

A decorative graphic consisting of several red lines and circles. The lines are horizontal, vertical, and diagonal, forming a network-like structure. The circles are placed at various points along these lines, some at the ends and some at junctions. The overall style is clean and modern, with a focus on geometric shapes and a single color (red) against a white background.

MUNTHNER A. DAHLEH

Economics and Market Design for Data

Thanks to my Collaborators



Anish Agarwal



Tuhin Sarkar



Devavrat Shah



Mardavij Roozbehani

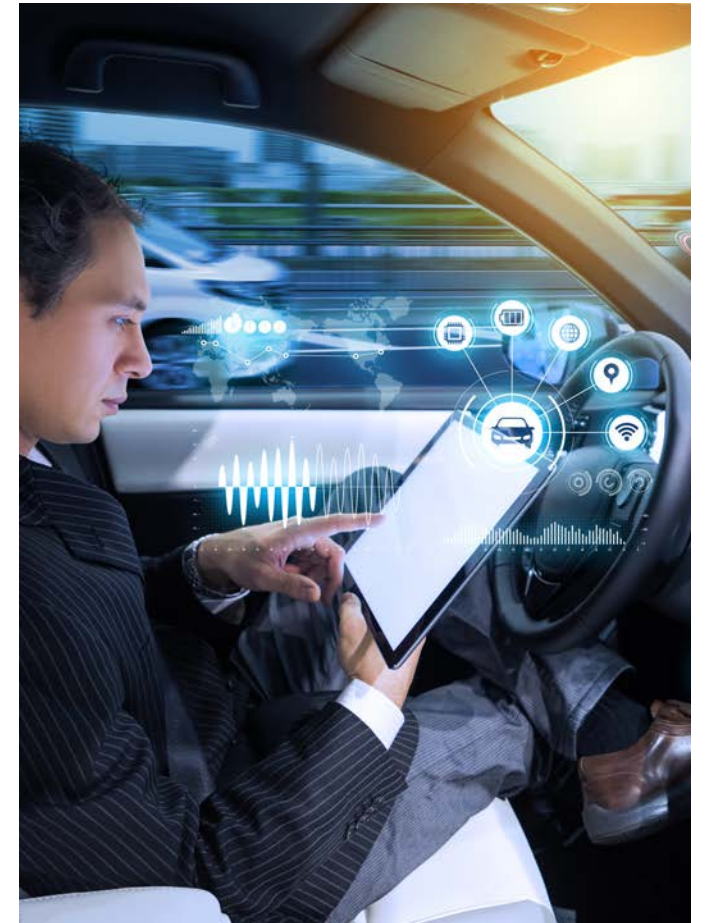


Flora Meng



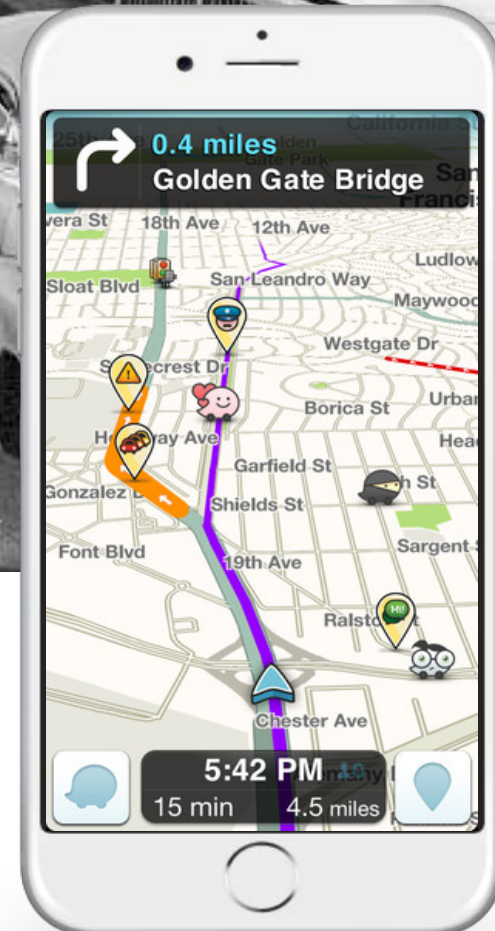
Maryann Rui

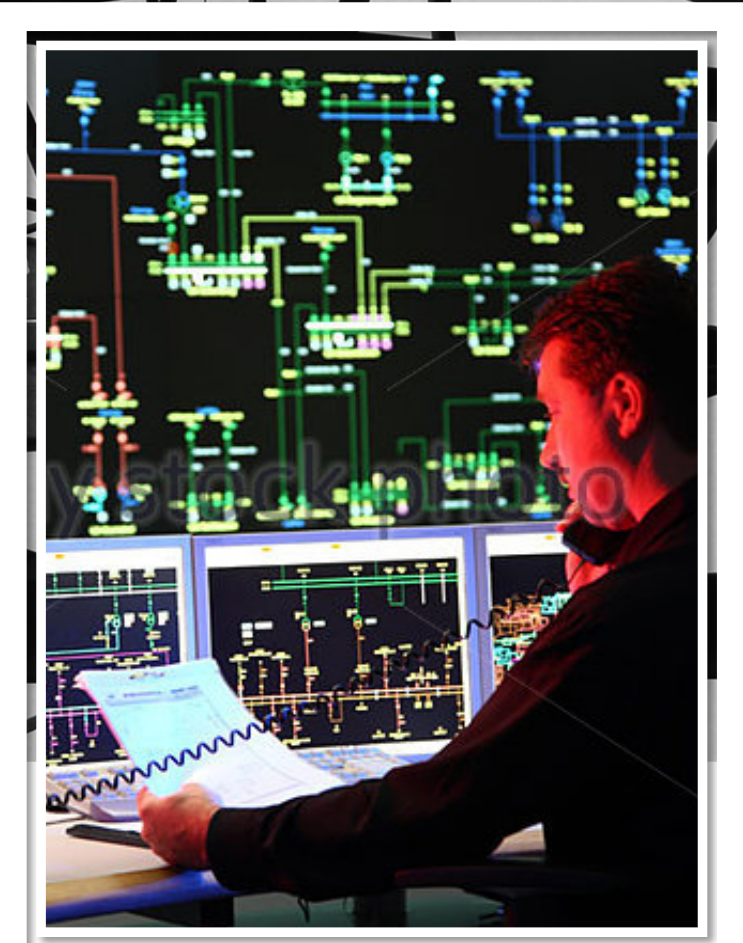
Computing and Data





Congestion Going From Bad To Worse



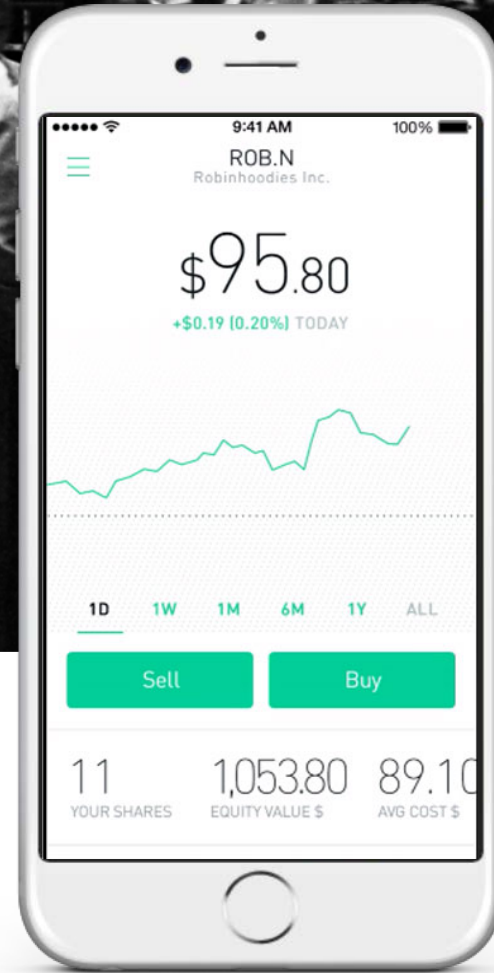


What if the Lights Go Out?

