CUNY College of Staten Island
The City University of New York

Trust and Governance
in
the Intelligent Society

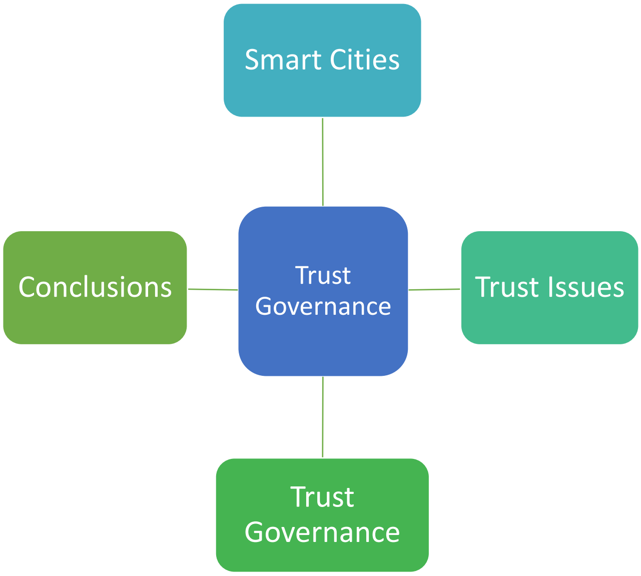
Soon Ae Chun
City University of New York
College of Staten Island & The Graduate Center

June 10, 2020

제7회 한미 지방행정 비교 포럼
Consulate General of Republic of Korea in New York

Agenda

- **Smart Cities**
- Trust Issues
  - Trust risks in data
  - Trust risks in algorithms
  - Trust risks in collaboration, Value Chain
  - Trust risks in connected devices
- Governance of Trust
- Conclusions



Smart Cities

Trust Governance

Conclusions

Trust Issues

Trust Governance

Smart Cities

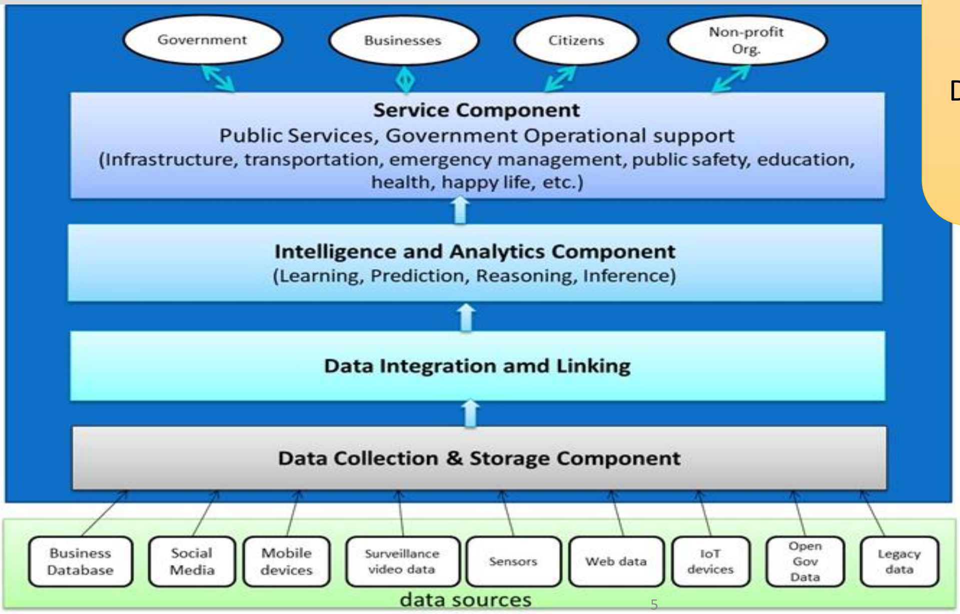
- A set of **applications and innovations** to make our cities
  - safer, more environmentally friendly and sustainable, and more efficient.
- Municipality that
  - **uses ICT** (Information and Communication Technologies)
  - to improve the quality of government services and citizen welfare.

Smart Cities – Innovations in Six Domains

- Smart Environment
  - Sensor technology, behavioral economics
  - Smart metering, smart grid, environmental monitoring
- Smart Mobility
  - Smart Transportation, Self driving cars, movements
  - Sensors provide a **real-time view of traffic flow**
  - inform transportation routing
- Smart Security
  - Smart cybersecurity to guard municipal datasets
  - Safe Neighborhoods
    - **Surveillance through drones**
    - Law enforcement officers use drones, wearable computing, facial-recognition, and predictive video
    - Crime predictions and prevention
- Smart Government/Education
  - Smart services, eco-systems, open data
  - Open platform for engagement, Collective Intelligence
  - virtual learning and augmented reality,
  - Blended, Adaptive, personalized learning
  - Life-long learning
- Smart Economy
  - Tech/tools to support business needs
  - Smart permits/licensing, workforce/talent nets, job training programs
- Smart Living/Health
  - Tools/techs for Smart health, smart connected home, healthcare, eldercare
  - Building systems (lighting, climate)
  - Monitoring systems for maintenance

Deloitte Report 2020 (source: Smart City | Smart Nation Providing the keys to unlock your city's potential)

Smart City – Generic Architecture


Government
Businesses
Citizens
Non-profit Org.
Service Component
Public Services, Government Operational support
(Infrastructure, transportation, emergency management, public safety, education, health, happy life, etc.)
Intelligence and Analytics Component
(Learning, Prediction, Reasoning, Inference)
Data Integration amd Linking
Data Collection & Storage Component
Business Database
Social Media
Mobile devices
Surveillance video data
Sensors
Web data
IoT devices
Open Gov Data
Legacy data
data sources
5

Information Services, Knowledge/Insights For Data/Evidence-based Decision Support, Recommendations, Nudging

Cases of Smart Cities

- **Multiple Cities in NC Regions: North Carolina Regional Water Level Monitoring Data Sharing Pilot**
  - Installing **real-time data sharing architecture** technology that provides real-time data associated with heavy rainfall events
  - More proactive response, reducing the amount of time required to respond to a **stormwater emergency**, enhances data-driven mitigation and infrastructure decisions, and improves prediction.
- **Markham, Ontario, Canada - Smart City Accelerator Program**
  - **Ultrasonic Storm/Flood sensors** were installed *under manhole covers* and *in rivers* to gather data on water levels that would signal imminent flooding conditions caused by storms.
  - In-pipe pressure sensors provide insight into potential watermain leaks.
  - *Sensors on multiple nearby hydrants* allow the city to triangulate **anomalies to pinpoint pipe integrity issues**. Location accuracy dramatically **reduces dig repair costs**. The real benefit is **proactively fixing cracked pipes** before they become catastrophic breaks.
  - Acoustic sensors inside watermains listens to the water stream itself and can pinpoint issues to **proactively schedule repairs**.

Cases of Smart Cities

- **Miami-Dade County, FL — Adaptive Signal Control Technology – 300 Traffic Controllers**
  - Implementing a small-scale **Adaptive Signal Control Technology** resulting in a 10% reduction in travel time along the corridor
  - a noticeable improvement in traffic flow when the roads are congested
- **San Diego, CA - Emergency Vehicle Mobility**
  - The Fire, Police and Transportation Departments
  - Use **vehicle location technology** to communicate with the city's traffic control center to **clear intersections of traffic** and provide emergency vehicles with a green signal.
    - A Novel Spatially-Aware Approach to Emergency Vehicle Pre-emption for First Responders.
  - Significantly cheaper than installing equipment at each signalized intersection and allows for a system-wide view and control.

Cases of Smart Cities

- **Washington DC: Smart Buildings**
  - an open source energy data platform
    - Data from controllers, switches, meters, sub-meters and various **devices present in buildings** can be collected and managed in a secure, single platform.
    - Informs DC's **municipal power purchase agreements** and the city's participation in Demand Response programs.
    - To reduce carbon dioxide emission by 70,000 tons.
- **Virginia Beach, VA: Flood disaster preparedness and recovery**
  - Internet of Things (IoT) sensors to mitigate the impact of sea-level rise and flooding
  - Appointed a Chief Data Officer who improved the Open Data and transparency site with meaningful data and visualizations.

Cases of Smart Cities

- **Westminster, CO: Cybersecurity**
  - City Council's goals and commitment to security
  - Fulfilled by adopting National Institute of Standards and Technology (NIST) cybersecurity framework
- **Chattanooga and Hamilton County, TN — 911 Project – Predicting Hotspots for Accidents** (Police and Law Enforcement and Emergency Management)
  - Leveraged machine learning, multilayer perceptron (MLP) neural network models
  - To analyze both historic and current 911 data to identify accident trends.
  - Mitigation strategies

Cases of Smart Cities

- **LA: Digital services to city residents**
  - Chatbots, automated assistants and social media platforms to better assist citizens
  - Clean Streets initiative, a program of the collaborative Comprehensive Homeless Strategy – 311 calls, mobile app, CRM
- **Topeka, KS — Open Data and Project Portal**
  - Data portals for **budget, checkbook and projects**
    - The checkbook shows all expenditures and
    - the project portal shows all active projects (the associated costs and timelines for the city's active Capital Improvement Projects)
  - Changing the way the City operates and its internal culture
  - Build transparency, accountability and trust.
- **Santa Clara County, CA — Assessment Appeals Data Management System**
  - Processes several thousands of **property tax appeals applications** each year from residents
  - Automated the entire process with centralized appeal data,
  - A smooth applicant experience, robust reporting capabilities, increased efficiency by 50% and increased online filings.
- **Bellville, WA: The eCityGov Alliance** by nine cities – **shared services**
  - Provides services such as recruiting, permitting and mapping; project management; help desk and application hosting services .

