

An Overview of
Data and AI Applications

Ikhlaq Sidhu, content author

Basic Concept of Working with Data



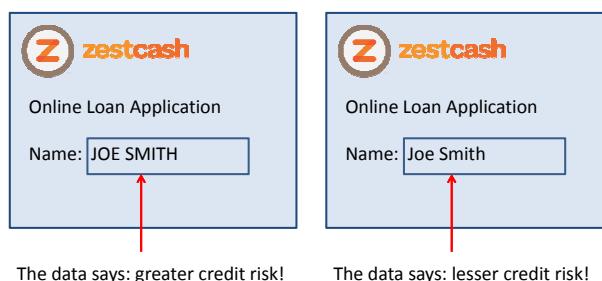
- Data Wrangling
- In Production

Ikhaaq Sidhu, content author

Example: Data and Information is a competitive advantage

Real-life Example: ZestCash

- “All data is credit data”



Reference: Shomit Ghose



Harrah's Casino: Knowing your customer

Reference: *Supercrunchers*

PLAY & WIN ▶

- Service provider of Gambling and Casinos
- Entry Card
- Pain points
- Intervention

Why: More Simply

Customer Insight/ Engagement	Operations: Reliable & Predictable	Security & Fraud
 TARGET	 	Compliance 360° Financial Firms Network Security
Harrah's		

Who Will Control the Automobile?



- Google? or Ford?
 - Whoever has the better software and data science team
 - Winner will get the vast (and incredibly valuable) streams of auto data

Shomit Ghose

Ikhaaq Sidhu, content author

The Core of Every Business in Future is Data and AI

The two key components of a business are resources (assets) and information (data)



Less value
over time

More value
Over time

If you buy data, then everyone else has it also.

Ikhaaq Sidhu, content author

Data and AI Approaches

AI, Machine Learning, and Data Science

- What is Machine Learning, Data Science, and AI
- Today's technology in Industry

Ikhsaq Sidhu, content author

Implementation: SW Tools / Stack

Ikhsaq Sidhu, content author

The Most Common Open Source Tools: AI/ML Stack

Start with Python as an interface

Jupyter Notebooks for prototyping

- Python: The interface
- NumPy, SciPy: Working with Arrays
- Pandas: Working in Tables, SQL to Pandas
- Sklearn: ML
- Matplotlib: Visualizing Data
- TensorFlow, Keras: Neural Networks
- SQL to Pandas
- NLP / NLTK: Natural Language
- Spark: For large data sets (GB, TB+)



