Digitization and Artificial Intelligence: Challenges and Opportunities for Policy

SUSAN ATHEY

THE ECONOMICS OF TECHNOLOGY PROFESSOR, STANFORD GSB

FACULTY DIRECTOR, GOLUB CAPITAL SOCIAL IMPACT LAB; ASSOCIATE DIRECTOR, STANFORD INSTITUTE FOR HUMAN CENTERED AI

HTTPS://ATHEY.PEOPLE.STANFORD.EDU/

Themes

Important advances in AI

- General purpose prediction algorithms
- Generative adversarial networks
- Iterative experimentation and bandits
- Reinforcement learning
- Automation vs. autonomy

Potential for beneficial impacts

- Efficiency for providing and receiving services
- Education and training
- Safety and monitoring
- Productivity vs. labor demand

Additional policy considerations

- Government efficiency and effectiveness, for good or bad
- Guiding and regulating AI
- Incentivizing beneficial AI

Economics of Beneficial Artificial Intelligence

AI development is endogenous

- Technical capabilities and limitations
- R&D into replacing vs. augmenting humans
- R&D that anticipates and avoids unintended consequences

AI creates opportunities to benefit humanity

- Social sector
- Government
- Traditional private sector

Private and social incentives may not be aligned

- Cost savings vs. labor externalities
- Market power vs. consumer benefits
- Scale economies and low marginal cost services

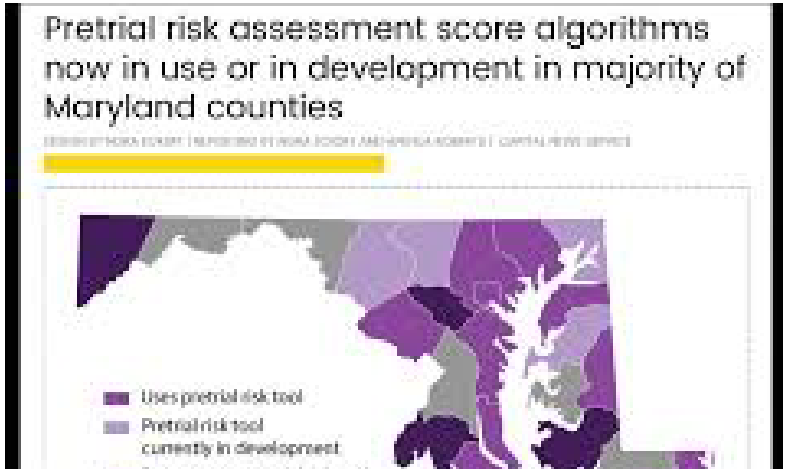
Replacing vs. Augmenting Humans

Replacing

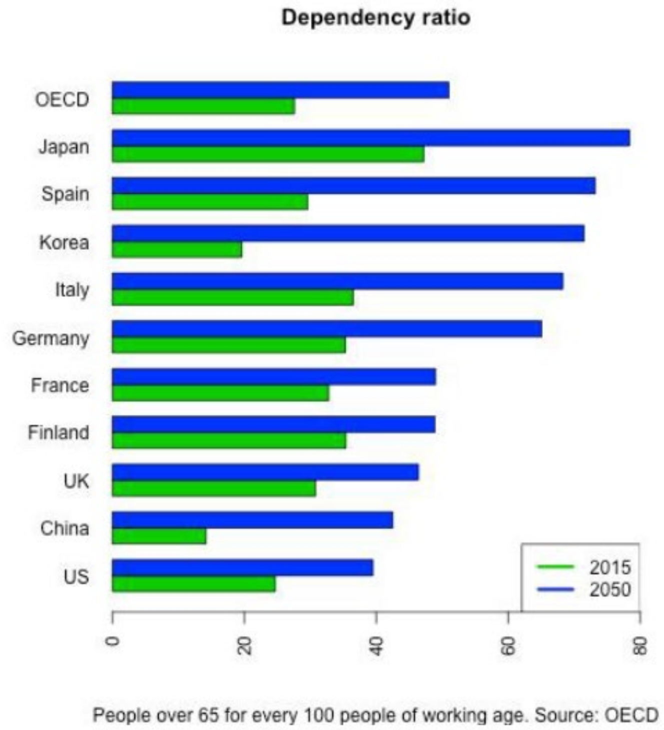
- Customers get their needs met through mobile app
- Automation on assembly line
- Checkout/ordering

Augmenting

- Risk scores/prioritization
- Automating components of tasks
- Automated performance improvement
- Safety, compliance, memory aids



Bots vs. Tots



Source: OECD.

