worldusabilityday

Catch me if you can

How to fight Fraud, Waste and Abuse using Machine Learning AND Machine Teaching?

Cupid Chan

2020-11-12

Human-Centered Design
Center of Excellence

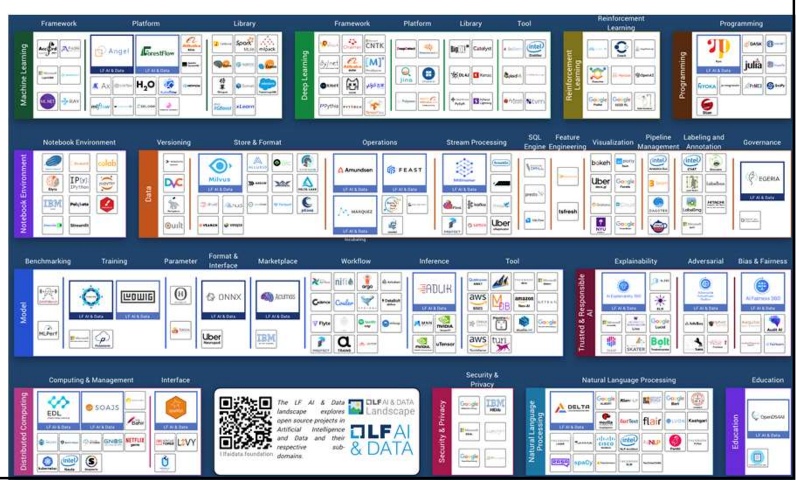
CCSQ WORLD USABILITY DAY 1

1

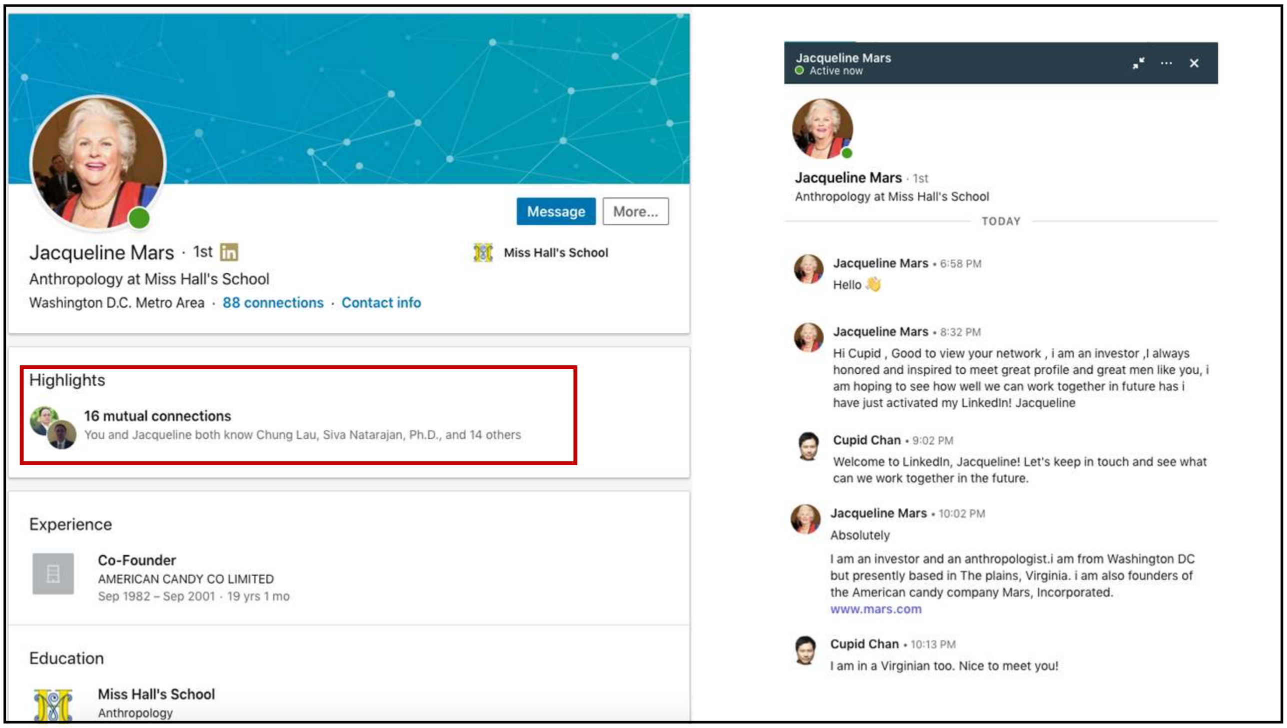
ODPi
odpi.org

\+ LF AI

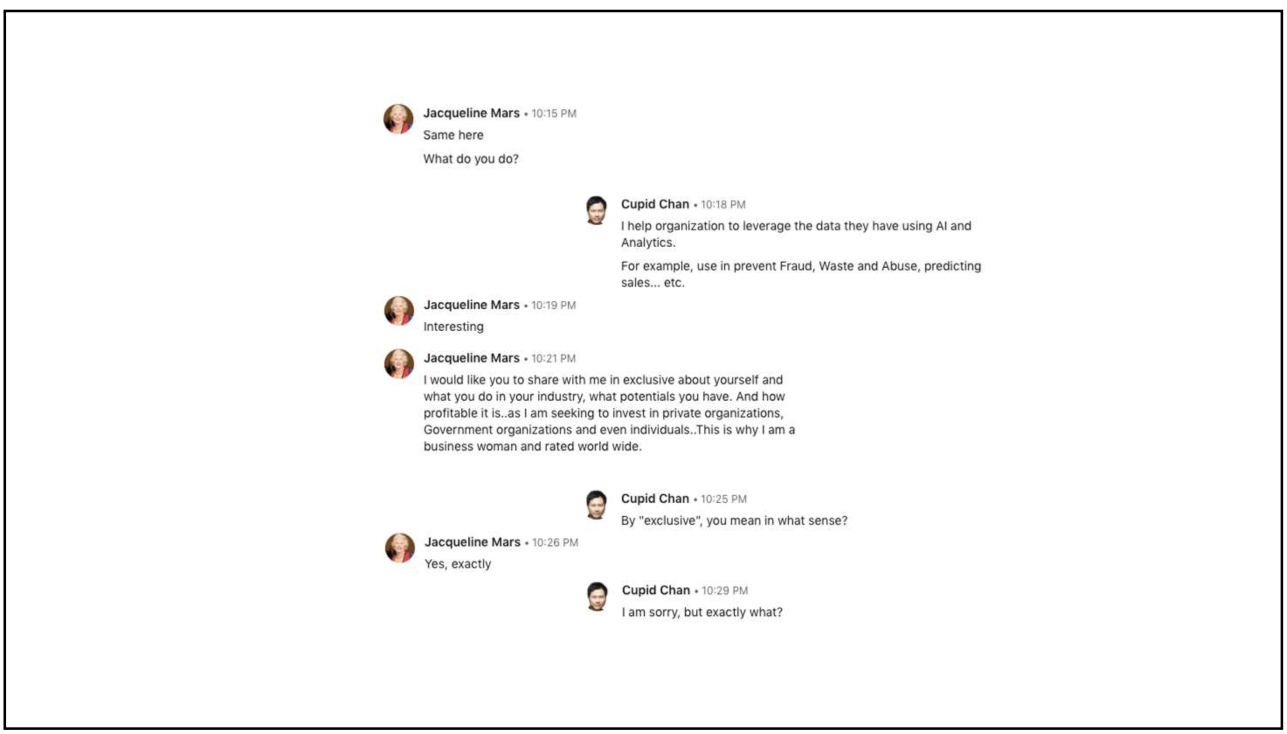
LF AI & DATA



2


Message
More...
Jacqueline Mars · 1st
Anthropology at Miss Hall's School
Washington D.C. Metro Area · 88 connections · Contact info
Miss Hall's School
Highlights
16 mutual connections
You and Jacqueline both know Chung Lau, Siva Natarajan, Ph.D., and 14 others
Experience
Co-Founder
AMERICAN CANDY CO LIMITED
Sep 1982 – Sep 2001 · 19 yrs 1 mo
Education
Miss Hall's School
Anthropology
Jacqueline Mars
Active now
Jacqueline Mars · 1st
Anthropology at Miss Hall's School
TODAY
Jacqueline Mars • 6:58 PM
Hello
Jacqueline Mars • 8:32 PM
Hi Cupid , Good to view your network , i am an investor ,I always honored and inspired to meet great profile and great men like you, i am hoping to see how well we can work together in future has i have just activated my LinkedIn! Jacqueline
Cupid Chan • 9:02 PM
Welcome to LinkedIn, Jacqueline! Let's keep in touch and see what can we work together in the future.
Jacqueline Mars • 10:02 PM
Absolutely
I am an investor and an anthropologist.i am from Washington DC but presently based in The plains, Virginia. i am also founders of the American candy company Mars, Incorporated.
www.mars.com
Cupid Chan • 10:13 PM
I am in a Virginian too. Nice to meet you!

3


Jacqueline Mars • 10:15 PM
Same here
What do you do?
Cupid Chan • 10:18 PM
I help organization to leverage the data they have using AI and Analytics.
For example, use in prevent Fraud, Waste and Abuse, predicting sales... etc.
Jacqueline Mars • 10:19 PM
Interesting
Jacqueline Mars • 10:21 PM
I would like you to share with me in exclusive about yourself and what you do in your industry, what potentials you have. And how profitable it is..as I am seeking to invest in private organizations, Government organizations and even individuals..This is why I am a business woman and rated world wide.
Cupid Chan • 10:25 PM
By "exclusive", you mean in what sense?
Jacqueline Mars • 10:26 PM
Yes, exactly
Cupid Chan • 10:29 PM
I am sorry, but exactly what?

