

AI in Social Research: Surprises and Pitfalls

Zhiyu Li

Department of Government
Harvard University

February 12, 2025

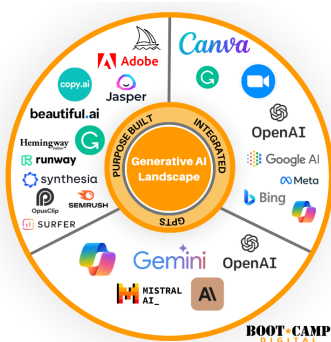
Warm-up: Survey

Warm-up: Survey

- Do you use AI in your daily life?

Warm-up: Survey

- Do you use AI in your daily life?
 - Common AI tools



Source: <https://bootcampdigital.com/blog/generative-ai-landscape-and-ecosystem/>

Warm-up: Discussion

Warm-up: Discussion

- What is Artificial Intelligence?

Warm-up: Discussion

- What is Artificial Intelligence?
- Four Approaches of Definition (Russell and Norvig 2016):
 - Acting humanly: The Turing Test approach
 - Natural language processing
 - Knowledge representation
 - Automated reasoning
 - Machine learning
 - Computer vision
 - Robotics
 - Thinking humanly: The cognitive modeling approach
 - How human minds work
 - Acting Rationally
 - Rational agent
 - Thinking Rationally
 - Logic

Warm-up: Discussion

How Can We Use AI to Study Social Phenomena?

Study 1: Mobile Data & Poverty - Motivation

Study 1: Mobile Data & Poverty - Motivation

- How can we know population characteristics in a country?
 - e.g., in China?
 - e.g., in Somalia?
 - e.g., in North Korea?

Study 1: Mobile Data & Poverty - Motivation

- How can we know population characteristics in a country?
 - e.g., in China?
 - e.g., in Somalia?
 - e.g., in North Korea?
- **Title:** Predicting Poverty and Wealth From Mobile Phone Metadata

Study 1: Mobile Data & Poverty - Motivation

- How can we know population characteristics in a country?
 - e.g., in China?
 - e.g., in Somalia?
 - e.g., in North Korea?
- **Title:** Predicting Poverty and Wealth From Mobile Phone Metadata
- **Year:** 2015

Study 1: Mobile Data & Poverty - Motivation

- How can we know population characteristics in a country?
 - e.g., in China?
 - e.g., in Somalia?
 - e.g., in North Korea?
- **Title:** Predicting Poverty and Wealth From Mobile Phone Metadata
- **Year:** 2015
- **Author(s):** Joshua Blumenstock, Gabriel Cadamuro, and Robert On

Study 1: Mobile Data & Poverty - Motivation

- How can we know population characteristics in a country?
 - e.g., in China?
 - e.g., in Somalia?
 - e.g., in North Korea?
- **Title:** Predicting Poverty and Wealth From Mobile Phone Metadata
- **Year:** 2015
- **Author(s):** Joshua Blumenstock, Gabriel Cadamuro, and Robert On
- **Citation:** Blumenstock, J., Cadamuro, G. and On, R., 2015. Predicting poverty and wealth from mobile phone metadata. Science, 350(6264), pp.1073-1076.

Data in Rwanda

Summary statistic	Phone survey	Call detail records	DHS(2007)	DHS(2010)
Number of unique individuals	856	1.5 million	7377	12,792
Data collection period	Jul-09	May 2008–May 2009	Dec. 2007–Apr. 2008	Sept. 2010–Mar. 2011
Number of questions in survey	75	N/A	1615	3396
Primary geographic units	30 districts	30 districts	30 districts	30 districts
Secondary geographic units	300 cell towers	300 cell towers	247 clusters	492 clusters

Figure: Summary statistics for primary data sets.

Result 1

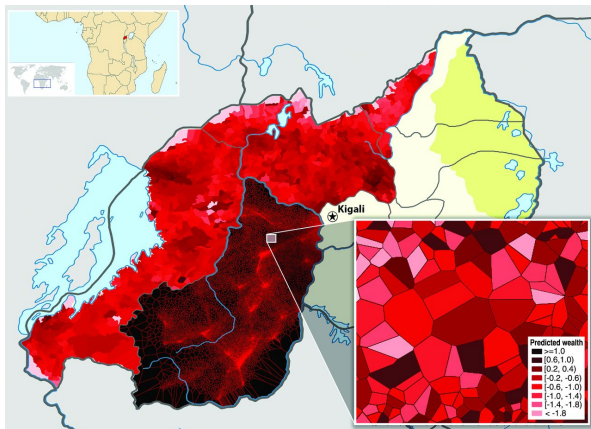
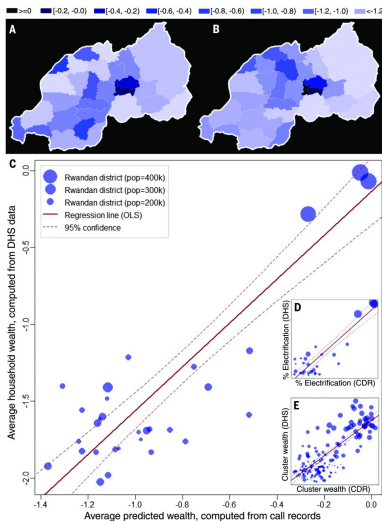


Figure: Construction of high-resolution maps of poverty and wealth from call records.

