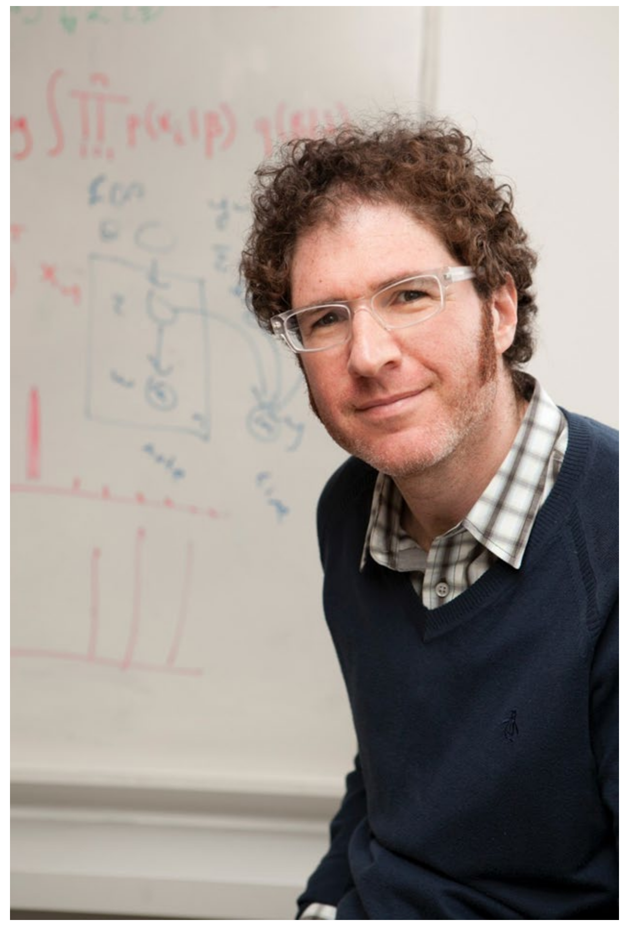
# Learning a Low-Dimensional Representation of Job History for Economic Prediction

**Keyon Vafa**
Columbia University

# Collaborators:



**Emil Palikot**
Stanford University

**Tianyu Du**
Stanford University

**Ayush Kanodia**
Stanford University

**Susan Athey**
Stanford University

**David Blei**
Columbia University

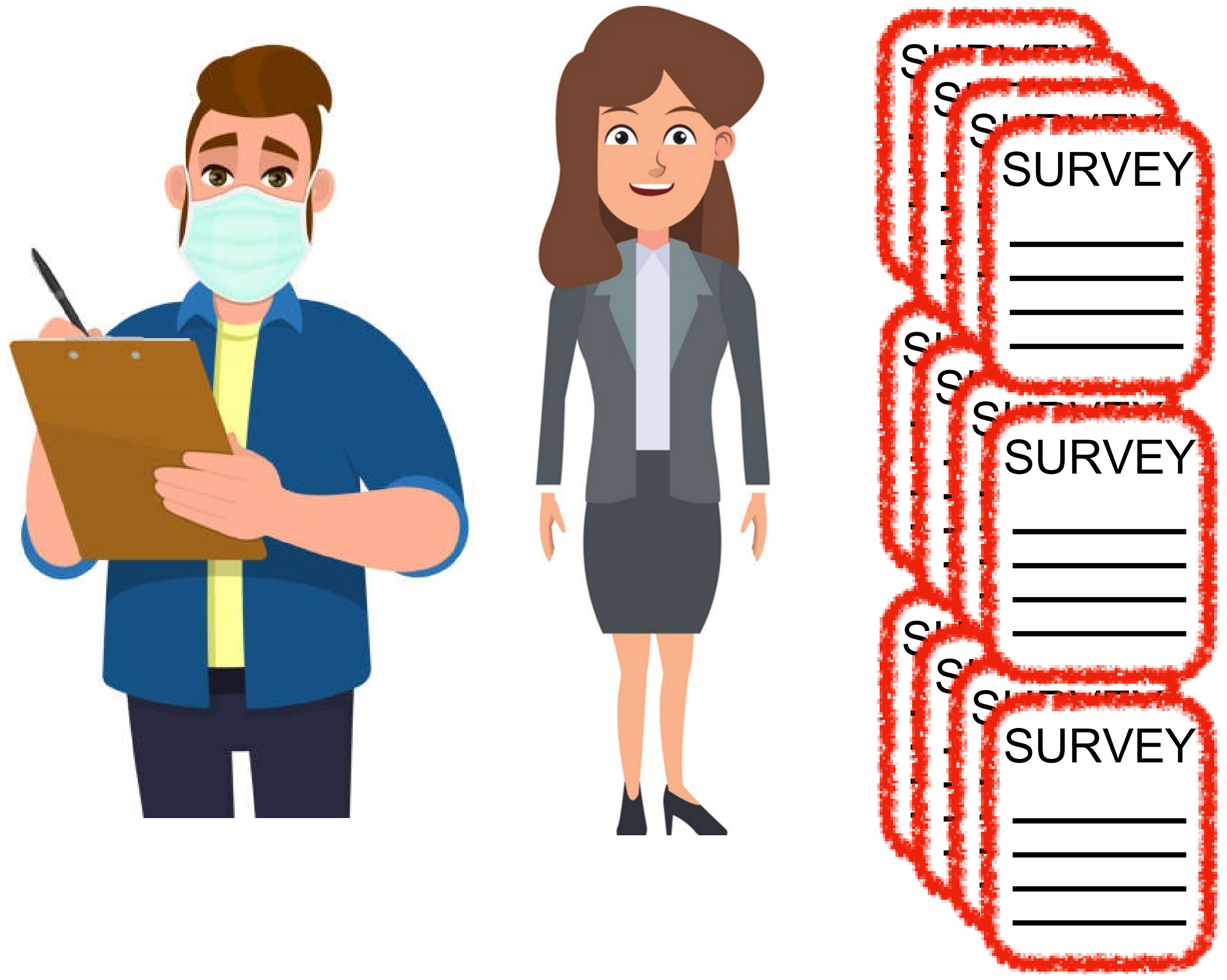
# Longitudinal Survey Datasets

1990


1991


…

2022




…

x thousands of respondents

Survey questions include:

- What is your **occupation**?
- What is your most recent **educational degree**?
- What is your **wage**?

