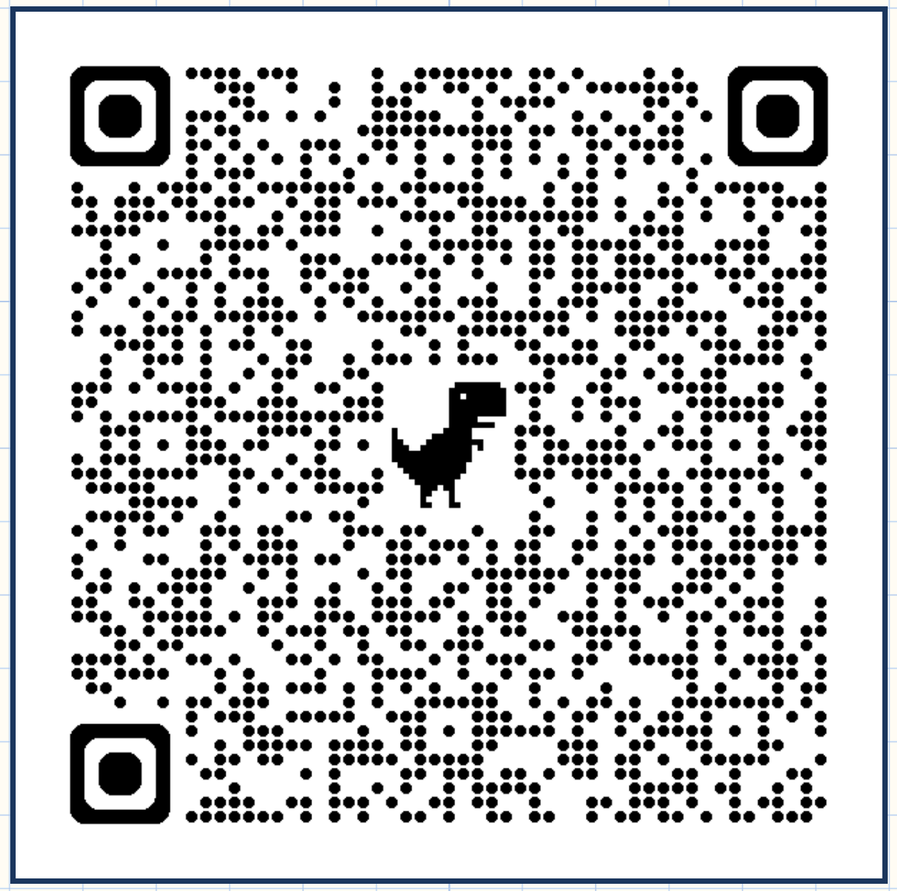
AAAS 2025

VALID STATISTICAL INFERENCE

ON AI/ML PREDICTED OUTCOMES

STEPHEN SALERNO, PHD (HE/HIM)
FRED HUTCHINSON CANCER CENTER

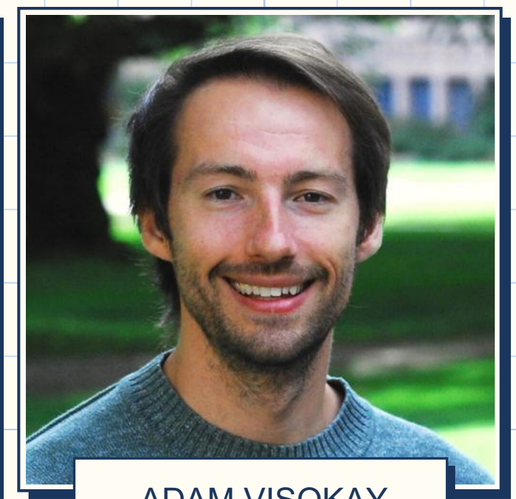
THESE SLIDES



ACKNOWLEDGEMENTS

