



# Data-X Master Class

A Framework for Harnessing Data, AI, and Intelligent Systems for Transformative Innovation

Ikhlaq Sidhu  
Chief Scientist & Founding Director  
IEOR Emerging Area Professor Award  
Sutardja Center

Alexander Fred-Ojala  
Co-creator and Lecturer, Data-X  
Research Director, Data Lab  
Sutardja Center

# Opening Comments and Introductions

Day 1:  
Executive Level  
Topics

Day 2+: Mostly  
Technical



Ikhlaq Sidhu, content author



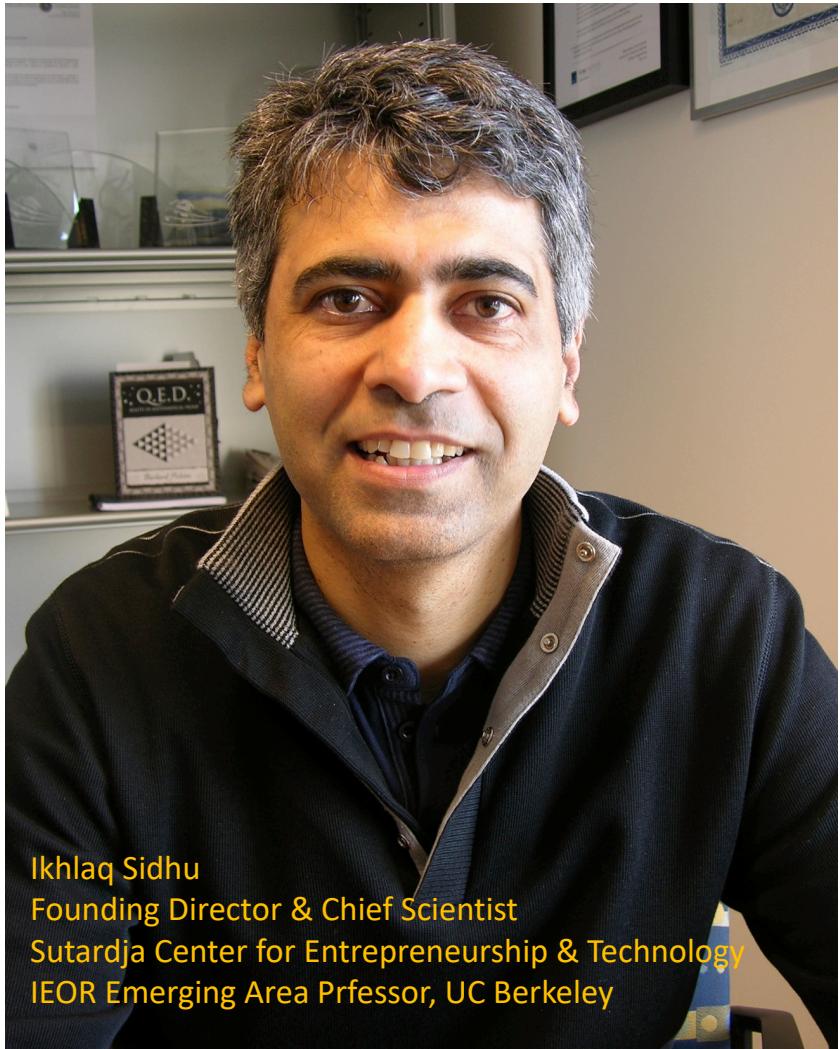
**HELLO  
MY NAME IS**

# Alexander Fred-Ojala

- **Research Director**  
Data Lab, SCET, UC Berkeley
- **Co-creator of Data-X**  
UC Berkeley class: Applied Data Science w Venture Applications
- **Co-founder**  
UC Berkeley / SCET's Blockchain lab
- **Founding Team of 3 companies**  
InnoQuant (COO), Auranest (CMO),  
Wheely's (YCombinator alumni)
- **Degree in Mathematical Statistics**  
UC Berkeley & Lund University, Sweden



Co-founder Blockchain-X



Ikhlaq Sidhu  
Founding Director & Chief Scientist  
Sutardja Center for Entrepreneurship & Technology  
IEOR Emerging Area Professor, UC Berkeley

## About Me: Ikhlaq Sidhu

- Electrical Engineering & HW and SW Design – 7 years, resulted in HP Laserjet design
- Applied Researcher in Networking & Communications – 7 years, Founded the Advanced Technologies Research Center (USR/3Com), 75 patents and Inventor of Year at 3Com Corp
- Led Development of 3 products lines - including VC Venture.
- Professor & Director, Entrepreneurship & Technology Innovation – 12+ years. SCET and Many spin outs ...
- Today: Perspective as engineer, innovator, executive, and academic => Life lessons that I can share

# SUTARDJA CENTER AT BERKELEY

## METRICS AT A GLANCE

### Undergraduate:

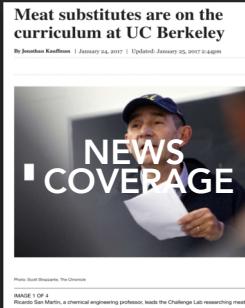
- 12-14 Courses, 1500+ Undergraduates

### X at Berkeley: Graduate, Labs and Professional

- 80+ Grad students
- 100+ Executives
- Labs: Data-X, Blockchain, Sustainable Food

### Ecosystem:

- 14+ Global Partners
- 500+ Executives
- 50+ Investors



# IEOR 135/290 Applied Data Science with Venture Applications

Supported by the Data-X Lab

**Course Number:** IEOR 135/290  
**Section:** 33036  
**Instructor:** Ikhlaq Sidhu  
**Offered:** Spring 2024  
**Units:** 3 (Lecture and Discussion)

**Course Description:** This course provides an understanding at the intersection of **Computer Science Tools** and **Statistical Concepts** and their **current computational applications** to real world problems. The course will be presented in application-oriented modules organized as Collect, Combine, and Visualize. Applications of this course include industry sectors such as finance, health, engineering, transportation, energy, and many others.

**Computer Science Tools =** Python pandas, Python NumPy, Python SciPy, SQL, R

**Statistical Concepts =** Descriptive Statistics, Probability, Inference, Transforms, Prediction

Please complete this form if you're interested in receiving additional information about this course.

Ikhlaq Sidhu, Author

QR code

# Data-X: IEOR 135/290 Applied Data Science with Venture Applications by the Numbers

- Developed over two years
- 300+ alumni students
- 100+ industry experts in network
- 80+ great projects completed
- 6+ published research papers
- 10+ students got hired because of Data-X
- Amazing testimonials:

*I think this class is so awesome because it teaches the tools and concepts that are most commonly used in workplace teams that are involved with data science and applied machine learning.*

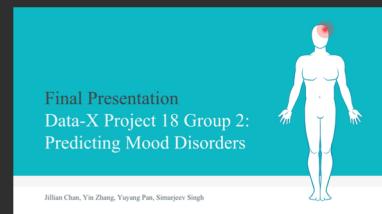
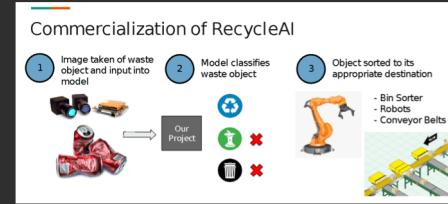


