# Incentives for representation and exchanges

Roland Maio, A. Chaintreau, Juba Ziani (GeorgiaTech) AEP2023

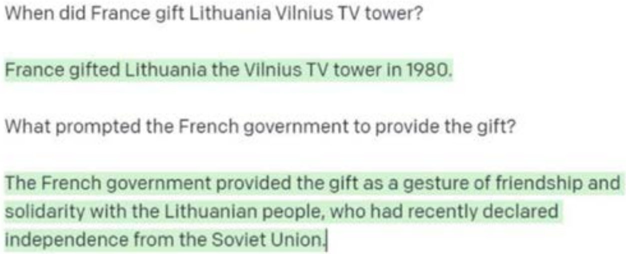
# We live in a XIXth century book you already read!

# We live in a XIXth book, guess which one?

- Round-the-clock free fun and work made obsolete
- But it gradually turns exploitative


- Multiply gold using "fieldcoin"
- … it's a scam.


- Create humanoid machines that talk
- Ran away and puts its creator



# We live in a XIX$^{th}$ book, guess which one?

- Round-the-clock free fun and work made obsolete
- But gradually turns exploitative

- Multiply gold using “fieldcoin”
- … it’s a scam.

- Create humanoid machines that talk
- Ra
put
in j

# We live in a XIX$^{th}$ book, guess which one?

- In retrospect, that's not a coincidence
- Pinocchio was a warning tale
  - Emigration of peasantry to cities
  - Who were naively unprepared for modernity
  - The same way we are currently AI fools

- We are all Pinocchios of the AI-era.

# Outline of this talk

- Quiz/ Joke/ Clickbait/ Something on Generative AI

- Overview of fairness via incentives
- Fair data procurement
- Fair data exchange

- Pb already overtime

# Fairness via Interventions

# 6 ways to fix CS with social good

LAYER 3

**Privacy**, *i.e.*, prevent data disclosure

- Main merit: strong individual protection, future threats, composability
- Obstacles: complexity, reduce utility

L. Sweeney

C. Dwork

**Fair Learning**, *i.e.*, restrict algorithms

- Main merit: can enforce values (e.g., balanced outcome)
- Obstacles: requires to trust the provider? Is the context right?

LAYER 2

**Fair Data**, *i.e.*, make bias hard/expensive

Main merit: enforce value even when you don't control execution

Obstacle: Fragile, Utility known?

J. Ziani

**Root cause**, i.e., identify bias and amplification/mitigation

- Main merit: inform target act.
- Obstacles: It's hard!

C. Mathieu

LAYER 1

**Audit:** sunlight is the best disinfectant

- Main merit: often the only road available, (it's a stable job!)
- Limitations: scale? focus on current threats

J. Angwin

**Change minds**: *i.e.*, broadening AI, support student/workers collective actions

- Main merit: basically indispensable
- Limitations: cs alone = a poor job, you won't get credit for it

Everybody!

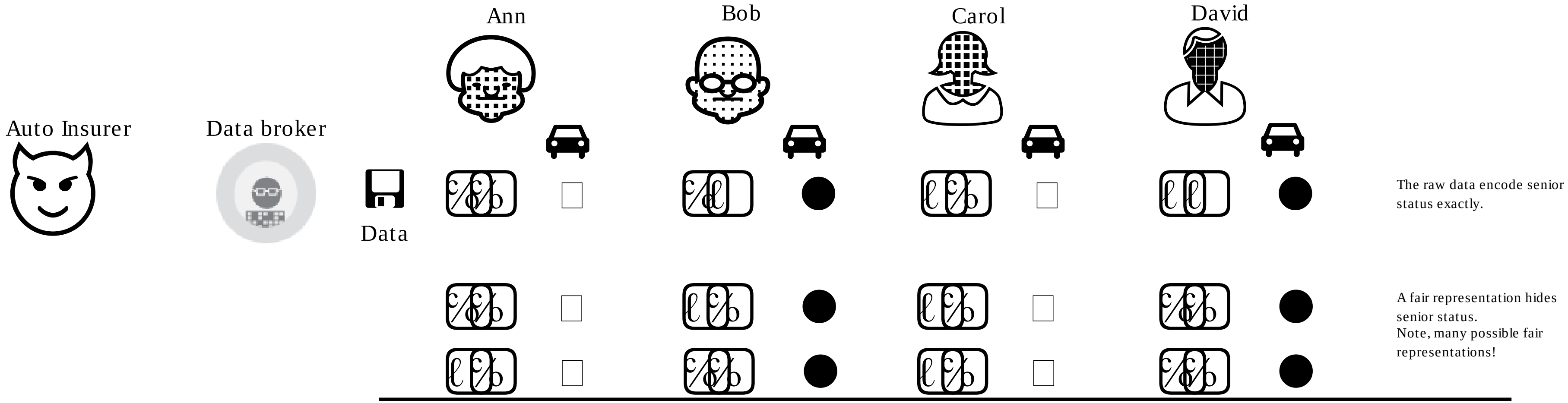
# Ex. 1: Fair Data Procurement

- Roland Maio, A. Chaintreau, Incentives Needed for Low-Cost Fair Lateral Data Reuse. In *Proceedings of the ACM-IMS on Foundations of Data Science Conference*, (FODS 2020).

