

Kamuda Açık ve Büyük Veri

Dünyadan Örnekler, Dersler, Riskler

Doç. Dr. İsmail Sengör Altingövde
ODTÜ Bilgisayar Mühendisliği Bölümü

Bu sunumda...

- Zaman aralığı:
 - 2014-2017
- Kaynakça
 - Akademik çalışmalar
 - Raporlar
 - Medya alıntıları
- Tüm kaynakça:

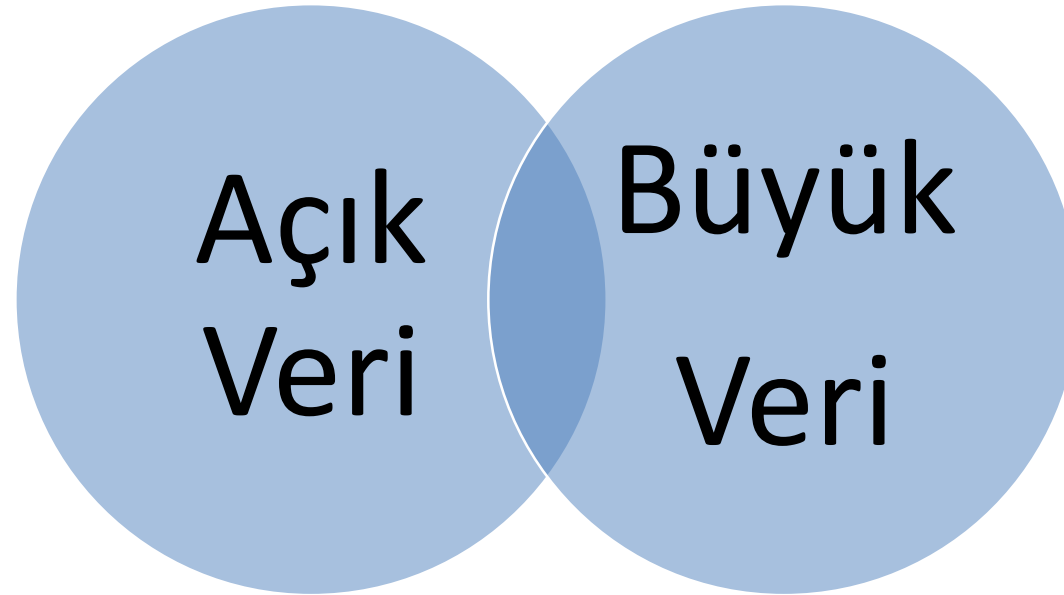
<http://user.ceng.metu.edu.tr/~altingovde/bigdatagov.html>



Kamunun (Büyük) Veri Macerası

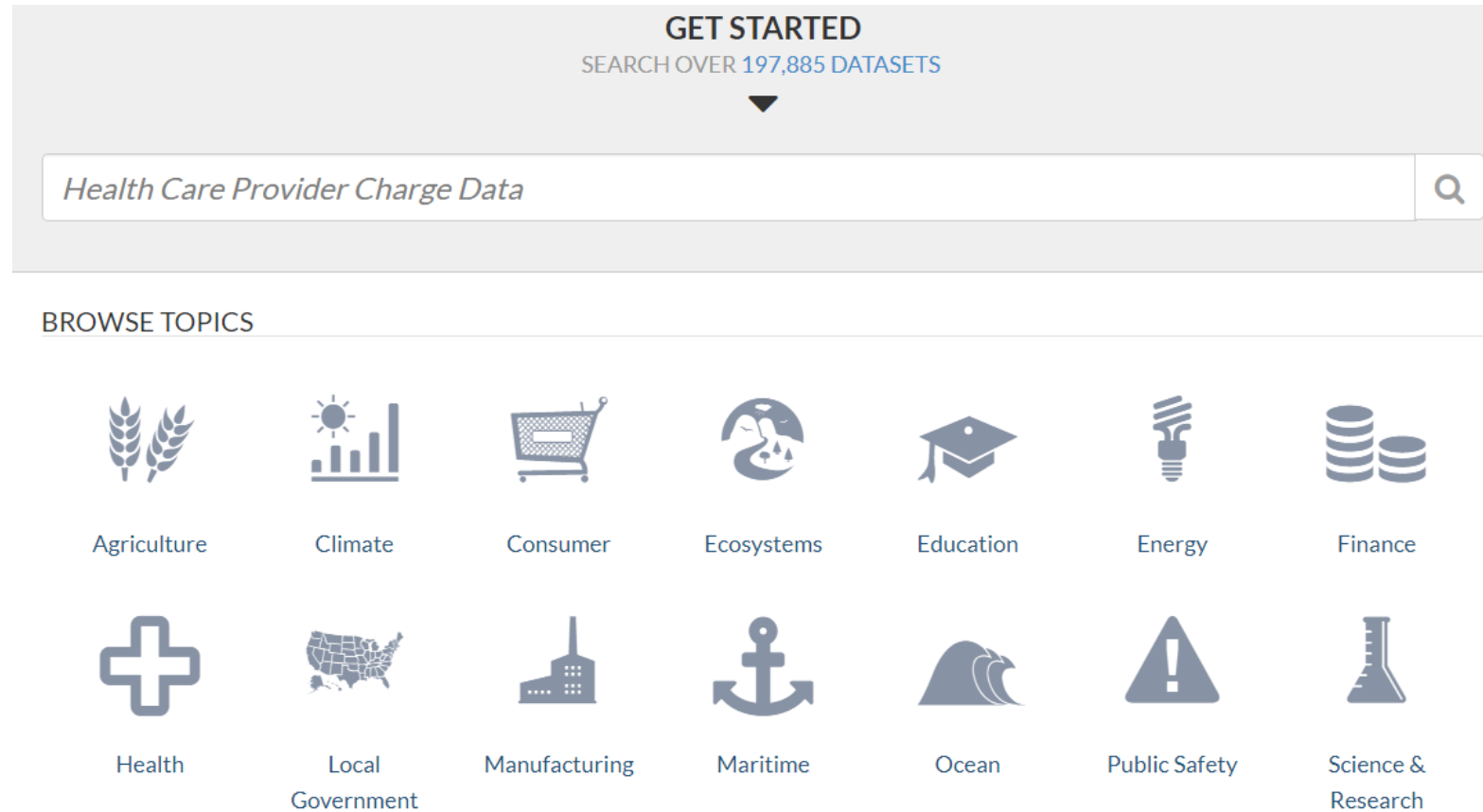
- Üretici: Milyonlarca veri kümesi [Açık/büyük veri]
- Tüketici: Verinin kamu hizmetlerini iyileştirmede kullanılması
- Kolaylaştırıcı: Hukuki, teknolojik ve ArGe altyapısı ve düzenlemeleri