<https://www.youtube.com/watch?v=Q0jGAZAdZqM>

<https://conda.io/docs/user-guide/install/download.html>

Ikhaaq Sidhu, content author

API ML Tools

Use your own data to train models

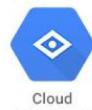


TensorFlow



Cloud Machine Learning Engine

Ready to use Machine Learning models



Cloud Vision API



Cloud Speech API



Cloud Jobs API



Cloud Translation API



Cloud Natural Language API



Cloud Video Intelligence API



Coming soon



Other (big) data services



Dialogflow



Cloud Dataproc

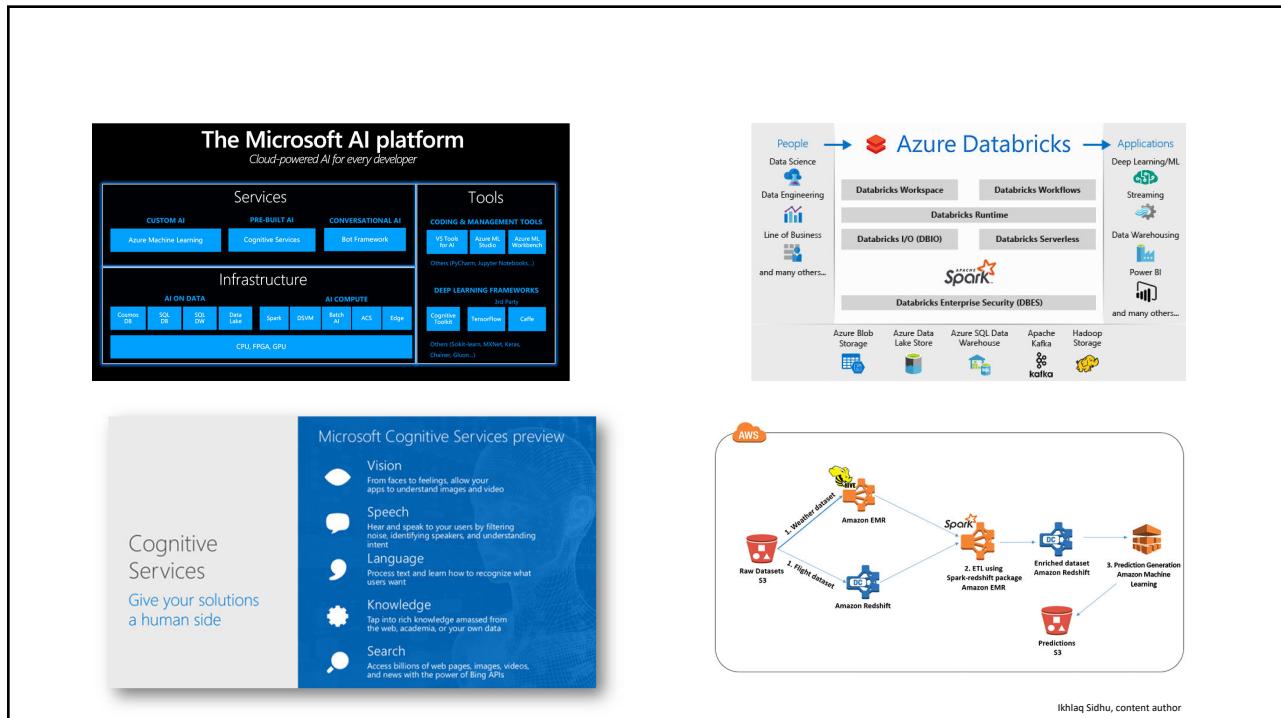


Cloud Dataflow



BigQuery

Ikhaaq Sidhu, content author



Where Does Data Come From?

Ikhlaq Sidhu, content author

Where Does Data Come From?

Real-life Example: ZestCash

- "All data is credit data"



The data says: greater credit risk!



The data says: lesser credit risk!

Public datasets on AWS

To enable more innovation, AWS hosts a selection of datasets that anyone can access for free. Data in our public datasets is available for rapid access to our flexible and low-cost computing resources.



Your Own Web Site

Public Data Sets
Stock market, etc.

IOT/Sensors

Other Web Sites

Ikhaq Sidhu, content author

Web Scraping

```

1  from bs4 import BeautifulSoup
2  import requests
3  page_link ='https://www.website_to_crawl.com'
4  # fetch the content from url
5  page_response = requests.get(page_link, timeout=5)
6  # parse html
7  page_content = BeautifulSoup(page_response.content, "html.parser")
8
9  # extract all html elements where price is stored
10 prices = page_content.find_all(class_='main_price')
11 # prices has a form:
12 #[<div class="main_price">Price: $66.68</div>,
13 # <div class="main_price">Price: $56.68</div>]
14
15 # you can also access the main_price class by specifying the tag of the class
16 prices = page_content.find_all('div', attrs={'class':'main_price'})
```

<https://github.com/ikhlaqsidhu/data-x>

https://github.com/ikhlaqsidhu/data-x/tree/master/03-tools-webscraping-crawling_api_afo

Web Scraping

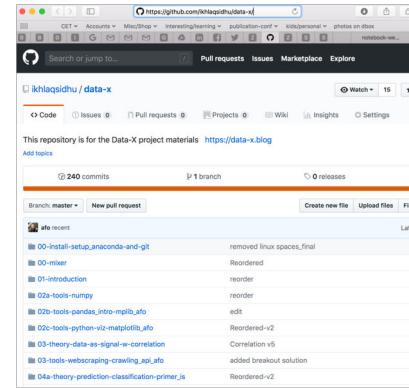
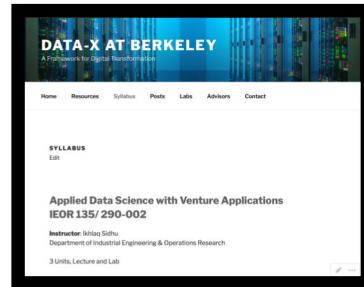


Extract data from any website

Ikhaq Sidhu, content author

Many Course Resources Are Already Available at For students and mentors

- Lectures and Slides
- Code Samples
- Articles and Readings
- Projects
- Mentors



data-x.blog

<https://github.com/ikhlaqsidhu/data-x/>

IkhlAQ Sidhu, content author

Formatting Data

IkhlAQ Sidhu, content author

An ML High Level Framework

- In Real Life**
- Objects
 - Events / Experiments
 - People / Customers
 - Products
 - Stocks
 - ...

