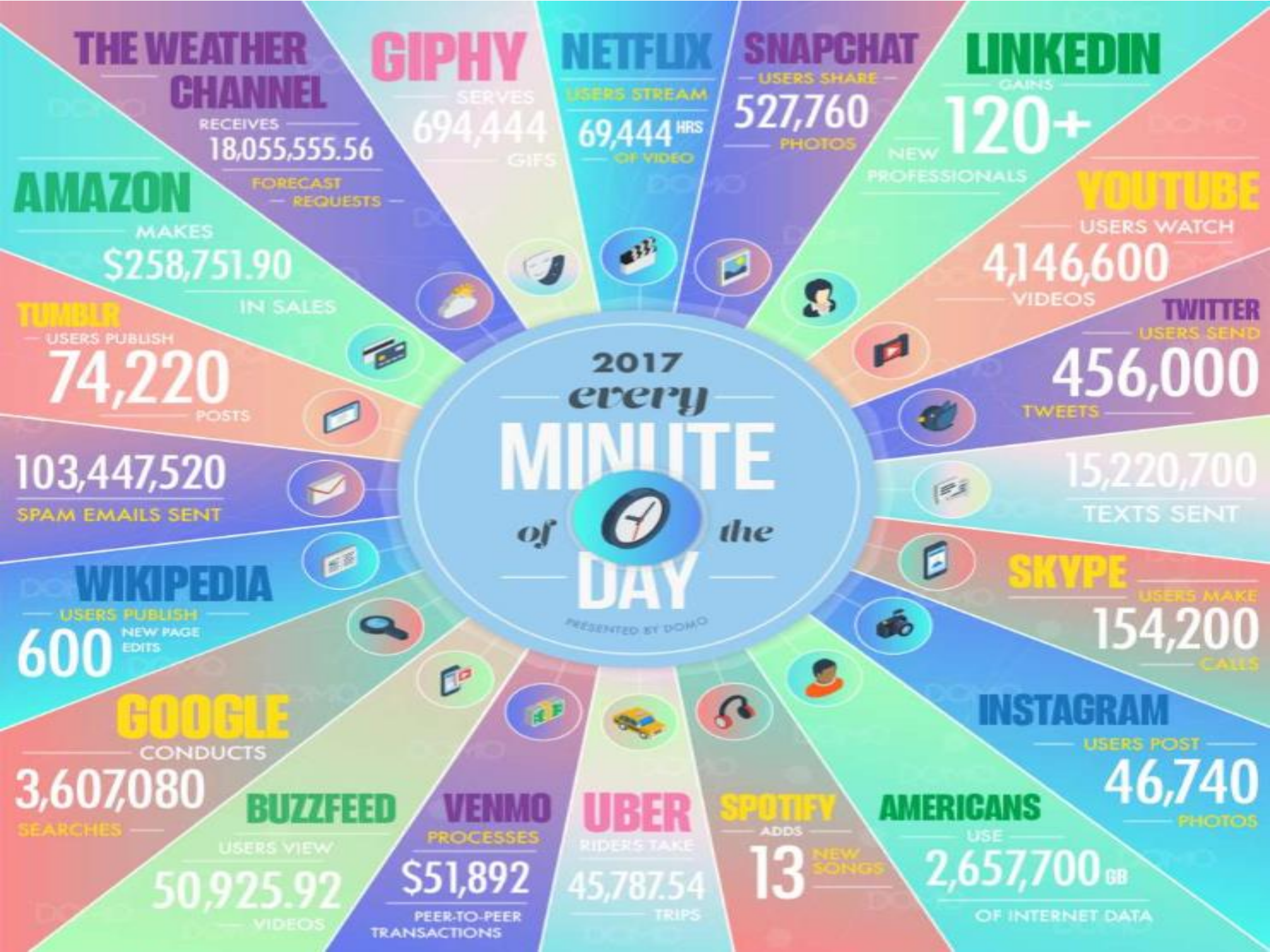
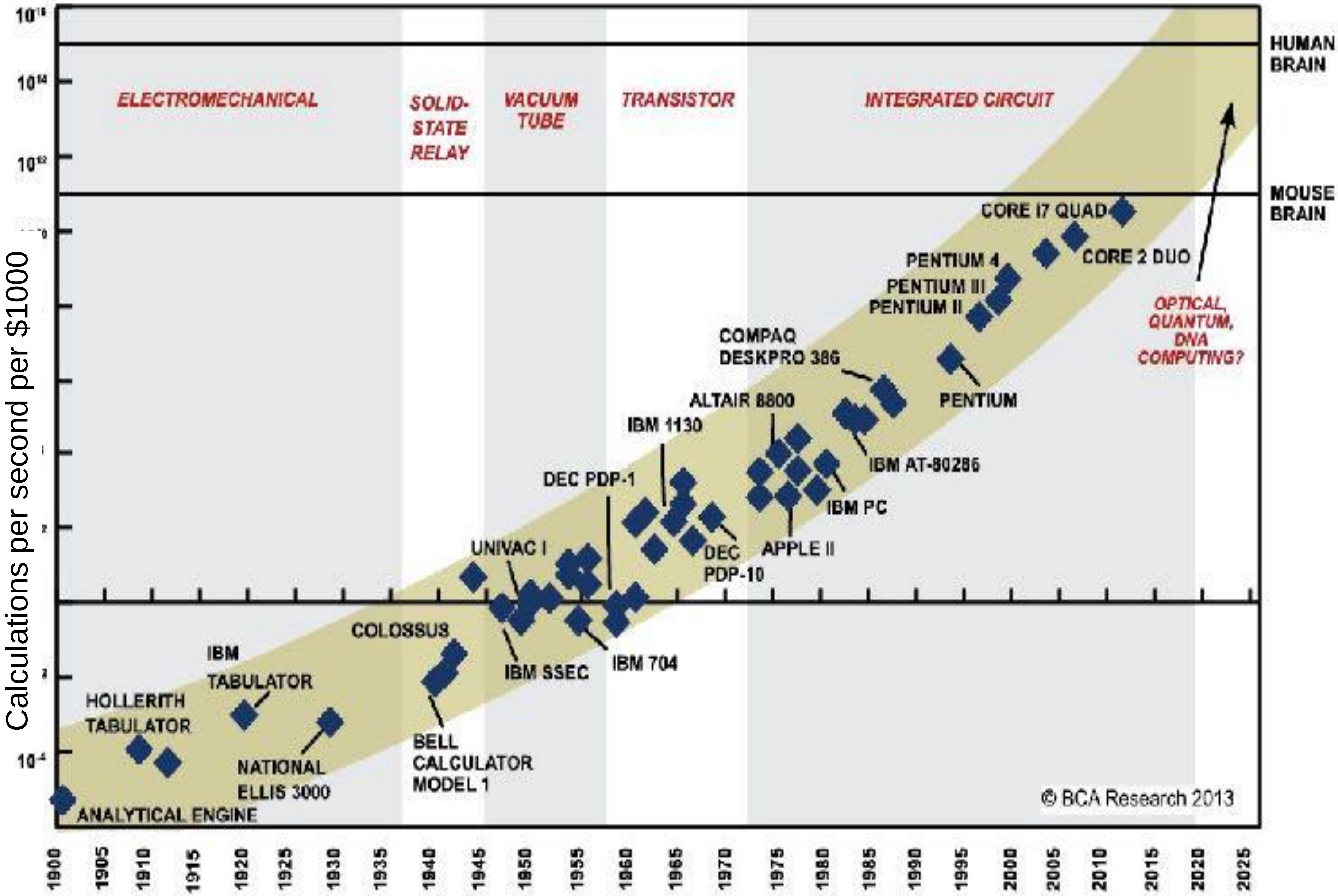
Cognitive analytics – health applications

