

# Data, Machines and ... People!

Enrico Bertini  
Associate Professor



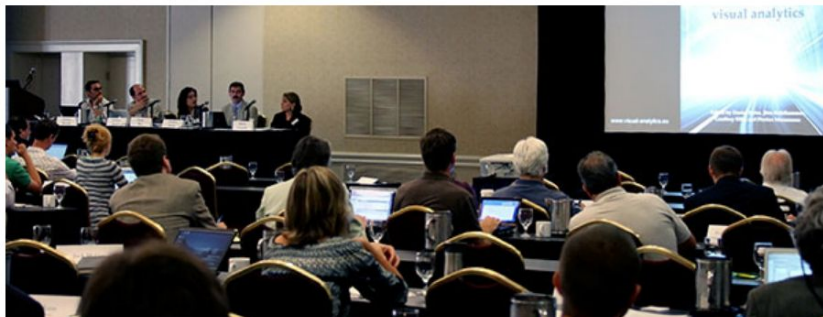
**NYU**

**TANDON SCHOOL  
OF ENGINEERING**

How many of you have heard the term  
“*Data Science*” before?



## VAC Consortium 2010 Meeting

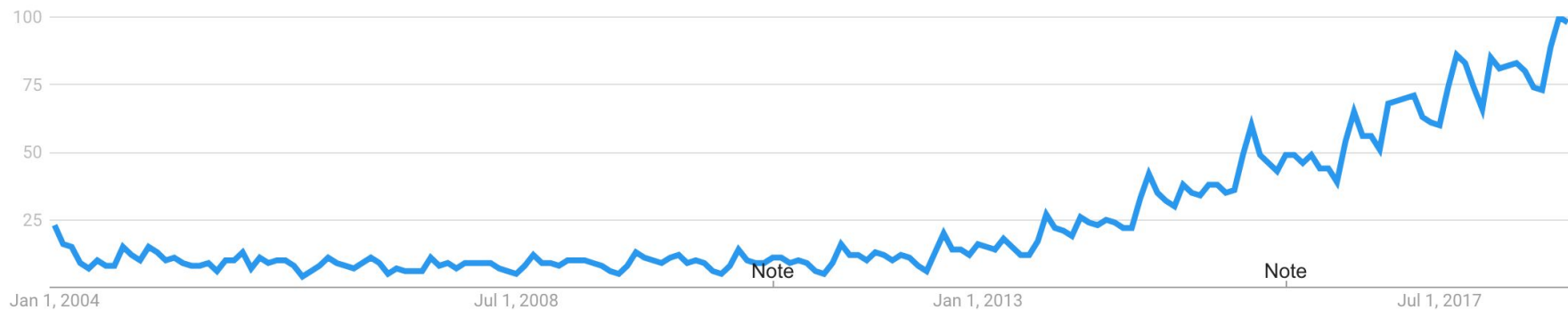


More than 150 participants from academia, government, and industry attended the Visual Analytics Community Consortium Meeting August 31-September 1, 2010 at the University of Maryland. The theme of this year's meeting was "Broadening Our Horizons." Dr. Jennifer Leaning kicked off the first day with a keynote address, which was followed by a panel discussion on Global Crisis Management that focused on the information challenges of the Haiti earthquake, the Gulf of Mexico oil spill, and other recent events. In another first-day session, NVAC leaders introduced the new community website, [vacommunity.org](#); described new community resources being developed; and elicited feedback from attendees.

Day two kicked off with keynote addresses from Joe Kielman, who described new applications for visual analytics, and Ben Shneiderman, who spoke about "Understanding Innovation Trajectories for Visual Analytics." These addresses were followed by a panel discussion about the use of interactive visualization and visual analytics in consumer applications. Representatives from UKVAC, NVAC, and Defence R&D Canada described current research efforts in the afternoon. The day concluded with panel discussions on the uses and opportunities for visual analytic technologies in a variety of fields.