Cases of Smart Cities

- **Sonoma County, CA — Accessing Coordinated Care and Empowering Self-Sufficiency**
  - Integrated health and human services (e.g. safe, secure housing for the more than 4,000 fire victims) - to provide holistic services to individuals with complex needs.
  - A **composite client picture** (**360-degree view**) from various department back-end systems
    - E.g. from Public Assistance, Behavioral Health, Substance Abuse, Health Services, Human Services, Child Support Services, Social Services, Justice, and Housing.
  - reduced duplications of services; improved workflow efficiencies; reduced likelihood of drop off; increased service and program awareness; reduced recidivism; better client results.
- **Oklahoma: OK Benefits** - Dept of Human Services
  - **Master Person Index**—a single repository for identifying every person who interacts with DHS
  - DHS delivers benefits and services to facilitate family-focused, outcome-driven decision-making.

Living Lab – Curiosity Lab

- **Peachtree Corners, GA — Curiosity Lab**
- A 5G-enabled **fully connected public infrastructure**
  - Test tracks for autonomous vehicle and smart city living laboratory.
  - One of the world's only real-world testing environments
  - Where people, **autonomous vehicles and smart city technology interact** each and every day, the Lab is fully operational and in-use by multiple companies.

Digital Twins

- **Digital replica of physical space**
  - 3D models of the built environment
  - street network model
  - simulation models (e.g. urban mobility, wind flow, etc.)
  - Data sets
  - implemented in a visualization platform for virtual reality
- Enable:
  - real-time data analysis (what-if scenarios)
  - Simulation or dynamic modeling of behaviors,
  - interactions and evolutions of complex systems

Digital Twins

- **Virtual Singapore** - Enable users from different sectors
  - To develop and test new tools, applications and services, improve planning and decision-making and research ways to solve city challenges
  - Data sources: 3D city models down to the building materials and terrain attributes, city data, real-time operating data from sensor networks and more.
  - Enables users and city leaders to **run simulations** of everything
    - *population growth , public events, natural disasters*, to determine the best response.
    - *Infrastructure changes*, such as building a new stadium, can also be modeled
      - to test how it might impact traffic, pollution, population density and more.

Source: engineering and Technology 2019

Digital Reality Capture

Reality capture (digitize world around you)
digital twins

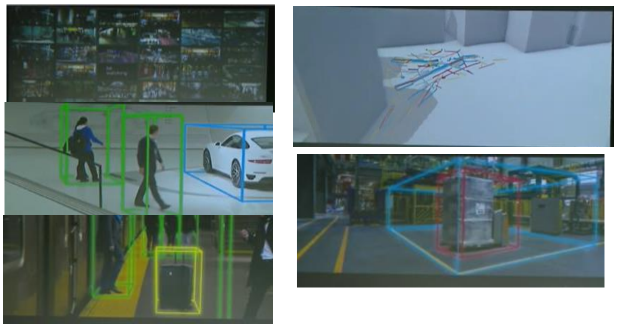
- Ariel photos, Lidar, Remote Sensing
- (airplane, satellites…)
- Aerial imagery captured using a drone.
- A terrestrial laser scanner to capture ground level from four positions
- photos captured inside a full-body scanner.

SW

- creates 3D models out of unordered photographs (terrestrial and/or aerial) or laser scans,
- cultural heritage (art and architecture),
- full body scanning,
- gaming, surveying, mapping, visual effects (VFX) and virtual reality (VR)

Simulations

3D Replica of a city
- flyby simulations



Source: Hexagon Digital Reality.

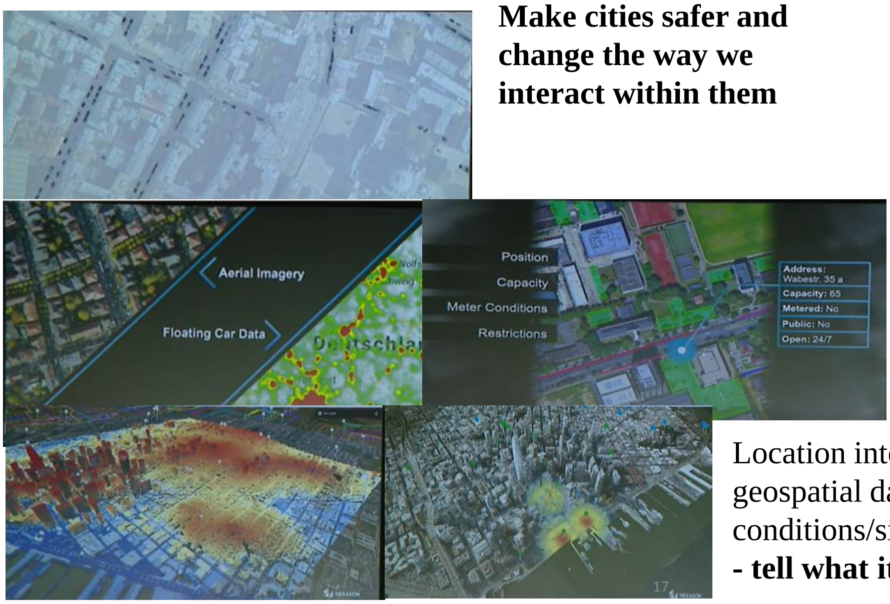
Switzerland Digital Twins Platform
Mesh the city model with layers of geospatial data for **real time situation awareness**

- Real time traffic information,
- Noise/pollution heatmaps,
- Piping systems underground (using Radar),

Beyond security camera – use thermal camera (LIDAR) – 3D monitoring

- **people tracking/change detection**
- **3D fencing** of objects of interest (heritage museum objects, objects in trains),
- Safety of city: by making all objects machine readable/detectable

Digitization and Machine Intelligence



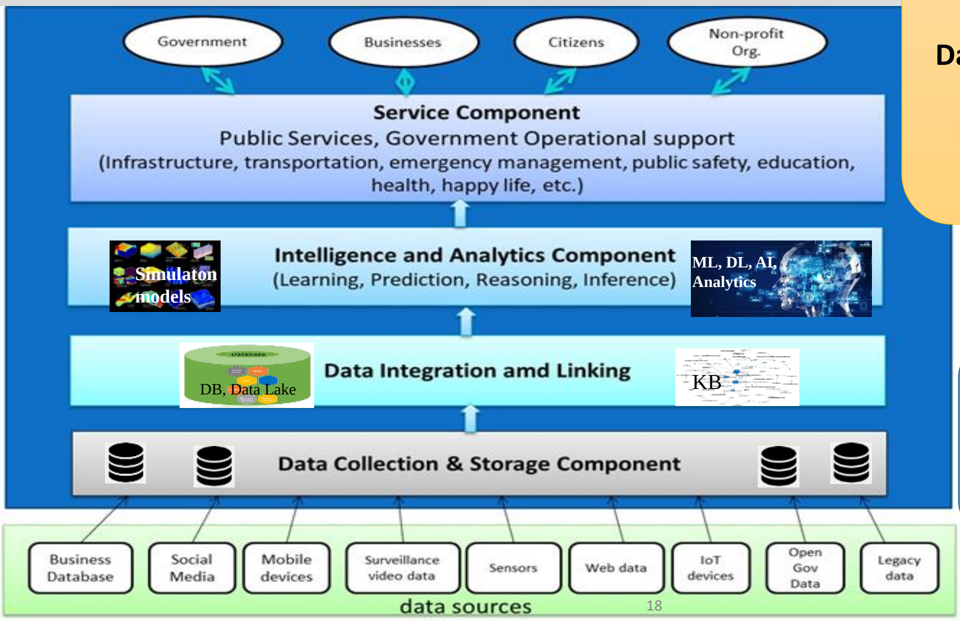
Make cities safer and change the way we interact within them

Machine Learning/AI Algorithms on areal photos

- Classify areal photos into roofs, trees, buildings, cars, etc.
-Predictive parking spaces

Location intelligence – connecting geospatial data for indoor, outdoor conditions/situations
- tell what it was, is, will be

Smart City – Generic Architecture


Government
Businesses
Citizens
Non-profit Org.
Service Component
Public Services, Government Operational support
(Infrastructure, transportation, emergency management, public safety, education, health, happy life, etc.)
Simulaton models
Intelligence and Analytics Component
(Learning, Prediction, Reasoning, Inference)
ML, DL, AI, Analytics
DB, Data Lake
Data Integration amd Linking
KB
Data Collection & Storage Component
Business Database
Social Media
Mobile devices
Surveillance video data
Sensors
Web data
IoT devices
Open Gov Data
Legacy data
data sources
18