MARELIZE GORGENS, MOHINI KAK AND DAVID WILSON
WORLD BANK

WORLD BANK GROUP



2017 every MINUTE of the DAY
PRESENTED BY DOMO
THE WEATHER CHANNEL
RECEIVES 18,055,555.56 FORECAST REQUESTS
GIPHY
SERVES 694,444 GIFS
NETFLIX
USERS STREAM 69,444 HRS OF VIDEO
SNAPCHAT
USERS SHARE 527,760 PHOTOS
LINKEDIN
GAINS 120+ NEW PROFESSIONALS
AMAZON
MAKES $258,751.90 IN SALES
YOUTUBE
USERS WATCH 4,146,600 VIDEOS
TUMBLR
USERS PUBLISH 74,220 POSTS
TWITTER
USERS SEND 456,000 TWEETS
103,447,520 SPAM EMAILS SENT
15,220,700 TEXTS SENT
WIKIPEDIA
USERS PUBLISH 600 NEW PAGE EDITS
SKYPE
USERS MAKE 154,200 CALLS
GOOGLE
CONDUCTS 3,607,080 SEARCHES
INSTAGRAM
USERS POST 46,740 PHOTOS
BUZZFEED
USERS VIEW 50,925.92 VIDEOS
VENMO
PROCESSES $51,892 PEER-TO-PEER TRANSACTIONS
UBER
RIDERS TAKE 45,787.54 TRIPS
SPOTIFY
ADDS 13 NEW SONGS
AMERICANS
USE 2,657,700 GB OF INTERNET DATA



SOURCE: RAY KURZWEIL, "THE SINGULARITY IS NEAR: WHEN HUMANS TRANSCEND BIOLOGY", P.67, *THE VIKING PRESS*, 2006. DATAPOINTS BETWEEN 2000 AND 2012 REPRESENT BCA ESTIMATES.

expectations
Connected Home
Deep Learning
Machine Learning
Virtual Assistants
IoT Platform
Autonomous Vehicles
Smart Robots
Nanotube Electronics
Edge Computing
Cognitive Computing
Augmented Data Discovery
Blockchain
Smart Workspace
Commercial UAVs (Drones)
Conversational User Interfaces
Cognitive Expert Advisors
Brain-Computer Interface
Volumetric Displays
Quantum Computing
Digital Twin
Serverless PaaS
5G
Human Augmentation
Enterprise Taxonomy and Ontology Management
Neuromorphic Hardware
Deep Reinforcement Learning
Software-Defined Security
Virtual Reality
Artificial General Intelligence
4D Printing
Augmented Reality
Smart Dust
As of July 2017
Innovation Trigger
Peak of Inflated Expectations
Trough of Disillusionment
Slope of Enlightenment
Plateau of Productivity
time
Years to mainstream adoption:
less than 2 years
2 to 5 years
5 to 10 years
more than 10 years
obsolete before plateau

Cognitive analytics in health lags

- Of all global data generated globally each year, an estimated 30% is health data
- Yet cognitive analytics in health is nascent
- In developing countries, examples are scarce – and met with skepticism
- Misperceptions of effort and investment needed
- Unaddressed privacy concerns
- Lack of electronic medical records exacerbates problem
- Lack of patient-reported health outcomes a further challenge
- But, examples have started to emerge – primarily in clinical and private health care

WORLD BANK GROUP

Yet the time is ripe for cognitive analytics in health

- Improvements in computing power – which doubles every year
- Ubiquity of cloud computing
- Improvements in natural language computing
- Expansion of **electronic health records**
- Combined **payments and clinical data**,
  with adequate privacy shields
- **Increased health data** from healthcare and personal devices
- Growth in **genomic sequencing databases**

Cognitive analytics framework

RIGHT LIVING

Informed lifestyle choices that promote well-being and the active engagement of consumers in their own care

RIGHT CARE

Evidence-based care that is proven to deliver needed outcomes for each patient while ensuring safety

RIGHT PROVIDER

Care provider (e.g., nurse, physcian) and setting that is most appropriate to deliver prescribed clinical impact

RIGHT VALUE

Sustainable approaches that continuously enhance healthcare value by reducing cost at the same or better quality

RIGHT INNOVATION

Innovation to advance the frontiers of medicine and boost R&D productivity in discovery, development, and safety

McKinsey, 2013

Cognitive analytics framework

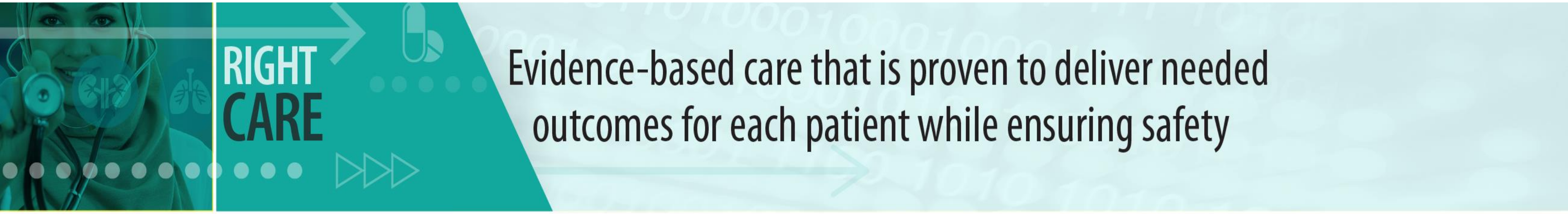
RIGHT LIVING

Informed lifestyle choices that promote well-being and the active engagement of consumers in their own care