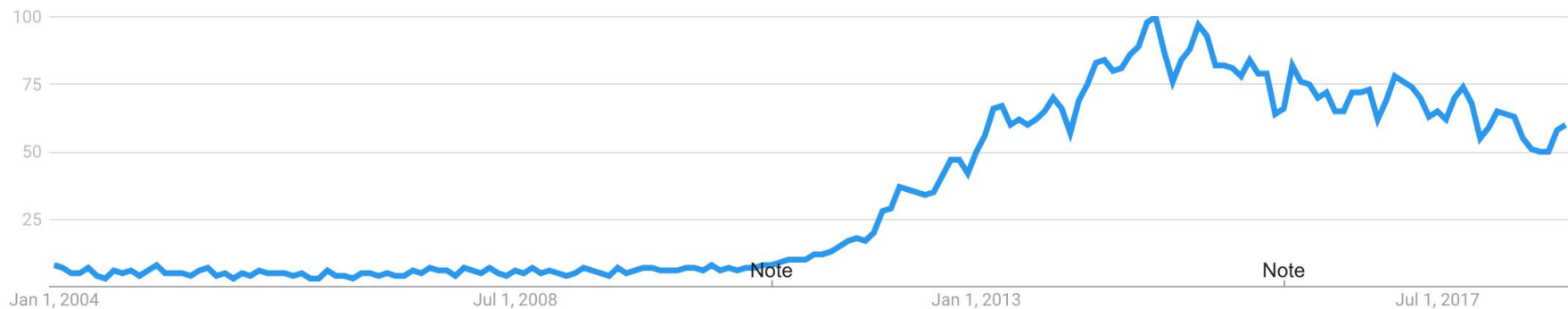
# Data Science

Interest over time 



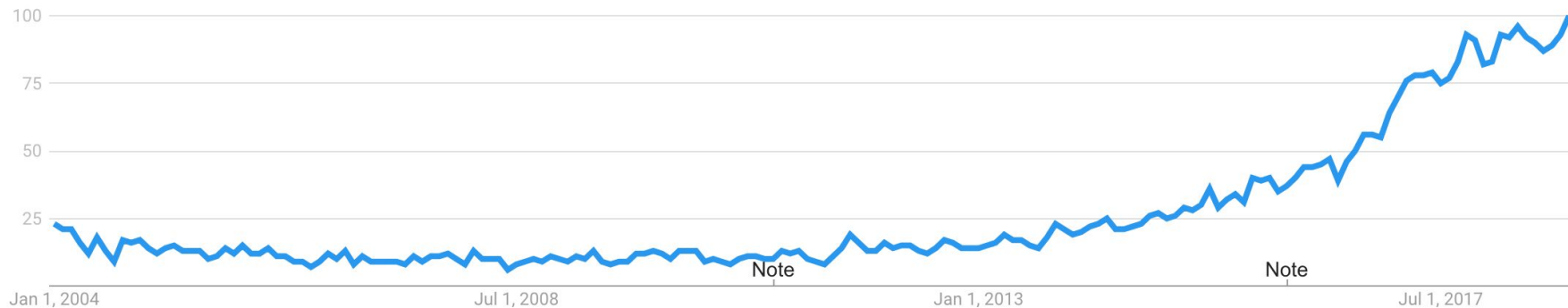
# Big Data

Interest over time 



# Machine Learning

Interest over time 

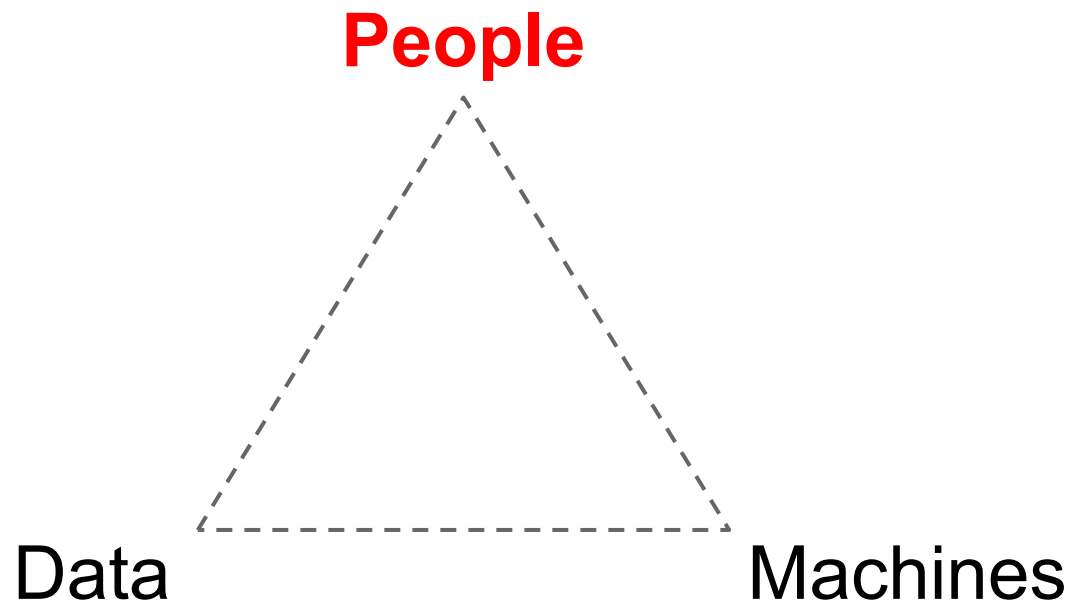


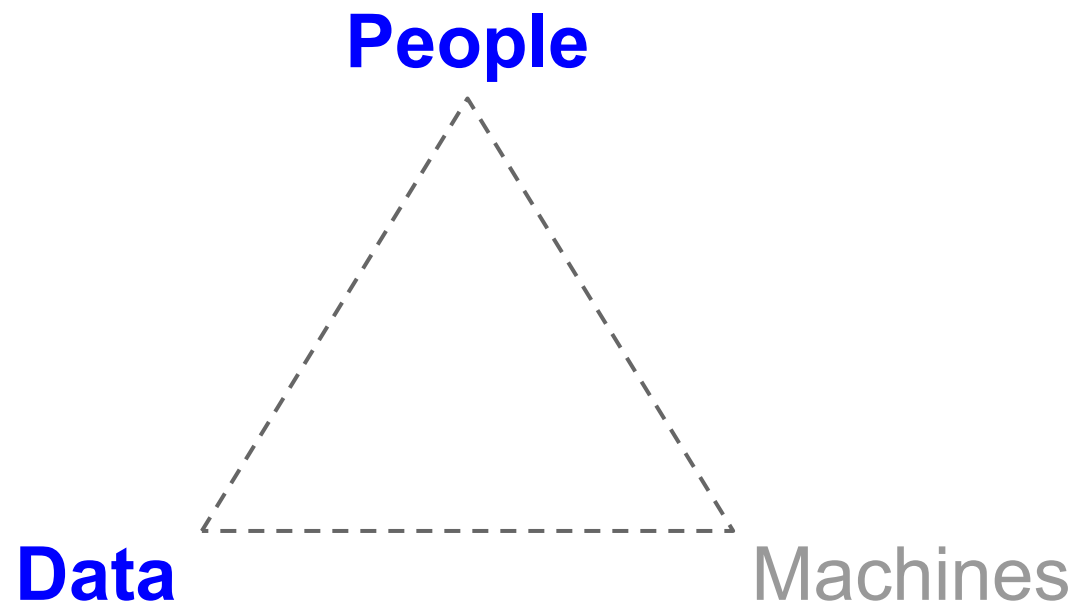
Why are we so excited?