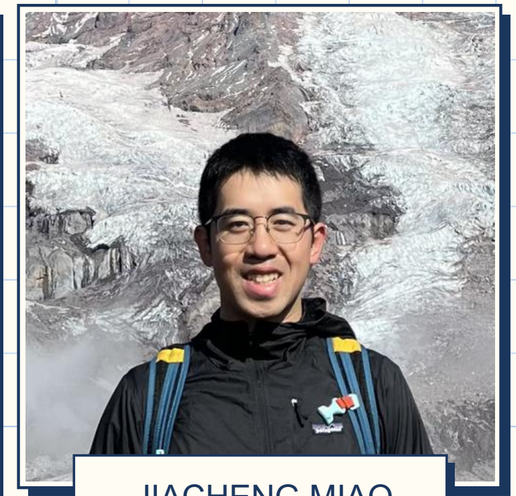
KENTARO HOFFMAN

AWAN AFIAZ



ADAM VISOKAY

SHUXIAN FAN

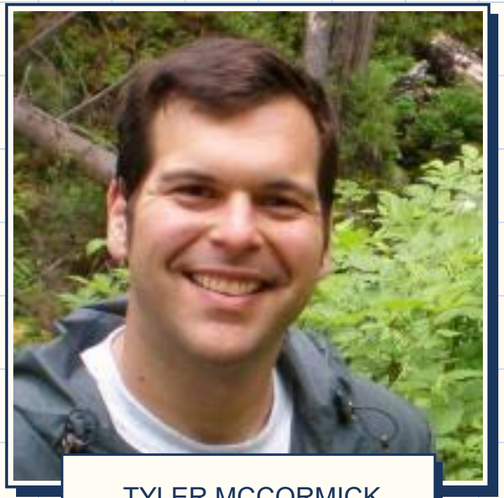


JIACHENG MIAO

ANNA NEUFELD

QIONGSHI LU

LI LIU



TYLER MCCORMICK

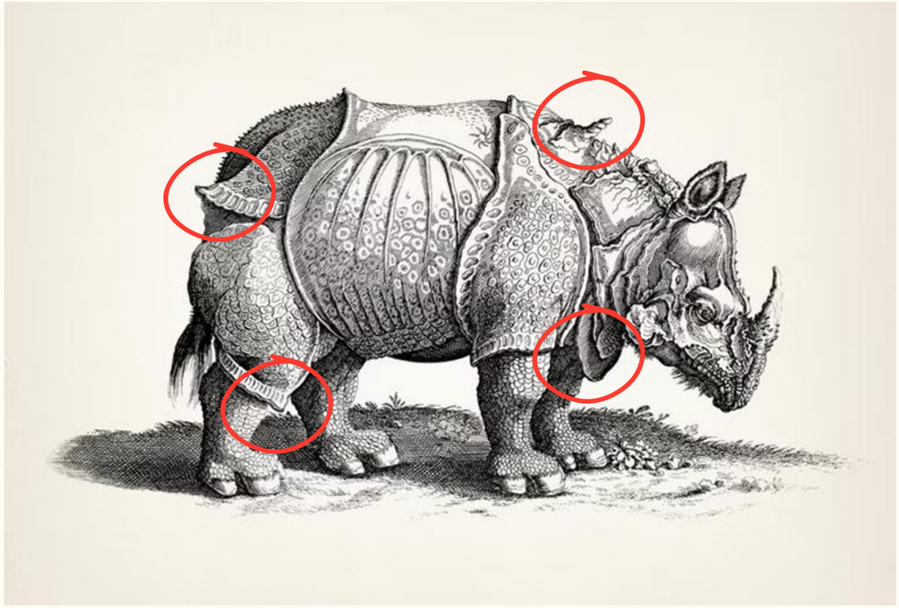
JEFF LEEK

QUESTION ...