- **Disease prevention**
- Several health insurers offer discount for wearing linked fitness band, further discounts for meeting exercise targets – and provide individually tailored health advice
- **Data-enabled adherence**
- Virta diabetes smartphone app monitors glucose, weight, blood pressure, activity diet, energy and mood and provides individually tailored patient management, enabling 87% of patients to reduce or eliminate insulin – matching highly invasive bariatric surgery. **Such apps could reduce US diabetes costs by $100 million annually**

McKinsey, 2013

Cognitive analytics framework



- **Proven care pathways**
- Artificial intelligence (IBM Watson, Microsoft Hanover) already complements radiological, heart disease and cancer screening and patient management
- MIT and Stanford deep learning algorithms support ICU staff to predict and manage patients at risk of imminent emergencies
- **Coordinated care**
- Simple informatic-based care coordination enabled Scotland's health providers to collaborate to monitor and manage diabetes patients, **reducing amputations by 30%.** Similar informatics enabled Finland to **reduce COPD hospitalization by 40%**
- Lumiata developed artificial intelligence tools for risk management and care coordination. **Using such tools, St Francis Healthcare Partners, Connecticut reduced re-hospitalization by 30%**

McKinsey, 2013

Cognitive analytics framework

RIGHT PROVIDER

Care provider (e.g., nurse, physcian) and setting that is most appropriate to deliver prescribed clinical impact

- **Artificial intelligence solutions** match patients to the right providers (nurse, physician, specialist) and care settings for the best outcomes.

McKinsey, 2013

Cognitive analytics framework

RIGHT VALUE

Sustainable approaches that continuously enhance healthcare value by reducing cost at the same or better quality

- Accolade and Health Advocate artificial intelligence solutions help patients to identify the best value medical and pharmaceutical services, keeping annual cost increases below 1%
- Kaiser Permenante's HealthConnect platform **reduced office visits by 26% and increased telephone consultations 8-fold**
- New smart senior living homes enable clients to remain in their homes, where smart sensors enable nurses to remotely track their falls, activity, sleep and eating and to provide support as required, reducing hospitalization and improving quality of life
- Artificial intelligence may transform task shifting and primary care, **providing nurses and front-line workers with highly specialized, remote cloud based decision support capabilities**

McKinsey, 2013

Cognitive analytics framework

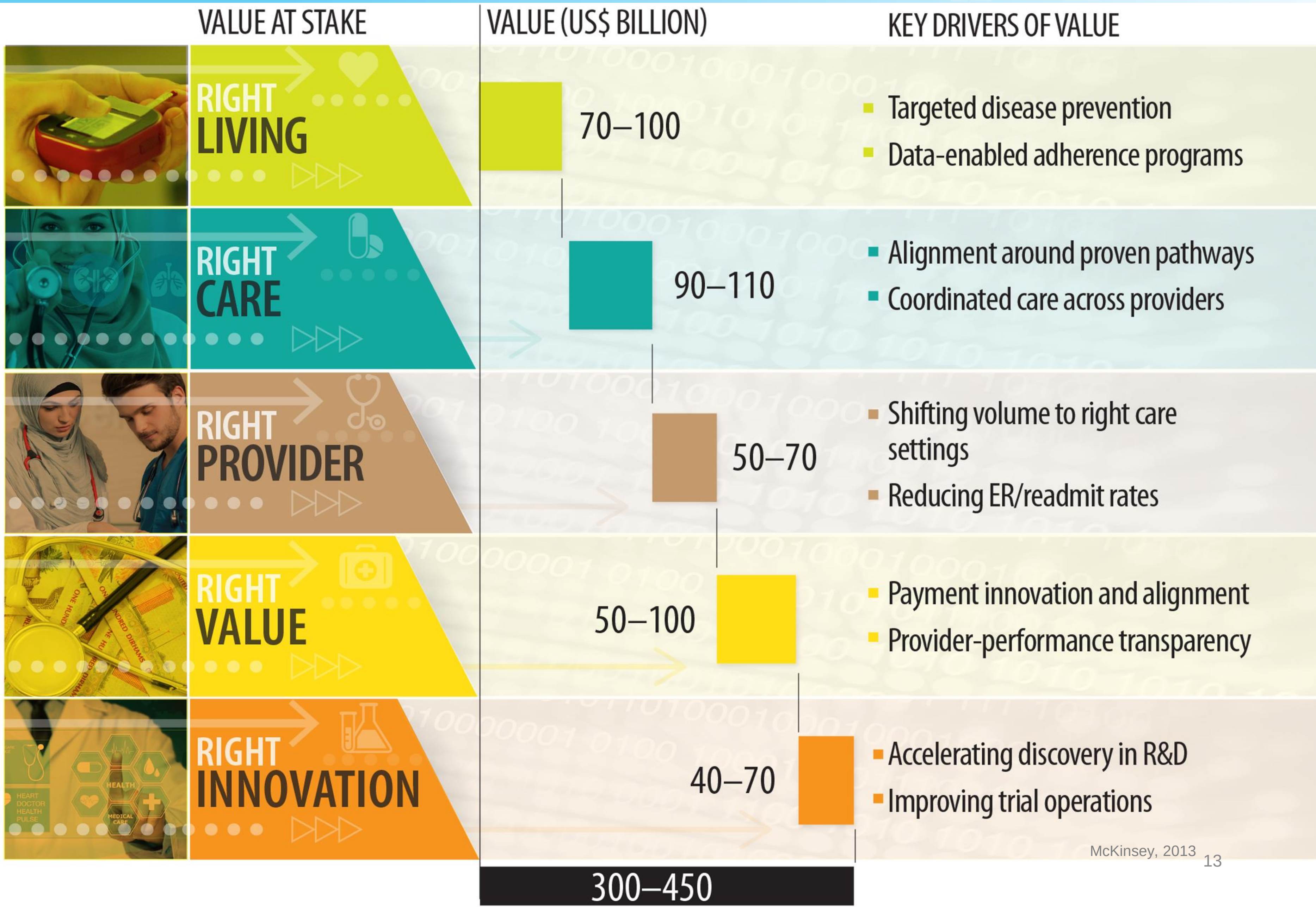
RIGHT INNOVATION

Innovation to advance the frontiers of medicine and boost R&D productivity in discovery, development, and safety