Information Services, Knowledge/Insights, For **Data/Evidence-based Decision Support,** Recommendation, Nudging

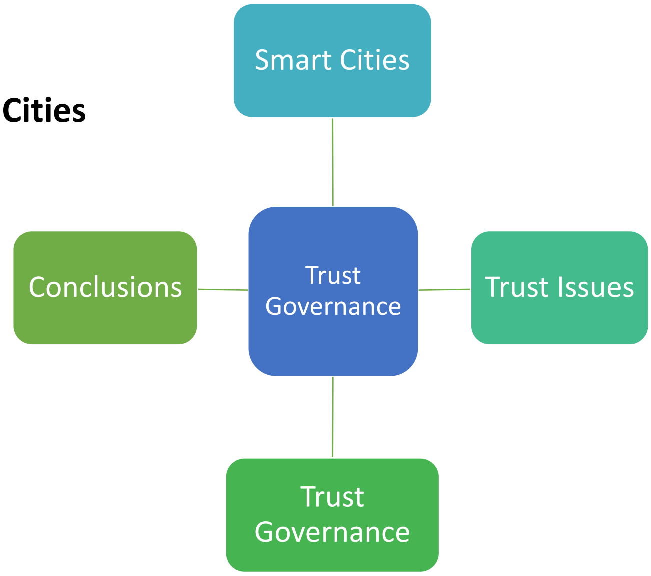
Intelligent Machines Learn, think and do

Automated Decision Making

Intelligent Society

Agenda

- Smart Cities
- **Trust Issues in Smart/Intelligent Cities**
  - Trust risks in data
  - Trust risks in algorithms
  - Trust risks in collaboration, Value Chain
  - Trust risks in connected devices
- Governance of Trust
- Conclusions



Smart Cities

Conclusions

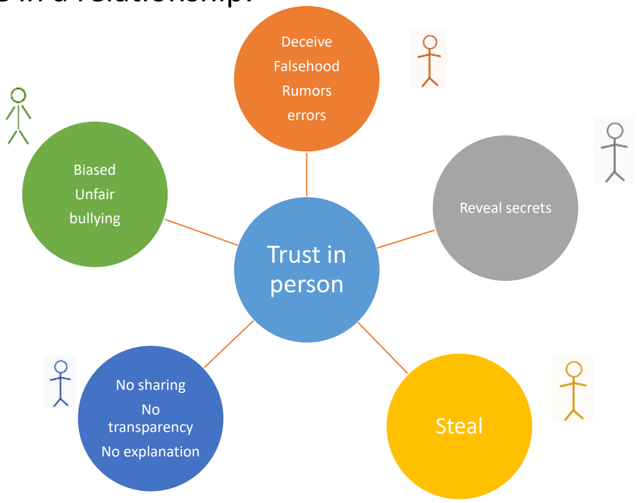
Trust Governance

Trust Issues

Trust Governance

Trust in people

- What do you mean by trusting someone in a relationship:
  - You can **rely on them**
  - You are **comfortable confiding in them**
  - You **feel safe with them**.
- Can we trust a person?
  - Show false identity (deceive who they are)
  - Speak falsehood, rumors, errors
  - Reveals secrets/confidential matters
  - Steals something
  - Is biased or prejudiced
  - Is not fair
  - Bullying, threatening
  - Not sharing
  - Is not transparent in decisions/actions and not explaining



Deceive
Falsehood
Rumors
errors

Biased
Unfair
bullying

Reveal secrets

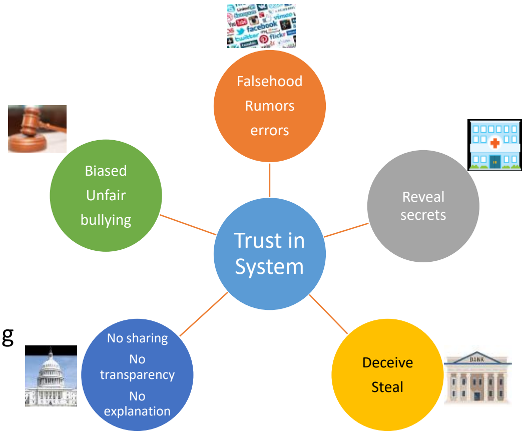
Trust in person

No sharing
No transparency
No explanation

Steal

Trust in Systems (Institutions)

- What do you mean by trusting ~~someone~~ systems (proxy of institutions) in a relationship:
  - You can **rely on them**
  - You are **comfortable confiding in them**
  - You **feel safe with them**.
- Can we trust a ~~person~~ System?
  - Shows false identity (deceive who they are)
  - Speaks falsehood, rumors, errors
  - Reveals secrets/confidential matters
  - Steals something
  - Is biased or prejudiced
  - Is not fair
  - Bullying, threatening
  - Not sharing
  - Is not transparent in decisions/actions and not explaining
- People interact with institutions through
  - “Computer-mediated” systems
  - These may exhibit “Trust eroding behaviors”, Trust Risks



Trust in System

Falsehood
Rumors
errors

Reveal secrets

Deceive
Steal

No sharing
No transparency
No explanation

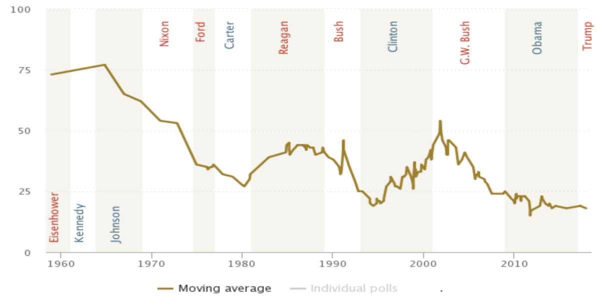
Biased
Unfair
bullying

Trust as Social Glue

- Societies can’t run without trust;
  - It is social glue (Bostman 2017)
- The institutional trust that is really important to society
  - is disintegrating at an alarming rate
  - whether it’s banks, the media, government, churches . . .

Public trust in government near historic lows

% who trust the govt in Washington always or most of the time

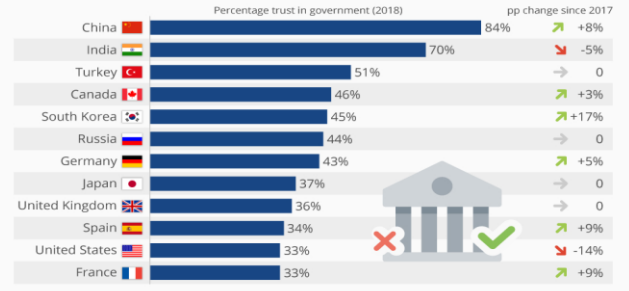


PEW RESEARCH CENTER

Source: MPR news

Where Trust In Government Is Highest and Lowest

% trusting the government and change from 2017 to 2018 (selected countries)



@StatistaCharts Source: Edelman Trust Barometer

Source: Statista

statista

AI driven Intelligent Society- Challenges

Intelligent Cities/Society =

Devices + Hyper-connectivity + DATA + Super Intelligent Machines + Knowledge/Decision Support

- Surveillance and Privacy issues
  - Video capture of faces/mobility & Vision AI models
- Profiling, behavior tracking issues
  - Continuous data from IoT and mobility data,
  - Connected data from different sources, etc.
- **Data contaminations**
  - IoT device security issues
    - Insulin pumps, smart home locks, pace makers, etc.
  - Social media content manipulations – fake news
  - Fake videos and images
- **Intelligent Machines**, AI Models, advanced Analytics Algorithms
  - Digital reality capture to model
    - Biases, discrimination, hatred, etc.
  - Its recommendations or decisions not explained
  - Intelligent machines that fake identity or behaviors or generate fake data

➔ TRUST CHALLENGES

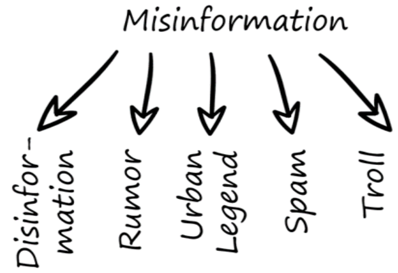
Source: Human-AI Collaboration: Technical & Ethical Challenges 2017

Trust Risks in Data

- Data in digital society is a new currency.
  - Evidence/data based knowledge products, decision support
    - relies on reliable, trustworthy, high quality data
  - Social Media + mobile phones
    - extremely easy to share content, react to the news, and comment on them with opinions about the information on social media
- Influencing
  - contaminated information, misleading the users and confusing facts
  - whether satirical or malicious - is already shaping global debate and changing opinions, perhaps to the point of swaying elections.
  - Falsified information is spread through various channels of social media,
    - which is subsequently used by the general public **to make decisions**.