What **can** we achieve?

What do we **want** to achieve?







# “Intelligence Amplification”

How can we use machine to help people do remarkable things with data?

# **The world is full of complexity!**

Biology

Finance

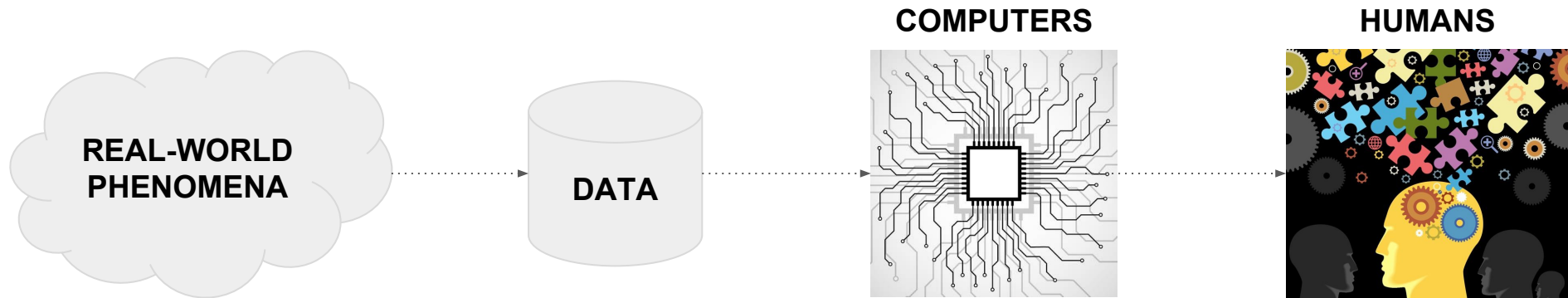
Social Media

Cyber Security

Urban Systems

...

If we are able to capture relevant information about phenomena through data, then maybe we can understand these phenomena better?



**We are not interested in data!**

We are interested in the **phenomena** described by data.  
And our ability to **understand** them better with data.

