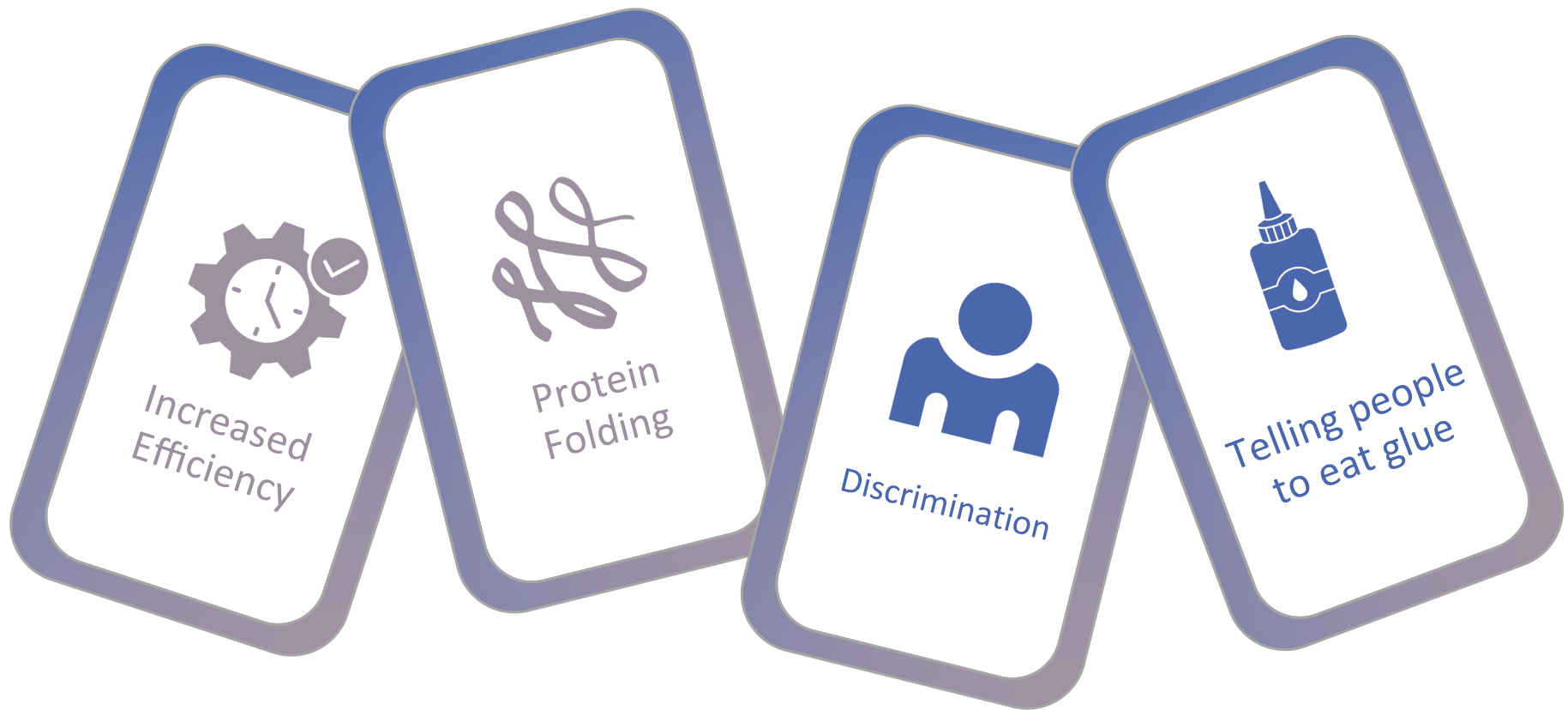


Human-Centered Data Collection for Machine Learning: Lessons from Survey Research