4

LinkedIn Story to be continued…

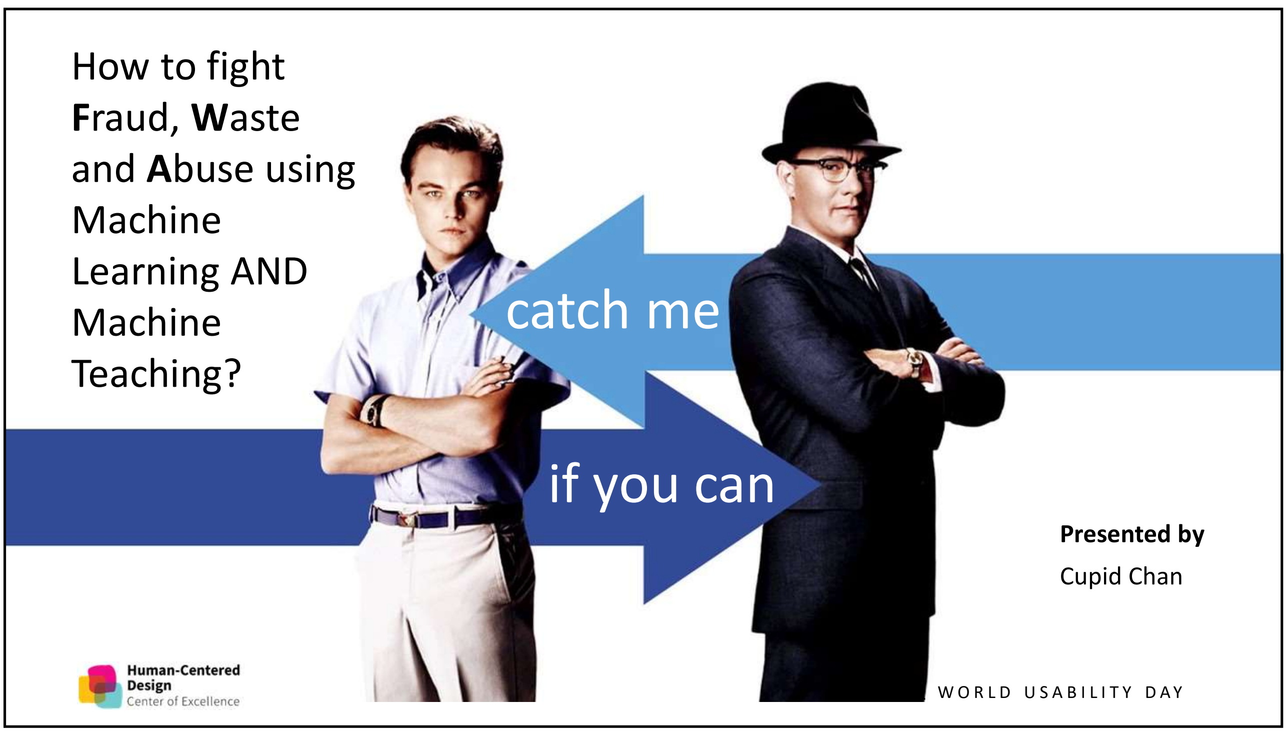
5

10 seconds Polling Question

Have you watched the movie "Catch me if you can"?

6

7


How to fight **F**raud, **W**aste and **A**buse using Machine Learning AND Machine Teaching?
catch me
if you can
Presented by
Cupid Chan
Human-Centered Design
Center of Excellence
WORLD USABILITY DAY

8

Fraud, Waste, Abuse and Error

Fraud
Intentional Deception

Abuse
Bending the rules

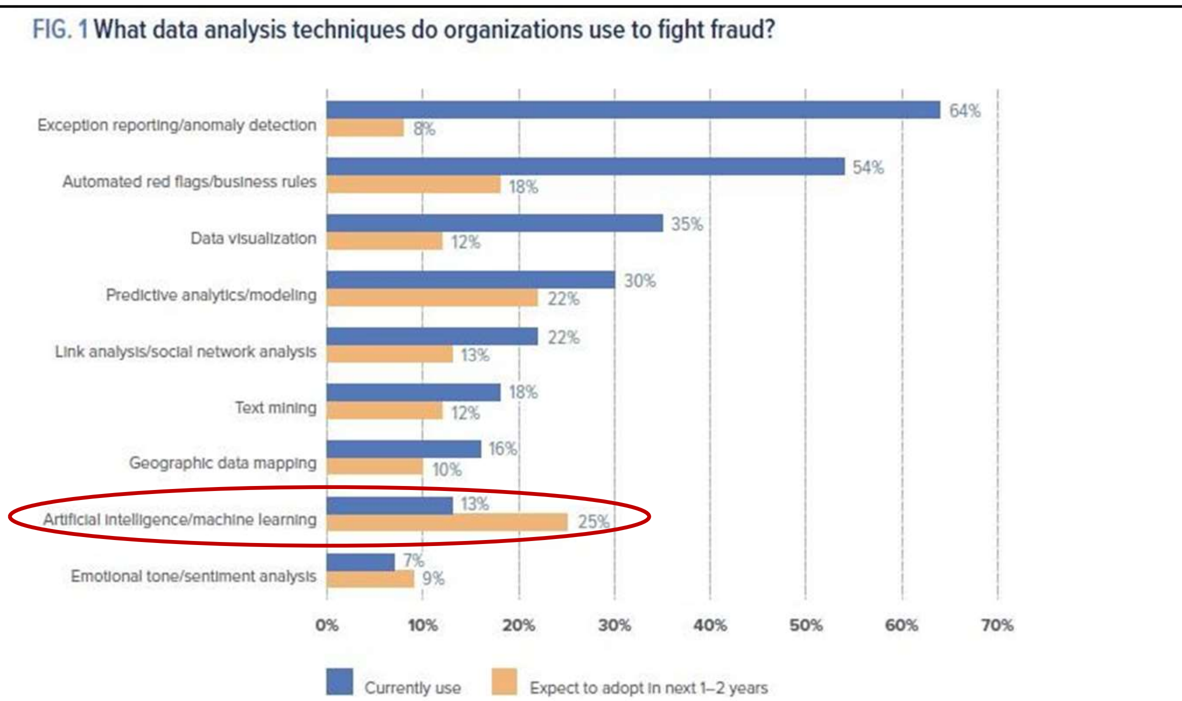
Waste
Inefficiency

Error
Unintentional fault

Human-Centered Design
Center of Excellence

CCSQ WORLD USABILITY DAY 9

9



https://www.forbes.com/sites/louiscolumbus/2019/08/01/ai-is-predicting-the-future-of-online-fraud-detection/#1a19bfa374f5

10

10 seconds Polling Question - 1

Which of the following popular fraud costing the most?

- Credit card Fraud
- Healthcare Fraud
- Identity Fraud

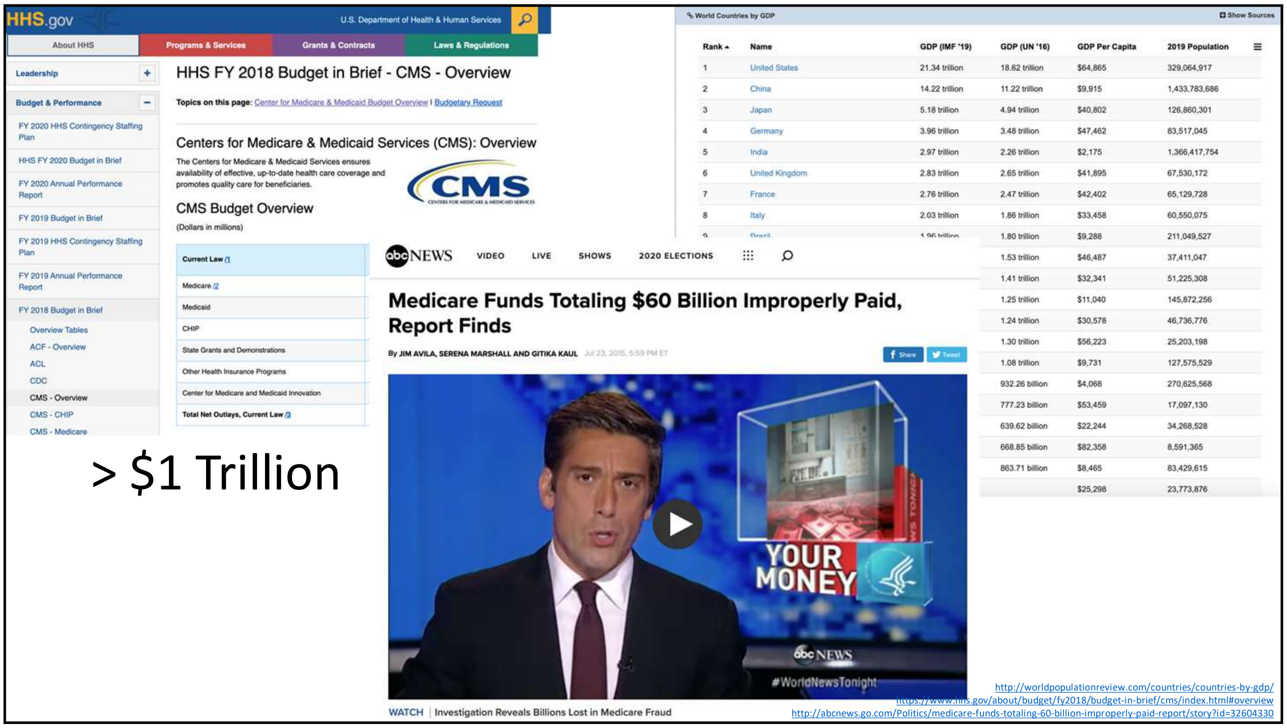
Human-Centered Design Center of Excellence

