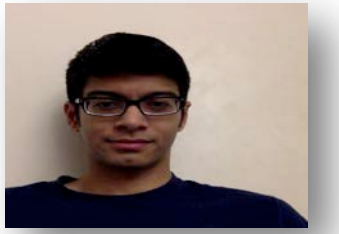
MUNTHER A. DAHLEH

Economics and Market Design for Data

Thanks to my Collaborators

Anish Agarwal



Tuhin Sarkar

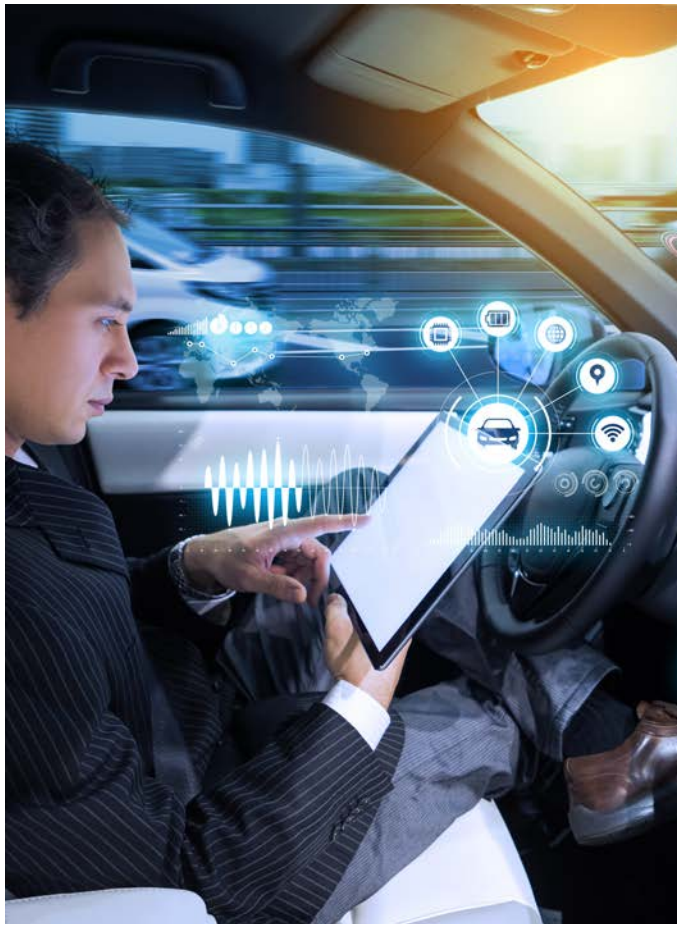
Devavrat Shah

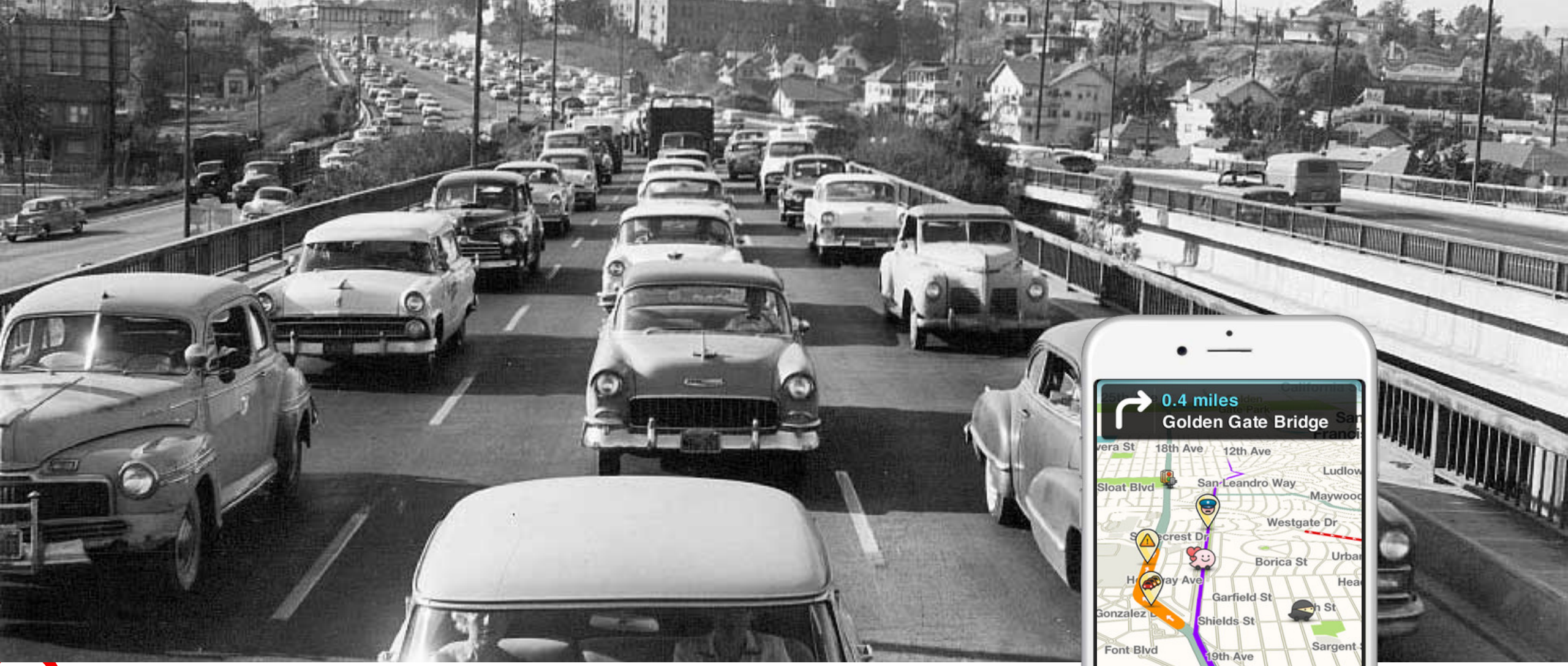