Kamunun (Büyük) Veri Macerası



- Açık Veri
 - Şeffaflık / kontrol / hesap verebilirlik
 - Tekrar-kullanım: iş dünyası ve özel taraflarca

Açık Veri



2009'dan itibaren: data.gov (ABD), data.gov.uk (BK), data.gov.au (AUS), data.gov.sg (SG)

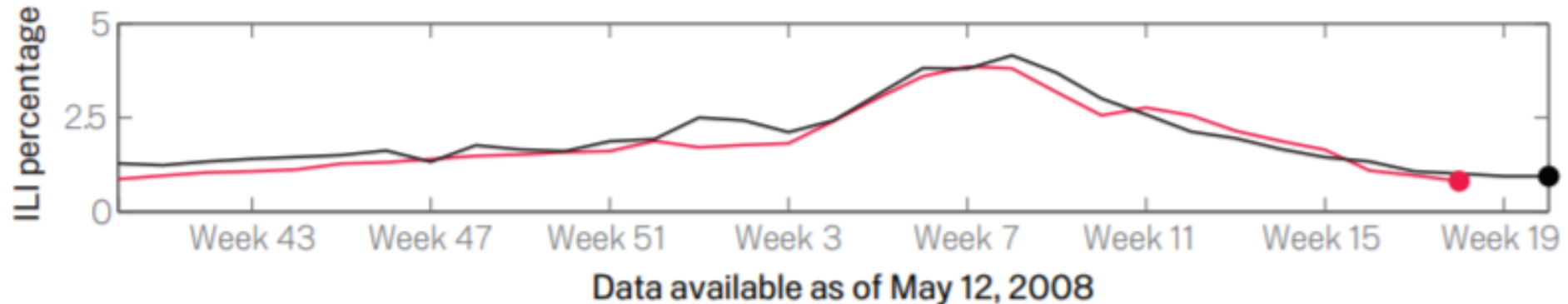
Büyük Veri: Akademik Çalışmalardan Örnekler

- Sağlık



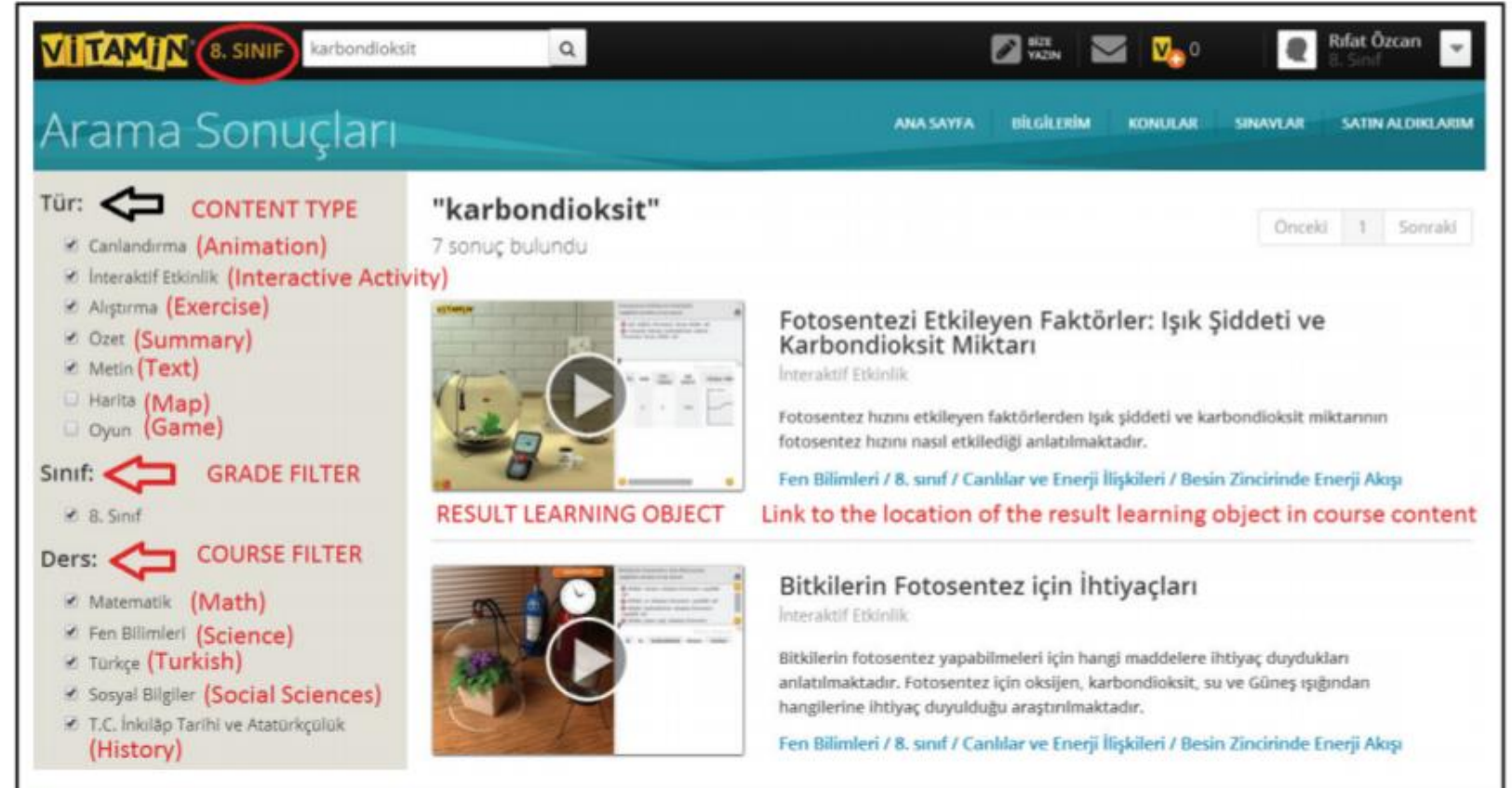
Detecting influenza epidemics using search engine query data