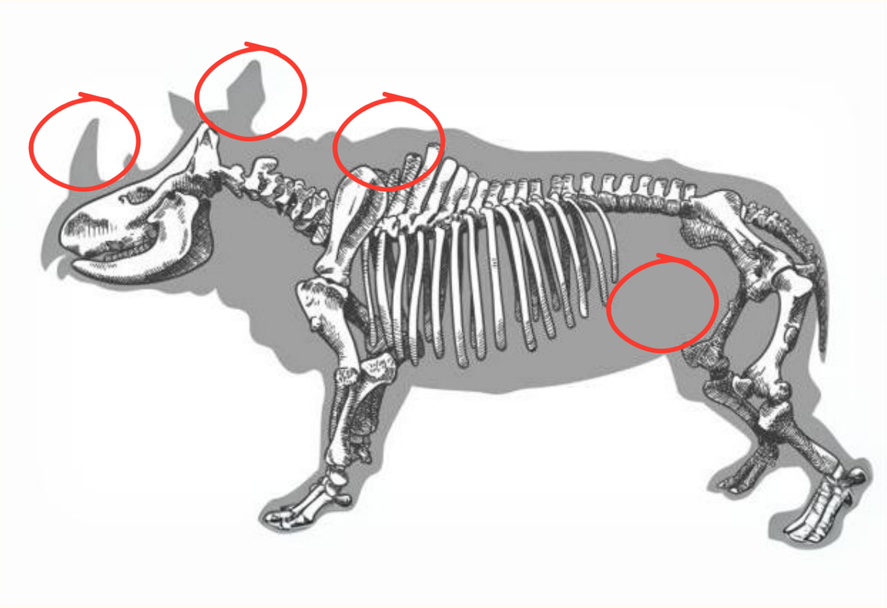
If you had never seen a rhino before, how would you draw it?

... LET'S LOOK AT TWO ATTEMPTS

ARTIST RENDERINGS BASED ON LIMITED INFORMATION



Albrecht Dürer's The Rhinoceros, *Woodcutting* (1515)



C.M. Kösemen's Rhinoceros, *All Yesterdays* (2012)

USING AI-GENERATED DATA IS APPEALING FOR FINANCIAL AND LOGISTIC REASONS

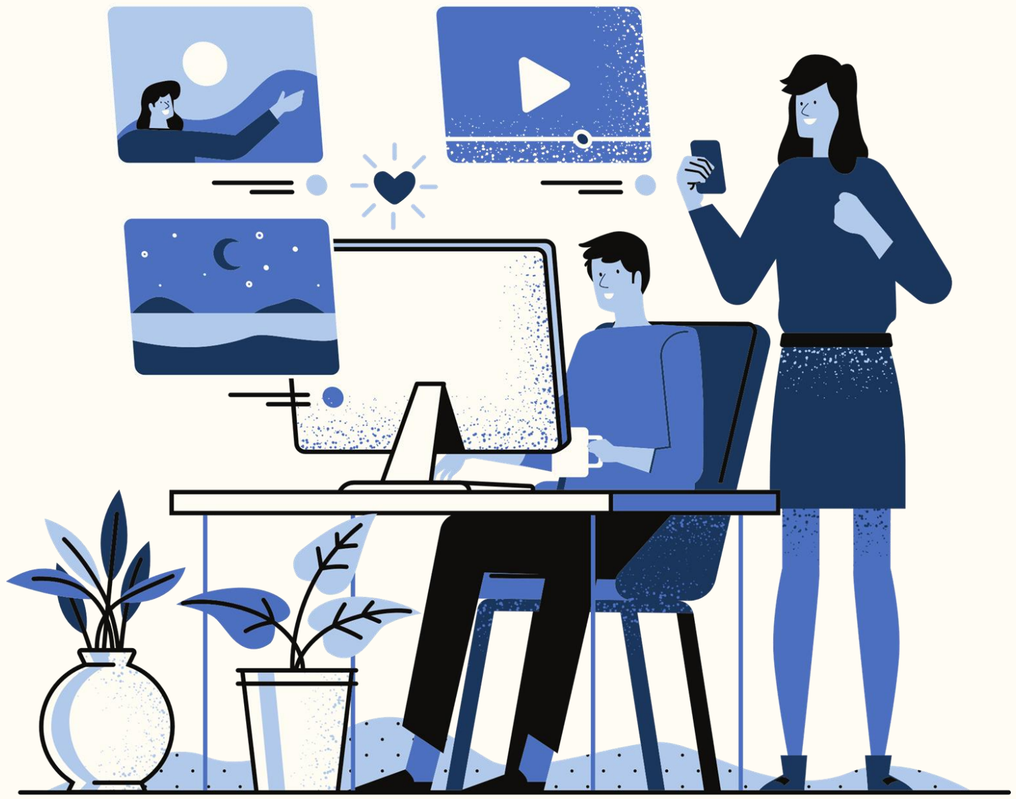
As AI becomes more accurate & accessible, the distinction between AI/ML-generated and empirical data has blurred.