Features, but also loss of information

Sex	Age	Market	Occupation	Job Title	Checking	Savings	Deöffent Mark
Female	27.17	Married	Semi-professional	0 No	Yes	Good	
Male	25.92	Married	Blue Collar	0.375 No	Yes	Good	
Male	23.08	Married	Blue Collar	1 No	Yes	Good	
Male	39.58	Married	Semi-professional	0 No	Yes	Good	
Male	30.85	Married	Blue Collar	0.125 No	Yes	Good	
Male	17.25	Married	Blue Collar	0.4 No	No	Good	
Female	16.5	Single	Semi-professional	0.165 No	No	Good	
Male	27.33	Married	Semi-professional	0 No	No	Good	
Male	31.25	Married	Semi-professional	0 No	Yes	Good	
Male	39.5	Married	Blue Collar	1.5 No	No	Good	
Male	36.5	Married	Blue Collar	3.5 No	No	Good	
Male	52.42	Married	Blue Collar	3.75 No	No	Good	

Copyright © Plug&Score

Person 1
Person 2
Person 3
...
Person N

In Sample Out of Sample

Some data has observed results

- Characteristics
- Patterns
- Models
- Predictions
- Similarities
- Differences
- Distance

Ikhaq Sidhu, content author

- In Real Life**
- Objects
 - Events / Experiments
 - People / Customers
 - Products
 - Stocks
 - ...

Features, but also loss of information

Sex	Age	Market	Occupation	Job Title	Checking	Savings	Deöffent Mark
Female	27.17	Married	Semi-professional	0 No	Yes	Good	
Male	25.92	Married	Blue Collar	0.375 No	Yes	Good	
Male	23.08	Married	Blue Collar	1 No	Yes	Good	
Male	39.58	Married	Semi-professional	0 No	Yes	Good	
Male	30.85	Married	Blue Collar	0.125 No	Yes	Good	
Male	17.25	Married	Blue Collar	0.4 No	No	Good	
Female	16.5	Single	Semi-professional	0.165 No	No	Good	
Male	27.33	Married	Semi-professional	0 No	No	Good	
Male	31.25	Married	Semi-professional	0 No	Yes	Good	
Male	39.5	Married	Blue Collar	1.5 No	No	Good	
Male	36.5	Married	Blue Collar	3.5 No	No	Good	
Male	52.42	Married	Blue Collar	3.75 No	No	Good	

Copyright © Plug&Score

Person 1
Person 2
Person 3
...
Person N

In Sample Out of Sample

Some data has observed results

- Characteristics
- Patterns
- Models
- Predictions
- Similarities
- Differences
- Distance

$$X = \begin{bmatrix} -2 & 4 & 7 & 31 \\ 6 & 9 & 12 & 6 \\ 12 & 11 & 0 & 1 \\ 9 & 10 & 2 & 3 \end{bmatrix}$$

Ikhaq Sidhu, content author

A Fundamental Idea: From Table to Score

$X =$

The diagram illustrates a transformation from a matrix X to a vector Y through a function $F(X)$. On the left, a matrix X is shown with columns labeled **Cust**, **F1**, **F2**, and **F3**. The rows are labeled A, B, C, D, E, F, and .. (ellipsis). An arrow labeled $F(X)$ points from the matrix X to the vector Y on the right. The vector Y has two columns: **Cust** and **Credit Score**. The data points correspond to the rows of X : A (552), B (381), C (760), D (330), E (452), F (678), and .. (..).

Cust	F1	F2	F3
A	4	2	2
B	4.5	1.5	3
C	3	3	5
D	1	2	2
E	3	1.5	5
F	3.5	3.5	1
..

X

Cust	Credit Score
A	552
B	381
C	760
D	330
E	452
F	678
..	..

Y

Ikhaq Sidhu, content author

A Fundamental Idea: From Table to Score

$X =$

The diagram illustrates a transformation from a matrix X to a vector Y through a function $F(X)$. On the left, a matrix X is shown with columns labeled **Cust**, **F1**, **F2**, and **F3**. The rows are labeled A, B, C, D, E, F, and .. (ellipsis). An arrow labeled $F(X)$ points from the matrix X to the vector Y on the right. The vector Y has two columns: **Cust** and **Credit Score**. The data points correspond to the rows of X : A (552), B (381), C (760), D (330), E (452), F (678), and .. (..).

Cust	F1	F2	F3
A	4	2	2
B	4.5	1.5	3
C	3	3	5
D	1	2	2
E	3	1.5	5
F	3.5	3.5	1
..

X

Cust	Credit Score
A	552
B	381
C	760
D	330
E	452
F	678
..	..