Jeremy Ginsberg¹, Matthew H. Mohebbi¹, Rajan S. Patel¹, Lynnette Brammer², Mark S. Smolinski¹ & Larry Brilliant¹



Büyük Veri: Akademik Çalışmalardan Örnekler

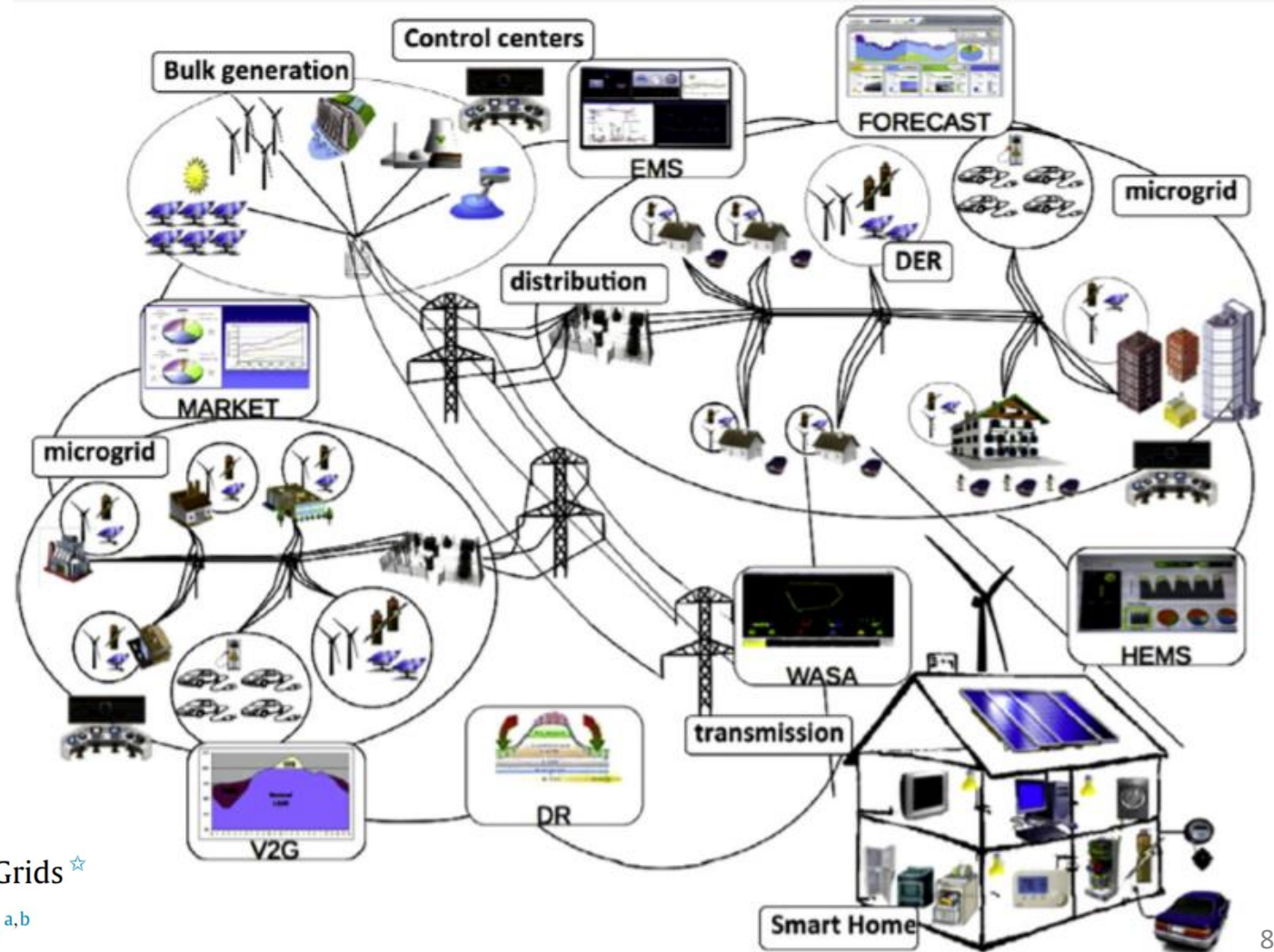
- Eğitim
 - Pearson «learnign curve» veri bankası
 - Öğrenme analitiği (eğitimsel veri madenciliği)



A. Usta, I. S. Altingövde, I. B. Vidinli, R. Ozcan, Ö. Ulusoy: “How k-12 students search for learning?: analysis of an educational search engine log”, Proceedings of the 37th Int’l ACM SIGIR, pp. 1151-1154, 2014.⁷