Figure 1 Demand and supply of labour

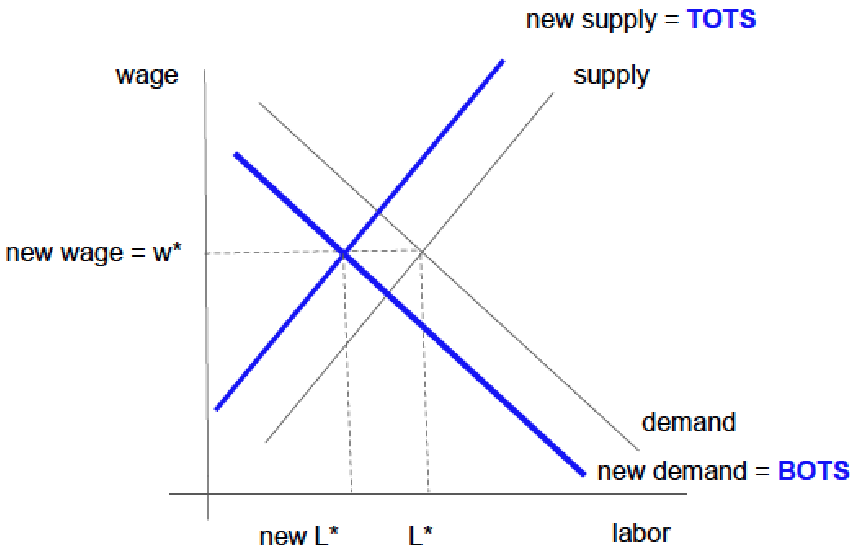
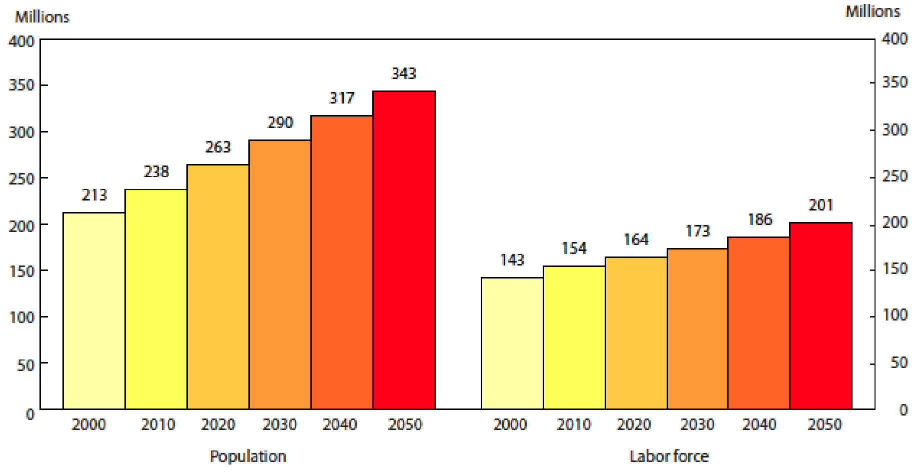


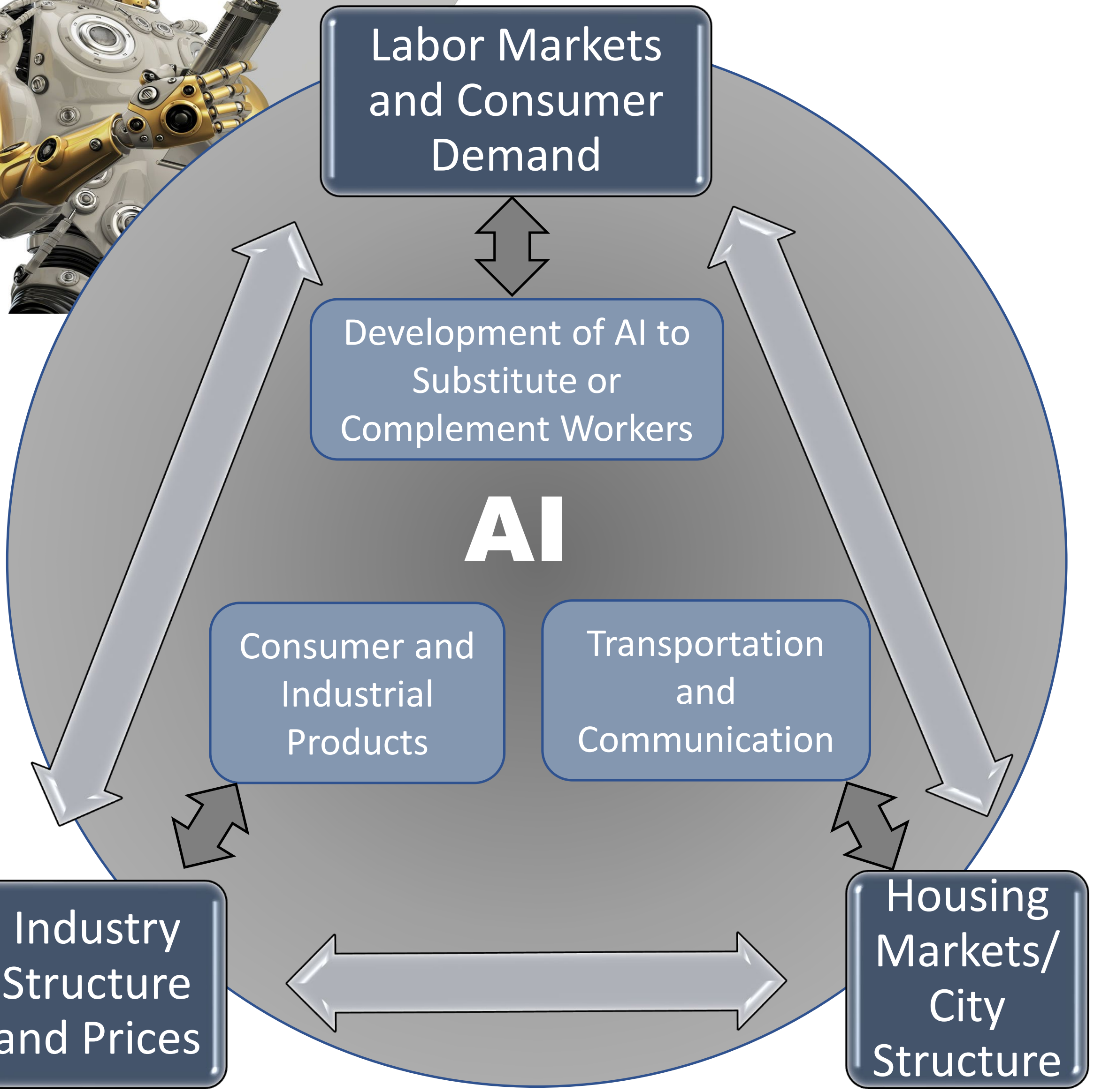
Figure 9 Population and labour force growth projected to 2050