Stephanie Eckman
December 16, 2025

Benefit and Harm from AI



Training Data Collection

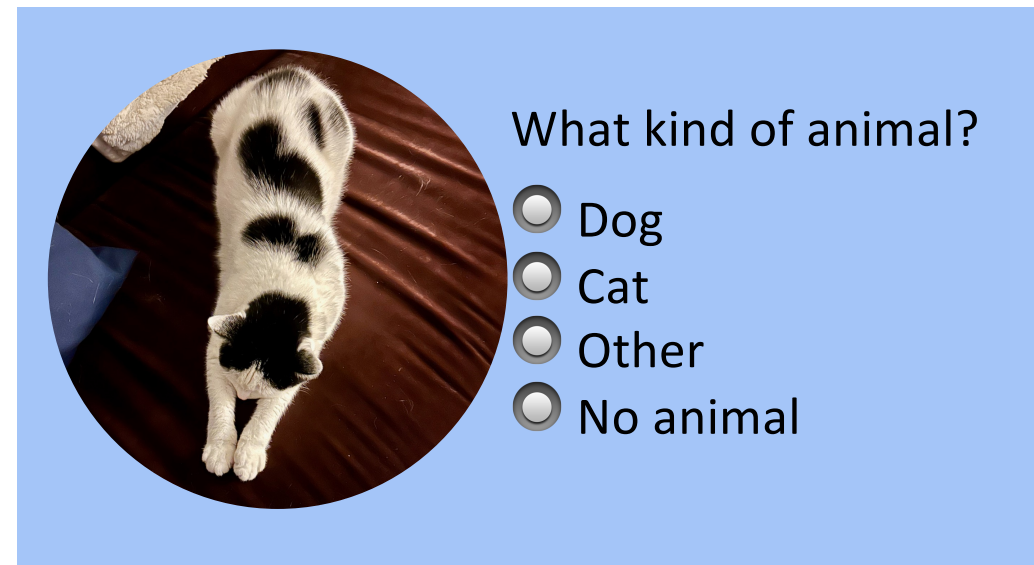
Tend to be objective tasks

Crowdworkers

- Mechanical Turk
- Appen
- Scale AI

Experts

Researchers / Graduate Students



Human Feedback Data Collection

Trained on everything

Human feedback data

- Subjective, opinion tasks
- Collected from crowdworkers

Choose the best response to this query:

What's the weather like?

- ☐ It is 28°C and humid.
- ☐ It's a warm, humid 28 degrees Celsius out there today – feels pretty sticky!

Core Insight



What kind of animal?

- ☐ Dog
- ☐ Cat
- ☐ Other
- ☐ No animal

Choose the best response to this query:

What's the weather like?

☐ It is 28°C and humid.

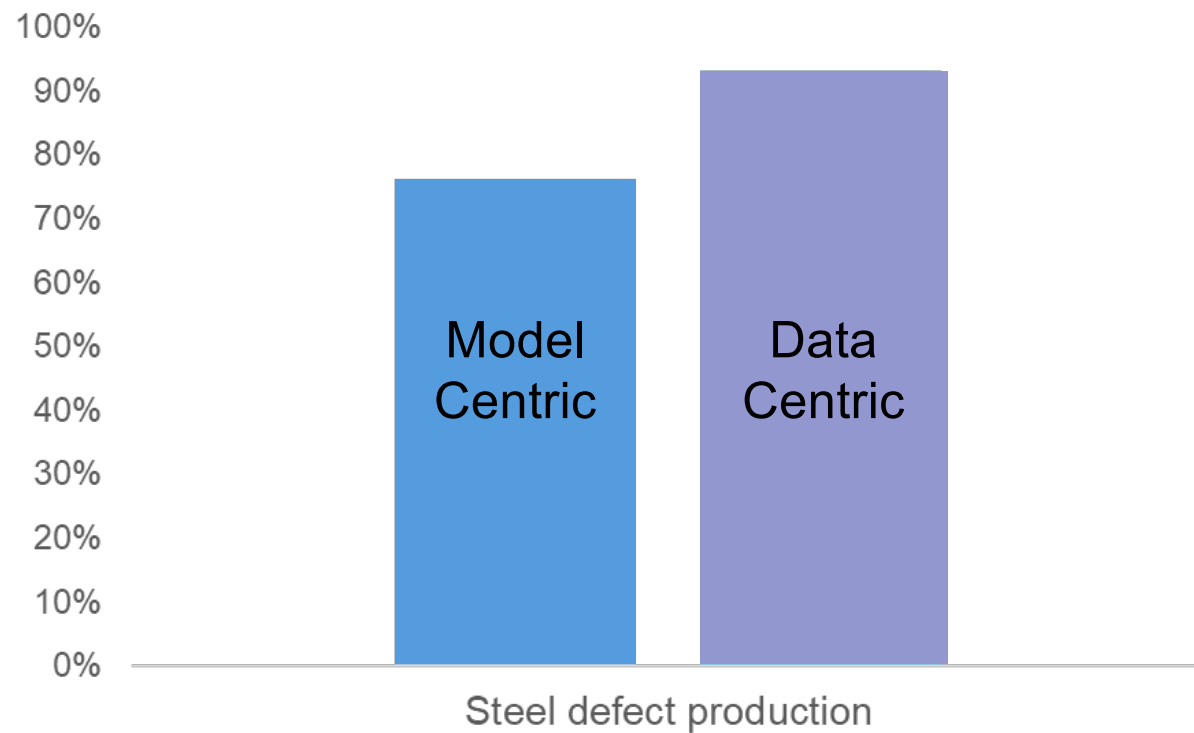
☐ It's a warm, humid 28 degrees Celsius out there today – feels pretty sticky!

A lot like surveys!

Impact of Data Centric AI

"reasonable algorithm
with good data
is preferable to
a great algorithm with
not-so-good data"

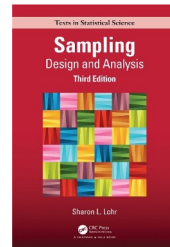
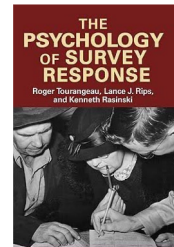
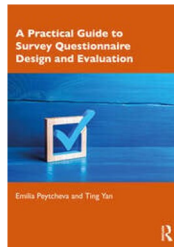
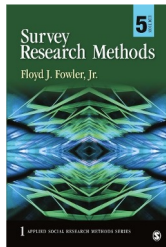
- Andrew Ng



