MEASURING LABOR MARKET MATCH QUALITY WITH LARGE LANGUAGE MODELS

Yi Chen[1] Hanming Fang[2] Yi Zhao[3] Zibo Zhao[1]

[1]ShanghaiTech University

[2]University of Pennsylvania

[3]Tsinghua University

May, 2025

MEASURING LABOR MARKET MATCH QUALITY

The match quality between a worker and a job essentially evaluates the "suitability" of the correspondence among a set of categorical variables:

- Worker:
  - Major: STEM, economics, literature, …
  - Previous occupation: engineer, accountant, public servant, …
  - Previous industry: IT, finance, agriculture, …
  - Even MBTI is a categorical variable (INTJ, ESTP), …
- Job:
  - Occupation, industry, …

FEATURES CATEGORICAL VARIABLES

Three common types of variables:

- Ordinal variable: a clear ordering of the categories without meaningful interval (e.g., self-rated health)
- Cardinal variable: values with meaningful interval (e.g., temperature)
- Categorical variable: no intrinsic ordering to the categories (e.g., occupations)
- We cannot easily say one category is better than another (STEM > economics?)

Researchers often use a set of dummy variables for different categories to handle categorical variables, known as the fixed effect (FE) approach.

LIMITATIONS OF THE FE APPROACH (I)

Disregards the valuable information contained in the textual labels associated with categorical variables:

- Consider three occupations in our data: “Software engineer”, “System tester”, and “Sales representative”
- Intuitively, we know “Software engineer” and “System tester” are more similar to each other than “Sales representative”.
- The FE approach fails to capture subtle similarities among them.
  - Papageorgiou (2022) investigates whether larger cities help workers find better-matched jobs by using job switching as an indicator of match quality.

For the FE approach, the labels of the categories don’t matter, as long as they are different.

LIMITATIONS OF THE FE APPROACH (II)

Analysis (such as computing sample average) based on categories with limited observations in a category can be highly unstable:

Occupation (CLDS 2016)

目前或2015年1月以来最近一份职业	Freq.	Percent	Cum.
10100	9	0.06	0.06
10200	1	0.01	0.07
10401	1	0.01	0.08
10402	2	0.01	0.09
10403	1	0.01	0.10
10500	45	0.32	0.41
10600	2	0.01	0.43
10601	96	0.67	1.10
10602	7	0.05	1.15
20100	2	0.01	1.16
20108	1	0.01	1.17
20114	2	0.01	1.18
20200	21	0.15	1.33
20201	3	0.02	1.35
20205	3	0.02	1.37
20206	2	0.01	1.39
20207	5	0.04	1.42

Major (CLDS 2016)

```
. tab I2_1_33_w16
```

本科专业	Freq.	Percent	Cum.
22	1	0.13	0.13
99997	1	0.13	0.26
99999	3	0.39	0.65
ACCA	1	0.13	0.78
不清楚	1	0.13	0.91
中医	5	0.65	1.57
中医学	1	0.13	1.70
中文	6	0.78	2.48
中文系	5	0.65	3.13
中药	1	0.13	3.26
中药学	1	0.13	3.39
中西医结合	1	0.13	3.52
临床医学	9	1.17	4.70
临床药学	1	0.13	4.83
交通	1	0.13	4.96
交通运输	3	0.39	5.35
交通运输管理	1	0.13	5.48
产品设计	1	0.13	5.61
人力资源	1	0.13	5.74

HOW TO OVERCOME THESE LIMITATIONS?

- The recent development of large language models (LLMs), including GPTs, presents a novel approach to overcome the limitations of traditional measures and and provide additional insights on the job–worker match quality.
- LLMs are proficient at interpreting and analyzing textual content, allowing for direct examination of the textual labels of categories.

APPLICATION TO LABOR MARKET MISMATCH

- **Vignette-based job analysis method:** Employ GPT-3.5-turbo to simulate a human resources specialist tasked with assessing the match between workers and jobs.
- **Prompts for the "major–occupation" mismatch:**
  Pretend that you are an HR specialist. Based solely on the provided information (without considering any additional information or assumptions such as education level, working experience, previous jobs, on-the-job learning, or training), please assess whether the applicant graduated from [major title] is capable of performing the [job title]. Please respond with "Definitely can" or "Probably can" or "Probably cannot" or "Definitely cannot."

FOUR TRADITIONAL MEASURES OF LABOR MARKET MISMATCH

- **Job switching (JS) method:** examines workers' tendencies to switch jobs, assuming that this results in loss of occupation- and/or industry-specific human capital in the labor market (Kambourov and Manovskii, 2009; Sullivan, 2010; Papageorgiou, 2022).
- **Realized matches (RM) method:** derives the match index the actual distribution of educational or skill levels within occupations, assuming that workers self-select into better-matched positions (Nieto et al., 2015; Altonji et al., 2016; Sellami et al., 2018).
- **Worker-assessment (WA) method:** relies on individuals' personal opinions regarding their job match (Robst, 2007; Zhu, 2014).
- **Job analysis (JA) method:** relies on evaluations by job analysts who define required education or skills for jobs (Guvenen et al., 2020; Lise and Postel-Vinay, 2020).

STEP 1: VALIDATION

Cross-validate our GPT measure of match quality with various traditional measures with two complementary data sets.

- Zhaopin.com (1,048,575 applications to 29,914 unique job postings)
  - Large online sample
  - Application instead of final match
  - Traditional method: job switching, realized matches
- China Labor-Force Dynamic Survey (CLDS) (2,431 employees with a college degree or above).
  - Small offline sample
  - Final match
  - Traditional method: Job analysis

STEP 2: APPLICATION

An example that GPT method can construct new properties that cannot be measured by traditional data-driven methods.

- Major versatility: the ability to qualify students for various occupations.
- Consider two hypothetical majors:

  Major A: equips students with generalized skills that can be applied to a wide range of occupations.

  Major B: does not prepare students for any specific occupation.
- Students from two majors can display similar application patterns! (apply to a wide range of occupations)

LITERATURE: USING LLMS IN ECONOMICS

- As research/teaching assistant: Cowen and Tabarrok (2023); Korinek (2023)
- As natural language processor: Hansen and Kazinnik (2023); Lopez-Lira and Tang (2023); Yang and Menczer (2023)
- As simulated agents (*Homo Silicus*): Argyle et al. (2023); Chen et al. (2023); Eloundou et al. (2024); Horton (2023)

Our paper adds to the various roles that can be assigned to the GPT, specifically focusing on measuring the match quality in the labor market, and validates the GPT method in this novel application.