Y

```
#Setting up for Supervised learning
# First clean: use mapping + buckets

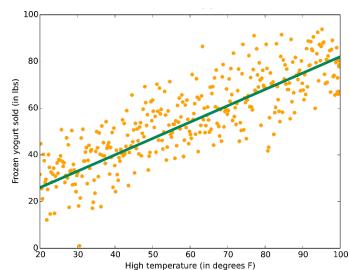
# X = matrix of data - e.g 1000 rows
# Y = In sample responses

# Typically we want to split in to training data and test data

X_train = X[0:500]
Y_train = Y[0:500]
X_test = X[501:1000]
Y_test = Y[501:1000]
```

Ikhaq Sidhu, content author

Linear Regression Illustration



```
#Setting Linear Regression in sklearn
from sklearn import linear_model

model= linear_model.LinearRegression()
model.fit(X_train, Y_train)

Y_pred_train = model.predict(X_train)
Y_pred_test = model.predict(X_test)

# Compare Y_pred_test with Y_test for
error.
```

Illustration Source: <https://docs.microsoft.com/en-us/azure/machine-learning/machine-learning-algorithm-choice>

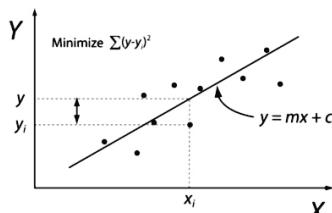
Ikhaq Sidhu, content author

Prediction

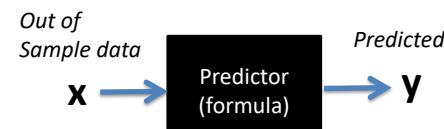
Data We Might Have
(In Sample)

X	Y
2	3
5	9
6	11
8	?
10	?
?	?

Data View



Math View



Our Goal: Working with
out of sample data

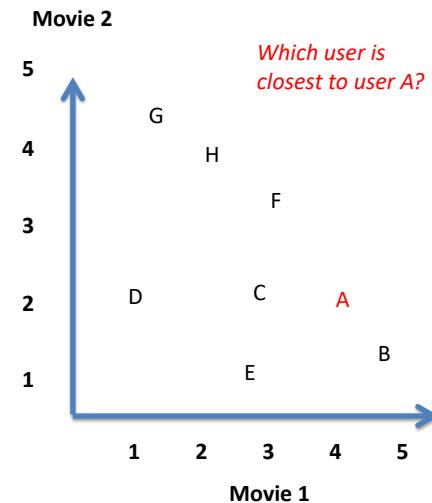
Systems View

Ikhaq Sidhu, content author

A Fundamental Idea: From Table to N- Dimensional Space

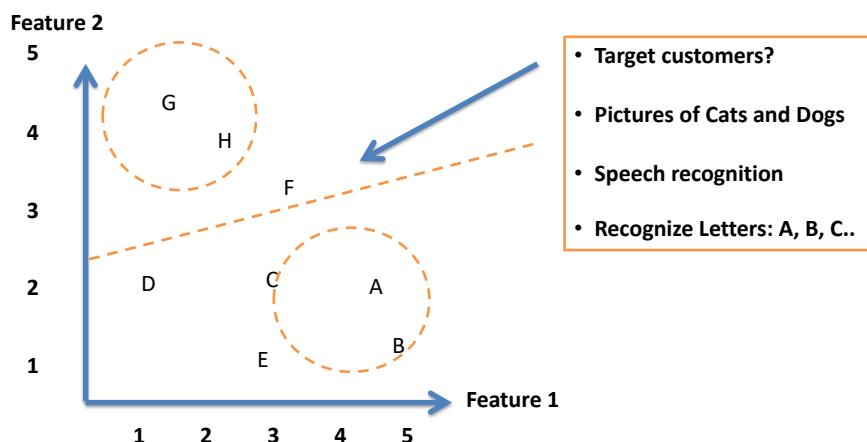
Element	F1	F2	F3
A	4	2	2
B	4.5	1.5	3
C	3	3	5
D	1	2	2
E	3	1.5	5
F	3.5	3.5	1
..

X =



Ikhaq Sidhu, content author

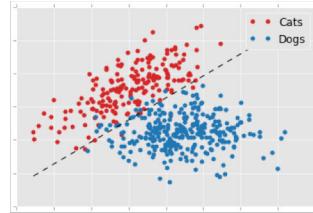
Clustering to Classification



Ikhaq Sidhu, content author

Traditionally 2 Tasks: Classification & Predictive Scoring

Extracted Data
often in
Table
Format



Classification:
Cats and Dogs, Speech Recognition
Movie Recommendation



The most famous
application has been
recommendation:
“which other user is
most like you”



Scoring:
Credit Score, Movie Rating
Health Score, Any Isoquant...