CCSQ WORLD USABILITY DAY 11

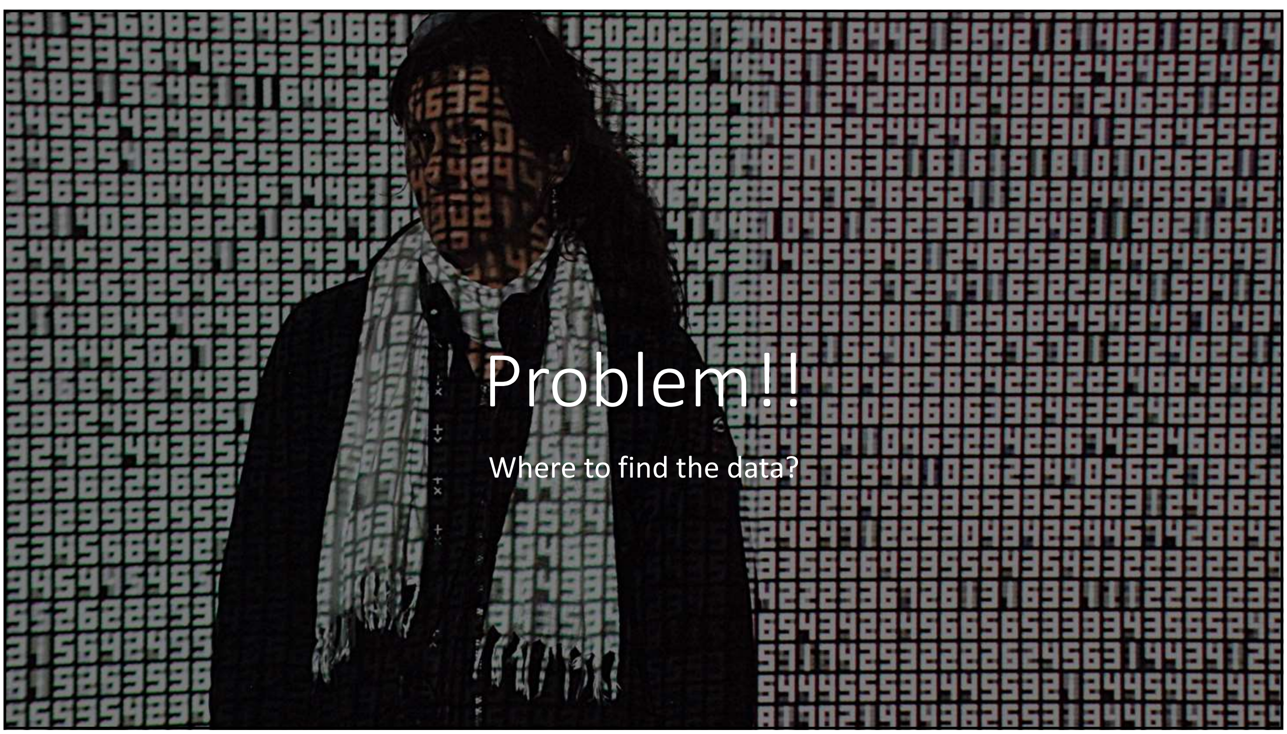
11

https://www.bcbsm.com/health-care-fraud/fraud-statistics.html
https://losspreventionmedia.com/credit-card-fraud-news-2018-update/
https://www.javelinstrategy.com/coverage-area/2019-identity-fraud-report-fraudsters-seek-new-targets-and-victims-bear-brunt

12


HHS.gov
U.S. Department of Health & Human Services
HHS FY 2018 Budget in Brief - CMS - Overview
Centers for Medicare & Medicaid Services (CMS): Overview
CMS Budget Overview
World Countries by GDP
Rank | Name | GDP (IMF '19) | GDP (UN '16) | GDP Per Capita | 2019 Population
1 | United States | 21.34 trillion | 18.62 trillion | $64,865 | 329,064,917
2 | China | 14.22 trillion | 11.22 trillion | $9,915 | 1,433,783,686
3 | Japan | 5.18 trillion | 4.94 trillion | $40,802 | 126,860,301
4 | Germany | 3.96 trillion | 3.48 trillion | $47,462 | 83,517,045
5 | India | 2.97 trillion | 2.26 trillion | $2,175 | 1,366,417,754
6 | United Kingdom | 2.83 trillion | 2.65 trillion | $41,895 | 67,530,172
7 | France | 2.76 trillion | 2.47 trillion | $42,402 | 65,129,728
8 | Italy | 2.03 trillion | 1.86 trillion | $33,458 | 60,550,075
Medicare Funds Totaling $60 Billion Improperly Paid, Report Finds
By JIM AVILA, SERENA MARSHALL AND GITIKA KAUL
YOUR MONEY
#WorldNewsTonight
WATCH | Investigation Reveals Billions Lost in Medicare Fraud
> $1 Trillion
http://worldpopulationreview.com/countries/countries-by-gdp/
https://www.hhs.gov/about/budget/fy2018/budget-in-brief/cms/index.html#overview
http://abcnews.go.com/Politics/medicare-funds-totaling-60-billion-improperly-paid-report/story?id=32604330

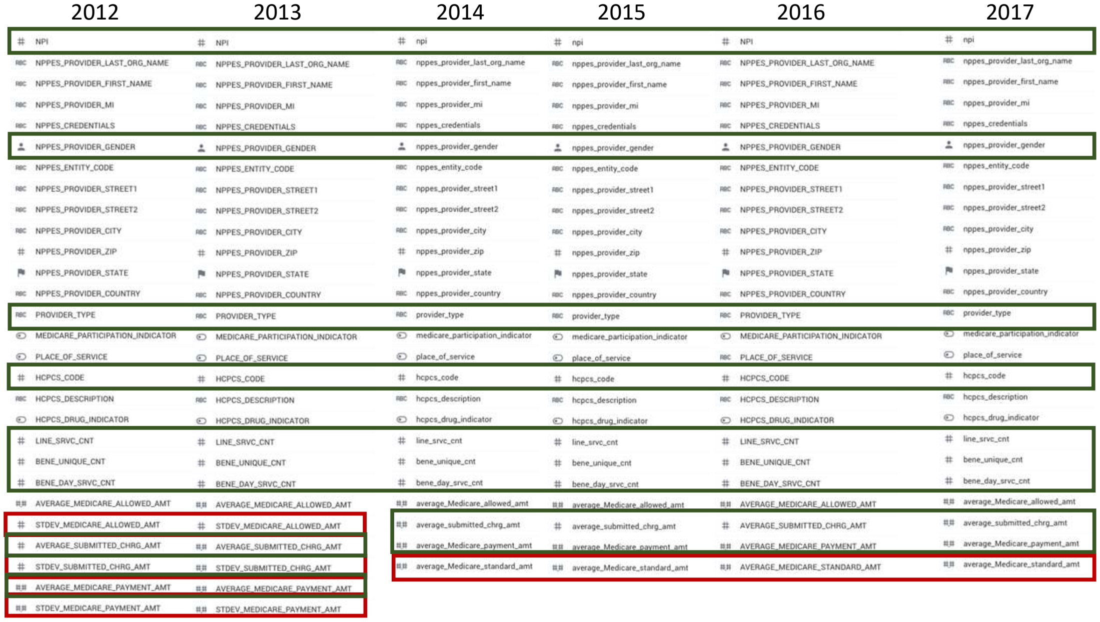
13


Problem!!
Where to find the data?