LITERATURE: MEASURING LABOR MARKET MISMATCHES

We contribute to this literature by proposing a novel method:

- Unlike the job switching and realized matches methods, our GPT method can recover the overlooked information in categorical variables by considering textual labels.
- Since GPT is pre-trained on vast datasets, our method isn't limited by sample size, unlike the realized matches method.
- Compared to the job analysis method, our approach treating GPT as the job analyst is more cost-effective than employing humans, especially in developing countries.

LITERATURE: TEXTUAL ANALYSIS IN STUDYING LABOR MARKET

- Use the bag-of-words/dictionary method (the meaning of words doesn't matter) to extract information from job descriptions/job titles: Deming and Kahn (2018); Atalay et al. (2020); Deming and Noray (2020); Marinescu and Wolthoff (2020)
- Measuring differences and similarities between documents (using *k*-means clustering or word2vec or TF-IDF): Imbert et al. (2022) and Biasi and Ma (2022).

Our study leverages the capabilities of recently developed LLMs, which can capture contextual nuances, semantic relationships, and diverse language patterns, to explore their application in empirical economic research.

WHAT IS LLM?

WHAT IS LLM?

Large language model (LLM): an algorithm designed to understand and generate human language by predicting word sequences.

- Utilize extensive data and parameters, enabling them to excel in comprehending and generating natural language with unparalleled proficiency.
- Can perform a wide range of language-related tasks, such as text generation, translation, summarization, question answering, and more.

GPT VERSUS CHATGPT

We specifically focus on Generative Pre-training Transformer (GPT) and use GPT-3.5-turbo in our study

- Developed by OpenAI, “brain” behind ChatGPT, its predecessor (GPT-3) boasts 175 billion parameters and is trained on a dataset containing around 500 billion tokens.
- **Generative:** Model’s ability to generate text or other forms of data.
- **Pre-trained:** A pre-training process on a massive dataset of text from the internet.
- **Transformer:** Represents a significant advancement in natural language processing (NLP).
- **ChatGPT:** fine-tuned and specialized for conversational applications
- **GPT:** more general-purpose

HOW GPT WORKS

- Next-token prediction problem
- Predict the next word in a sentence or sequence of tokens.
- Consider GPT as a language expert who has learned from vast amounts of text data.

DATASETS USED TO TRAIN GPT 3.5

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	20%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

Next-step: Finetuning and RLHF (reinforcement learning with human feedback)

DATA AND MEASURES OF MATCH QUALITY

DATA

We use two complementary datasets:

- An administrative data from Zhaopin.com which comprises 1,048,575 applications to 29,914 unique job postings on Zhaopin.com in 2013.
- The 2016 and 2018 waves of the CLDS survey data which includes 2,431 employers with a college degree or above.

SAMPLING PROCESS (ZHAOPIN.COM DATA)

Data is extracted in January 2014.

Step 1 Random sample 61,674 job applicants who initiated their search cycle in August 2013.

Step 2 Track all their applications until November 30, 2013.

- Sampled applicants filed 281,618 applications in total.

Step 3 Sample 29,914 unique job postings in Step 2.

Step 4 Collect all applications to those postings from January 1, 2013 to November 30, 2013.

- Sampled postings received 1,048,575 applications.
- This study only uses this dataset.

AVAILABLE INFORMATION (ZHAOPIN.COM DATA)

- Applicant's information
  - Demographics: gender, age, education (including major), marriage, ...
  - Current/previous work status: industry, occupation, experience, employment, ...
  - Expectation about future job: location, wage
- Posting information
  - Industry, occupation, job title
  - requirement (education/experience), offered wage (optional), # persons demanded
  - Firm: size, ownership type
- **50** categories of industries, **92** detailed categories of majors, **588** detailed categories of occupations, numerous job titles

CLASSIFICATION SYSTEMS FOR OCCUPATIONS AND INDUSTRIES (ZHAOPIN.COM DATA)

Job Title	Detailed Occupation Category	Broad Occupation Category	Industry Category
Software test engineer	Software test engineer	Software personnel/Internet developer/ System integration staff	Computer software
Game tester	Software test engineer	Software personnel/Internet developer/ System integration staff	Internet business/E-commerce
Software R&D engineer	Software R&D engineer	Software personnel/Internet developer/ System integration staff	Computer software
Video algorithm engineer	Software R&D engineer	Software personnel/Internet developer/ System integration staff	Internet business/E-commerce
Accountant	Accountant	Financial personnel/Auditors/ Taxation staff	Computer software
Human resources specialist	Administrative officer/ Administrative assistant	Administrative staff /Logistics personnel/ Secretarial staff	Computer software
Accountant	Accountant	Financial personnel/Auditors/ Taxation staff	Internet business/E-commerce
Human resources specialist	Administrative officer/ Administrative assistant	Administrative staff /Logistics personnel/ Secretarial staff	Internet business/E-commerce

CHINA LABOR-FORCE DYNAMIC SURVEY DATA (CLDS)

- The 2016 and 2018 waves of the CLDS
  - A national longitudinal social survey targeted at the labor force.
  - Consists of 37,623 respondents,
  - Jobs in the CLDS data are categorized according to the official classification.
- Complementarity between two data sources
  - Online job market v.s. entire labor market
  - Search process v.s. realized matching
  - Expected wage v.s. real wage
  - Non-official occupational classification systems v.s. official one
- Focus on the subsample consisting of 2,431 are employed and hold a college degree or above with major information

TRADITIONAL MEASURE I: JS METHOD (ZHAOPIN.COM DATA)

If individuals apply for a job j in the same occupation/industry category as their most recent job i, we consider the pair as "matched"

$$\text{Industry match}_{i,j} = \mathbb{1}\{\text{Industry category of job } j = \text{that of job } i\}$$

$$\text{Occupation match}_{i,j} = \mathbb{1}\{\text{Occupation category of job } j = \text{that of job } i\}$$

TRADITIONAL MEASURE II: RM METHOD (BOTH DATA)

Use Duncan index to compute the major–occupation match index

$$\text{Duncan match}_{m,o} = \text{Milliles}(\theta_{m,o} - \theta_m)$$

- Applicant in major category *m* applies to a job in occupation category *o*
- Milliles: a function divides the ratio difference into 1,000 (100) quantiles for Zhaopin.com data (CLDS data), and further scales it from 0 to 1.

$$\theta_{m,o} = \frac{\#\text{applicants in occupation category } o \text{ with major category } m}{\#\text{applicants in occupation category } o}$$

$$\theta_m = \frac{\#\text{applicants in major category } m}{\#\text{applicants}}$$

- Intuition: If individuals in major category *m* *disproportionally* apply for jobs in occupation category *o*, the pair is considered as a good match.