Mardavij Roozbehani

Flora Meng

Maryann Rui

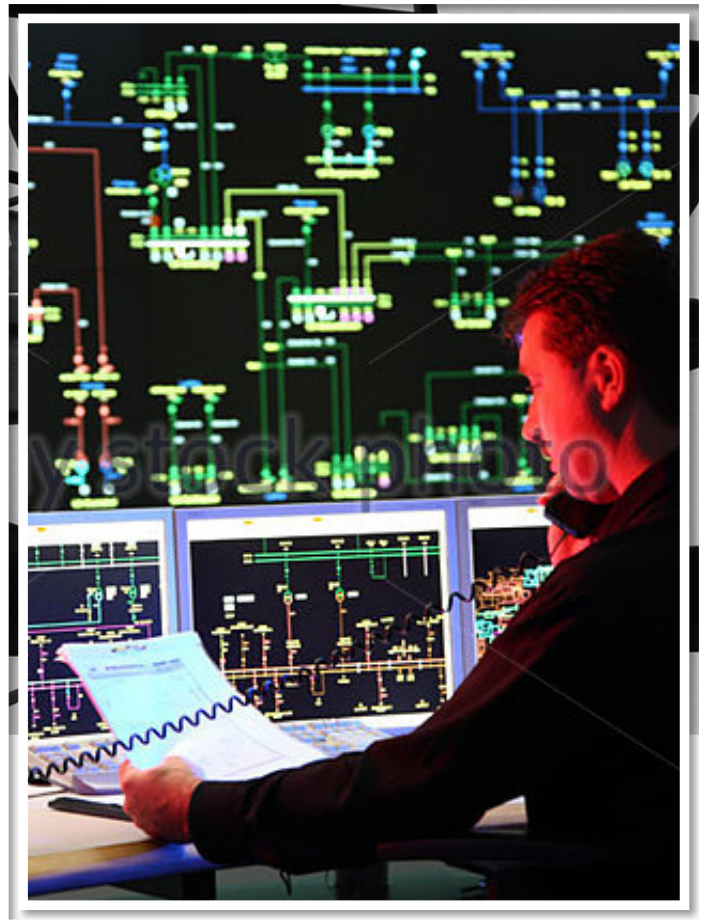
Computing and Data





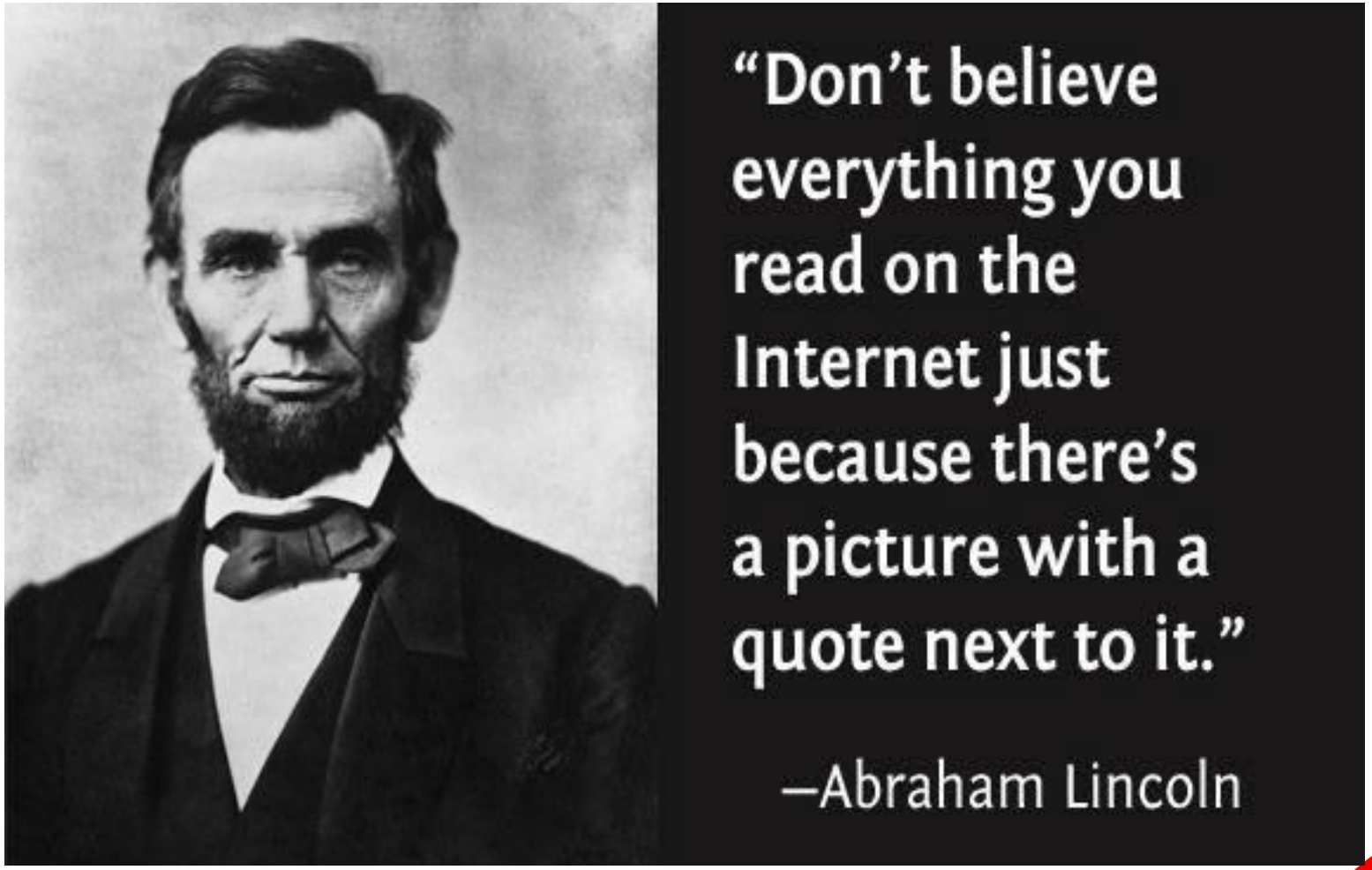
Congestion Going From Bad To Worse

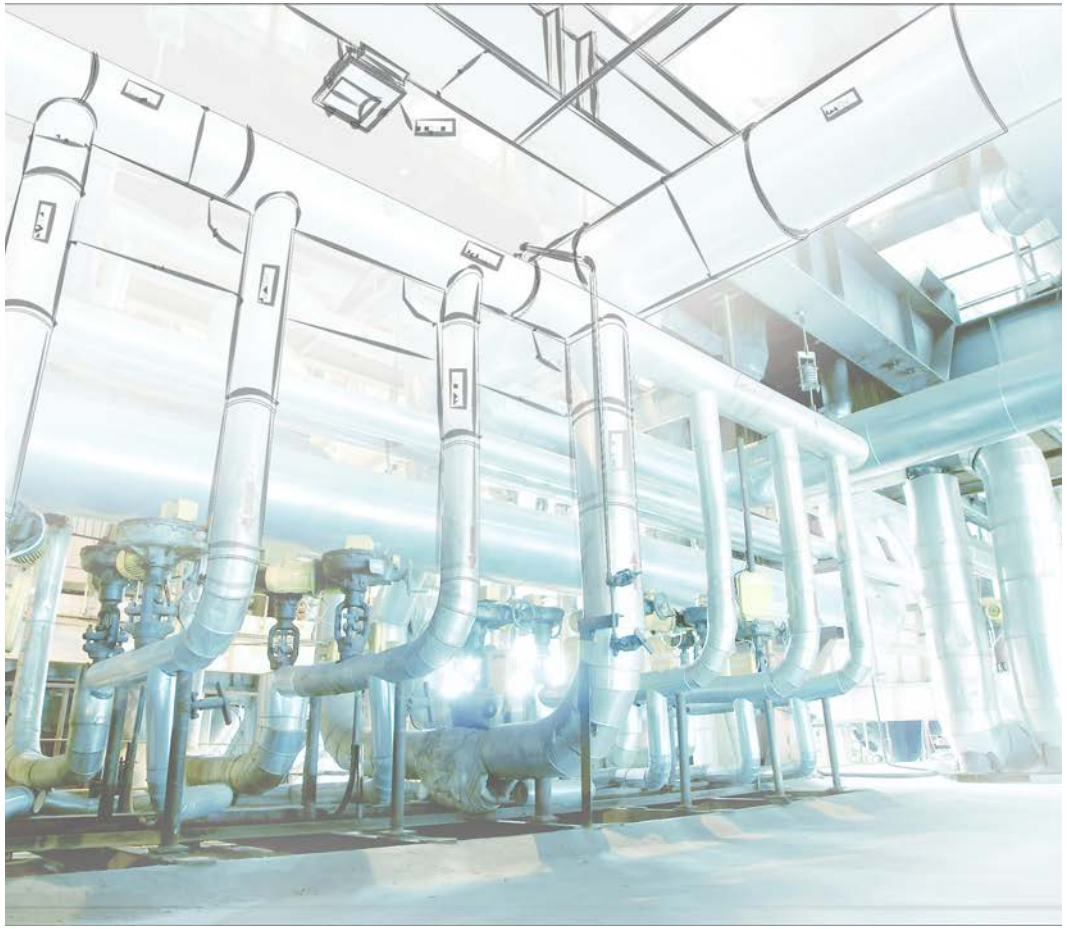
What if the Lights Go Out?