- FAKE NEWS
  - False information/Disinformation
  - fake facts, rumors, conspiracy theories, and opinions
- DEEP FAKE
- SOCIAL BOTS

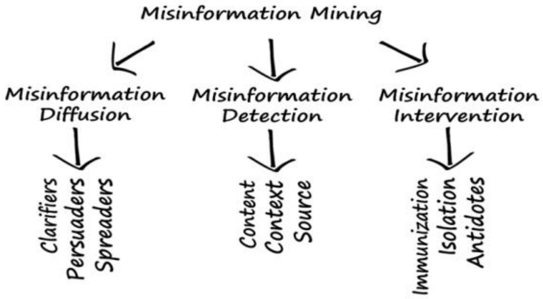
Types of Mis/Dis-Information



Source: [KDKnuggets]

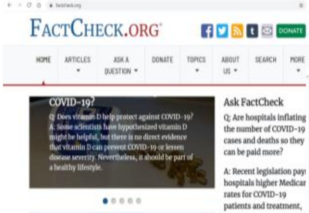
- **Disinformation**
  - fake or inaccurate information that is **intentionally spread**.
  - Misinformation: fake and inaccurate information in social media, regardless of spreader intentions (difficult to tell intentions on social media)
- **A rumor**
  - a story circulating from person to person, of which the truth is unverified or doubtful.
  - rumor detection and truthfulness prediction
- **An urban legend**
  - a fictional story that contains themes related to local popular culture, inaccurate or false stories due to the distortion and exaggeration
- **Spam**
  - unsolicited messages sent to a large number of recipients, containing irrelevant or inappropriate information, which is unwanted.
  - Spamming messages are usually involved with spreading ads, malware, and even leading to scams.
  - spam that conveys misinformation can directly leads to information and financial loss.
- **A troll**
  - a user who posts messages that are deliberately offensive or provocative, with the aim of upsetting other people.

Techniques to Counter Misinformation



- Diffusion
  - studies how misinformation spreads in a network
    - Spreaders disseminate misinformation
    - Clarifiers illuminate the falsehood, and
    - Persuaders try to change users' beliefs
- Detection
  - techniques to find/classify misinformation.
    - the content,
    - context
    - source of misinformation
- Intervention
  - preventive and anti-epidemic measures can be used to fight against misinformation
    - Immunization - Warnings of specific misinformation are sent
    - Isolation - suspend malicious users
    - Antidotes - statement from relevant authorities is essential to effectively terminating the spread

Falsified Information



FACTCHECK.ORG

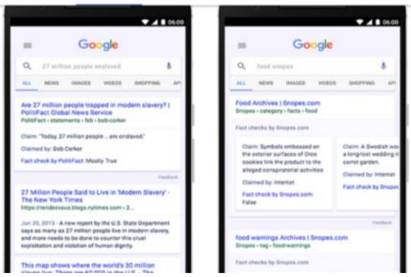
COVID-19?

Q: Does vitamin D help protect against COVID-19?
A: Some scientists have hypothesized vitamin D might be helpful, but there is no direct evidence that vitamin D can prevent COVID-19 or lessen disease severity. Nevertheless, it should be part of a healthy lifestyle.

Ask FactCheck

Q: Are hospitals inflating the number of COVID-19 cases and deaths so they can be paid more?

A: Recent legislation pay hospitals higher Medicar rates for COVID-19 patients and treatment, but there is no evidence



Are 27 million people trapped in modern slavery? | PolitiFact Global News Service

Claim: "Today, 27 million people ... are enslaved."

Claimed by: Bob Corker

Fact check by PolitiFact: Mostly True

27 Million People Said to Live in 'Modern Slavery' - The New York Times

Food Archives | Snopes.com

Fact checks by Snopes.com

Claim: Symbols embossed on the exterior surfaces of Oreo cookies link the product to the alleged conspiratorial activities

Claimed by: Internet

Fact check by Snopes.com: False

food warnings Archives | Snopes.com

College of Staten Island
The City University of New York

Preventive Intervention

- **Censorship** of media or entirely removing poor information
  - is one common practice of minimizing false information.

Detecting Falsified information

- **Fact checkers**
  - https://www.factcheck.org
    - Sites that posted bogus content
    - snopes.com
      - True, mostly true, false, unproven, mixture, outdated, etc.

Google fact checks the search results (2017)

Show "Fact Check" tag for a news article (verified by fact checkers or publishers)

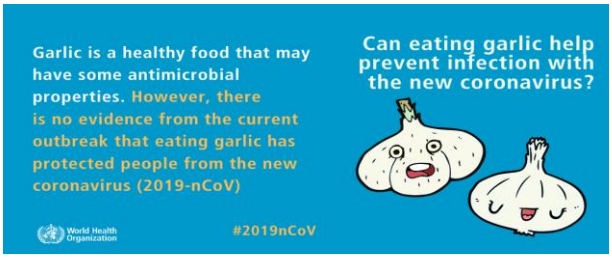
-Snopes.com fact check results shown

- Schema.org/ClaimReview – markup document should be used by publishers

---

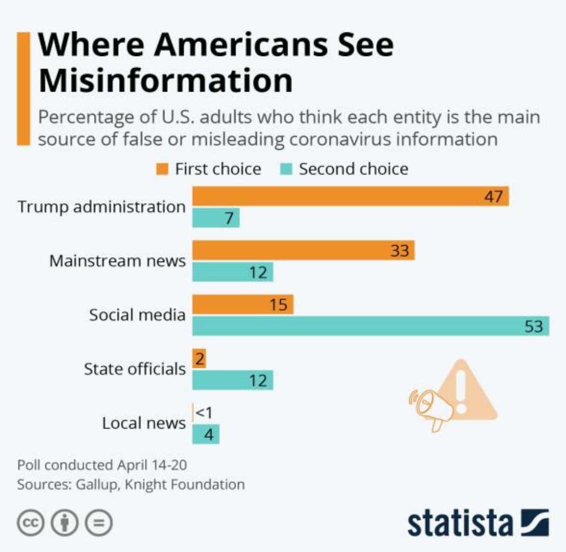
Infodemic

- Criminals are exploiting the COVID-19 crisis (UN)
  - selling fake coronavirus cures online
  - a cyberattack on hospitals' critical information systems,
- Need to its fight against a proliferation of **false information** about the virus.



Can eating garlic help prevent infection with the new coronavirus?

Garlic is a healthy food that may have some antimicrobial properties. However, there is no evidence from the current outbreak that eating garlic has protected people from the new coronavirus (2019-nCoV)



Source: Statista – Gallup poll

Reliable sources of information during the COVID-19 pandemic

- Critical to have up-to-date, accurate news.
  - “We’re all making daily decisions about everything from running to self-isolation to how much toilet paper to buy.”
  - In order to best protect ourselves and our society, we need to know not only the latest facts, but also what actions to take

Source: How to find credible info in a pandemic

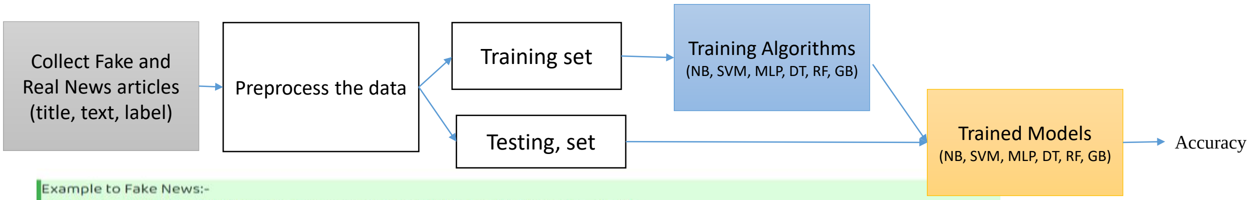
- **News Sources**
  - NPR News Special Series The Coronavirus Crisis:
  - New York Times: The Coronavirus Outbreak
  - BBC News: Coronavirus pandemic
  - POLITICO
- **Medical Sources**
  - **WHO, CDC,** Johns Hopkins University School of Medicine: Coronavirus Resource Center
- **COVID Black: A Taskforce on Black Health and Data**
- **Ask a Scientist:**
  - NJ State Government – COVID info hub uses a network of scientists to answer COVID-19 questions

Social Media Sources

- @CDCgov and @WHO
- @AriadneLabs, a health system innovation center that’s a collaboration between Brigham and Women’s Hospital and the Harvard T.H. Chan School of Public Health, and @Asaf_Bitton, its executive director
- @HarvardChanSPH, the Harvard T.H. Chan School of Public Health
- @Laurie_Garrett, a Pulitzer Prize-winning science journalist
- @HelenBranswell, who writes about infectious diseases for STAT
- @ScottGottliebMD, a former commissioner of the U.S. Food and Drug Administration

---

Detect: Supervised Machine Learning Models



Collect Fake and Real News articles (title, text, label)

Preprocess the data

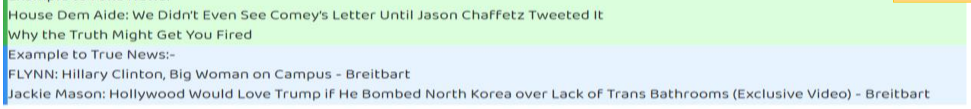
Training set

Testing, set

Training Algorithms (NB, SVM, MLP, DT, RF, GB)