▸ Example

TRADITIONAL MEASURE III: JA METHOD (CLDS DATA)

- In 2021, the government employed job analysts to establish matched majors for all occupations listed in the official occupation classification.

一、本科层次

（一）普通高等教育本科

序号	职业编码	职业名称	专业代码	专业名称
1	1-06-01-02	企业经理	120201K	工商管理
2	2-01-01-00	哲学研究人员	010101	哲学
3	2-01-01-00	哲学研究人员	010102	逻辑学
4	2-01-01-00	哲学研究人员	010104T	伦理学
5	2-01-02-00	经济学研究人员	020101	经济学
6	2-01-02-00	经济学研究人员	020104T	资源与环境经济学
7	2-01-02-00	经济学研究人员	020105T	商务经济学
8	2-01-02-00	经济学研究人员	020306T	信用管理
9	2-01-02-00	经济学研究人员	030205T	政治学、经济学与哲学
10	2-01-03-00	法学研究人员	030101K	法学

- The CLDS data are categorized according to the official classification, allowing us to utilize the JA method.

$$\text{JA match}_{m,o} = \mathbb{1}\{\text{major } m \in \text{matched majors for occupation } o\}$$

HOW DO OTHER STUDIES DESIGN PROMPTS?

1. Lopez-Lira and Tang (2023): Whether a news is good for a stock?
   - Ans: Yes/No/Unknown
2. Hansen and Kazinnik (2023): Decipher Fedspeak.
   - Ans: Dovish/Hawkish/Mostly Dovish/Mostly Hawkis/Neutral
3. Yang and Menczer (2023): Rate news outlet credibility.
   - Ans: 0/0.1/0.2/.../1
4. Eloundou et al. (2024): Whether using LLM can decreases the time of a task by 50%?
   - Ans: 0/1 (or Level 1/2/3/4)

OUR GPT PROMPTS

Pretend that you are an HR specialist. Based solely on the provided information (without considering any additional information or assumptions such as education level, working experience, previous jobs, on-the-job learning, or training), please assess whether the applicant graduated from [major, m] is capable of performing the [job, j]. Please respond with "Definitely can" or "Probably can" or "Probably cannot" or "Definitely cannot".

$$\text{GPT match}_{m,j} = \begin{cases} 0, \text{response} = \text{“Definitely/Probably cannot”} \\ 1, \text{response} = \text{“Probably/Definitely can”} \end{cases}$$

▸ Why don't we use more complicated prompts? ▸ Some Examples

BACK TO THE LIMITATIONS OF TRADITIONAL MEASURES (I)

- JS method: if two occupations do not fall into the same category, they are viewed as completely different (e.g., “Software test engineer” and “Software R&D engineer”).

$$\text{Occupation match}_{i,j} = \mathbb{1}\{\text{Occupation category of job } j = \text{that of job } i\}$$

- The GPT method can exploit the meaning of labels, and captures similarities between different categories.

COMPARISONS BETWEEN TRADITIONAL AND GPT MEASURES

Detailed Occupation Category of Applied Job	Detailed Occupation Category of Current Job	Same-occupation Dummy	GPT Response	GPT Occupation-occupation Match
Software engineer	Software engineer	1	Probably can	1
	System tester	0	Probably can	1
	Sales representative	0	Probably cannot	0
Industry Category of Applied Job	**Industry Category of Current Job**	**Same-industry Dummy**	**GPT Response**	**GPT Industry-industry Match**
IT services	IT services	1	Probably can	1
	Computer software	0	Probably can	1
	Computer hardware	0	Probably cannot	0

BACK TO THE LIMITATIONS OF TRADITIONAL MEASURES (II)

- RM method: if the segment is very small, the computed ratio would not be reliable.

$$\theta_{m,o} = \frac{\#\text{applicants in occupation category } o \text{ with major category } m}{\#\text{applicants in occupation category } o}$$

- The occupation title as an extreme.
- The GPT method is pre-trained on vast datasets and does not have any requirement on the segment size.

SUMMARY STATISTICS

Data	Zhaopin.com			CLDS Data		
	Mean	S.D.	Obs.	Mean	S.D.	Obs.
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Individual Characteristics						
Female	0.48	0.50	847,801	0.51	0.50	2,431
Age	26.90	4.74	847,801	35.68	9.64	2,428
Married	0.28	0.45	847,801	0.71	0.45	2,431
Bachelor degree or above	0.44	0.50	847,801	0.54	0.50	2,431
Years of schooling	15.53	0.76	847,801	15.69	0.90	2,431
Working experience	5.70	3.60	847,801	19.99	9.73	2,428
Monthly wage of the most recent job	4,457	3,076	846,535	5,055	3,837	2,206
Monthly wage of expected job	4,709	3,243	846,740			
Unemployed	0.73	0.44	847,801			
Panel B: Match Measures						
Same-occupation dummy	0.22	0.42	847,801			
GPT occupation–occupation match	0.69	0.46	843,296			
Same-industry dummy	0.26	0.44	847,801			
GPT industry–industry match	0.48	0.50	773,203			
Duncan major-occupation match	0.71	0.33	816,161	0.60	0.31	2,431
JA major–occupation match				0.35	0.48	2,382
GPT major–title match	0.54	0.50	832,623	0.47	0.50	2,303

▸ More graphs for the “major-occupation” match

USING THE GPT TO MEASURE MATCH QUALITY

PAIRWISE CORRELATION BETWEEN TRADITIONAL AND GPT MEASURES

Panel A: Zhaopin.com	Same-occupation dummy (1)	GPT occupation-occupation match (2)	Same-industry dummy (3)	GPT industry-industry match (4)	Duncan major-occupation match (5)	GPT major-title match (6)
Same-occupation dummy	1					
GPT occupation-occupation match	**0.354*** **	1				
Same-industry dummy	0.130***	0.100***	1			
GPT industry-industry match	0.117***	0.100***	**0.655*** **	1		
Duncan major-occupation match	0.103***	0.103***	0.075***	0.089***	1	
GPT major-title match	0.098***	0.078***	0.081***	0.088***	**0.436*** **	1

Panel B: CLDS Data	Duncan major-occupation match	JA major-title match	GPT major-title match
Duncan major-occupation match	1		
JA major-occupation match	0.507***	1	
GPT major-title match	**0.429*** **	**0.555*** **	1

- GPT measures are highly correlated with traditional ones.
- How to interpret the difference?