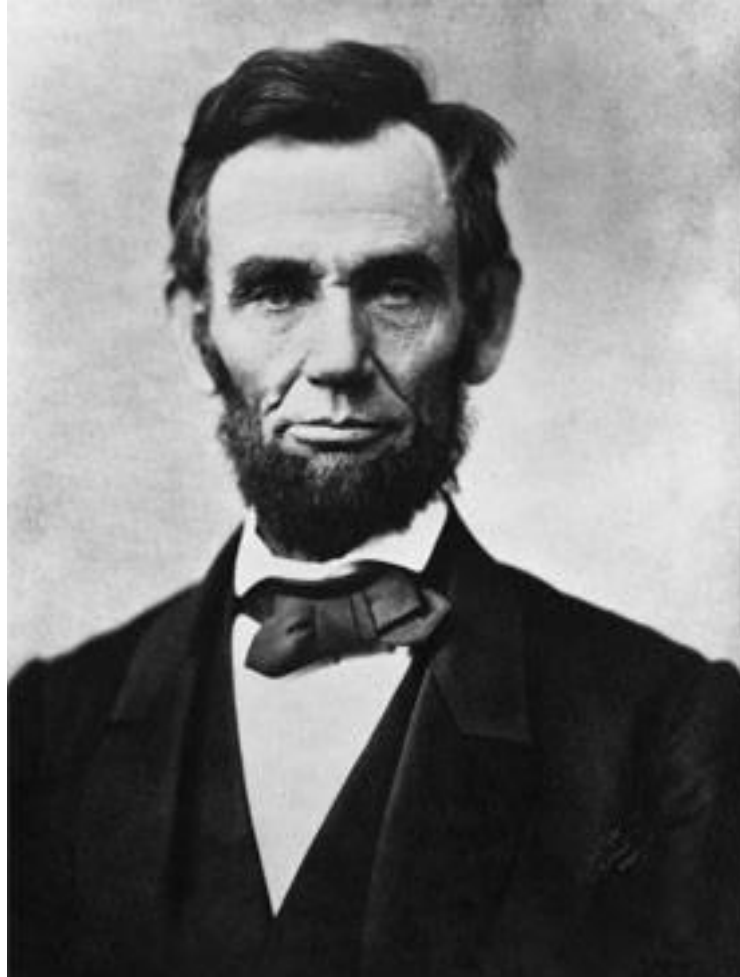




Too Big To Fail



Misinformation



“Don’t believe everything you read on the Internet just because there’s a picture with a quote next to it.”

—Abraham Lincoln



Physical and Engineered Systems



Institutions

Paradigm Shift



**Physical and
Engineered Systems**



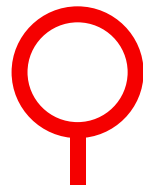
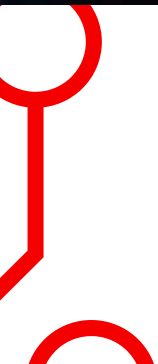
Society