# Longitudinal Survey Datasets

Longitudinal surveys are constructed to be **nationally representative**.

Survey datasets in the United States: **NLSY** and **PSID**

National Longitudinal Survey of Youth | 1979

NLSY79

America's Family Tree since 1968

PSID

# Applications using Longitudinal Survey Datasets

Consumption Inequality over the Last Half Century:
Some Evidence Using the New PSID Consumption Measure[†]

By Orazio Attanasio and Luigi Pistaferri*

**Birth Order, Educational Attainment, and Earnings**
An Investigation Using the PSID

**Jasmin Kantarevic**
**Stéphane Mechoulan**

**Boys with high body masses have an increased risk of developing asthma: findings from the National Longitudinal Survey of Youth (NLSY)**

D M Mannino, J Mott, J M Ferdinands, C A Camargo Jr, M Friedman, H M Greves & S C Redd

**Rising College Expectations Among Youth in the United States**
A Comparison of the 1979 and 1997 NLSY

John R. Reynolds
Jennifer Pemberton

**THE EFFECT OF THE SEX COMPOSITION OF JOBS ON STARTING WAGES IN AN ORGANIZATION: FINDINGS FROM THE NLSY***

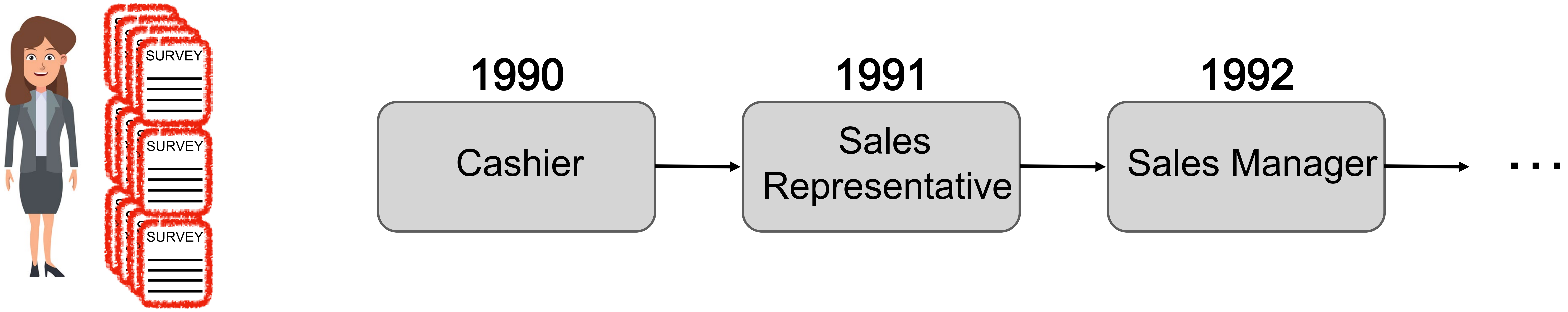
PAULA ENGLAND, LORI L. REID, AND BARBARA STANEK KILBOURNE

An empirical analysis of earnings dynamics among men in the PSID: 1968–1989

John Geweke [a, b], Michael Keane [b, c]

# Job Sequences

Over time, these datasets produce job sequences of individual workers.




Occupation modeling: Given an individual's job history, what is the distribution over their next job?

Useful for unemployment analyses (Hall, 1972), measuring occupational mobility (Kambourov and Manovskii, 2008), etc.

# Predicting Future Jobs

More accurate job predictions => more accurate economic analyses

Major challenge of building predictive models from survey datasets: they are **small** (containing only thousands of workers).

In practice, economists fit simple **linear** models that are either **Markov** or depend on history via hand-constructed summary statistics.

**Challenge 1:** Are there better models that can capture complex career trajectories?

# Resume Datasets

Recently: large resume datasets have become available.


**RESUME**

2015-2016: Data Scientist

2017: Software Engineer

These contain many job sequences:

| 2015 | 2016 | 2017 | |
|---|---|---|---|
| Data Scientist → | Data Scientist → | Software Engineer → | … |

x millions

# Suitability for Job Sequence Models


Pros: Very large datasets

Cons: Not representative, may be inaccurate (imputation errors and false information)


Pros: Representative of public and carefully curated

Cons: Small datasets

**Challenge 2:** Is there any way to leverage abundance of resumes for survey dataset analyses?

# Challenges for Modeling Job Sequences

**Challenge 1: Are there better models (beyond Markov and linear) that can capture complex long-term career trajectories?**

Challenge 2:Can we take advantage of job sequence data from resumes when analyzing survey datasets?

# Challenge 1: Modeling Complex Careers

Modeling sequences of jobs in a career is not so different from modeling sequences of words in a sentence.

**Transformer neural networks** have successfully modeled sequences of words in natural language processing (NLP).

The → dog → was → barking → and → ???