Trained Models (NB, SVM, MLP, DT, RF, GB)

Accuracy



Example to Fake News:-

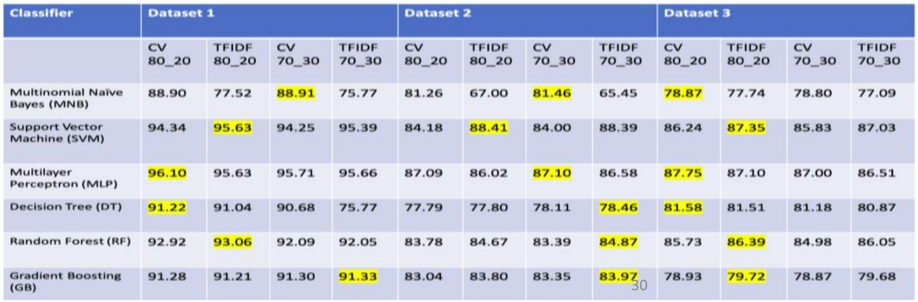
House Dem Aide: We Didn’t Even See Comey’s Letter Until Jason Chaffetz Tweeted It

Why the Truth Might Get You Fired

Example to True News:-

FLYNN: Hillary Clinton, Big Woman on Campus - Breitbart

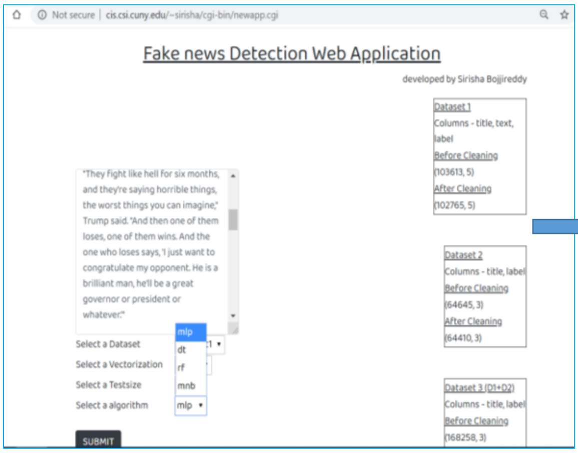
Jackie Mason: Hollywood Would Love Trump if He Bombed North Korea over Lack of Trans Bathrooms (Exclusive Video) - Breitbart



Classifier	Dataset 1				Dataset 2				Dataset 3			
	CV 80_20	TFIDF 80_20	CV 70_30	TFIDF 70_30	CV 80_20	TFIDF 80_20	CV 70_30	TFIDF 70_30	CV 80_20	TFIDF 80_20	CV 70_30	TFIDF 70_30
Multinomial Naïve Bayes (MNB)	88.90	77.52	**88.91**	75.77	81.26	67.00	**81.46**	65.45	**78.87**	77.74	78.80	77.09
Support Vector Machine (SVM)	94.34	**95.63**	94.25	95.39	84.18	**88.41**	84.00	88.39	86.24	**87.35**	85.83	87.03
Multilayer Perceptron (MLP)	**96.10**	95.63	95.71	95.66	87.09	86.02	**87.10**	86.58	**87.75**	87.10	87.00	86.51
Decision Tree (DT)	**91.22**	91.04	90.68	75.77	77.79	77.80	78.11	**78.46**	**81.58**	81.51	81.18	80.87
Random Forest (RF)	92.92	**93.06**	92.09	92.05	83.78	84.67	83.39	**84.87**	85.73	**86.39**	84.98	86.05
Gradient Boosting (GB)	91.28	91.21	91.30	**91.33**	83.04	83.80	83.35	**83.97**	78.93	**79.72**	78.87	79.68

Fake News Detection

Try the Demo site to detect the fake news or not.
Enter a news title and article, or title only.



Not secure | cis.csi.cuny.edu/~sirisha/cgi-bin/newapp.cgi

Fake news Detection Web Application

developed by Sirisha Bojjireddy

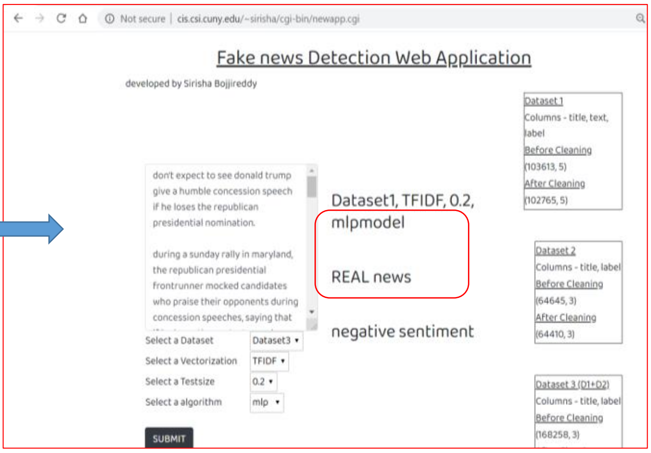
"They fight like hell for six months, and they're saying horrible things, the worst things you can imagine," Trump said. "And then one of them loses, one of them wins. And the one who loses says, 'I just want to congratulate my opponent. He is a brilliant man, he'll be a great governor or president or whatever.'"

Select a Dataset
Select a Vectorization
Select a Testsize
Select a algorithm: mlp

Dataset 1
Columns - title, text, label
Before Cleaning
(103613, 5)
After Cleaning
(102765, 5)

Dataset 2
Columns - title, label
Before Cleaning
(64645, 3)
After Cleaning
(64410, 3)

Dataset 3 (D1+D2)
Columns - title, label
Before Cleaning
(168258, 3)



Not secure | cis.csi.cuny.edu/~sirisha/cgi-bin/newapp.cgi

Fake news Detection Web Application

developed by Sirisha Bojjireddy

don't expect to see donald trump give a humble concession speech if he loses the republican presidential nomination.

during a sunday rally in maryland, the republican presidential frontrunner mocked candidates who praise their opponents during concession speeches, saying that

Select a Dataset: Dataset3
Select a Vectorization: TFIDF
Select a Testsize: 0.2
Select a algorithm: mlp

Dataset1, TFIDF, 0.2, mlpmodel

REAL news

negative sentiment

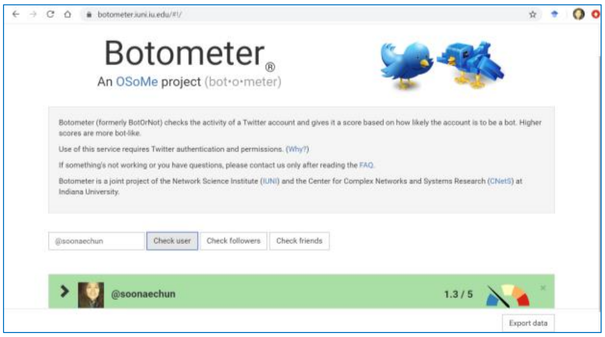
Dataset 1
Columns - title, text, label
Before Cleaning
(103613, 5)
After Cleaning
(102765, 5)

Dataset 2
Columns - title, label
Before Cleaning
(64645, 3)
After Cleaning
(64410, 3)

Dataset 3 (D1+D2)
Columns - title, label
Before Cleaning
(168258, 3)

Trust Risks: Social Bots

- Automated agents
  - create and control **social media fake accounts**
  - imitate humans to **manipulate discussions**
  - **alter** the popularity of users,
  - **pollute** content and **spread misinformation**
  - Perform **terrorist propaganda**
  - Perform **recruitment** actions



botometer.iuni.iu.edu/#!/

Botometer®
An OSoMe project (bot·o·meter)

Botometer (formerly BotOrNot) checks the activity of a Twitter account and gives it a score based on how likely the account is to be a bot. Higher scores are more bot-like.

Use of this service requires Twitter authentication and permissions. (Why?)

If something's not working or you have questions, please contact us only after reading the FAQ.

Botometer is a joint project of the Network Science Institute (IUNI) and the Center for Complex Networks and Systems Research (CNetS) at Indiana University.