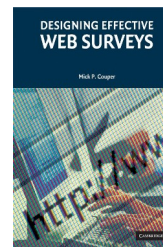
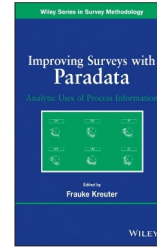
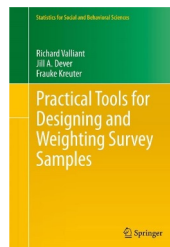
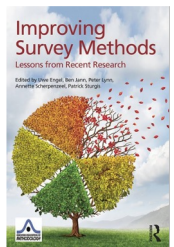
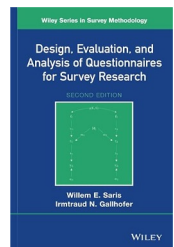
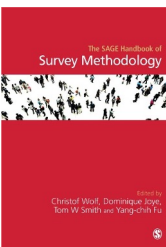
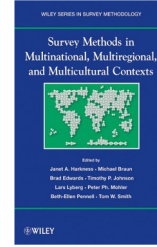
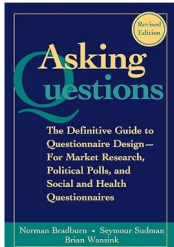
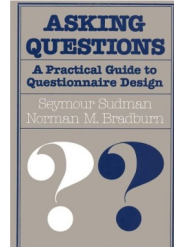
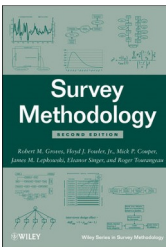
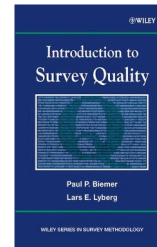
Science of Data Collection, Data Quality



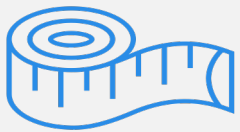
Questions



Quality



Survey Data Concerns



Measurement: Are the **labels** answers correct?



Representation: Who **labels** responds?

Example Instruments

East Repair Inc.

1912 Harvest Lane
New York, NY 12210

LOGO

Bill To

John Smith
2 Court Square
New York, NY 12210

Ship To

John Smith
3787 Pineview Drive
Cambridge, MA 12210

Invoice Date

11/02/2019

P.O.#

2312/2019

Due Date

26/02/2019

Qty	Description	Unit Price	Amount
1	Front and rear brake cables	100.00	100.00
2	New set of pedal arms	15.00	30.00
3	Labor 3hrs	5.00	15.00

Subtotal

145.00

Sales Tax 6.25%

9.06

Total

\$154.06

John Smith



Entities



Relations



Classification

Keyword Search



DATA_ID

1

DATE

2

INVOICE_ID

3

INVOICE_NUMBER

4

SELLER_ID

5

SELLER

6

MONTANT_HT_ID

7

MONTANT_HT

8

TVA_ID

9

TVA

E

TTC_ID

R

TTC

T

East Repair Inc. SELLER

1912 Harvest Lane LOGO New York, NY 12210 Bill To

Ship To SELLER_ID

Invoice Date

DATA_ID

11/02/2019 DATE

John Smith

John Smith SELLER

P.O.# 2312/2019 2 Court Square 3787 Pineview Drive

Due Date 26/02/2019 New York, NY 12210 Cambridge, MA 12210 Qty Description Unit Price Amount 1 Front and

rear brake cables 100.00 100.00 2 New set of pedal arms 15.00 30.00 3 Labor 3hrs 5.00 15.00 Subtotal

MONTANT_HT_ID

145.00

MONTANT_HT

Sales Tax

TVA_ID

6.25%

9.06

TVA

Total

TTC_ID

\$154.06

TTC

John Smith

Terms & Conditions Payment is due within 15 days Please make checks payable to: East Repair Inc.



Search ...

- ☐ Bird
- ☒ Not_Bird
- ☐ Unable_to_tell

SUBMIT

Submit

Skip

«

Page 3 / 11

»

Instruction

Summarize the following news article:

Article text here

Output A

Article summary

Rating (1 = worst, 7 = best)

1

2

3

4

5

6

7

Fails to follow the correct instruction / task ? ☐ Yes ☐ No

Inappropriate for customer assistant ? ☐ Yes ☐ No

Contains sexual content ☐ Yes ☐ No

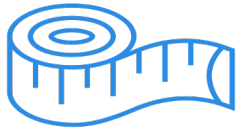
Contains violent content ☐ Yes ☐ No

Encourages or fails to discourage
violence/abuse/terrorism/self-harm ☐ Yes ☐ No

Denigrates a protected class ☐ Yes ☐ No

Gives harmful advice ? ☐ Yes ☐ No

Expresses moral judgment ☐ Yes ☐ No



Measurement:
Are the labels correct?

Finding: Design of Instrument Impacts Data

A

Tweet tweet tweet

Tweet tweet tweet

Tweet tweet tweet

Does this tweet contain hate speech?

☐ Yes

☐ No

Does this tweet contain offensive language?

☐ Yes

☐ No

B

Tweet tweet tweet

Tweet tweet tweet

Tweet tweet tweet

Does this tweet contain hate speech?