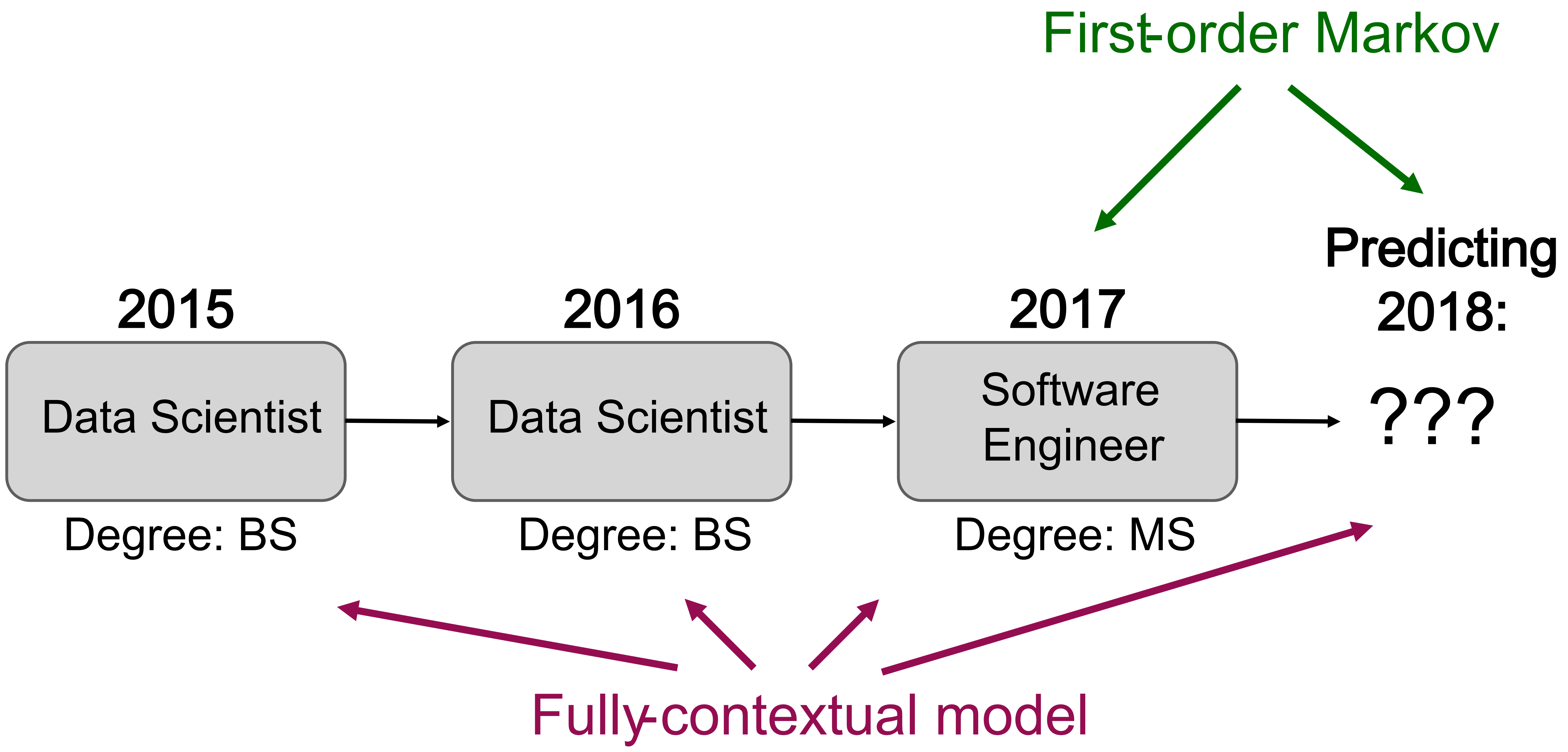
Transformer predicts: ???

We develop a transformer for modeling sequences of jobs in a career.

| 2010 | 2011 | 2012 | 2013 | 2014 | Transformer predicts: |
|---|---|---|---|---|---|
| Tutor → | Paralegal → | Judicial Clerk → | Judicial Clerk → | Lawyer → | ??? |
| Degree: BA | Degree: BA | Degree: JD | Degree: JD | Degree: JD | |

# Conditioning on Full History

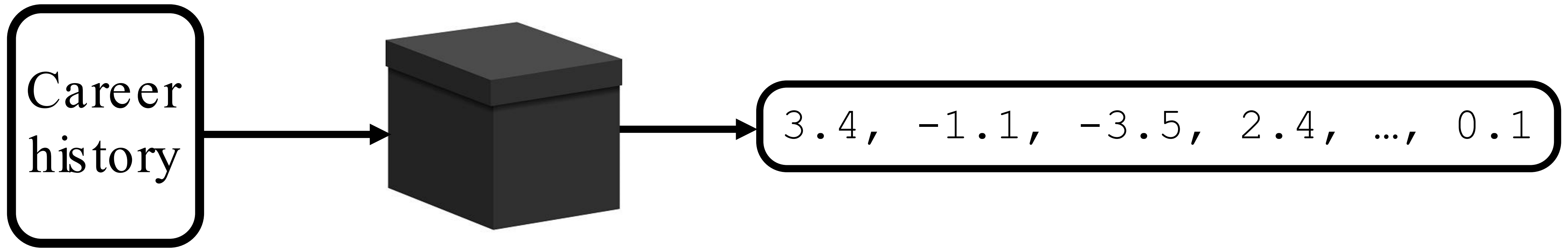
Common occupation models are Markov; ideally a model would **condition on full job history** to predict future jobs.