14

Medicare Provider Utilization and Payment Data: Physician and Other Supplier

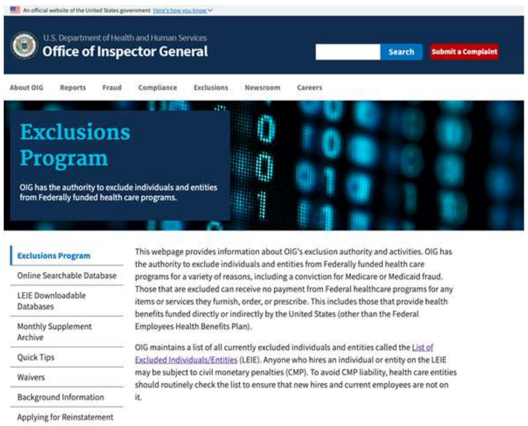
15

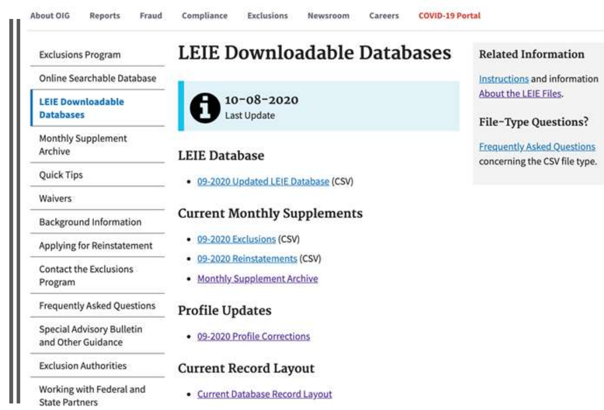


2012	2013	2014	2015	2016	2017
NPI	NPI	npi	npi	NPI	npi
NPPES_PROVIDER_LAST_ORG_NAME	NPPES_PROVIDER_LAST_ORG_NAME	nppes_provider_last_org_name	nppes_provider_last_org_name	NPPES_PROVIDER_LAST_ORG_NAME	nppes_provider_last_org_name
NPPES_PROVIDER_FIRST_NAME	NPPES_PROVIDER_FIRST_NAME	nppes_provider_first_name	nppes_provider_first_name	NPPES_PROVIDER_FIRST_NAME	nppes_provider_first_name
NPPES_PROVIDER_MI	NPPES_PROVIDER_MI	nppes_provider_mi	nppes_provider_mi	NPPES_PROVIDER_MI	nppes_provider_mi
NPPES_CREDENTIALS	NPPES_CREDENTIALS	nppes_credentials	nppes_credentials	NPPES_CREDENTIALS	nppes_credentials
NPPES_PROVIDER_GENDER	NPPES_PROVIDER_GENDER	nppes_provider_gender	nppes_provider_gender	NPPES_PROVIDER_GENDER	nppes_provider_gender
NPPES_ENTITY_CODE	NPPES_ENTITY_CODE	nppes_entity_code	nppes_entity_code	NPPES_ENTITY_CODE	nppes_entity_code
NPPES_PROVIDER_STREET1	NPPES_PROVIDER_STREET1	nppes_provider_street1	nppes_provider_street1	NPPES_PROVIDER_STREET1	nppes_provider_street1
NPPES_PROVIDER_STREET2	NPPES_PROVIDER_STREET2	nppes_provider_street2	nppes_provider_street2	NPPES_PROVIDER_STREET2	nppes_provider_street2
NPPES_PROVIDER_CITY	NPPES_PROVIDER_CITY	nppes_provider_city	nppes_provider_city	NPPES_PROVIDER_CITY	nppes_provider_city
NPPES_PROVIDER_ZIP	NPPES_PROVIDER_ZIP	nppes_provider_zip	nppes_provider_zip	NPPES_PROVIDER_ZIP	nppes_provider_zip
NPPES_PROVIDER_STATE	NPPES_PROVIDER_STATE	nppes_provider_state	nppes_provider_state	NPPES_PROVIDER_STATE	nppes_provider_state
NPPES_PROVIDER_COUNTRY	NPPES_PROVIDER_COUNTRY	nppes_provider_country	nppes_provider_country	NPPES_PROVIDER_COUNTRY	nppes_provider_country
PROVIDER_TYPE	PROVIDER_TYPE	provider_type	provider_type	PROVIDER_TYPE	provider_type
MEDICARE_PARTICIPATION_INDICATOR	MEDICARE_PARTICIPATION_INDICATOR	medicare_participation_indicator	medicare_participation_indicator	MEDICARE_PARTICIPATION_INDICATOR	medicare_participation_indicator
PLACE_OF_SERVICE	PLACE_OF_SERVICE	place_of_service	place_of_service	PLACE_OF_SERVICE	place_of_service
HCPCS_CODE	HCPCS_CODE	hcpcs_code	hcpcs_code	HCPCS_CODE	hcpcs_code
HCPCS_DESCRIPTION	HCPCS_DESCRIPTION	hcpcs_description	hcpcs_description	HCPCS_DESCRIPTION	hcpcs_description
HCPCS_DRUG_INDICATOR	HCPCS_DRUG_INDICATOR	hcpcs_drug_indicator	hcpcs_drug_indicator	HCPCS_DRUG_INDICATOR	hcpcs_drug_indicator
LINE_SRVC_CNT	LINE_SRVC_CNT	line_srvc_cnt	line_srvc_cnt	LINE_SRVC_CNT	line_srvc_cnt
BENE_UNIQUE_CNT	BENE_UNIQUE_CNT	bene_unique_cnt	bene_unique_cnt	BENE_UNIQUE_CNT	bene_unique_cnt
BENE_DAY_SRVC_CNT	BENE_DAY_SRVC_CNT	bene_day_srvc_cnt	bene_day_srvc_cnt	BENE_DAY_SRVC_CNT	bene_day_srvc_cnt
AVERAGE_MEDICARE_ALLOWED_AMT	AVERAGE_MEDICARE_ALLOWED_AMT	average_Medicare_allowed_amt	average_Medicare_allowed_amt	AVERAGE_MEDICARE_ALLOWED_AMT	average_Medicare_allowed_amt
STDEV_MEDICARE_ALLOWED_AMT	STDEV_MEDICARE_ALLOWED_AMT	average_submitted_chrg_amt	average_submitted_chrg_amt	AVERAGE_SUBMITTED_CHRG_AMT	average_submitted_chrg_amt
AVERAGE_SUBMITTED_CHRG_AMT	AVERAGE_SUBMITTED_CHRG_AMT	average_Medicare_payment_amt	average_Medicare_payment_amt	AVERAGE_MEDICARE_PAYMENT_AMT	average_Medicare_payment_amt
STDEV_SUBMITTED_CHRG_AMT	STDEV_SUBMITTED_CHRG_AMT	average_Medicare_standard_amt	average_Medicare_standard_amt	AVERAGE_MEDICARE_STANDARD_AMT	average_Medicare_standard_amt
AVERAGE_MEDICARE_PAYMENT_AMT	AVERAGE_MEDICARE_PAYMENT_AMT				
STDEV_MEDICARE_PAYMENT_AMT	STDEV_MEDICARE_PAYMENT_AMT				

16

List of Excluded Individuals and Entities (LEIE)





17

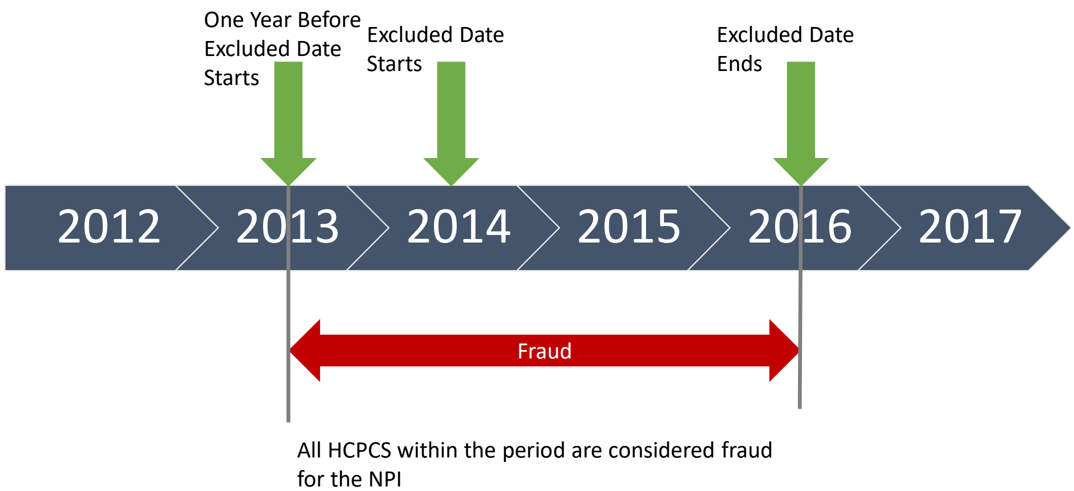
Mandatory Exclusions

Social Security Act	42 USC §	Amendment
1128(a)(1)	1320a-7(a)(1)	Conviction of program-related crimes. Minimum Period: 5 years
1128(a)(2)	1320a-7(a)(2)	Conviction relating to patient abuse or neglect. Minimum Period: 5 years
1128(a)(3)	1320a-7(a)(3)	Felony conviction relating to health care fraud. Minimum Period: 5 years
1128(a)(4)	1320a-7(a)(4)	Felony conviction relating to controlled substance. Minimum Period: 5 years
1128(c)(3)(G)(i)	1320a-7(c)(3)(G)(i)	Conviction of second mandatory exclusion offense. Minimum Period: 10 years
1128(c)(3)(G)(ii)	1320a-7(c)(3)(G)(ii)	Conviction of third or more mandatory exclusion offenses. Permanent Exclusion

https://oig.hhs.gov/exclusions/authorities.asp

18

Define what is fraud based on the data set

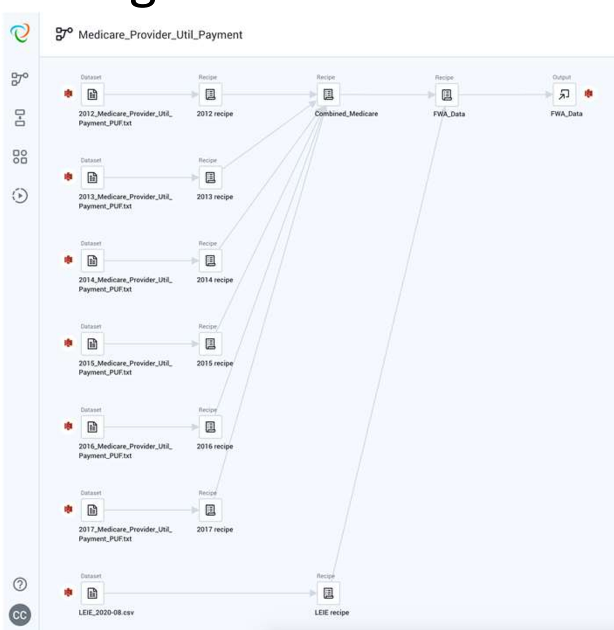


Human-Centered Design
Center of Excellence

CCSQ WORLD USABILITY DAY 19

19

Overall data ingestion flow



Human-Centered Design
Center of Excellence

CCSQ WORLD USABILITY DAY 20

20

Final data set structure

ABC	provider_type	Provider's specialty, e.g. Internal Medicine, Dermatology
	nppes_provider_gender	Provider Gender
ABC	hcpcs_code	Procedure or Service performed by the provider
#	line_srvc_cnt	Number of procedures or services the provider performed
#	bene_unique_cnt	Number of distinct Medicare beneficiaries receiving the service/procedure
#	bene_day_srvc_cnt	Number of distinct Medicare beneficiaries per day by the provider
#.#	average_submitted_chrg_amt	Average charge the provider submitted for the service or procedure
#.#	average_medicare_payment_amt	Average payment made to a provider per claim for the service
	fraud	Fraud label based on the logic described before

Human-Centered Design Center of Excellence

CCSQ WORLD USABILITY DAY 21

21

Let's predict using Machine Learning!

Based on my rich experience in AI 😎, I can build a model guaranteed with 99.9% accuracy within 10 seconds!

EVERYTHING

Is <u>NOT</u> Fraud

Human-Centered Design Center of Excellence

CCSQ WORLD USABILITY DAY 22

22

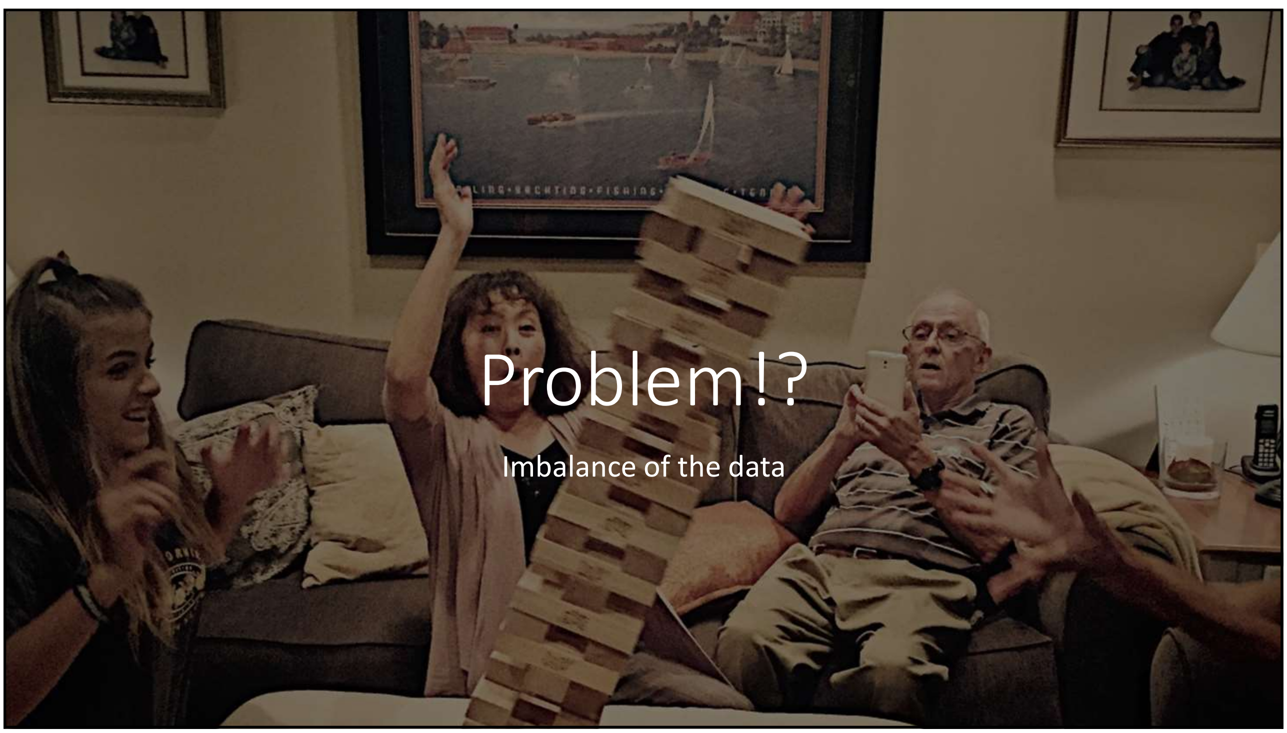
Confusion Matrix

- Total: 54,337,938
- Normal: 54,333,245
- Fraud: 4,693 (0.0086%)