☐ Yes

☐ No

Tweet tweet tweet

Tweet tweet tweet

Tweet tweet tweet

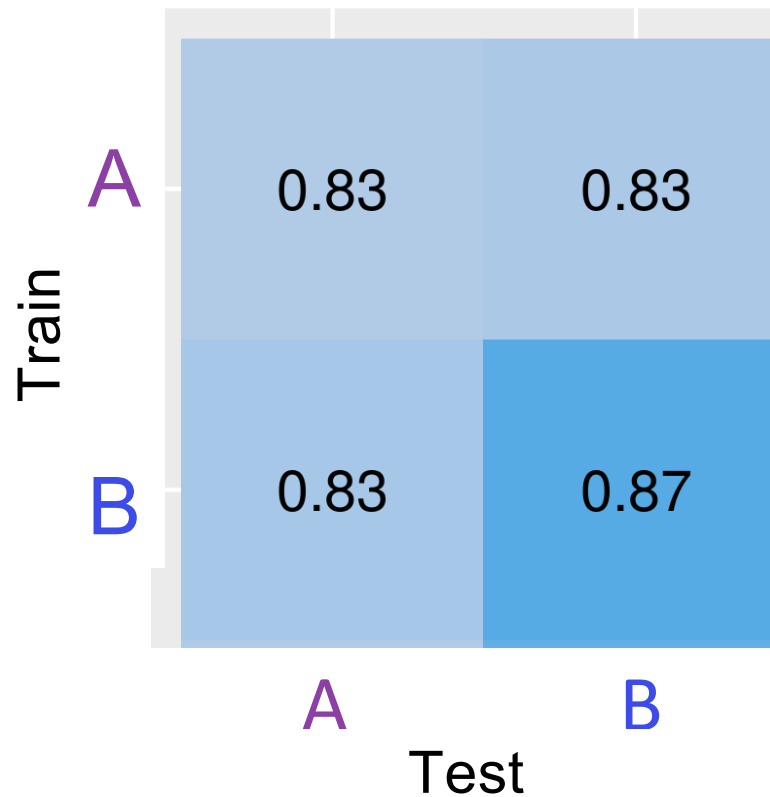
Does this tweet contain offensive language?

☐ Yes

☐ No

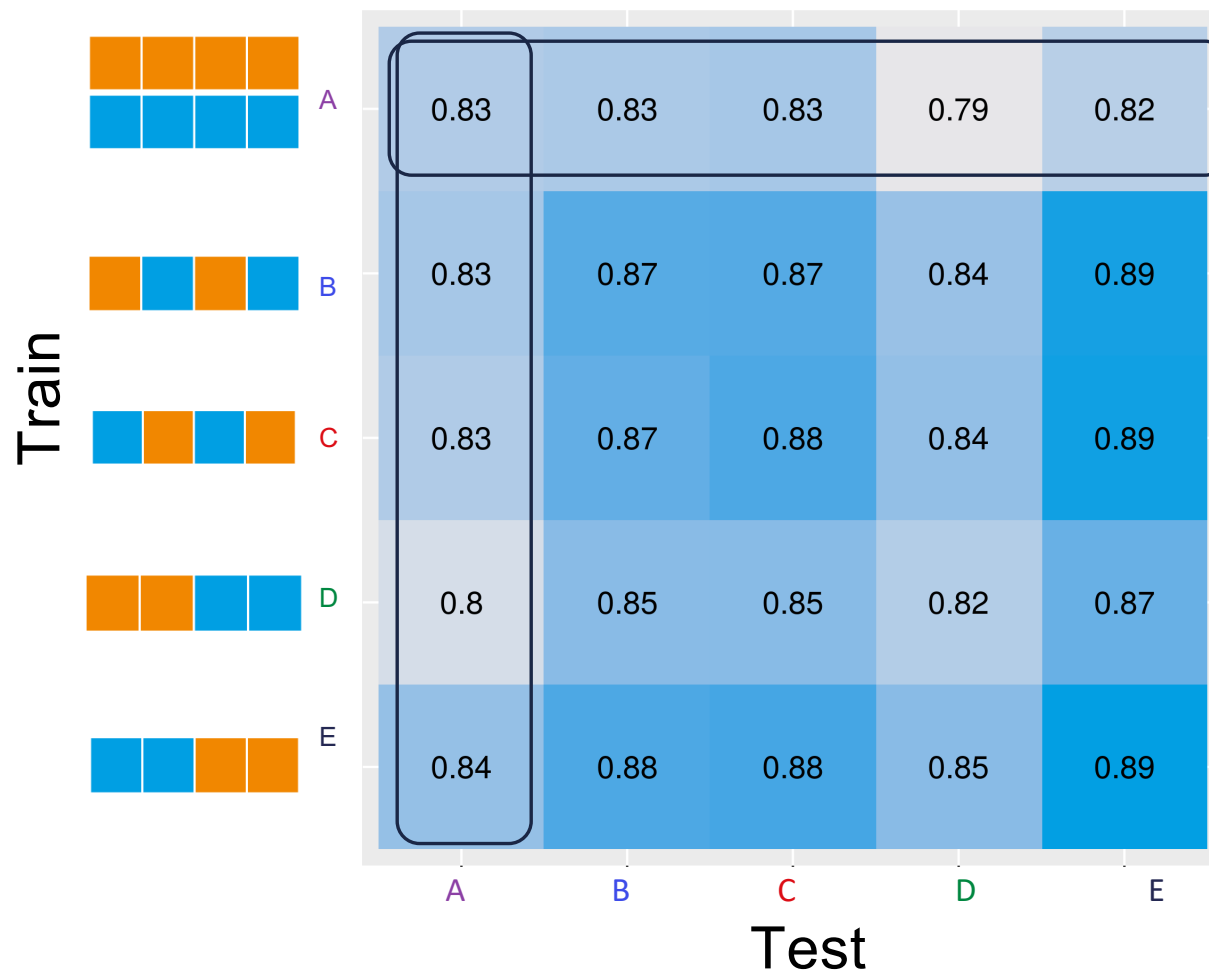
Version	Offensive Language	Hate Speech
A	52%	27%
B	59%	30%
Corr.	65%	60%