@soonaechun

@soonaechun 1.3 / 5

- BotorNot 2014 (Davis et al 2016)
  - A publicly-available service
    - over one million requests via our website and APIs (between 2014-16)
  - Use more than 1000 features
  - Determine whether a Twitter account exhibits similarity to the known characteristics of social bots
    - Engineered Social Tempering (Ferrare et al 2016)
  - known as Botometer

Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016). Botornot: A system to evaluate social bots, WWW

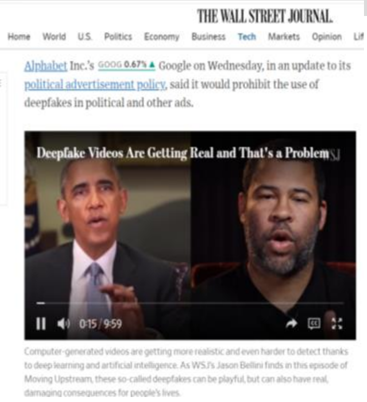
Trust Risk in AI - Deep Fake

- An innovative new deep learning technology
  - **Deep Learning AI technique: Generative adversarial networks (GANs)**
    - that enables to create **realistic-looking photos and videos of people**
  - **Visual deception ("what we see is not real")**
  - saying and doing things that **they did not actually say or do**.
    - President Obama:
      - using an expletive to describe President Trump,
    - Mark Zuckerberg deep fake:
      - admitting that Facebook's true goal is to manipulate and exploit its users,
    - Bill Hader
      - morphing into Al Pacino on a late-night talk show.
- Number is growing from early 2019 rapidly

> "deepfakes threaten to grow from an Internet oddity to a widely **destructive political and social force**. Society needs to act now to prepare itself."
>
> *From Forbes, May 25, 2020*

Deep Fakes - Scenarios

- Deep fake porns
  - FakeApp
    - Gather a photoset of a person,
    - choose a pornographic video to manipulate, and run the app
  - harass and shame people.
    - A video may not be real - but the psychological damage most certainly would be.



THE WALL STREET JOURNAL.

Home World U.S. Politics Economy Business Tech Markets Opinion Lif

Alphabet Inc.'s GOOG 0.67% ▲ Google on Wednesday, in an update to its political advertisement policy, said it would prohibit the use of deepfakes in political and other ads.

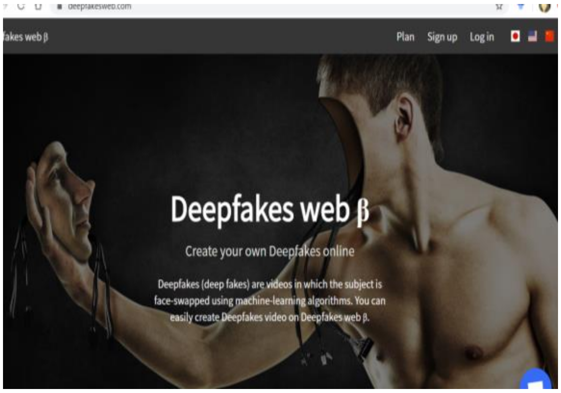
Deepfake Videos Are Getting Real and That's a Problem

Computer-generated videos are getting more realistic and even harder to detect thanks to deep learning and artificial intelligence. As WSJ's Jason Bellini finds in this episode of Moving Upstream, these so-called deepfakes can be playful, but can also have real, damaging consequences for people's lives.

Reference Our Result

- Political Deep Fakes
  - a politician engaging in bribery or sexual assault right before an election;
  - U.S. soldiers committing atrocities against civilians overseas;
  - President Trump declaring the launch of nuclear weapons against North Korea.
  - Belgian prime minister giving a speech that linked the COVID-19 outbreak to environmental damage
  - Gabon's president Ali Bongo – video considered fake led to coup.

Deepfakes Applications



Deepfakes web β

Create your own Deepfakes online

Deepfakes (deep fakes) are videos in which the subject is face-swapped using machine-learning algorithms. You can easily create Deepfakes video on Deepfakes web β.

- Create your own Deepfakes online
  1. Upload videos(or images)
  2. Wait until completion
  3. done ($2/hr)

- Zao: create deepfake videos within seconds.
  - You can choose a video clip from its library which includes scenes from Chinese drama series, Big Bang Theory, popular Hollywood movies
- Avenge Them
- Etc.

Deep Fake Impacts: Seeing is NOT believing any more

- Political and social dangers
  - distorting democratic discourse;
  - manipulating elections;
  - eroding trust in institutions;
  - weakening journalism;
  - exacerbating social divisions;
  - undermining public safety; and
  - inflicting hard-to-repair damage on the reputation of prominent individuals, including elected officials and candidates for office.” (Brookings Institute)
- Reality Apathy
  - “It’s too much effort to figure out what’s real and what’s not, so you’re more willing to just go with whatever your previous affiliations are.” (Aviv Ovadya)

“In the old days, if you wanted *to threaten the United States*, you needed 10 aircraft carriers, and nuclear weapons, and long-range missiles.

Today....*all you need is the ability to produce a **very realistic fake video*** that could undermine our elections, that could throw our country into tremendous crisis internally and weaken us deeply.”

US Senator Marco Rubio

“If we can't believe the videos, the audios, the image, the information that is gleaned from around the world, that is a serious national security risk.”

- Hani Farid, deep fake expert

“Even though there’s footage of you doing or saying something, you can say it was a deepfake and it's very hard to prove otherwise.”

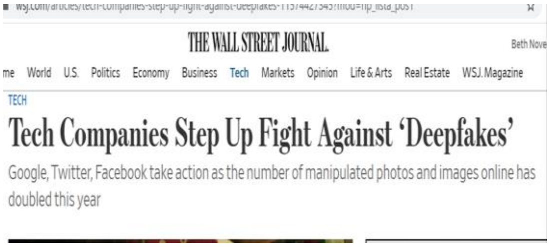
Deepfake Detection

- AI as a solution to harmful deepfake applications
  - Deepfake detection systems
    - that assess lighting, shadows, facial movements, and other features in order to flag images that are fabricated.
    - blinking irregularities were often a telltale sign that a video was fake
  - Innovative defensive approach
    - add a filter to an image file that makes it impossible to use that image to generate a deepfake.
  - Software to verify authenticity of images and videos
    - Truepic and Deeptrace

Laws and Initiatives on Deepfakes

- Legislative, political, and social steps
- California Law
  - Illegal to create or distribute deepfakes of politicians within 60 days of an election
  - Challenges
  - First amendment – freedom of speech
  - the anonymity and borderlessness of the Internet
- copyright, defamation and the right of publicity.
  - broad applicability of the fair use doctrine, the usefulness of these legal avenues may be limited
- Tech platform companies
  - Voluntary restrictions on the deep fakes
  - Free speech and censorship concerns
  - **Terms-of-service agreements** are

“the single most important documents governing digital speech in today’s world.” As a result, these companies’ content policies may be “the most salient response mechanism of all” to deepfakes.



Tech Companies Step Up Fight Against ‘Deepfakes’

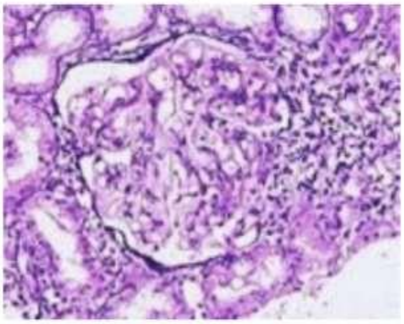
Google, Twitter, Facebook take action as the number of manipulated photos and images online has doubled this year

Trust risks in Automated Algorithms (AI)

Wrong decisions can be costly
and dangerous

"Autonomous car crashes, because it wrongly recognizes …"

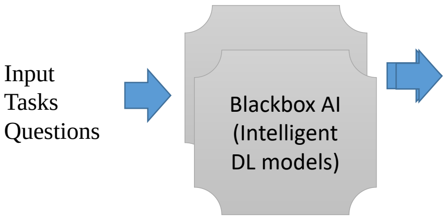
"AI medical diagnosis system misclassifies patient's disease …"



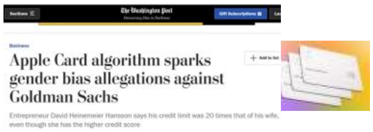
Credit: Samek, Binder, Tutorial on Interpretable ML, MICCAI'18

Trust Risks: Black-box AI

Creates confusion and doubt



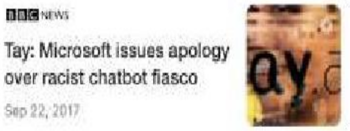
predict who will commit crime,
who will be a good employee,
who will default on a loan, etc



Business

Apple Card algorithm sparks gender bias allegations against Goldman Sachs

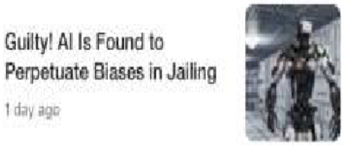
Entrepreneur David Heinemeier Hansson says his credit limit was 20 times that of his wife, even though she has the higher credit score



BBC NEWS

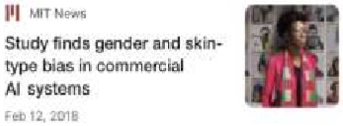
Tay: Microsoft issues apology over racist chatbot fiasco

Sep 22, 2017



Guilty! AI Is Found to Perpetuate Biases in Jailing

1 day ago

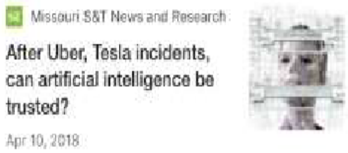


MIT News

Study finds gender and skin-type bias in commercial AI systems

Feb 12, 2018

facial-analysis software shows error rate of 0.8 percent for light-skinned men, 34.7 percent for dark-skinned women.