# **Illuminating the Path**

The Research and Development Agenda  
for  
Visual Analytics

Edited by James J. Thomas and Kristin A. Cook

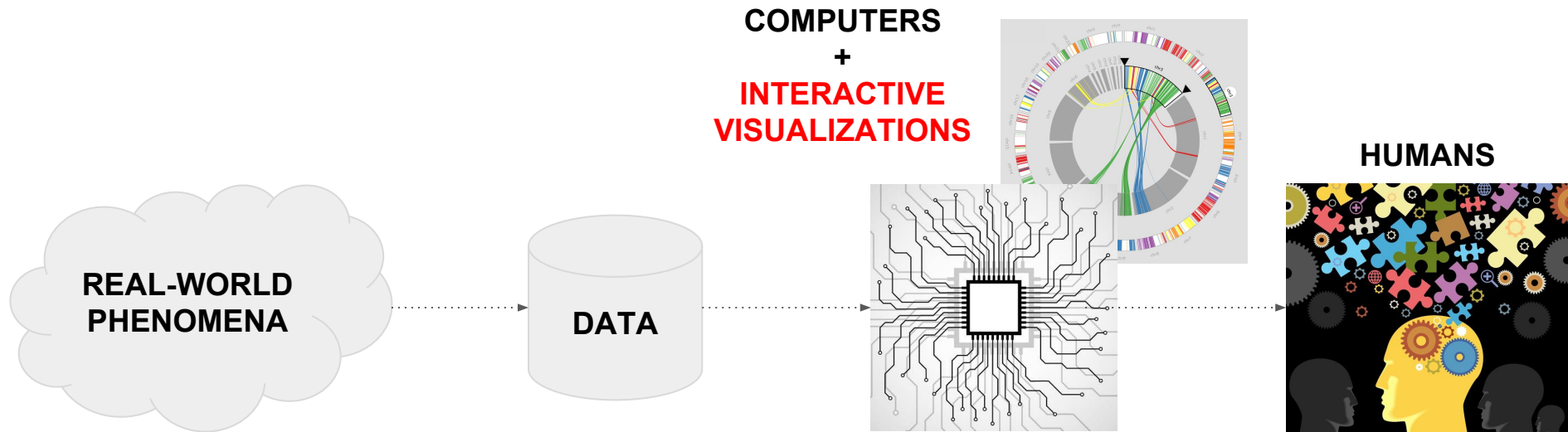


National Visualization and Analytics Center™



## **Visual Analytics**

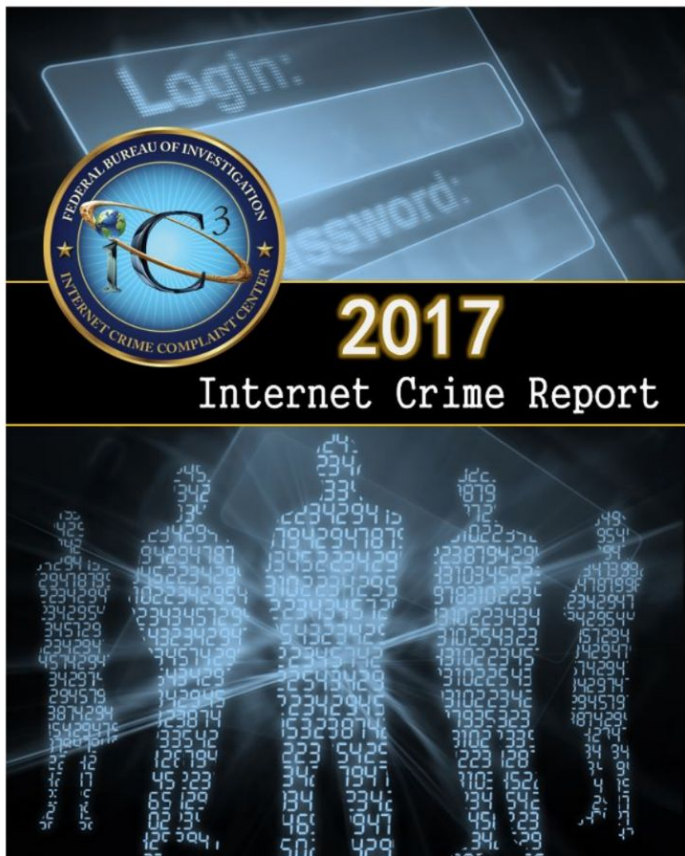
*“The science of analytical reasoning supported by interactive visual interfaces”*



**Visual perception** is the most effective channels we have to obtain information from the world.



# Example: Help Investigators Fight Scammers



## Top 10 Crime Types Reported to IC3 in 2017 (by Victim Loss)



From the 2017 Internet Crime Report

## Contact Graph

2326664144@domain.com	10
2042564431@domain.com	10
1126249745@domain.com	9
2094334163@domain.com	9
2414733712@domain.com	9
2974993531@domain.com	9
709917923@domain.com	9
2661471402@domain.com	9
4102881722@domain.com	9
3214692907@domain.com	9
2177946313@domain.com	9
217150053@domain.com	9
26983959@domain.com	9
1505032305@domain.com	8
Barry	8
1954659885@domain.com	8
2120310937@domain.com	8
Sheldrake	8
3180820129@domain.com	8
1196705573@domain.com	8
1877894974@domain.com	8
1763919603@domain.com	8
4064748577@domain.com	8
393679104@domain.com	8
4274787355@domain.com	8
1801780542@domain.com	8
3471838824@domain.com	8

## Entities

Contents		
meet	address	st
send	fun	rent
boytee6	clark	receive
rev	spam	online
international	woman	original
scam	craigslist.org	flag
mf	relay	unwanted
service	site	hey
reply.craigslist	support	chat
comment	tracking	cid
old	policy	qiw
limited	term	privacy
free	check	love
single	find	road
year	number	pic
dg	sale	member
picture	interested	victoria
move	building	kitt
basseterre	revi	item
photo	customer	girl
thing	notification	error
phone	add	longer
membership	program	condition
follow	hous.craigslist	list
receiving	understand	info
waiting	dec	electronic
hear	friend	team
reply	nov	opt
visit	received	talk
agree	private	update
open	guy	saint
flirt	room	newsletter

## Entities

Subject		
profile	meet	jan
asap	woman	find
chat	6	private
local	dec	7
im	2	guy
dating	girl	sex
unlocked	400	hey
hookup	hot	xdating
500	hello	3
Women	man	matche
place	16	fun
friend	request	apartment
check	samsung	attention
card	house	received
sexy	wanted	mom
1	bs	today
50	ben	boytee
300	brand	free
galaxy	hoty	love
needed	tonight	week
deal	bedroom	will
800	gift	gold
imp	enjoy	selected
area	black	good
time	5	apple
phone	top	waiting
350	cbd	clinical
ga	member	older
register	trial	online
best	jacksonville	russian
silver	yeah	younger
23	big	sale
wanna	a	a