Finding: Instrument Impacts Performance



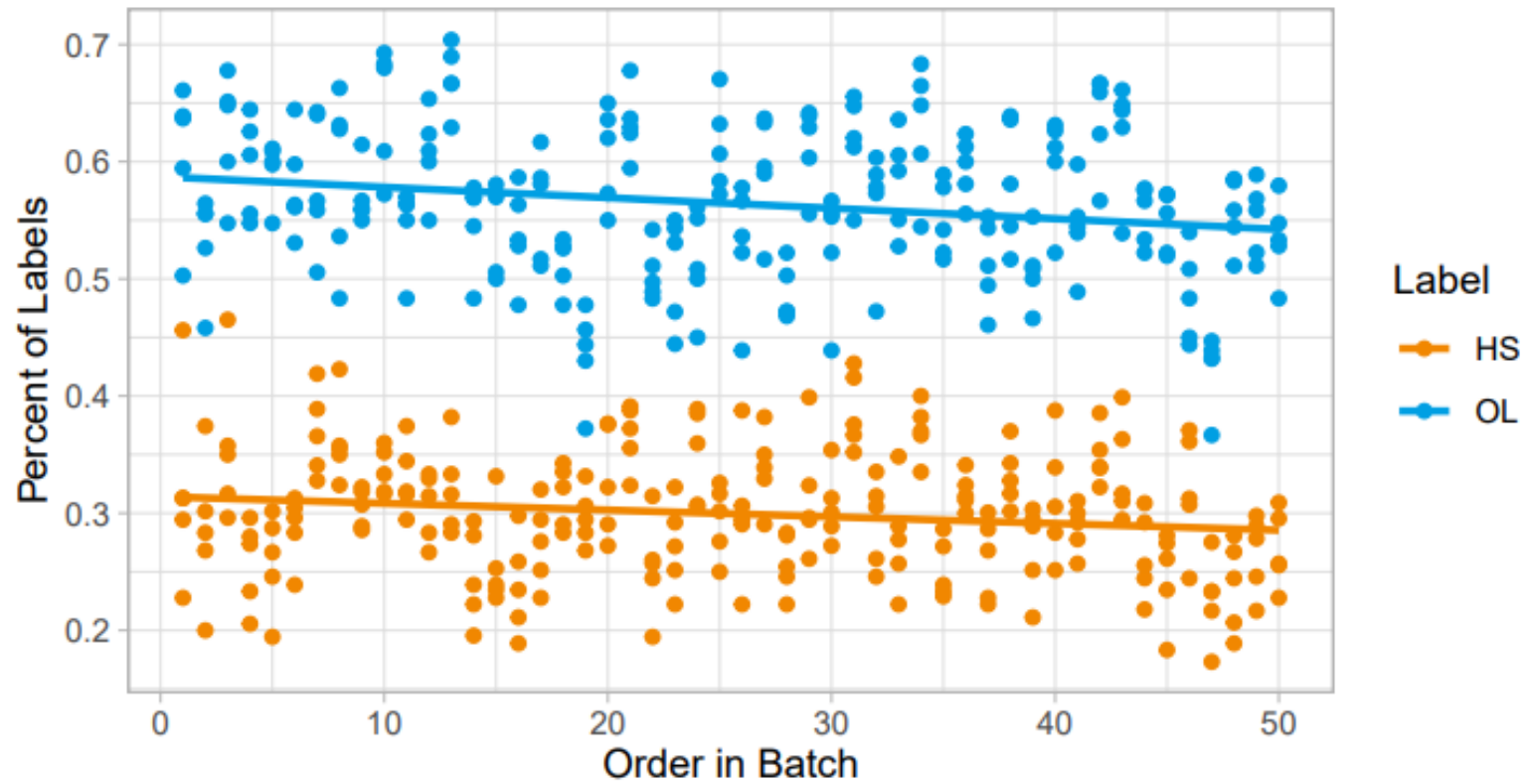
Balanced Accuracy:
Higher is better

Finding: Instrument Impacts Performance



Balanced Accuracy:
Higher is better

Finding: Order Effects



Human-AI Collaboration

Type of Emissions	2019	2020
Total	869	533
Scope 1	0	0
Scope 2	0	0
Scope 3	869	533

The 2020 Scope 3 emissions are 533, according to the AI

Is this correct?