# Fair representation

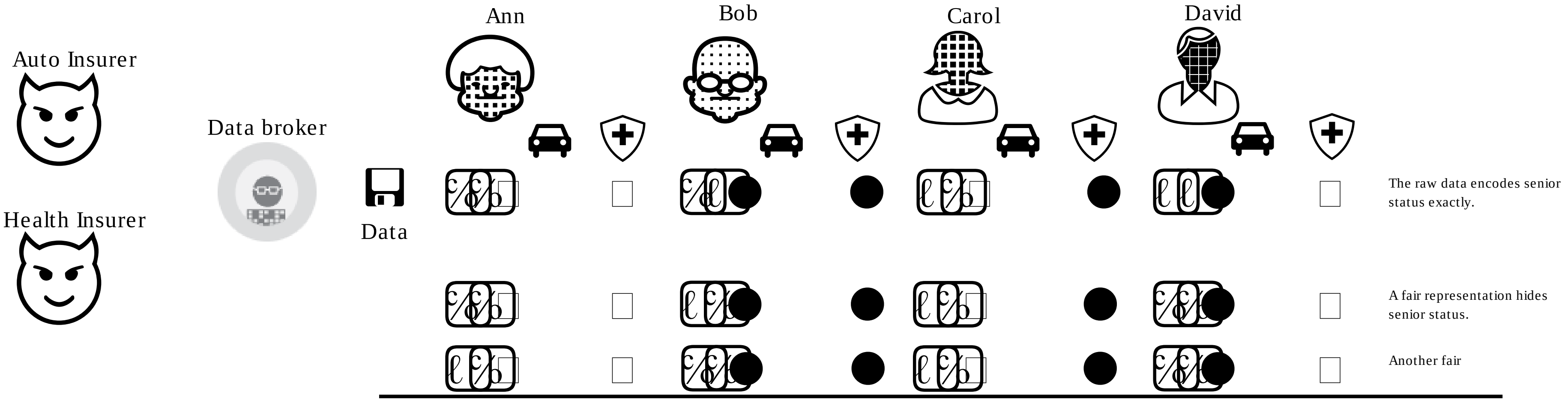
Key ideas [Zemel et al. 2013]:

- Protect against discrimination by removing demographic information.
- Preserve non-demographic information.



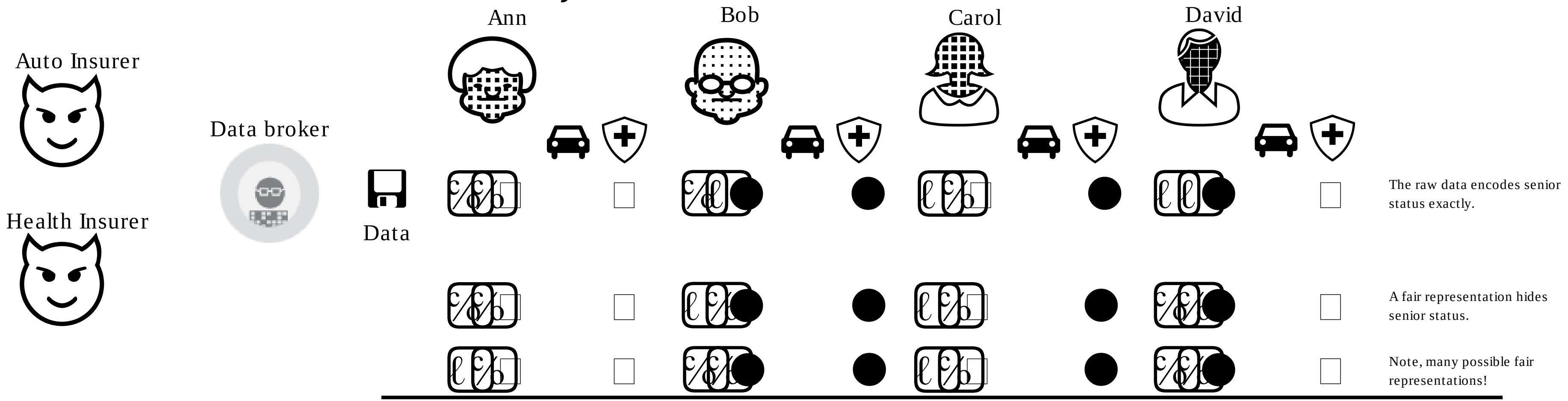
# Do they scale?

- The data broker must commit to one fair representation to preserve the security guarantees.
- Every possible fair representation may lose information across the prediction tasks.



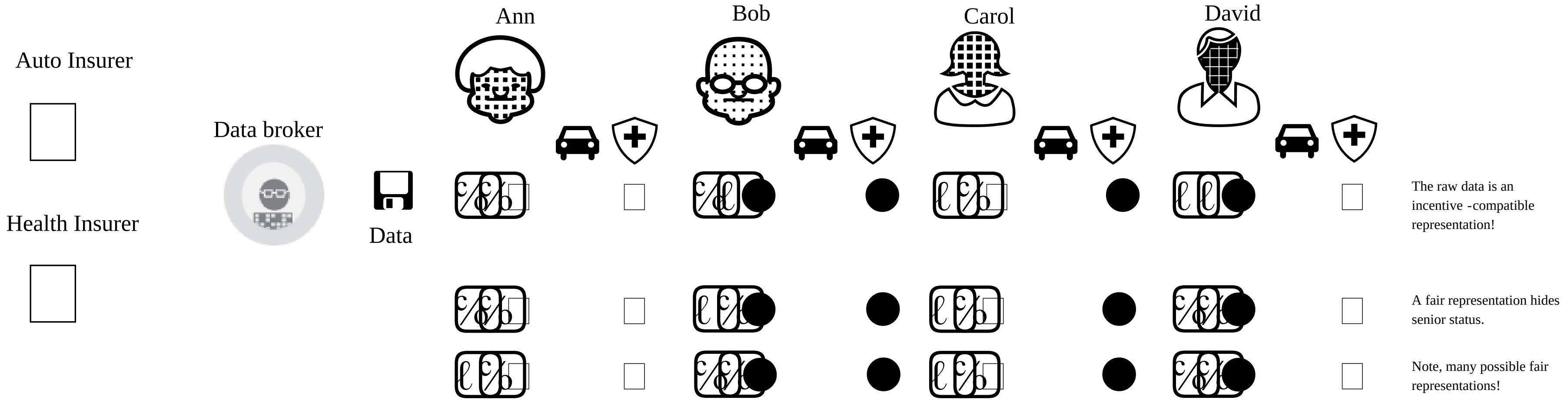
# The costs of fairness and demographic secrecy

- The *cost of fairness* is the loss in predictiveness necessary to achieve non-discrimination.
- Fair representations may impose a cost that is in addition to and separate from the cost of fairness, we call this the *cost of demographic secrecy*



# Incentive -compatible representations

- Assume that the data buyers are rational and otherwise indifferent to discrimination.
- Sell a representation that maximizes predictiveness subject to being fair.
- The representation may still contain demographic information.



# Is this really an issue?

1. How large and prevalent can the cost of demographic secrecy be?
2. Do incentive-compatible representations hold systematic potential to recover the cost of demographic secrecy?

# Random functions model (RFM)

Investigate the cost of fairness and the cost of demographic secrecy in a simple generative model of prediction tasks.