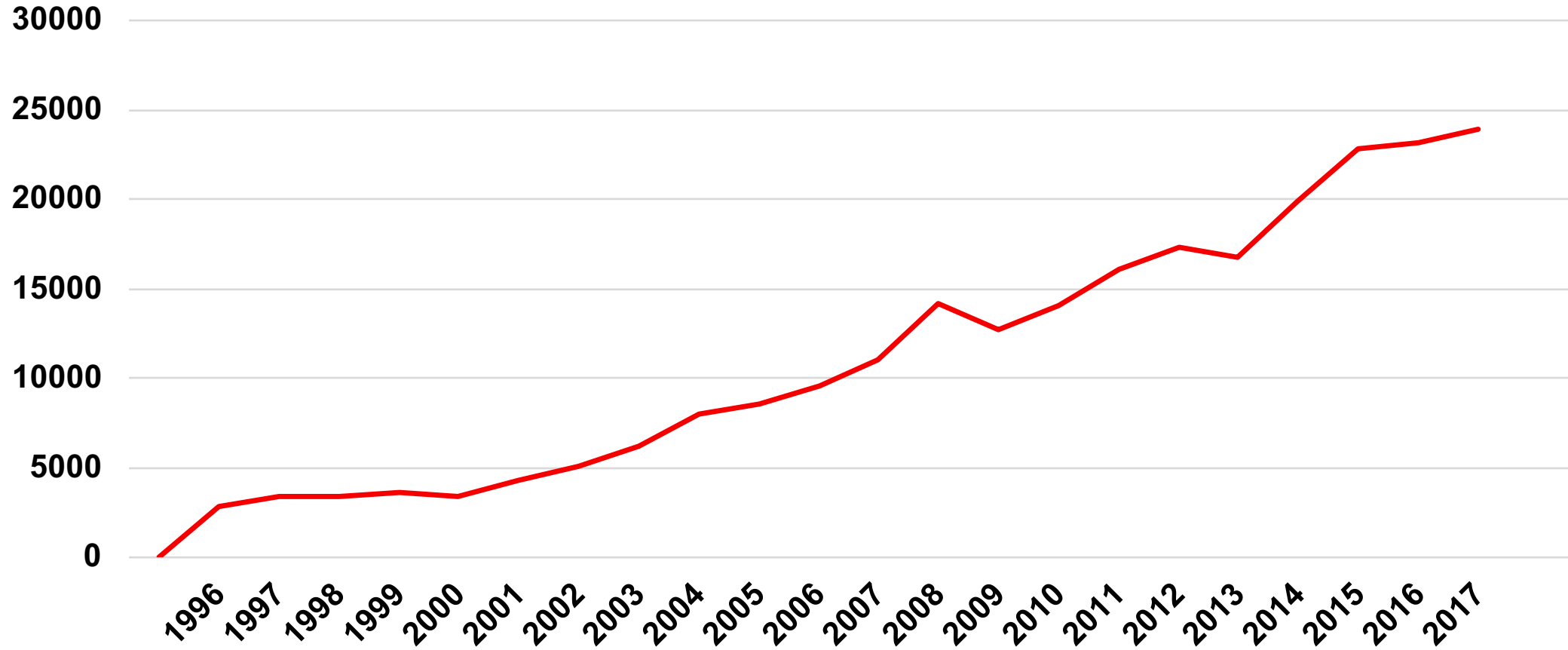
Institutions



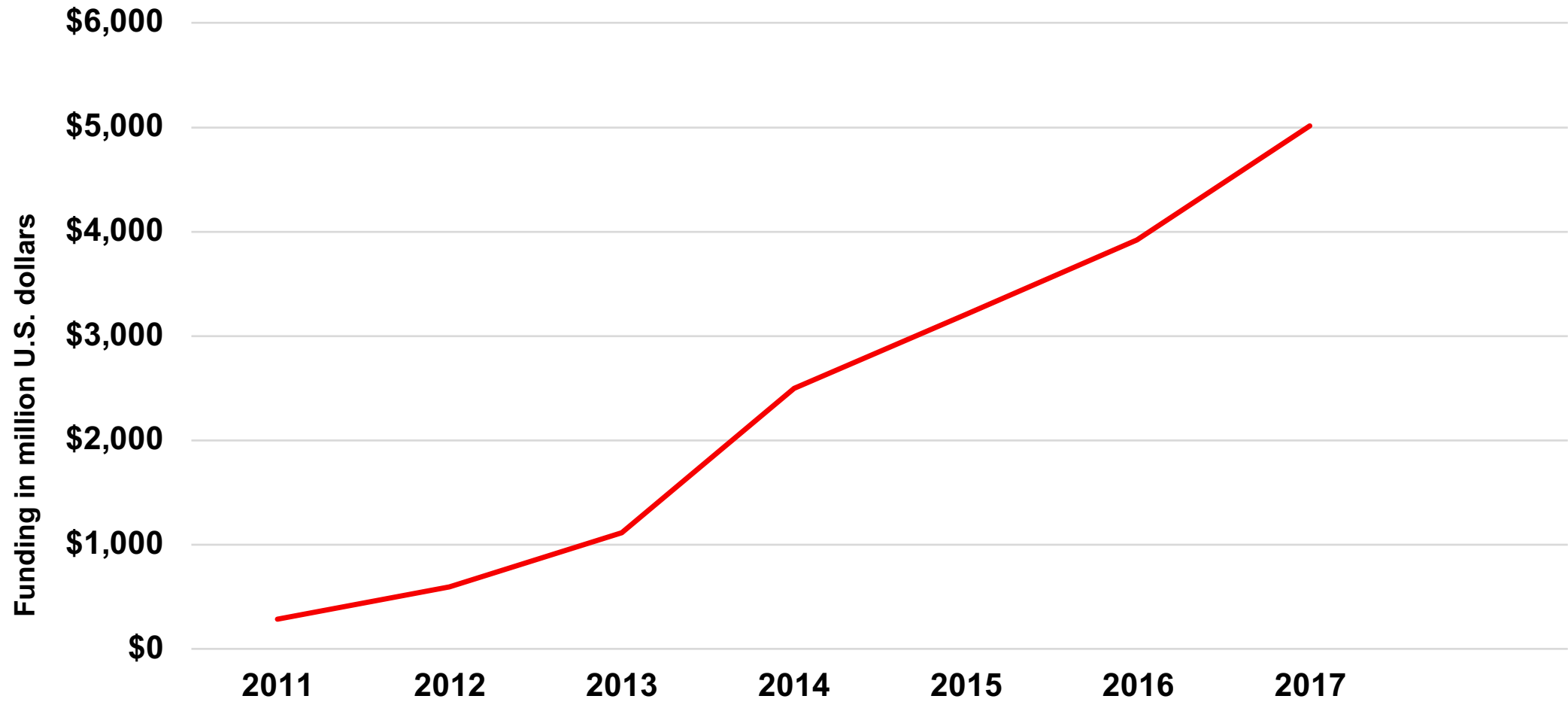
Data as a Commodity



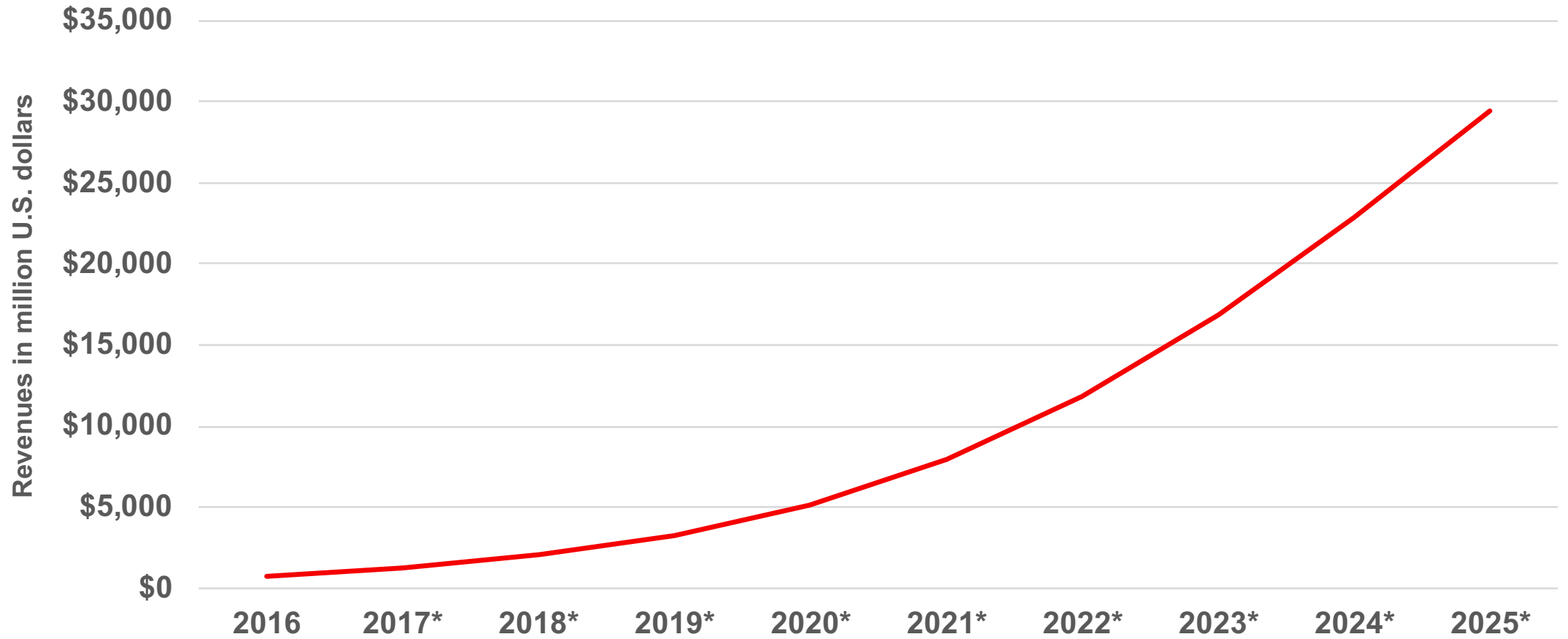
Papers Published on AI (Scopus.com)