Büyük Veri: Akademik Çalışmalardan Örnekler

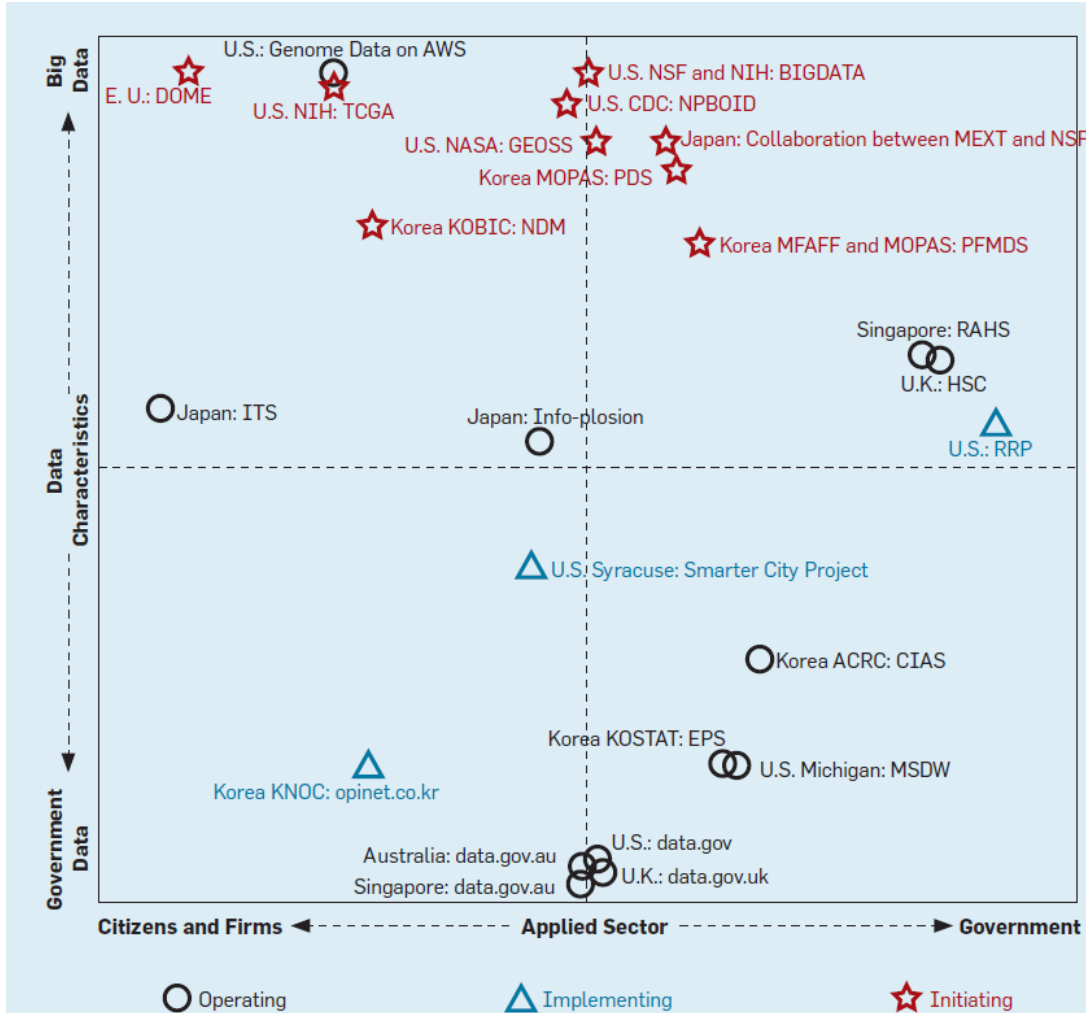
- Enerji:
 - Akıllı grid



Big Data Analytics for Dynamic Energy Management in Smart Grids ☆

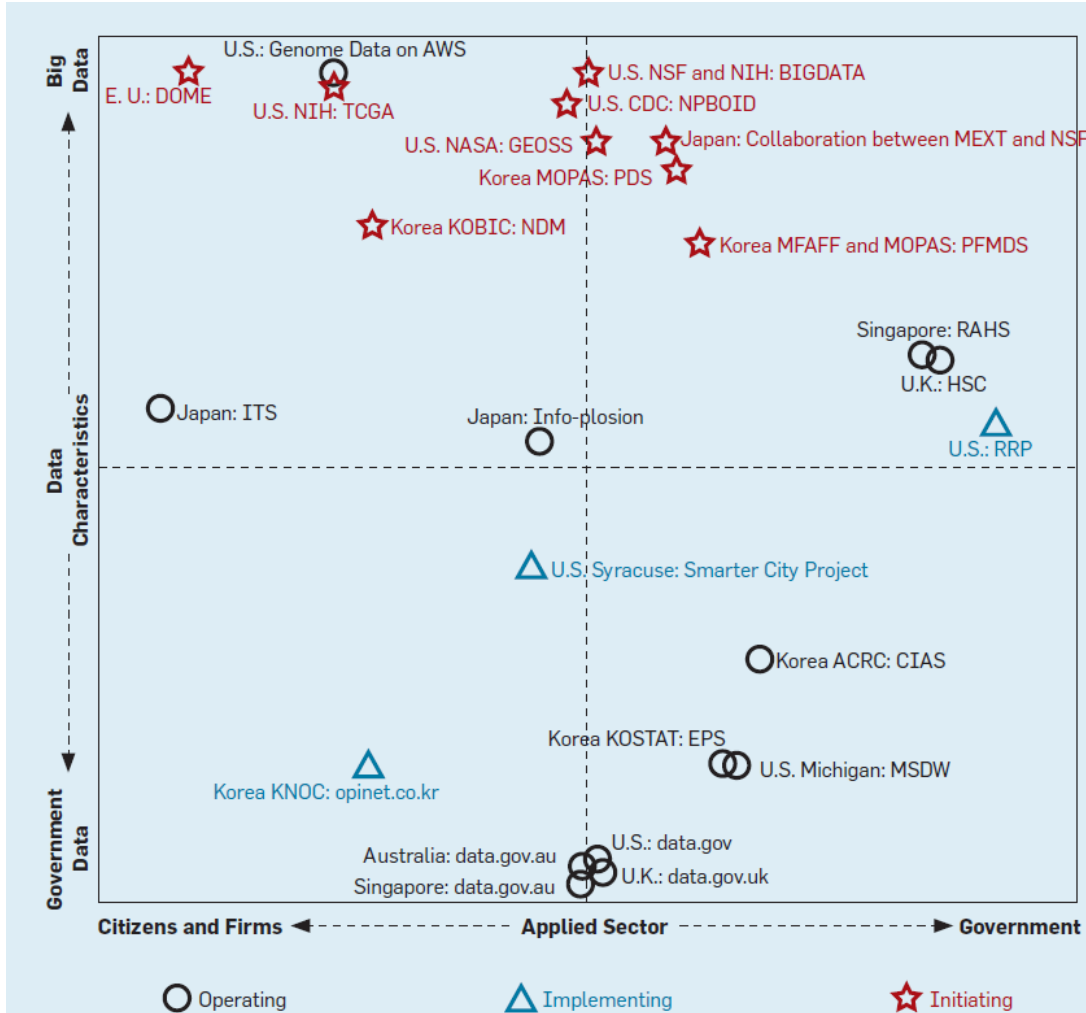
Panagiotis D. Diamantoulakis^{b,*}, Vasileios M. Kapinas^b, George K. Karagiannidis^{a,b}