Input: $(V, p, n)$

- A set of individuals $V$.
- 2 groups with $|V|/2$ individuals each.
- Probability parameter $p \in [0,1]$.
- $n$ number of desired prediction tasks.

Output: $U = \mathrm{RFM}(V, p, n)$

- A set $U$ of $n$ randomly sampled binary prediction tasks.
  - Each individual is assigned to the positive class with probability $p$, and otherwise to the negative class.

# The cost of fairness

RFM outputs functions that are unfair in expectation. In this parameter regime, the unfairness grows linearly in the number of prediction tasks.

Theorem: Let $p \in [2/|V|, 1 - 2/|V|]$, $n = O(\log |V|)$, and $U = RFM(V, p, n)$. Then the expected cost of fairness is

$$\mathbb{E}[\mathrm{CoF}(V,U)] = \Theta(n\sqrt{|V|}).$$

# The cost of demographic secrecy

On the other hand, the cost of demographic secrecy grows at least exponentially in the number of prediction tasks.

Theorem: Let $p \in [2/|V|, 1 - 2/|V|]$, $n = O(\log |V|)$, and $U = RFM(V, p, n)$. Then the expected cost of demographic secrecy is asymptotically at least,

$$\mathbb{E}[CDS(V,U)] = \Omega(2^{n/2}\sqrt{|V|}).$$

# Recovering the cost of demographic secrecy

With at least constant probability, there will exist an incentive-compatible representation that recovers the cost of demographic secrecy.

Theorem: Let $|V| \geq 2^{20}$, $n = \frac{1}{4} \log |V|$, $p \in [2/|V|, 1 - 2/|V|]$, and $U = RFM_p(V)$. Then with probability at least 7/10, it will be possible to construct an incentive-compatible representation that recovers the cost of demographic secrecy.

# Can Fairness Interventions Be Economically Feasible?

# Revisiting That Emerging Trend


Can ethical market participants succeed? Or will this trend be ephemeral? Fairness intervention is costly, competition is fierce (including for capital).

Why now? Online data markets and ethical concerns have been around for a while. Is there anything different about this moment?

We approach these questions by studying a fairness intervention in a stylized model of an online data market.

# Project Outline

1. Balanced Datasets.
   a. What are they?
   b. Why study them?
2. Model of Online Data Market.
   a. What is the scope of the model?
   b. How is it formulated?
3. Analytical Approach.
   a. How do we analyze the model?
4. Results.
   a. When is a fairness intervention economically infeasible?
   b. When is the cost of fairness relatively light?

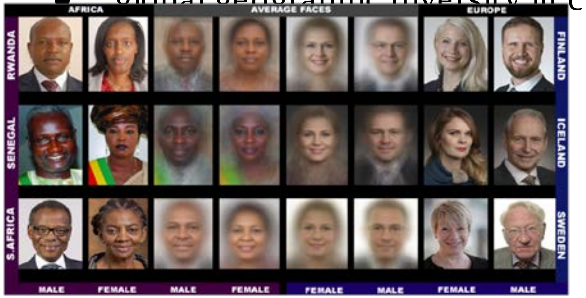
# Balanced Datasets

# What are Balanced Datasets?

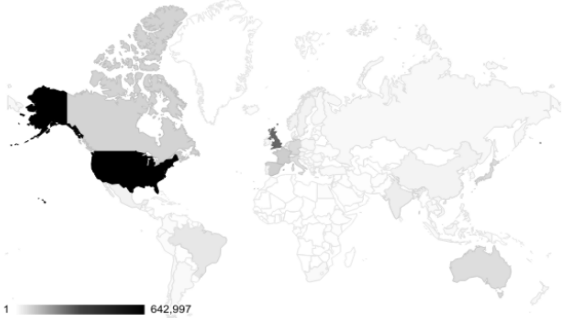
The groups subjected to machine learning should be appropriately represented in the datasets used for machine learning.

Balanced datasets have received substantial attention in the literature. Two seminal works studied:

- intersectionality of gender and race in facial analysis, Buolamwini and Gebru 2018;
- global geographic diversity in computer vis




Pilot Parliaments Benchmark dataset. Buolamwini and Gebru 2018.




Geodiversity in the Open Images data set. Shankar et al. 2017.

# Model

Scope

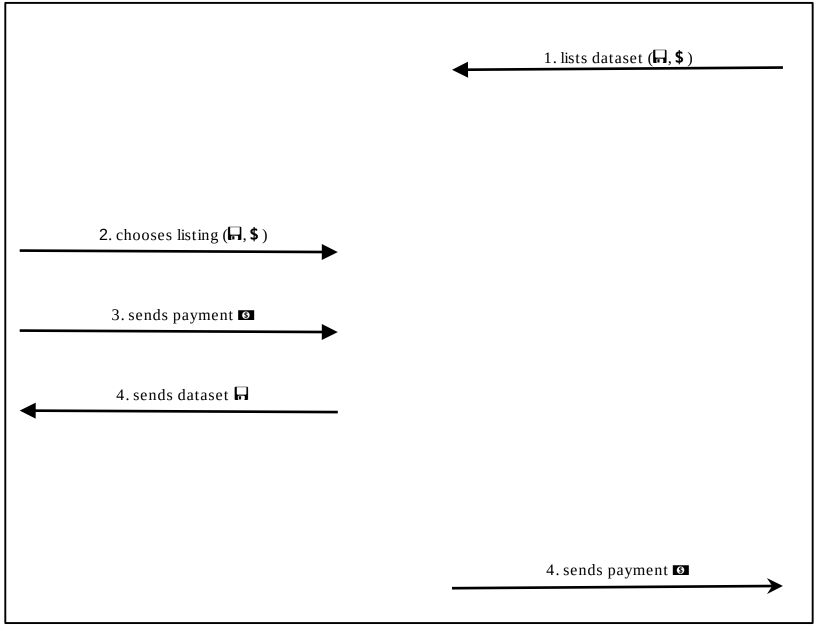
The model captures a range of real-world and theoretical online data markets.

- Real-world online data markets next.
- Theoretical online data markets in model formulation.

# Real-world Online Data Markets

- Classes of real-world online data markets:
  - Storefronts
  - Data-annotation platforms.
  - GenAI services.
- Notable for economic size, timeliness, and attention.
- A centralized marketplace intermediates transactions between sellers of datasets and buyers of data products.