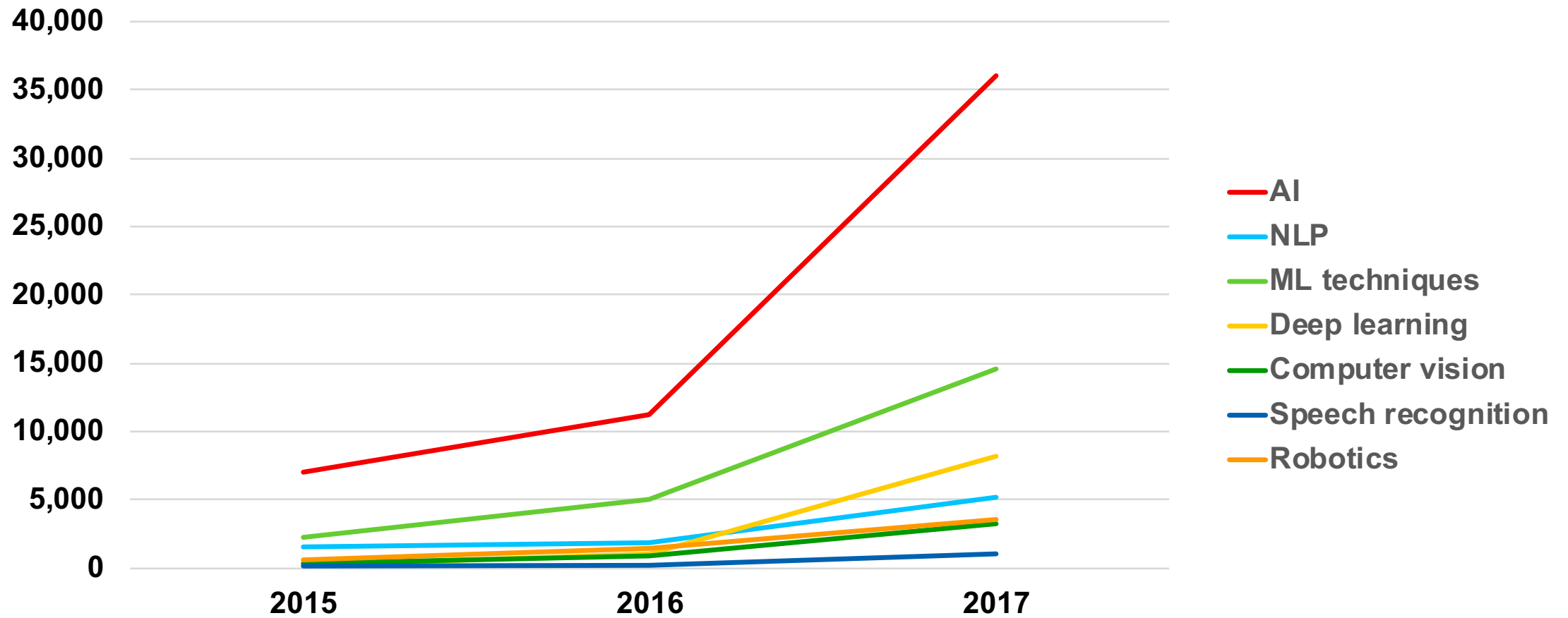
AI Funding in the U.S.



Revenues from AI Market in North America



Job Openings (Monster.com)



A thick red line starts from the top left corner, goes down vertically, then turns right at a 90-degree angle, and finally turns diagonally upwards and to the right, forming a partial frame around the text.

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

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

A thick red curved line starts from the bottom right corner and curves upwards and to the left, ending near the center of the bottom edge.

Bloomberg



Cambridge
Analytica

EQUIFAX

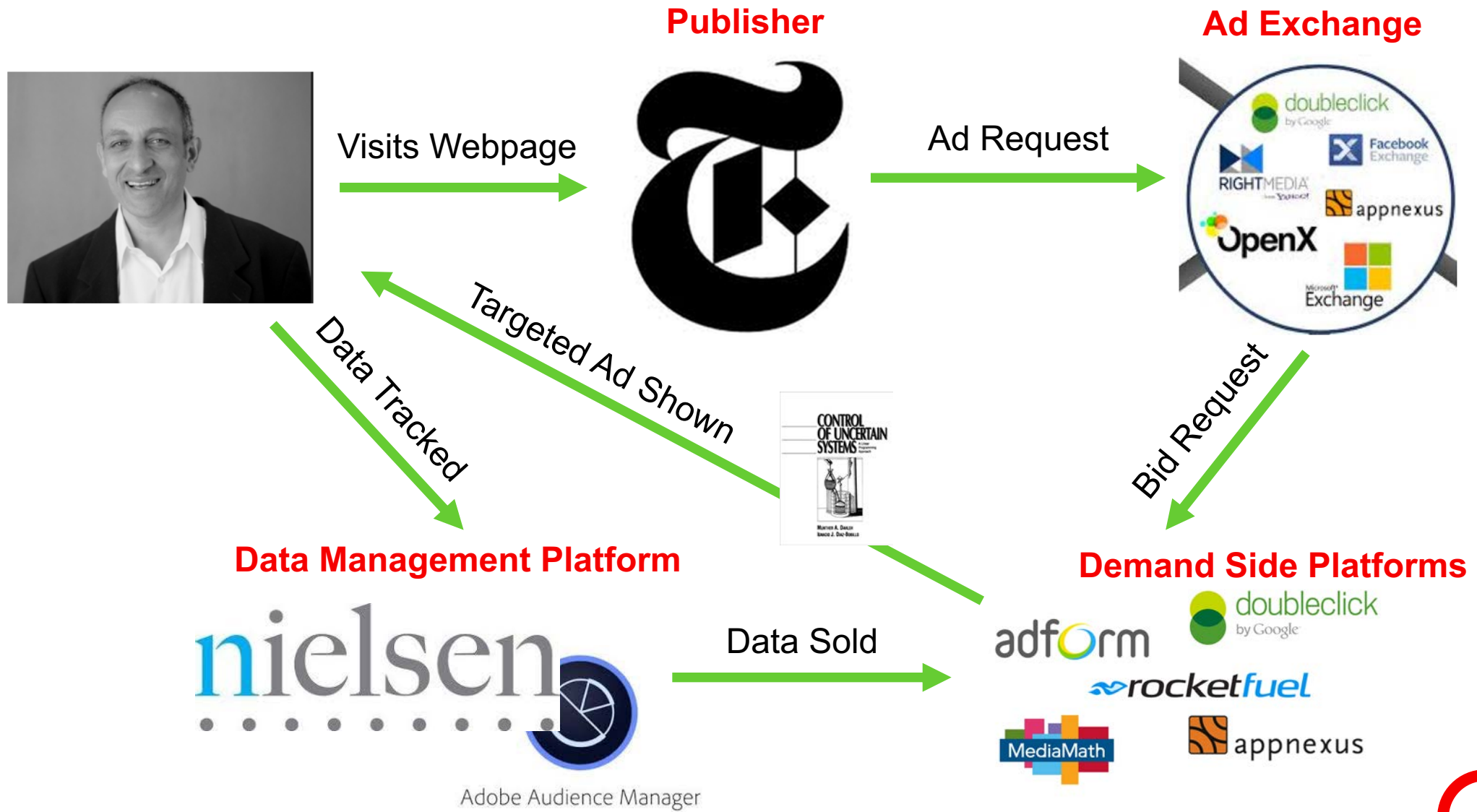


**Why design
a digital
marketplace
now?**



Today customers are not part of the AD market.