Too Big To Fail

Misinformation

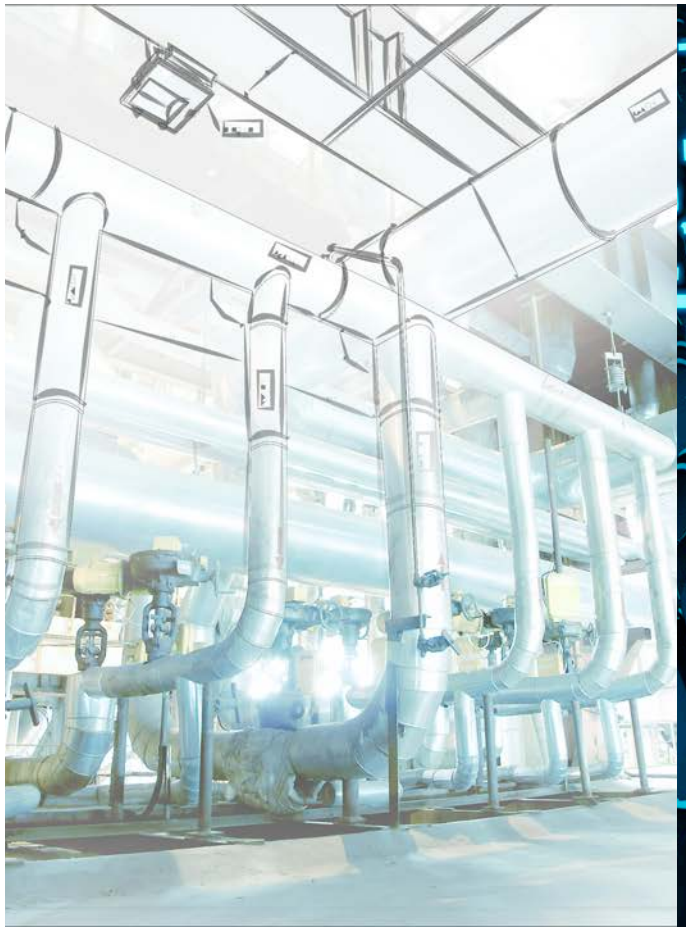




Physical and Engineered Systems

Institutions

Paradigm Shift



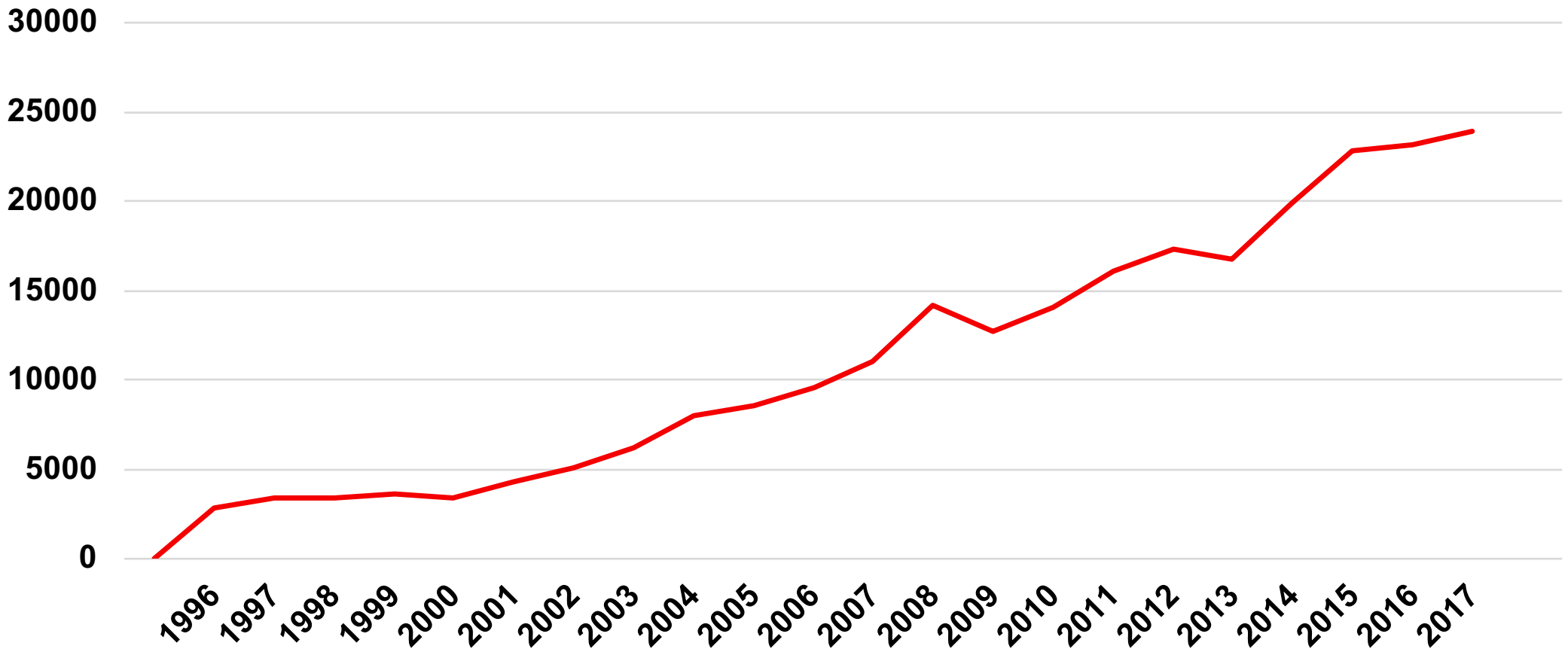
