Appendix A Proofs

The appendix A contains the algebra and proofs for theoretical model.

A.1 Proof of the Model with Traditional AI

Now we know the task specific productivity profiles are

$$\alpha_L(i) = (1-i)^{\eta},$$

$$\alpha_H(i) = i^{\eta},$$

$$\alpha_M(i) = (1-i)^{\mu}.$$

The competitive equilibrium can be characterized by the following conditions:

1. Law of one price for skills

$$F = \tilde{p}(i) A_M \alpha_M(i) \text{ for any } i < \tilde{I}_1,$$

$$\tilde{W}_L = \tilde{p}(i) A_L \alpha_L(i) \text{ for any } \tilde{I}_2 > i > \tilde{I}_1,$$

$$\tilde{W}_H = \tilde{p}(i) A_H \alpha_H(i) \text{ for any } i > \tilde{I}_2.$$

2. Goods market (or task market) clearing condition demands:

$$\tilde{p}(i)\tilde{y}(i) = \tilde{p}(i')\tilde{y}(i') = \tilde{Y}.$$

In the three intervals, the following equations are satisfied

$$\tilde{p}(i) A_L \alpha_L(i) \tilde{l}(i) = \tilde{p}(i') A_L \alpha_L(i') \tilde{l}(i'),$$

$$\tilde{p}(i) A_H \alpha_H(i) \tilde{h}(i) = \tilde{p}(i') A_H \alpha_H(i') \tilde{h}(i'),$$

$$\tilde{p}(i) A_M \alpha_M(i) \tilde{m}(i) = \tilde{p}(i') A_M \alpha_M(i') \tilde{m}(i').$$

Combining with the condition of the law of one price for skills, we obtain

$$\tilde{m}(i) = \tilde{m}(i') \text{ for any } i < \tilde{I}_1,$$

$$\tilde{l}(i) = \tilde{l}(i') \text{ for any } \tilde{I}_2 > i > \tilde{I}_1,$$

similarly

$$\tilde{h}(i) = \tilde{h}(i') \text{ for any } i > \tilde{I}_2.$$

3. No arbitrage condition across skills and models: the threshold task $\tilde{I}_1$ can be profitably produced using either models or unskilled workers, and the threshold task $\tilde{I}_2$ can be profitably produced using either skilled or unskilled workers,
$$A_L\alpha_L\left(\tilde{I}_1\right)\tilde{l}\left(\tilde{I}_1\right)=A_M\alpha_M\left(\tilde{I}_1\right)\tilde{m}\left(\tilde{I}_1\right),$$
$$A_L\alpha_L\left(\tilde{I}_2\right)\tilde{l}\left(\tilde{I}_2\right)=A_H\alpha_H\left(\tilde{I}_2\right)\tilde{h}\left(\tilde{I}_2\right).$$
4. Labor market clearing condition:
$$\int_0^I \tilde{l}(i)\,di=L,$$
$$\int_I^1 \tilde{h}(i)\,di=H.$$

With previous conditions, we obtain that
$$\tilde{l}(i)=L/\left(\tilde{I}_2-\tilde{I}_1\right),$$
$$\tilde{h}(i)=H/\left(1-\tilde{I}_2\right).$$

We can prove the proposition 2 by contradiction. If $\tilde{I}_2\le I$, we know that
$$H/\left(1-\tilde{I}_2\right)\le H/(1-I),$$
and
$$L/\left(\tilde{I}_2-\tilde{I}_1\right)>L/I.$$
With the No arbitrage condition at the threshold, we know that
$$A_L\alpha_L\left(\tilde{I}_2\right)\tilde{l}\left(\tilde{I}_2\right)=A_H\alpha_H\left(\tilde{I}_2\right)\tilde{h}\left(\tilde{I}_2\right),$$
i.e.
$$\frac{A_L}{A_H}=\frac{\alpha_H\left(\tilde{I}_2\right)}{\alpha_L\left(\tilde{I}_2\right)}\frac{H/\left(1-\tilde{I}_2\right)}{L/\left(\tilde{I}_2-\tilde{I}_1\right)} \tag{A.1}$$
must hold.

But we already know that
$$A_L\alpha_L(I)L/I=A_H\alpha_H(I)H/(1-I),$$
i.e.
$$\frac{A_L}{A_H}=\frac{\alpha_H(I)}{\alpha_L(I)}\frac{H/(1-I)}{L/I}. \tag{A.2}$$
It is easy to see that
$$\frac{\alpha_H(I)}{\alpha_L(I)}\ge\frac{\alpha_H\left(\tilde{I}_2\right)}{\alpha_L\left(\tilde{I}_2\right)},\text{ and }\frac{H/(1-I)}{L/I}>\frac{H/\left(1-\tilde{I}_2\right)}{L/\left(\tilde{I}_2-\tilde{I}_1\right)}.$$
Then equations (A.1) and (A.2) can not hold at the same time. By contradiction, we know that $\tilde{I}_2>I$. Wage premium is determined by the productivity difference at the threshold
$$\frac{\tilde{W}_H}{\tilde{W}_L}=\frac{A_H\alpha_H\left(\tilde{I}_2\right)}{A_L\alpha_L\left(\tilde{I}_2\right)}>\frac{A_H\alpha_H(I)}{A_L\alpha_L(I)}.$$
In sum, we have proved that a shift of threshold and an increase in wage premium.