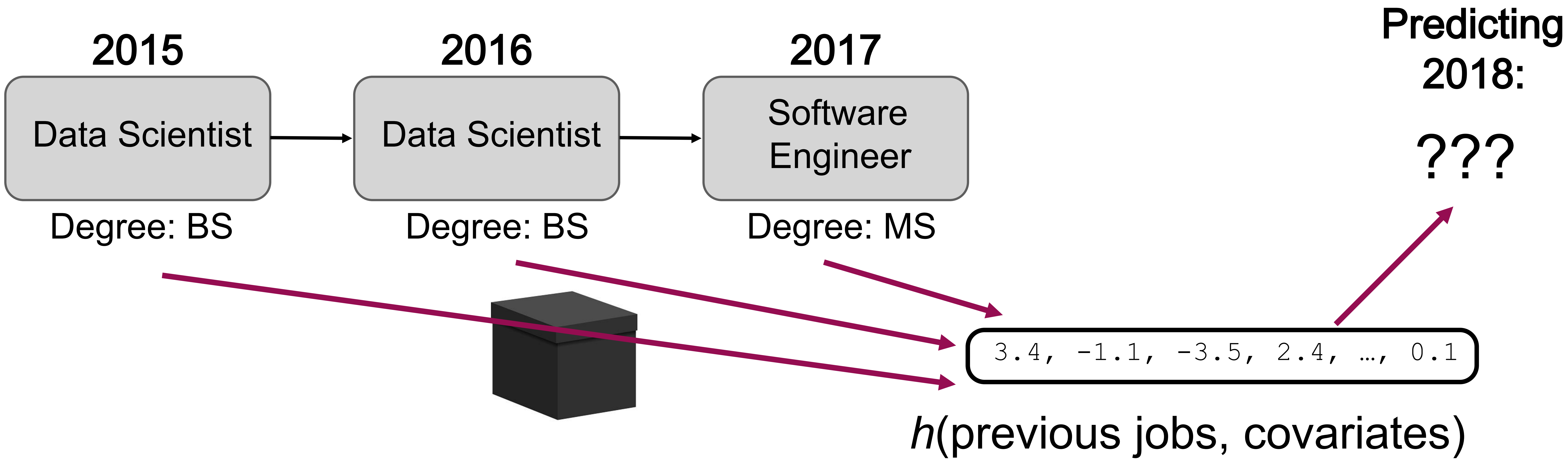

Challenge: Infeasible to condition on full history directly (exponentially many possible combinations and permutations).

# Transformers are Based on Representations

Idea: Learn **low-dimensional representation** of job history.




Summarize previous jobs and covariates with a low-dimensional representation that carries all relevant information for predicting the next job.