Physical and Engineered Systems

Society

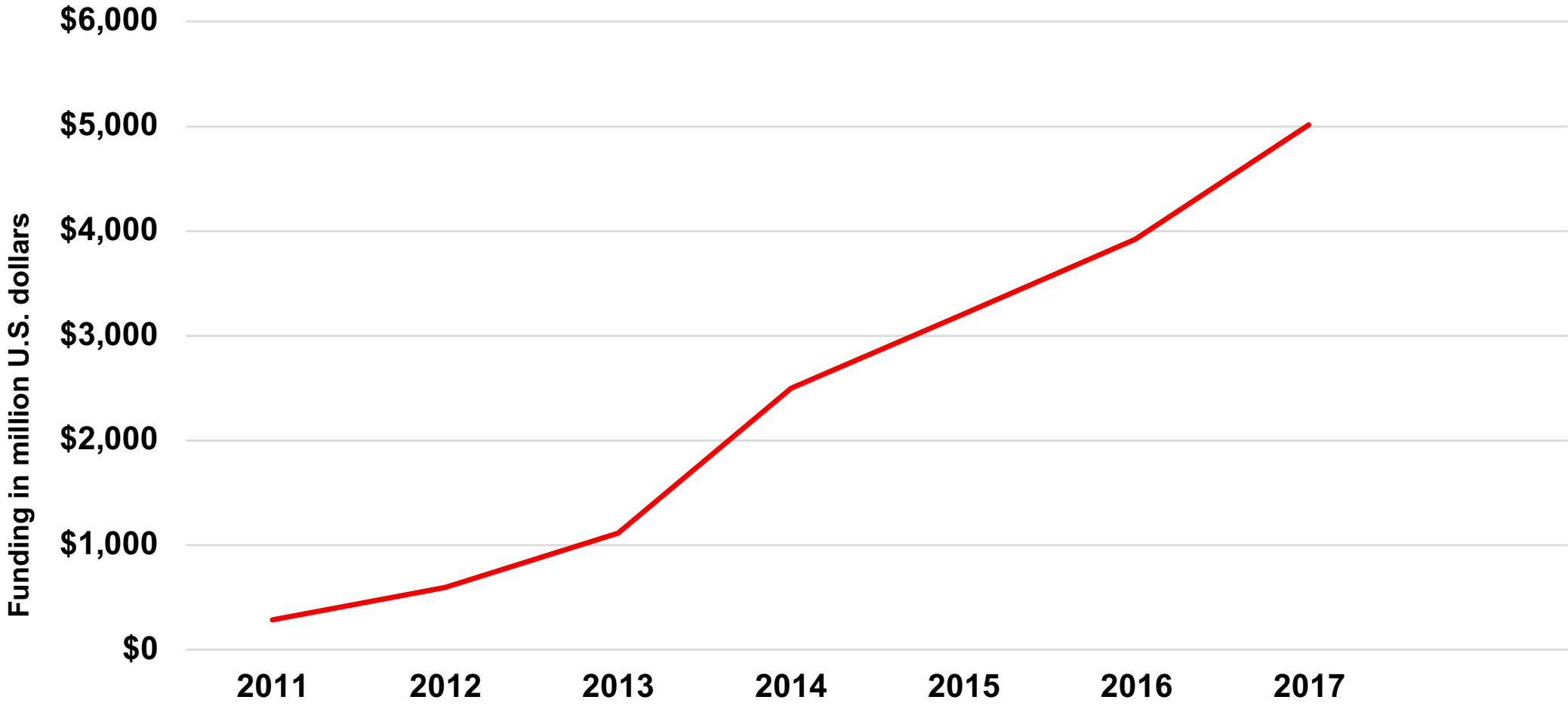
Institutions

Data as a Commodity

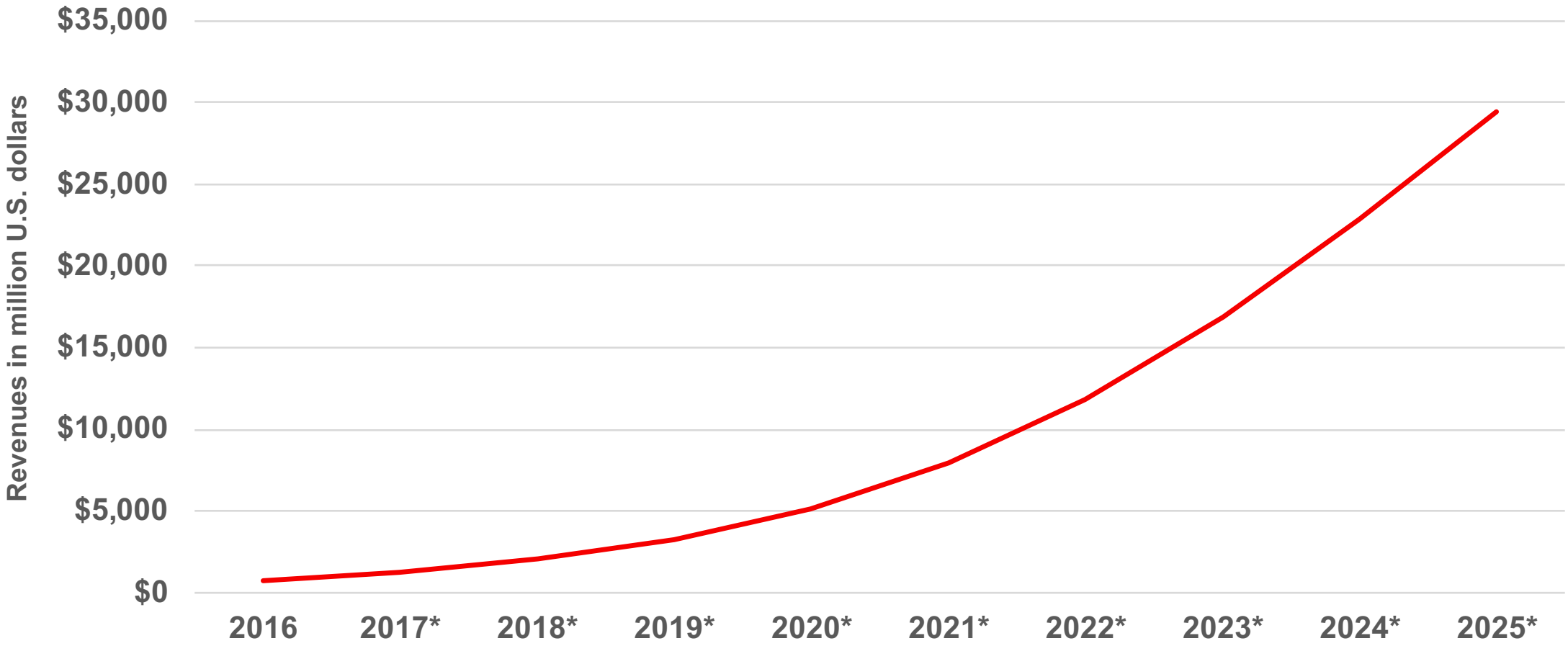
Papers Published on AI (Scopus.com)