A.2 Proof of the Model with Deep Learning AI and LLMs AI Technology

Now we know the task specific productivity profiles are

$$\alpha_L(i) = (1-i)^\eta,$$

$$\alpha_H(i) = i^\eta,$$

$$\alpha_M(i) = \left(1 - C - \frac{a}{N(i)^\alpha} - \frac{b}{G^\beta}\right)^\eta.$$

The competitive equilibrium can be characterized by the following conditions:

1. Law of one price for skills

$$\hat{W}_L = \hat{p}(i) A_L \alpha_L(i) \text{ for any } i < \hat{I}_1,$$

$$F = \hat{p}(i) A_M \alpha_M(i) \text{ for any } \hat{I}_3 > i > \hat{I}_2,$$

$$\hat{W}_H \hat{=} \tilde{p}(i) A_H \alpha_H(i) \quad \text{for any } i > \hat{I}_3 \text{ and } \hat{I}_2 > i > \hat{I}_1.$$

2. Goods market (or task market) clearing condition demands:

$$\hat{p}(i)\hat{y}(i) = \hat{p}(i')\tilde{y}(i') = \hat{Y}.$$

In the four intervals, the following equations are satisfied

$$\hat{p}(i) A_L \alpha_L(i) \hat{l}(i) = \hat{p}(i') A_L \alpha_L(i') \hat{l}(i'),$$

$$\hat{p}(i) A_H \alpha_H(i) \hat{h}(i) = \hat{p}(i') A_H \alpha_H(i') \hat{h}(i'),$$

$$\hat{p}(i) A_M \alpha_M(i) \hat{m}(i) = \hat{p}(i') A_M \alpha_M(i') \hat{m}(i').$$

Combining with the condition of the law of one price for skills, we obtain

$$\hat{m}(i) = \hat{m}(i') \text{ for any } \hat{I}_3 > i > \hat{I}_2,$$

$$\hat{l}(i) = \hat{l}(i') \text{ for any } i < \hat{I}_1,$$

similarly

$$\hat{h}(i) = \hat{h}(i') \text{ for any } i > \hat{I}_3 \text{ and } \hat{I}_2 > i > \hat{I}_1$$

3. No arbitrage condition across skills and models:

$$A_L \alpha_L(\hat{I}_1) \hat{l}(\hat{I}_1) = A_H \alpha_H(\hat{I}_1) \hat{h}(\hat{I}_1),$$

$$A_L \alpha_L(\hat{I}_1) \hat{l}(\hat{I}_1) = A_M \alpha_M(\hat{I}_1) \hat{m}(\hat{I}_1),$$

4. Labor market clearing condition:

$$\int_0^I \hat{l}(i)\,di = L,$$

$$\int_I^1 \hat{h}(i)\,di = H.$$

With previous conditions, we obtain that

$$\hat{l}(i) = L/\hat{I}_1,$$

$$\hat{h}(i) = H/\left(1-\hat{I}_3+\hat{I}_2-\hat{I}_1\right).$$

We can prove the proposition 3 by contradiction. If $\hat{I}_1 \geq I$, we know that

$$H/\left(1-\hat{I}_3+\hat{I}_2-\hat{I}_1\right) > H/(1-I),$$

and

$$L/\hat{I}_1 \leq L/I.$$

With the No arbitrage condition at the threshold, we know that

$$A_L\alpha_L\left(\hat{I}_1\right)\hat{l}\left(\hat{I}_1\right) = A_H\alpha_H\left(\hat{I}_1\right)\hat{h}\left(\hat{I}_1\right),$$

i.e.

$$\frac{A_L}{A_H} = \frac{\alpha_H\left(\hat{I}_1\right)}{\alpha_L\left(\hat{I}_1\right)}\frac{H/\left(1-\hat{I}_3+\hat{I}_2-\hat{I}_1\right)}{L/\hat{I}_1} \tag{A.3}$$

must hold.

But we already know that

$$A_L\alpha_L(I)L/I = A_H\alpha_H(I)H/(1-I),$$

i.e.

$$\frac{A_L}{A_H} = \frac{\alpha_H(I)}{\alpha_L(I)}\frac{H/(1-I)}{L/I}. \tag{A.4}$$

It is easy to see that

$$\frac{\alpha_H(I)}{\alpha_L(I)} \leq \frac{\alpha_H\left(\hat{I}_1\right)}{\alpha_L\left(\hat{I}_1\right)}, \text{ and } \frac{H/(1-I)}{L/I} < \frac{H/\left(1-\hat{I}_3+\hat{I}_2-\hat{I}_1\right)}{L/\hat{I}_1}.$$

Then equations (A.3) and (A.4) can not hold at the same time. By contradiction, we know that $\hat{I}_1 < I$. Wage premium is determined by the productivity difference at the threshold

$$\frac{\hat{W}_H}{\hat{W}_L} = \frac{A_H\alpha_H\left(\hat{I}_1\right)}{A_L\alpha_L\left(\hat{I}_1\right)} < \frac{A_H\alpha_H(I)}{A_L\alpha_L(I)}.$$

In sum, we have proved that a shift of threshold and a decrease in wage premium.

Appendix B Methodology Notes

B.1 LLMs Scoring Details

The scores fall into the following categories:

- E0 indicates no exposure as the LLMs are neither sufficiently useful for the occupation, nor considered a result of the intrinsic nature of the occupation, (e.g., involving physical activities);
- E1 is applied if a 50 reduction in completion time is already feasible with the existing large language model interfaces;
- E2 is applied if such a productivity gain is feasible, but only once the current capabilities of the model can be deployed through applications with further inputs or if it is trained on domain-specific issues or data;
- E3 is applied when the productivity increase would require image processing capabilities in addition to current text processing.

B.2 Expert Rubrics

You are an expert evaluator of the potential for large language models to replace human labor. Large language models are deep learning models used to process and generate natural language text. The latest large language models can generate and describe images and videos based on natural language texts. In this context, you are asked to score each occupation based on whether the occupation could reduce the need for human labor time and participation to achieve the same output or effect in the same amount of time with the aid of large language models.