# IEOR 135/290 Applied Data Science with Venture Applications

## Sample Data-X Projects

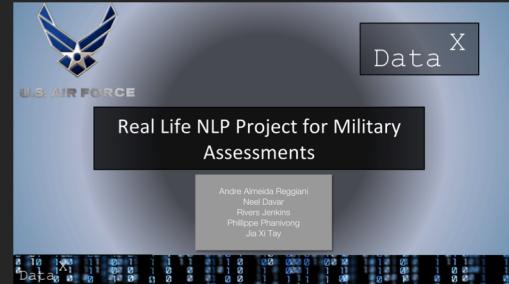
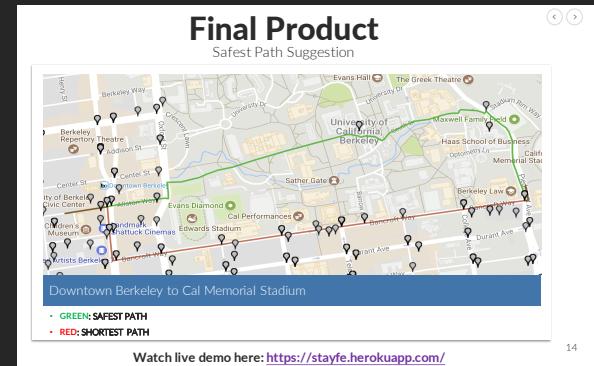
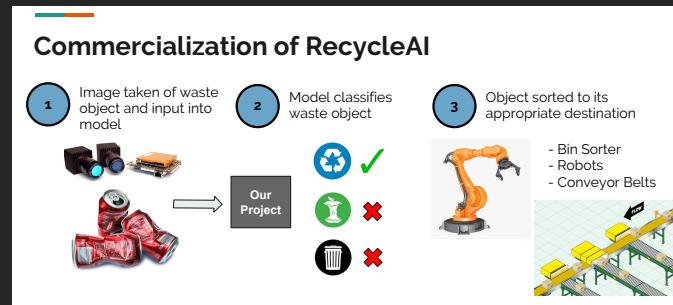
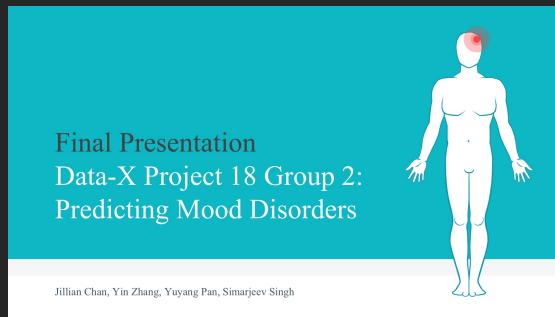
- Detection of fake news
- Prediction of long-term energy prices to solve Wall Street problem
- Prediction applications for stock market, sports betting, and more
- AI for crime detection, traffic guidance, medical diagnostics, etc.
- A version of Zillow that is recalculated with the effects of AirBnB income

and many many more...



# DATA-X PROJECT EXAMPLES

• • •



# Data-X: Applied Data Science with Venture Applications



Make the Tools

Most CS / Math



Use the Tools  
(Optimally)

This Course



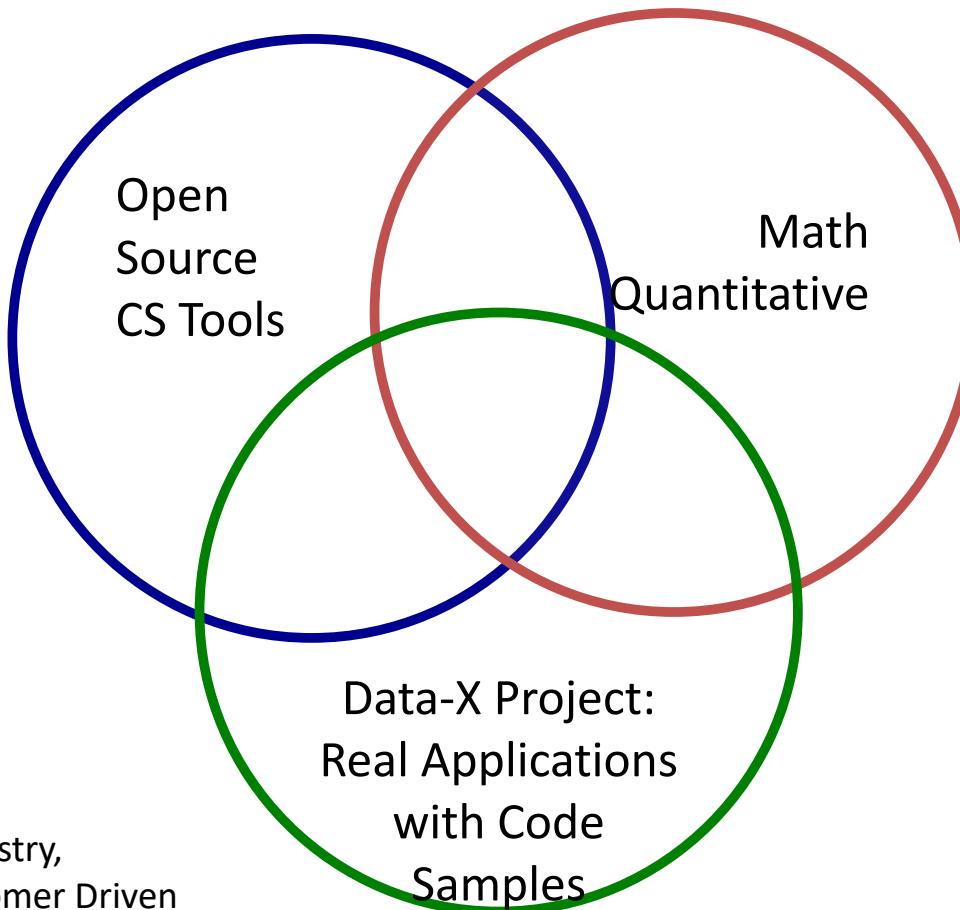
Architect  
the System



Why and how  
you build

Ikhlaq Sidhu, content author

# What is in the full Data-X course at Berkeley?



Holistic Perspective: Industry,  
Social Applications, Customer Driven

# What is in the full Data-X course at Berkeley?

## Common Open Source CS Tools:

- Numpy, SciPy
- Pandas
- TensorFlow, Sklearn
- SQL to Pandas
- NLP / NLTK
- Matplotlib

Often: Working Code First  
Fill In Theory After

## Quantitative

- Prediction: Regression
- ML Classification: Logistic, SVM.. Trees, Forests, Bagging, Boosting,..
- Entropy / Information Topics
- Deep Learning examples, including CCNs
- Correlations
- Markov Processes
- LTI Systems: Fourier, Filters where applicable
- Control Models where applicable

## Building Block Code Samples

- Webscraping
- Stock market live download, simple trading
- Convolutional Neural Networks
- Next Word Predictor, Spell Checking
- Recommendation
- Web Crawler
- Chatbot, E-mail
- Social net interfaces including twitter

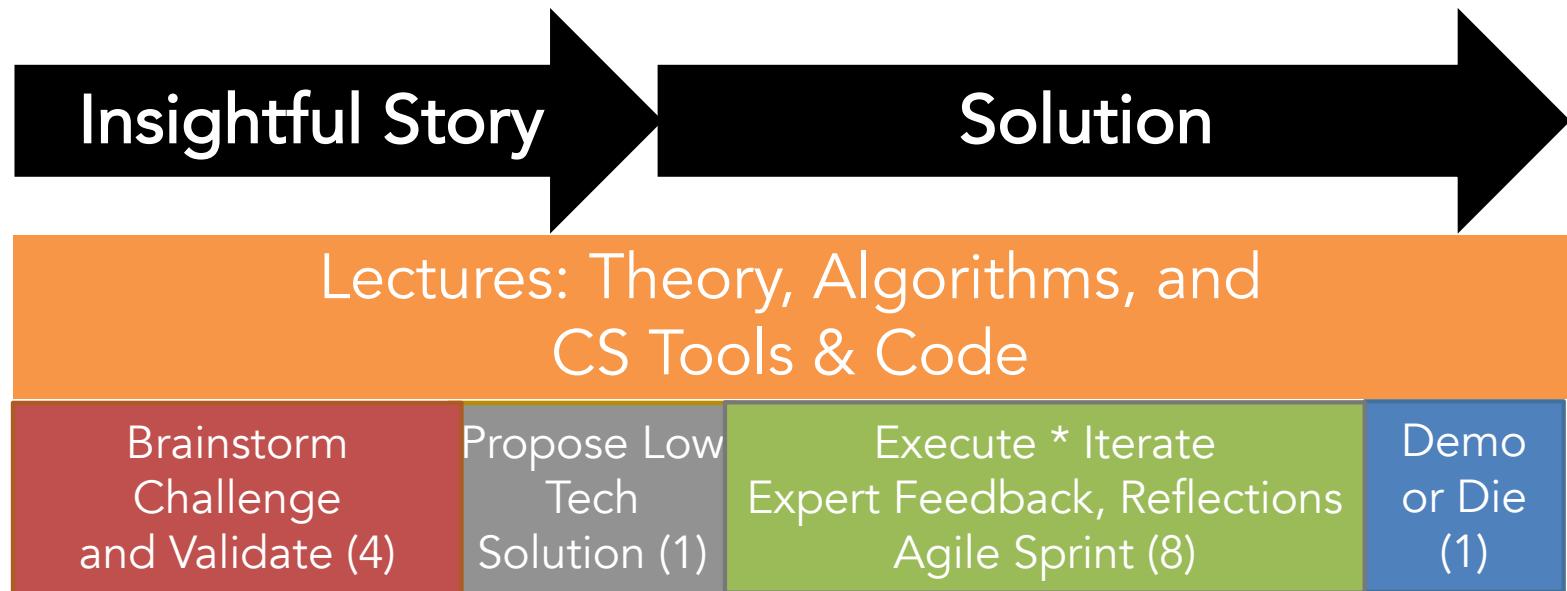
This class will help you combine math and data concepts

The course updates with new tools to stay current. You may learn and use tools not presented in the class project.