WAGE PREMIUM OF MATCHES MEASURED BY TRADITIONAL AND GPT METHODS (ZHAOPIN.COM DATA)

Dependent Variable	Monthly Wage of Expected Job (Log)											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Same-occupation dummy	0.010**		0.004							0.007**		0.003
	(0.004)		(0.004)							(0.003)		(0.004)
GPT occupation-occupation match		0.017***	0.016***								0.015***	0.013***
		(0.003)	(0.003)								(0.003)	(0.003)
Same-industry dummy				0.022***		0.017***				0.020***		0.015**
				(0.005)		(0.006)				(0.005)		(0.006)
GPT industry-industry match					0.018***	0.008***					0.016***	0.007**
					(0.003)	(0.003)					(0.003)	(0.003)
Duncan major-occupation match							0.008***		0.008***	0.006***		0.005***
							(0.002)		(0.002)	(0.002)		(0.002)
GPT major-title match								0.005***	0.002		0.003**	0.001
								(0.002)	(0.001)		(0.001)	(0.001)
Basic control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major category FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation category of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry category of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	794,461	794,461	794,461	743,568	743,568	743,568	801,107	801,107	801,107	712,586	712,586	712,586
Additional R-squared for controlling match measure ($\times 10^{-2}$)	0.0312	0.1039	0.1089	0.1428	0.1030	0.1553	0.0222	0.0076	0.0232	0.1744	0.1860	0.2397

GPT MEASURES CAN PROVIDE EXTRA INFORMATION

Dependent Variable	Monthly Wage of Expected Job (Log)											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Same-occupation dummy	0.010**		0.004							0.007**		0.003
	(0.004)		(0.004)							(0.003)		(0.004)
GPT occupation-occupation match		0.017***	0.016***								0.015***	0.013***
		(0.003)	(0.003)								(0.003)	(0.003)
Same-industry dummy				0.022***		0.017***				0.020***		0.015**
				(0.005)		(0.006)				(0.005)		(0.006)
GPT industry-industry match					0.018***	0.008***					0.016***	0.007**
					(0.003)	(0.003)					(0.003)	(0.003)
Duncan major-occupation match							0.008***		0.008***	0.006***		0.005***
							(0.002)		(0.002)	(0.002)		(0.002)
GPT major-title match								0.005***	0.002		0.003**	0.001
								(0.002)	(0.001)		(0.001)	(0.001)
Basic control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major category FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation category of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry category of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	794,461	794,461	794,461	743,568	743,568	743,568	801,107	801,107	801,107	712,586	712,586	712,586
Additional R-squared for controlling match measure ($\times 10^{-2}$)	0.0312	0.1039	0.1089	0.1428	0.1030	0.1553	0.0222	0.0076	0.0232	0.1744	0.1860	0.2397

WHERE DOES THE EXTRA INFORMATION COME FROM?

- The GPT measures of the "occupation–occupation" and "industry–industry" matches still exhibit statistically significant positive wage effects when both traditional and GPT measures are included in the regressions.
- For the Occupation match$_{i,j}$, no variation if conditional on different occupations (because this is how we define the measure)
  - Similar logic applies to Industry match$_{i,j}$
  - GPT measure can provide extra information because it uses the label information attached to different categories.

Dependent Variable:	Monthly Wage of Expected Job (Log)				
Regressions Conditional on:	Applied job in a different occupation category		Applied job in a different industry category		Interaction between Duncan and GPT Measures
	(1)	(2)	(3)	(4)	(5)
Same-occupation dummy	Omitted	Omitted		0.001	
				(0.003)	
GPT occupation-occupation match	**0.016*** **	0.013***		0.015***	
	(0.003)	(0.003)		(0.003)	
Same-industry dummy		0.018**	Omitted	Omitted	
		(0.007)			
GPT industry-industry match		0.007**	**0.008*** **	0.006**	
		(0.003)	(0.003)	(0.003)	
Duncan major-occupation match		0.004**		0.007***	0.007***
		(0.002)		(0.001)	(0.002)
GPT major-title match		0.002		0.001	0.002
		(0.001)		(0.001)	(0.001)
Duncan major-occupation match×					-0.002***
GPT major-title match					(0.001)
Basic control	Yes	Yes	Yes	Yes	Yes
Major category FE	Yes	Yes	Yes	Yes	Yes
Occupation category of applied job FE	Yes	Yes	Yes	Yes	Yes
Industry category of applied job FE	Yes	Yes	Yes	Yes	Yes
City of applied job FE	Yes	Yes	Yes	Yes	Yes
Observations	612,717	545,836	531,498	508,099	801,107

WHEN DOES THE GPT METHOD WORK BETTER?

DETAILED CATEGORIES (588) ⟶ BROAD CATEGORIES (58)

Dependent Variable	Monthly Wage of Expected Job (Log)					
	(1)	(2)	(3)	(4)	(5)	(6)
Same-occupation dummy	0.015***		0.007			
	(0.005)		(0.010)			
GPT occupation-occupation match		0.015***	0.010			
		(0.006)	(0.010)			
Duncan major-occupation match				0.012***		0.011***
				(0.003)		(0.002)
GPT major-title match					0.006**	0.001
					(0.003)	(0.002)
Basic control	Yes	Yes	Yes	Yes	Yes	Yes
Major category FE	Yes	Yes	Yes	Yes	Yes	Yes
Broad occupation category of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry category of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes
City of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	794,075	794,075	794,075	801,108	801,108	801,108
Additional R-squared for controlling match measure ($\times 10^{-2}$)	0.0722	0.0798	0.0878	0.0455	0.0101	0.0459

- GPT methods perform better when categories are detailed, because more categories can be conceptually similar.
- The FE approach is sufficient when the number of categories is small (e.g., male vs. female).

WHY DOESN'T THE TRADITIONAL NLP METHOD WORK?

- Traditional NLP: similarities between the textual labels of the categorical variables.
- First, convert text into numerical vectors based on word frequencies.
- Then, compute the cosine similarity:

$$\text{Cosine Similarity} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|},$$

- Intuition: "Accounting" (*Kuaiji xue*) fits (*Kuaiji* in Chinese); "researcher assistant" is similar to "assistant researcher."