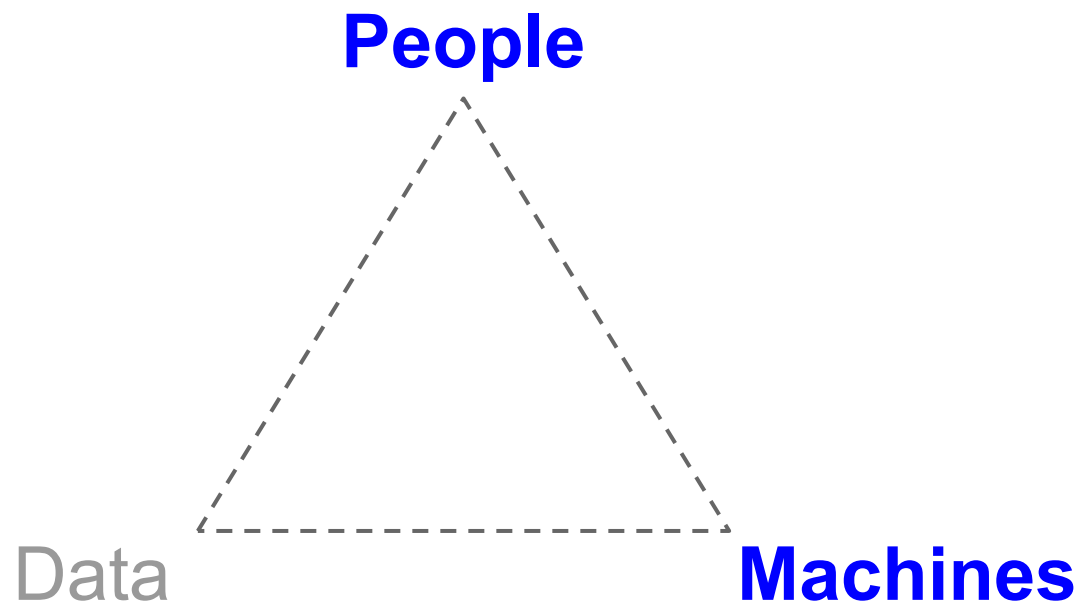
## Communications

Jan 29, 2018 MINT iPhone 6S UNLOCKED 16GB - \$230 Do you still have this item for sale?
Jan 29, 2018 Your ad needs one more step to be published: Brand new unlocked Apple iPhone 7 Your ad needs one more step to be published: Brand new unlocked Apple iPhone 7 plus Your ad needs one more step to be published This message refers to the ad: "Brand new unlocked Apple iPhone 7 plus" of ID . Hey OLXer, Unfortunately, your Ad doesn't meet the OLX requirements and we are unable to publish it because of the following reason: "Your
Jan 29, 2018 Meetups this week with: Financial Modeler, Agile Nigeria Meetup Members, Grinder Top suggestions for you this week Check out a few highlights based on your current Meetup groups and interests. 1 How to become an Advanced Financial Modeler:
Jan 29, 2018 Welcome on OLX Kuwait Welcome on OLX Kuwait? utm_source= * Hi, * Glad to have you with us! We have some useful information for you below, which is worth reading Useful options My
Jan 29, 2018 Activate account Activate account Hi, if you want to complete the process registration, click on the button below. Click below to activate account: * Activate an account * To further enhance the results of your ad visit our pro-tips and safety rules Useful Links: Help I Contact us I My Account This message is generated automatically, please do not reply to this.
Jan 29, 2018 Re: Like new 32gb Rose Gold 6s+ - \$350 (Windsor) Am getting this for myself and I will be paying you through PayPal cos I wont be able to come and pick it up I will offer you \$900 Including shipping fee.so get back to me with your PayPal email address so that I can make the payment now On Mon, Jan 29, 2018 at 1:36 AM, craigslist
Jan 29, 2018 Re: Iphone 6 Plus Unlock 16GB - \$200 (meriden) Am getting this for myself and I will be paying you through PayPal cos I wont be able to come and pick it up I will offer you \$500 Including shipping fee.so get back to me with your PayPal email address so that I can make the payment now On Mon, Jan 29, 2018 at 12:10 AM, craigslist
Jan 29, 2018 FG suspends tax officials 'collecting bribe' from whistle-blowers Your daily news update is waiting for you. Come visit Pulse.ng If you believe this has been sent to you in error, please click to safely unsubscribe.
Jan 29, 2018 Are You Getting Ready For The Marathon? Do you like it when a girl takes it nice and slow, stripping away each piece of clothing so that you can savor every moment? During our Striptease Promo and Discount, we're celebrating our girls that cherish the art of the striptease and know how to make their audience wild with lust. Any model who earns at least 10K credits, 20K credits, or 30K credits today wins \$100, \$200 or
Jan 29, 2018 Iphone 6 Plus Unlock 16GB - \$200 (meriden) Do you still have this item for sale?

Start Date : 2015-03-01

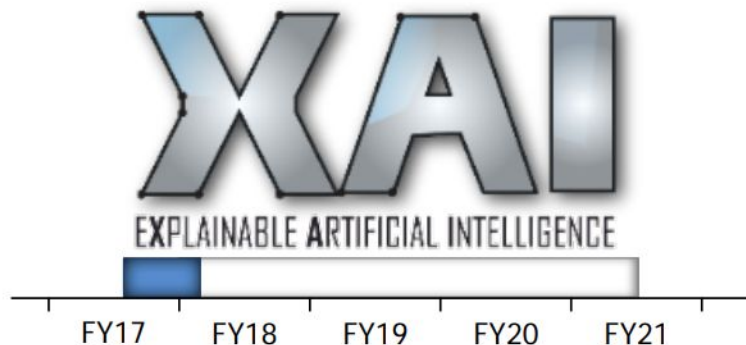
Interval: MONTH





As more automated decision affect our lives, how do we retain the power to influence, verify and understand these decisions?

# Explainable Artificial Intelligence (XAI)



David Gunning

DARPA/I2O

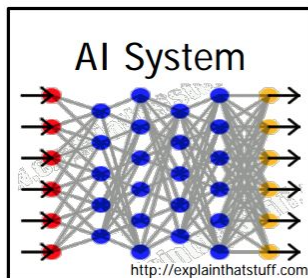
Program Update November 2017







# The Need for Explainable AI



- We are entering a new age of AI applications
- Machine learning is the core technology
- Machine learning models are opaque, non-intuitive, and difficult for people to understand



- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

Machine learning is only as good as the  
data we feed to them ...

# Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

*by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica*

May 23, 2016

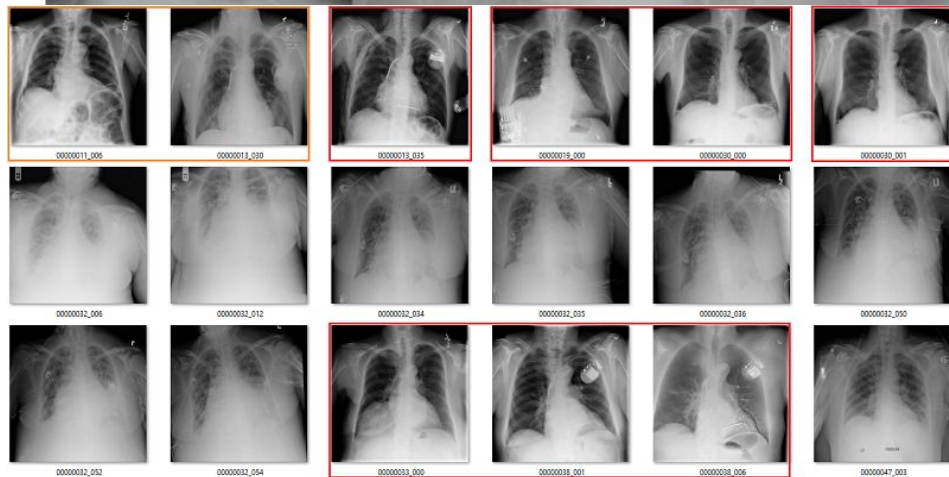
ON A SPRING AFTERNOON IN 2014, Brisha Borden was running late to pick up her god-sister from school when she spotted an unlocked kid's blue Huffy bicycle and a silver Razor scooter. Borden and a friend grabbed the bike and scooter and tried to ride them down the street in the Fort Lauderdale suburb of Coral Springs.

Just as the 18-year-old girls were realizing they were too big for the tiny conveyances — which belonged to a 6-year-old boy — a woman came running after them saying, “That’s my kid’s stuff.” Borden and her friend immediately dropped the bike and scooter and walked away.





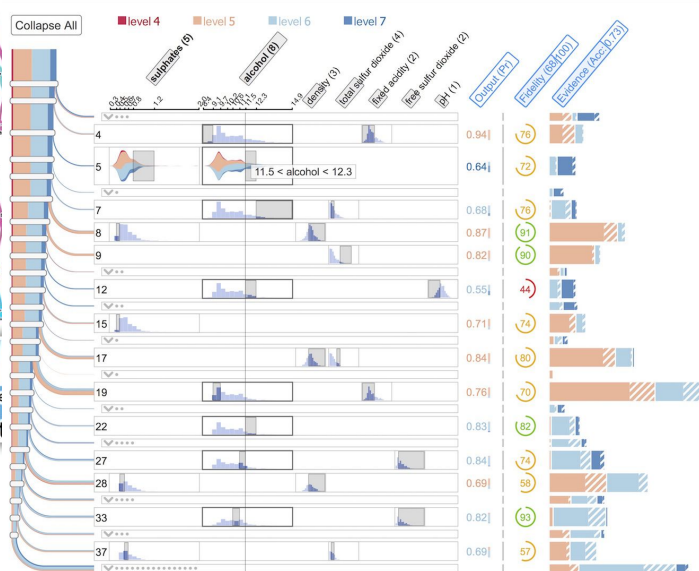
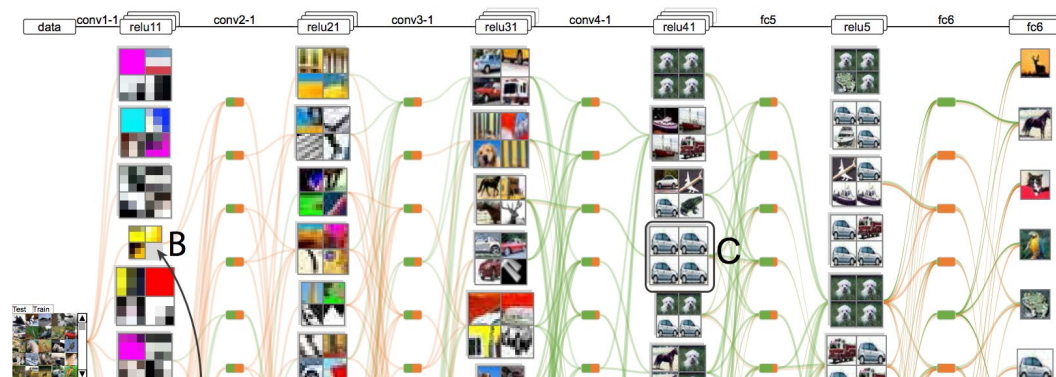
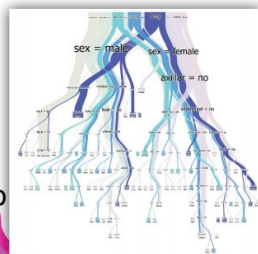
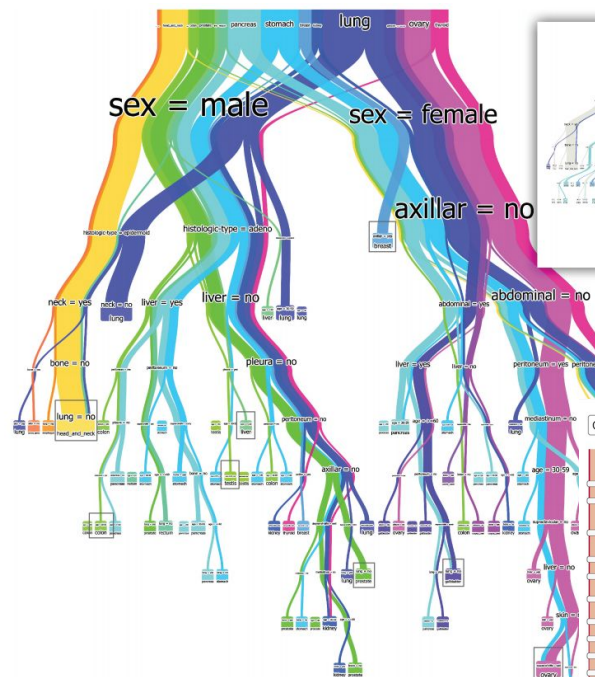
# Exploring the ChestXray14 dataset: problems