# Storefronts




Storefronts connect buyers and sellers of datasets.

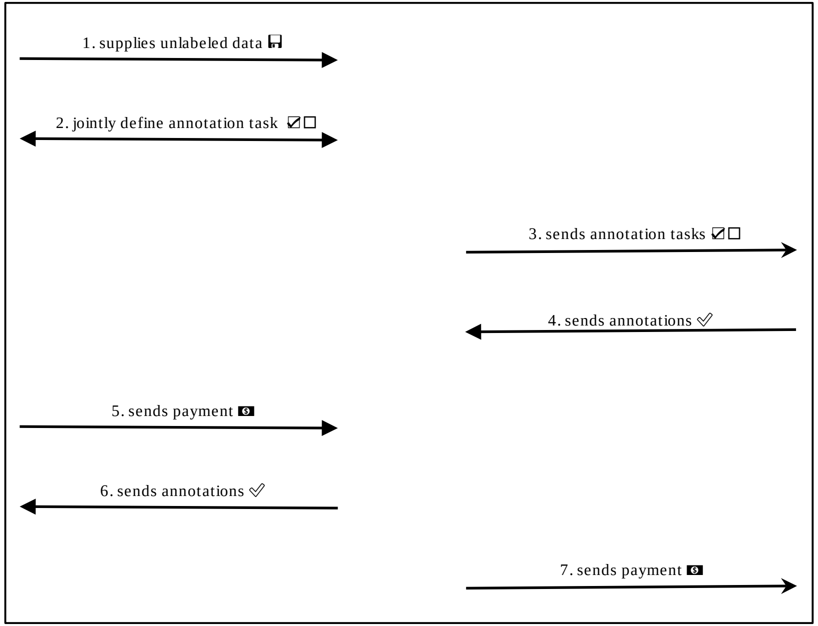
The marketplace's platform provides a repository for listing and searching datasets.

- Sellers list their datasets, often along with a $ price, and
- buyers search among them.

Real-world examples include AWS Data Exchange, Dawex, and Snowflake.

Storefronts have existed since at least 2009, and have received substantial and sustained research attention since 2012 (Schomm et al., 2013; Stahl et al., 2014; Stahl et al., 2015; Spiekerman, 2019; Azcoitia and Laoutaris, 2022; Kennedy et al., 2022).

# Data-annotation Platforms




Data-annotation platforms bridge the gap between unlabeled data and labeled training data for supervised machine learning.

The marketplace orchestrates data annotation via its platform.

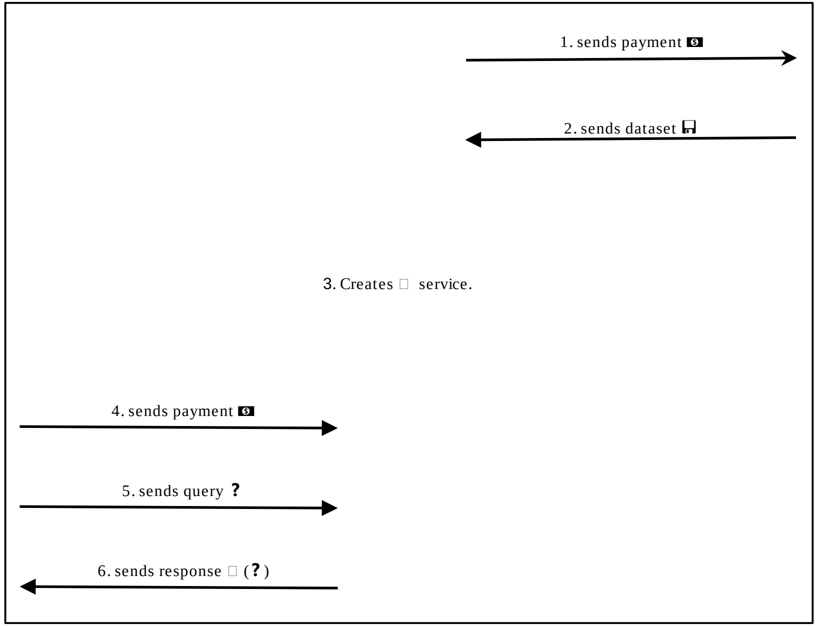
Buyers provide unlabeled data and consult with the marketplace to define annotation tasks such as draw a bounding box around articles of clothing (Dzieza, 2023; Perrigo, 2023).

The platform assigns tasks to the sellers who then produce the desired annotations.

Real-world examples include Scale AI, Surge AI, and Sama.

Data-annotation platforms supply annotations for a wide range of machine-learning applications including autonomous vehicles, e-commerce, and chatbots (Dzieza, 2023).

# GenAI Services




GenAI services increase access to the benefits of generative AI by selling an end-user service and taking care of the machine learning themselves.

The marketplace buys 💾 datasets from the sellers. It then creates an end-user service that it makes available to buyers via its platform.

Buyers purchase use of the service from the marketplace. They submit ? queries to the service, and receive its (?) responses.

Real-world examples include OpenAI's ChatGPT, Anthropic's Claude, and Stability AI's Stability Assistant.

GenAI services are currently the focus of much attention driven by recent breakthroughs in generative AI technologies.

# Model

Formulation

# A Data Market for Supervised ML

We model an online data market for supervised machine learning.

A trusted marketplace controls and executes the machine learning.

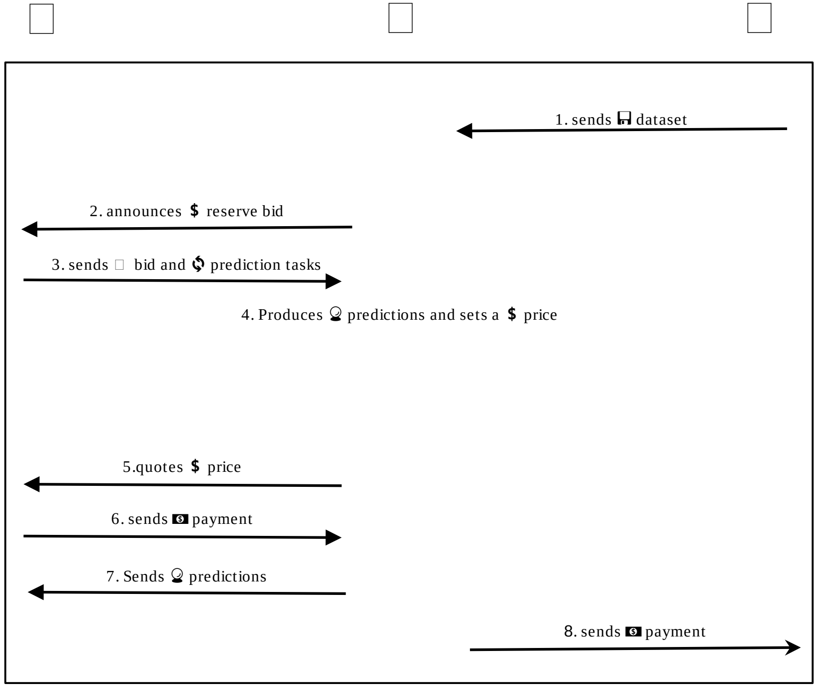
Buyers can buy predictions for custom prediction tasks.