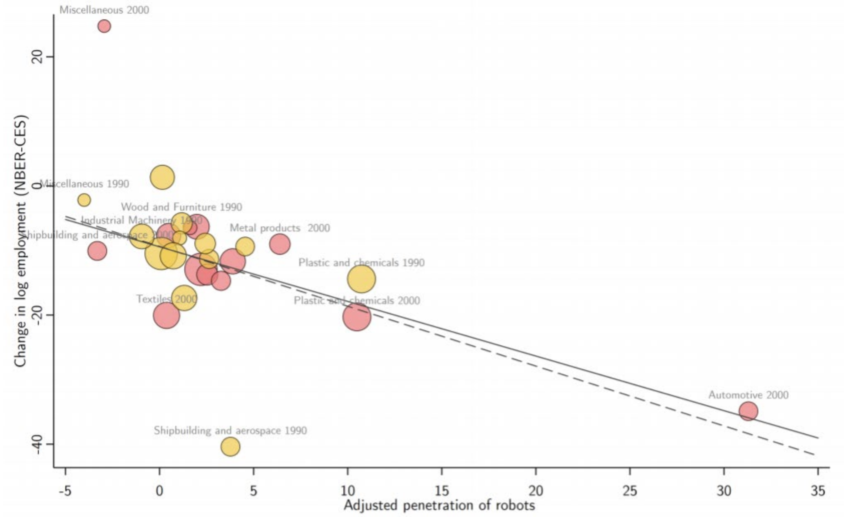


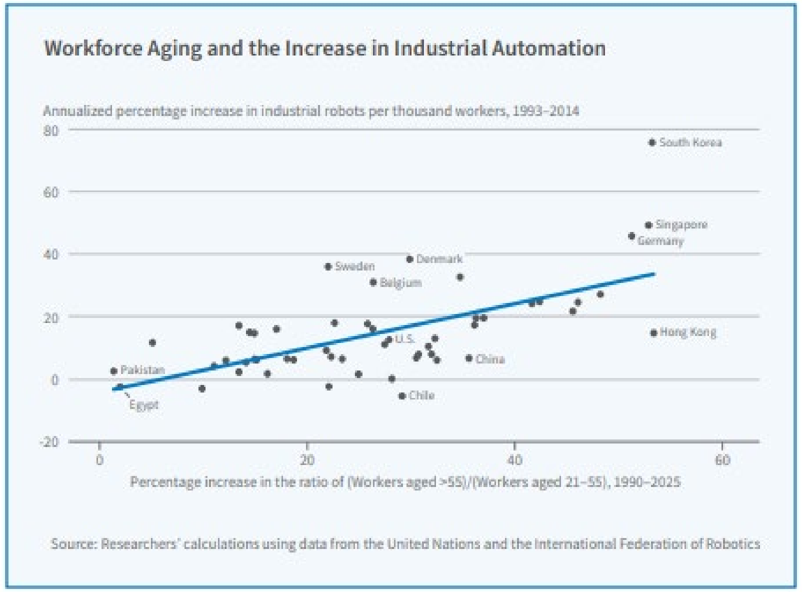
Source: Bureau of Labor Statistics.

Source: Hal Varian
https://voxeu.org/article/automation-versus-procreation-aka-bots-versus-tots

AI and the Economy







Acemoglu and Restrepo (2018, 2020)

Using ML/AI for Good: Meeting the Consumer where they Are

Access & Targeting Through Digital Provision of Services

ACCESS AND THE VALUE OF TIME

Access to specialists traditionally limited by location and local scale economies
- Medical
- Educational
- Repair

Transportation costs & time off from work

Taking advantage of underutilized, most convenient time:
- While commuting
- While children sleep

Crosses political boundaries: rural, urban

PERSONALIZATION

Get relevant services

Automate access to government services

Targeting benefits in emergencies

Education & training at right level & pedagogy

Open issues: credentialing, mix of in-person, live-remote, asynchronous, etc.

Guiding AI

Crucial to place large emphasis on guiding AI

Universities, government, philanthropy play important roles

- Fairness, bias, access
- Invest and innovate towards societal benefit
- Robustness, stability, unintended consequences
- Market power
- Cost/ Benefit Approach to Regulation
- Data privacy and security

Application: Monitoring and Incentives

Marketplaces need to provide incentives and screen for quality

Ratings are noisy, often missing and biased, uncomfortable and time consuming for customers