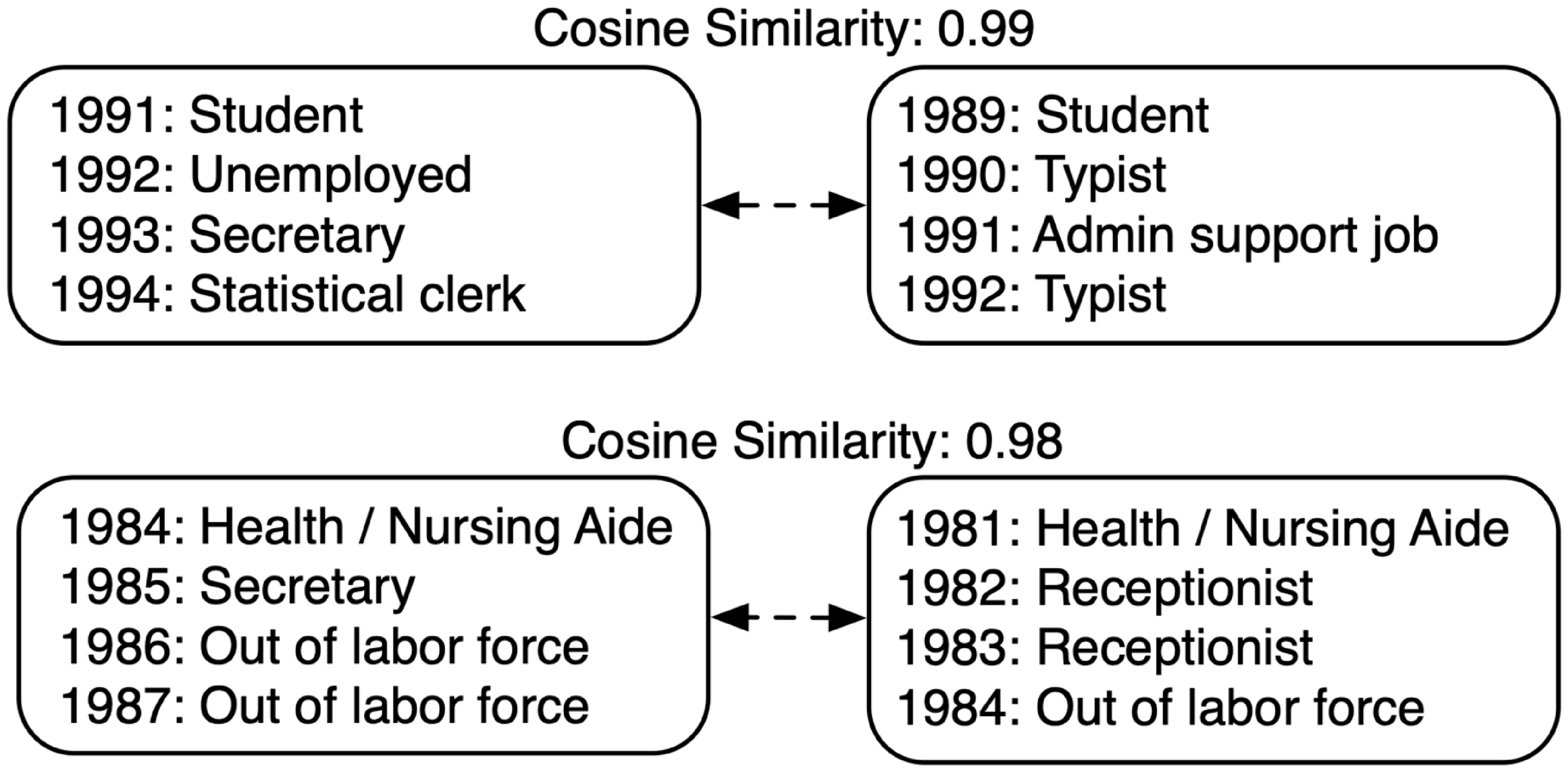


# Representations Encode Career Similarity

Intuition: If two individuals have similar but non-identical career paths, their predicted next jobs should still be similar.

Predicting jobs based on representations allows for sharing information.

Similarities of career histories based on learned representations:



# Challenges for Modeling Job Sequences

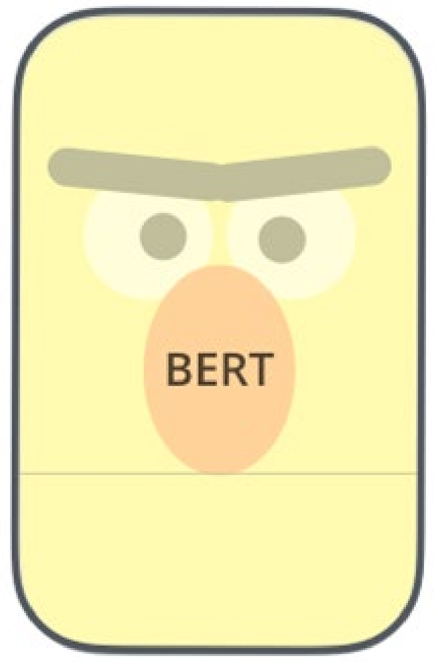
Challenge 1: Are there better models (beyond Markov and linear) that can capture complex long-term career trajectories?

**Challenge 2: Can we take advantage of job sequence data from resumes when analyzing survey datasets?**

# Transfer Learning

Modern methods in natural language processing (NLP) rely on *transfer learning*.

These methods first train a model on large, unlabelled text corpora (e.g. Google books, Wikipedia) before adjusting these models on tasks of interest involving small amounts of labelled data (e.g. sentiment analysis).




Intuition: models learn about underlying grammar/vocabulary on large corpora before specializing on a particular task.

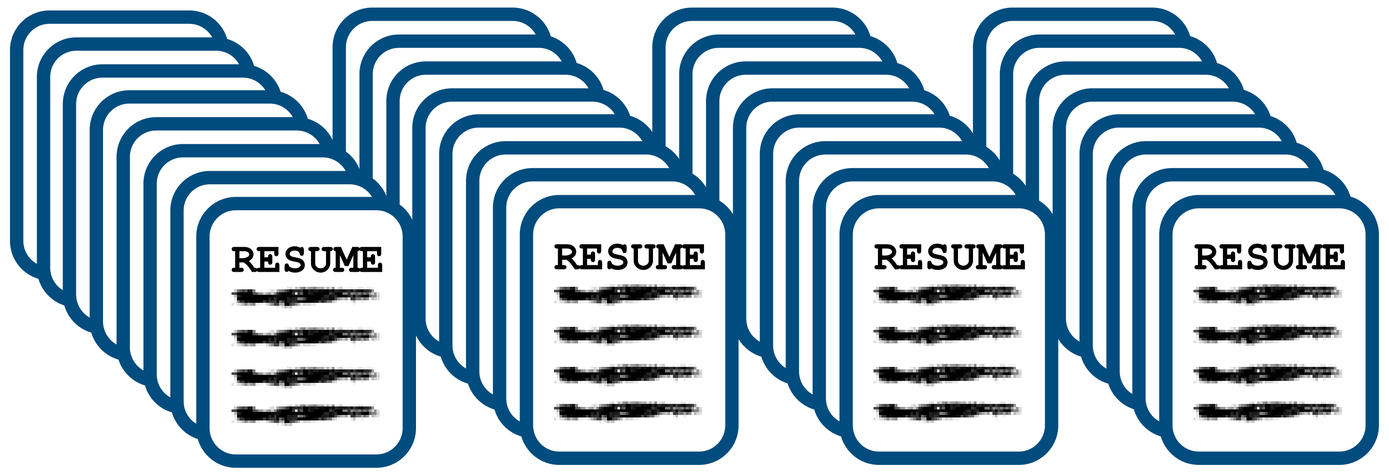
# Challenge 2: Leveraging Resumes for Surveys

**Transfer learning:** Learning representations of careers from resumes that can be adjusted (or finetuned) on survey datasets.