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Positive (FN)	Sensitivity/Recall $\frac{TP}{TP+FN}$
	Negative	False Negative (FP)	True Negative (TN)	Specificity $\frac{TN}{TN+FP}$
		Precision $\frac{TP}{TP+FP}$	Negative Predictive Value $\frac{TN}{TN+FN}$	Accuracy $\frac{TP+TN}{TP+TN+FP+FN}$

Human-Centered Design Center of Excellence

23



24

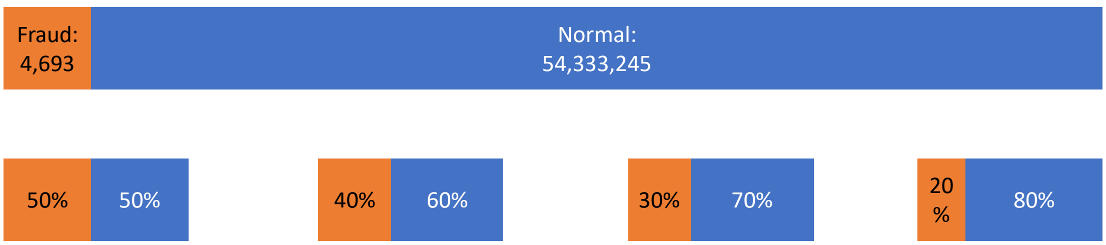

Data Discrimination – Minority Class
• Credit Card Fraud
• Manufacturing Defect
• Rare Disease Diagnosis
• Natural Disasters

25


Potential Solution for Class Imbalance
• Decreasing Majority
• Random Under Sampling (RUS)
• Increasing Minority
• Random Over Sampling
• Replicate Minority Observation
• Synthetic Minority Oversampling Technique

26

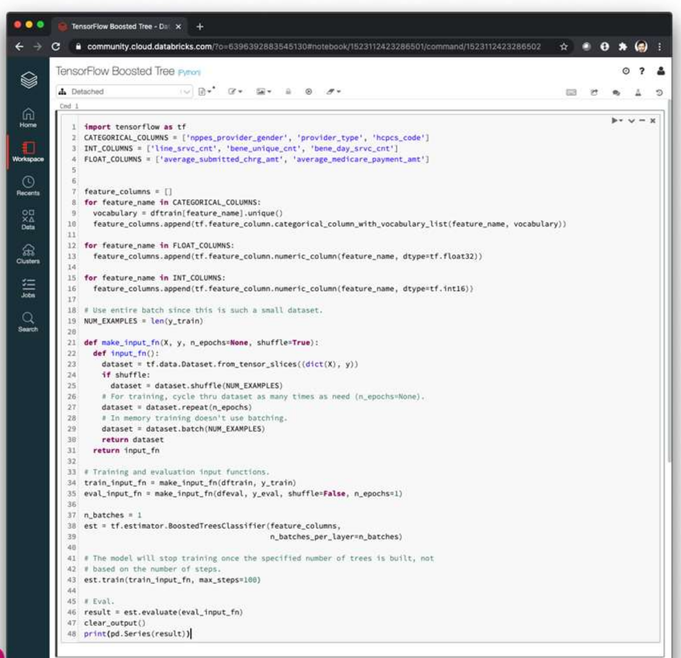
Random Under Sampling (RUS)



Human-Centered Design
Center of Excellence

CCSQ WORLD USABILITY DAY 27

27



Using 50:50 Class Distribution

TensorFlow Boosted Tree Classifier

Accuracy: 0.760789

Precision: 0.706313

Recall: 0.857778

AUC: 0.845611

Center of Excellence

CCSQ WORLD USABILITY DAY 28

28

Potential Improvement

- Add back Geographical information to the data set in analysis
- Add beneficiary data to form a graph analysis. Right now we only analyze from Provide side
- More granular e.g. by type
- Add more metrics (Medicare Standard Amount, Medicare Allowed Amount)
- A lot of missing NPI in LEIE. looking up missing NPI numbers in the National Plan and Provider Enumeration System (NPPES) registry

Human-Centered Design Center of Excellence

CCSQ WORLD USABILITY DAY 29

29

Even with those improvements, there are still limitations

- Tagged data not always be available
- Not good for emerging anomalies with entirely new and more sophisticated forms

Human-Centered Design Center of Excellence

CCSQ WORLD USABILITY DAY 30

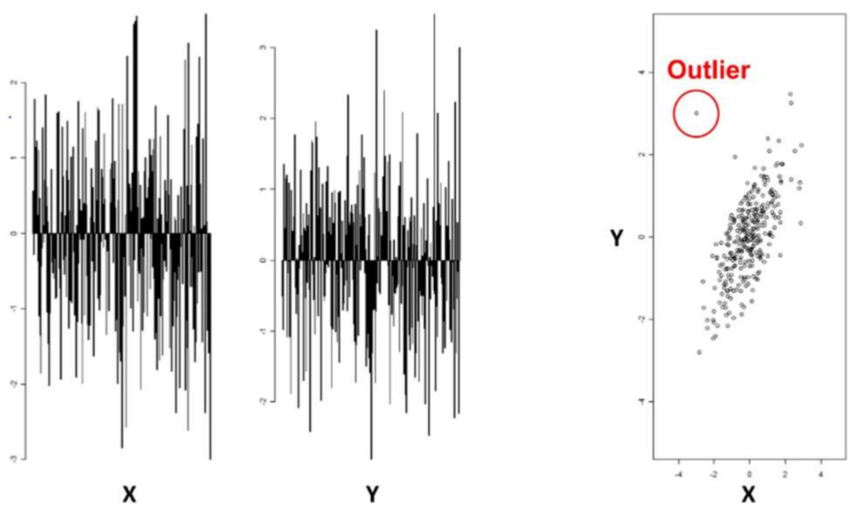
30

Unsupervised Learning

- Good for detecting Outlier, but it doesn't mean that is Fraud.
- It provides hint to start finding Fraud.

31

Where is the Outlier?



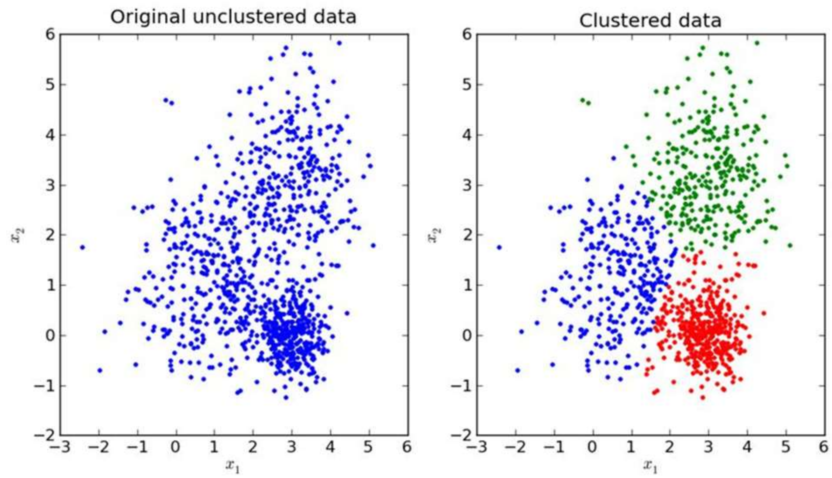
Human-Centered Design Center of Excellence

CCSQ WORLD USABILITY DAY 32

https://miro.medium.com/max/1944/1*BaOiTGWAoFi-3_-E-rJGdg.jpeg

32

Unsupervised Clustering



Human-Centered Design Center of Excellence

CCSQ WORLD USABILITY DAY 33

http://www.frankichamaki.com/data-driven-market-segmentation-more-effective-marketing-to-segments-using-ai/

33

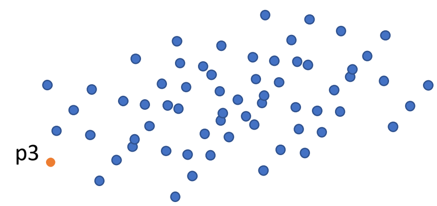
https://www.youtube.com/watch?v=BorcaCtjmog&t=4s

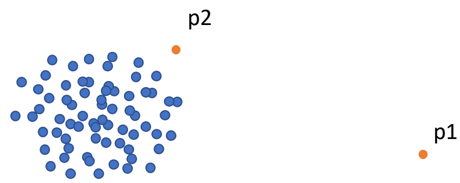
34

10 seconds Polling Question: Which point is considered outliner?

Nearest Neighbor Approach

- By distance
- Having the largest distance away from closest points
- Only p1 is considered outlier





Density based Approach

- By density
- Having the lowest density among closest points
- Both p1 and p2 are considered outlier

Human-Centered Design Center of Excellence

CCSQ WORLD USABILITY DAY 35

35

10 seconds Polling Question -3

You must know her for this 3rd approach

Who is she?