# What is actually in this class?

- The ML stack most commonly used in creating ML/AI/Data applications
- Application and **systems** viewpoint of data and ML
- Implementation, architecture, and relevant processes to build real systems
- Statistical, rule based, and **hybrid decision systems**
- Connection with relevant **mathematical foundations** (optimization, entropy, correlation, LTI, prediction, classification)
- Practical insight into **advanced techniques** and tools: (eg. CNNs, NLP, scraping, recurrent networks, etc.)
- System **modeling** for data applications
- Application talks: Recommender systems, Blockchain, Spark etc.

## How the Data-X Course Works:



**Open-ended, real-world project:** Typically 5 students, with available advisor network

# X at Berkeley

SCET Post Grad and Thought Leadership

- **Labs**
  - Data-X, Blockchain, Food
  - Technical depth for classes
  - New Model: Our labs also create teams
- **Professional Programs**
  - ELPP over-subscribed, very rich network
  - Silicon Valley Innovation Leadership Week

The screenshot shows the official website for the Sutardja Center for Entrepreneurship & Technology at Berkeley Engineering. The header features the center's logo and navigation links for About, Courses, Programs, X at Berkeley, News, Explore, and Contact. The main title "X at Berkeley" is prominently displayed in large yellow letters against a dark background with a grid pattern. Below the title, a section titled "Labs and Executive Programs" is visible. A sub-section titled "We focus on the 'Future of X'" discusses the center's focus on new technologies, societal problems, and emerging industry areas. It highlights the "main approach" of building diverse teams of students, researchers, industry experts, and investors. Below this, sections for "Labs" and "Executive Programs" are shown, each with images and descriptions of their respective areas of focus.

**Sutardja Center for Entrepreneurship & Technology**  
Berkeley Engineering

About Courses Programs X at Berkeley News Explore Contact

# X at Berkeley

## Labs and Executive Programs

We focus on the "Future of X"

For us, 'X' means new technologies, societal problems, and emerging industry areas that will change the world. Our Labs are at the most applied edge of the spectrum of all university research.

While most labs focus on pure research contribution, we work on this also, however, our main approach is to build teams of diverse experts from students, researchers, industry experts, and investors. In this way, our labs produce teams that develop publications, new technologies, business models, new ventures, and social impact.

Results from our labs inform and support the Sutardja Center's undergraduate, graduate, professional, and executive courses and programs.

### Labs

 Data Lab

 Blockchain Lab

 Sustainable Foods Lab

### Executive Programs

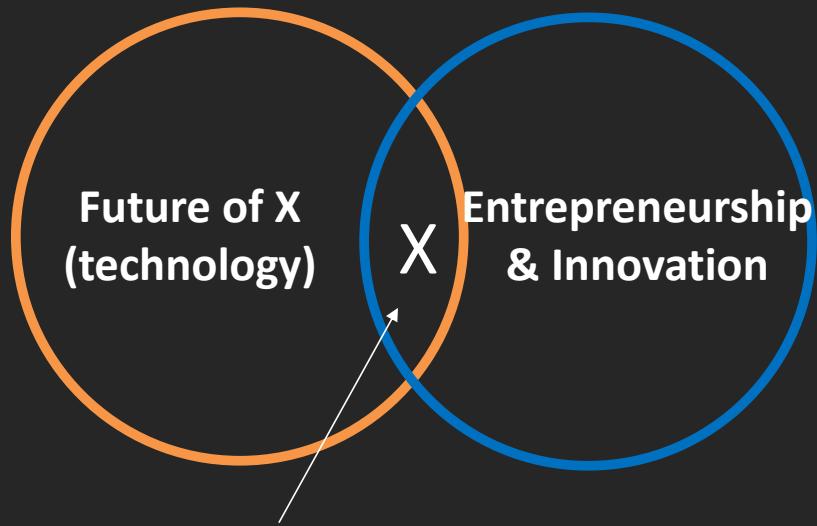
 Executive Programs



Ikhlaq Sidhu, content author

# X at Berkeley: Labs

A New Model for Applied Research Labs



X at Berkeley: Applied Research that results in new courses, team formation, and industry-wide thought leadership

The screenshot shows the homepage of the Sutardja Center for Entrepreneurship & Technology. At the top, there is a navigation bar with links for About, Courses, Programs, X at Berkeley, News, Explore, and Contact. The main title "X at Berkeley" is displayed in a large, stylized yellow font. Below it, the subtitle "Labs and Executive Programs" is visible. A section titled "We focus on the 'Future of X'" discusses the center's focus on new technologies, societal problems, and emerging industry areas. It highlights their approach of building diverse teams of students, researchers, industry experts, and investors to develop publications, new technologies, business models, new ventures, and social impact. Results from their labs inform and support undergraduate, graduate, professional, and executive courses and programs. Below this, there are sections for "Labs" and "Executive Programs", each featuring images and descriptions of specific labs and programs.

Perrin and Ting  
**Sutardja Center**  
for Entrepreneurship & Technology  
Berkeley Engineering

About Courses Programs X at Berkeley News Explore Contact

# X at Berkeley

## Labs and Executive Programs

We focus on the "Future of X"

For us, 'X' means new technologies, societal problems, and emerging industry areas that will change the world. Our Labs are at the most applied edge of the spectrum of all university research. While most labs focus on pure research contribution, we work on this also, however, our main approach is to build teams of diverse experts from students, researchers, industry experts, and investors. In this way, our labs produce teams that develop publications, new technologies, business models, new ventures, and social impact. Results from our labs inform and support the Sutardja Center's undergraduate, graduate, professional, and executive courses and programs.

### Labs

 Data Lab

 Blockchain Lab

 Sustainable Foods Lab

### Executive Programs

 Executive Programs



Ikhlaq Sidhu, content author

# Our Agenda and Purpose

1. Short Version of full course
  1. What, Why, and How for AI
  2. Learn the ML tools/stack
  3. Code samples / basic theory
  4. Starting point to go further

2. Project Options:
  1. Learning basics
  2. Defining Work Project
  3. Improving Existing Project

3. Additional Goals:
  1. Teams – meet each other
  2. Diversity in perspective
  3. Pay it forward
  4. Global resources
  5. Staying connected

Why:  
Relevance to  
Organization

What:  
Technology

How:  
Implementation  
with Innovation  
Behaviors

# Our Agenda and Purpose

1. Short Version of full course
  1. What, Why , and How for AI
  2. Learn the ML tools/stack
  3. Code samples / basic theory
  4. Starting point to go further
2. Project Options:
  1. Learning basics
  2. Defining Work Project
  3. Improving Existing Project
3. Additional Goals:
  1. Teams – meet each other
  2. Diversity in perspective
  3. Pay it forward
  4. Global resources
  5. Staying connected

	<b>Day 1: Executive Overview: Also for technical leaders</b>	<b>Day 2: Tools and solutions: Focused for technical leads</b>	<b>Day 2.5 Deep dive: Focused for technical leads</b>
Session 1  9 - 10.20am	<p>Program Introduction with Objectives (30 mins)</p> <p>Executive Level Overview of AI, ML, and Research (60 mins)</p>	<p>Introduction to tools and installations (45 mins)</p> <p>Standard AI Development Stack: Numpy, pandas, matplotlib (45 min)</p> <p>Includes a Breakout HW Session (to practice before tomorrow). Additional help with installation if needed</p>	<p><b>Advanced Topics:</b> Neural Networks, TensorFlow, Computer Vision, Keras and Cats vs Dogs! (75mins)</p> <p>Q&amp;A on Advanced topics in Data Science (15mins)</p>
Session 2  10.40 - noon	<p><b>Define Projects (70 mins)</b></p> <ul style="list-style-type: none"> <li>A. Learning basics</li> <li>B. Defining Work Project</li> <li>C. Improving Existing Project</li> </ul> <p>Finalize projects</p>	<p><b>Getting and Working with Data.</b> Web scraping solutions (80mins)</p> <p>Possible discussion and work sessions on low tech demo</p>	<p><b>Short intro to Big Data:</b> Spark and Databricks</p> <p><b>Reflection and Next Steps:</b> Lowtech demo, final remarks</p> <p>Learning Summary and Discussion Future Roadmap</p>
Lunch  12 - 1.30pm	Possible discussion and work sessions.	Possible discussion and work sessions.	<p><b>Staying Connected</b></p> <p>Lunch and end of 2.5 day master class</p>
Session 3  1.30 - 3.00pm	<p>Innovation Leadership Overview (60 mins)</p> <p>Assigned Lowtech demo for Day 2 (30 min)</p>	<p><b>Machine Learning in Python:</b> Real World Data and Example (60 mins)</p> <p>ML Algorithm Comparison (20mins)</p>	
Session 4  3.30 - 5.00pm	<p><b>Blockchain and Web 3.0</b></p> <ol style="list-style-type: none"> <li>1. The Breakthrough and how Bitcoin works</li> <li>2. Current state of the Blockchain</li> <li>3. Smart Contracts and business use cases</li> <li>4. Future and Challenges</li> </ol>	<p><b>Project (80mins)</b></p> <p>Possible discussion and work sessions on low tech demo</p>	

or

# An Overview of Data and AI Applications

Ikhlaq Sidhu, content author

# Basic Concept of Working with Data



- Data Wrangling
- In Production

## Example: Data and Information is a competitive advantage

### Real-life Example: ZestCash

- “All data is credit data”



The data says: greater credit risk!