Overview of Online Ad Data Market



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



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

Data Management Platform

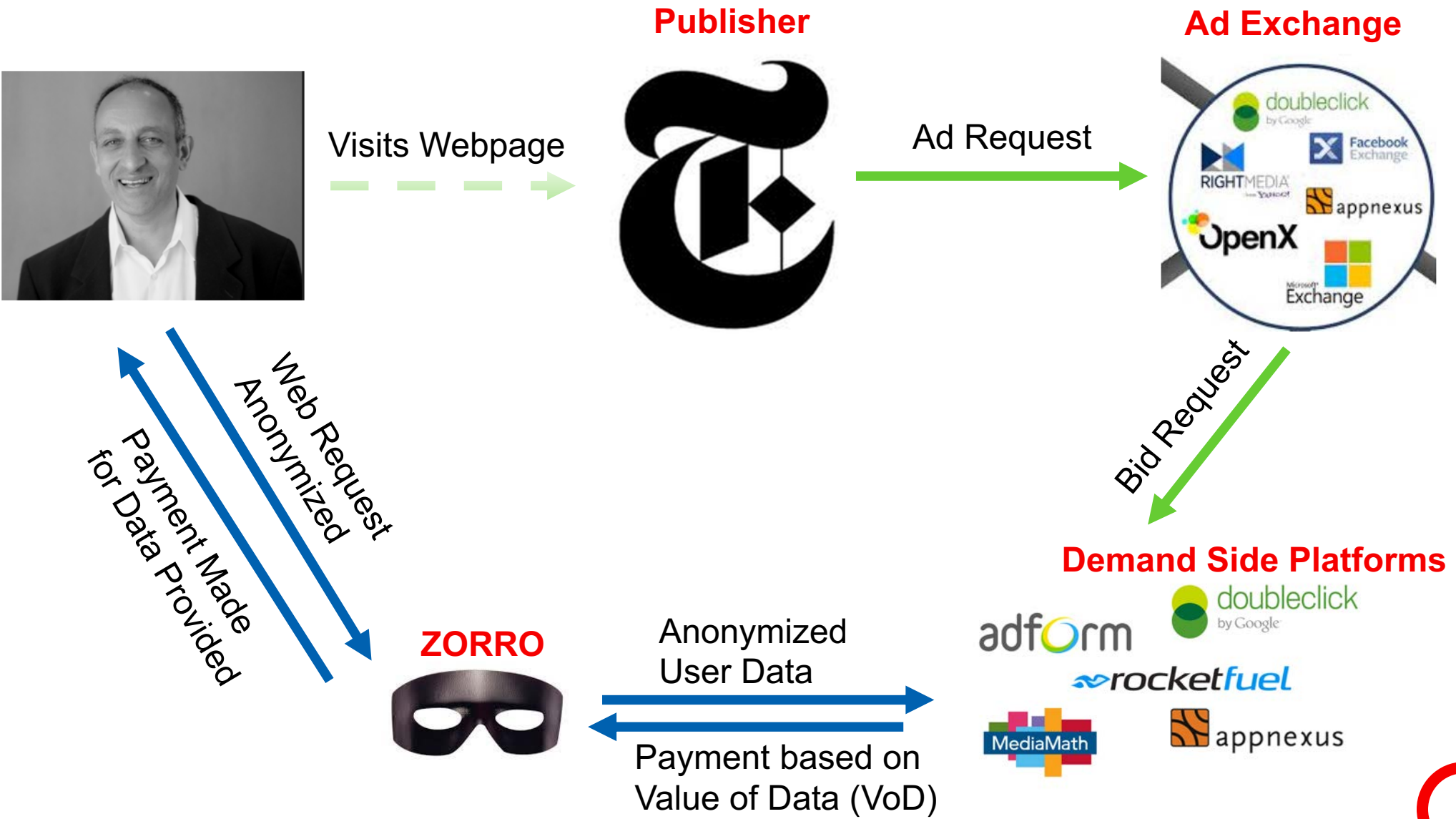


Data Sold

Demand Side Platforms



Zorro: System to Price Online Cookie Data



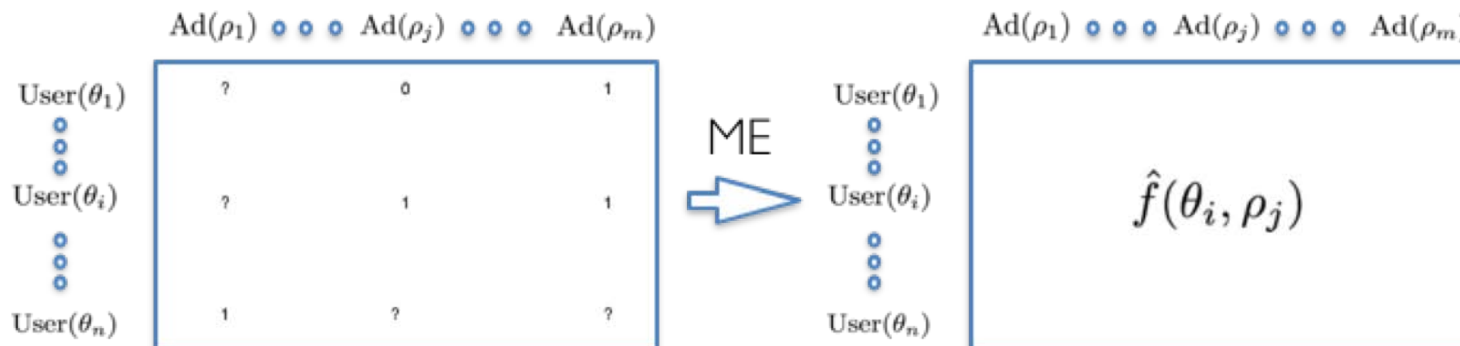
Economic Impact of Zorro is Vast

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



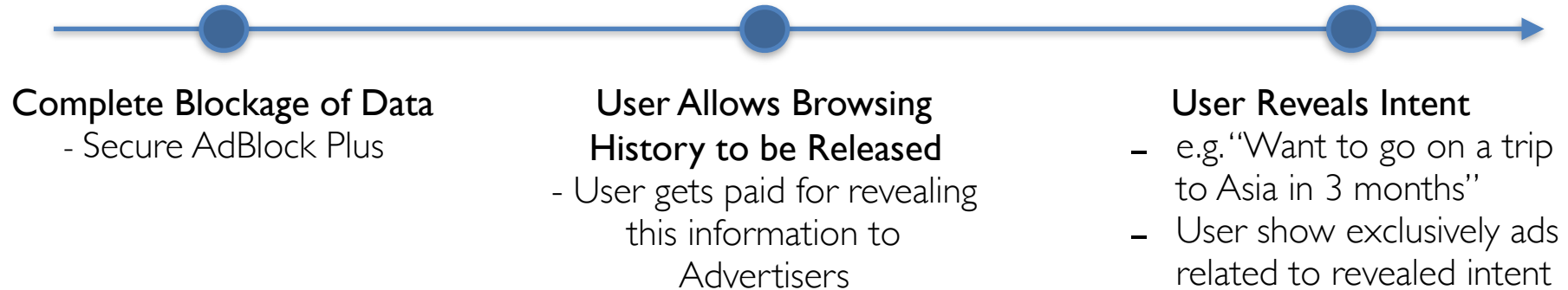
θ_i : User Browsing history and/or Intent Data