Alternative: direct monitoring and feedback to sellers

Approaches

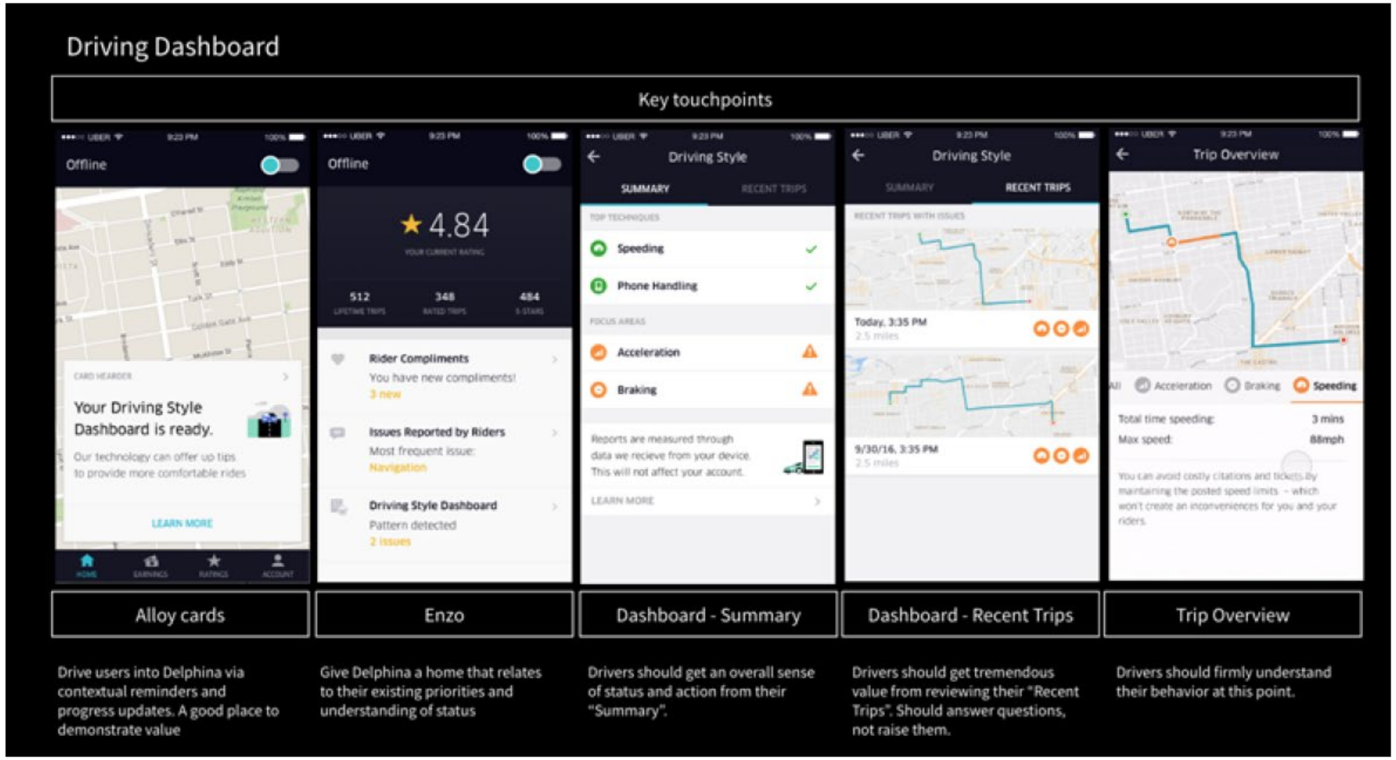
Gather data passively

Gather customer satisfaction data from a sample, or passively from customer behavior

Train a model to estimate quality of service

Provide feedback and coaching to seller, require training, explicit incentives

Nudging Drivers to Better Performance

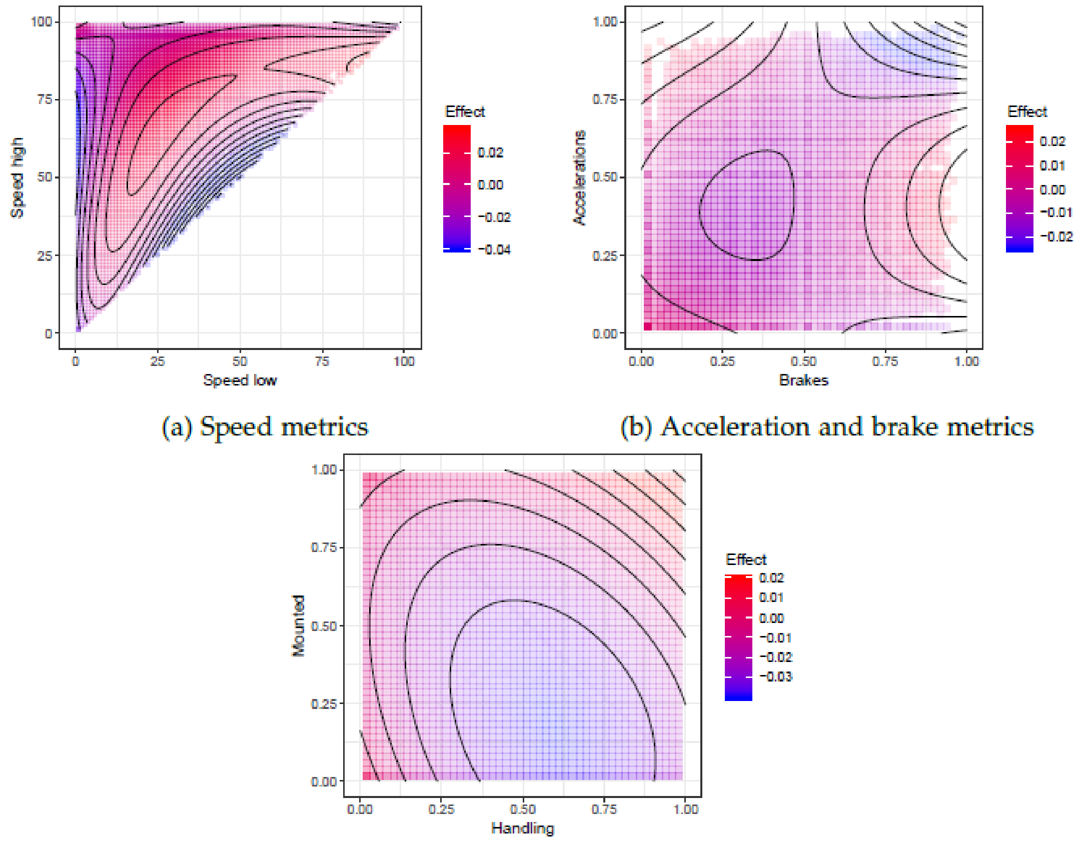


Experiment:

- Randomly select drivers have access to app
- Small effect improving driver safety on average
- Much larger effect for drivers whose performance was poor prior to experiment

Monitoring Workers or Service Providers for Quality: UberX drivers provide higher quality than taxi's