EXAMPLES:

- **Genomics: Phenotyping**
- **Emergency Medicine: Sepsis Prediction**
- **Sociology: Predicting Demographics**
- **Demography: Verbal Autopsy**
- **Oncology: Automated Karyotyping**
- **Economics: Labor Market Outcomes**
- **Marketing: Digital ‘Click’ Visits**
- **Technology: Launching App Features**
- **Politics: Election Forecasting**

... & MANY MORE

HOWEVER, GOOD PREDICTIONS DO NOT MEAN VALID INFERENCE

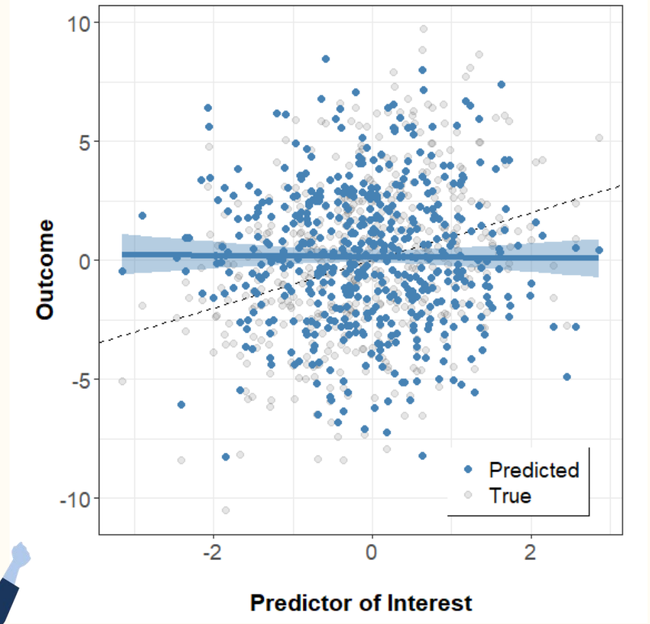


- **Algorithmic bias in the relationship between predicted outcomes and their true, unobserved counterparts**
- **Lack of robustness in the prediction model to resampling or uncertainty about the training data**
- **Inability to appropriately propagate bias & uncertainty in predictions to the downstream inferential procedure**

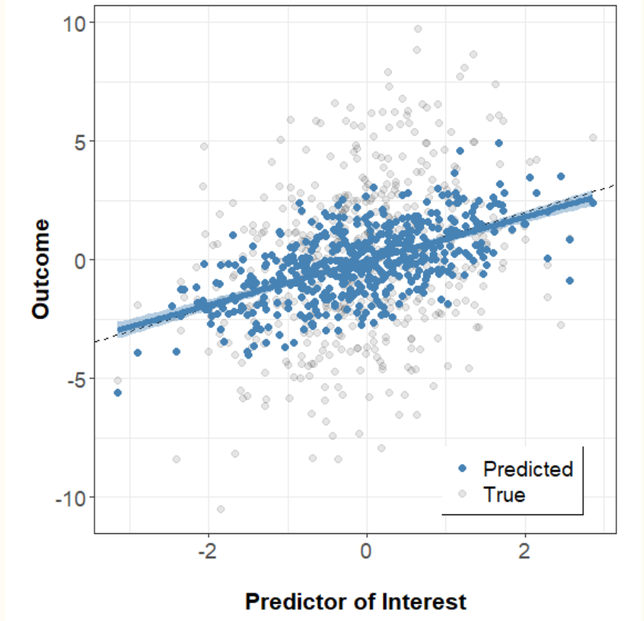
LET’S SEE....