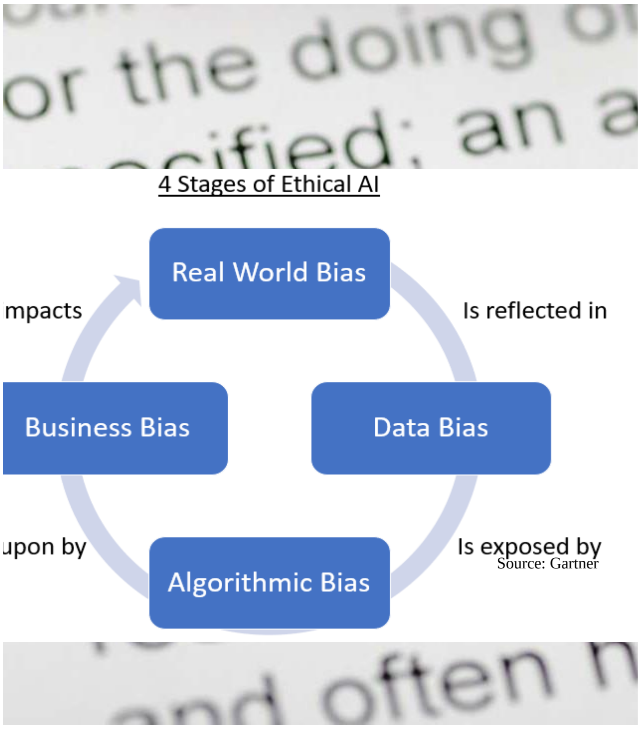


Missouri S&T News and Research

After Uber, Tesla incidents, can artificial intelligence be trusted?

Apr 10, 2018

Source:: explainable AI in industry WWW 2020



Source: Gartner

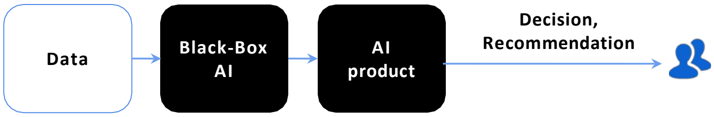
Data bias:

a systematic distortion in data that compromises its use for a task.

Societal biases embedded in behavior can be amplified by algorithms

What is Explainable AI?

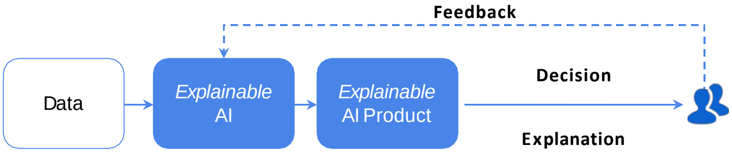
Black Box AI



Confusion with Today's AI Black Box

- Why did you do that?
- Why did you not do that?
- When do you succeed or fail?
- How do I correct an error?

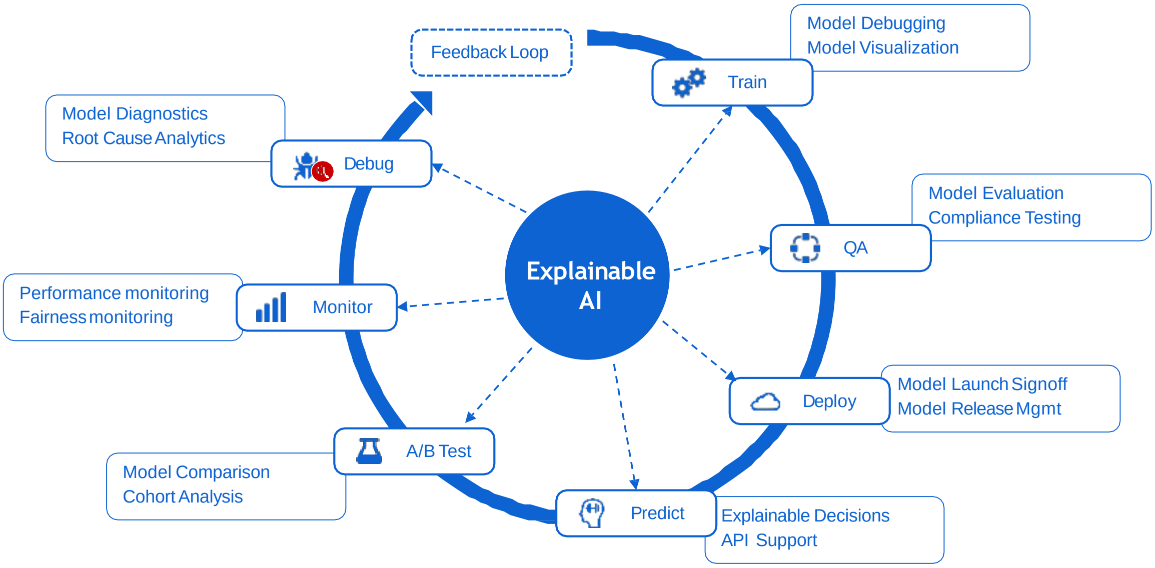
Explainable AI



Clear & Transparent Predictions

- I understand why
- I understand why not
- I know why you succeed or fail
- I understand, so I trust you

“Explainability by Design” for AI products



Explainable AI

Feedback Loop

Train
- Model Debugging
- Model Visualization

QA
- Model Evaluation
- Compliance Testing

Deploy
- Model Launch Signoff
- Model Release Mgmt

Predict
- Explainable Decisions
- API Support

A/B Test
- Model Comparison
- Cohort Analysis

Monitor
- Performance monitoring
- Fairness monitoring

Debug
- Model Diagnostics
- Root Cause Analytics

Source: Gade et al. www2020

Trust Risk in AI models - Laws against Discrimination

- **Algorithmic decision-making** can also threaten human rights, such as the right to anti-discrimination.
  - No-discrimination law
  - Data protection law could help to protect people.

Correctional Offender Management Profiling for Alternative Sanctions

- COMPAS ‘correctly **predicts** recidivism 61 percent of the time
- However,
- **blacks are almost twice as likely as whites to be labeled a higher risk** but not actually re-offend.
- It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes
  - Source:

Source:

Citizenship

Immigration Reform and Control Act

Sex

Equal Pay Act of 1963; Civil Rights Act of 1964

Age

Age Discrimination in Employment Act of 1967

Race

Civil Rights Act of 1964

Disability status

Rehabilitation Act of 1973; Americans with Disabilities Act of 1990

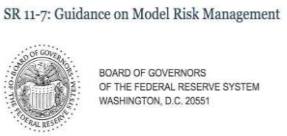
GDPR Concerns Around Lack of Explainability in AI

Fairness

Privacy

Transparency

Explainability

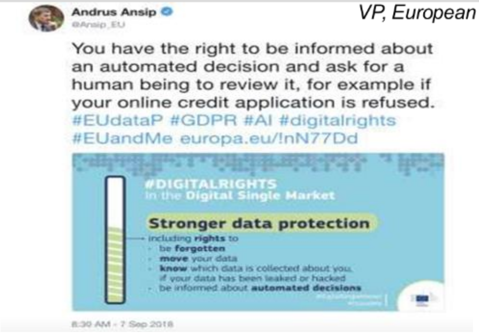


SR 11-7: Guidance on Model Risk Management

BOARD OF GOVERNORS
OF THE FEDERAL RESERVE SYSTEM
WASHINGTON, D.C. 20551

GDPR

CALIFORNIA CONSUMER PRIVACY ACT OF 2018



Andrus Ansip
@Ansip_EU

You have the right to be informed about an automated decision and ask for a human being to review it, for example if your online credit application is refused. #EUdataP #GDPR #AI #digitalrights #EUandMe europa.eu/!nN77Dd

#DIGITALRIGHTS
in the Digital Single Market

Stronger data protection
including **rights** to
- be **forgotten**
- **move** your data
- **know** which data is collected about you, if your data has been leaked or hacked
- be informed about **automated decisions**

8:30 AM - 7 Sep 2018

VP, European Commision

*Companies should commit to ensuring systems that could fall under GDPR, including AI, will be compliant. The threat of **sizeable fines of €20 million or 4% of global turnover** provides a sharp incentive.*

*Article 22 of GDPR empowers individuals with the **right to demand an explanation of how an AI system made a decision that affects them.***

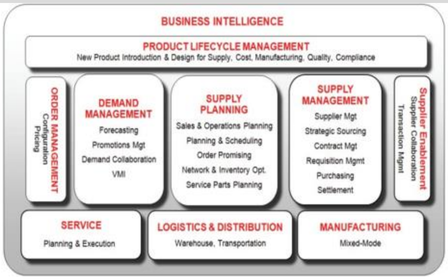
- European Commission

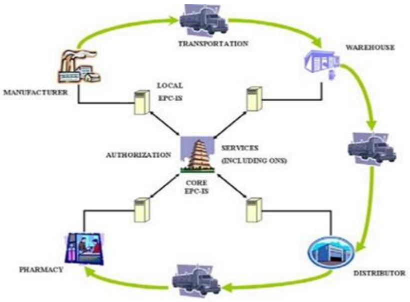
AI Explainability: Growing Global AI Regulations

- GDPR: Article 22 empowers individuals with **the right to demand an explanation of how an automated system** made a decision that affects them.
- Algorithmic Accountability Act 2019: Requires companies to **provide an assessment of the risks** posed by the automated decision system to the **privacy** or **security** and the risks that contribute to **inaccurate, unfair, biased, or discriminatory decisions** impacting consumers (introduced)
- Washington Bill 1655: Establishes **guidelines for the use of automated decision systems** to protect consumers, improve transparency, and create more market predictability. (in house committee)
- Massachusetts Bill H.2701: Establishes **a commission on automated decision-making, transparency, fairness, and individual rights**. (recommended)
- Illinois House Bill 3415: States **predictive data analytics determining creditworthiness or hiring decisions** may not include information that correlates with **the applicant race or zip code**. (in house committee)

Trust in Supply Chain, Value Chain

- Supply chain activities cover everything
  - from product development, sourcing, production, and logistics
- Involves a lot of companies and actors:
  - Suppliers, production companies, distribution companies, transportation companies, retail stores, customer
- SCM:
  - the information systems needed to coordinate these activities.
- **Knowledge product value chain**
  - **Procurements for Smart Cities projects involve** Multiple partners involved
  - data capture technology providers, data managing providers, data scientist, system integrators, end-users, etc.
- Trust in Collaboration
  - **Visibility/Transparency**





Trust Risks in Data Sharing & Collaboration

- Data Sharing may create Concerns faced with challenges and Risks.