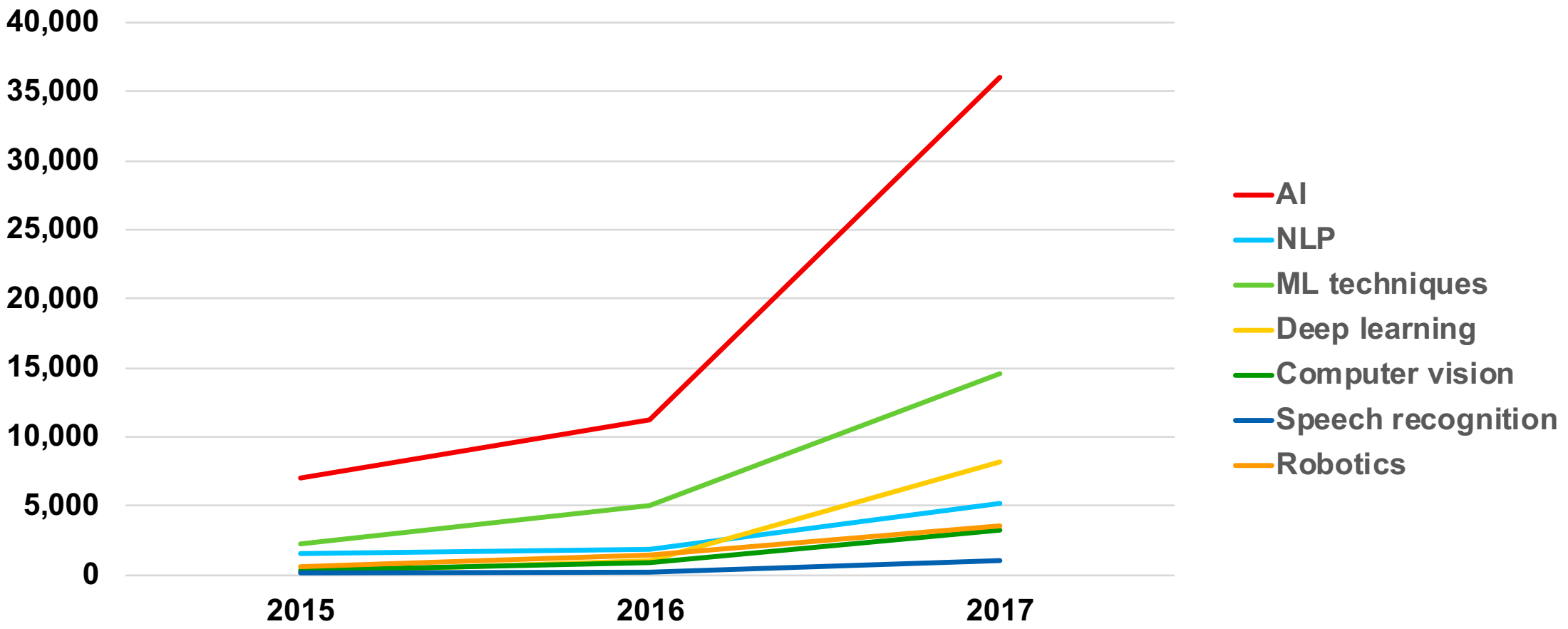
AI Funding in the U.S.



Revenues from AI Market in North America



Job Openings (Monster.com)



“Personal data is the new oil of the internet and the new currency of the digital world.”

Meglena Kuneva, Head of the EU Delegation to the Council of Europe

Bloomberg

23andMe®

Cambridge Analytica

EQUIFAX

Why design a digital marketplace now?

Today customers are not part of the AD market.

Overview of Online Ad Data Market

Visits Webpage

Publisher

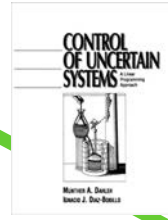
Ad Request

Ad Exchange

doubleclick by Google
Facebook Exchange
RIGHTMEDIA
appnexus
OpenX
Microsoft Exchange

Data Tracked

Targeted Ad Shown



Bid Request

Data Management Platform

nielsen

Adobe Audience Manager

Data Sold

Demand Side Platforms

adform
doubleclick by Google
rocketfuel
MediaMath
appnexus

Market has Two Major Inefficiencies

(1) Users: Do not decide what information about them is released and do not get paid for loss of privacy

Data Tracked

Data Management Platform

nielsen

Adobe Audience Manager

(2) Advertisers: Stuck in inefficient long-term contracts where they purchase user data without knowing how much value it provides

Data Management Platform

nielsen

Adobe Audience Manager

Data Sold

Demand Side Platforms

adform

doubleclick by Google

rocketfuel

MediaMath

appnexus

Zorro: System to Price Online Cookie Data

Visits Webpage

Publisher

Ad Request

Ad Exchange

Bid Request

Demand Side Platforms

Web Request
Anonymized

Payment Made
for Data Provided

ZORRO

Anonymized
User Data

Payment based on
Value of Data (VoD)

Economic Impact of Zorro is Vast

What is VoD?

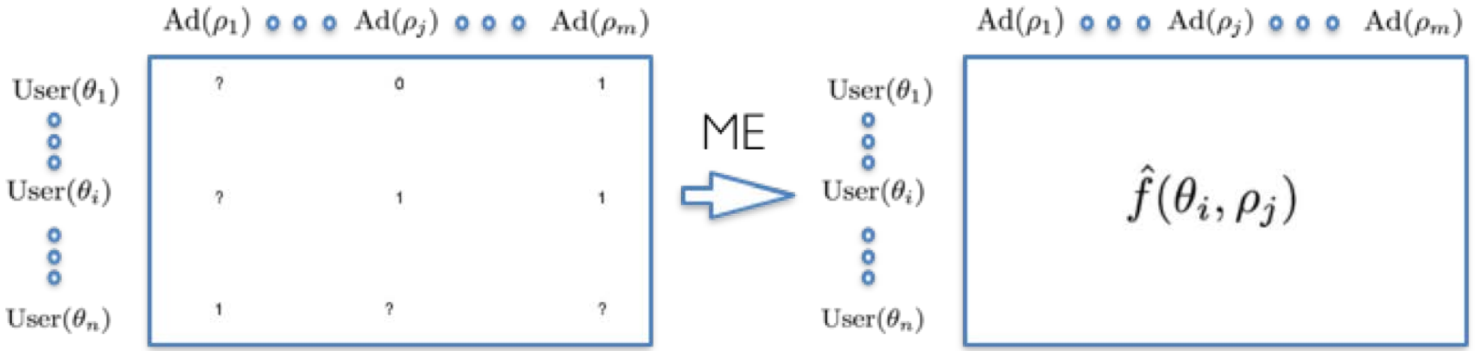
- If Zorro in place, then not getting user data will cost advertisers at least ~16B USD in lost efficiency by not being able to target ads (54B USD display advertising industry, 27% YoY growth)

- On large dataset of 190M clicks, VoD of user data through Zorro is between 30% to 70%

How can we estimate VoD without access to advertiser's model?

- Can do so by applying matrix estimation. Only need user click data, not advertiser model

- R^2 of 0.6. On closely related content recommendation dataset (Movielens), R^2 of 0.4 - 0.5



θ_i : User Browsing history and/or Intent Data

ρ_j : Latent parameters defining advertiser "type"

Value of Data: $|\hat{f}(\theta_i, \rho_j) - \mathbb{E}_\theta[\hat{f}(\theta_i, \rho_j)]|$

Releasing a Google Chrome Extension Soon…

Zorro Chrome Extension

Complete Blockage of Data
- Secure AdBlock Plus

User Allows Browsing History to be Released
- User gets paid for revealing this information to Advertisers