The data says: lesser credit risk!

Reference: Shomit Ghose

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



*Reference: Supercrunchers*

## Top 8 Business Models Using Data

1. Knowing your customer, **better targeting and relationship.**  
E.g. Target, Disney, Netflix
2. Improving **physical** product or servicer with **complimentary information:**  
E.g. UPS, FedEx
3. Data-driven **reliability** or security  
E.g. GE, BMW, Siemens
4. Information **Brokers, Arbitrage, and Trading Opportunities:**  
E.g. Investment funds.
5. Improving the **customer journey/experience..**  
E.g. Harrah's

---

6. Functional Applications: **HR/Hiring**, Operations etc..  
Eg Walmart, Baseball, Sports
7. Efficiency or better performance **per dollar cost.**  
E.G. General IT, SAP, etc
8. Risk Management, regulation, and **compliance**  
Eg. Compliance 360



## Why: More Simply

Customer  
Insight/  
Engagement

Operations:  
Reliable &  
Predictable

Security &  
Fraud



Compliance 360°

Financial Firms

Network Security

# 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

# The Core of Every Business in Future is Data and AI

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



=



*resources*

+



Information and  
automated decisions

Less value  
over time

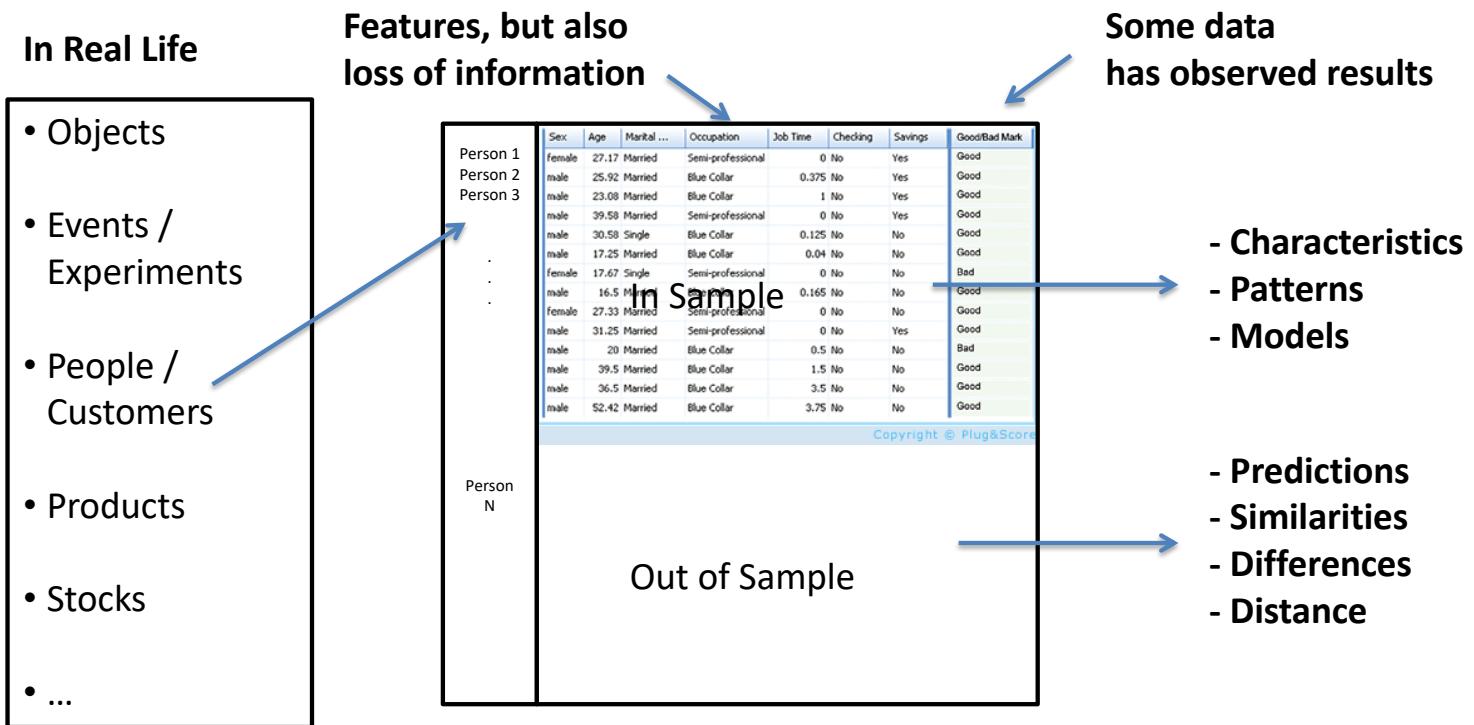
More value  
Over time

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

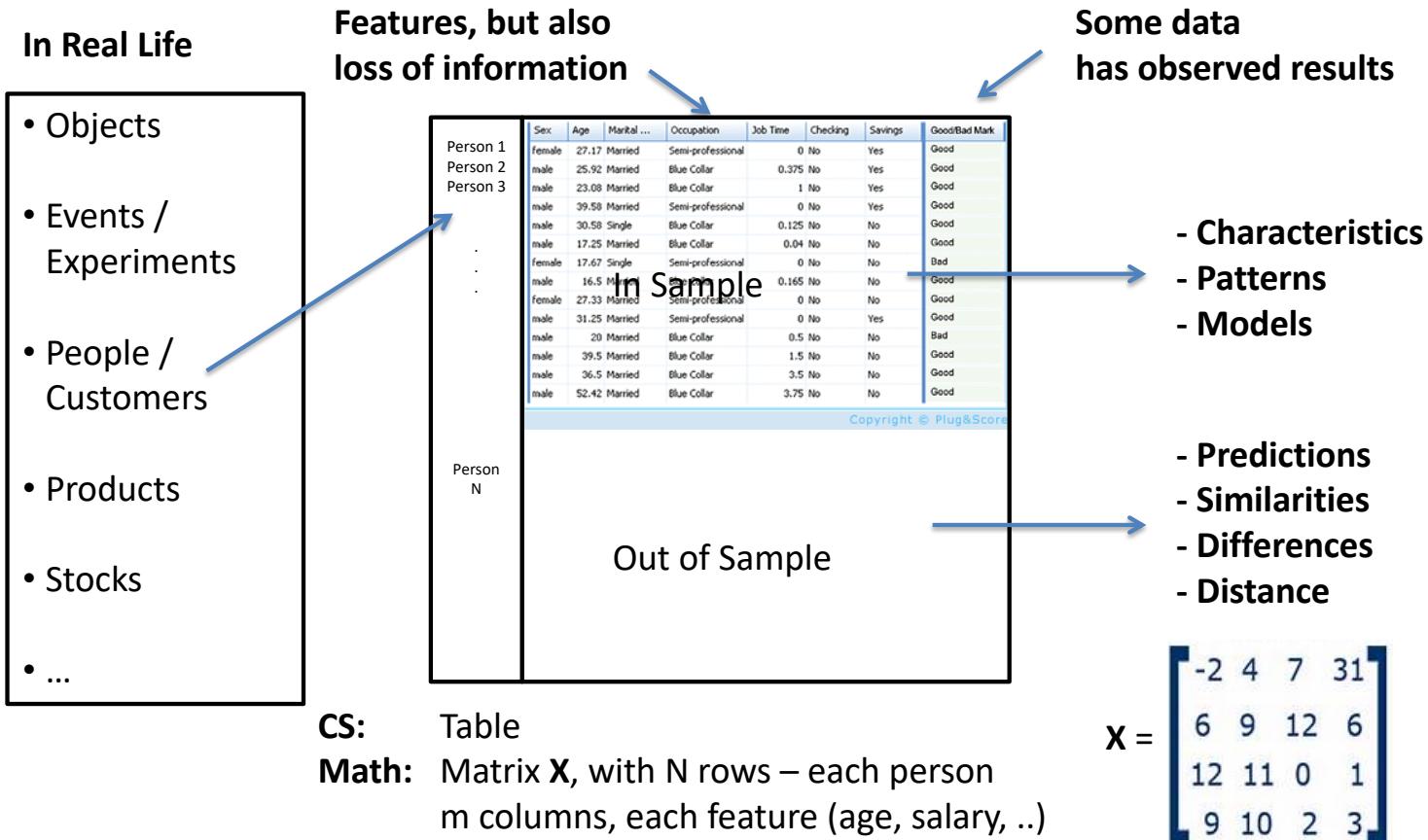
Ikhlaq Sidhu, content author

## Data and AI Approaches

# An ML High Level Framework



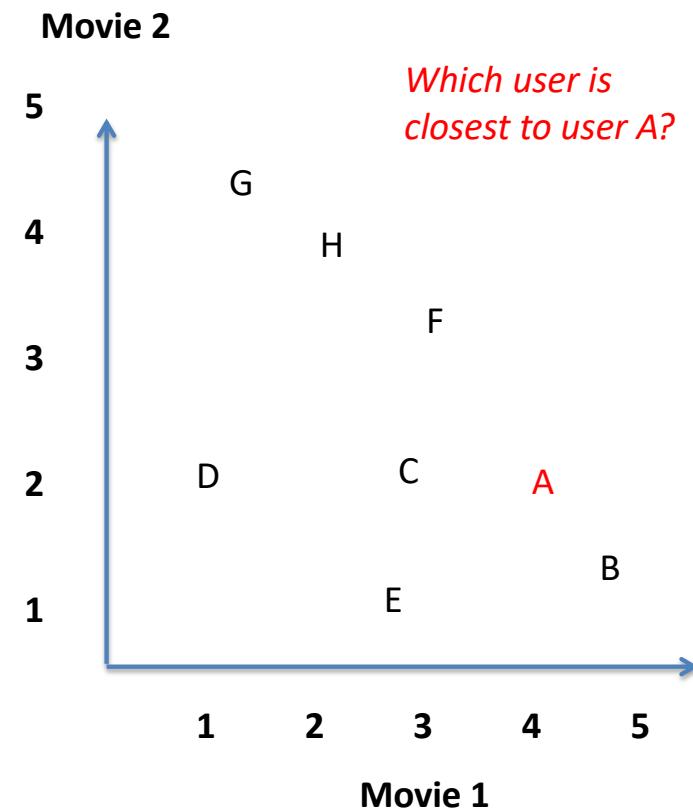
# An ML High Level Framework



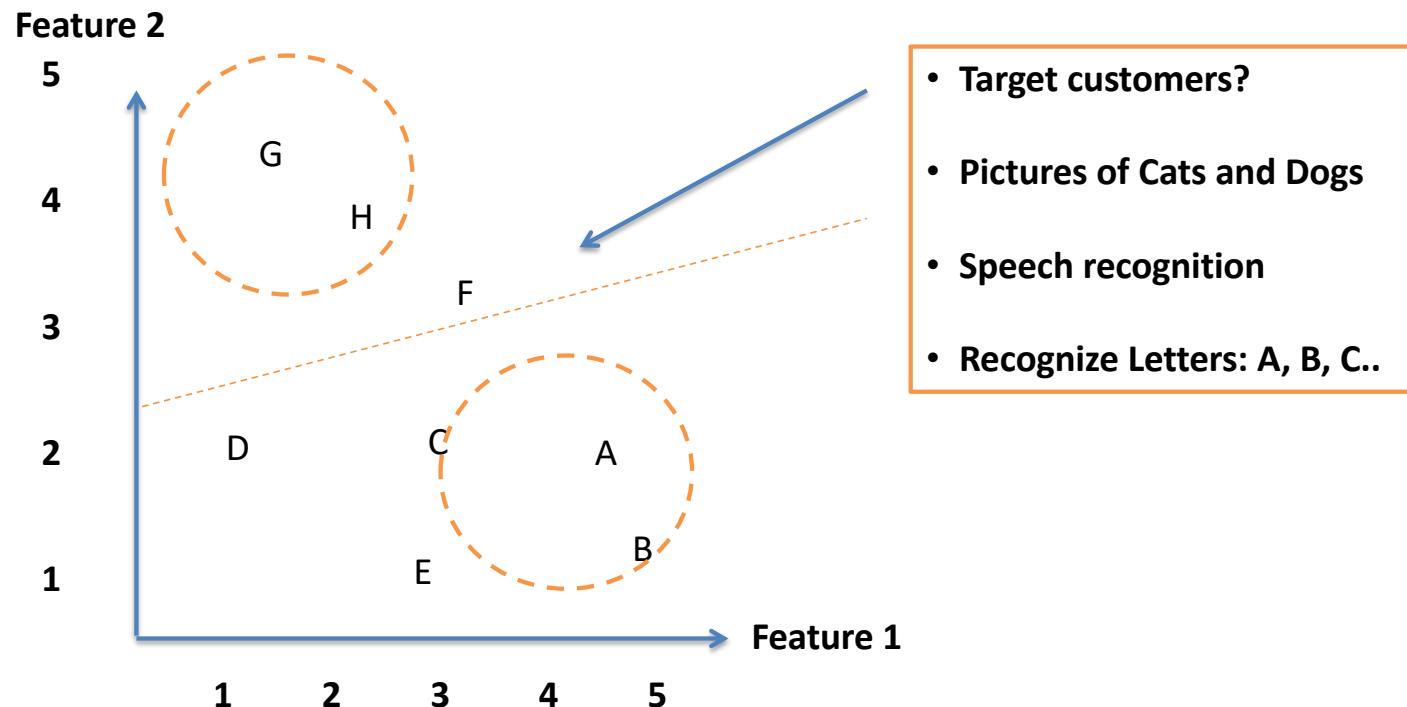
## A Fundamental Idea: From Table to N- Dimensional Space

$X =$

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
..	..	..	..



# Clustering to Classification



# A Fundamental Idea: From Table to Score

X =

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

$F(X)$

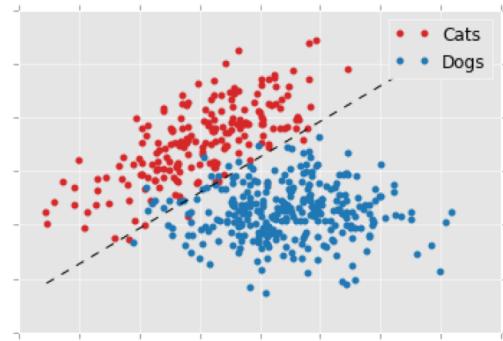
Cust	Credit Score
A	552
B	381
C	760
D	330
E	452
F	678
..	..