- She committed a $100M fraud in a European bank last year by guessing the admin password using AI algorithm
- She is an actually a man dressed in disguise to fool the airport security smuggling 500 fake passports (with valid passport numbers produced by AI) in US last year
- The person inventing this 3rd approach AI algorithm

http://stylegan.xyz /paper
https://github.com/NVlabs/stylegan

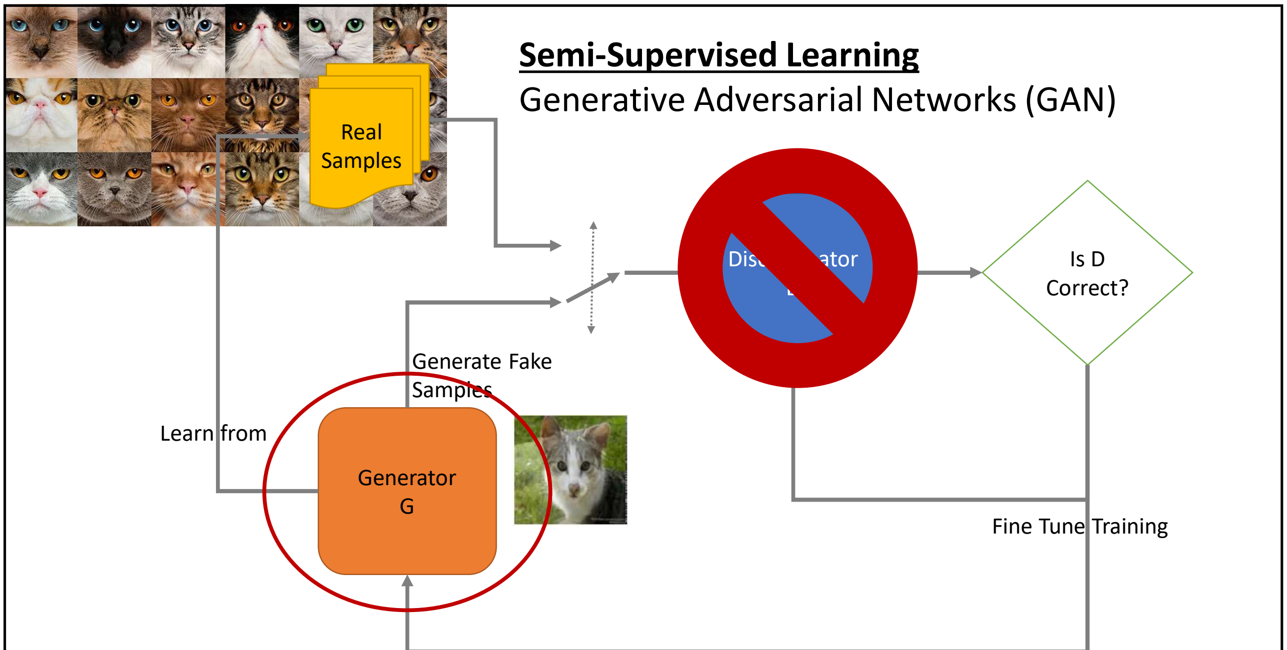
Human-Centered Design Center of Excellence

CCSQ WORLD USABILITY DAY 36

36

Semi-Supervised Learning

Generative Adversarial Networks (GAN)



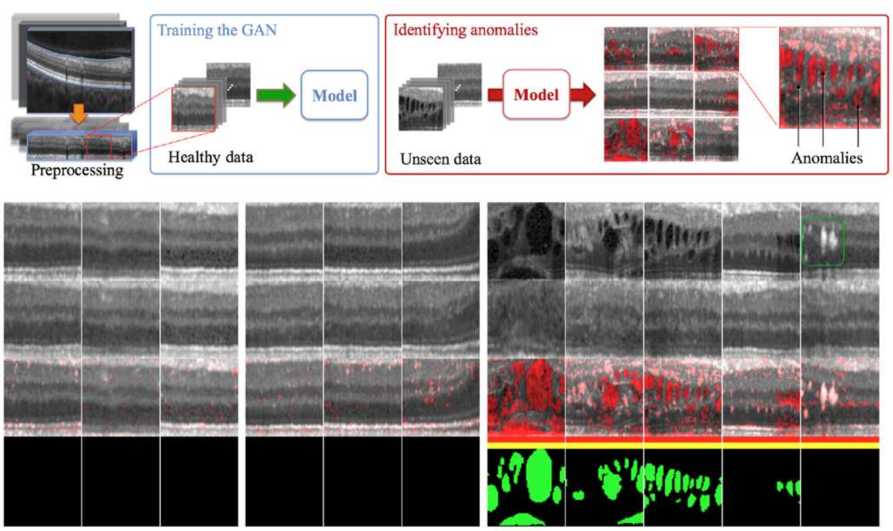
Human-Centered Design Center of Excellence

https://www.youtube.com/watch?v=pJBS5SEmIMc
https://www.27machines.com/machine-learning-basics/

CCSQ WORLD USABILITY DAY 37

37

AnoGAN



Human-Centered Design Center of Excellence

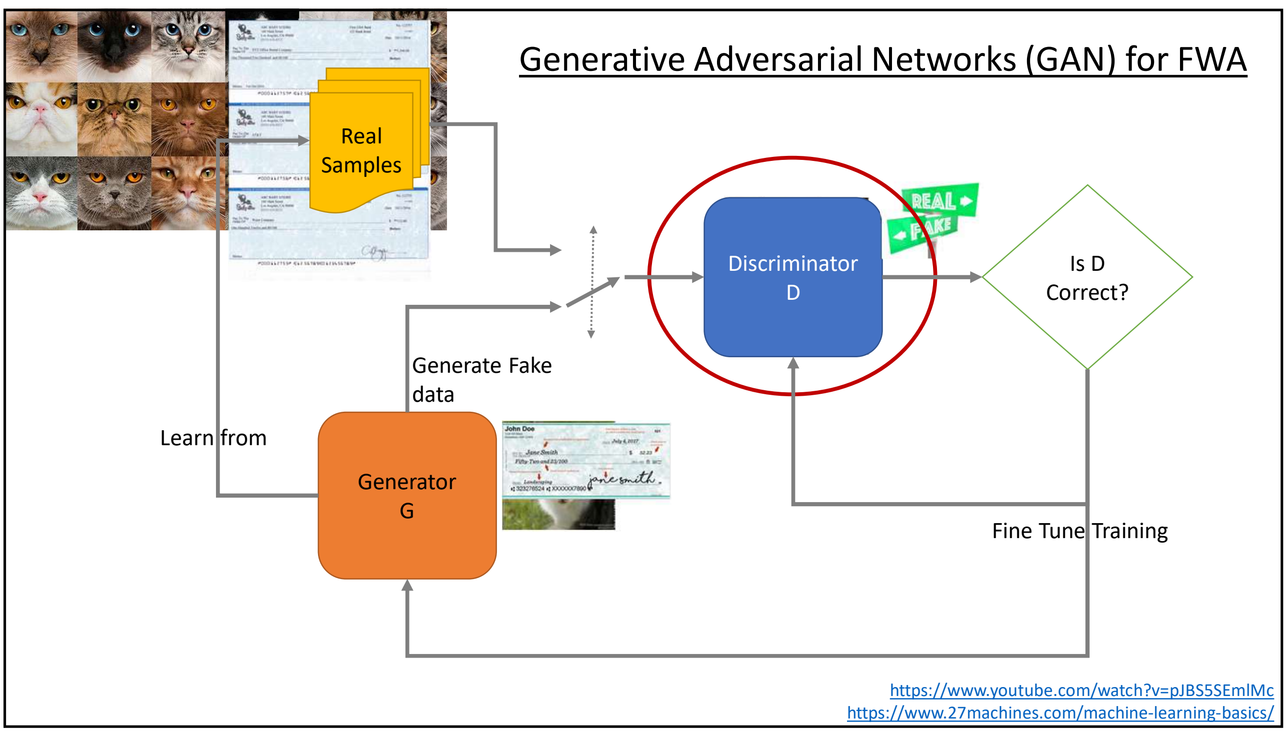
Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery: Thomas Schlegl, Philipp Seeb ock, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, and Georg Langs (Medical University Vienna, Austria) http://arxiv.org/abs/1703.05921

CCSQ WORLD USABILITY DAY 38

38

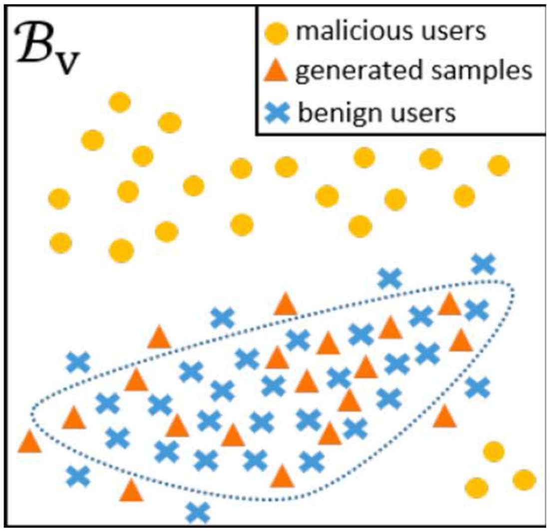

But we are talking about FWA, right?!