Ikhaq Sidhu, content author

We have now switched
to Neural Networks as
Function Approximators

X

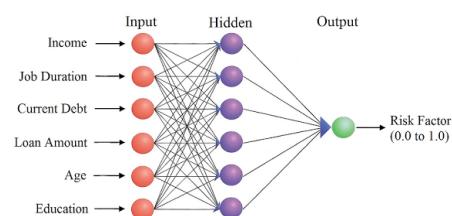


ML Algorithms Guess
this function $F(x)$

Y

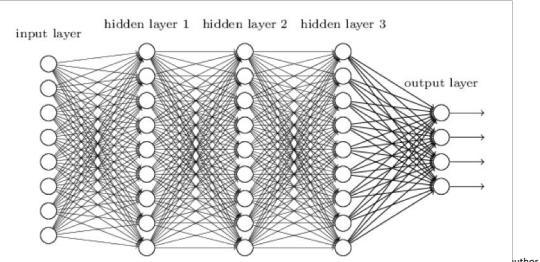
"Non-deep" feedforward
neural network

X



Deep neural network

X



Y

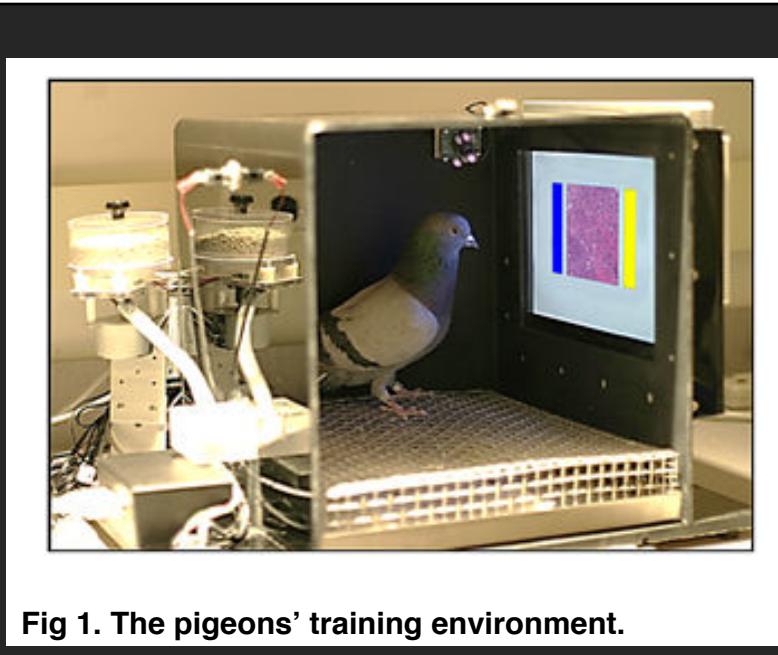
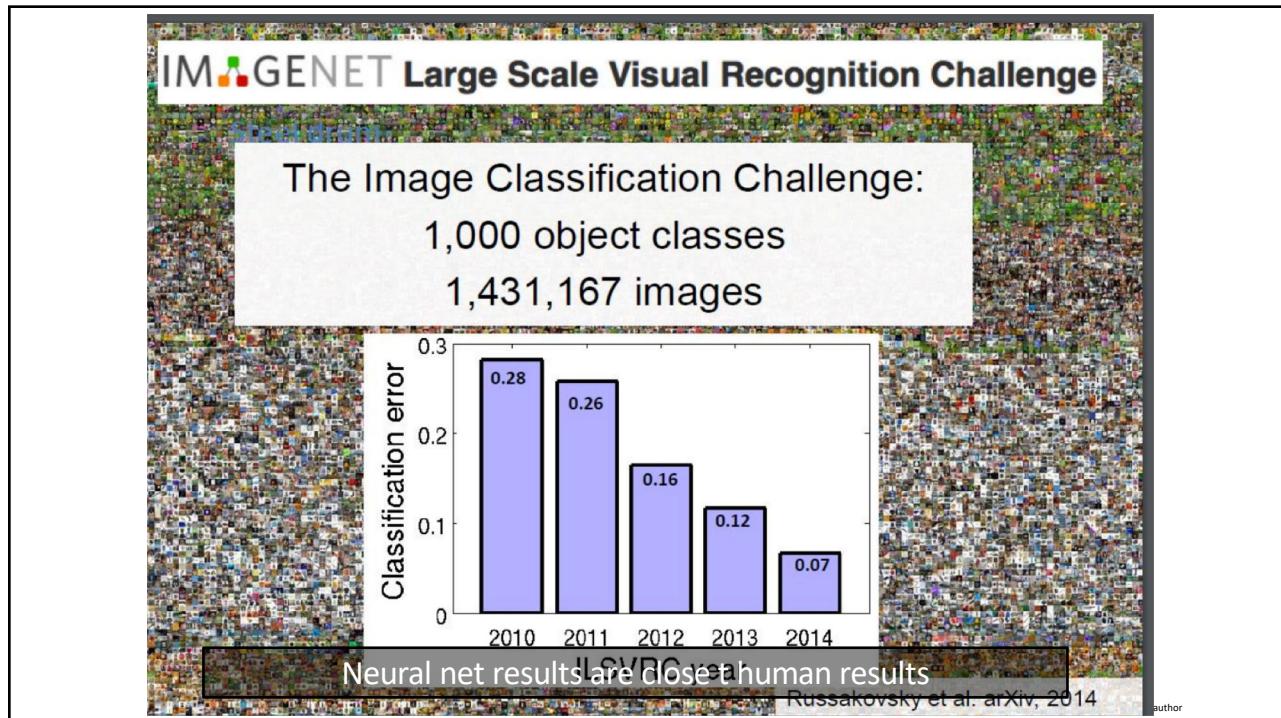


Fig 1. The pigeons' training environment.