Result 2



Study 2: Analyzing the Meanings of Class

- Is the meaning of “automobile” the same today as it was 100 years ago?

Study 2: Analyzing the Meanings of Class

- Is the meaning of “automobile” the same today as it was 100 years ago?
 - university?

Study 2: Analyzing the Meanings of Class

- Is the meaning of “automobile” the same today as it was 100 years ago?
 - university?
 - cloud?

Study 2: Analyzing the Meanings of Class

- Is the meaning of “automobile” the same today as it was 100 years ago?
 - university?
 - cloud?
- **Title:** The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings

Study 2: Analyzing the Meanings of Class

- Is the meaning of “automobile” the same today as it was 100 years ago?
 - university?
 - cloud?
- **Title:** The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings
- **Year:** 2019

Study 2: Analyzing the Meanings of Class

- Is the meaning of “automobile” the same today as it was 100 years ago?
 - university?
 - cloud?
- **Title:** The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings
- **Year:** 2019
- **Author(s):** Austin C. Kozlowski, Matt Taddy, and James A. Evans

Study 2: Analyzing the Meanings of Class

- Is the meaning of “automobile” the same today as it was 100 years ago?
 - university?
 - cloud?
- **Title:** The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings
- **Year:** 2019
- **Author(s):** Austin C. Kozlowski, Matt Taddy, and James A. Evans
- **Citation:** Kozlowski, A.C., Taddy, M. and Evans, J.A., 2019. The geometry of culture: Analyzing the meanings of class through word embeddings. American Sociological Review, 84(5), pp.905-949.

Word Embedding & Semantic Space

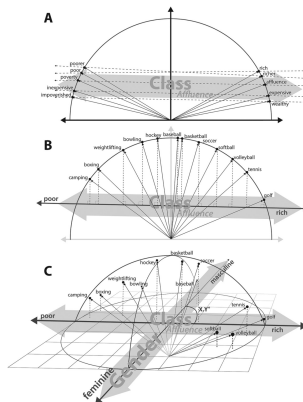


Figure: Conceptual Diagram of Semantic Space

Result 1

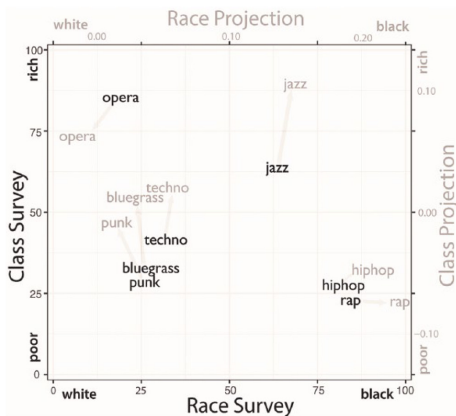


Figure: Projection of Music Genres onto Race and Class Dimensions of the Google News Word Embedding (Gray) and Average Survey Ratings for Race and Class Associations (Black)

Result 2

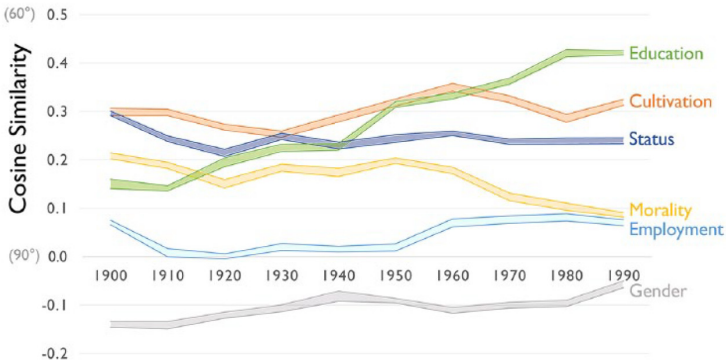


Figure: Cosine Similarity between the Affluence Dimension and Six Other Cultural Dimensions of Class by Decade

Result 3

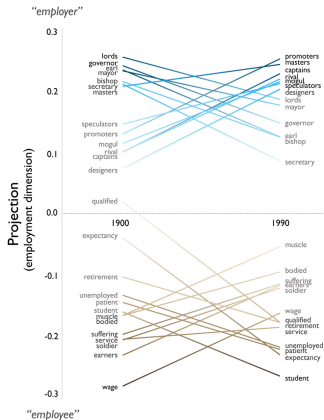


Figure: Words That Project High and Low on the Employment Dimension of Word Embedding Models

Study 3: Measuring the predictability of life outcomes

- Can you imagine yourself being where you are now, but 6 years ago? 9 years ago?