Kamuda Büyük Veri: 2014



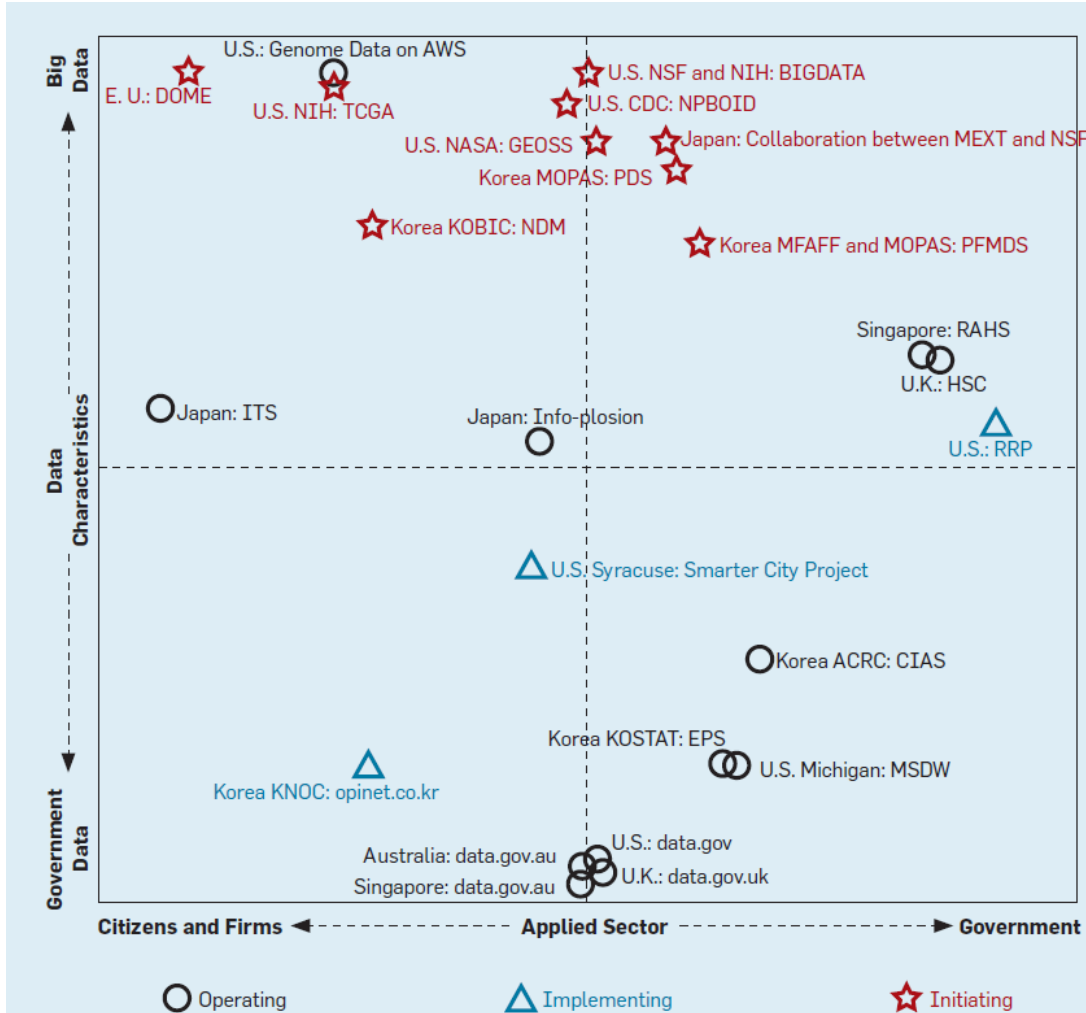
- ABD:
 - İnsan gen datası (TBs) Amazon'da
 - İade Değerlendirme: Vergi sahtekarlığı/kaçığı tespiti
 - Akıllı Şehir: Syracuse + IBM, boşalacak dairelerin tespiti
- BK:
 - Horizon Scanning Center (2004)
 - BK + AB: DOME projesi, SKA radyo teleskopu verileri için

Kamuda Büyük Veri: 2014



- Güney Kore:
 - Sosyal Refah Yönetimi (35 ajans, 385 veri tipi)
 - Şap Hastalığı Önleme (gümrük, çiftlik, çalışan verileri)
 - Afet Tahmini
 - Kişiyi Özel Tedavi
- Singapur:
 - RAHS: Kuş Gribi salgın riskleri
- Japonya:
 - MEXT: Doğal afet önleme & yönetimi¹⁰

Kamuda Büyük Veri: 2014



- Projelerin çoğu kısmen büyük veri; veriler yapılandırılmış, gerçek-zamanlı / hareket değil
- Kamuda ana hedeflere rağmen uygulamalar vatandaş katılımı ve iş sektörlerine yönelik



Ulusal öncelikler!

- ABD (güvenlik, sağlık)
- Japonya (doğal afetler)

Kamuda Büyük Veri: 2017

Hizmet Sunumu

Politika Belirleme

Vatandaş Katılımı

Kamuda Büyük Veri: 2017

Hizmet Sunumu



London Counter Fraud Hub (BK)



Proaktif dinleme (Brezilya)
Ebola Açık Veri İnsiyatifi



Shangai'da 3300 km su boru hattında sensör projesi
Hindistan'da uydu görüntüsünden elektrik kullanımının takibi projesi



PredPol: Tahminleyici suç önleme (ABD'de 50 şehirde; Brezilya ve BK'da benzer)

Kamuda Büyük Veri: 2017

Politika Belirleme



Seul'de 3 milyar çağrı /metin veri noktasıyla gece otobüsleri için rota belirleme



Avustralya-LinkedIn ortaklığıyla iş/beceri taleplerini belirleme
Dünya Bankası + Flowminder + Haiti mobil tel verisiyle iş fırsatları belirlenmesi



Billion Prices Project (MIT): Çevrim içi fiyat bilgisiyle gerçek-zamanlı enflasyon verisi edinilmesi



Global Forest Watch uydu görüntüsü ve açık verilerle ormanlık alanları takip ediyor

Kamuda Büyük Veri: 2017

Vatandaş Katılımı



Brezilya'da 2014 Dünya Kupası protesto tweet'lerinin analizi (Brezilya'nın kupa harcamaları eleştirildi)



Haiti depreminde arama-kurtarma işlemlerinde kalabalık-gücü kullanan uygulama

Devlet Ne Yapmalı?