The scores range from 0 to 1:

- 0 means the occupation cannot reduce human labor input with the help of large language models.
- 0.2 means human labor input could be reduced by 20%.
- 0.4 means human labor input could be reduced by 40%.
- 0.6 means human labor input could be reduced by 60%.
- 0.8 means human labor input could be reduced by 80%.
- 1 means human labor input could be reduced by 100%, meaning the occupation would no longer require human participation.

Your scores represent the potential proportion of human labor input that could be saved by large language models for each occupation. Please score carefully based on the current capabilities of large language models and what you think is possible in the future.

Appendix C Appendix Figures

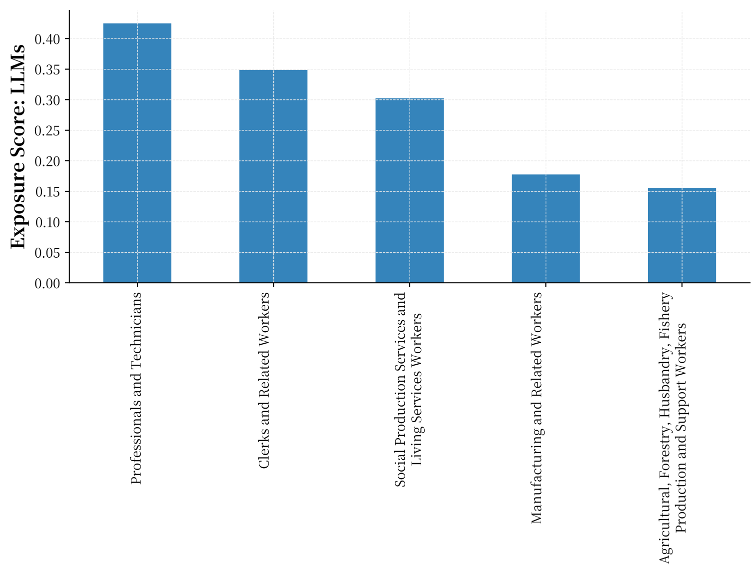

Figure 10: Exposure Score on Large-category Level: LLMs.

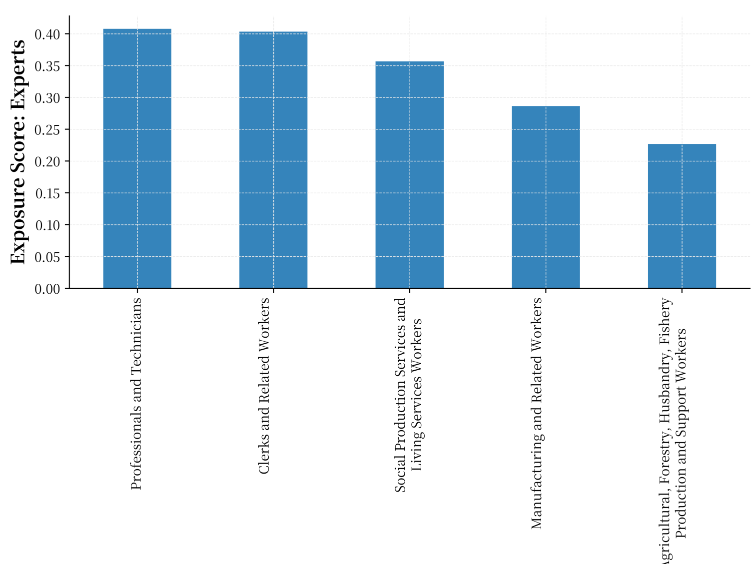

Figure 11: Exposure Score on Large-category Level: Experts.

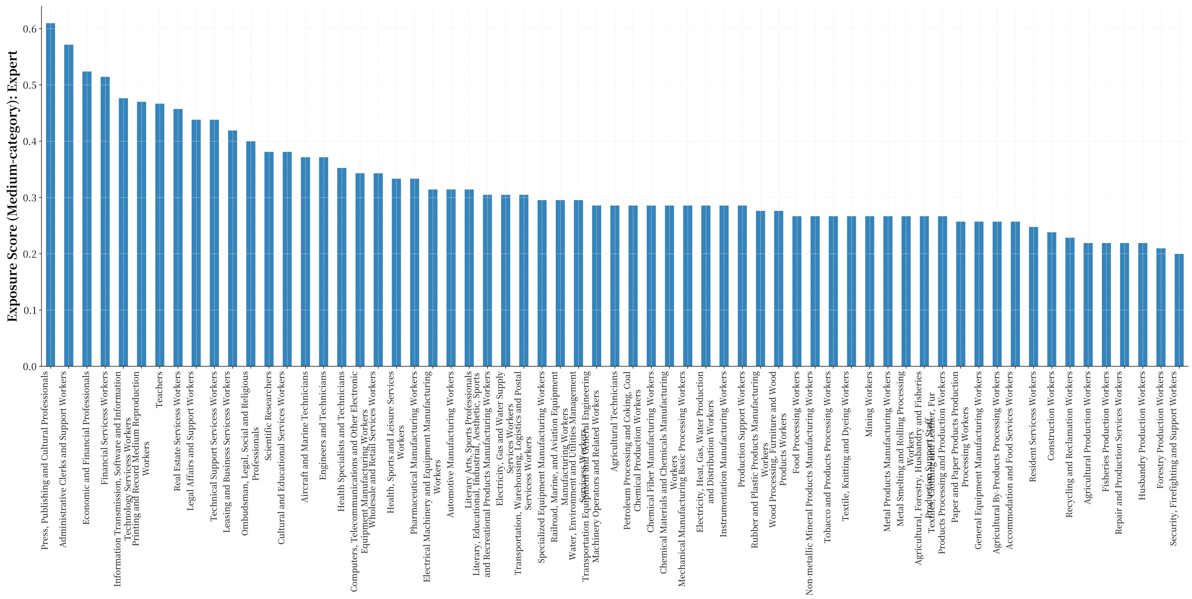

Figure 12: AI Medium-category Occupation Exposure: Expert.