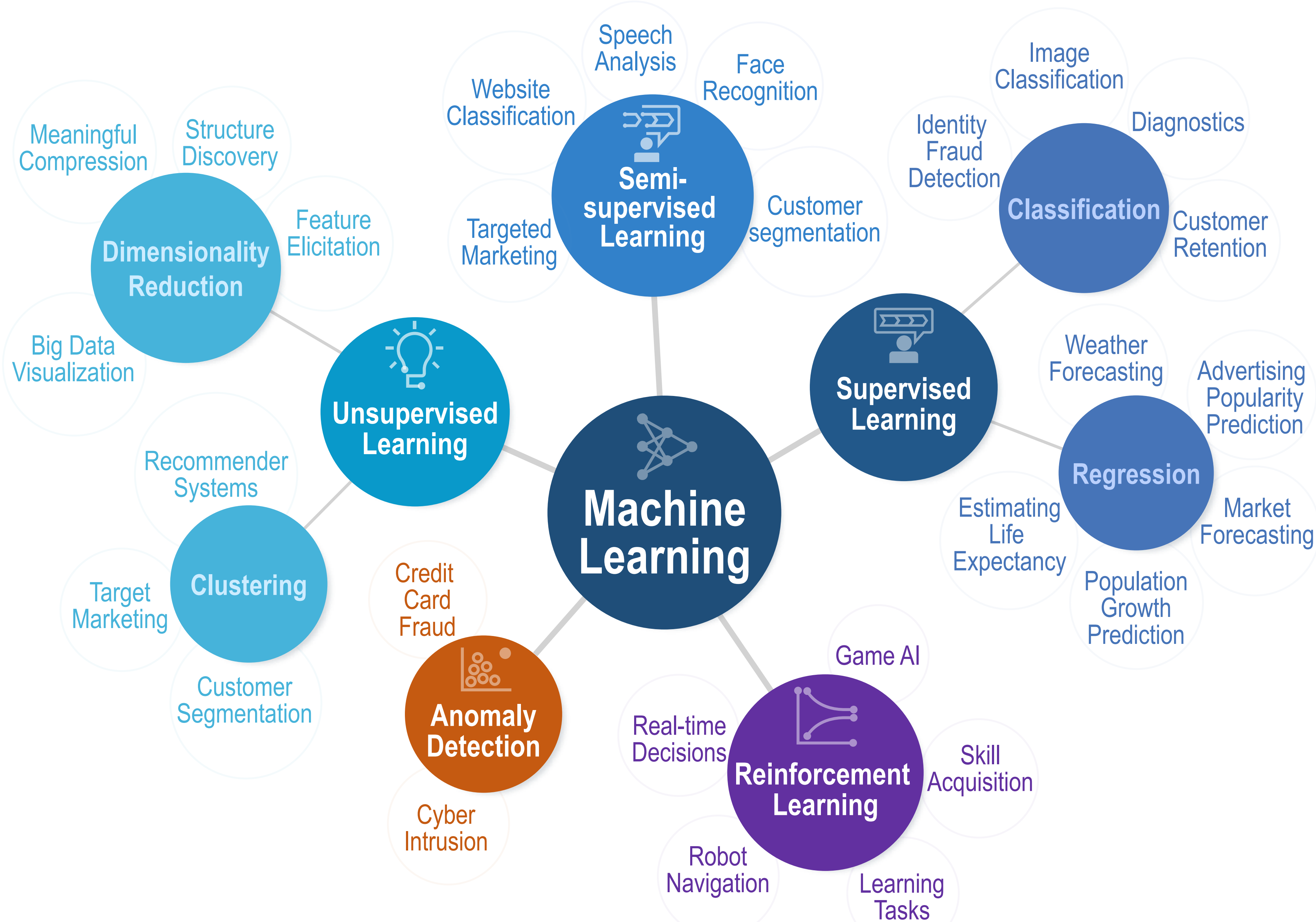
- AI is pinpointing **priority innovations and accelerating drug discovery** though better bio-process predictive modeling, expanded (social media) yet better targeted (genetic) trial recruitment and real-time trial safety monitoring

McKinsey, 2013

Potential US savings from cognitive analytics

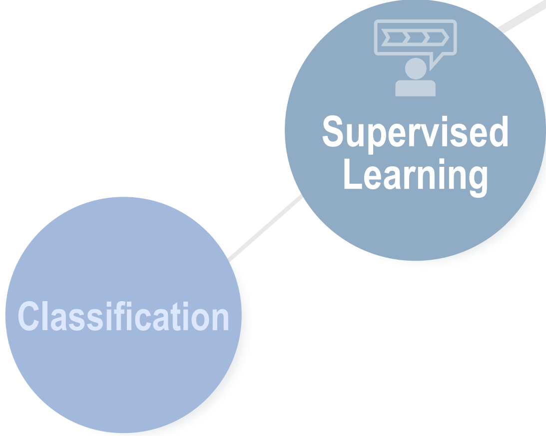


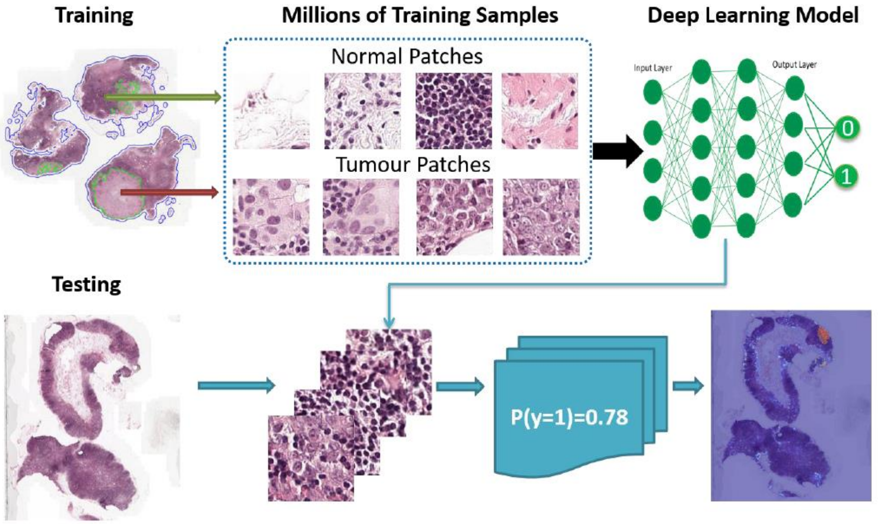
McKinsey, 2013


Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Feature Elicitation
Big Data Visualization
Clustering
Recommender Systems
Target Marketing
Customer Segmentation
Semi-supervised Learning
Website Classification
Speech Analysis
Face Recognition
Targeted Marketing
Customer segmentation
Supervised Learning
Classification
Image Classification
Identity Fraud Detection
Diagnostics
Customer Retention
Regression
Weather Forecasting
Advertising Popularity Prediction
Estimating Life Expectancy
Market Forecasting
Population Growth Prediction
Anomaly Detection
Credit Card Fraud
Cyber Intrusion
Reinforcement Learning
Game AI
Real-time Decisions
Skill Acquisition
Robot Navigation
Learning Tasks