39


Generative Adversarial Networks (GAN) for FWA
Real Samples
Discriminator D
REAL
FAKE
Is D Correct?
Generate Fake data
Learn from
Generator G
John Doe
Jane Smith
Fine Tune Training
https://www.youtube.com/watch?v=pJBS5SEmIMc
https://www.27machines.com/machine-learning-basics/

40

Traditional GAN



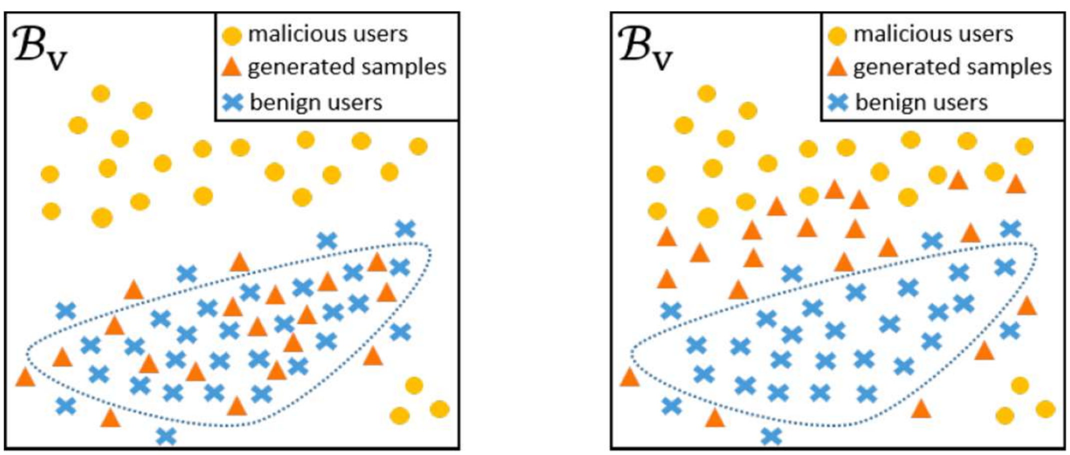
Human-Centered Design Center of Excellence

CCSQ WORLD USABILITY DAY 41

https://arxiv.org/pdf/1803.01798.pdf

41

One-Class Adversarial Nets (OCAN) GAN



(a) Regular GAN **(b) Complementary GAN**

Human-Centered Design Center of Excellence

CCSQ WORLD USABILITY DAY 42

https://arxiv.org/pdf/1803.01798.pdf

42

Advantage of One-Class Adversarial Nets

- No need for fraud data
  - No need to manually prepare a mixed training data set, which is usually has a very few fraud data to start with
- Discriminator will take in either real benign or generated malicious
  - More adaptive to different kinds of malicious behavior
- Adapt to newly emerged normal user pattern

Human-Centered Design Center of Excellence

CCSQ WORLD USABILITY DAY 43

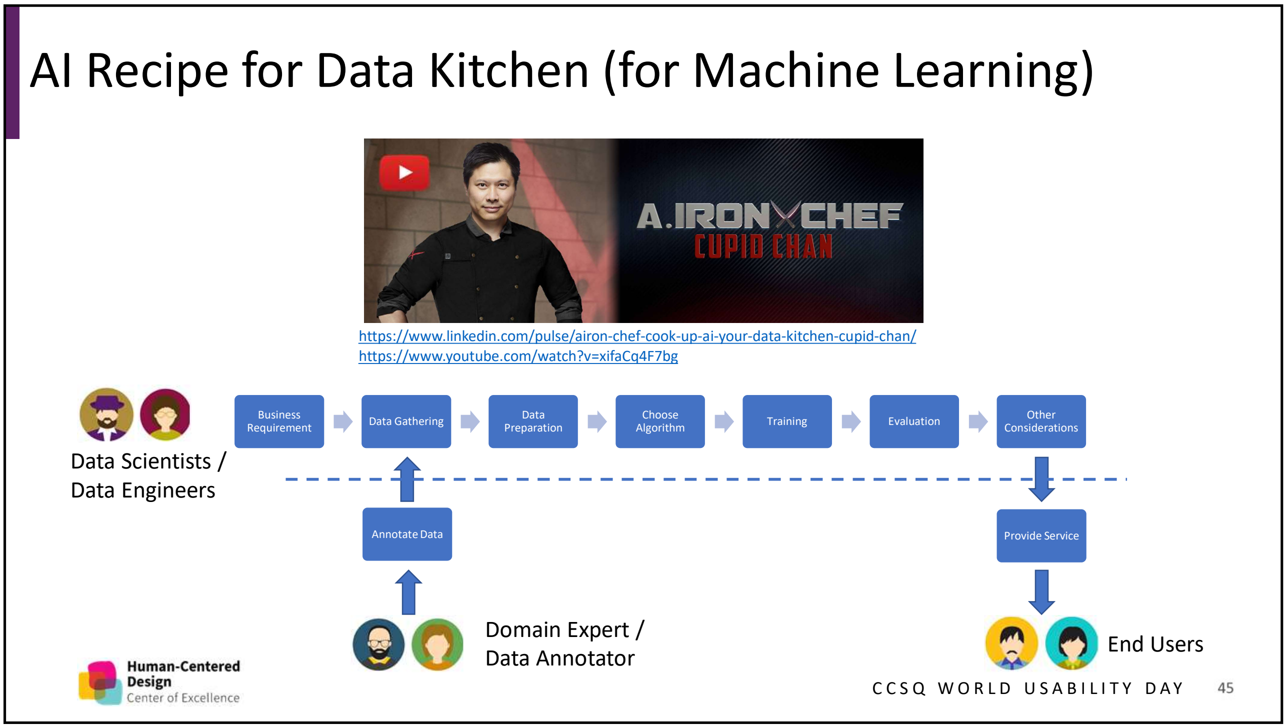
43

Recap: What we have talked about so far…

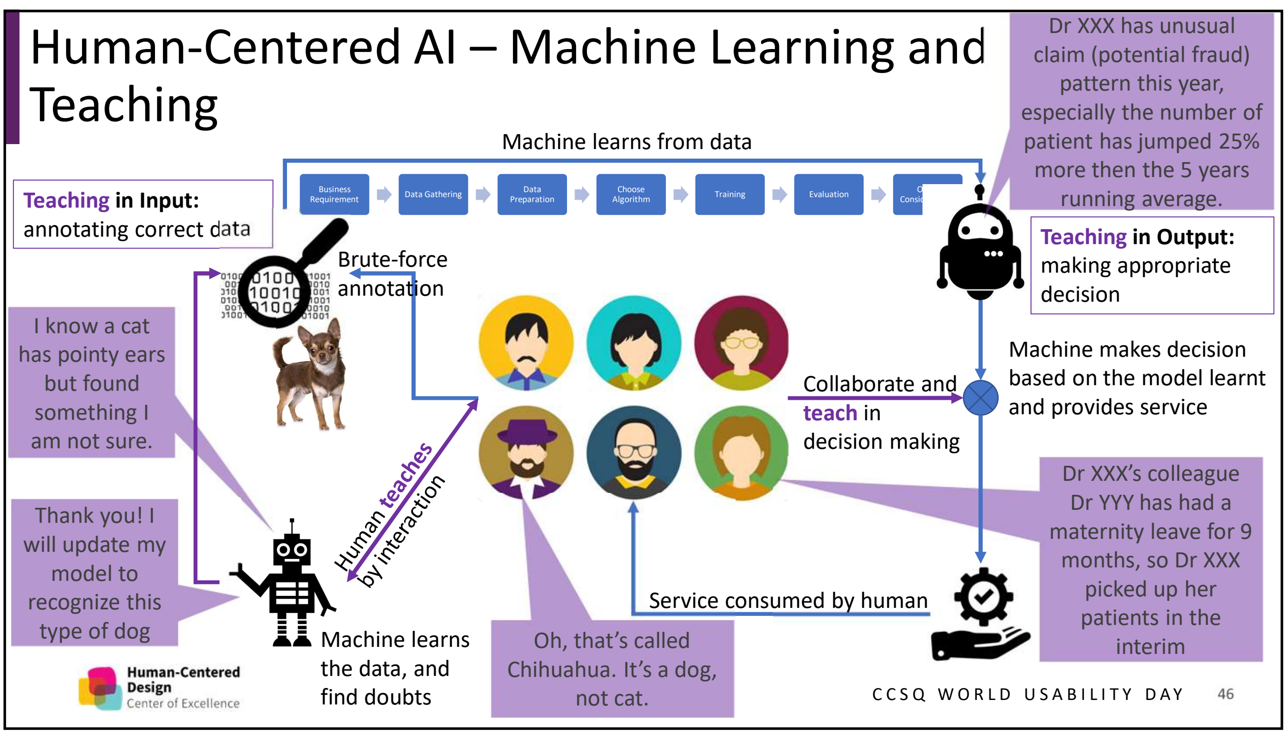
Human-Centered Design Center of Excellence

CCSQ WORLD USABILITY DAY 44

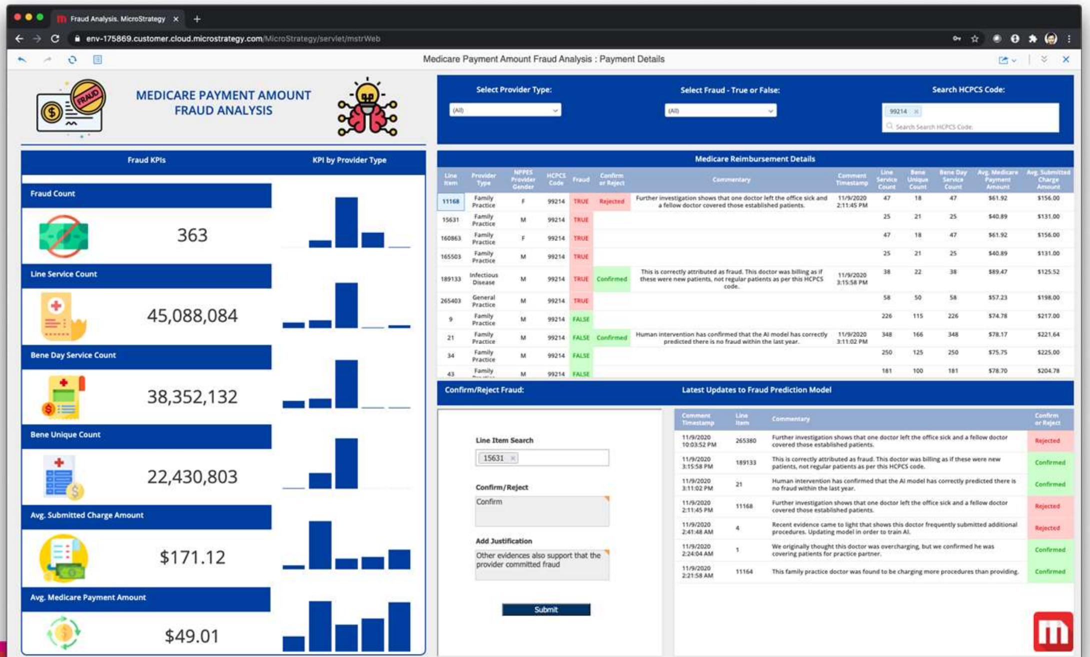
44


AI Recipe for Data Kitchen (for Machine Learning)
A.IRON CHEF
CUPID CHAN
https://www.linkedin.com/pulse/airon-chef-cook-up-ai-your-data-kitchen-cupid-chan/
https://www.youtube.com/watch?v=xifaCq4F7bg
Data Scientists / Data Engineers
Business Requirement
Data Gathering
Data Preparation
Choose Algorithm
Training
Evaluation
Other Considerations
Annotate Data
Provide Service
Domain Expert / Data Annotator
End Users
Human-Centered Design Center of Excellence
CCSQ WORLD USABILITY DAY 45