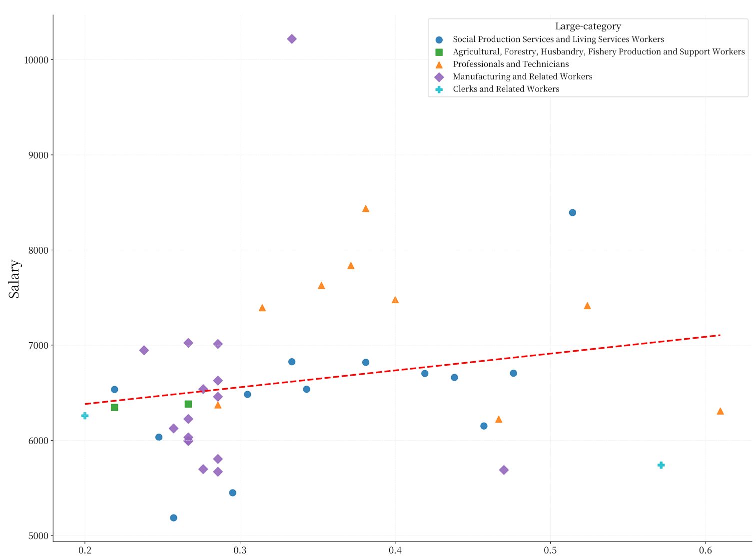

Figure 13: Salary and Exposure Score (Medium-category): Experts.

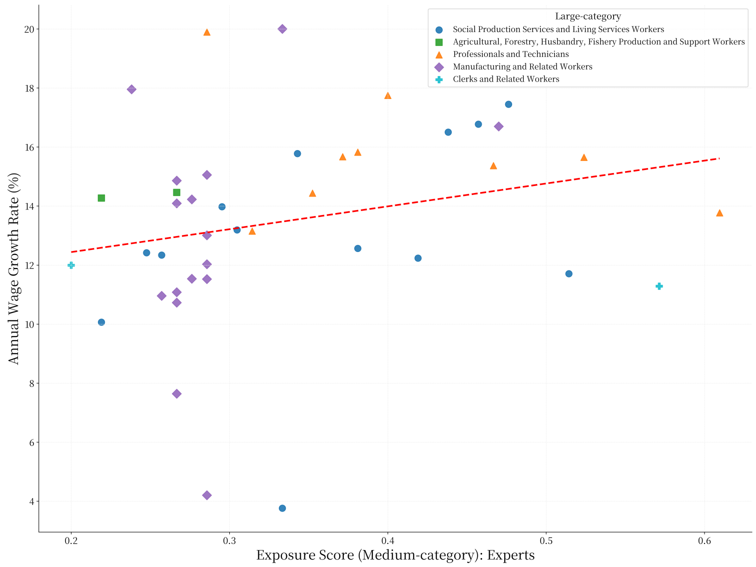

Figure 14: Annual Wage Growth Rate and Exposure Score(Medium-category): Experts.

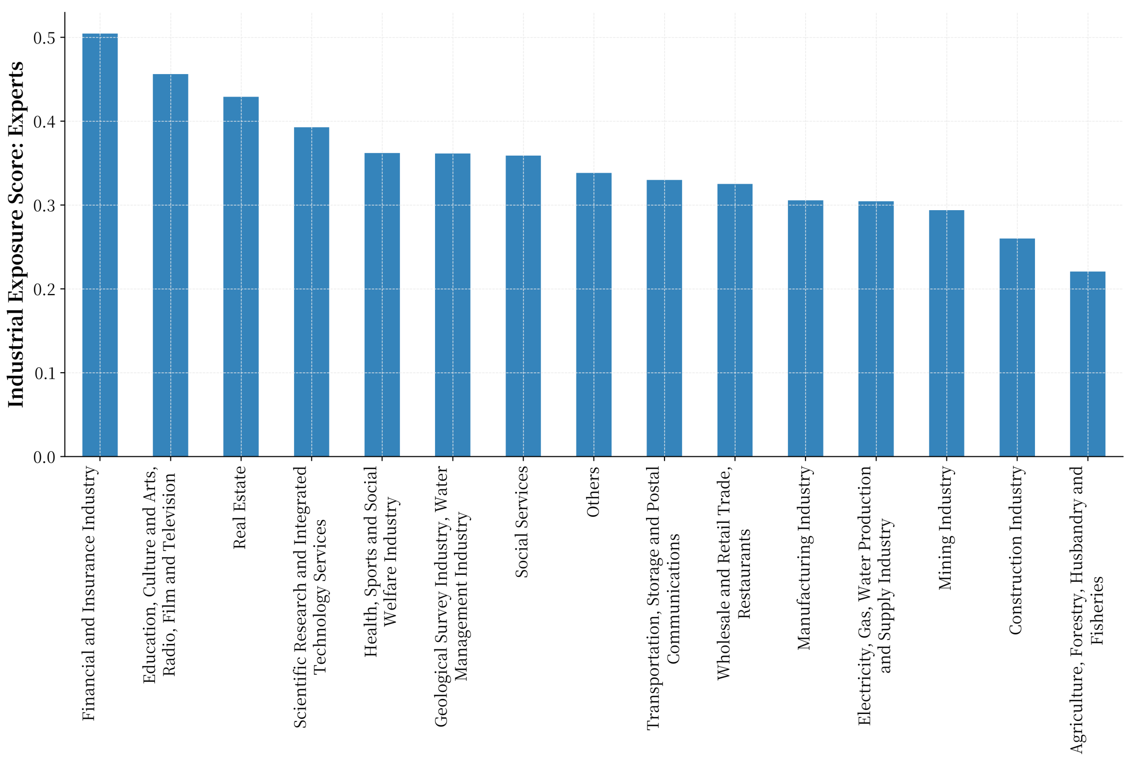

Figure 15: Industrial Exposure Score: Experts.