Label	PPV (visual)	PPV (text mining)
Consolidation	35%	97%
Cardiomegaly	80%	97%
Pneumothorax	60%	89%
Pneumonia	35%	89%
Fibrosis	24%	92%
No finding	60%	NA

A couple of weeks ago, I mentioned I had some concerns about the ChestXray14 dataset. I said I would come back when I had more info, and since then I have been digging into the data. I've talked with Dr Summers via email a few times as well. Unfortunately, this exploration has only increased my concerns about the dataset.

We need methods and tools to help us  
understand and verify AI systems!



# Predicting Admission from ER

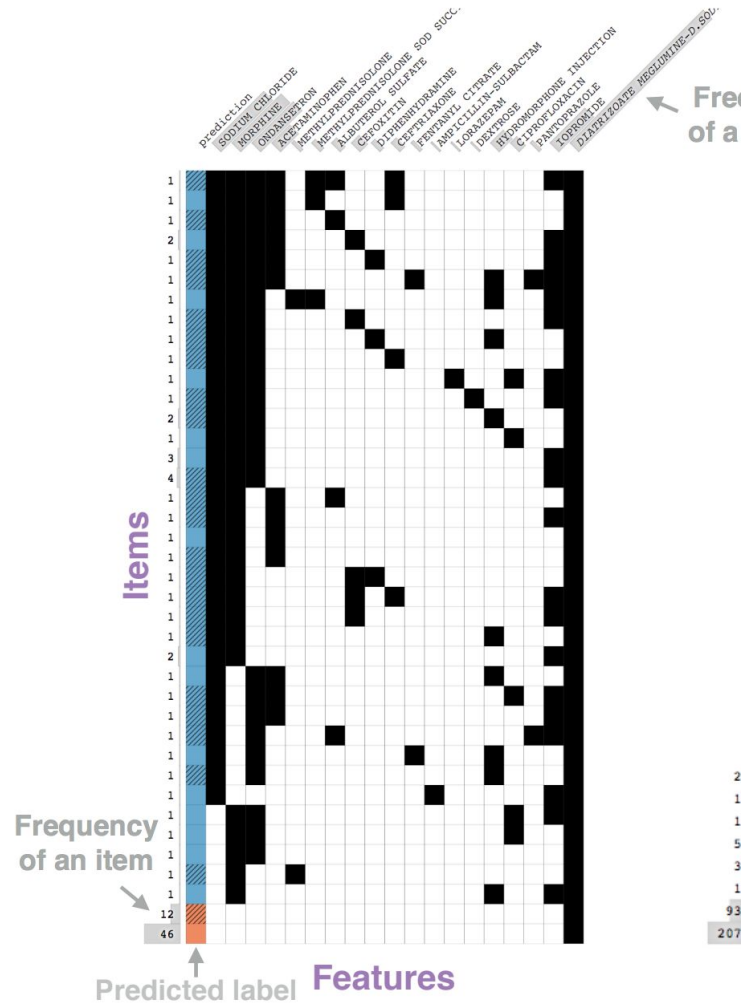


Will she/he be  
admitted?



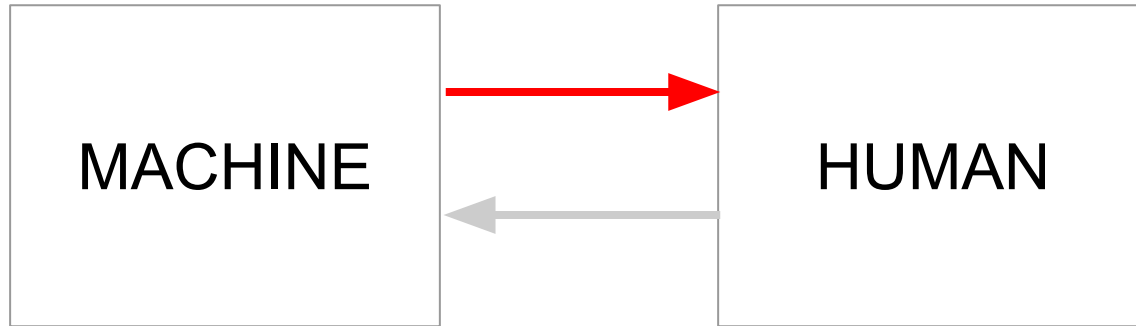
# Explainer: visual analytics to understand black-box ML models.