TRADITIONAL NLP METHODS WORK POORLY

Dependent Variable	Monthly Wage of Expected Job (Log)				
	(1)	(2)	(3)	(4)	(5)
Duncan major-occupation match	0.008***		0.008***		0.008***
	(0.002)		(0.002)		(0.002)
TF-IDF major-title match		-0.000	-0.000		
		(0.000)	(0.000)		
BoW major-title match				-0.000	-0.001
				(0.000)	(0.000)
Basic control	Yes	Yes	Yes	Yes	Yes
Major category FE	Yes	Yes	Yes	Yes	Yes
Occupation category of applied job FE	Yes	Yes	Yes	Yes	Yes
Industry category of applied job FE	Yes	Yes	Yes	Yes	Yes
City of applied job FE	Yes	Yes	Yes	Yes	Yes
Observations	801,107	801,107	801,107	801,107	801,107
Additional R-squared for controlling match measure ($\times 10^{-2}$)	0.0222	0.0000	0.0222	0.0001	0.0223

ROBUSTNESS CHECKS WITH DIFFERENT PROMPTS AND LLMS

Our results are robust to different prompts and LLMs.

- A prompt tasking GPT to simulate a career advisor and evaluate job fitness from the job seekers' perspective. ▸ Prompts from job seeker's perspective
- A more complex prompt, instructing GPT to answer step-by-step, known as "Chain of Thought" (CoT). ▸ CoT prompting
- A LLM (ERNIE Bot) developed by a Chinese company—Baidu, that may possess more local knowledge about the Chinese labor market. ▸ ERNIE Bot
- A recent LLM (Claude 3 Haiku) developed by the second-largest LLM startup—Anthropic, that is released in 2024. ▸ Claude 3 Haiku

THE WIDE APPLICABILITY OF GPT METHOD

Utilize the CLDS data with a focus on assessing the most demanding major–occupation matches.

- The RM method, employed for Zhaopin.com data, cannot be used for CLDS data due to a small sample size.
- Our GPT method does not impose any requirement on sample size.

THE WIDE APPLICABILITY OF THE GPT METHOD (CLDS DATA)

Dependent Variable	Monthly Wage of Current Job (Log)			
	(1)	(2)	(3)	(4)
Duncan major-occupation match	0.025**			-0.010
	(0.009)			(0.017)
JA major-occupation match		0.056***		0.037
		(0.017)		(0.021)
GPT major-title match			0.058***	0.048**
			(0.016)	(0.018)
Basic control	Yes	Yes	Yes	Yes
Survey year FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes	Yes	Yes
Major category FE	Yes	Yes	Yes	Yes
Occupation category FE	Yes	Yes	Yes	Yes
Industry category FE	Yes	Yes	Yes	Yes
Observations	2,035	2,002	2,035	2,002
Additional R-squared for controlling match measure ($\times 10^{-2}$)	0.0759	0.4509	0.4403	0.6730

MEASURING THE VERSATILITY OF MAJORS WITH GPT

AN ARTIFICIAL EXAMPLE

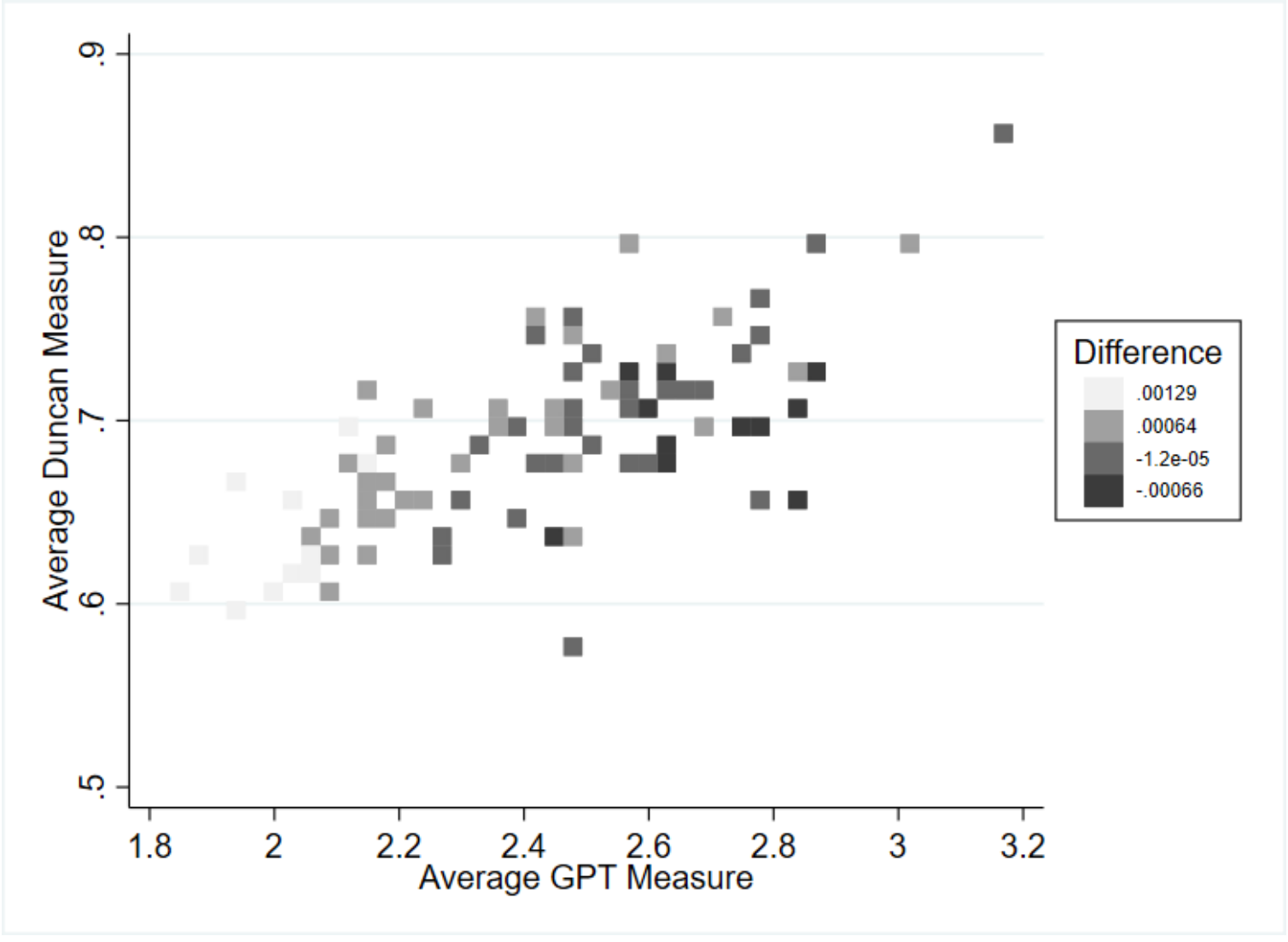
Majors	Occupations Applied For	Proportion of Applicants (%)	Duncan Major-Occupation Index	GPT Major-Title Match
Specialized major	O1	91	1	1
	O2	3	0.1	0
	O3	3	0.1	0
	O4	3	0.1	0
Versatile major	O1	25	0.4	1
	O2	25	0.4	1
	O3	25	0.4	1
	O4	25	0.4	1
Unprepared major	O1	25	0.4	0
	O2	25	0.4	0
	O3	25	0.4	0
	O4	25	0.4	0

- Although the second and third majors share the same Duncan index due to similar application patterns, they differ fundamentally in versatility.