Y

Ikhlaq Sidhu, content author

## Traditionally 2 Tasks: Classification & Predictive Scoring

Extracted Data  
often in  
Table  
Format



Classification:  
Cats and Dogs, Speech Recognition  
Movie Recommendation



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



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

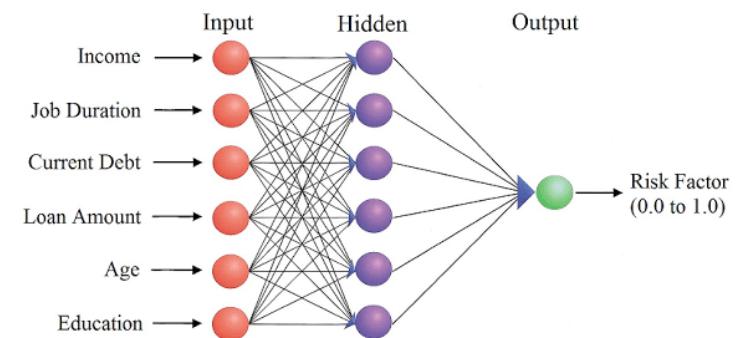
Ikhlaq Sidhu, content author

We have now switched  
to Neural Networks as  
Function Approximators



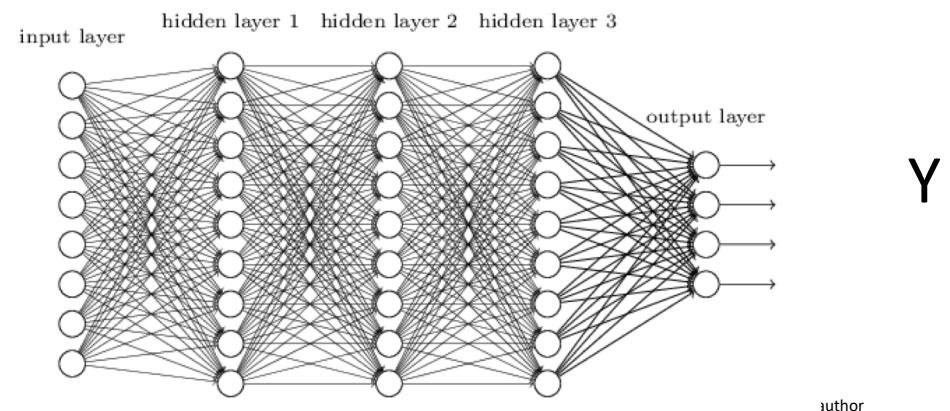
"Non-deep" feedforward  
neural network

X



Deep neural network

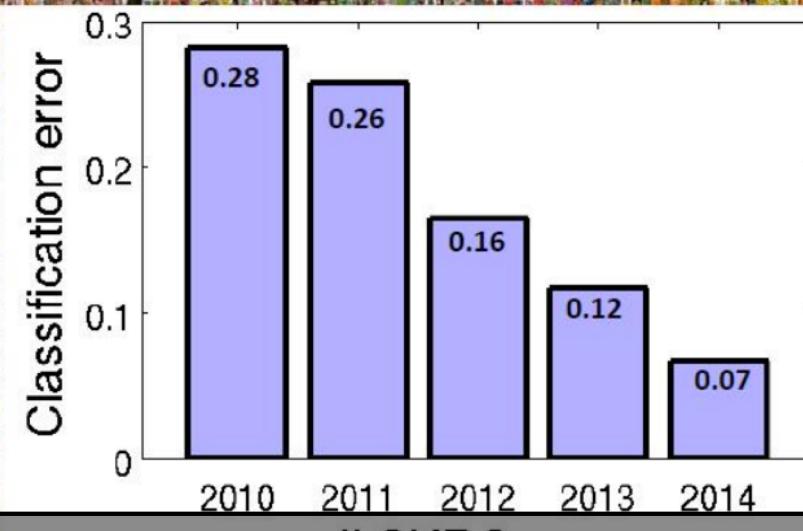
X



# IMAGENET Large Scale Visual Recognition Challenge

Street drum

The Image Classification Challenge:  
1,000 object classes  
1,431,167 images



Neural net results are close to human results

Russakovsky et al. arXiv, 2014

author

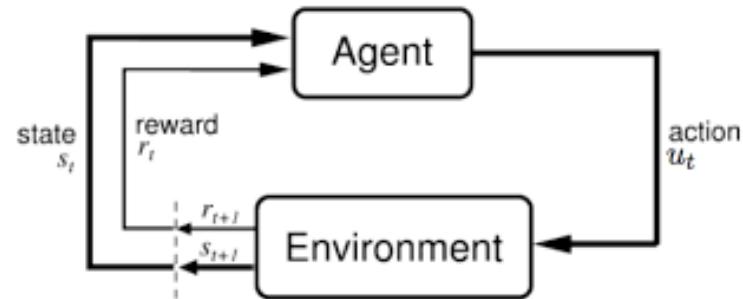
# Data and AI Future Directions

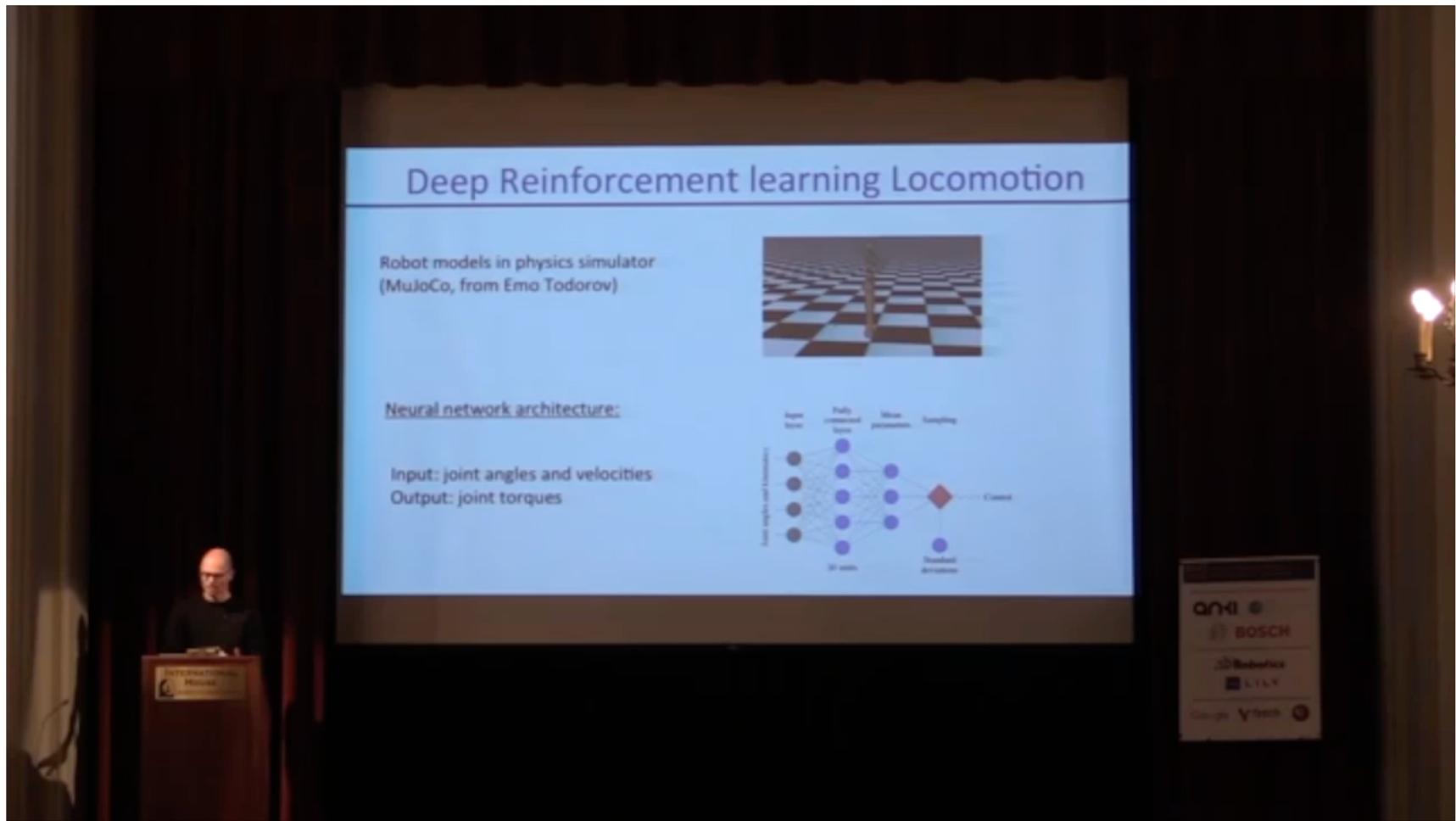
Ikhlaq Sidhu, content author

## Peter Abbeel – Deep Reinforcement Learning



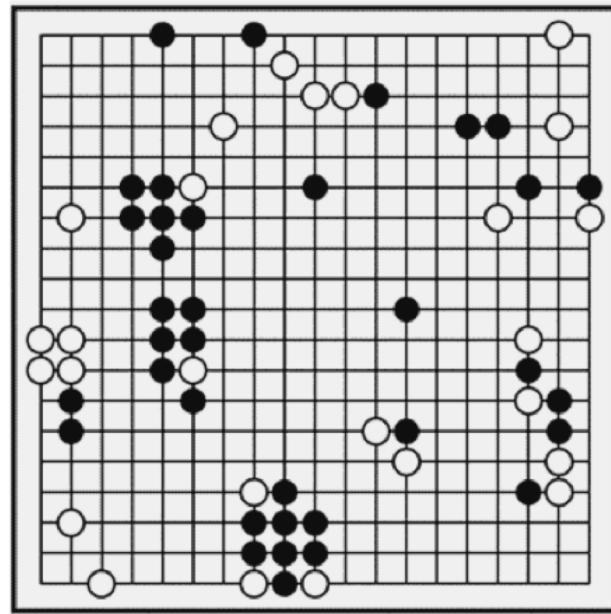
Peter Abbeel  
Professor at UC Berkeley





Ikhlaq Sidhu, content author

# Recent AI News



Source: Ken Goldberg, CPAR, People and Robotics Initiative

Ikhlaq Sidhu, content author

Does this mean AI Can Do  
Everything Better than Humans

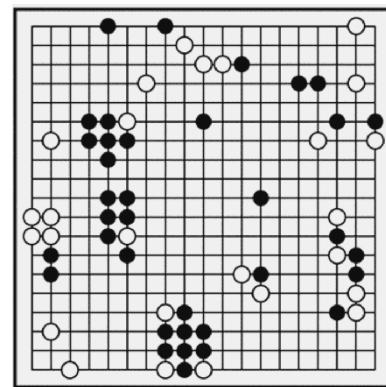
## Even then, AI Cannot Solve Real Life Problems Better Than Humans And in fact, AI Can not even Work without Humans