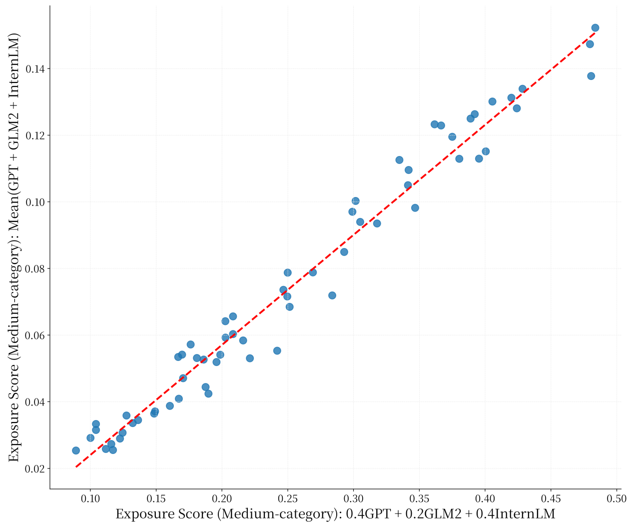

Figure 16: Exposure Scores with Different Weighted Assignments

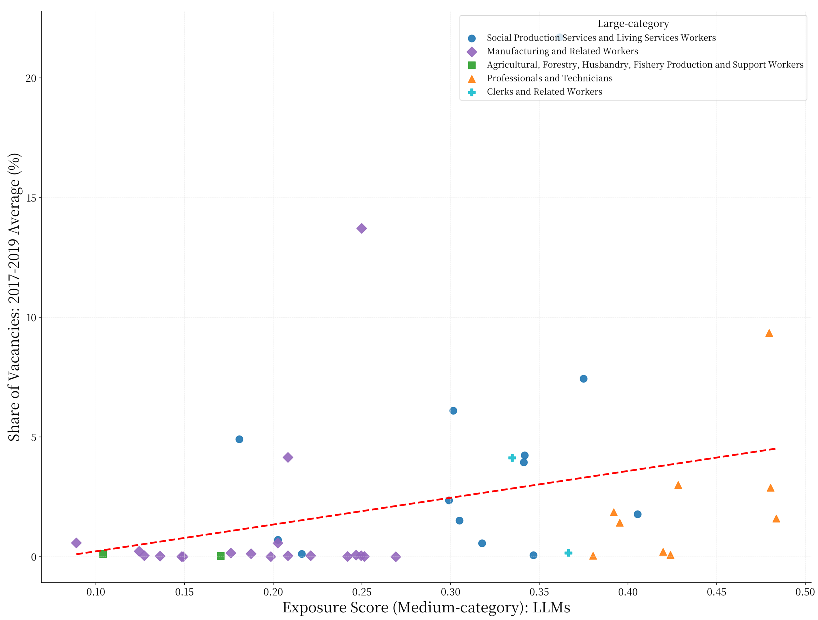

Figure 17: Share of Vacancies and Exposure Score(2017-2019): LLMs.

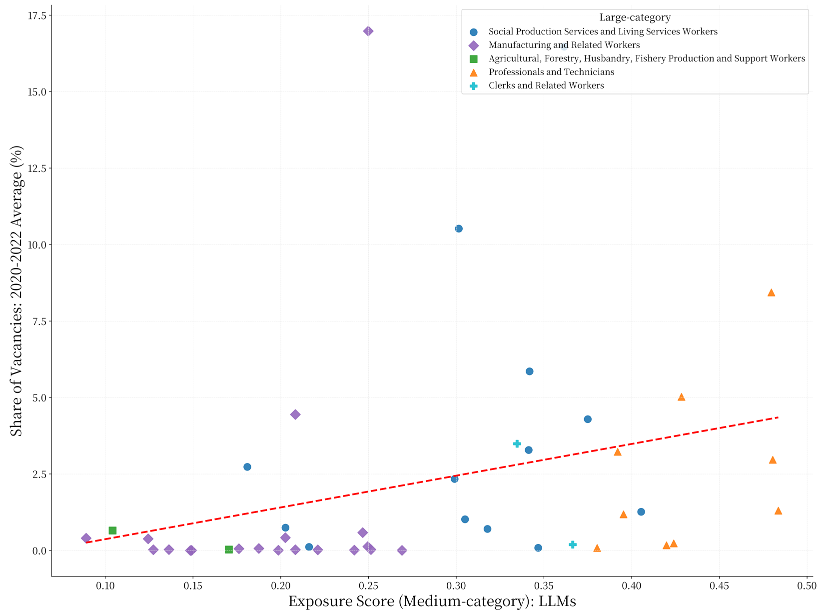

Figure 18: Share of Vacancies and Exposure Score(2020-2022): LLMs.