Step 1: Learn representations of careers from resumes:

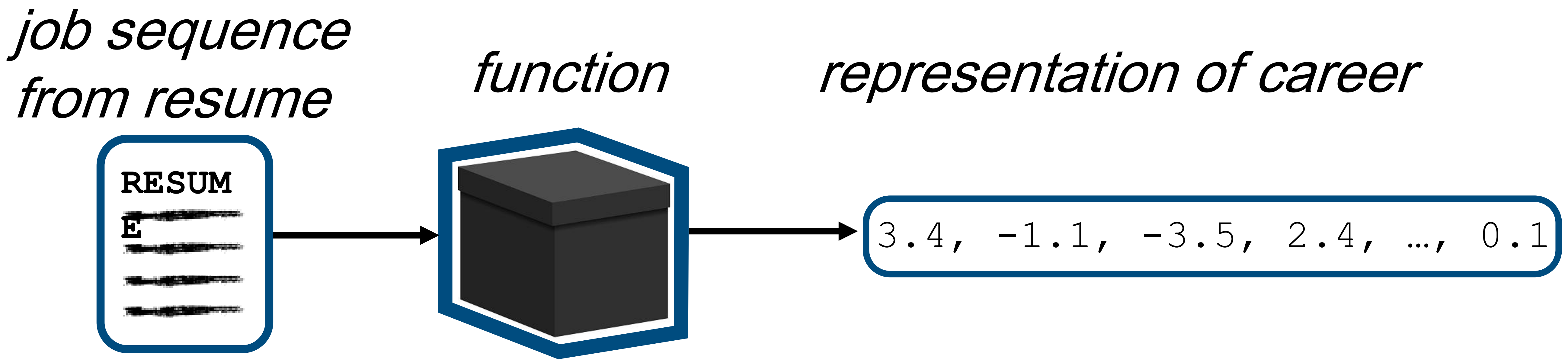
**Input:**

Millions of resumes




**Output:**

Representation function of resumes



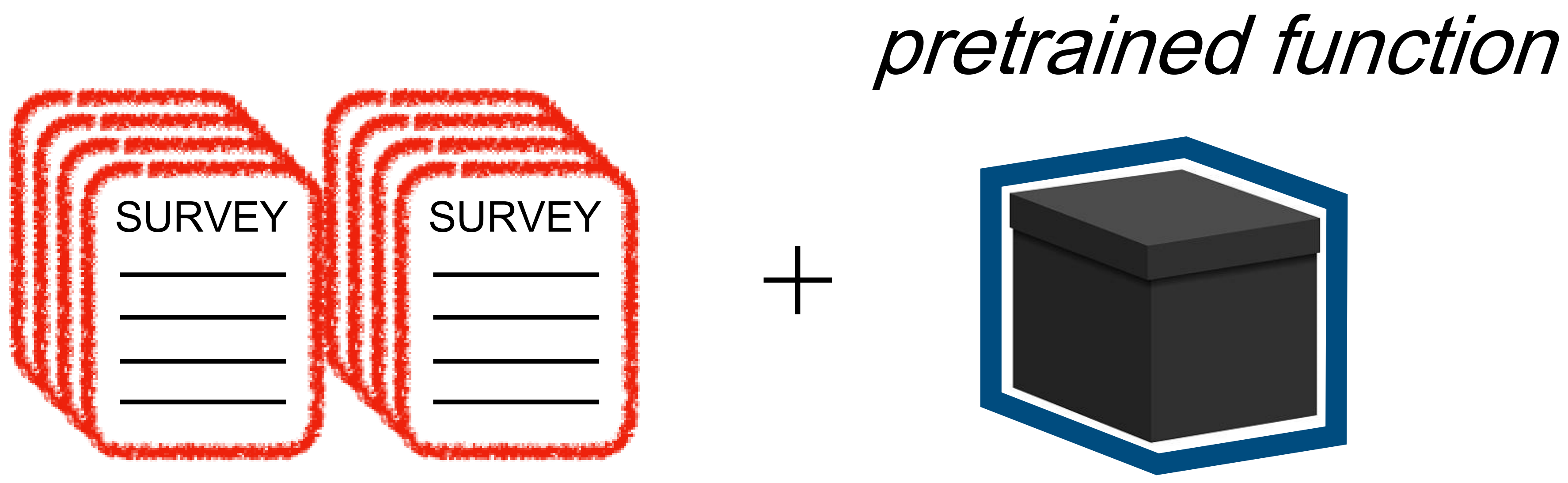
# Challenge 2: Leveraging Resumes for Surveys

**Transfer learning:** Learning representations of careers from resumes that can be adjusted (or fine-tuned) on survey datasets.

Step 2: Fine-tune representation function to predict on survey datasets.