VERSATILE MAJORS ARE UNFAIRLY PENALIZED BY TRADITIONAL RM MEASURES

Dependent Variable:	Monthly Wage of Expected Job (Log)				
Regressions Conditional on:	Applied job in a different occupation category		Applied job in a different industry category		Interaction between Duncan and GPT Measures
	(1)	(2)	(3)	(4)	(5)
Same-occupation dummy	Omitted	Omitted		0.001	
				(0.003)	
GPT occupation-occupation match	**0.016*** **	0.013***		0.015***	
	(0.003)	(0.003)		(0.003)	
Same-industry dummy		0.018**	Omitted	Omitted	
		(0.007)			
GPT industry-industry match		0.007**	**0.008*** **	0.006**	
		(0.003)	(0.003)	(0.003)	
Duncan major-occupation match		0.004**		0.007***	0.007***
		(0.002)		(0.001)	(0.002)
GPT major-title match		0.002		0.001	0.002
		(0.001)		(0.001)	(0.001)
Duncan major-occupation match ×					-0.002***
GPT major-title match					(0.001)
Basic control	Yes	Yes	Yes	Yes	Yes
Major category FE	Yes	Yes	Yes	Yes	Yes
Occupation category of applied job FE	Yes	Yes	Yes	Yes	Yes
Industry category of applied job FE	Yes	Yes	Yes	Yes	Yes
City of applied job FE	Yes	Yes	Yes	Yes	Yes
Observations	612,717	545,836	531,498	508,099	801,107

VERSATILE MAJORS ARE UNFAIRLY PENALIZED BY TRADITIONAL RM MEASURES



CONCLUSION

- We propose a novel method that utilizes recent advances in large language models (LLMs) to recover overlooked information in categorical variables.
- We highlight several advantages:
    1. It can establish correlation among different categories.
    2. It performs well in small categories.
    3. It has wide applicability with a lower cost.
- We demonstrate the capacity of GPT method to provide extra information conditional on traditional measures of matching quality.
- We demonstrate how GPT can assist in measuring the versatility of academic majors, a task that traditional methods struggle to address.

AN EXAMPLE OF DUNCAN MAJOR-OCCUPATION MATCH INDEX

Detailed Occupation Category	Detailed Major Category (Proportion of Applicants in Major Category in the Data, %)	Proportion of Applicants in Major Category within Occupation Category (%)	Proportion difference (%)	Duncan Index
Tour consultant	Mechanical (9.17)	2.11	-7.06	0.016
Tour consultant	Tourism management (2.38)	30.56	28.18	0.99
Mechanical designer	Mechanical (9.17)	82.85	73.68	1
Mechanical designer	Tourism management (2.38)	0.04	-2.33	0.082

◂ Back

USING PROMPTS FROM JOB SEEKERS' PERSPECTIVE

Dependent Variable	Monthly Wage of Expected Job (Log)											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: GPT with Alternative Prompt												
Same-occupation dummy	0.007*		0.002							0.005		0.001
	(0.004)		(0.006)							(0.004)		(0.006)
GP occupation-occupation match		0.012***	0.011*								0.010***	0.009
using alternative prompt		(0.004)	(0.006)								(0.004)	(0.006)
Same-industry dummy				0.019***		0.024***				0.018***		0.022***
				(0.003)		(0.007)				(0.003)		(0.007)
GPT industry-industry match					0.017***	-0.005					0.016***	-0.005
using alternative prompt					(0.003)	(0.006)					(0.003)	(0.006)
Duncan major-occupation match							0.005**		0.004*	0.004		0.002
							(0.002)		(0.002)	(0.002)		(0.002)
GPT major-title match								0.005*	0.004		0.005	0.004
using alternative prompt								(0.003)	(0.003)		(0.003)	(0.003)
Additional R-squared for controlling match measure ($\times 10^{-2}$)	0.0236	0.0508	0.0526	0.1104	0.078	0.1118	0.0062	0.0072	0.0109	0.1273	0.1228	0.1544
Panel B: GPT with Baseline Prompt												
Same-occupation dummy	0.007*		0.001							0.005		-0.000
	(0.004)		(0.004)							(0.004)		(0.004)
GPT occupation-occupation match		0.016***	0.015***								0.015***	0.014***
		(0.004)	(0.004)								(0.004)	(0.004)
Same-industry dummy				0.019***		0.017***				0.018***		0.016***
				(0.003)		(0.005)				(0.003)		(0.005)
GPT industry-industry match					0.016***	0.003					0.014***	0.003
					(0.003)	(0.004)					(0.002)	(0.004)
Duncan major-occupation match							0.005**		0.005**	0.004		0.003
							(0.002)		(0.002)	(0.002)		(0.003)
GPT major-title match								0.001	-0.001		-0.001	-0.002
								(0.003)	(0.003)		(0.003)	(0.003)
Additional R-squared for controlling match measure ($\times 10^{-2}$)	0.0236	0.0902	0.0907	0.1104	0.0715	0.1117	0.0062	0.0002	0.0063	0.1273	0.1489	0.1858
Observations	96,432	96,432	96,432	96,432	96,432	96,432	96,432	96,432	96,432	96,432	96,432	96,432
Baseline control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major category FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation category of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry category of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