Appendix D Appendix Tables

Table 2: Full List if Medium-category Occupations of AI Exposure Score

Occupation Code	Medium-category Occupation Title	Expert	GLM	GPT4	InternLM	GLM + InternLM + GPT4
2-01	Scientific Researchers	0.3810	0.6227	0.3472	0.3021	0.4240
2-02	Engineers and Technicians	0.3714	0.5612	0.3600	0.3640	0.4284
2-03	Agricultural Technicians	0.2857	0.5881	0.2308	0.3221	0.3803
2-04	Aircraft and Marine Technicians	0.3714	0.6235	0.0893	0.1384	0.2837
2-05	Health Specialists and Technicians	0.3524	0.6776	0.2778	0.2307	0.3954
2-06	Economic and Financial Professionals	0.5238	0.6673	0.4467	0.3248	0.4796
2-07	Ombudsman, Legal, Social and Religious Professionals	0.4000	0.5505	0.4808	0.2284	0.4199
2-08	Teachers	0.4667	0.8164	0.5313	0.0938	0.4805
2-09	Literary Arts, Sports Professionals	0.3143	0.4573	0.5196	0.1991	0.3920
2-10	Press, Publishing and Cultural Professionals	0.6095	0.6176	0.5179	0.3155	0.4836
3-01	Administrative Clerks and Support Workers	0.5714	0.3196	0.5568	0.1278	0.3348
3-02	Security, Firefighting and Support Workers	0.2000	0.3551	0.4545	0.2898	0.3665
3-03	Legal Affairs and Support Workers	0.4381	0.4583	0.5000	0.2083	0.3889
4-01	Wholesale and Retail Services Workers	0.3429	0.3199	0.5515	0.2132	0.3615
4-02	Transportation, Warehousing, Logistics and Postal Servicess Workers	0.3048	0.3047	0.5160	0.0838	0.3015

4-03	Accommodation and Food Services Workers	0.2571	0.4570	0.4531	0.2148	0.3750
4-04	Information Transmission, Software and Information Technology Servicess Workers	0.4762	0.4074	0.3241	0.2940	0.3418
4-05	Financial Servicess Workers	0.5143	0.4808	0.4327	0.3029	0.4054
4-06	Real Estate Servicess Workers	0.4571	0.3393	0.3929	0.1652	0.2991
4-07	Leasing and Business Servicess Workers	0.4190	0.4724	0.3669	0.1845	0.3413
4-08	Technical Support Servicess Workers	0.4381	0.4199	0.3066	0.1885	0.3050
4-09	Water, Environment and Utilities Management Servicess Workers	0.2952	0.3267	0.1583	0.1229	0.2027
4-10	Resident Servicess Workers	0.2476	0.2885	0.1417	0.1125	0.1809
4-11	Electricity, Gas and Water Supply Services Workers	0.3048	0.6753	0.1458	0.3802	0.4005
4-12	Repair and Production Services Workers	0.2190	0.4203	0.1250	0.1029	0.2161
4-13	Cultural and Educational Services Workers	0.3810	0.5036	0.1685	0.2813	0.3178
4-14	Health, Sports and Leisure Services Workers	0.3333	0.6070	0.1595	0.2737	0.3467
5-01	Agricultural Production Workers	0.2190	0.3259	0.1964	0.0357	0.1860
5-02	Forestry Production Workers	0.2095	0.2054	0.2321	0.0714	0.1696
5-03	Husbandry Production Workers	0.2190	0.1518	0.1429	0.0179	0.1042
5-04	Fisheries Production Workers	0.2190	0.2852	0.0000	0.0625	0.1159
5-05	Agricultural, Forestry, Husbandry and Fisheries Production Support Staff	0.2667	0.3163	0.0650	0.1300	0.1704
6-01	Agricultural By-Products Processing Workers	0.2571	0.3958	0.1750	0.0167	0.1958
6-02	Food Processing Workers	0.2667	0.4807	0.2054	0.0681	0.2514
6-03	Tobacco and Products Processing Workers	0.2667	0.4826	0.3333	0.0625	0.2928

6-04	Textile, Knitting and Dyeing Workers	0.2667	0.2259	0.1193	0.0369	0.1274
6-05	Textiles, Clothing and Leather, Fur Products Processing and Production Workers	0.2667	0.1534	0.0682	0.0455	0.0890
6-06	Wood Processing, Furniture and Wood Products Workers	0.2762	0.1989	0.2727	0.0568	0.1761
6-07	Paper and Paper Products Production Processing Workers	0.2571	0.1979	0.2292	0.0729	0.1667
6-08	Printing and Record Medium Reproduction Workers	0.4700	0.2656	0.2813	0.0781	0.2083
6-09	Literary, Educational, Industrial, Aesthetic, Sports and Recreational Products Manufacturing Workers	0.3048	0.2997	0.0481	0.0609	0.1362
6-10	Petroleum Processing and Coking, Coal Chemical Production Workers	0.2857	0.5016	0.0000	0.0674	0.1897
6-11	Chemical Materials and Chemicals Manufacturing Workers	0.2857	0.5307	0.0000	0.1327	0.2211
6-12	Pharmaceutical Manufacturing Workers	0.3333	0.6215	0.0000	0.1042	0.2419
6-13	Chemical Fiber Manufacturing Workers	0.2857	0.3203	0.0000	0.0313	0.1172
6-14	Rubber and Plastic Products Manufacturing Workers	0.2762	0.4583	0.0000	0.1042	0.1875
6-15	Non-metallic Mineral Products Manufacturing Workers	0.2667	0.2829	0.0174	0.0349	0.1118
6-16	Mining Workers	0.2667	0.2885	0.0000	0.1080	0.1322
6-17	Metal Smelting and Rolling Processing Workers	0.2667	0.3379	0.0660	0.0436	0.1492
6-18	Mechanical Manufacturing Basic Processing Workers	0.2857	0.3448	0.2500	0.0302	0.2083
6-19	Metal Products Manufacturing Workers	0.2667	0.3438	0.0625	0.0391	0.1484