ILLUSTRATING PROBLEMS WITH SIMULATED DATA

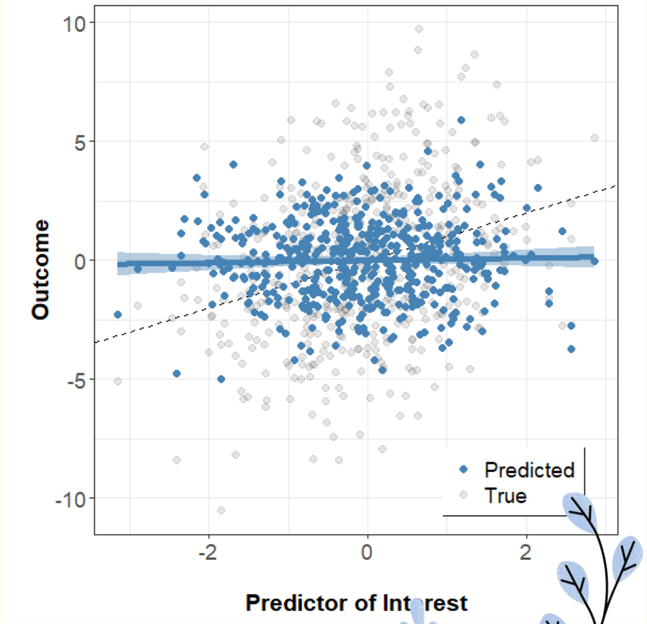
1. BIASED PREDICTIONS



2. UNDERESTIMATED VARIANCE

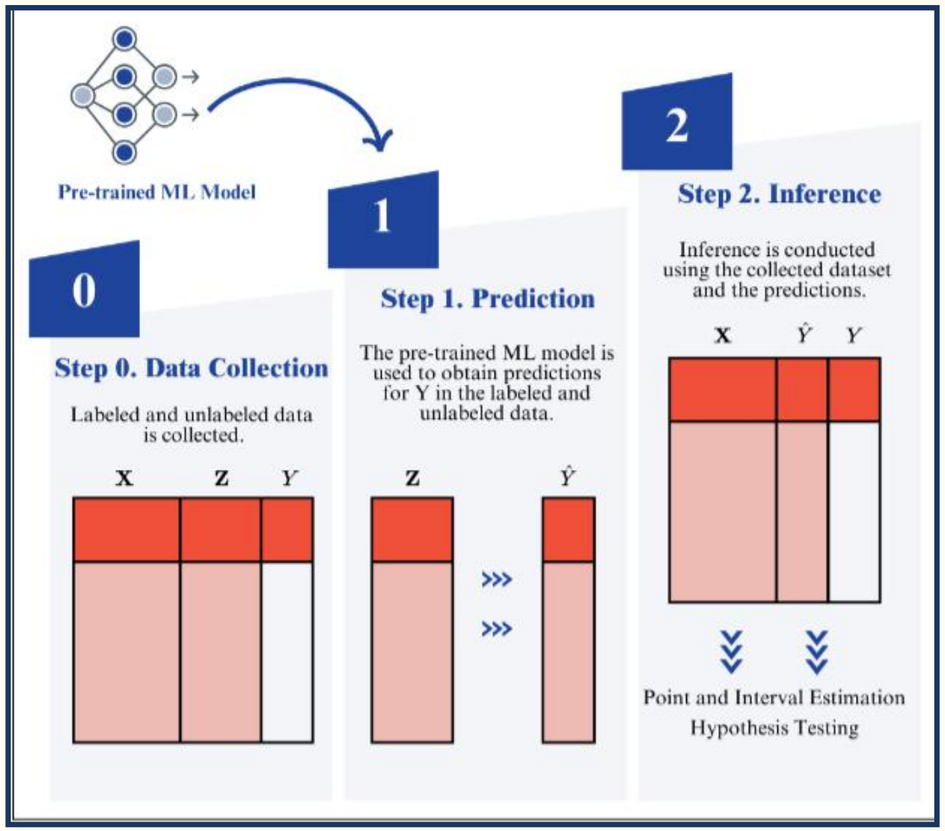


3. BOTH BIAS AND VARIANCE ISSUES



INFERENCE ON PREDICTED DATA

OVERVIEW



FROM GRONSBELL J, GAO J, SHI Y, MCCAW ZR, & CHENG D. (2024). ANOTHER LOOK AT INFERENCE AFTER PREDICTION. ARXIV PREPRINT ARXIV:2411.19908.