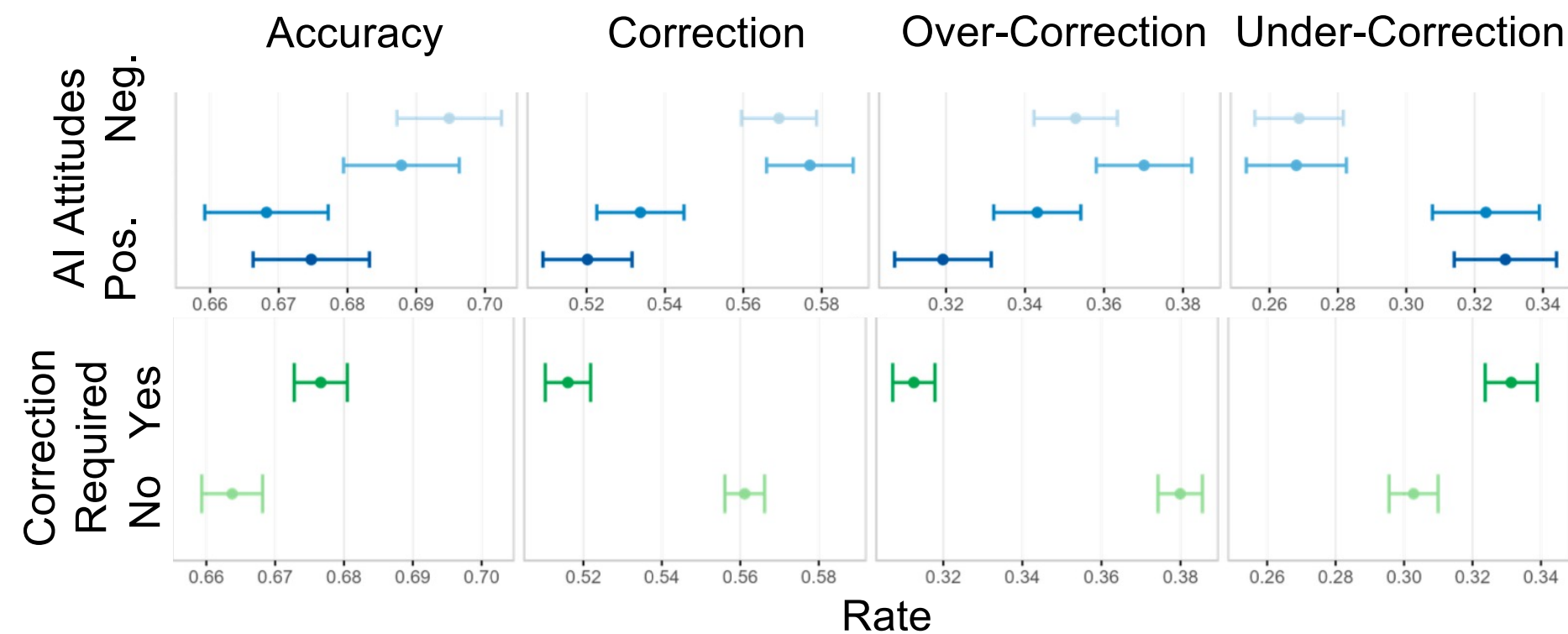
☐

Correct

☐

Incorrect

Finding: Pre-Labeling Can Lead to Biases



Findings Known in Psychology, Surveys

Context Effects: Contrast & Assimilation

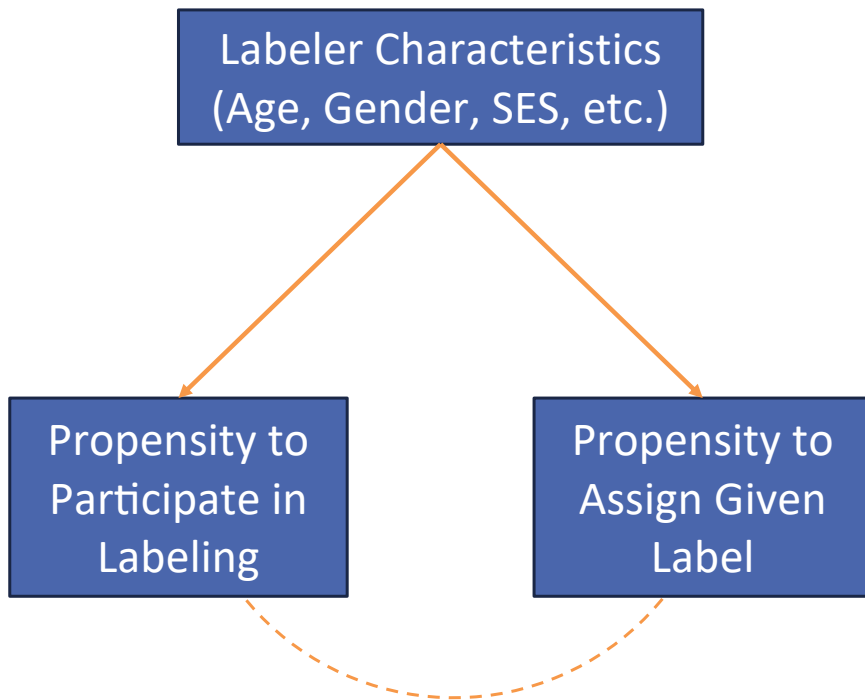
Satisficing: Reduce cognitive load

Motivated Misreporting



Representation:
Who labels?