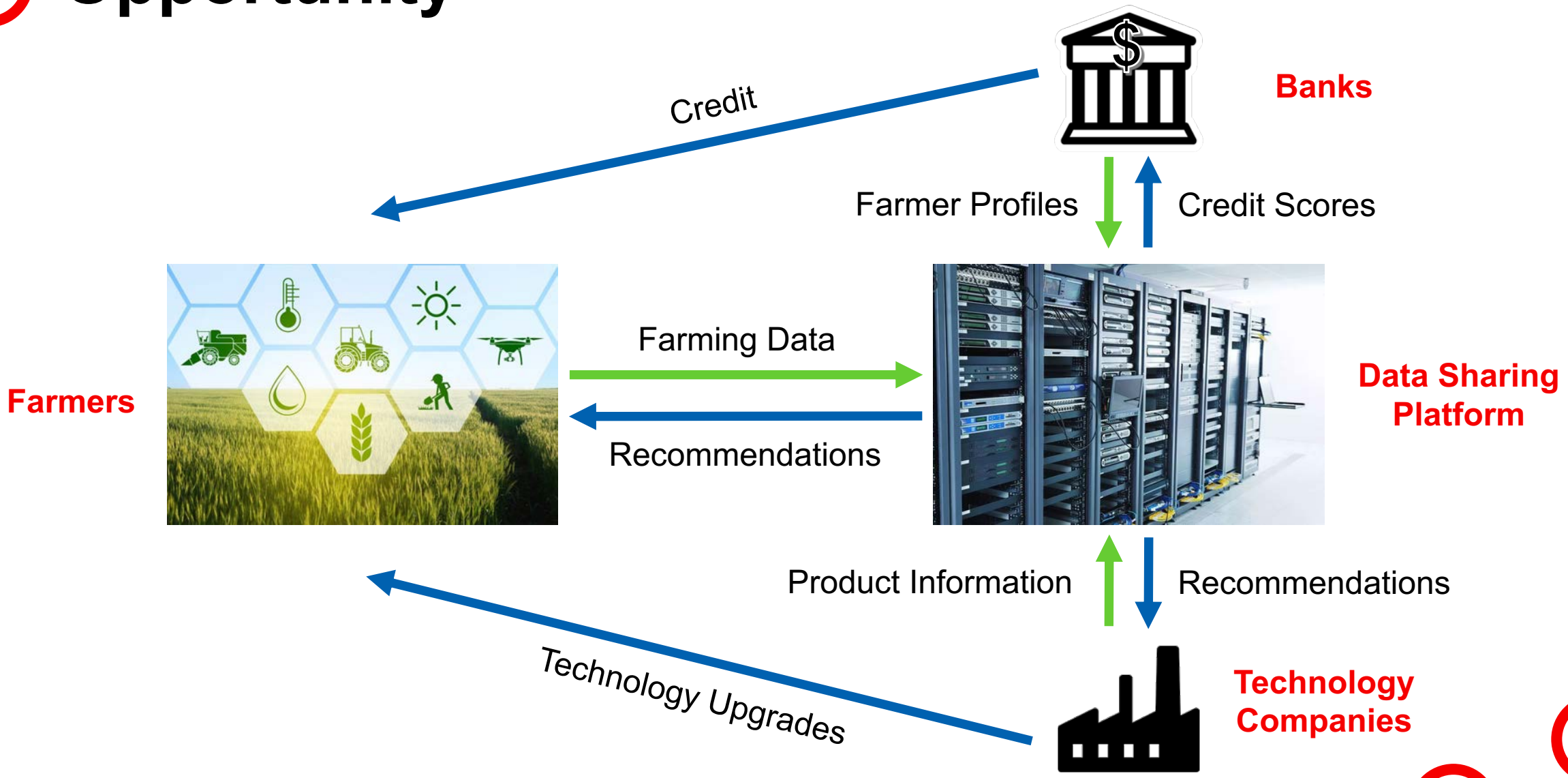
User Reveals Intent
- e.g. "Want to go on a trip to Asia in 3 months"
- User show exclusively ads related to revealed intent

User gets explicit reward for revealing personal information. Can also provide forward-looking intent, large loss of value in current system.

Challenge: Sub-saharan Africa Farming

- Farmers are poor despite their assets
- Risk of financing
- Barrier to upgrading production
- Data sharing can empower farmers