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



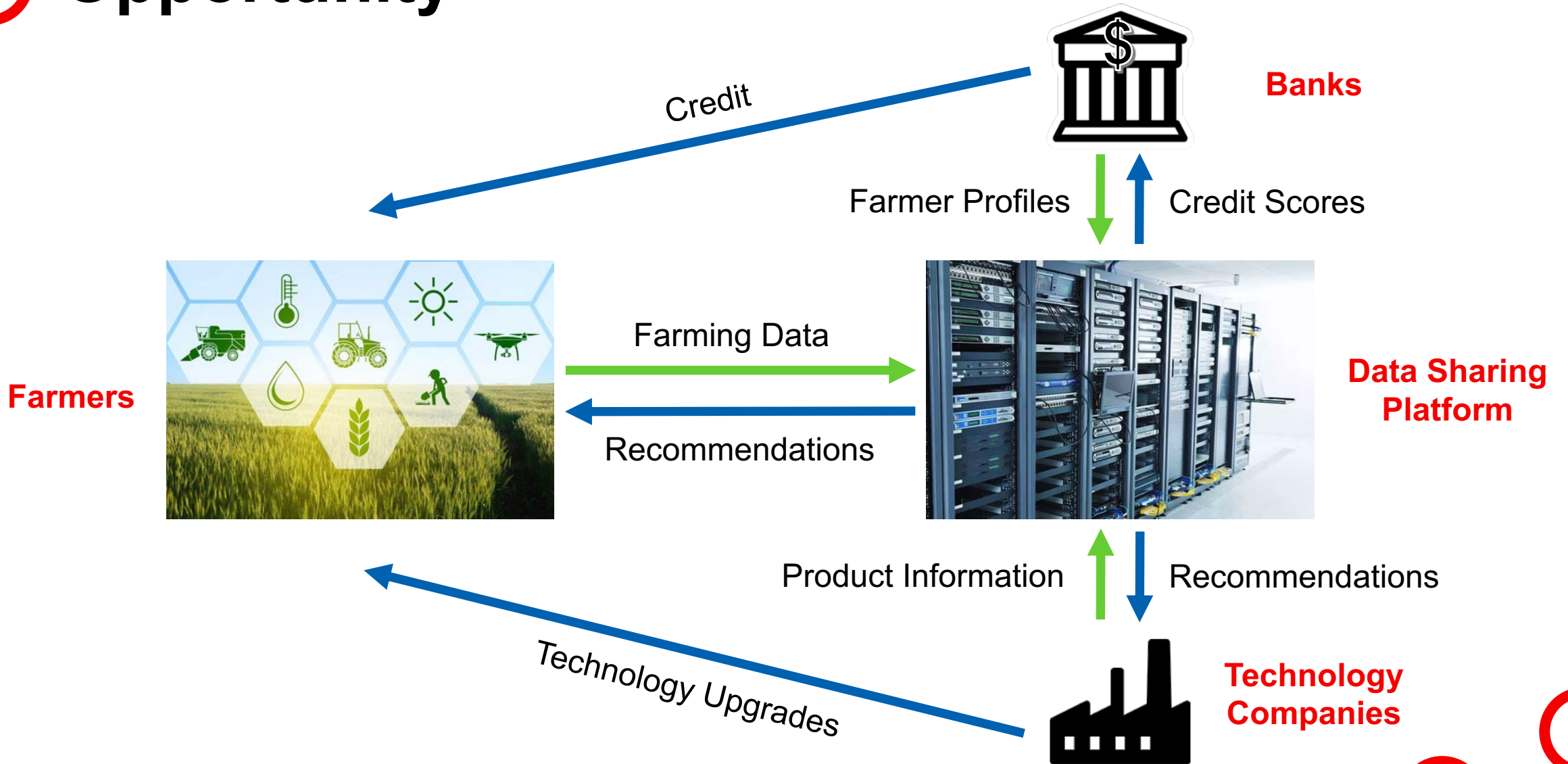
User gets explicit reward for revealing personal information. Can also provide forward-looking intent, large loss of value in current system.