Predicted Star Ratings as a Function of Telematics



(a) Speed metrics

(b) Acceleration and brake metrics

(c) Cell phone use metrics

Experimental Estimates of Informational Nudges

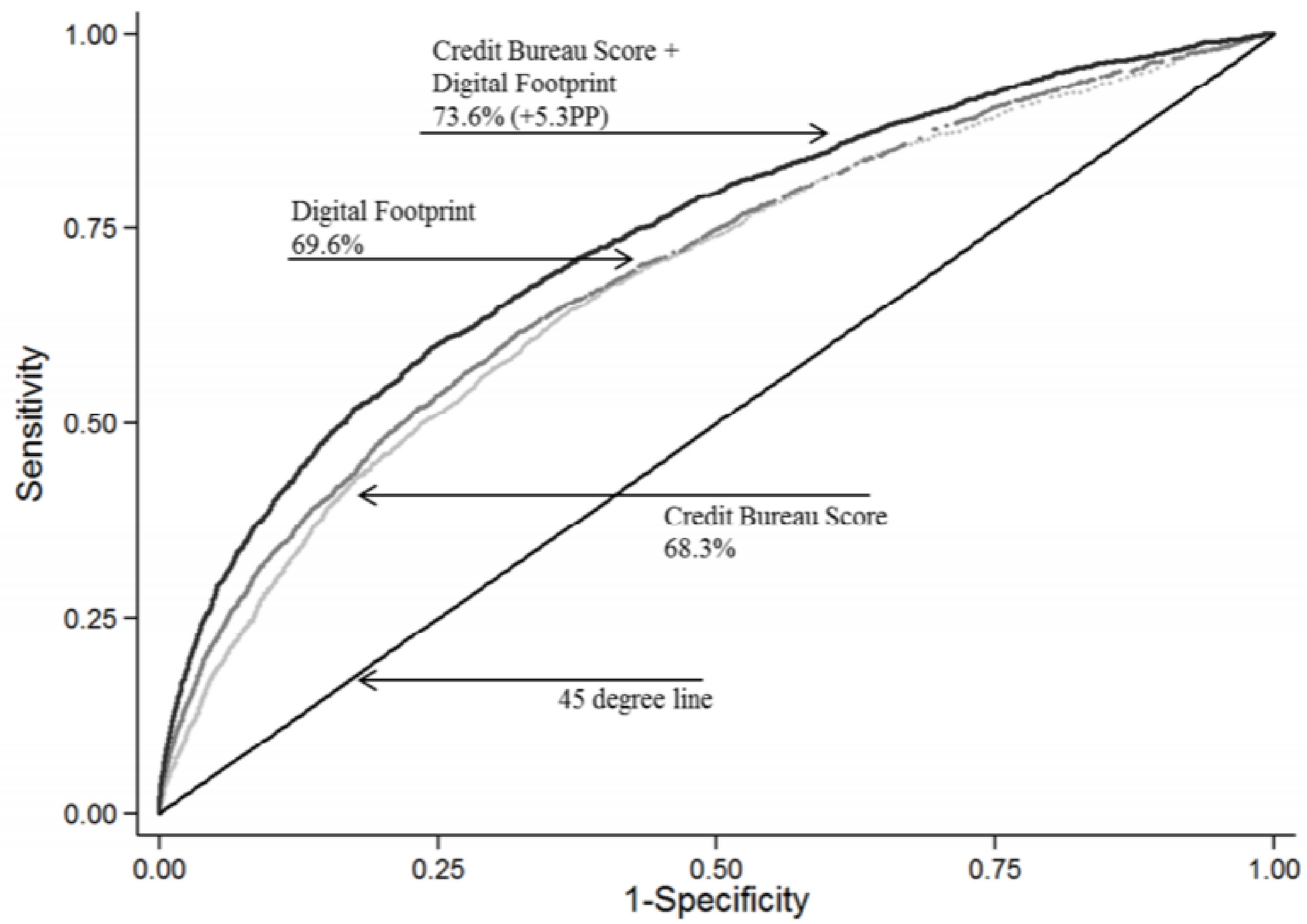
	Dependent variable:		
	Score F	Score S	Score NS
	(1)	(2)	(3)
	Panel A: Intent to treat estimator		
Bottom 10th Perc. Before	−0.0271***	−0.0082***	−0.0255***
	(0.0005)	(0.0004)	(0.0004)
Treatment x Not Bottom 10th Perc.	0.0001	0.0001	0.00004
	(0.0002)	(0.0001)	(0.0001)
Treat x Bottom 10th Perc.	0.0015**	0.0006	0.0014**
	(0.0006)	(0.0005)	(0.0005)
Observations	4,254,109	4,254,109	4,254,109
	Panel B: 2SLS estimator		
Bottom 10th Perc. Before	−0.0008*	−0.0005**	0.0002
	(0.0005)	(0.0002)	(0.0004)
App Int. x Not Bottom 10th Perc.	0.0003	0.0001	0.0002
	(0.0002)	(0.0001)	(0.0002)
App Int. x Bottom 10th Perc.	0.0028***	0.0007	0.0027***
	(0.0010)	(0.0005)	(0.0009)
Observations	4,254,109	4,254,109	4,254,109

Using digital footprints for credit scoring

"On the Rise of FinTechs – Credit Scoring Using Digital Footprints," Berg, Burg, Gombovic, Puri, forthcoming

Figure 3: AUC (Area Under Curve) for scorable customers for various model specifications

This figure illustrates the discriminatory power of three different model specifications by providing the rec operating characteristics curve (ROC-curve) and the area under curve (AUC). The ROC-curves are estimated usi logit regression of the default dummy on the credit bureau score (light gray), the digital footprint (gray), both c bureau score and digital footprint (dark gray). The sample only includes customers with credit bureau scores. sample period is from October 19, 2015 to December 2016. For variable definitions see Appendix Table 1.



Using digital footprints for credit scoring

- Manipulability
- Stability

DRIVERS' E-FAIL Admiral hikes insurance costs for drivers using Hotmail email addresses

It follows our story yesterday on how insurers charge drivers called Mohammed more

EXCLUSIVE

Katie Hodge | Ben Leo
23 Jan 2018, 0:01 | Updated: 23 Jan 2018, 20:25

2 COMMENTS

CAR insurer Admiral last night admitted hiking premiums for drivers applying via Hotmail.