- 🔑 Devlet seviyesinde büyük veri kullanımı için **veri kullanımı ve sahipliği, kişisel gizlilik, sivil özgürlükler** ve **eşitlik** konularında düzenleme ve yönetmelikler gerekli
- 🔑 Büyük veri kontrol kuleleri (2014'den miras [Kim vd. 2014]!)
- 🔑 Algoritmik şeffaflık
- 🔑 Kamu-özel sektör işbirlikleriyle özel sektördeki –belki daha değerli- verinin de kamu ve sosyal fayda için kullanılması

ABD'den Güncel Örnekler

Mart 2017

Big Data in Government: Progress & Best Practices 2017

- Dr. Ed Kearns, Chief Data Officer, National Oceanic and Atmospheric Administration (**NOAA**)
- Renee Wynn, Chief Information Officer, **NASA**
- Frank Konieczny, Chief Technology Officer, **U.S. Air Force**
- Ram Pisipati, Data Integration & Analytics Director, **Leidos**
- Alan Ford, Director, Government Systems, **Teradata**
- Stephen Moore, Vice President, **Alpha Six Corporation**



ABD'den Güncel Örnekler: Projeler

Mart 2017

- NASA:
 - Pilotların bayılması durumunda yönetimi devralacak bir sistemin eğitilmesi
- Ulusal Okyanus ve Atmosfer Dairesi (NOAA):
 - Açık veri çerçevesinde hava durumu vb. verilerin Amazon üzerinden açılması
- ABD Hava Kuvvetleri
 - Veri-Sürücülü Araç Yönetimi: ~80 bin taşıt (7 milyar USD), müşteri talep tahmini, stok yönetiminin iyileştirilmesi, vb. (Teradata ile)
 - Potansiyel subayların tahmini
- Leidos + NASA
 - 1960'dan günümüze veriler kullanılarak mürettebat sağlığı için iyileştirmeler: riskler ve uzun vadeli etkiler
- Terradata
 - cms.gov için Terradata'nın 600 TB veri üzerinde sahtekarlık tespiti analizi (3 yılda 10 milyar USD)

ABD'den Güncel Örnekler: Dersler

- 🔑 Verilerin farklı birimlerin hükümlerinden alınıp birleştirilmesi (kontrol kulesi?!)
- 🔑 Açık veri sunuluyorsa **güvenlik/gizlilik** + **doğruluk** + veri ile beraber bulut mimaride çalışacak temel **yazılım araçları** sağlanması
- 🔑 Teknolojik esneklik:
 - Belli bir ürün/teknolojiden bağımsız olmak
 - Farklı analitik araçların sisteme eklenebilir olması
 - Çevik programlama
- 🔑 Veri analitiği sonrası bulguları yorumlayabilecek ve tavsiyeye dönüştürebilecek uzman gereksinimi

ABD'den Güncel Örnekler: Zorluklar

Mart 2017

- Süreç/kültür:
 - Algılar: makine öğrenmesi her şeyi çözer (mi)?! **Ateorik olma!**
 - Kaynak yetersizliği → Endüstri işbirliği
- Teknolojik
 - Eski sistemlerdeki veriyi anlamak?
 - Veriyi yedeklemek değil kullanılabilir şekilde korumak (hangi formatta, ne zaman faydalı, vb.)
 - Farklı mimarilerde (Hadoop, veri ambarı, vb.) saklanan veriyi tek bir şekilde görmek (her şeyi 0'dan yapmadan!)
 - Devlet-seviyesinde güvenlik kurallarının teknolojiye ve personele entegrasyonu

Büyük Veri: Ne Ters Gidebilir?

- Gizlilik: AOL 2006 skandalı

☰ 📰 🔍 The New York Times

TECHNOLOGY

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and
TOM ZELLER Jr. AUG. 9, 2006

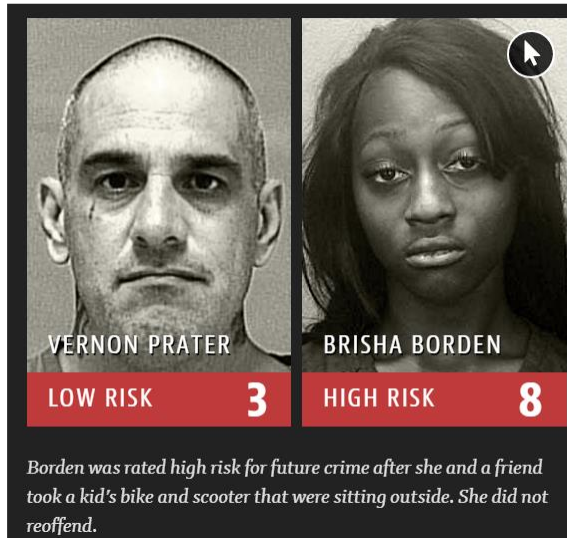
👤 🐦



Thelma Arnold's identity was betrayed by AOL records of her Web searches, like ones for her dog, Dudley, who clearly has a problem.
Erik S. Lesser for The New York Times

Büyük Veri: Ne Ters Gidebilir?