6-20	General Equipment Manufacturing Workers	0.2571	0.2856	0.0000	0.0877	0.1244
6-21	Specialized Equipment Manufacturing Workers	0.2952	0.3006	0.0000	0.0670	0.1225
6-22	Automotive Manufacturing Workers	0.3143	0.3813	0.0000	0.1000	0.1604
6-23	Railroad, Marine, and Aviation Equipment Manufacturing Workers	0.2952	0.1629	0.0000	0.1373	0.1001
6-24	Electrical Machinery and Equipment Manufacturing Workers	0.3143	0.3898	0.0197	0.0921	0.1672
6-25	Computers, Telecommunications and Other Electronic Equipment Manufacturing Workers	0.3429	0.3801	0.1212	0.0947	0.1987
6-26	Instrumentation Manufacturing Workers	0.2857	0.4323	0.1250	0.2500	0.2691
6-27	Recycling and Reclamation Workers	0.2286	0.1250	0.1250	0.0625	0.1042
6-28	Electricity, Heat, Gas, Water Production and Distribution Workers	0.2857	0.4236	0.2407	0.0845	0.2496
6-29	Construction Workers	0.2381	0.2525	0.2072	0.1480	0.2026
6-30	Transportation Equipment and General Engineering Machinery Operators and Related Workers	0.2857	0.3760	0.2438	0.1203	0.2467
6-31	Production Support Workers	0.2857	0.3179	0.2315	0.2002	0.2499

Table 3: List of Industry ID

Ind.	**ID**
Education, Culture and Arts, Radio, Film and Television	1
Scientific Research and Integrated Technology Services	2
Wholesale and Retail Trade, Restaurants	3
Social Services	4

Others	5
Financial and Insurance Industry	6
Transportation, Storage and Postal Communications	7
Real Estate	8
Construction Industry	9
Electricity, Gas, Water Production and Supply Industry	10
Manufacturing Industry	11
Health, Sports and Social Welfare Industry	12
Agriculture, Forestry, Husbandry and Fisheries	13
Mining Industry	14
Geological Survey Industry, Water Management Industry	15

Table 4: Full List of Occupation Share of Industry

Ind. Occ.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Transporta-tion, Warehousing, Logistics and Postal Services Workers	0%	0%	1%	7%	8%	1%	72%	0%	1%	1%	3%	1%	0%	4%	0%
Accommoda-tion and Food Services Workers	0%	0%	21%	10%	5%	1%	0%	1%	0%	1%	1%	0%	0%	0%	5%
Information Transmission, Software and Information Technology Servicess Workers	0%	23%	0%	1%	4%	1%	6%	0%	0%	1%	0%	0%	0%	0%	0%
Repair and Production Services Workers	0%	7%	0%	2%	5%	0%	2%	0%	0%	2%	1%	0%	0%	1%	0%
Agricultural Production Workers	0%	2%	0%	0%	3%	1%	0%	1%	1%	0%	0%	1%	96%	0%	0%

Printing and Record Medium Reproduction Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%
Resident Servicess Workers	2%	8%	0%	14%	7%	0%	0%	0%	0%	1%	1%	2%	0%	0%	0%
Engineers and Technicians	0%	18%	0%	0%	0%	1%	1%	1%	3%	2%	2%	1%	0%	2%	14%
Construction Workers	0%	0%	0%	1%	3%	0%	1%	9%	78%	5%	1%	0%	0%	0%	5%
Wholesale and Retail Services Workers	1%	3%	69%	14%	17%	3%	4%	9%	3%	1%	3%	8%	0%	2%	5%
Technical Support Servicess Workers	3%	2%	0%	0%	2%	0%	0%	1%	0%	0%	1%	0%	0%	0%	18%
Teachers	77%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%	0%	0%	0%
Cultural and Educational Services Workers	1%	3%	0%	1%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Literary Arts, Sports Professionals	4%	2%	0%	0%	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Press, Publishing and Cultural Professionals	2%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%
Ombudsman, Legal, Social and Religious Professionals	0%	0%	0%	5%	1%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Leasing and Business Servicess Workers	1%	2%	1%	11%	9%	1%	1%	5%	0%	4%	1%	1%	0%	2%	0%
Economic and Financial Professionals	2%	12%	2%	4%	5%	26%	2%	11%	1%	1%	3%	2%	0%	1%	5%
Admini-strative Clerks and Support Workers	4%	5%	0%	11%	5%	2%	3%	4%	1%	3%	3%	7%	0%	4%	9%
Health, Sports and Leisure Services Workers	0%	2%	0%	1%	0%	0%	0%	0%	0%	0%	0%	6%	0%	0%	0%

Agricultural, Forestry, Husbandry and Fisheries Production Support Staff	0%	2%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Security, Firefighting and Support Workers	0%	2%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Production Support Workers	0%	3%	0%	0%	4%	0%	1%	0%	2%	17%	8%	1%	0%	8%	0%
Scientific Researchers	0%	3%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Computers, Telecommunications and Other Electronic Equipment Manufacturing Workers	0%	2%	0%	0%	2%	0%	0%	0%	0%	1%	6%	0%	0%	0%	0%
Recycling and Reclamation Workers	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	5%
Agricultural By-Products Processing Workers	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Literary, Educational, Industrial, Aesthetic, Sports and Recreational Products Manufacturing Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	4%	0%	0%	0%	0%
Wood Processing, Furniture and Wood Products Workers	0%	0%	0%	0%	1%	0%	0%	0%	2%	0%	4%	0%	0%	0%	0%
Mechanical Manufacturing Basic Processing Workers	0%	0%	0%	0%	1%	0%	0%	0%	1%	1%	6%	0%	0%	0%	0%
Rubber and Plastic Products Manufacturing Workers	0%	0%	0%	0%	1%	0%	0%	0%	0%	1%	2%	0%	0%	0%	0%