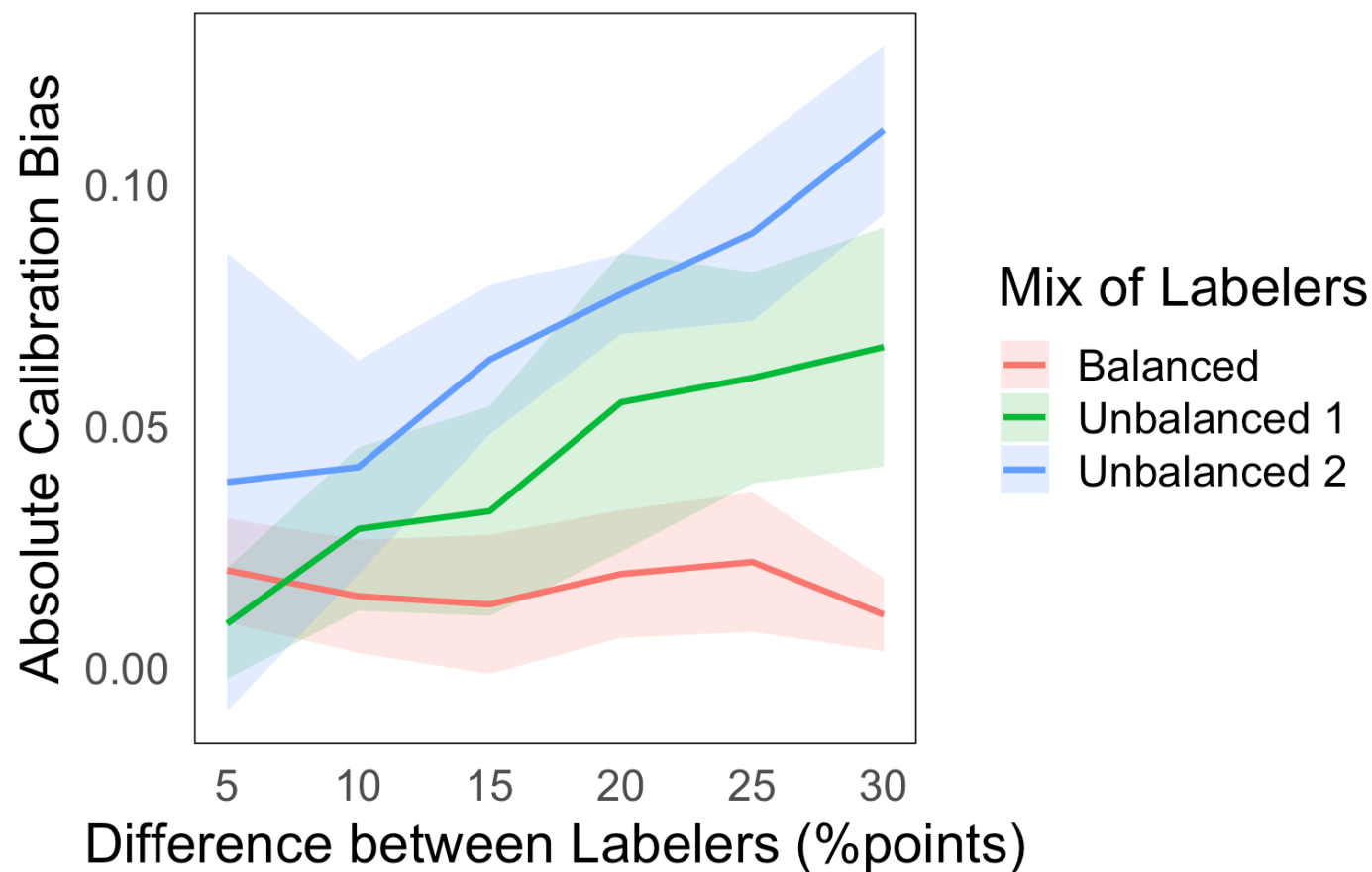
Finding: Who Labels Matters



We could:

- Recruit diverse labelers
- Train labelers to be consistent
- Use weights to adjust training data