- Algoritmik yanlılık? (bias)



A beauty contest was judged by AI and the robots didn't like dark skin

The first international beauty contest decided by an algorithm has sparked controversy after the results revealed one glaring factor linking the winners

Sonuç

Olumsuz:

- Büyük veri kavramında karışıklıklar
- Bazı projeler operasyonel veya prototip seviyesinde
 - Hedefenen yarara gerçekleştirmiş değil
- Veriden elde edileni anlayacak eleman eksikliği



Dersler:

- Gizlilik / algoritmik şeffaflık
- Veri için sınırların aşılması
- Sıfırdan başlama!
 - Veri ambarı + büyük veri araçları
 - Artırımsal ve hızlı ilerleme

Olumlu:

- Pek çok büyük ölçekli analitik proje de var!
 - Ciddi kazanç potansiyeli

CUMHURİYET BAYRAMI'MIZ Kutlu Olsun!



data.gov: Örnek Uygulamalar [2017]



FoodEducate

[Consumer Price Indexes by Item](#)
[Consumer Price Index](#)
[My Pyramid Food Raw Data](#)
[Food-a-pedia](#)

Süpermarket alışverişinde
barkod tarama ile besin değerlerini öğren!



[Neighborhood Scout](#)

[FHFA House Price Indexes](#)
[Housing Affordability Data System](#)
[Certified Schools SEVIS](#)
[Neighborhood Environment](#)

Mahalle bazlı suç, okul ve emlak raporu



[College Affordability
and Transparency
Center](#)

[Department of Education – National
Center for Education Statistics, Integrated
Postsecondary Education Data System](#)

Üniversite maliyetleri hakkında daha iyi
bilgilendirme



[Power Profiler](#)

[Greenhouse Gas Emissions Inventory](#)
[National Greenhouse Gas Emission Inventory](#)
[Energy-related Environmental Data](#)

Bölgenizde elektrik üretim yolları ve sera
etkisini öğrenip, eviniz için etkiyi belirleme

BK'dan Güncel Örnekler

Şubat 2017



Big Data in government: the challenges and opportunities

John Manzoni, Chief Executive of the UK Civil Service and Permanent Secretary of the Cabinet Office

BK'dan Güncel Örnekler

- In DWP, providing job seekers with more targeted advice, and opportunities that closely match their personal profiles.
- Healthcare is another exciting area. Moorfields Eye Hospital and DeepMind Health are partners in a research project that could lead to earlier detection of eye diseases.
 - Moorfields will share a data set of one million anonymised scans with DeepMind, who will analyse them using machine-learning technology. This can detect and learn patterns from data in seconds, to quickly diagnose whether a condition is urgent.

Dersler ve Zorluklar

- We need to think about the collection and storage of data as part of core national infrastructure, in the same way as we think of our road and rail systems, energy supply, and telecommunications networks as infrastructure.
 - short of the key data science skills in Government.
 - at every level should have an appreciation of the power of data, we're developing a programme in data literacy for non-data specialists.
- One of the fundamental challenges is that of public trust and 'buy-in'.
 - data security, openness, user need and public benefit.
 - Safe people Safe projects Safe settings Safe data Safe output

Büyük Veri: Ne Ters Gidebilir?

- Sosyal veri – her zaman doğru/tam olmayabilir!

The Parable of Google Flu: Traps in Big Data Analysis

Citation:

David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. 2014. "The Parable of Google Flu: Traps in Big Data Analysis." *Science*, 343, 14 March, Pp. 1203-1205. Copy at <http://j.mp/2ovtFps>

Download Citation

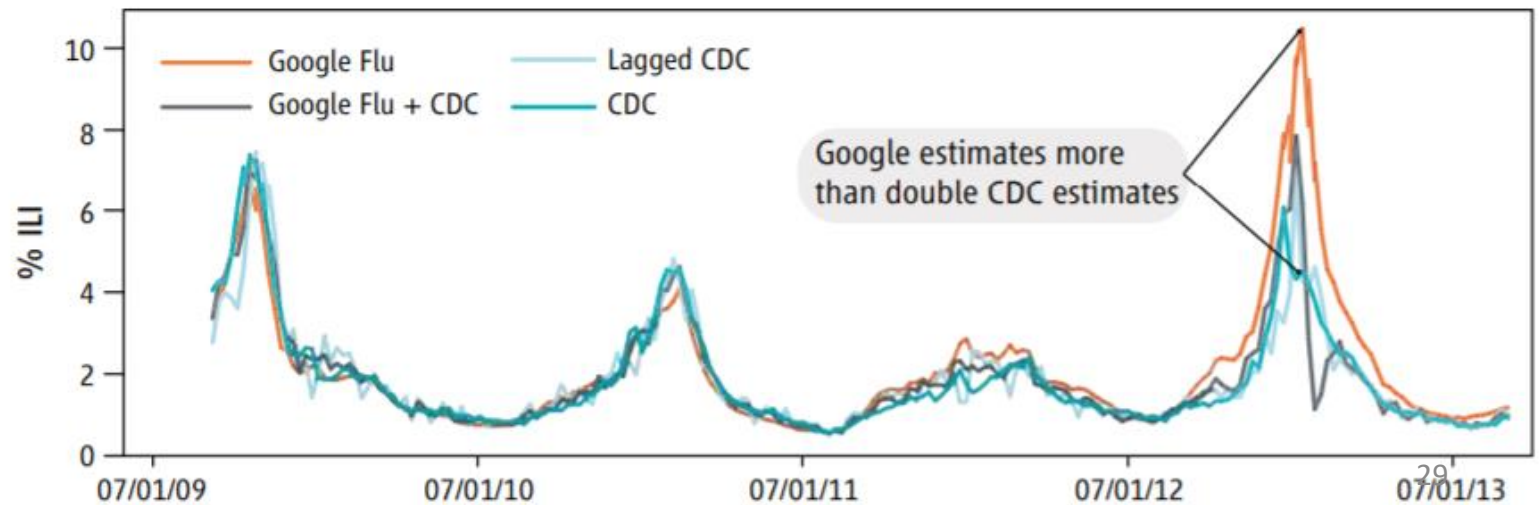
 Article 1.21 MB



Abstract:

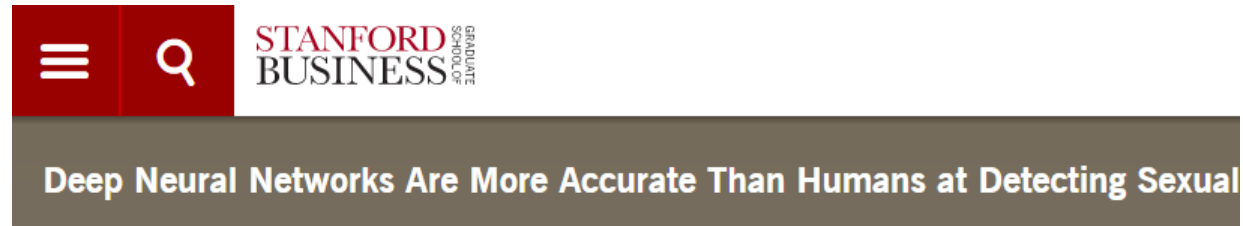
Large errors in flu prediction we offers lessons for the use of big

In February 2013, Google Flu Tre not for a reason that Google exe



Büyük Veri: Ne Ters Gidebilir?