Challenge: Sub-saharan Africa Farming

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



Opportunity

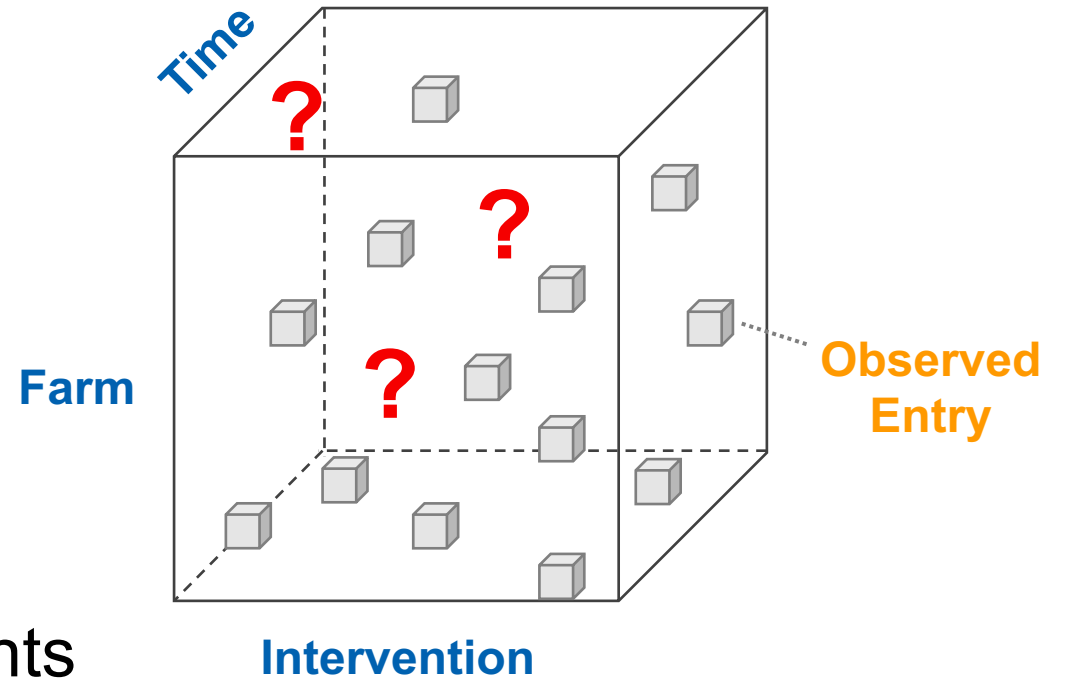


Data Sharing Platform



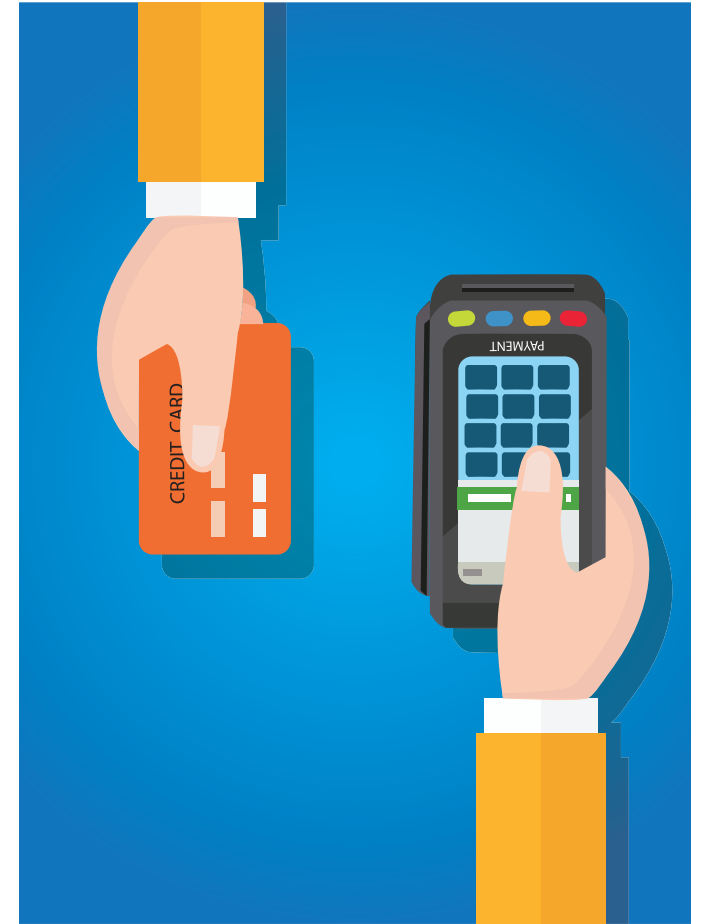
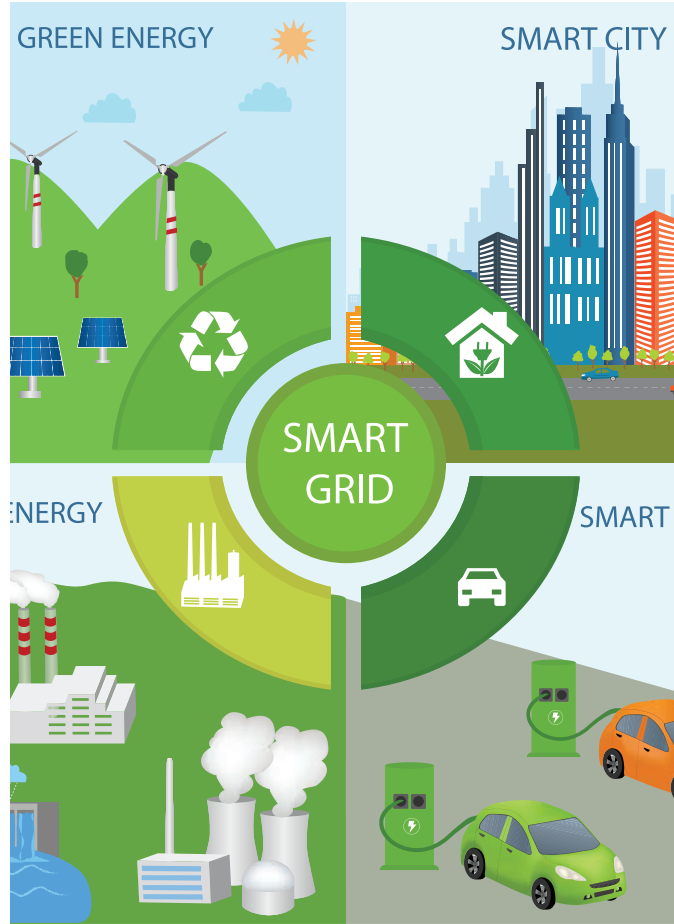
- Incentives to share high-quality data
- Types of data needed for various stakeholders

Algorithmic Challenges



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

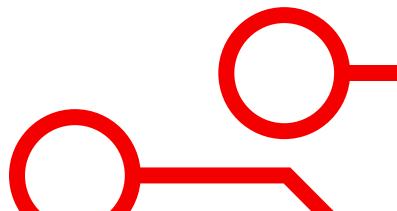
Two-Sided Markets





○ 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%

Stock Market

Index	Var%	Var	High	Low	Open	Close	Volume	Capital	NO	Transaction
2018.39	-4.12	2024.21	2016.93	2022.71	2022.51	9M	8M	4241		
Symbol	Name	Close	Open	Bid QTY	Bid Price	Ask Price	Ask QTY	High	Low	Last Price
SPIC	SPCZ.INVST.COMD	0.77	0.75	1500	0.72	0.74	838649	0.75	0.74	0.74
JNTH	TAJGATERINGHOUSING	2.47	2.47	1000	2.42	2.46	3410	2.47	2.45	2.46
DERA	DEERA	1.11	1.14	97183	1.16	1.19	5100	1.16	1.09	1.16
IDMC	AD-DULAYL PARK	0.56	0.57	1000	0.53	0.54	163940	0.57	0.54	0.54
RJAL	ROYAL JORDANIAN	0.72	0.73	41042	0.69	0.70	5000	0.75	0.69	0.69
SURA	SURA	0.48	0.49	42035	0.48	0.49	16611	0.50	0.47	0.49
REDV	REAL ESTATE DV	0.25	0.25	67800	0.23	0.24	182409	0.25	0.24	0.24
JOPH	JOR PHOSPHATE MN	10.59	10.45	370	10.45	10.75	2245	10.75	10.42	10.75
AMAL	AL-AMAL INV.	0.97	0.97	1045	0.94	0.95	9950	0.97	0.94	0.95
FRST	FIRST JORDAN	0.14	0.15	1181287	0.14	0.15	575360	0.15	0.14	0.14
JOFC	JOR CERAMIC FAC	1.70	1.69	2000	0.89	1.62	15475	1.69	1.62	1.62
UAIC	UNI ARAB INVEST	0.04	0.04	4201142	0.03	0.04	574923	0.04	0.03	0.04
JOJA	RESOURCES INVEST	0.30	0.30	7605	0.29	0.30	1975	0.30	0.30	0.30
JOIR	JORDAN IND.RES.	0.29	0.28	120675	0.27	0.28	37000	0.28	0.28	0.28

Prediction Markets

Markets Portfolio Reporting My Markets Cash: 0.958.74 Rep: 47.00 Ebit: 4.38 Sign Out

Portfolio

Will SpaceX successfully complete a manned flight to the International Space Station by the end of 2018? [View Market](#)

Yes/No Market (2 outcomes) End Date: Jan 1, 2019 Trading Fee: 2.0% Volume: 17.00 shares

TOP PREDICTIONS	YOUR TRADING
Yes: 55%	Positions: 1
No: 45%	Profit / Loss: -0.90
	Trades: 1
	Unrealized P/L: -0.02

Which political party's candidate will win the 2016 U.S. Presidential Election? [View Market](#)

Multiple-Choice Market (4 outcomes) End Date: Jan 3, 2017 Trading Fee: 2.0% Volume: 118.00 shares

TOP PREDICTIONS	YOUR TRADING
Democratic: 33%	Positions: 1
Republican: 23%	Profit / Loss: -0.35
Libertarian: 22%	Trades: 1
Other: 22%	Unrealized P/L: -0.01