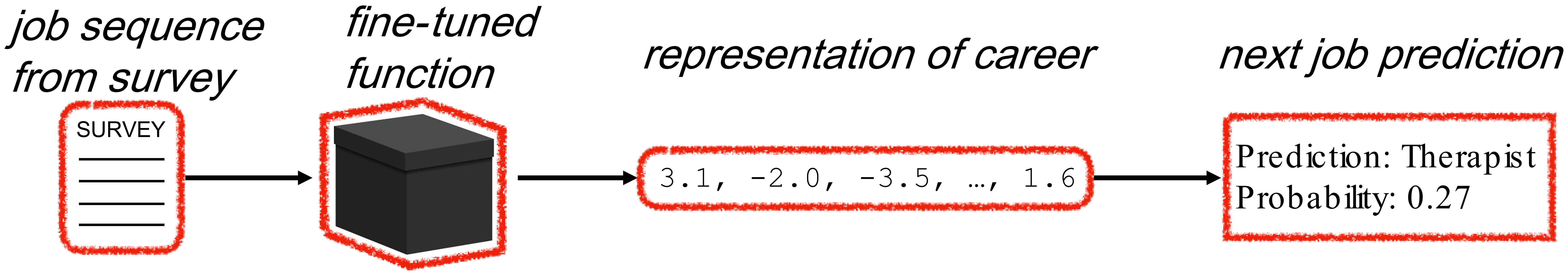
**Input:**

Survey dataset and resume representation function




**Output:**

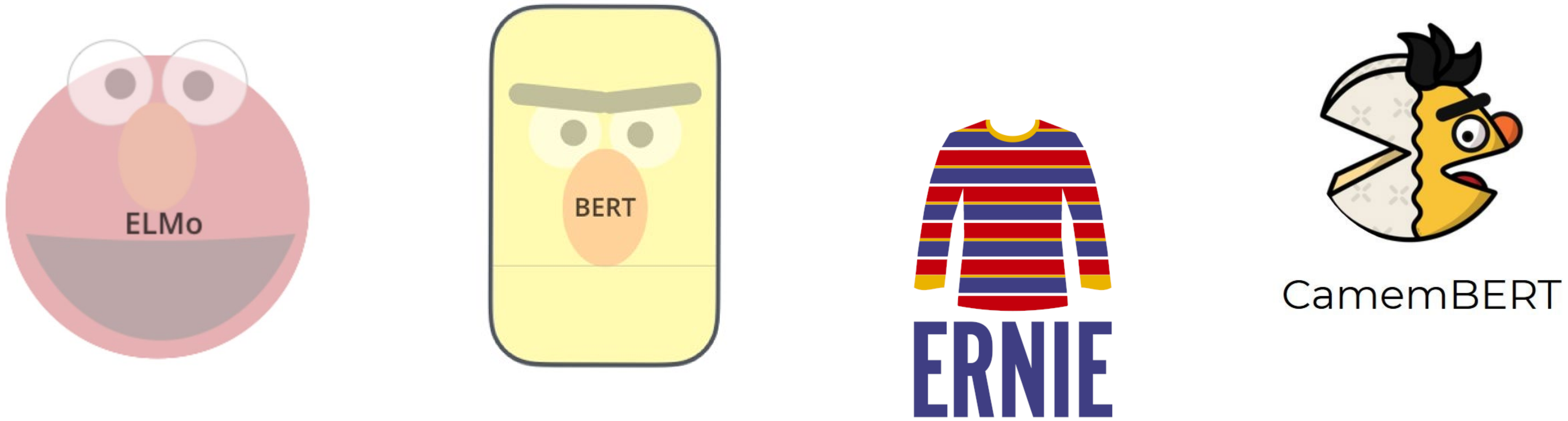
Fine-tuned representation and prediction function



# Transfer Learning Motivation

Initializing with pretrained representations ensures that the model does not need to re-learn representations from small survey datasets.

Instead, it only needs to *fine-tune* representations to account for dataset differences.




We call our approach **CAREER**

(short for **C**ontextual **A**ttention-based **R**epresentations of **E**mployment **E**ncoded from **R**esumes)

# Dataset Comparison

survey datasets

| | Resumes | NLSY79 | NLSY97 | PSID |
|---|---|---|---|---|
| Number of individuals | 24 million | 12 thousand | 9 thousand | 12 thousand |
| Unemployed/ out-of-labor-force/ student available? | No | Yes | Yes | Yes |
| Median year | 2007 | 1991 | 2007 | 2011 |
| Proportion manual laborers | 7% | 17% | 13% | 12% |
| Covariates available | year, location, education | year, location, education, gender, race/ethnicity | year, location, education, gender, race/ethnicity | year, location, education, gender, race/ethnicity |