Supervised learning: classification

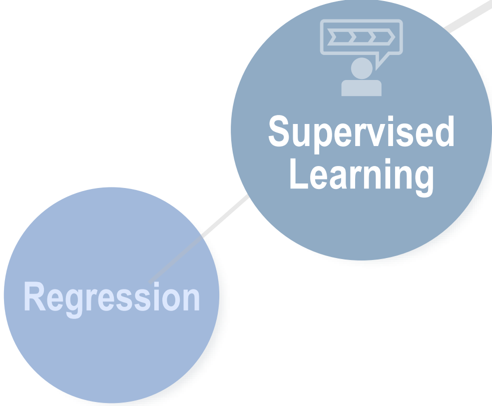
- Extensively used in medical diagnostics utilizing MRI/CT scans to detect cancer tissue, clogged arteries, and more
- Machine learns to identify which tissues are cancerous.
- Popular uses in contemporary times include handwriting and speech recognition, spam detection



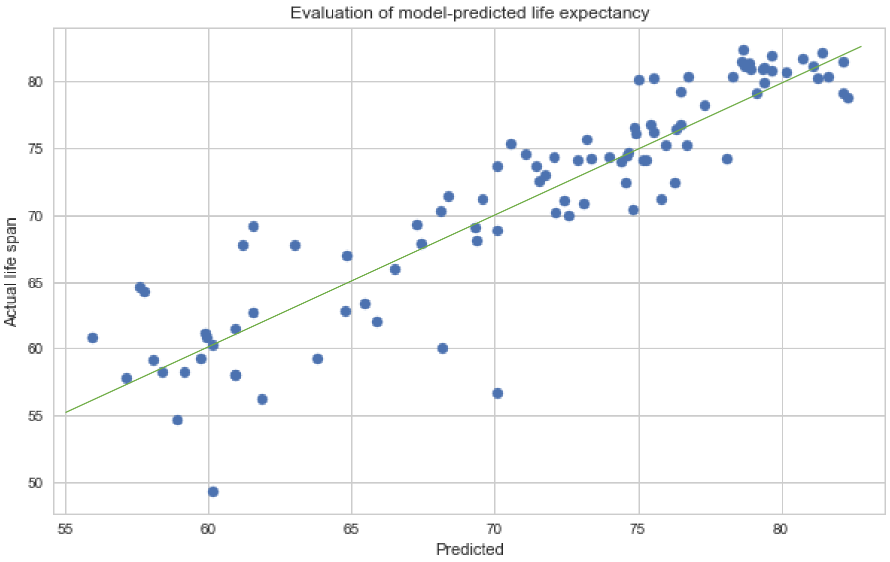


"Deep Learning Drops Error Rate for Breast Cancer Diagnoses by 85%"

Supervised learning: regression



- Applications range from prediction of life expectancy and infection rates, to market predictions
- Machine identifies which factors are statistically significant (developmental indicators such as sanitation, GDP, health expenditure) and determines correlation between factors



"Using the entire subset of 21 factors, the model predicts life expectancy with 89.7% accuracy"

Image and quote source: "Machine learning to predict life expectancy", Caitlin McDonnell, for Towards Data Science

Semi-supervised learning

Semi-supervised Learning

- Most common example is ranking models for search engines
- Search results are ranked based on "relevance"
- Initial factors of relevance are coded by humans
- Machine learns relevance as it gets more inputs from users

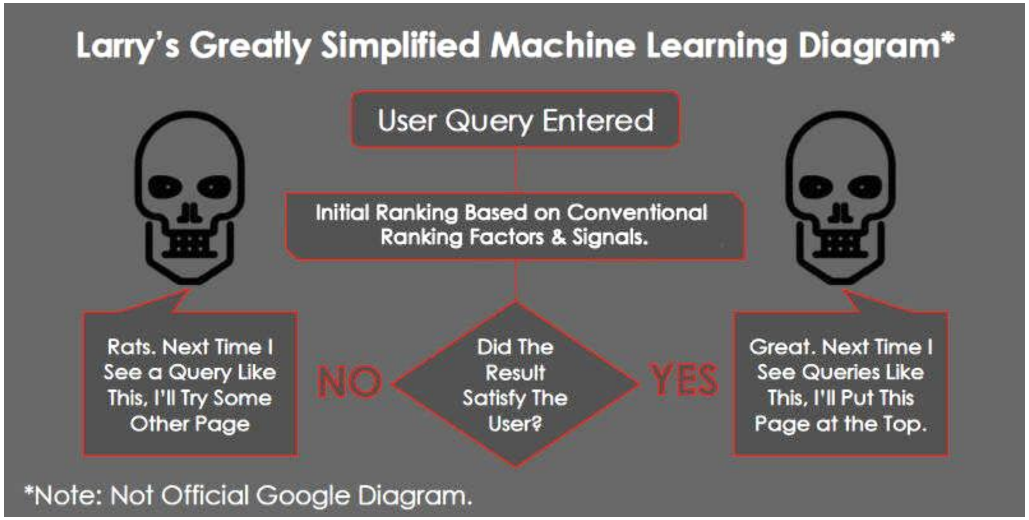
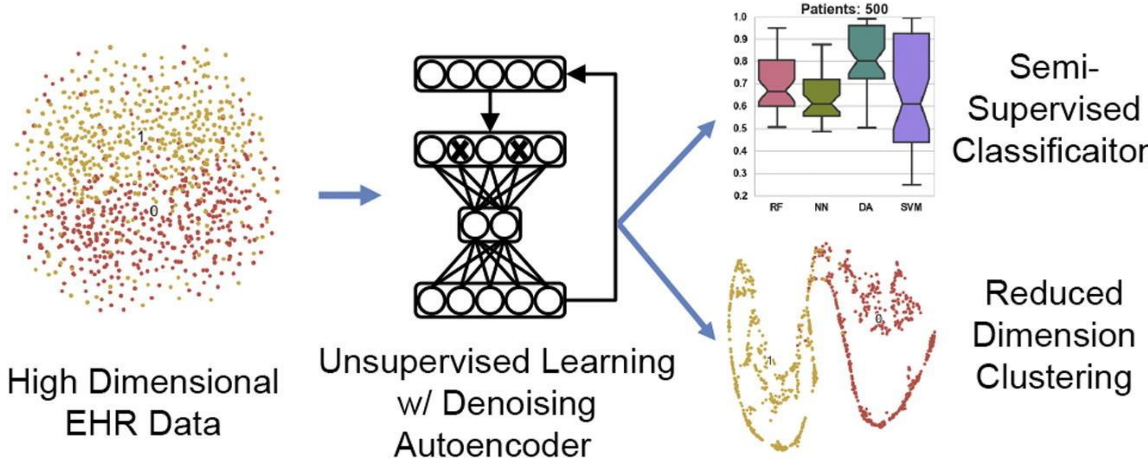