A promising approach is to leverage a small subset of data with the outcome and its associated features measured:

- **Correct for bias by modeling the true and predicted outcomes in the subset**
- **More accurate variance, accounting for uncertainty in the predictions**
- **Guaranteed to be unbiased and at least as efficient as using only the subset of complete data**

RECENT METHODS

AND APPLICATIONS

Methods have been developed in quick succession, including those for:

- **General Inference Problems**
- **Genome-Wide Association Studies**
- **Causal Inference**
- **Ranking Methods**
- **Federated/Decentralized Data**
- **Bayesian Estimation**

With recent methods having guaranteed performance based in statistical theory

THE IPD R PACKAGE

Implements recent IPD methods with a user-friendly wrapper function and ‘tidy’ helper functions

- Provides domain experts access to these tools for use in their work
- Enables data scientists a means to develop/compare new methods
- Will be continuously updated to include m & func

Software Note

R Package

AN EXAMPLE:
VERBAL AUTOPSY

- **Fewer than 1/3 of deaths worldwide are assigned a medically-certified cause (Horton, 2007)**
- **Involves conducting a structured interview with family and caregivers**
- **Cause of death is predicted via time-consuming, resource-intensive process**

UNPROCESSED VA TEXT NARRATIVE
Deceased started to ill while at working place, He came home while experiencing cough with chest pain, difficult in breathing, tiredness and blood vision. The after visited Belfast clinic to get treatment but no improvement. Afterwards deceased complained of stomach pain. Then after experienced diarrhea. He was given traditional medicine but did not change. Afterwards he vomiting worms and diarrhea continued. He continued using traditional medicine and the condition remains the same. Three days before death deceased sneezed a thing like a worm. He died at home and he also experienced hot body. It was examined that his chest and throat developed wounds. Treatment given but no change. His lower lip also had rash that at time chapping and a lot of blood will comes out. After treatment that lip became healed He was taken to traditional healer, but condition unchanged. He was taken Tintswalo hospital, where he was admitted Oxygen supplier was given but he finally passed away on the third day at hospital. A week before death he complained about body pain. At the beginning deceased also had cough and complained of headache during the night only throughout the illness. A month before death he experienced hiccup which continued until death but recurrent, he skips days not defecating When defecate the stool were hard then after yellowish and black few days before death. Deceased also developed ring worms on both checks but healed before death
PROCESSED VA TEXT NARRATIVE
['cough', cough',' chest',' pain',' tiredness',' blood',' vision',' stomach',' pain',' ' vomit',' worms','diarrhea',' sneezed',' worms',' hot',' chest',' throat',' ' lip',' rash',' chapping',' blood',' lip',' pain',' cough',' headache',' hiccup'," defecating',' defecate',' stool',' yellowish',' ring','worms']

Mapundu et al. 2024

AN EXAMPLE: VERBAL AUTOPSY

VA Example

METHOD

- **Gold standard verbal autopsy data: n=6,763, 6 sites, 5 CODs**
- **Extended recent method (PPI++) to multiclass regression:**

$$\mathbb{E}[\ell_\theta(X_L, Y_L)] + \lambda\left(\mathbb{E}[\ell_\theta(X_U, \hat{Y}_U^{AI})] - \mathbb{E}[l_\theta(X_L, \hat{Y}_L^{AI})]\right)$$