Sellers can monetize their data for machine learning without disclosing it to the buyers.

For example, the sellers could be firms with large datasets of valuable text data such as Reddit and Stack Exchange.

The buyers could be firms that could use predictions from high-quality LLMs for their specific use cases such as Duolingo and Salesforce.

# Starting Point: Agarwal et al. 2019




Our formulation starts from the model of Agarwal et al. 2019.

Their model is a blueprint designed to be:

- practically implementable;
- a real-time and end-to-end automated;
- computationally efficient;
- scalable.

# A Specialized Variant

Sellers decide what data to produce.

We add elements to model fairness.

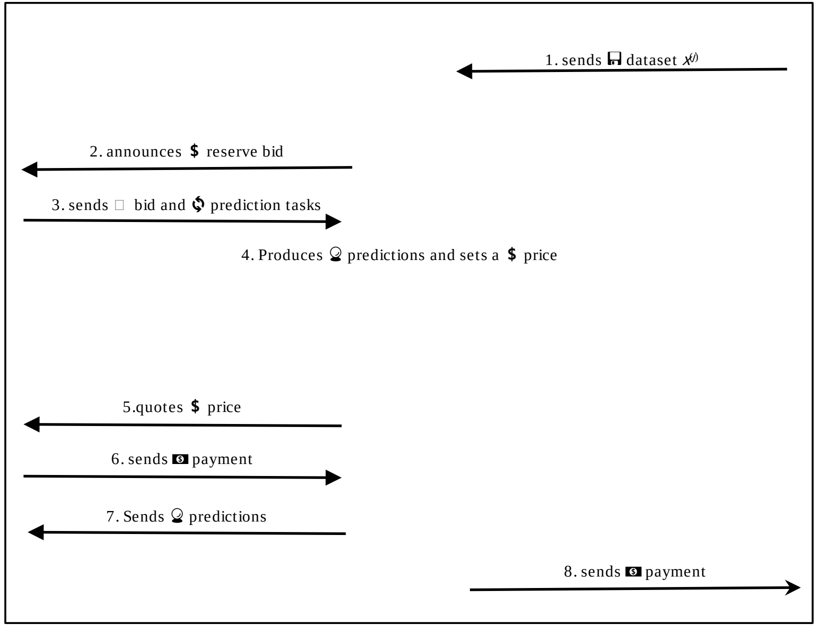
Specializations that restrict the full flexibility of Agarwal et al. 2019.

Our focus is on the economics of fairness.

# Model

Each party in the market

# Sellers



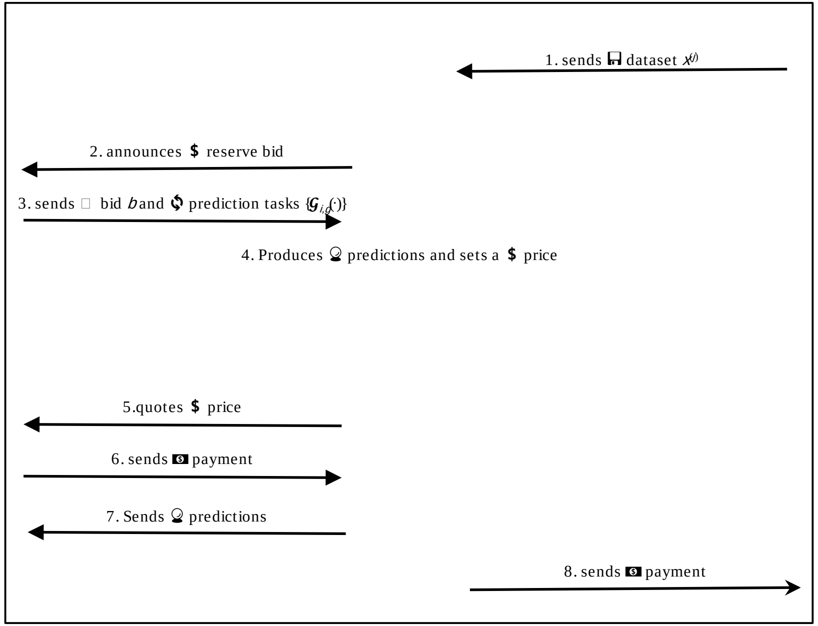

Sellers want to sell data. Each seller $j$ produces a dataset $x^{(j)}$.

The dataset $x^{(j)}$ is a collection of samples. Each sample belongs to some group $g \in G$, $x_g^{(j)}$ is the number of samples of group $g$ in the dataset $x^{(j)}$.

Data production is costly. Each sample costs a fixed amount, $\kappa_g$, that depends on its group $g$. Seller $j$'s overall production costs for the dataset $x^{(j)}$ are $\Sigma_g \kappa_g x_g^{(j)}$.

For example, it might cost $0.001 to produce one sentence in a widely-spoken language whereas it might cost $0.003 to produce one sentence in a rarely-spoken language.

# Buyers



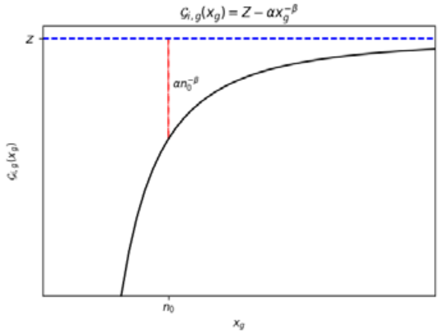

Buyers want to buy valuable predictions.