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

Ikhlaq Sidhu, content author

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



Massive Data



## Google Operations

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

Ikhlaq Sidhu, content author



PIETER  
ABBEEL\*



PETER  
BARTLETT\*



TREVOR  
DARRELL\*



ANCA  
DRAGAN\*



ALYOSHA  
EFROS\*



JOHN  
DENERO



LAURENT  
EL GHAOUI



RON  
FEARING



JACK  
GALLANT



JOSEPH  
GONZALEZ



KEN  
GOLDBERG\*



MICHAEL I.  
JORDAN\*



MATT  
KLEIN\*



LEVINE\*



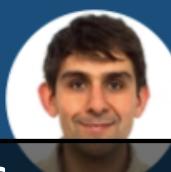
星辰



TOM  
MALIK\*



GRIFFITHS



HARDT



MARTA  
HEARDST



KURT  
KEUTZER



REN  
NG



BEN  
RECHT\*



STUART  
RUSSELL\*



RUZENA  
BAJCSY



ALEXANDRE  
BAYEN



JOHN  
CANNY



BRUNO  
OLSHAUSEN



CHRISTOS  
PAPADIMITRIOU



SHANKAR  
SASTRY



DAWN  
SONG



CLAIRE  
TOMLIN

At Berkeley, we have a lot of research on  
“How Machines Will Work as Part of Larger Systems  
that Work with People”



MARTIN  
WAINWRIGHT



LAURA  
WALLER



BIN  
YU



AVIDEH  
ZAKHOR



JEROME A.  
FELDMAN†



NELSON  
MORGAN†



LOTFI  
ZADEH†

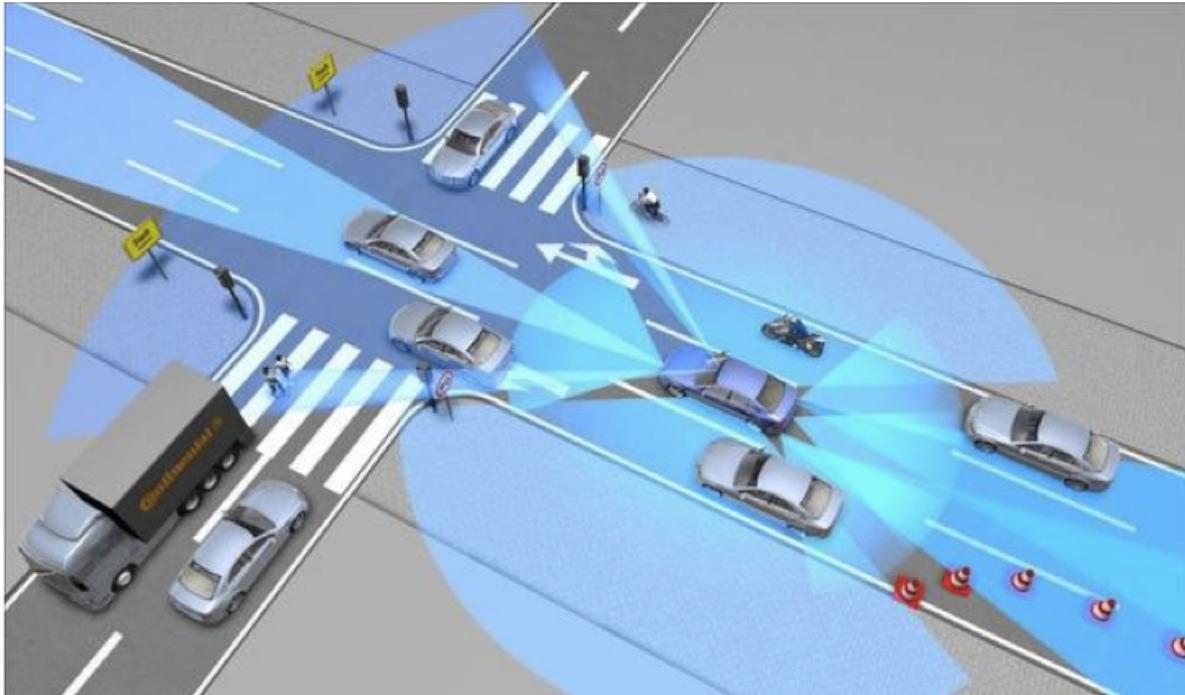
37

faculty

# My Drive Home From Berkeley



# 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

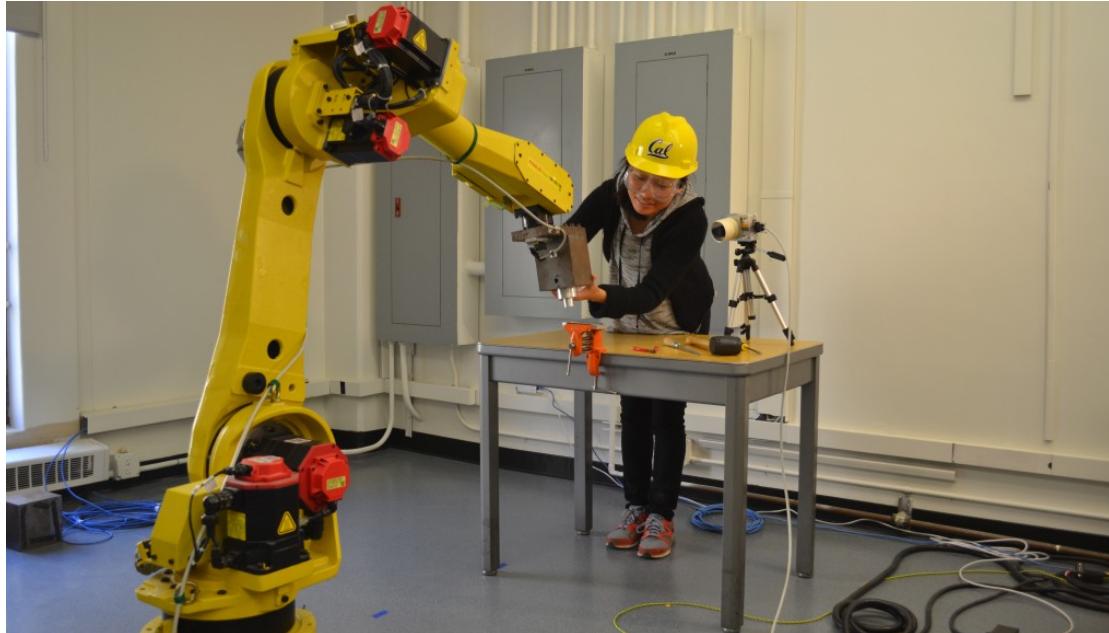


**Francesco Borrelli**  
UC Berkeley

UC Berkeley People and Robotics Initiative

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

UC Berkeley 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

A Case Example  
to Understand



### 3 Factors to Understand:

1. Data vs Algorithm: Data is more valuable
2. Algorithm -> System: ML is only a small fraction of the system
3. Algorithm, Data, and Computing: Data is growing faster than computing

# 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

Ikhlaq Sidhu, content author

Ikhlaq Sidhu, content author

Before We Go On to discuss  
Tools, Implementation, and Projects

# We have students who experience this:

I want the computer to do X:

- recognize faces
- write an email reply automatically
- ... anything else a human would do

Then they build a 15 layer neural network (or something else with 1000 parameters) , train it with the data they have.

And then they say: I get 15% accuracy, but the state of the art is 90%

What happened?