			tp	fp	fn	tn	pos / total		odds ratio			
			3187				1328 / 4784		0.1	1	10	100
SODIUM CHLORIDE	SODIUM CHLORIDE	SODIUM CHLORIDE	93		207		99	313				
		IBUPROFEN			218		13	231				
		VANCOMYCIN	136		49		136	185				
		ACETAMINOPHEN			137		22	159				
		ASPIRIN	63		68		63	131				
		OXYCODONE-ACET.			102		14	116				
		ONDANSETRON			101		14	115				
		DIA. MEGLUMINE.	19	26		46	31	103				
		KETOROLAC			84		13	97				
		MORPHINE	29		45		29	74				
ONDANSETRON	MORPHINE	SODIUM CHLORIDE	13		40		13	53				
		SODIUM CHLORIDE	19		27		19	46				
		FAMOTIDINE	9		36		9	45				
		FUROSEMIDE	36		8		36	44				
		DIA. MEGLUMINE.	25		18		25	43				
		LIDOCAINE			40		3	43				
		ONDANSETRON	13		28		13	41				
		SODIUM CHLORIDE	15		24		15	39				
		MORPHINE	10		27		10	37				
		DIPH	3		31		3	34				
SODIUM CHLORIDE	ONDANSETRON	FAMOTIDINE	4		27		4	31				
		SODIUM CHLORIDE			27		2	29				
		DIA. MEGLUMINE.	10		15		10	25				
		ALBUTE. SULFATE	10		15		10	25				
		ALBUTE. SULFATE	7		18		7	25				
		IPRATR. BROMIDE					0	25				
		METOCLOPRAMIDE					0	25				
		KETOROLAC			25		0	25				
		IBUPROFEN			25		0	25				



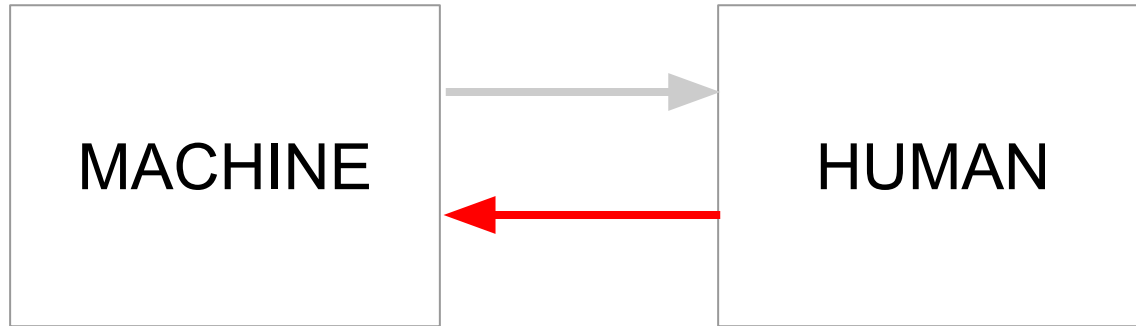
How can human + machine intelligence  
**cooperate** to solve problems together?

Machines provide the information  
we need





We communicate our knowledge  
and intentions to the machine

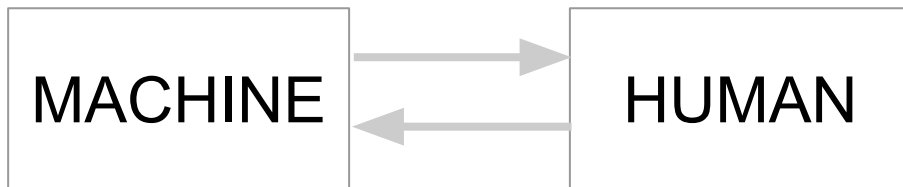




# Very Hard!

Output: hard to interpret/verify, uncertain.

Input: complex, ambiguous, small  
bandwidth.



Search

Coverage: 13966 19348

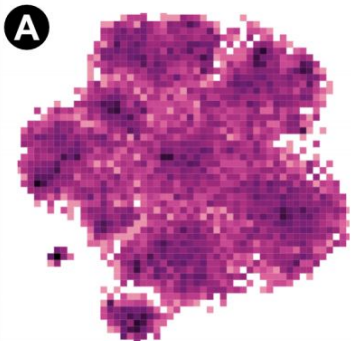
Text

Clear Filters

Add Label

Save

< Projection LDA Word Si >



0 24

car, sale, dealership

car sale dealership manager price  
told vehicle asked buy day

car, time, told

car time told change work will  
day oil tire service

car, great, repair

car great repair guy auto brought  
cost drive mechanic vehicle

car, repair, cost

<

law

B

Documents 1467 Exclusive 38%

Positive

lawyer 224 / 224

loan 89 / 89

legal 154 / 154

divorce 61 / 61

law 340 / 340

case 652 / 652

firm 266 / 266

court 258 / 258

accident 216 / 216

spa

Documents 1664 Exclusive 47%

nail spa pedicure massage facial

hotel

Documents 4492 Exclusive 41%

room stay stayed night resort  
hotel

apartment

Documents 772 Exclusive 42%

apartment rent tenant bedroom  
resident

haircuts

Documents 2048 Exclusive 48%

haircut salon hair cut barber  
cutting

tattoo

Documents 319 Exclusive 47%

artist tattoo talented ink

doctor (medical)

Documents 308 Exclusive 37%

(doctor OR dr) AND (practice OR prescription OR health)

Info

Intersection

Show:

ALL: All

ANY: All

NOT: All

Keywords

Discriminant

time recommend  
place experience  
great best  
service friendly  
will going  
good year  
staff work  
nice i'm  
day people  
well price

C

Documents: 19348

↑

im going to give this hotel casino a 2 star. the reason for my 2 star is... the hotel and casino and amenities and restaurants are all good. i had no complaints about that at all. my only gripe...

This hotel has beautiful, lush, green grounds; there were palm trees and other vegetation, some of which had lights that were on at night - so pretty. The grounds also had some wildlife.

D

Absolute worst service department. The only reason I gave them a second chance was purely out of convenience since I live down the street. They dropped my 2014 Honda off of a lif...

< 1 2 3 4 5 ... 194 >

>

Concluding ...

It's crucial to keep in mind the **role humans play**  
on the road to more automation.

# Challenges

Big Data → Big Understanding

How to make sense of machines.

How to cooperate with machines.

# Big Data