# Prediction Performance

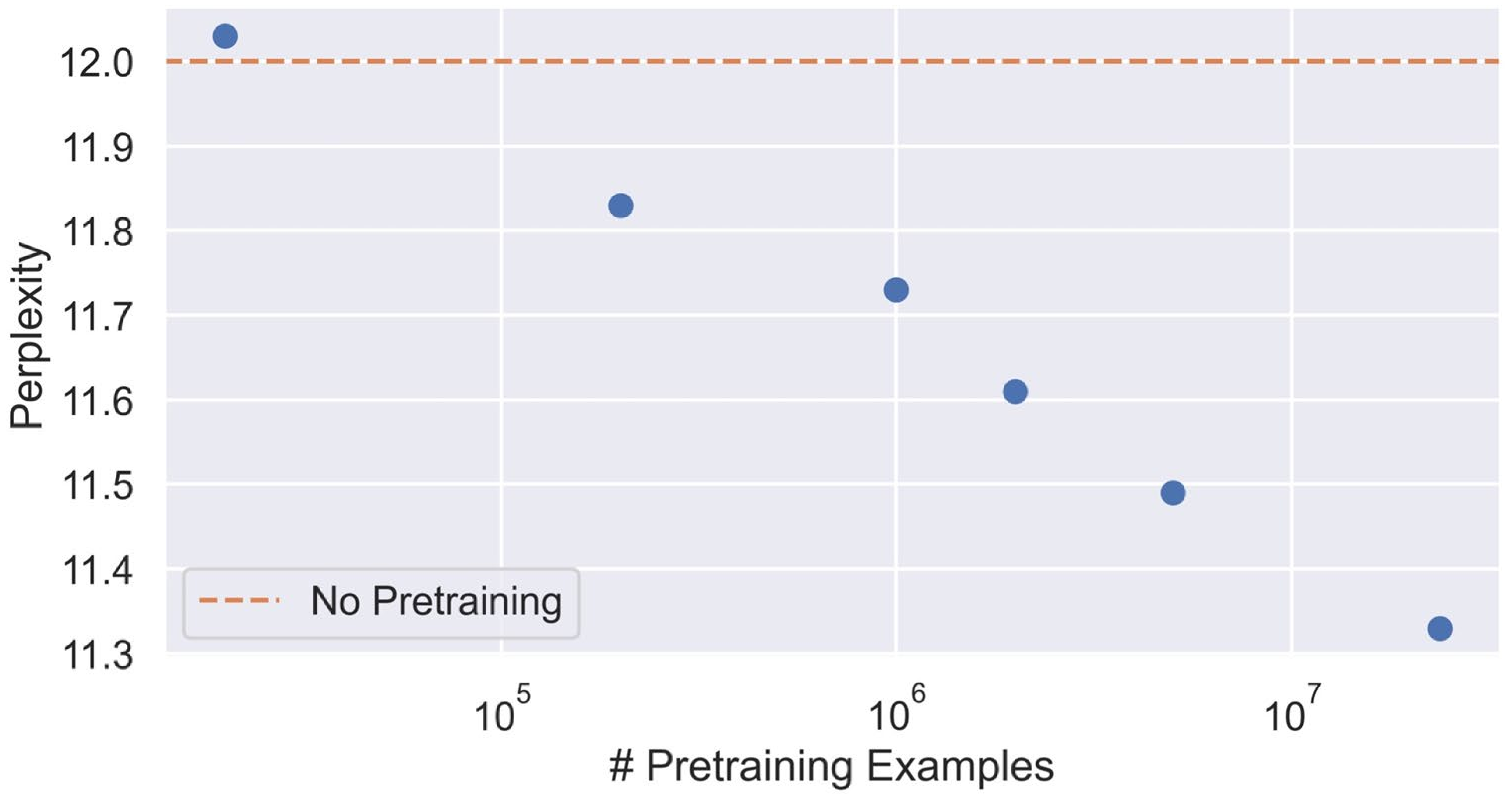

We pretrain **CAREER** on resumes and finetune on survey datasets.

| | PSID | NLSY79 | NLSY97 |
|---|---|---|---|
| Markov regression (Hall, 1972) | 18.97 ±0.10 | 15.03 ±0.03 | 20.81 ±0.02 |
| Bag-of-jobs (Ruiz et al., 2020) | 16.21 ±0.08 | 13.09 ±0.03 | 16.20 ±0.01 |
| NEMO (Li et al., 2017) | 17.58 ±0.04 | 12.82 ±0.04 | 18.38 ±0.08 |
| CAREER (vanilla) | 15.26 ±0.08 | 12.20 ±0.04 | 16.19 ±0.04 |
| CAREER (two-stage) | 14.79 ±0.04 | 12.00 ±0.00 | 15.22 ±0.03 |
| CAREER (two-stage + pretrain) | **13.88 ±0.01** | **11.32 ±0.00** | **14.15 ±0.03** |

**CAREER** outperforms baselines at predicting jobs on survey data.

# Scaling Law

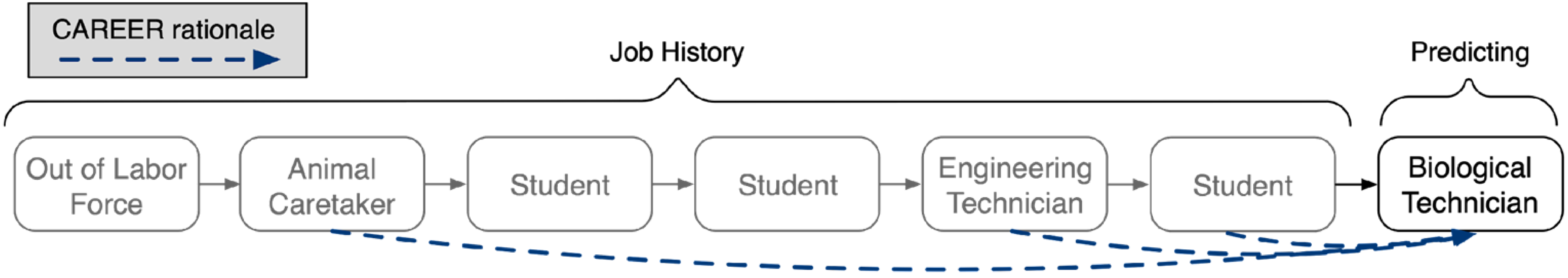
How does CAREER's predictive performance on survey datasets vary as a function of pretraining data size (number of resumes)?




Relationship follows power law, similar to scaling laws found in NLP.

# Example Prediction

Held-out sequence:




Rank of true next job (biological technician) of possible next jobs:

Regression: 41st

Bag-of-jobs: 38th

**CAREER:**2nd

# Future Work: CAREER for Economic Adjustment

Many analyses on survey datasets don't involve predicting next jobs directly, but rather estimating a quantity that **adjusts** for history.

These models typically estimate adjusted quantities by building outcome models that adjust for history.

For example: the adjusted gender wage gap involves predicting wage from covariates and summary statistics about experience (e.g. years worked).

**CAREER** learns a low-dimensional representation of job history that can be plugged into these models. Wage prediction MSE:

| | 1981 | 1990 | 1999 | 2007 | 2009 | 2011 |
|---|---|---|---|---|---|---|
| Full specification from Blau & Kahn (2017a) | 0.148 | 0.152 | 0.178 | 0.198 | 0.204 | 0.203 |
| Full specification + CAREER | **0.139** | **0.134** | **0.160** | **0.181** | **0.183** | **0.182** |

# Summary

Many economic analyses rely on building predictive models on job sequences from survey datasets.

While modern machine learning methods may struggle on small survey datasets, recent years have seen the emergence of large-scale resume data.

Inspired by approaches in NLP, CAREER leverages resume data to learn useful representations of job sequences for downstream survey predictions.

Future work: Using CAREER to estimate adjusted economic quantities.

Link to code and paper:

`https://github.com/keyonvafa/career-code`

Thank you!