Image source: Larry Kim via Search Engine Land

Semi-supervised learning

Semi-supervised Learning

- Currently being used in the extraction of electronic phenotypes from patient health records
- These phenotypes are paired with genetic data to identify genes underlying common human diseases
- Unsupervised learning occurs on the full list of patients to identify patients with a particular disease
- This is followed by supervised learning to fine tune genetic association of diseases



Source: "Semi-supervised learning of the electronic health record for phenotype stratification". Brett K.Beaulieu-Jones, Casey S.Greene, the Pooled Resource Open-Access ALS Clinical Trials Consortium. Journal of Biomedical Informatics, Volume 64, December 2016, Pages 168-178

Unsupervised learning: dimensionality reduction

Unsupervised Learning

Dimensionality Reduction

- Especially important in big data visualization (presence of a large number of dimensions)
- These unsupervised algorithms preserve the variance and characteristics of data while making it easier to be analyzed and visually represented

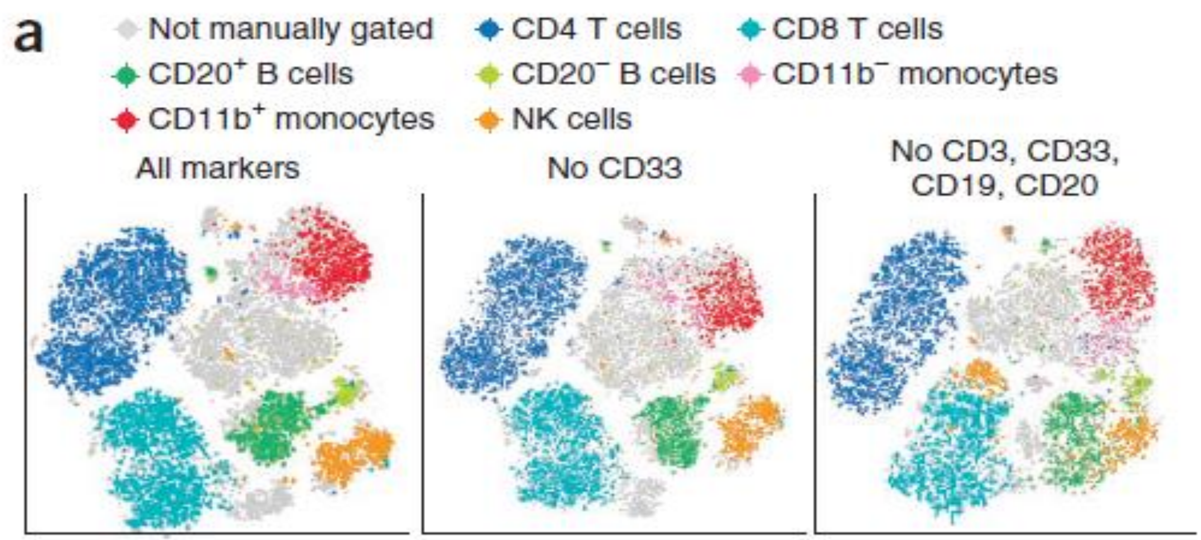
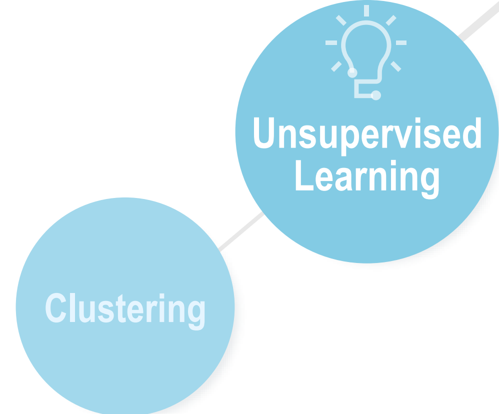


Fig: Application of viSNE (software to map high-dimensional cytometry data onto 2D plots, while conserving the structure) Image shows a healthy human bone marrow sample, stained with 13 markers and measured with mass cytometry. Despite removing 4 canonical markers (right), viSNE separates most major subtypes using the remaining 9 channels.

Image source: "viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia", El-ad David Amir, Kara L Davis, Michelle D Tadmor, Erin F Simonds, Jacob H Levine, Sean C Bendall, Daniel K Shenfeld, Smita Krishnaswamy, Garry P Nolan & Dana Pe'er. Nature Biotechnology 19 May 2013; doi:10.1038/nbt.2594

Unsupervised learning: clustering



- Used frequently in online targeted marketing and by social networking sites
- Facebook's "mutual friends" uses clustering algorithms and suggests "People you may know"