- Ayrımcılık?

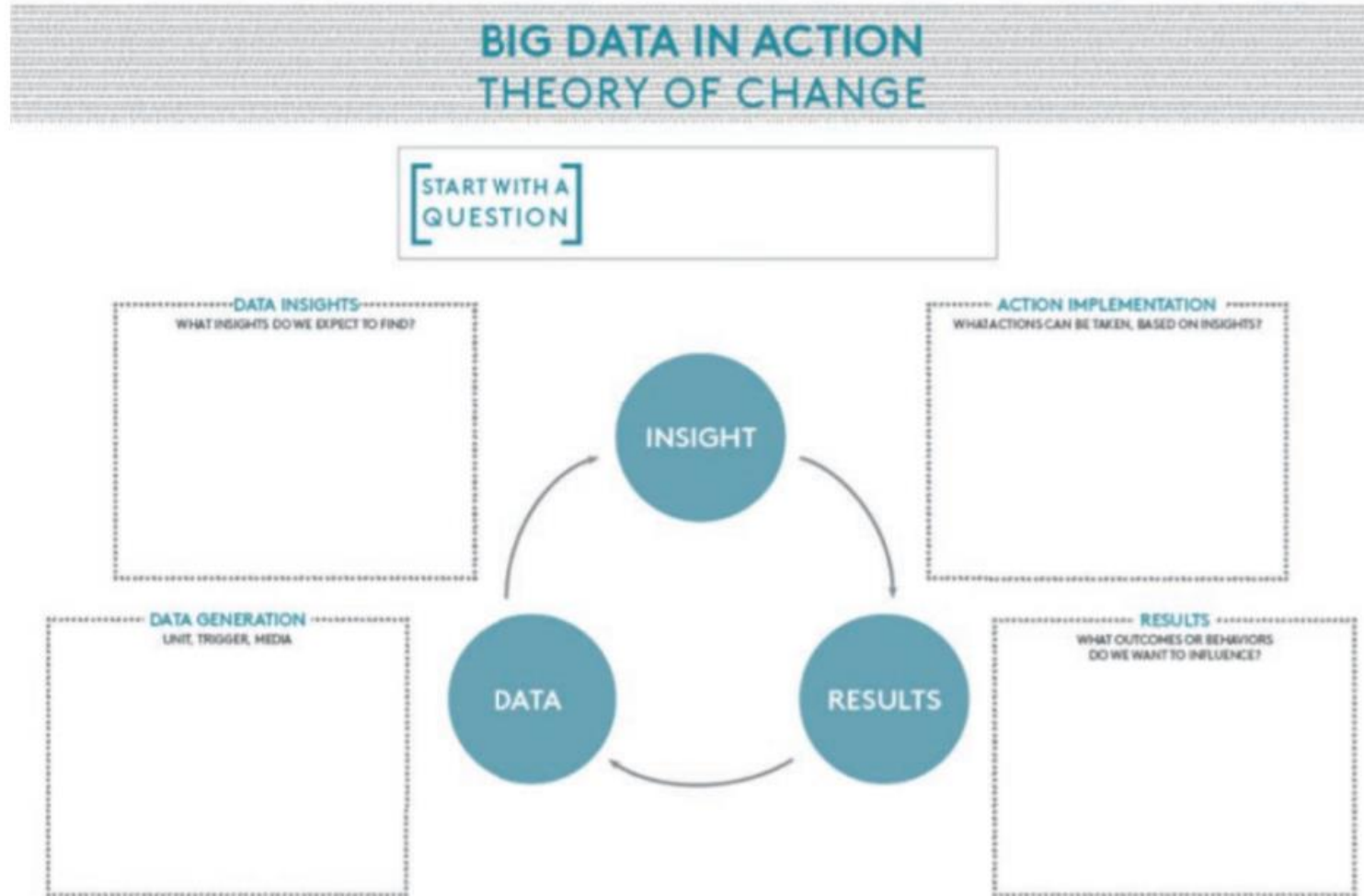


Deep Neural Networks Are More Accurate Than Humans at Detecting Sexual Orientation From Facial Images

By **Michal Kosinski**, Yilun Wang

Journal of Personality and Social Psychology (in press). September 7, 2017.

Çözüm Geliştirme Alıştırması



Kamuda Büyük Veri: 2017



- İşler ve Beceriler: Büyük Veri ile İş Pazarını anlamak?
 - Daha iyi eşleştirme sağlanabilir
 - Devletler becerilere talebi daha iyi anlamak ve kaynakları ilgili eğitime yönlendirmek yoluyla dinamik, proaktif ve dirençli iş gücü sağlayabilir
- Muhasebe: Büyük Veri ile Devlet Görevlilerinin Dürüstlüğü?
- Sağlık: Büyük Veri ile Önleyici Yaklaşımlar?
- Ulaşım: Büyük Veri ile Kalabalık Şehirlerde Ulaşımın İyileştirilmesi?

Kamuda Büyük Veri: 2017



- Çözüm Prototipi Üretme:
 - Organizasyonunuzun önümüzdeki 3 yılda büyük veri ile ilgili vizyonu?
 - Hangi devlet hizmetleri önümüzdeki 3 yılda büyük veri ile en temelden değişme potansiyeli taşıyor?
 - Belirlenen alanlar için bu hizmetler nasıl değişebilir?