Cannot replicate ad-space/stock

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

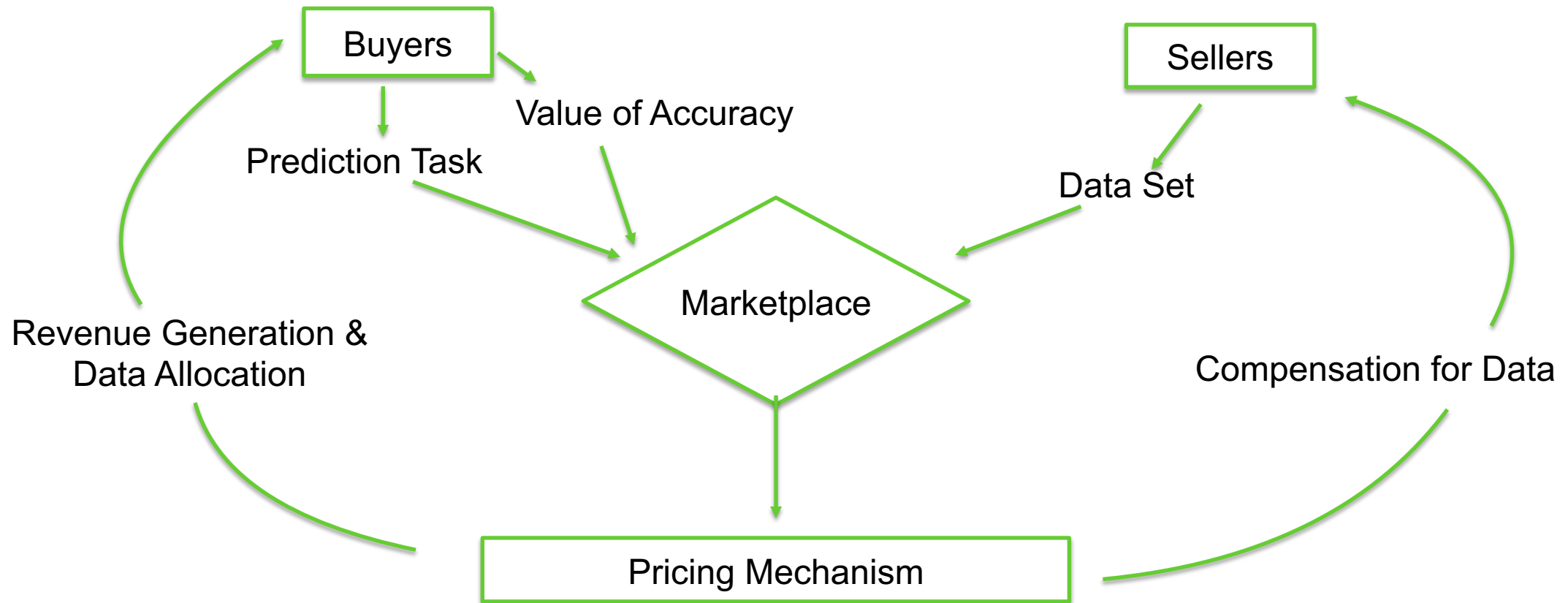
Data has no analogy to historical click-through rate/historical stock performance

- 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

Prediction tasks and experts can be matched easily

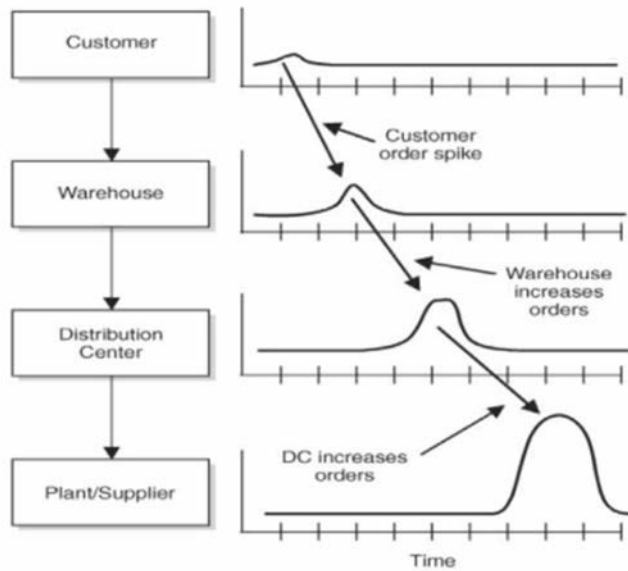
- Prediction Markets - no need to actively match experts with prediction tasks
- Data Markets - difficult to know which data is useful for which prediction task a priori

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”

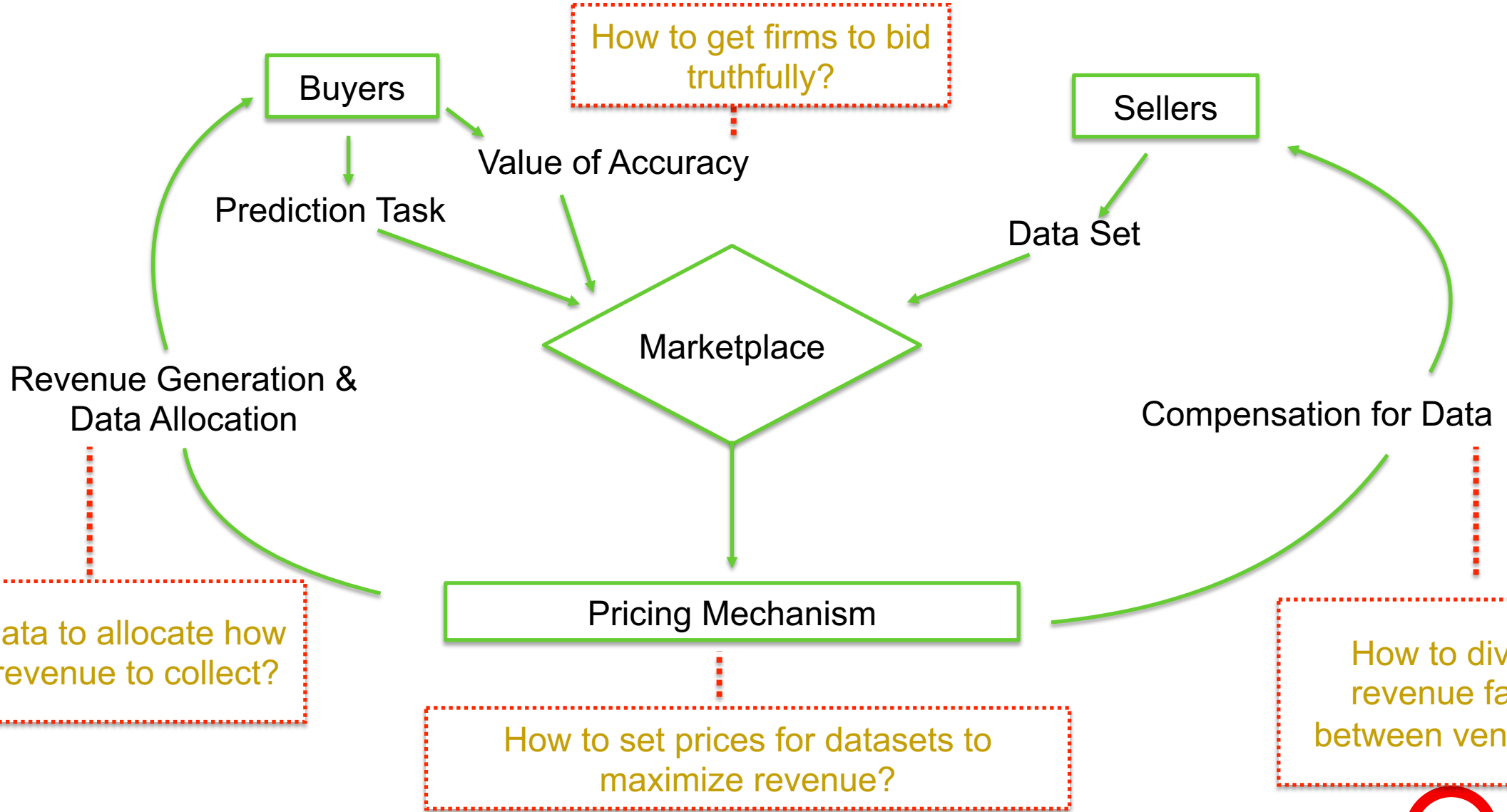
What do Data Purchasers (Buyers) need to decide?

- 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



How to get firms to bid truthfully?

Buyers

Sellers

Marketplace

Pricing Mechanism

Prediction Task

Value of Accuracy

Data Set

Revenue Generation & Data Allocation

Compensation for Data

What data to allocate how much revenue to collect?

How to set prices for datasets to maximize revenue?

How to divide revenue fairly between vendors?



○ 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

