0000 0000 0000 0000 0000 0000 0000 0000
*
0000130 0000 0000 0000 0000 0000 0000 0000
000013e
```



Message so Far

- Show don’t tell
- Give data and results
- Derive data representation
- Derive model relationships

AI Bias and Sources of Modeling Failure

- Computers do not see images with semantic understanding
- Computers "see" things we cannot detect, lighting, type of camera and are not attuned to
- Computers don't filter "irrelevant" information, background
- Basic statistics: Training data must include random cross section of population
  - Extended to include non-human characteristic picked up by ML
- Solutions
  1. Methodological rigor in the collection and management of the training data
  2. Technical tools to analyze and diagnose the behavior of the model
  3. Training, education and caution in the deployment of ML in products

AI Bias and Sources of Modeling Failure

- Problem domain
  - Technical problem inherent with algorithm?
  - Problem with implementation?
  - Institutional failure by large bureaucracy?
- Understanding decisions in big organization very complex
  - Like evaluating ML models "people are black boxes too"
- AI is just a machine—think washing machine not HAL9000—powerful but limited

Cheat Sheet

Term	Definition
Statistics	Learning from Data
Descriptive Statistics	Facts, no interpretation
EDA	Current sample only, not generalization
Predictive modeling	Estimate value for a new unit, (MAS I and II)
Inferential statistics	Understand importance of explanatory data over population
Artificial Intelligence	Problem domain; solving problems people find easy but computers find hard
Machine Learning	Algorithm allowing computer to extract and improve model from raw data
Knowledge base	Human curated explanation of complex real-world system
Expert System	Rules based explanation of complex real-world system
Neural network	Multi-layered, inter-connected hinge function regression model, loosely modeled on brain function
Deep learning	Approach to AI; type of ML; learn complicated concepts by building out of many simpler ones
Representation Learning	Use ML to process figure data representation

References

Slide	Reference
	https://www.zillow.com
	Godfathers of AI, The Verge 2019, https://www.theverge.com/2019/3/27/18280665/ai-godfathers-turing-award-2018-yoshua-bengio-geoffrey-hinton-yann-lecun
	Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, (2016). https://deeplearningbook.org
	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. https://arxiv.org/pdf/1206.5538.pdf
	Ben Evans, Notes on AI Bias (2019) https://www.ben-evans.com/benedictevans/2019/4/15/notes-on-ai-bias
	Anna Munster, Raw data, pure data, databodies, and data worlds (2017) https://prezi.com/ktz-4hw2qbly/raw-data-pure-data-databodies-and-data-worlds/
	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. https://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-16-0123.1

```
## May 2019 for CAS talk AI buzzword bingo image
from imageio import imread
import matplotlib.pyplot as plt
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
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 Represenatation', '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()
```