◂ Back

USING COT PROMPTING

Dependent Variable	Monthly Wage of Expected Job (Log)											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: GPT with CoT Prompt												
Same-occupation dummy	0.011**		0.009*							0.010**		0.009*
	(0.004)		(0.005)							(0.004)		(0.005)
GPT occupation-occupation match using CoT prompt		0.011***	0.008**								0.010***	0.007**
		(0.003)	(0.004)								(0.003)	(0.003)
Same-industry dummy				0.017***		0.017***				0.016***		0.016***
				(0.004)		(0.005)				(0.004)		(0.005)
GPT industry-industry match using CoT prompt					0.008*	-0.001					0.008*	-0.001
					(0.004)	(0.005)					(0.004)	(0.005)
Duncan major-occupation match							0.007**		0.006**	0.005*		0.005
							(0.003)		(0.003)	(0.003)		(0.003)
GPT major-title match using CoT prompt								0.003	0.002		0.003	0.002
								(0.002)	(0.002)		(0.002)	(0.002)
Additional R-squared for controlling match measure ($\times 10^{-2}$)	0.0544	0.0408	0.0769	0.0847	0.0207	0.0848	0.0101	0.0031	0.0115	0.1374	0.0618	0.1568
Panel B: GPT with Baseline Prompt												
Same-occupation dummy	0.011***		0.008*							0.010**		0.007*
	(0.004)		(0.004)							(0.004)		(0.004)
GPT occupation-occupation match		0.014***	0.011***								0.013***	0.010**
		(0.004)	(0.004)								(0.004)	(0.004)
Same-industry dummy				0.017***		0.014***				0.016***		0.013***
				(0.004)		(0.004)				(0.004)		(0.004)
GPT industry-industry match					0.015***	0.004					0.014***	0.004
					(0.004)	(0.004)					(0.004)	(0.004)
Duncan major-occupation match							0.007**		0.007**	0.005**		0.005
							(0.003)		(0.003)	(0.003)		(0.003)
GPT major-title match								0.002	0.000		0.001	-0.000
								(0.003)	(0.003)		(0.003)	(0.003)
Additional R-squared for controlling match measure ($\times 10^{-2}$)	0.0544	0.0647	0.0879	0.0847	0.0594	0.0868	0.0101	0.0014	0.0101	0.1374	0.1163	0.1652
Observations	101,141	101,141	101,141	101,141	101,141	101,141	101,141	101,141	101,141	101,141	101,141	101,141
Baseline control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major category FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation category of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry category of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

◂ Back

USING ERNIE BOT

Dependent Variable	Monthly Wage of Expected Job (Log)											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: ERNIE Bot with Baseline Prompt												
Same-occupation dummy	0.008*		0.004							0.007		0.003
	(0.005)		(0.006)							(0.005)		(0.006)
ERNIE Bot occupation-occupation match		0.012***	0.010***								0.012***	0.009**
		(0.003)	(0.004)								(0.003)	(0.004)
Same-industry dummy				0.017***		0.016***				0.016***		0.014***
				(0.003)		(0.004)				(0.003)		(0.004)
ERNIE Bot industry-industry match					0.012***	0.002					0.011***	0.002
					(0.003)	(0.004)					(0.003)	(0.004)
Duncan major-occupation match							0.006**		0.006**	0.005*		0.004
							(0.003)		(0.003)	(0.003)		(0.003)
ERNIE Bot major-title match								0.003	0.002		0.002	0.001
								(0.002)	(0.002)		(0.002)	(0.002)
Additional R-squared for controlling match measure ($\times 10^{-2}$)	0.0277	0.0524	0.0579	0.0828	0.0402	0.0837	0.0093	0.0027	0.0101	0.1095	0.0877	0.1331
Panel B: GPT with Baseline Prompt												
Same-occupation dummy	0.008*		0.005							0.007		0.005
	(0.005)		(0.005)							(0.005)		(0.005)
GPT occupation-occupation match		0.012***	0.010***								0.011***	0.008**
		(0.003)	(0.003)								(0.003)	(0.003)
Same-industry dummy				0.017***		0.016***				0.016***		0.015**
				(0.003)		(0.006)				(0.003)		(0.006)
GPT industry-industry match					0.011***	0.002					0.010***	0.001
					(0.003)	(0.005)					(0.003)	(0.005)
Duncan major-occupation match							0.006**		0.007**	0.005*		0.005*
							(0.003)		(0.003)	(0.003)		(0.003)
GPT major-title match								0.001	-0.001		0.000	-0.002
								(0.002)	(0.001)		(0.002)	(0.002)
Additional R-squared for controlling match measure ($\times 10^{-2}$)	0.0277	0.0441	0.054	0.0828	0.0381	0.0833	0.0093	0.0004	0.0095	0.1095	0.0759	0.1298
Observations	100,260	100,260	100,260	100,260	100,260	100,260	100,260	100,260	100,260	100,260	100,260	100,260
Baseline control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major category FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation category of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry category of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

◂ Back

USING CLAUDE 3 HAIKU

Dependent Variable	Monthly Wage of Expected Job (Log)											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Claude 3 Haiku with Baseline Prompt												
Same-occupation dummy	0.005		-0.001							0.004		-0.002
	(0.004)		(0.004)							(0.004)		(0.003)
Claude 3 Haiku occupation-occupation match		0.022***	0.022***								0.021***	0.021***
		(0.003)	(0.003)								(0.003)	(0.003)
Same-industry dummy				0.044***		0.040***				0.043***		0.037***
				(0.003)		(0.004)				(0.003)		(0.004)
Claude 3 Haiku industry-industry match					0.012*	0.004*					0.011*	0.004*
					(0.007)	(0.003)					(0.006)	(0.002)
Duncan major-occupation match							0.004		0.002	0.003		0.001
							(0.003)		(0.003)	(0.003)		(0.004)
Claude 3 Haiku major-title match								0.007***	0.006**		0.006***	0.007**
								(0.002)	(0.002)		(0.002)	(0.003)
Additional R-squared for controlling match measure ($\times 10^{-2}$)	0.0120	0.2085	0.209	0.1477	0.0545	0.1541	0.0046	0.0153	0.0170	0.1586	0.2689	0.3579
Panel B: GPT with Baseline Prompt												
Same-occupation dummy	0.005		0.000							0.004		-0.001
	(0.004)		(0.003)							(0.004)		(0.003)
GPT occupation-occupation match		0.016***	0.016***								0.016***	0.015***
		(0.004)	(0.004)								(0.004)	(0.004)
Same-industry dummy				0.044***		0.044***				0.043***		0.042***
				(0.003)		(0.004)				(0.003)		(0.003)
GPT industry-industry match					0.011	-0.000					0.010	-0.000
					(0.009)	(0.002)					(0.008)	(0.002)
Duncan major-occupation match							0.004		0.003	0.003		0.001
							(0.003)		(0.003)	(0.003)		(0.004)
GPT major-title match								0.003	0.002		0.002	0.002
								(0.002)	(0.003)		(0.002)	(0.003)
Additional R-squared for controlling match measure ($\times 10^{-2}$)	0.0120	0.1104	0.1104	0.1477	0.0310	0.1477	0.0046	0.0036	0.0062	0.1586	0.1388	0.2469
Observations	90,780	90,780	90,780	90,780	90,780	90,780	90,780	90,780	90,780	90,780	90,780	90,780
Baseline control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Major category FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Occupation category of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry category of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
City of applied job FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

◂ Back

WHY DON'T WE USE MORE COMPLICATED PROMPTS?