45


Human-Centered AI – Machine Learning and Teaching
Dr XXX has unusual claim (potential fraud) pattern this year, especially the number of patient has jumped 25% more then the 5 years running average.
Machine learns from data
Business Requirement
Data Gathering
Data Preparation
Choose Algorithm
Training
Evaluation
Teaching in Input: annotating correct data
Teaching in Output: making appropriate decision
Brute-force annotation
I know a cat has pointy ears but found something I am not sure.
Machine makes decision based on the model learnt and provides service
Collaborate and teach in decision making
Human teaches by interaction
Thank you! I will update my model to recognize this type of dog
Dr XXX's colleague Dr YYY has had a maternity leave for 9 months, so Dr XXX picked up her patients in the interim
Service consumed by human
Machine learns the data, and find doubts
Oh, that's called Chihuahua. It's a dog, not cat.
Human-Centered Design Center of Excellence
CCSQ WORLD USABILITY DAY 46

46



Center of Excellence

CCSQ WORLD USABILITY DAY 47

47

When you talk about AI, think of IA!

Artificial Intelligence

VS

Intelligence Augmentation

In Speaker Packet: *There is zero tolerance for sales pitches masquerading as educational or informational.*

Human-Centered Design Center of Excellence

CCSQ WORLD USABILITY DAY 48

48

AI + BI = CI

AI (Artificial Intelligence) – Excellence in learning with speed

BI (Business Intelligence) – Historically proven to enable human intuited the direction

CI – (Cognitive Intelligence) - Intelligence by combining the Speed of how a Machine Learn and Direction Intuited from Human Insight


2017 CONFERENCE ON HEALTH IT AND ANALYTICS (CHITA)

CHITA
CONFERENCE ON HEALTH IT AND ANALYTICS

November 3rd – 4th 2017 | Washington, DC
Presented by the Center for Health Information & Decision Systems

The Conference on Health IT and Analytics (previously known as the Workshop on Health IT & Economics) is an annual health IT and analytics research summit, including a doctoral consortium that each year gathers prominent scholars from more than 40 research institutes, and leading policy and practitioner attendees in a vibrant setting to discuss opportunities and challenges in the design, implementation and management of health information technology and analytics. Its goal is to deepen our understanding of strategy, policy and systems fostering health IT and analytics effective use and to stimulate new ideas with both policy and business implications.

This forum provides a productive venue to facilitate interaction and collaboration among academia, government, and industry. Now in its 8th year, each year CHITA draws over 100 participants.

Hosted by the Center for Health Information & Decision Systems (CHIDS), support for CHITA is provided by the Robert H. Smith School of Business and the University of Michigan School of Public Health.

We hope that you will join us for this engaging, stimulating and fun event!

FEATURED SPEAKERS

Martin "Marty" Makary
Chief, Islet Transplant Surgery, Professor, NYT best-selling author
Johns Hopkins University School of Medicine

Warren D. D'Souza
Vice President, Enterprise Data and Analytics
University of Maryland Medical System

FEATURED PANELISTS

Cupid Chan
Chief Technology Officer
Index Analytics

Rema Padman
Professor of Management Science and Healthcare Informatics
Carnegie Mellon University

Anna T. Fernandez
Senior Associate, Health Informatics
Booz Allen Hamilton

Charles Gabrial
Analytics & Data Science Manager
Defense Health Agency

Hoo Chang Shin
Solutions Architect
NVIDIA

Hong Yu
Professor of Quantitative Health Sciences

Human-Centered Design
Center of Excellence

49

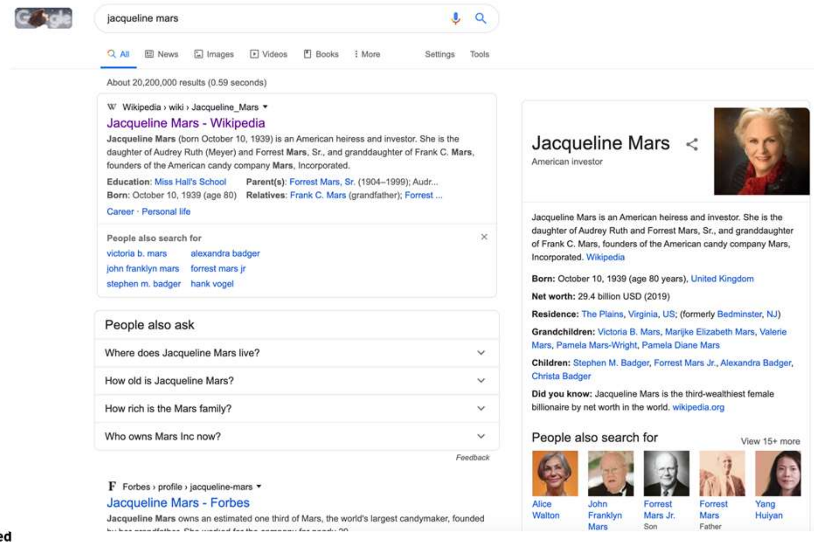
… LinkedIn Story continues

Human-Centered Design
Center of Excellence

CCSQ WORLD USABILITY DAY 50

50

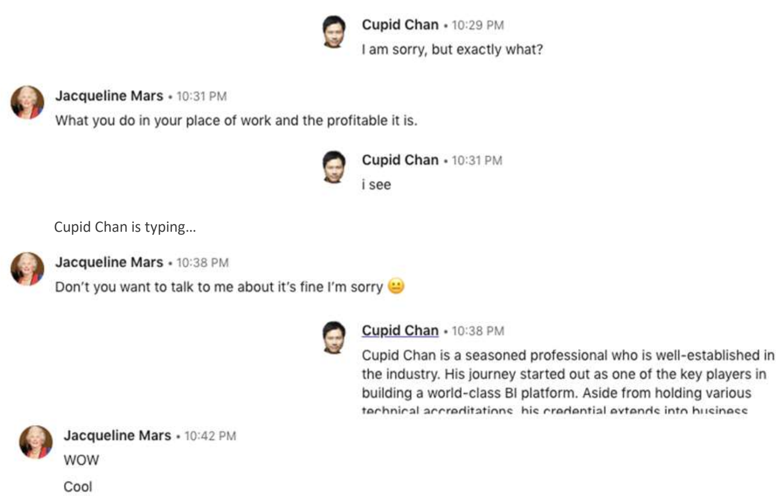
I decided to find my most knowledgeable friend for help



Human-Centered Design
Center of Excellence

CCSQ WORLD USABILITY DAY 51

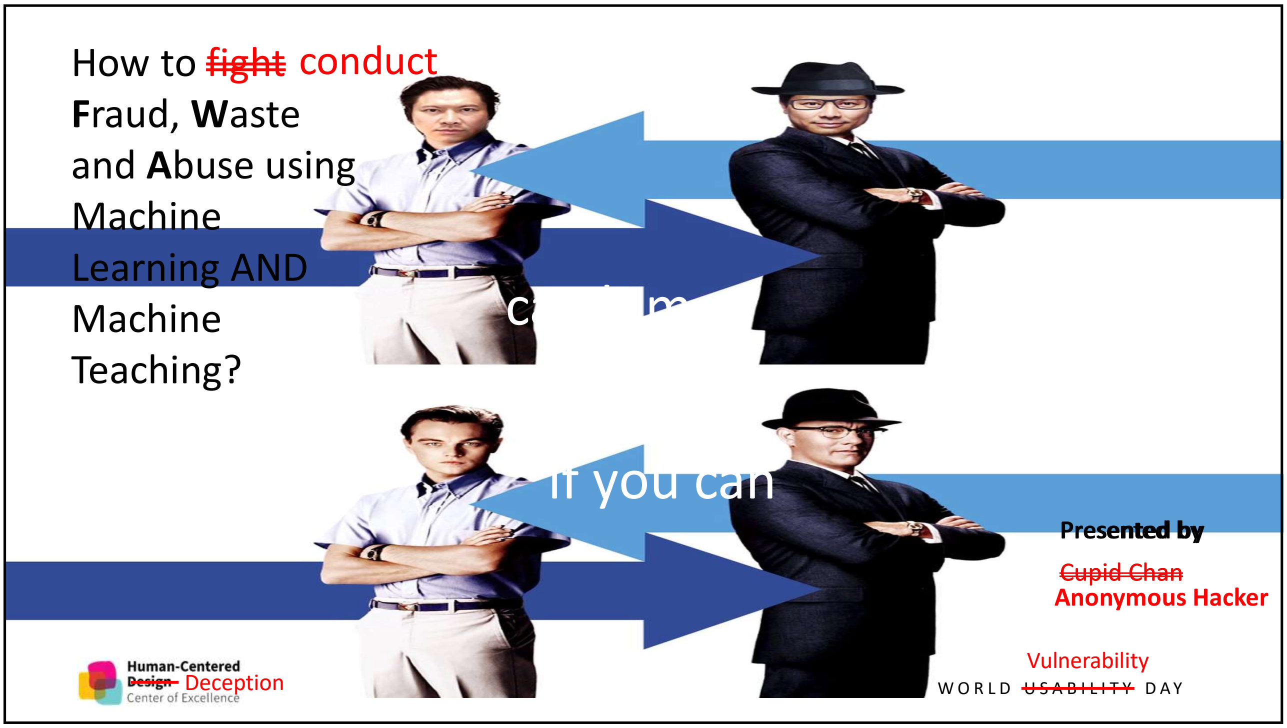
51



52

53

54



55

Cupid Chan

- Board of Directors and Technical Steering Committee, Chairperson of BI & AI Project, Linux Foundation ODPi
- Senior Fellow & Adjunct Professor, University of Maryland College Park

www.linkedin.com/in/cupidchan/
@cupidckchan

56