Finding: Who Labels Matters



PAIR: Adjust Labels to Match Population

Collect Labeler
Characteristics



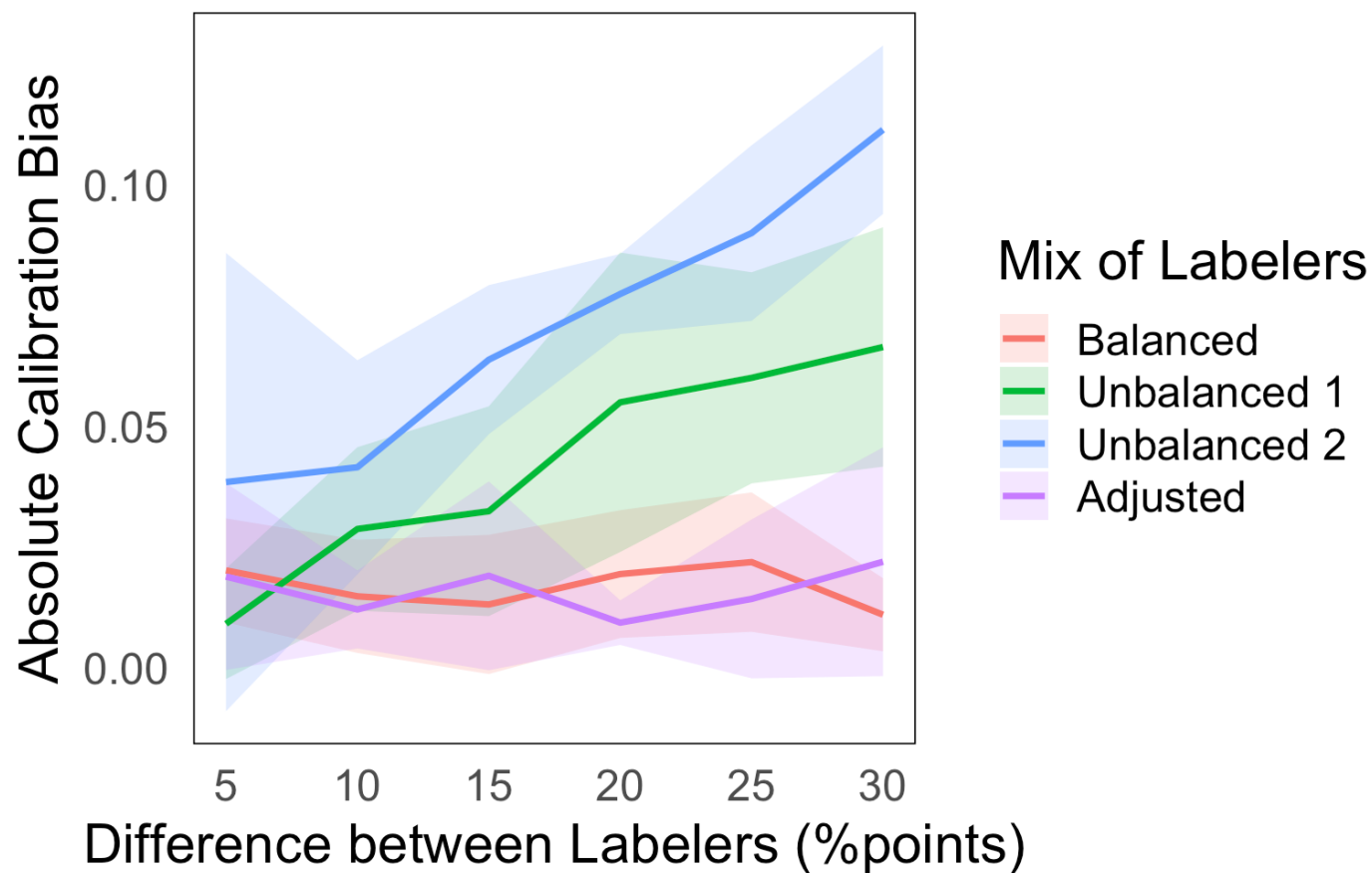
Calculate
Weights



Replicate
Instances



PAIR: Adjust Labels to Match Population



“The bias I am most nervous about is the bias of the human feedback raters”

Sam Altman
March 25 2023 “The Lex Fridman Podcast”



Practical Guidance for Label Collection



Recruit many, diverse labelers

Pay adequately



Simplify instructions: 8th grade reading level

Simplify instrument: Follow survey design practices



Test instructions and instrument

Document everything for replication and transparency

Research Group



Rob Chew
RTI International



Bolei Ma
LMU



Frauke Kreuter
LMU, UMD



Christoph Kern
LMU



Jacob Beck
LMU

Recent Papers

Eckman et al. 2024. Position: Insights from Survey Methodology can Improve Training Data for Machine Learning Models. ICML

<https://proceedings.mlr.press/v235/eckman24a.html>

Kern et al. 2023. Annotation Sensitivity: Training Data Collection Methods Affect Model Performance. EMNLP <https://aclanthology.org/2023.findings-emnlp.992/>

Beck et al. 2024. Order Effects in Annotation Tasks: Further Evidence of Annotation Sensitivity. UncertainNLP <https://aclanthology.org/2024.uncertainlp-1.8/>

Eckman et al. 2025. Aligning NLP Models with Target Population Perspectives using PAIR: Population-Aligned Instance Replication. NLPerspectives

<https://aclanthology.org/2025.nlperspectives-1.9.pdf>

Beck et al. 2025. Bias in the Loop. Working Paper <https://arxiv.org/abs/2509.08514>

Thank You

Stephanie Eckman

Researcher and Data Scientist

SODA Center, UMD

www.stepheckman.com



*An LA Love Story of Data,
Innovation, and the
Quest for Truth*

May 13 - 15, 2026
Los Angeles, CA

- Short Course: “Human-Centered Data Collection for Machine Learning: Lessons from Survey Research,” May 12, 2026
- Online Course at Joint Program in Survey Methodology, June 2026