- LLMs generate texts that rely heavily on training data (Wyntera et. al., 2023; Meng et. al., 2022). (latent space memorization)
- LLMs behave poorly when the prompts are quite far away from what has already been seen in the training set. Examples:
  - Reversal Curses. (Berglund et. al., 2023)
  - Repeat "poem, poem, poem, poem" indefinitely can extract training data. (Nasr et. al., 2023)
  - hacking the positional encoding from extrapolation to interpolation improve the context length of LLM (Chen et. al., 2023).
- LLMs can be conditioning out of context

WHY DON'T WE USE MORE COMPLICATED PROMPTS? (EMPIRICAL ARGUMENT)

- We also experimented with throwing in all the information about the applicants and jobs into the prompt.
- GPT would heavily rely on one single dimension.
  - In our case, GPT would consider "experienced" applicants fit most jobs.
- Borrowing the experience from previous studies using LLMs, we use simple prompt to generate intermediate outputs and perform further analysis by ourselves.
  - This approach also partly alleviate the "black-box" issue of LLMs.

◂ Back

AN EXAMPLE OF CONSTRUCTING THE MEASURE OF MAJOR–OCCUPATION MATCH USING GPT

Prompt of the Question (Chinese and English Translation)

[用户]
假设你是一名HR专家，仅根据提供的信息（不考虑其他信息或假设，例如教育水平、工作经验、以前的工作、在职学习或培训），评估是否‘管理学’毕业的申请人能够胜任‘汽车零部件销售’。请回答‘一定是’或‘大概是’或‘大概不是’或‘一定不是’。

[User]
Pretend that you are an HR specialist. Based solely on the information provided (without considering additional information or assumptions such as educational level, work experience, previous jobs, on-the-job learning or training), assess whether the applicant graduated from 'Management' is capable of performing 'Automotive Parts Sales'. Please respond with "Definitely can" or "Probably can" or "Probably cannot" or "Definitely cannot".

Answer (Chinese and English Translation)

[GPT]
大概是

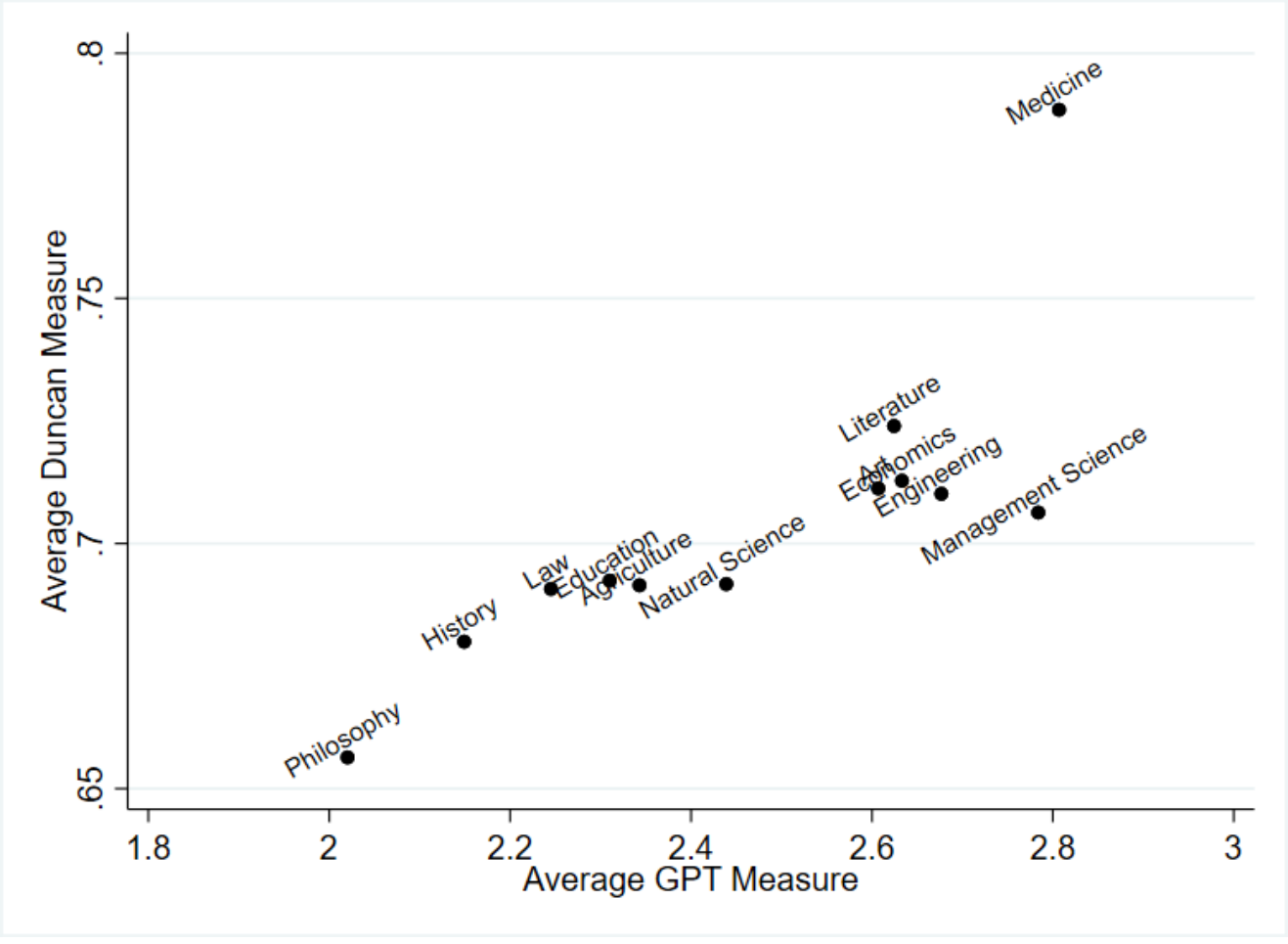
理由：
管理学毕业的申请人通常具备一定的组织管理、沟通和市场分析能力，这些技能在汽车零部件销售中可能会有用。然而，能否完全胜任还需要结合其具体的行业知识、产品理解和销售经验。

[GPT]
Probably Can

Reason:
Graduates in management typically possess skills in organization, communication, and market analysis, which could be useful in automotive parts sales. However, their ability to fully perform the role would also depend on specific industry knowledge, product understanding, and sales experience.

Back

THE "MAJOR–OCCUPATION" MATCH BY BROAD OCCUPATION CATEGORY



Back