Ikhlaq Sidhu, content author

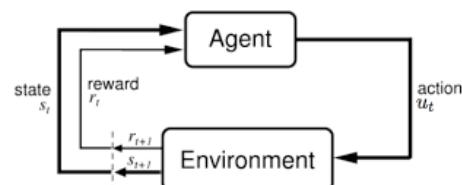
Data and AI Future Directions

Ikhsaq Sidhu, content author

Peter Abbeel – Deep Reinforcement Learning



Peter Abbeel
Professor at UC Berkeley



Ikhsaq Sidhu, content author

Deep Reinforcement learning Locomotion

Robot models in physics simulator (MuJoCo, from Emo Todorov)

Neural network architecture:

Input: joint angles and velocities
Output: joint torques

Diagram illustrating the neural network architecture:

```

graph LR
    Input[Input: joint angles and velocities] --> Layer1[Fully connected layer]
    Layer1 --> Layer2[Fully connected layer]
    Layer2 --> Layer3[Fully connected layer]
    Layer3 --> Sampling[Sampling]
    Sampling --> Output[Output: joint torques]
    
```

Companies mentioned in the slide footer: ANKI, BOSCH, Boston Dynamics, LIVIN, Google, Yandex.

Ikhlaq Sidhu, content author

Recent AI News

KBA

Google DeepMind Challenge Match
8–15 March 2016

AlphaGo

AlphaGo Lee Sedol

Source: Ken Goldberg, CPAR, People and Robotics Initiative

Ikhlaq Sidhu, content author

Does this mean AI Can Do Everything Better than Humans

Ikhaq Sidhu, content author

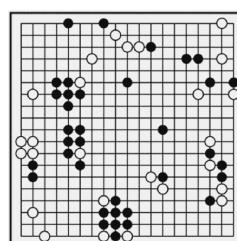
Even then, AI Cannot Solve Real Life Problems Better Than Humans And in fact, AI Can not even Work without Humans