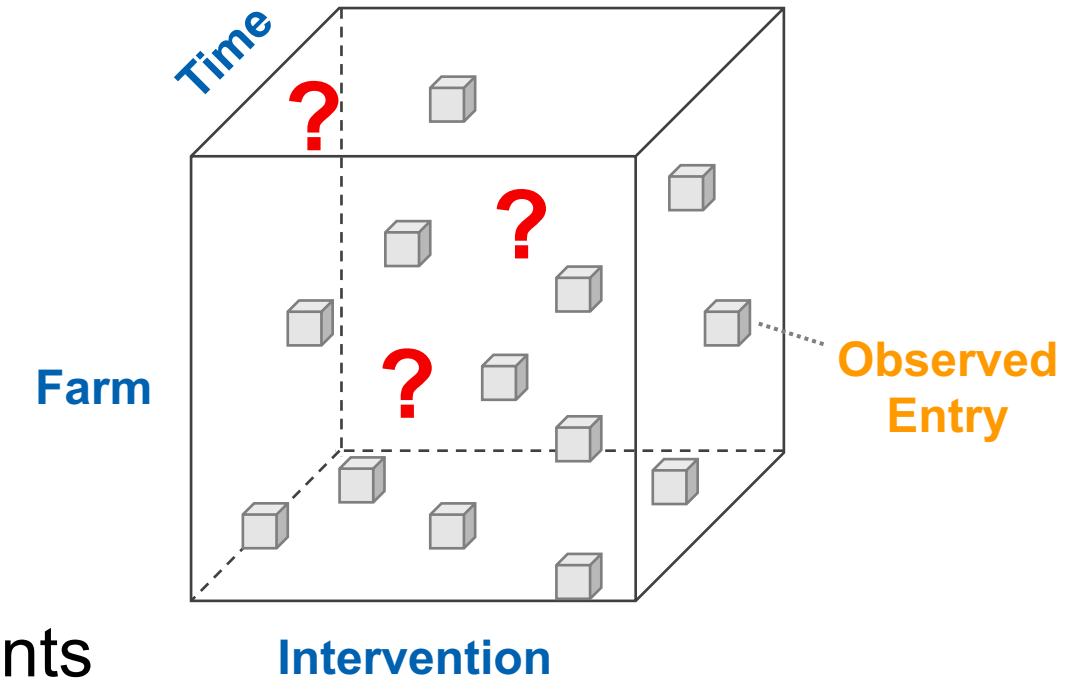
Opportunity



Data Sharing Platform

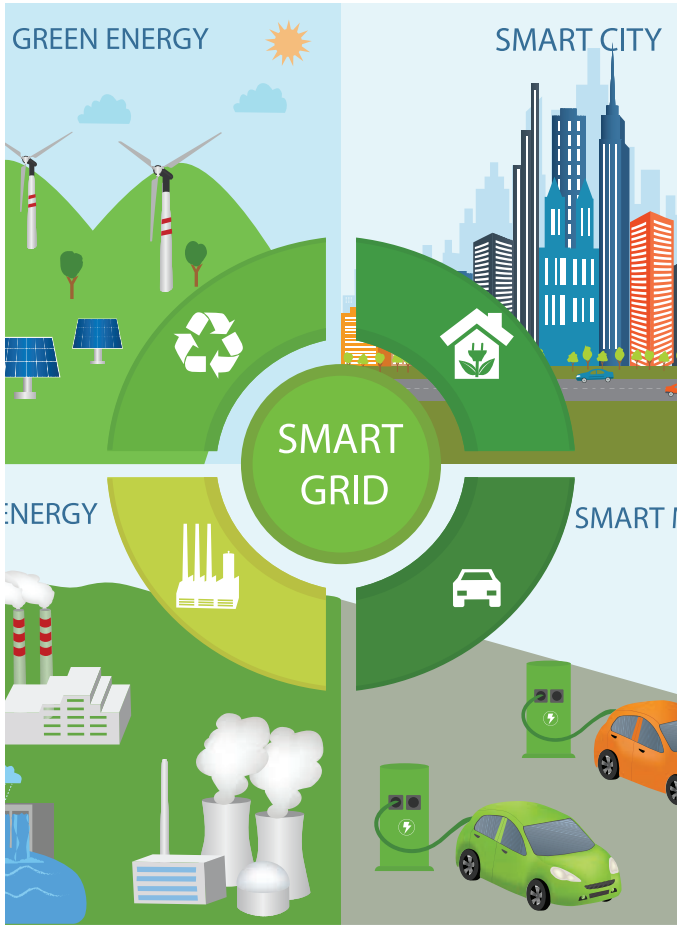
- Incentives to share high-quality data
- Types of data needed for various stakeholders

Algorithmic Challenges



- Sparse interventions & measurements
- Exploration vs. exploitation
- Inference of value added by technology

Two-Sided Markets



What makes data different as an asset class?

- Replication is at zero marginal cost
- Value is inherently combinatorial
- Prediction tasks and accuracy can vary widely
- Authenticity and value are difficult to verify
- Value depends on what others can access

Current real-time marketplaces are insufficient for data

Ad Auctions

Google AdWords			
Rank[4]	Relative Impressions	Relative CTR	Click Potential
1	100.0%	100.0%	100.0%
2	77.2%	77.4%	59.8%
3	71.3%	66.6%	47.5%
4	67.9%	57.4%	39.0%
5	65.8%	52.9%	34.8%
6	62.3%	50.2%	31.3%
7	60.6%	39.7%	24.0%
8	58.3%	34.3%	20.0%
9	58.6%	26.0%	15.3%
10	52.6%	26.3%	13.9%

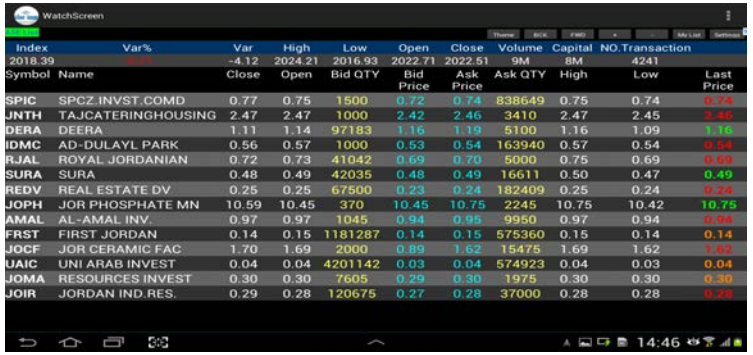
Cannot replicate ad-space/stock

- Market doesn't need to do "price discovery" & allocation
- Simply allocate good to highest bidder

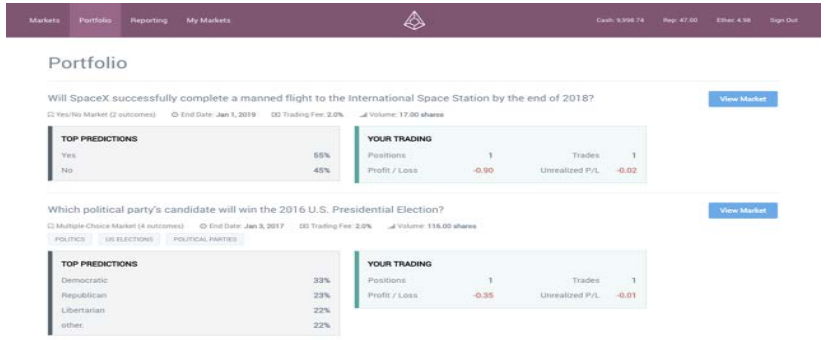
Data has no analogy to historical click-through rate/historical stock performance

- Buyers have strong priors on the value of an ad slot/value of a company
- Data Markets - buyers do not have such priors
    - Past success of data has little meaning for current buyer
    - Matching needs to be done based on which datasets are empirically the most useful for buyer