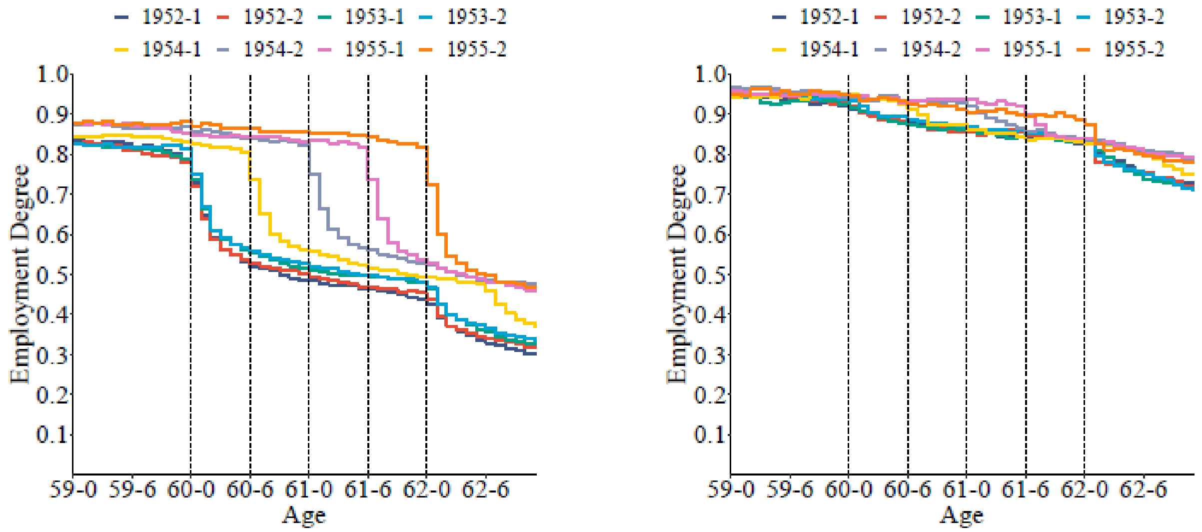
Using ML/AI for Good: Identify the vulnerable, target interventions

Danish Retirement Reform

Retirement reform pushed back age of eligibility for early retirement benefits by six months at a time over several years

When early retirement is available, large chunk of people take it shortly after eligibility

Estimating treatment effect of early retirement benefits on working boils down to predicting who takes the benefit, since without the benefit, very slow decline in employment with age



(a) Lower secondary education (b) Master's degree or equivalent

Figure 2: Employment Trends by Cohort and Education

Notes: This figure shows the trends in employment degree by cohort and education across age. Each cohort is marked by a different color. The figure plots individuals with a lower secondary education as the highest completed education (LHS) as well as individuals with a master's

Between work, public programs, and retirement: heterogeneous responses to a retirement reform*

Susan Athey* Rina Friedberg†

Nicolaj Mühlbach‡ Henrike Steimer§ Stefan Wager¶

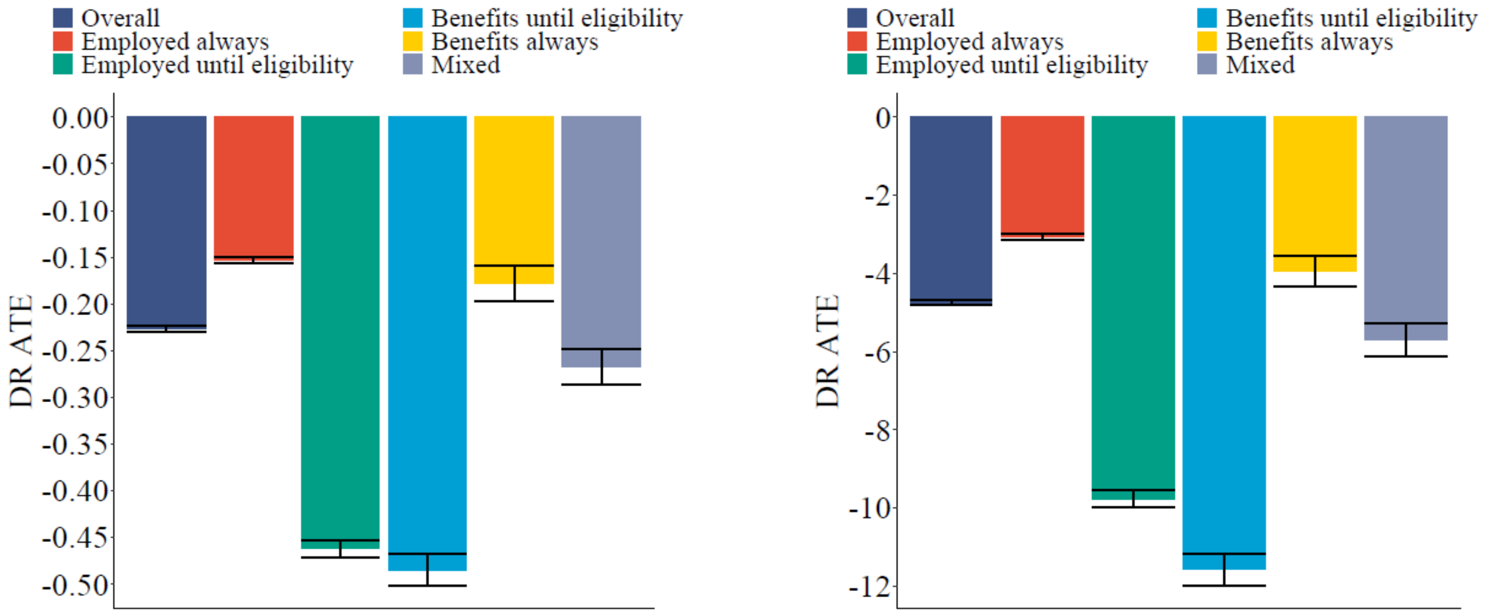
Danish Retirement Reform

Can predict who would have used benefits, retired early—which workers work less when benefits are available

Specifically, we use ML to characterize workers by their predicted "paths" based on characteristics

Bars correspond to different buckets of individuals, classified by their predicted path based on their histories

Our path prediction model successfully predicts who will stop working



(a) Binary IEM

(b) Continuous IEM

Figure 4: DR ATE on Employment

Notes: This figure shows the estimated DR ATE with a 95% CI on employment overall and by predicted retirement path for the main cohorts in the binary IEM (LHS) and the continuous IEM (RHS). For the binary model, the treatment effect is the probability of being mainly employed six months after becoming eligible to retire early, whereas for the continuous model, the treatment effect is the change in number of weeks employed six months after eligibility.