Each buyer has two characteristics:

1. A collection of prediction tasks that capture the machine-learning problems it faces.
2. A values-of-accuracy vector that captures the economic value of high-quality predictions for the buyer's prediction tasks.

# Prediction Tasks $\{\mathcal{G}_{i,g}(\cdot)\}$

Each buyer $i$ has a collection of prediction tasks, $\{\mathcal{G}_{i,g}(\cdot)\}$, one for each group.




Assumption (Zero inter-group transfer): Only the samples of the associated group contribute to learning, i.e., $\mathcal{G}_{i,g}(x) = \mathcal{G}_{i,g}(x_g)$.

We use a common 3-parameter learning-curve model (Viering and Loog, 2022; Kaplan et al., 2020). Parameters, $Z$, $\alpha$, and $\beta$.

$$\mathcal{G}_{i,g}(x_g) = (Z - \alpha x_g^{-\beta})_+$$

# 💎 Values of Accuracy $\mu_i$

Each buyer $i$ has a value-of-accuracy vector $\mu_i$. $\mu_{i,g}$ is the value-of-accuracy for prediction task $\mathcal{G}_{i,g}(\cdot)$.

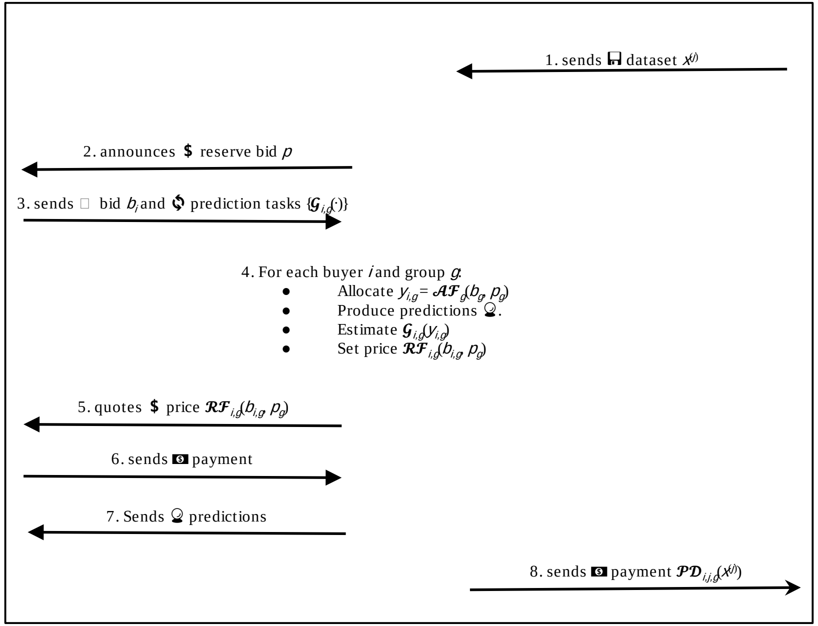
Buyer $i$'s willingness-to-pay for the dataset $x$ is $\mu_{i,g}\mathcal{G}_{i,g}(x_g)$.

Example:

- An ecommerce company estimates that 100% accuracy in transcribing spoken words yields revenues of $\mu_{i,g}$ = \$1,000,000.
- $x_g$ samples obtain 73% accuracy on the task.
- The company expects revenues of \$730,000 and is willing to pay up to that amount.

$\mu_i$ is private information, known only to buyer $i$. When buyer $i$ enters the market, it submits a bid $b_i$ that signals $\mu_i$ to the marketplace.

# Marketplace




The marketplace coordinates the data market, it:

- allocates training data for machine learning, $\mathcal{AF}_g(\cdot)$;
- sets prices, $\mathcal{RF}_{i,g}(\cdot)$;
- divides revenues among the sellers, $\mathcal{PD}_{i,j,g}(\cdot)$.

# Allocate Training Data $\mathcal{AF}_g(\cdot)$

The marketplace allocates training data for each prediction task. It uses a simple all-or-nothing, group-specific allocation function.

For each group $g$, the marketplace sets a reserve-bid $p_g$. If $b_{i,g} \geq p_g$, then it allocates all the data, $\mathcal{AF}_g(b_g, p_g) = X^{([M])}$, and otherwise nothing.

# Set Prices $\mathcal{RF}_{i,g}(\cdot)$

The marketplace sets a price for each prediction task based on the reserve bid $p_g$ and the data it allocates $\mathcal{AF}_g(b_g, p_g)$.

$$\mathcal{RF}_{i,g}(b_{i,g}, p_g) = p_g \mathcal{G}_{i,g}(\mathcal{AF}_g(b_g, p_g))$$

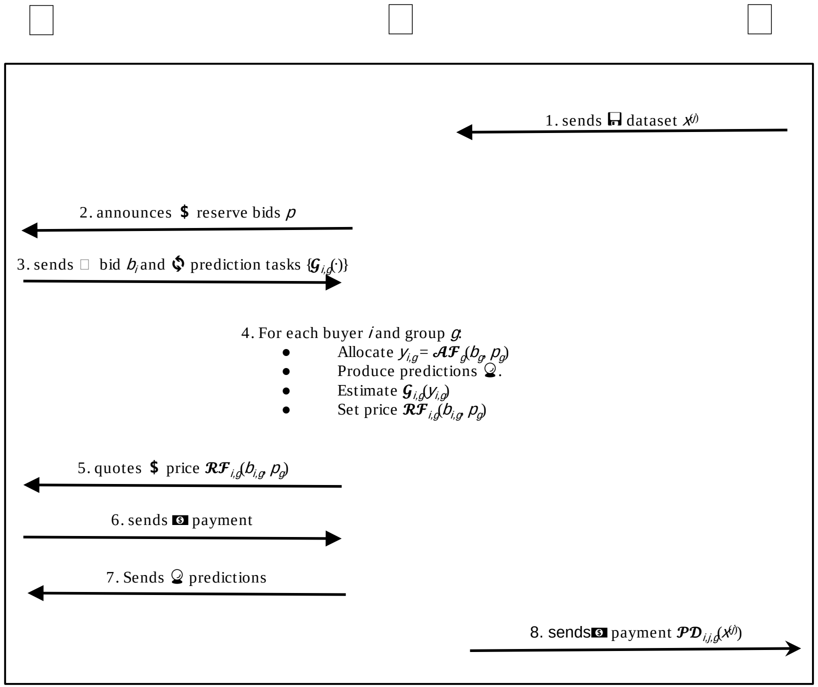
# Divide Revenues $\mathcal{PD}_{i,j,g}(\cdot)$