# Advanced Warning

Yes: Tools are powerful

Yes: Neural Networks have high accuracy

However: AI, ML alone won't solve your problem because

- Real problems are more complex
- Complex models require a lot of data
- And only part of the problem is about AI and ML

## You Will Still Need to Understand and Think

I've seen many  
technical projects with  
smart people go off  
track

• • •



## Why we can't deliver:

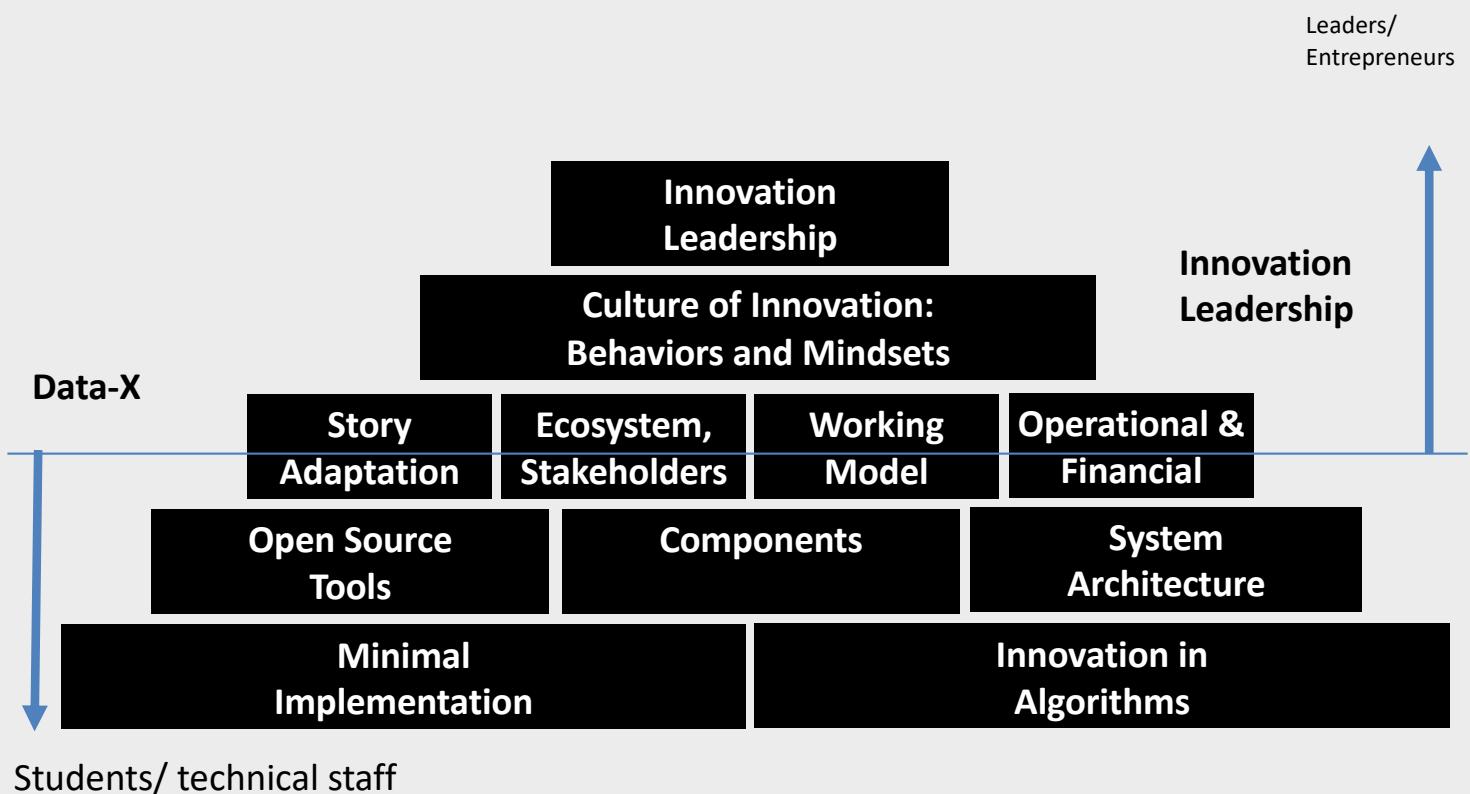
- Theoretical understanding **without a practical understanding** of implementation
- Narrow focus: **silos of disconnected expertise** not leading to any useful work product or innovation
- Over-design: **way too complex**
- Not even sure **what to create**. Wanting implementation specs that no one has.
- Expensive **cost over-runs** on development, sometimes even trying to create something that already exists
- Disconnected from technical **reality**
- People not on the same page (**misaligned**), cannot work with each other, **team breakdown**.

## Data-X Framework

At Berkeley, we have results:

People in our programs can build amazing, working projects in 3 months with a relatively little background in ML, AI, and other data technologies.

## A Solution for Rapid Implementation

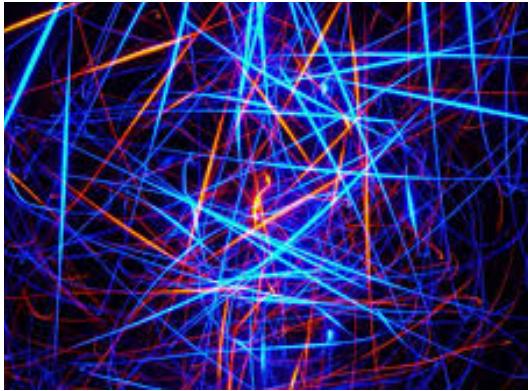


Applicable to all categories of digital transformation

Ikhlaq Sidhu, content author

End of Section

# Project Types



Learning the Basics  
Make something  
that is interesting

- Target: X teams
- Process: All write down a project idea / objective use NABC and Current Situation
- Share to collect team members just or this class



Defining a New  
Work Project



Improve Your  
Existing Company Project

# How to Communicate Your Project

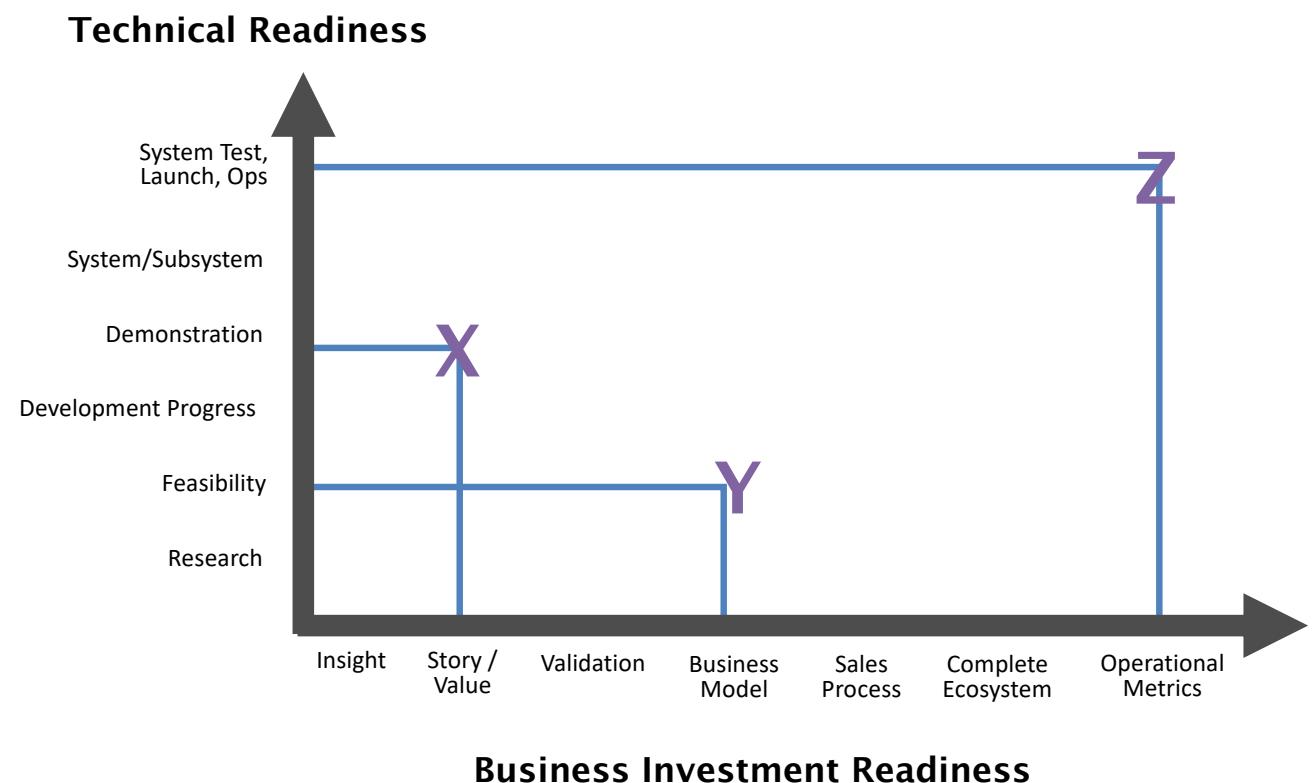
Need:  
For whom and why?

Approach:  
How will you do it?

Benefit:  
If its done, what is the benefit

Competition:  
Are there any other ways to do this or who  
else is doing something that competes

# Identify the Stage of Your Project



Ikhlaq 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