Danish Retirement Reform

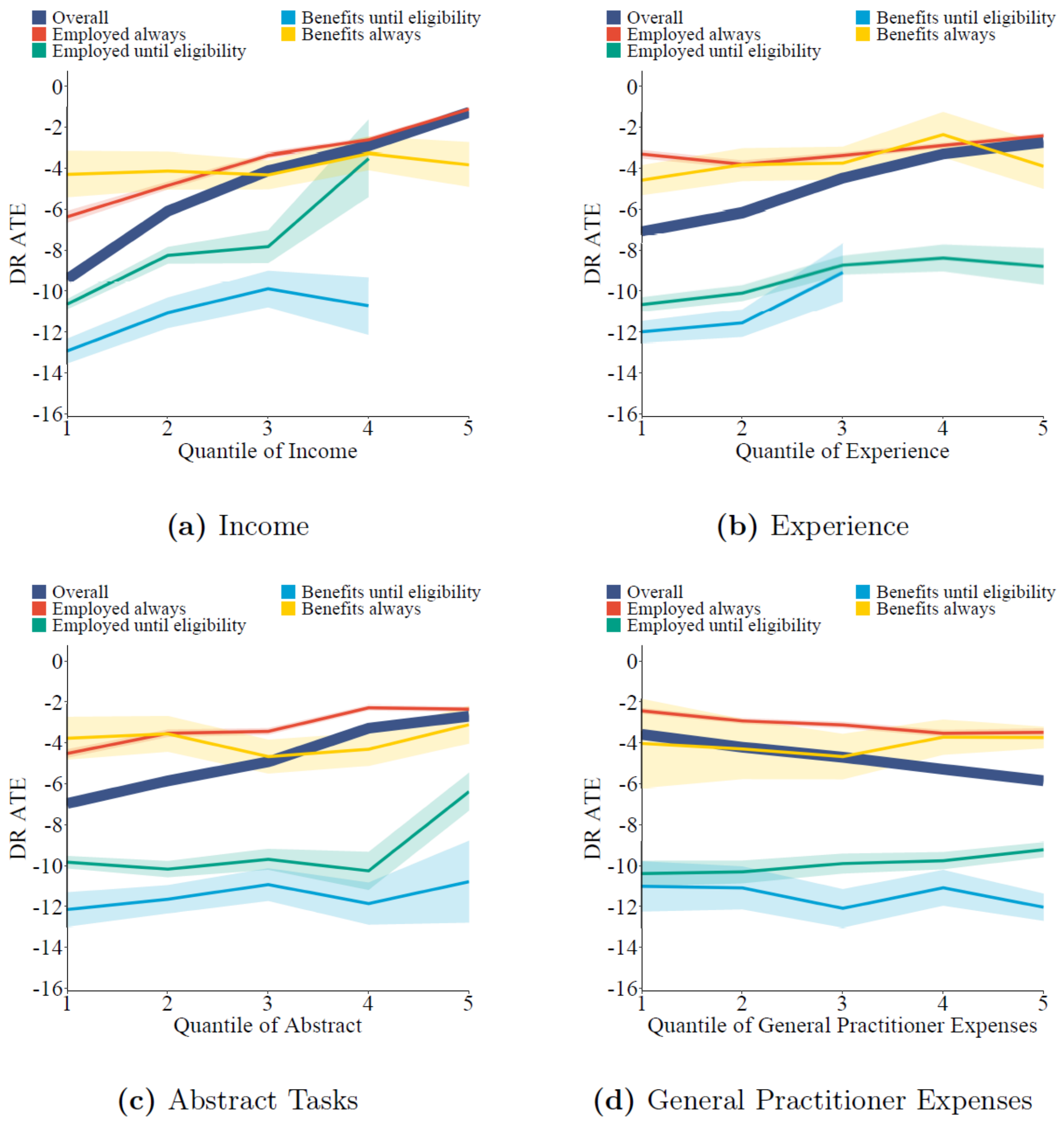


Figure 5: DR ATE on Employment by Continuous Covariates

Workers with low income, less experience, and less abstract jobs take benefits if available

ML identifies more complex groups: lots of heterogeneity within low-income group

Notes: This figure shows the DR ATE with a 95% CI overall and by predicted retirement path across quintiles of four selected covariates (*income*, *experience*, *abstract tasks*, and *general practitioner expenses*) as estimated by the continuous IEM using a six-month horizon. The sample is the training sample of the main cohorts and all predictions are out-of-bag.

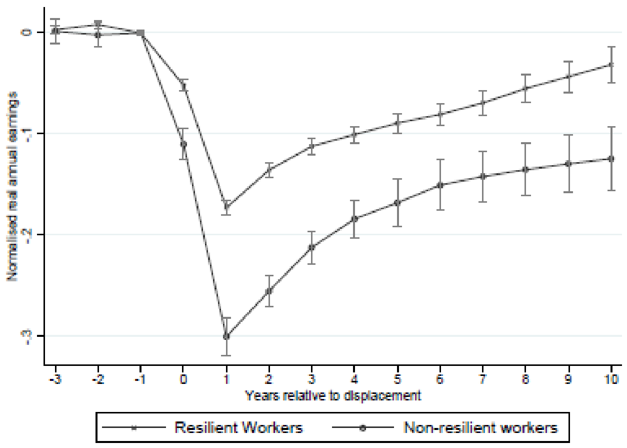
Swedish Plant Closures

Use machine learning to identify “resilient” and “non-resilient” groups.

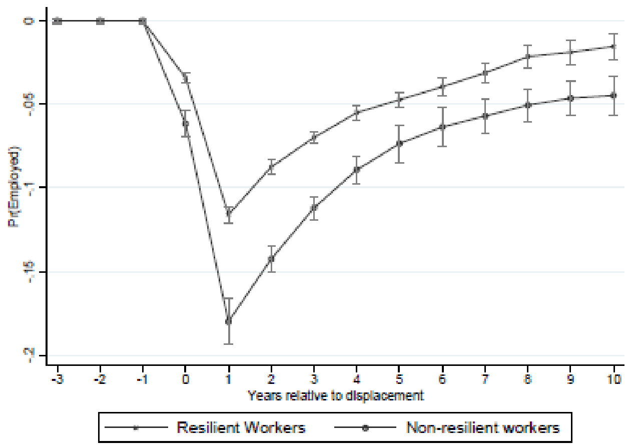
The separation in the figures shows that resilience is highly predictable.