Study 3: Measuring the predictability of life outcomes

- Can you imagine yourself being where you are now, but 6 years ago? 9 years ago?
- **Title:** Measuring the predictability of life outcomes with a scientific mass collaboration

Study 3: Measuring the predictability of life outcomes

- Can you imagine yourself being where you are now, but 6 years ago? 9 years ago?
- **Title:** Measuring the predictability of life outcomes with a scientific mass collaboration
- **Year:** 2020

Study 3: Measuring the predictability of life outcomes

- Can you imagine yourself being where you are now, but 6 years ago? 9 years ago?
- **Title:** Measuring the predictability of life outcomes with a scientific mass collaboration
- **Year:** 2020
- **Author(s):** Matthew J. Salganik et al.

Study 3: Measuring the predictability of life outcomes

- Can you imagine yourself being where you are now, but 6 years ago? 9 years ago?
- **Title:** Measuring the predictability of life outcomes with a scientific mass collaboration
- **Year:** 2020
- **Author(s):** Matthew J. Salganik et al.
- **Citation:** Salganik, M.J., Lundberg, I., Kindel, A.T., Ahearn, C.E., Al-Ghoneim, K., Almaatouq, A., Altschul, D.M., Brand, J.E., Carnegie, N.B., Compton, R.J. and Datta, D., 2020. Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, 117(15), pp.8398-8403.

Data

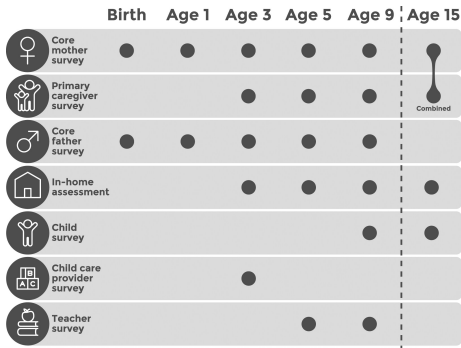


Figure: Data collection modules in the Fragile Families study

Data

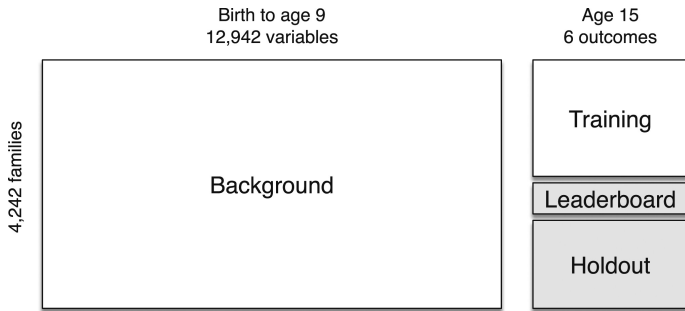


Figure: Datasets in the Fragile Families Challenge.

Result

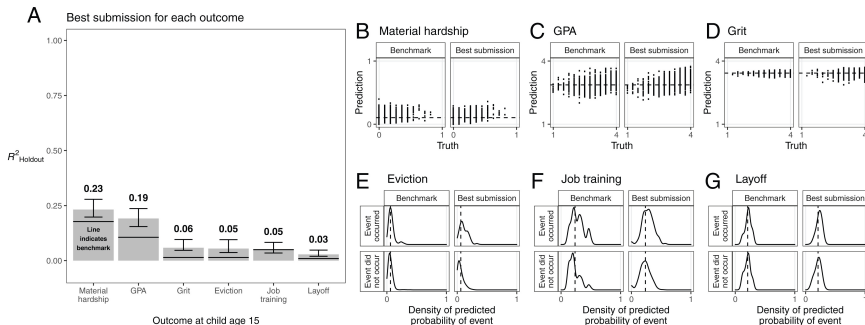


Figure: Performance in the holdout data of the best submissions and a four variable benchmark model.

Surprises and Pitfalls?

What Can AI Do and What Can It Not Do?

More Studies

- **Music:** Nie, K., 2023. Bowing to five pecks of rice: how online monetization programs shape artistic novelty. Chinese Sociological Review, 55(1), pp.66-95.
- **Satellite Imagery:** Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B. and Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. Science, 353(6301), pp.790-794.
- **Diffussion Pattern:** Goel, S., Anderson, A., Hofman, J. and Watts, D.J., 2016. The structural virality of online diffusion. Management science, 62(1), pp.180-196.
- **Agent:** Park, J.S., O'Brien, J., Cai, C.J., Morris, M.R., Liang, P. and Bernstein, M.S., 2023, October. Generative agents: Interactive simulacra of human behavior. UIST (pp. 1-22). [\[URL\]](#)

Q&A

Welcome