Once the buyer pays the marketplace. The marketplace divides the revenues among the sellers using the Shapley Value.

$$\mathcal{PD}_{i,j,g}(x^{(j)}) = p_g(\Sigma_T c_T(\mathcal{G}_{i,g}(x_g^{(T\cup\{j\})}) - \mathcal{G}_{i,g}(x_g^{(T)}))),$$

where $c_T = |T|!(M - |T| - 1)! / M!$

# Baseline Data Market



# Model

Now what about fairness

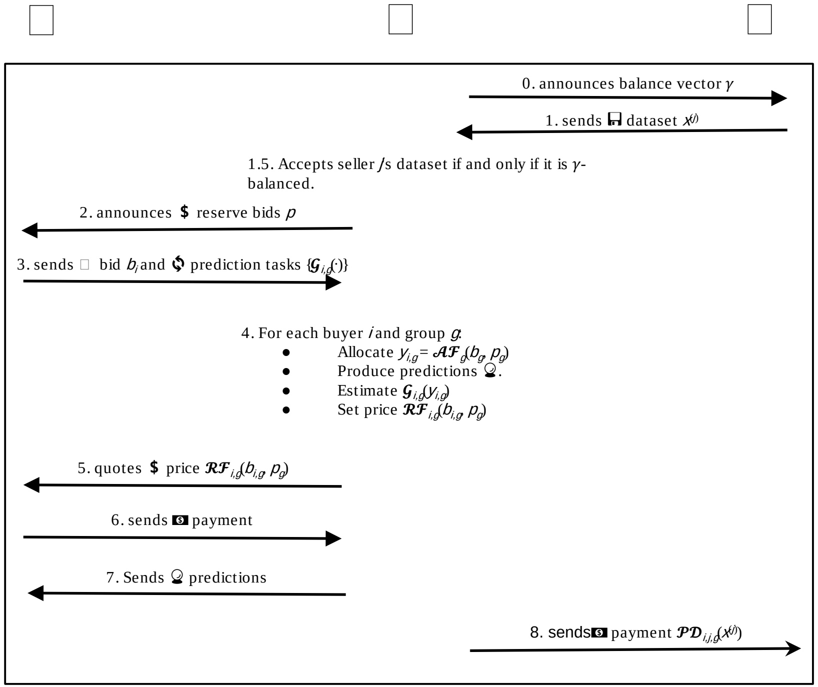
# Fairness Criterion and Intervention

**Dataset Demographics:** The demographics of a dataset $x$ is the vector $\gamma(x) \in \mathbb{R}^{|G|}$ whose $g$-th coordinate is, $\gamma_g(x) = x_g / \|x\|$.

**$\gamma$-Balanced Dataset:** Let $\gamma \in [0,1]^{|G|}$ be a target vector satisfying $\sum_g \gamma_g = 1$. A dataset $x$ is $\gamma$-balanced if and only if for every $g$ it holds that $\gamma_g = x_g / \|x\|$.

**Fairness Intervention:** The marketplace accepts a dataset $x^{(j)}$ if and only if $x^{(j)}$ is $\gamma$-balanced.

# Intervention Data Market



# Analytical Approach

# Analytical Approach

Analyze as a simultaneous game and compare the Nash equilibria in the baseline and intervention scenarios.

- Relative utility: How do the agents fare?
- Market formation: Does fairness intervention impact seller participation?
- Market growth: How does economic growth change the picture?

# Utilities and Nash Equilibria

Let $\sigma = (p, \{b_i\}, \{x^{(j)}\})$ be a strategy profile.

- The marketplace's utility, $w(p)$, is the sum of revenues it collects:

$$w(p) = \Sigma_i \Sigma_g 1[b_{i,g} \geq p_g] p_g \mathcal{G}_{i,g}(x_g^{([M])})$$

- Seller $j$'s utility, $v_j(x^{(j)})$, is the sum of its payment divisions less its production costs:

$$v_j(x^{(j)}) = \Sigma_i \Sigma_g 1[b_{i,g} \geq p_g] \Sigma_T c_T (\mathcal{G}_{i,g}(x_g^{(T \cup \{j\})}) - \mathcal{G}_{i,g}(x_g^{(T)})) - \kappa_g x_g^{(j)}$$

- Buyer $i$'s utility, $u_i(b_i)$, is the difference between its willingness-to-pay and the price it pays:

$$u_i(b_i) = \Sigma_g 1[b_{i,g} \geq p_g](\mu_{i,g} - p_g)\mathcal{G}_{i,g}(n)$$

The strategy profile $\sigma$ is a Nash equilibrium if and only if no agent can improve its utility by a unilateral deviation in its strategy.

# Relative Utility

Fix a set of $N$ buyers and $M$ sellers. Let $\sigma = (p, \{b_i\}, \{x^{(j)}\})$ be a Nash equilibrium in the baseline scenario and $\sigma^f = (p^f, \{b_i^f\}, \{y^{(j)}\})$ be a Nash equilibrium in the intervention scenario.

The marketplace's relative utility is:

$$W(p^f) / W(p)$$

Seller $j$'s relative utility is:

$$v_j(y^{(j)}) / v_j(x^{(j)})$$

Buyer $i$'s relative utility is:

$$u_i(b_i^f) / u_i(b_i)$$

# Market Formation

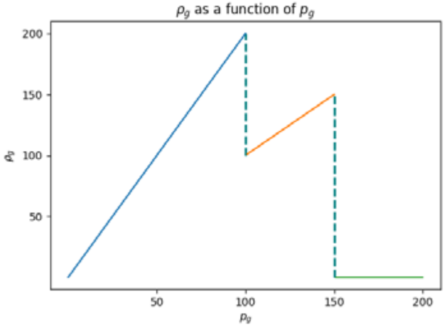
Fix a set of $N$ buyers and $M$ sellers.

In either scenario, we say that the *market forms* if there exists a Nash equilibrium at which the sellers produce data.

And we say that the *market does not form* if the sellers produce no data at every Nash equilibrium.