Ken Goldberg
Leading AI
Researcher at
Berkeley

Professor and
Department Chair,
IEOR

William S. Floyd Jr.
Distinguished Chair



discrete single agent
fully observed finite



continuous multi-agent
uncertain infinite time horizon

Ken Goldberg UC Berkeley

Ikhaq Sidhu, content author

AI Systems Only Work because of Human are Part of the System

Google Operations

Massive Data →

Result →

Feedback By clicks ←

People Write Web Pages

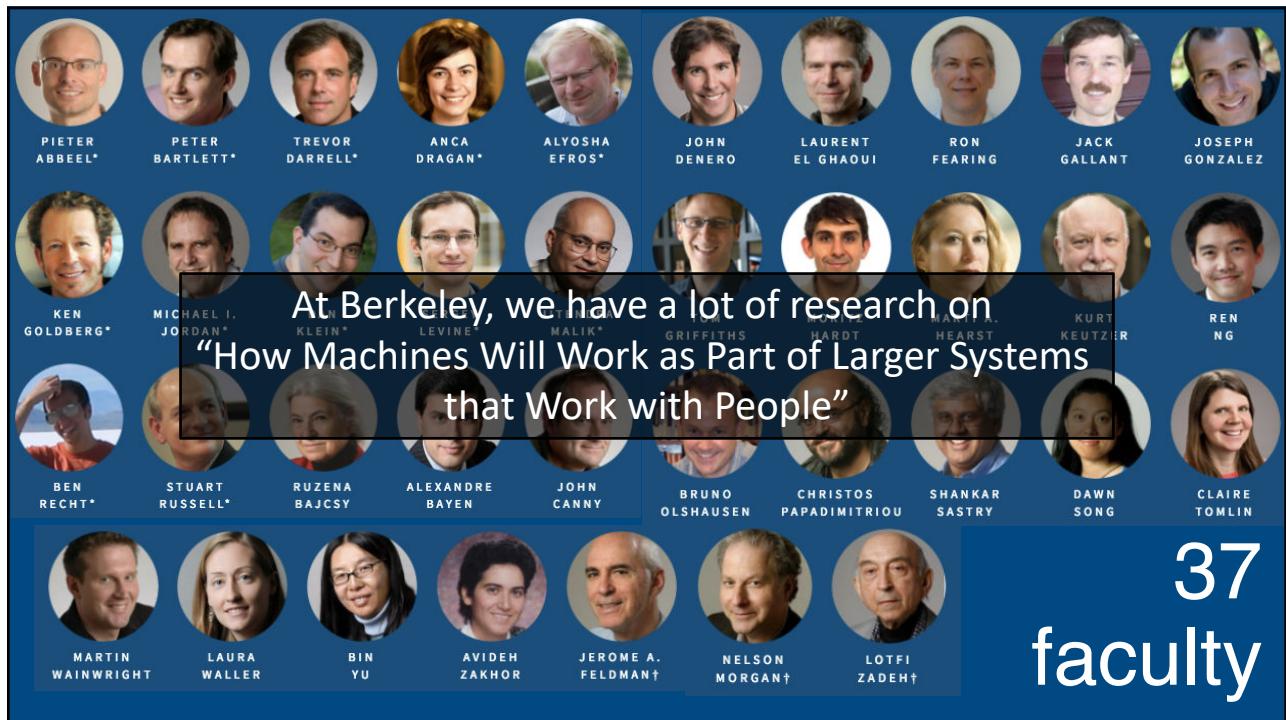
People at Google Tune the Results

People Click on What They Want

There is no “Intelligence”, “Desire”, or “Existence” in AI without People
There are only people who “invest in, design and operate the machines”

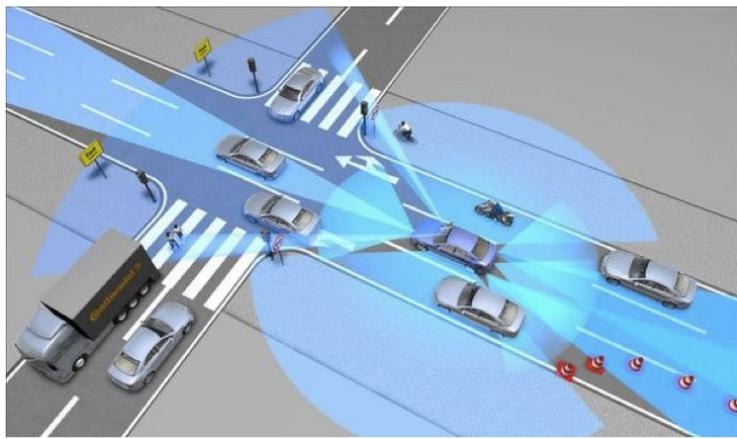
Acknowledgement to Ken Goldberg UC Berkeley

Ikhaq Sidhu, content author





Autonomous Driving and Driver-Assist



- Communicating intent
- Driver-in-the-loop modeling
- Two-way learning: knowledge transfer between vehicle and driver
- Safety in autonomous and assisted driving

Principal investigators:



Trevor Darrell
UC Berkeley



Anca Dragan
UC Berkeley



Ken Goldberg
UC Berkeley



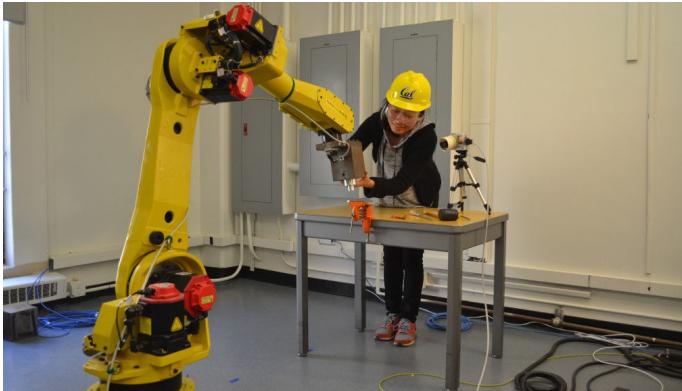
Ruzena Bajcsy
UC Berkeley



Francesco Borrelli
UC Berkeley

Source: Ken Goldberg, CPAR, People and Robotics Initiative

Safety in Human-Robot Interaction: Guarantees and Verification



Safety-constrained motion planning for efficiency in factory human-robot interaction

Learning and prediction for safety in HRI

Provably safe human-centric autonomy

Principal investigators:



Claire Tomlin
UC Berkeley



Masayoshi Tomizuka
UC Berkeley



Francesco Borrelli
UC Berkeley

Source: Ken Goldberg, CPAR, People-and-Robotics Initiative

Most Common Data/AI Research Trends in 2017

- Large-scale machine learning - amounts of data
- Deep learning - recognition, classification
- Reinforcement learning - time sequence, aided by Neural Networks
- Robotics - beyond navigation, to safe interaction
- Computer vision - most prominent perception, better than human
- Natural Language Processing - interacting with people/dialog
- Collaborative systems - autonomous systems w/people + machines using complimentary functions
- Crowdsourcing and human computation – harness human intelligence, uses other AI, vision, ML, NLP, ...
- Algorithmic game theory and computational social choice – systems using social computing, incentives, prediction markets, game theory, peer prediction, scoring rules, no regret learning
- Internet of Things (IoT) – using AI to unravel sensory information, interfaces, and protocols
- Neuromorphic Computing – new computing fabrics based on biological models

New
Data/AI
Systems

Semantics for AI

AI Umbrella

II: Intelligent Infrastructure

A web on computation, data, and physical entities that make the human environment more supportive, interesting, & safe.

IA: Intelligent Automation

Computation and data are used to create services that augment human intelligence and creativity

Michael Jordan, UC Berkeley

Ikhiq Sidhu, content author

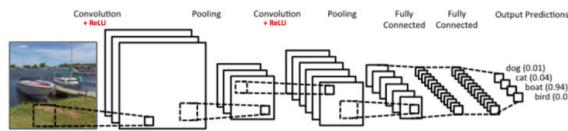
Unsupervised Image to Image



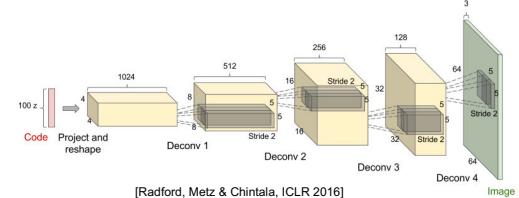
[CycleGAN: Zhu, Park, Isola & Efros, 2017]

Pieter Abbeel – UC Berkeley | Gradescope | Covariant.AI

Typical CNN converts image to output vector of features



- Ability to generate data *that look real* entails some form of understanding



[Radford, Metz & Chintala, ICLR 2016]

Ikhiq Sidhu, content author

IEEE SPECTRUM

Experts Bet on First Deepfakes Political Scandal

Researchers wager on a possible Deepfake video scandal during the 2018 U.S. midterm elections

By Jeremy Hsu

Ikhaq Sidhu, content author

Large Investment and Valuations

FINANCE • SELF-DRIVING CARS

GM Buying Self-Driving Tech Startup for More Than \$1 Billion

China Now Has the Most Valuable AI Startup in the World

Published: April 8, 2018, 6:00 PM EDT

It becomes the world's most valuable private AI startup.
The company drives China's ambition to dominate global AI.

SenseTime valued at over \$3 billion

This Chinese company is the most valuable AI startup in the world. Photo: AP

ORACLE
Intelligent Finance:
How CFOs Can Lead the Coming Disruption
+ GET REPORT

"My goal is to recreate a European sovereignty in AI."

INTERVIEW WITH WIRED
THURSDAY, MARCH 29TH, 2018

'Whoever leads in AI will rule the world': Putin to Russian children on Knowledge Day

Published time: 1 Sept, 2017 14:08
Edited by: Alexey Druzhin / Sputnik
Gut short URL
Russian President Vladimir Putin & Alexey Druzhin / Sputnik

Vladimir Putin spoke with students about science in an open lesson on September 1, the start of the school year in Russia. He said that the future belongs to those who lead in artificial intelligence, and that it will rule the world. "Whoever leads in AI will rule the world," he told his audience in Moscow, but not himself. "If comes with some risks, but also involves that are difficult to predict. Whoever becomes the leader in this sphere will become the ruler of the world," Russian President Vladimir Putin said.

However, the president said he would not like to see anyone "monopolize" the field.

If we become leaders in this area, we will choose that technology with entire society. The same way we chose our nuclear programme today," he told students from across Russia via satellite link-ups speaking from the Yaroslavl region.

Beijing to build \$2 billion AI research park: Xinhua

Reuters Staff
2 MIN READ
Beijing (Reuters) - Beijing is planning to build a 13.8 billion yuan (\$2.12 billion) artificial intelligence development park in the city's west, the official Xinhua news agency reported, as China pushes ahead to fulfill its ambition to become a world leader in AI by 2025.

The AI park will house up to 400 enterprises and have an estimated annual output of 50 billion yuan, Xinhua said, citing a report from authorities in Beijing's Mentougou district.

Ikhaq Sidhu, content author

Contact:

Ikhlaq Sidhu

Founding Faculty Director, Center for Entrepreneurship & Technology

IEOR Emerging Area Professor, UC Berkeley

[sidhu @berkeley.edu](mailto:sidhu@berkeley.edu), scet.berkeley.edu

Ikhlaq Sidhu, content author