Stock Market



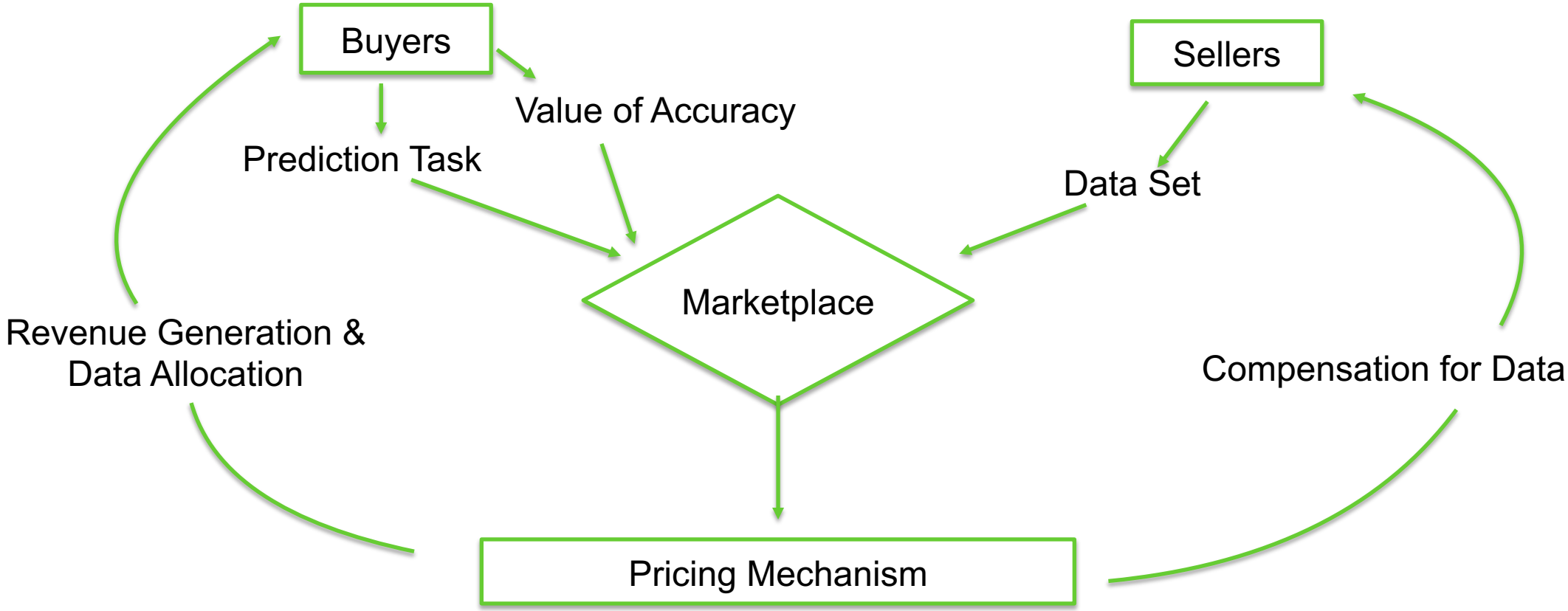
Prediction Markets



Prediction tasks and experts can be matched easily

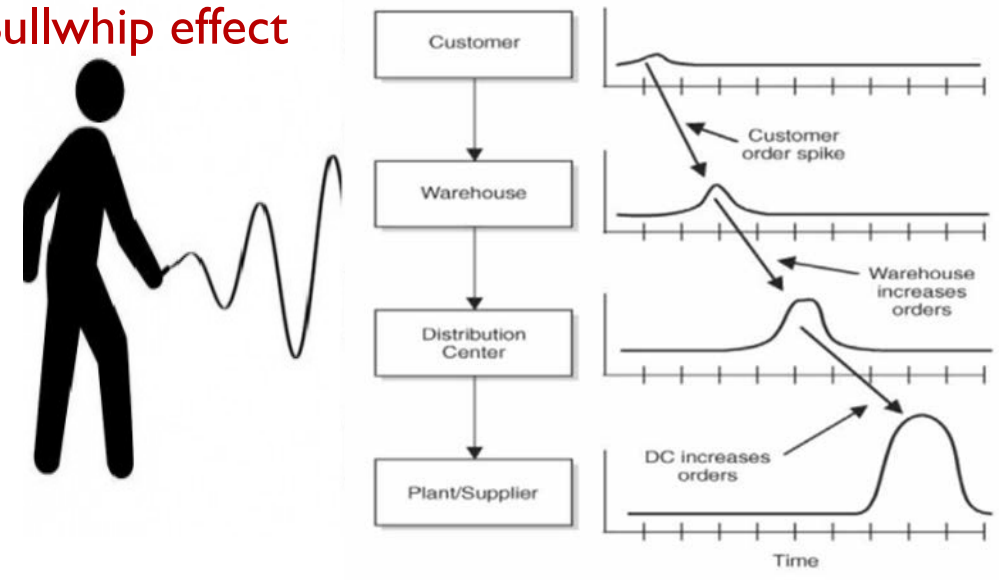
- Prediction Markets - no need to actively match experts with prediction tasks
- Data Markets - difficult to know which data is useful for which prediction task a priori

Robust Matching Mechanism



Logistics Market

Bullwhip effect



Data Buyers - Logistic/manufacturing companies

- Companies have well-defined cost models for not predicting demand well:
  "10% over/under-capacity costs me $10,000 per week"

- Can make bids to the market of the following form:
  "Willing to pay $1000 per % decrease in over-supply from previous week"

What do Data Purchasers (Buyers) need to decide?

- Machine Learning algorithm
  - Select from pre-set drop-down list

- Prediction Gain function
  - Standard in open-source libraries ---dependent on prediction task

- Time-scale of data

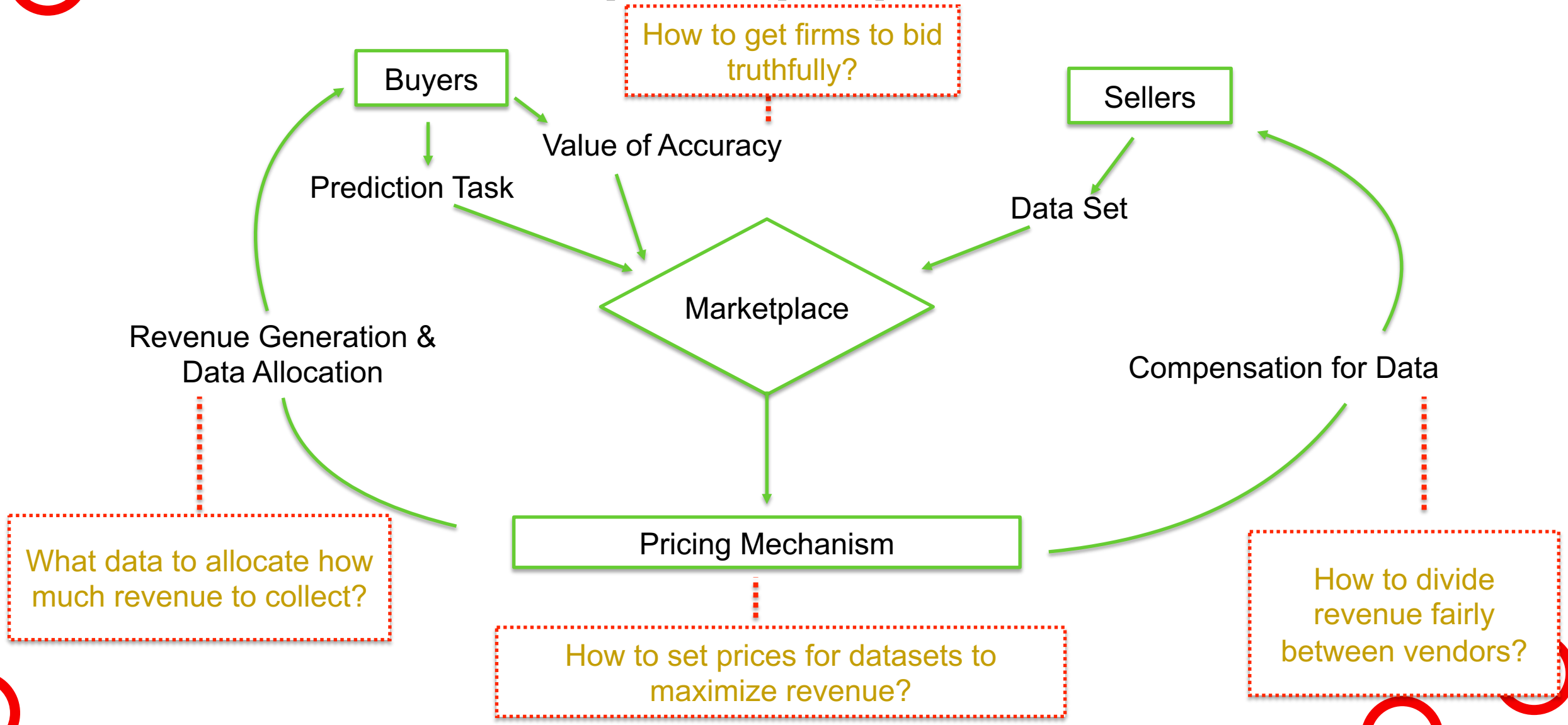
Data Sellers – predictive data for inventory demand

- Uber/Lyft/Hubway: Real-time routing info near shopping districts

- Macy's/Starbucks: Real-time foot-traffic into stores carrying goods supplied by manufacturer

- Grubhub/Yelp: Which type of food items/restaurants are "trending"

Desirable marketplace properties



Solution: Algorithmic Market

- Payment function based on Myerson Payment Rule (Generalized Auction)
- Revenue divided using ‘Shapley Value’
- Prices are decided based on ‘maximum regret algorithm’
- Replication is penalized using information-theoretic criterion
- Computation is real-time


NEW COLLECTION
UP TO 50% OFF
DESIGNER'S CHOICE COLLECTION
SHOP now
BRANDS
PREMIUM Credit