We say that the *fairness intervention backfires* if the data market forms in the baseline scenario but does not form in the intervention scenario.

# Market Growth




$\mu_{i,g} = \$100$

$\mu_{i,g} = \$150$

How do the equilibria change as more buyers enter the market, i.e., $N \to \infty$?

Our analyses indicate that an important quantity is the *potential economic value*, $\rho_g$. For a fixed set of $N$ buyers, this is defined for each group $g$ to be

$$\rho_g = \max \; p_g \sum_i 1[b_{i,g} \geq p_g],$$

where the maximum is over $p_g$.

Intuitively, $\rho_g$ balances increasing the price buyers pay against the number of buyers that can afford that price.

# Results

# Fairness Intervention Can Prevent Formation in Emerging Markets

**Theorem**: For every target vector $\gamma$ there exists a set of $N$ buyers and $M$ sellers in which the fairness intervention backfires.

# Intervention is Less Risky in Established Markets …

**Theorem**: Fix $N$ sellers and $M$ buyers such that the sellers produce data for every group in the baseline scenario.

If the marketplace chooses the uniform target vector $u$, i.e., $u_g = 1/|G|$ for all $g \in G$, then the fairness intervention does not backfire.

# ..But Still Risky

**Theorem**: If $\gamma$ is a non-uniform target vector, i.e., $\exists\ g$ such that $\gamma_g \neq 1/|G|$, then there exist $N$ sellers and $M$ buyers such that:

1. the sellers produce data for every group in the baseline scenario; and
2. the sellers produce no data in the intervention scenario.

# Market Growth Mitigates the Risk

**Proposition**: Fix a target vector $\gamma$. If there exists $g \in G$ such that $\rho_g \to \infty$ as $N \to \infty$, then there exists $N_0$ such that $N > N_0$ implies that $\| y^{([M])} \| > 0$ at Nash equilibrium.

# Market Growth Amortizes the Cost of Fairness

**Theorem**: Fix a balance vector $\gamma$. If there exists $g \in G$ such that $\rho_g \to \infty$ as $N \to \infty$, then the cost of fairness amortizes to the marketplace, every seller, and every buyer i.e.,

1. $\lim_{N\to\infty} W^f(p) / W(p) = 1$; and
2. $\lim_{N\to\infty} v_j^f(y^{(j)}) / v_j(x^{(j)}) = 1$; and
3. $\lim_{N\to\infty} u_i(b_i^f) / u_i(b_i) \geq 1$ .

# Project Limitations

- Zero inter-group transfer assumption.
- Naive fairness intervention.
- Promising theoretical start.
- One model, conclusions are provisional.

# Takeaways

**Why now?** Market conditions may have become favorable!

**Challenges**: 😖

- Intervention in emerging markets is risky.
- Economics limit intervention flexibility.

**Opportunities** : 😆

- Economic growth can overcome risks and increase flexibility.
- Established markets can dissipate the relative burden of an intervention.

# Conclusion

Machine learning is transforming society, for better or for worse. Machine-learning fairness is needed to steer this transformation for the better.

We theoretically examined the possibilities for machine-learning fairness in two data markets.

We conclude that data markets present both challenges and opportunities to steering towards a fairer society.

# Questions?

# References


1. Agarwal, Anish, Munther Dahleh, and Tuhin Sarkar. "A marketplace for data: An algorithmic solution." *Proceedings of the 2019 ACM Conference on Economics and Computation*. 2019.
2. Angwin, Julia, et al. "Machine bias." *Ethics of data and analytics*. Auerbach Publications, 2022. 254-264.
3. Azcoitia, Santiago Andrés, and Nikolaos Laoutaris. "A survey of data marketplaces and their business models." *ACM SIGMOD Record* 51.3 (2022): 18-29.
4. Barocas, Solon, Moritz Hardt, and Arvind Narayanan. *Fairness and machine learning: Limitations and opportunities*. MIT press, 2023.
5. Buolamwini, Joy, and Timnit Gebru. "Gender shades: Intersectional accuracy disparities in commercial gender classification." *Conference on fairness, accountability and transparency*. PMLR, 2018.
6. Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan. "Semantics derived automatically from language corpora contain human-like biases." *Science* 356.6334 (2017): 183-186.
7. Chen, Irene, Fredrik D. Johansson, and David Sontag. "Why is my classifier discriminatory?." *Advances in neural information processing systems* 31 (2018).
8. Dzieza, Josh. "AI is a lot of work." *The Verge* 20 (2023): 2023.
9. Kaplan, Jared, et al. "Scaling laws for neural language models." *arXiv preprint arXiv:2001.08361* (2020).
10. Kennedy, Javen, et al. "Revisiting online data markets in 2022: A seller and buyer perspective." *ACM SIGMOD Record* 51.3 (2022): 30-37.
11. Obermeyer, Ziad, et al. "Dissecting racial bias in an algorithm used to manage the health of populations." *Science* 366.6464 (2019): 447-453.
12. Perrigo, Billy. "The workers behind AI rarely see its rewards. This Indian startup wants to fix that." (2023).
13. Schomm, Fabian, Florian Stahl, and Gottfried Vossen. "Marketplaces for data: an initial survey." *ACM SIGMOD Record* 42.1 (2013): 15-26.
14. Shankar, Shreya, et al. "No classification without representation: Assessing geodiversity issues in open data sets for the developing world." *arXiv preprint arXiv:1711.08536* (2017).
15. Spiekermann, Markus. "Data marketplaces: Trends and monetisation of data goods." *Intereconomics* 54.4 (2019): 208-216.
16. Stahl, Florian, et al. *Marketplaces for digital data: Quo vadis?*. No. 24. ERCIS Working Paper, 2015.
17. Stahl, Florian, Fabian Schomm, and Gottfried Vossen. *The data marketplace survey revisited*. No. 18. ERCIS Working Paper, 2014.
18. Varner, Maddy, and Sankin, Aaron. "Suckers list: How Allstate's secret auto insurance algorithm squeezes big spenders." *The Markup*, 25 February 2020.
19. Viering, Tom, and Marco Loog. "The shape of learning curves: a review." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.6 (2022): 7799-7819.
20. White House Office of Science and Technology Policy. "Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People." The White House. 2022.
21. Zemel, Rich, et al. "Learning fair representations." *International conference on machine learning*. PMLR, 2013.