Gives evaluation of earnings for observations with large earnings loss in the first year (non-resilient) and smaller earnings loss in the first year (resilient)

(a) Earnings



(b) Employment status



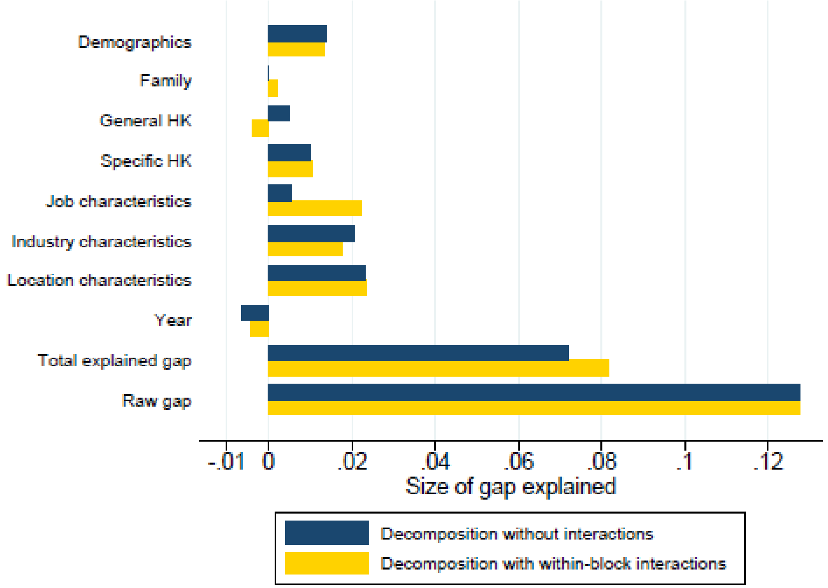
Resilience to Adverse Labor Market Shocks

Susan Athey (Stanford), Lisa K Simon (Stanford), Oskar N. Skans (UU), Johan Vikström (IFAU), Yaroslav Yakymovych (UU)

Swedish Plant Closures

What explains resilience?

- 1/3 remains unexplained, true relationship is highly non-linear.
- Industry and location characteristics are relatively more important than individual characteristics



Resilience to Adverse Labor Market Shocks

Susan Athey (Stanford), Lisa K Simon (Stanford), Oskar N. Skans (UU), Johan Vikström (IFAU), Yaroslav Yakymovych (UU)

FASFA Text Message Experiment

Experiment

- Run in 2017 and 2018 by ideas42 and the City University of New York (CUNY).
- Control group: business-as-usual emails from the college.
- Treatment groups: supplementary behavioral emails and text messages.
- Matriculated students from CUNY community colleges who had not yet renewed FAFSA in February of the study year.
- 2017: 25,167 students from 3 colleges
- 2018: 40,638 students from 5 colleges
- ATE: on-time submission increases by (in percentage points) 6.4 ± 0.6 (2017) and 12.1 ± 0.7 (2018), increasing early filing rates from 37% to 43% and 38% to 50%, respectively

FASFA Text Message Experiment

Text Message Content (using BMCC texts as an example):

Msg. #	Send Date and Time	Content
0	Wed, March 1 @ 6pm	**Part 1:** Hi {First Name}! This is the CUNY Student Persistence Team. To help you finish the year strong we will send you a few helpful texts. **Part 2:** Reply CANCEL if you don't want help setting yourself up for success. **Response to "cancel":** Thanks for letting us know. You will no longer receive texts from us.
1	Tues, March 14 @ 6pm	{First Name}, you must renew your FAFSA each year. This year it's easier -- you can use the same tax info as last year! Go to http://bit.ly/FAFSABMCC today.
2	Tues, March 28 @ 6pm	Renew your FAFSA and do it right the first time! Stop by the Financial Aid Lab (S115-C) and get help renewing today.
3	Wed, April 12 @ 6pm	Renew your FAFSA today! Many people renew in 30min or less at http://bit.ly/FAFSABMCC. Tip: use the IRS data retrieval tool to renew quickly.
4	Tues, April 25 @ 6pm	Unsure how to renew FAFSA? That's OK! Many students go before/after class to FinAid Lab (S115-C) for free help. Hrs: M/Th 10-6, F 10-5.
5	Tues, May 2 @ 6pm	{First Name Last Name}: FAFSA Status—NOT RENEWED. Renew now at http://bit.ly/FAFSABMCC

Using Machine Learning for Heterogeneous Treatment Effects

Data-Driven Approach to Heterogeneity

Estimate CATE → **Create subgroups** → **Compare group characteristics** → **Choosing an optimal policy**

- Estimate conditional average treatment effects (CATE) using machine learning algorithms (GRF package, Athey et al)

- Create subgroups based on the predicted treatment effect strength (e.g., quartiles)

- Observe how subgroups vary across a range of characteristics

- Is personalization worthwhile?

Results of FASFA Text Message Experiment

Who should be targeted?

- Those predicted (using ML) to be *unlikely* to file on time?
- Those predicted (using ML) to be *likely* to file on time?
- Those estimated (using causal ML) to have biggest treatment effects?

Findings

- Targeting those who were *likely* does as well as targeting based on treatment effects
- Nudging people over the finish line!

The Tech Toolkit for Incremental Innovation

Analytics

- Previous observational data
- Previous experiments

Innovation

- Algorithmic development: e.g. AI for personalization, recommendation systems
- Pilot experiment

Experimental Design

- Develop KPIs and validate externally
- Formulate hypotheses
- Pre-analysis planning
- Advanced experimentation (e.g. adaptive)

Analyze and Improve

- Generalizable insights
- Tactical insights
- New innovation plan