Regression on true COD in labeled subset

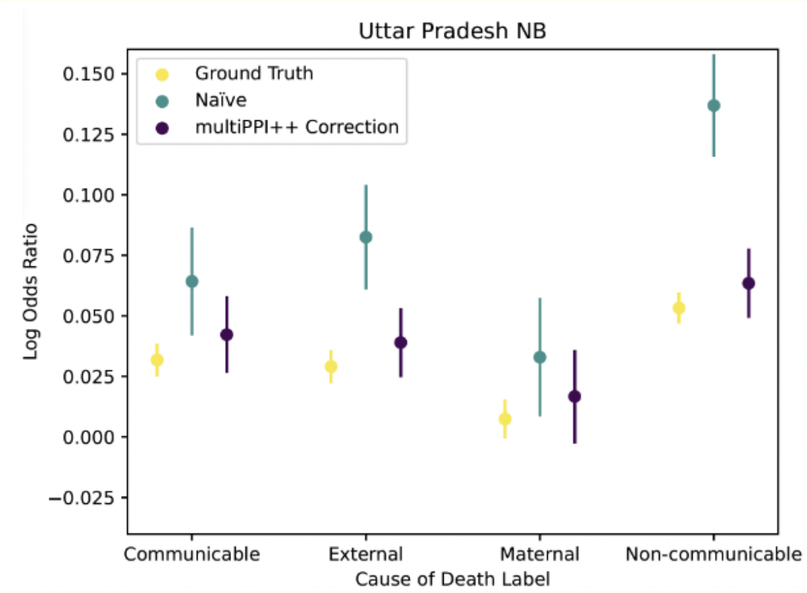
Weight contribution of predicted COD

Regression on predicted COD in unlabeled subset

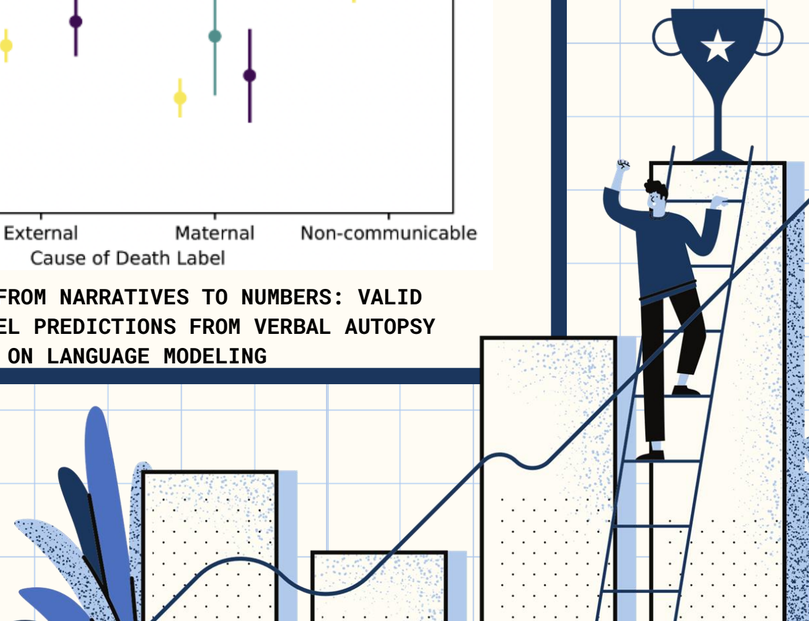
Regression on predicted COD in labeled subset

ANGELOPOULOS, A. N., BATES, S., FANNJIANG, C., JORDAN, M. I., & ZRNIC, T. (2023). PREDICTION-POWERED INFERENCE. SCIENCE, 382(6671), 669-674.

RESULTS



FROM FAN, S. ET AL. (2024). FROM NARRATIVES TO NUMBERS: VALID INFERENCE USING LANGUAGE MODEL PREDICTIONS FROM VERBAL AUTOPSY NARRATIVES. FIRST CONFERENCE ON LANGUAGE MODELING



REMEMBER THE PAST...

... OR REPEAT IT

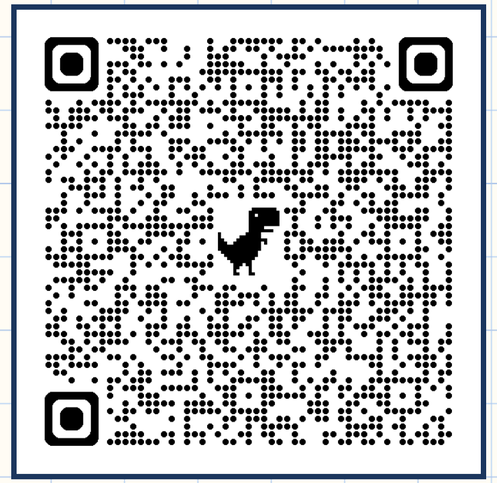
KEY TAKEAWAYS:

- **Increasing reliance on AI/ML raises questions about data quality/validity**
- **IPD is a rapidly evolving field, driven by need for rigorous methods**
- **Open-source collaboration is the fastest way to success!**

SNOPES.COM/ARTICLES/441423/MAGDEBURG-UNICORN/

QUESTIONS?

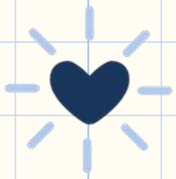
Editorial



This Talk

ssalerno@fredhutch.org

https://github.com/ipd-tools



THANK YOU!