Privacy & Security

Bias & Quality

Competitive Advantage Risks

Lack of Value Understanding

Verhulst & Young (2017) Concerns in Data Sharing

- PRIVACY AND SECURITY
  - disclosing personally or demographically identifiable information
- GENERALIZABILITY, DATA BIAS, AND QUALITY
  - Data from a particular demographic subset,
  - possibly ignoring "**data invisibles**"—individuals, often from vulnerable communities, who are unrepresented in private or public datasets
- COMPETITIVE ADVANTAGE RISKS
  - Raw data sharing may threaten Competitive Advantage
- VALUE CHALLENGES
  - Lack of understanding benefits/values of collaboration and data sharing

Trust Risks in IoT

- IoT: **connected things**
  - a network of physical devices
    - fitness trackers, smart thermostats, locks, vehicles, home appliances, and other items embedded with electronics, software, sensors
  - 20 bn devices by 2020 (Gertner)
  - collect and exchange data.

Promise convenience, efficiency and insight

Nest Smart Thermostat

Petnet Smart Pet Feeder

Kisi Smart Lock

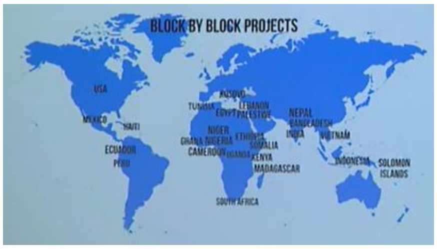
- A **platform for shared risks**
  - Unwitting **surveillance**
  - Any **device can be altered or controlled** through the Internet
  - Security cameras used as part of a **botnet** to attack the Internet. (Mirai DDOS botnets)
  - Absence of security norms and responsible privacy practices
- Regulations
  - may be needed but not effective and takes time
- **IoT Trust Framework (Online Trust Alliance (OTA)**
  - Guides manufacturer and service provider
    - design and business policy choices
    - from initial design through the entire product lifecycle,
  - Provides purchasers and distribution channels with
    - the appropriate filters to assess privacy and safety
  - Gives policymakers
    - the necessary security principles for informed advocacy and economic policy.

Trust for Smart City - Participatory /Co-Design Model

- Redesign of cities (UN Habitat project)

- Minecraft Tool
  - redesign of existing public spaces: Parks, roads/paths from home to school, etc. (35 countries)
  - **Mixed reality** – show the panel and use design options to redesign the reality
    - input from public, and finetune the public space redesign
    - Can view the mixed reality (e.g. bridge built in Minecraft)

Source: Smart City Expo World Congress/



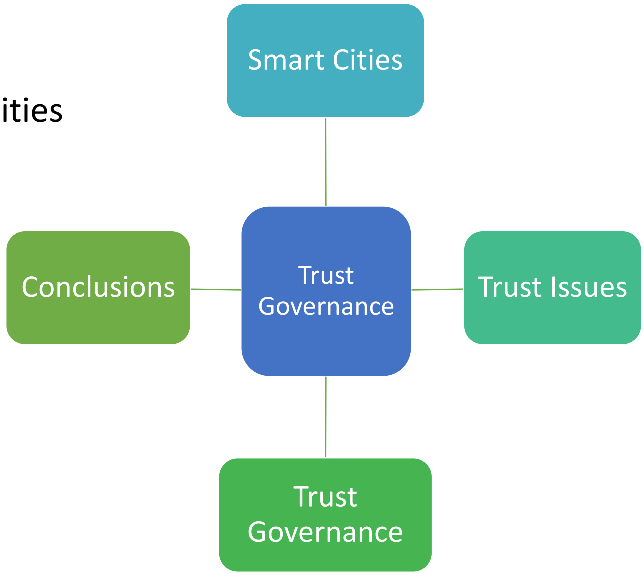
Trust: Inclusive Smart Cities

- **Historically marginalized communities** – suffers from **digital divide**
  - **low-income, elderly, immigrant, and disabled residents**,
- Smart city initiatives inadvertently deepening existing inequalities
  - Lack of transparency, fail to engage community members, or overlook residents' diverse needs and preferences.
  - In different phases of a smart city initiative: design, implementation, and reflection
- Smart city initiatives can gain trust by
  - Shifting from being technology-centric to **citizen-centric**
  - Putting **engagement and inclusion** at the center

Source: Inclusive smart cities

Agenda

- Smart Cities
- Trust Issues in Smart/Intelligent Cities
  - Trust risks in data
  - Trust risks in algorithms
  - Trust risks in collaboration, Value Chain
  - Trust risks in connected devices
- **Governance of Trust**
- **Conclusions**



Smart Cities

Conclusions

Trust Governance

Trust Issues

Trust Governance

Trust Governance for Intelligent Society/Smart Cities

- We have Trust issues in welcoming the Intelligent Society with Automated Machines
- **Trust Governance:**

  Maintain and enhance **trust** in the value chain
  - Technology governance
    - Devices (Supply Chain) – authenticity, security
  - Data Governance
    - Bias in data collection
    - Data integrity, truthfulness, privacy
  - Knowledge creation algorithm/process governance
    - Bias perpetuation issues
    - Marginalized data representation
    - Explainability, Transparency
  - Design Governance
    - Inclusive, Fair, participatory of diverse stakeholders
    - Services, Knowledge products, information creation, etc.

Governance is the act, process or power of governing.
- A combination of processes to manage and monitor the organization's activities in achieving its objectives.
- There is a code of conduct implemented to ensure appropriate behavior and establish credibility

Openness

Integrity

Accountability

Explainable

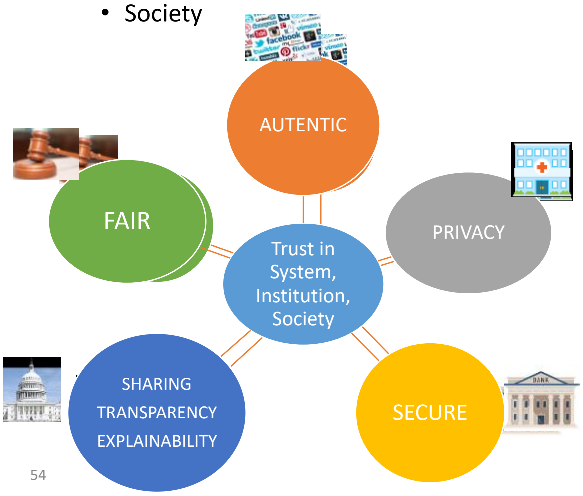
Inclusive

Fair/Ethical

Trust Governance of Intelligent Systems

- Approaches
  - Technical Control
    - against dangers of intelligent machines (e.g. fake detection)
  - Standard of Behaviors
    - Legal/Policy framework
    - **Algorithmic Trust Policy**
    - Living Lab/Digital Twin **Certifications**
  - Organizational
    - Create **Trust Officer/Unit**
    - Human Resources Training
      - Educational approach (digital, ethical)
      - Critical view of the new technologies
    - Future trust-equipped workforce training
  - External Reporting
    - Intelligent Machines/systems evaluation and reporting of trust behaviors

- Building trust in
  - Relationships among people
  - Digital Systems/ Institutions
  - Society



THANK YOU!

- Questions or comments
- Contact
  - Soon.chun@csi.cuny.edu

  @soonaechun

- Submit a case report, a research paper, or a commentary to the ACM journal:
  - **Digital Government Research & Practice**
  - https://dl.acm.org/journal/dgov

  @acmdgov

Acknowledgements:
The presentation is partly supported by grants from NSF CNS 1747728, NSF CNS1624503, and NRF-Korea: 2017S1A3A2066084.

Disclaimer:
This presentation contains images, figures and other open materials from public search results, and they are used for educational purposes in accordance with the Fair Use of copyrighted materials.