Water, Environment and Utilities Management Servicess Workers	0%	0%	1%	10%	3%	0%	0%	0%	0%	1%	0%	20%	0%	0%	14%
Textile, Knitting and Dyeing Workers	0%	0%	0%	0%	1%	0%	0%	0%	0%	1%	4%	0%	0%	0%	0%
Textiles, Clothing and Leather, Fur Products Processing and Production Workers	0%	0%	0%	0%	1%	0%	0%	0%	0%	0%	13%	0%	0%	0%	0%
Metal Products Manufacturing Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%	0%	0%	0%	0%
Non-metallic Mineral Products Manufacturing Workers	0%	0%	0%	0%	1%	0%	0%	0%	2%	0%	6%	0%	0%	0%	0%
Food Processing Workers	0%	0%	1%	0%	1%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%
Health Specialists and Technicians	0%	0%	0%	1%	0%	0%	0%	0%	0%	0%	0%	41%	0%	1%	0%
Real Estate Servicess Workers	0%	0%	0%	4%	0%	0%	0%	56%	0%	0%	0%	0%	0%	0%	0%
Forestry Production Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Electricity, Gas and Water Supply Services Workers	0%	0%	0%	1%	0%	0%	1%	0%	0%	19%	0%	0%	0%	0%	14%
Paper and Paper Products Production Processing Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%
Financial Servicess Workers	0%	0%	0%	0%	1%	63%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Specialized Equipment Manufacturing Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%

Instrumenta-tion Manufacturing Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Agricultural Technicians	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Chemical Materials and Chemicals Manufacturing Workers	0%	0%	0%	0%	2%	0%	0%	0%	0%	1%	1%	0%	0%	0%	0%
Pharmaceutical Manufacturing Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%
Electricity, Heat, Gas, Water Production and Distribution Workers	0%	0%	0%	0%	0%	0%	0%	1%	0%	30%	0%	0%	0%	0%	5%
Electrical Machinery and Equipment Manufacturing Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%	0%	0%	0%	0%
Petroleum Processing and Coking, Coal Chemical Production Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	0%	0%
Transporta-tion Equipment and General Engineering Machinery Operators and Related Workers	0%	0%	0%	0%	0%	0%	3%	0%	2%	0%	1%	0%	0%	5%	0%
Mining Workers	0%	0%	0%	0%	1%	0%	0%	0%	0%	5%	0%	0%	0%	68%	0%
Metal Smelting and Rolling Processing Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	3%	0%	0%	0%	5%
Railroad, Marine, and Aviation Equipment Manufacturing Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%
Aircraft and Marine Technicians	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

General Equipment Manufacturing Workers	0%	0%	0%	0%	0%	0%	0%	0%	1%	1%	1%	0%	0%	0%	0%
Chemical Fiber Manufacturing Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Automotive Manufacturing Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	0%	0%
Tobacco and Products Processing Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Fisheries Production Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%
Husbandry Production Workers	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%	0%

Table 5: Proportion of Occupations with ¿50% Consistent Scores for Different Models

Model	**Proportion**
GPT	83.0%
GLM2	64.4%
InternLM	83.2%

Appendix E Data Appendix

In this Appendix we provide further description on the corpus of online job vacancy postings collected by the City Data Group.

Appendix E.1 Web Sources and Data Provider

Our corpus of online job vacancy postings is provided by the labour market and analytics company 'The City Data Group'. The City Data Group has been scraping online job vacancy postings in China since January 2015. Each job vacancy posting is scraped from the internet, including major online job market platforms in China such as zhaoping.com, 51job, 58.com, Ganji.com, Lagou.com, and Kanzhun. Lightcast actively audits their list of web sources to ensure data from new websites is on-boarded in a timely manner. One of the main competitive advantages of the City Data Group's data product is the breadth of their sources. These data are often referred to in the literature as the 'near universe' of online job vacancy postings.

Appendix E.2 Recruitment Data Processing Workflow

Once an online job vacancy posting is scraped, the City Data Group processes this data to produce the online posting dataset. A description the data processing workflow are: When aggregating raw recruitment data into standardized formats, we processed different fields in distinct ways. Fields Stored in Their Original or Nearly Original Forms 1. Company Name; 2. Industry; if it is stated. 3. Benefits; 4. Job Title: Stored as a string. 5. Company Size (Lower and Upper Limits): measured in the number of employees. For instance, if a company size is listed as "10-99 employees," enscale_lower is 10, and enscale_upper is 99. 6. Experience Requirements: measured as the minimum required years. 7. Recruitment Numbers. 8. Job Descriptions: Stored as a string with full job descriptions, including: iob Responsibilities, job details, working conditions, etc.

Appendix E.3 Errors Checking and Missing Information

Overall, the data product is a highly informative and accurate product, but we also acknowledge that any dataset with hundreds of millions of observations scraped from different sources will never be perfect. Both the structured data and the plain text data require a number of pre-processing steps and the use of algorithmic feature extraction, which in a very small number of cases produce errors (e.g. misclassification of occupations, truncation of plain text, presence of erroneous text). In the following we highlight some of the errors we have encountered, and discuss the strategies we employed to ensure our results remain robust to such issues. Missing Values A relevant value (e.g. the educational requirement for a job, the salary for a job) might be missing for at least two reasons: (i) the employer does not mention this explicitly in the text of the job ad, and (ii) the algorithm used to extract this feature from the text failed. By our double checking, missing values are almost entirely due to lack of information, and not poor feature extraction. Erroneous Plain Text In a very small number of cases we observe that the plain text includes some parts of the website other than the job description. We exclude these erroneous pain text. Duplicate Recruitment Ads In some cases, a recruitment advertisement will be released duplicated within the same platform or across platform. We define duplicate job adverts as job adverts from the same firm recruiting the same type of employee are repeated during the thirty-day window period. In our empirical analysis, esp. counting vacancy number, we consider this concern by adjusting duplicate recruitment ads.

Appendix E.4 Representativeness of Online Job Vacancy Postings

The City Data Group frequently reviews the representativeness of the job vacancy postings it scrapes, to ensure the information renders an accurate picture at least the online recruitment labor market. Obviously, the online recruitment labor market does not equal to the whole labor market; to gauge the nuance difference between online recruiting labor market and the overall labor market, the one precondition is a representative aggregate recruitment dataset. Since in China there is no representative recruitment dataset such as JOLTS in United States, therefore, we cannot address the issue of representative of online job recruitment.