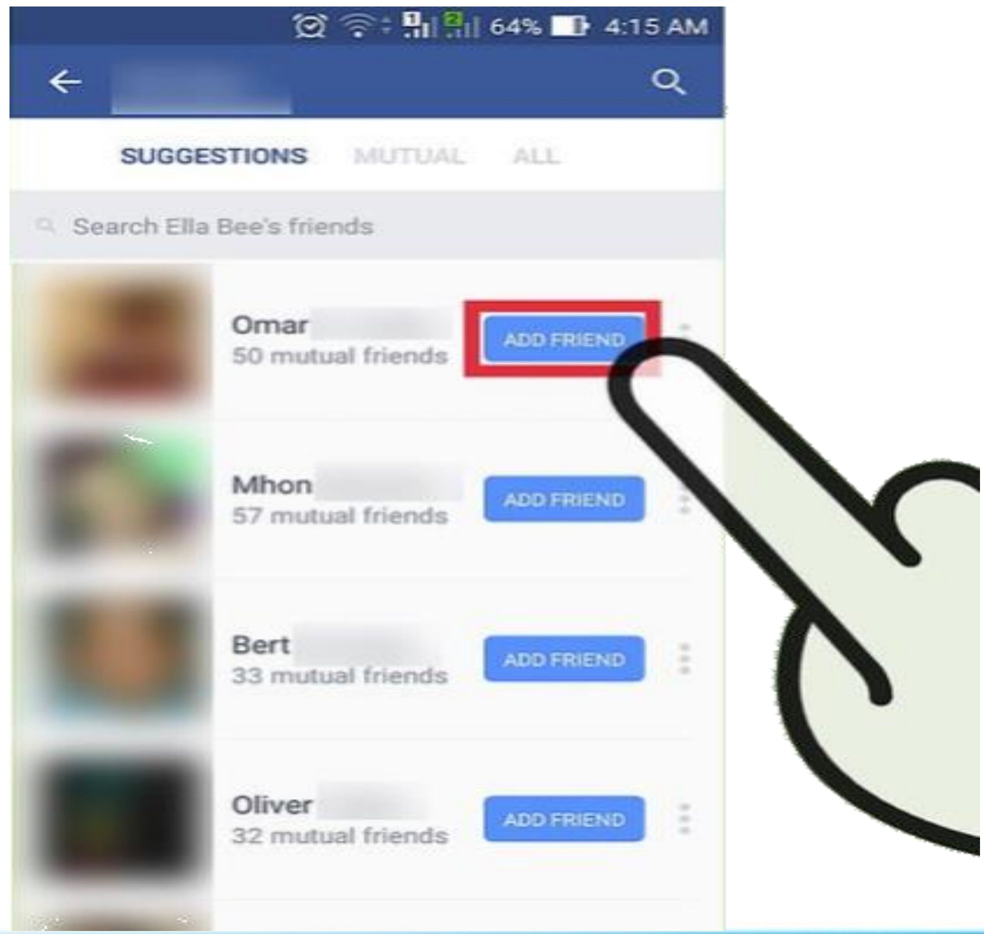


Image source: wikiHow "How to get mutual friends on Facebook"

Unsupervised learning: clustering

Unsupervised Learning

Clustering

- Used frequently where output data is unreliable or absent, and inferences have to be made only using input data
- For instance: clustering patients into those who have diabetes and those who don't, when the actual diagnosis isn't known
- Using only input factors (age, weight, blood pressure, etc.) the program clusters patients based on similarities in attributes

Source: "Step by Step to K-Means Clustering", Hanlei Zhu, Healthcare.ai

Image credit: "2014/365/307 Eight Bottles" by Alan Levine (@cogdog on Flickr), CC BY 2.0

Anomaly detection

- Used extensively by financial institutions e.g. banks, credit card companies
- Transactions are run through algorithm in real time, assessing probability of fraud based on trustworthiness of vendor, prior purchasing behavior, IP address of transaction, etc
- Anomalies in these factors immediately cause the transaction to be marked as fraudulent

Anomaly Detection

Anomaly detection

- Used extensively in wearable tech, such as blood pressure and heart rate monitors
- Critical to distinguishing faulty readings from actual emergency situations

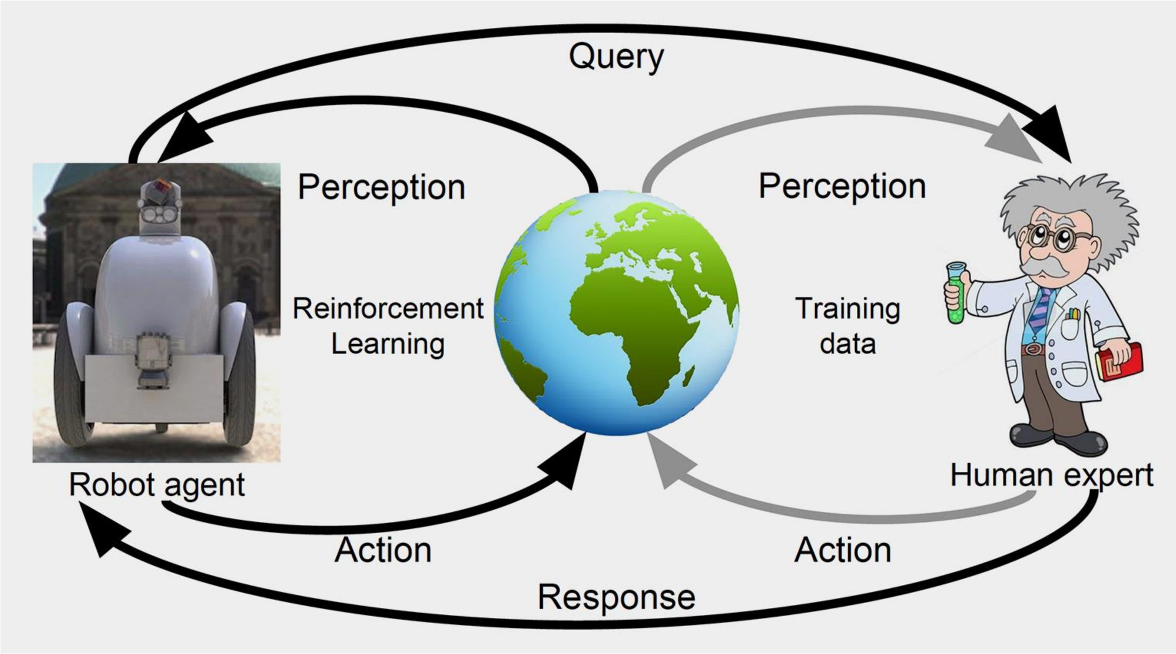
WORLD BANK GROUP

Source: "Anomaly Detection in Medical Wireless Sensor Networks using Machine Learning Algorithms". GirikPachauri, SandeepSharma. Procedia Computer Science, Volume 70, 2015, Pages 325-

Image credit: “Day 51 - 50!” by Simon James (@bearpark on Flickr). CC BY-SA

Reinforcement learning

- Being used in automated navigation systems, games, manufacturing, etc.
- Allows software agents and machines to automatically determine the ideal behavior within a specific context, to maximize its performance
- In the case of autonomous vehicles/robots, reinforcement learning is helping in decision making, speeding up the perception-to-decision process



Reinforcement Learning

WORLD BANK GROUP

Image credit: "Closing the Visual-Motor Loop with Deep Reinforcement Learning", Fei-Fei Li, Stanford University.

Reinforcement learning

- Systems process information, automatically decide on an action, and then compare the outcome against a “reward” definition
- Being used to optimize and increase the precision of tumor classification in medical imaging

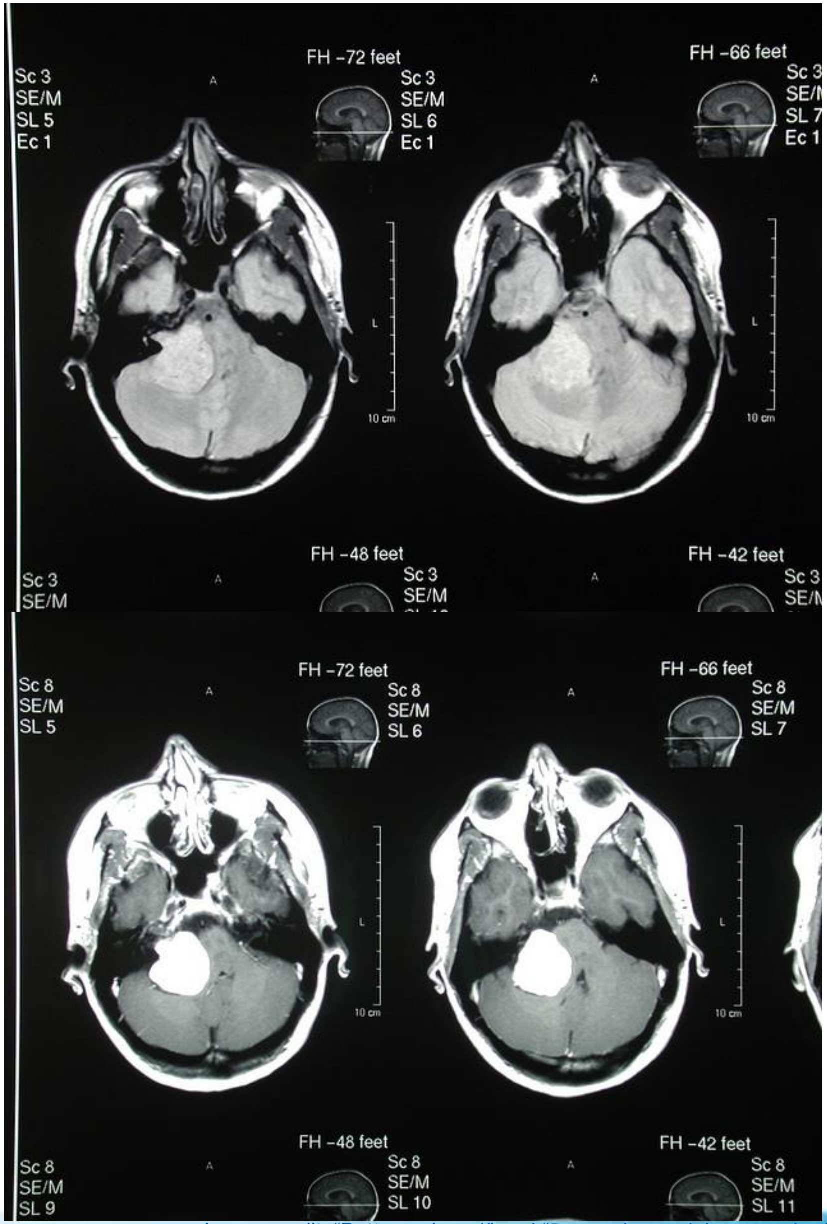


Image credit: “Bottom_view_1” and “Dye_enhanced_bottom_view” by Eric Lewis (@subewl on Flickr). CC BY-SA 2.0

Source: "Application on Reinforcement Learning for Diagnosis Based on Medical Image". Stelmo Magalhães Barros Netto, Vanessa Rodrigues Coelho Leite, Aristófanes Corrêa Silva, Anselmo Cardoso de Paiva, Areolino de Almeida Neto. January 1st 2008. DOI: 10.5772/5291

Potential of Big Data To Improve National Nutrition Mission

Importance of ICDS-CAS

- ICDS-CAS is a major initiative; amongst the largest digital health initiatives globally
- However, its full potential is currently untapped as dashboards rely heavily on the analytical ability of the user to interpret and use data for decision making
- It can be taken to the next level by incorporating cognitive computing

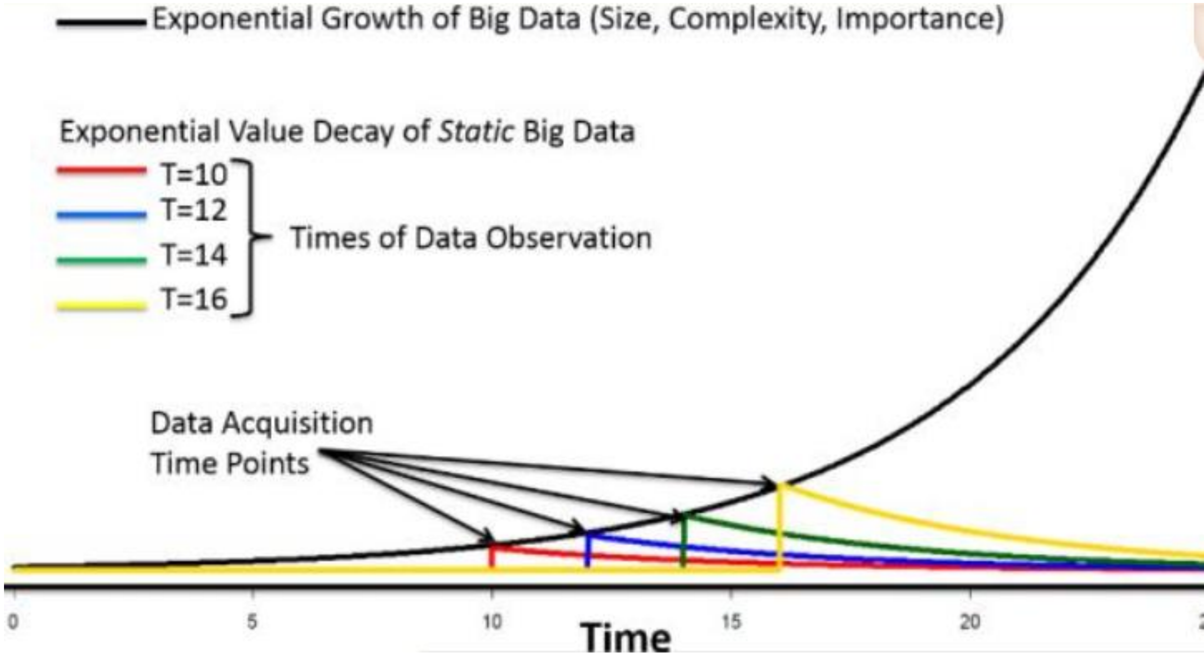
Role of cognitive computing in ICDS-CAS

- Anomaly detection: outlier AWCs and AWWs for learning and additional support
- Identify key problems in the ICDS-CAS and pre-emptively address technical challenges quicker than through manual review of helpdesk records
- Targeting of performance improvements: Identify best and worst performing AWCs
- Clustering of types of beneficiaries:
- Predicting future outcomes in health outcomes amongst mothers and children
- Correlate education, health and other data with health outcomes of beneficiaries
- Correlate AWW behaviour and daily tasks with health outcomes of beneficiaries
- Identify AWCs for data auditing to improve quality of routine data

WORLD BANK GROUP

Caveat: Cognitive analytics means data must be up-to-date

- Declining value of static data
- Potential of the ICT-RTM + to replace other data: only if such data are current
- Data fundamentals – quality, relevance, verification and use – don't disappear in face of big data – no magic bullet



WORLD BANK GROUP