Developing an occupational prestige scale using Large Language Models

Robert de Vries
University of Kent

Mark Hill
King's College London

Laura Ruis
University College London

Abstract

Large Language Models (LLMs), being trained on fractions of all online text, reflect societal biases and stereotypes – such as racial and gender biases. One key bias, occupational prestige, plays a significant role in determining social status, which can profoundly impact a person's well-being as well as their mental and physical health. In this paper, we propose a method of using LLMs to capture societal perceptions of occupational prestige. We create four occupational prestige scales, with each tapping a difference facet of prestige perceptions. These scales are validated against existing prestige scales based on human data. We conclude that it is possible to create valid measures of occupational prestige by prompting a commercially available LLM – though with some important limitations.

1 Introduction

A person's social status is their position on the axis along which they are ordered by subjective esteem and respect (Chan and Goldthorpe, 2007). Measures of social status tap into who is looked *up* to in society, and who is looked *down* on. Our social status contributes to our overall well-being, and relative status deprivation has been shown to significantly impact mental and physical health (Mishra and Carleton, 2015). Further, status shows a strong relationship with cultural outcomes, such as political attitudes (Chan and Goldthorpe, 2007). Social status may be determined by a variety of factors, including consumption patterns and behaviour (Waters and Waters, 2010). One particularly important aspect is the prestige accorded to one's occupation. As Chan (2010) notes, "in modern societies occupation is one of the most salient positional characteristics to which status attaches" (p.29). A person's occupation is one of the first personal details we learn about them, helping us to socially 'place' them relative to ourselves and others (Weeks and Leavitt, 2017).

These considerations underscore the necessity of measuring occupational prestige in society. Within existing research, there are two principal approaches to measuring occupational prestige. The first is to ask survey respondents to rank or score a list of occupations according to their prestige. The second is through 'social distance' measures derived from patterns of friendship or marriage. The underlying assumption is that occupations that often marry or befriend each other (i.e. are socially proximate) occupy similar positions in the social hierarchy. Both approaches have important limitations. Firstly, humans respondents cannot be expected to reliably rank or score hundreds or thousands of occupational titles, which motivates survey creators to group titles. In the most widely used occupational taxonomies (including ISCO, the US Census occupational classification system, and the UK Standard Occupational Classification [SOC] scheme), a Hollywood film producer and a local community theatre director would both class under the heading of 'Theatre and film producers and directors', despite the vast gulf in prestige between the two. Measuring occupational prestige implicitly through social distance markers overcomes this limitation, but has issues of its own, of which the most important is construct validity concerns. It is not clear to what extent social distance measures are capturing occupational esteem (Bukodi et al., 2011; Bihagen and Lambert, 2018).

Socially Responsible Language Modelling Research (SoLaR) Workshop at NeurIPS 2024.

In this paper, we propose a complementary method for estimating perceptions of occupational prestige using large language models (LLMs). LLMs have been shown to reflect societal biases, such as racial or gender biases (Kotek et al., 2023; Lee et al., 2024). For example, doctors are more often assumed to be men, while nurses tend to be women (Kirk et al., 2024), and violence is more strongly associated with African American women than European American women (An et al., 2023). These biases are most often researched in context of their problematic consequences, however, they also offer a window into human biases and stereotypes. In this work, we endeavour to capture the 'prestige bias' associated with occupations. Using LLMs to this end has clear benefits; LLMs can be iterrogated at extremely low cost compared to human participants, providing flexibility to address questions that would be prohibitively expensive with humans. For example, we can rank occupations at a fine-grained level, comparing occupational titles that interact with other demographic categories such as race and gender.

We extract prestige rankings over occupations from an LLM, and make the following observations:

- GPT-3 derived prestige scales correspond closely with existing scales derived from human participant data.
- Prompts based on related concepts (social attraction and job training requirements and pay) produce scales that are related to the prestige scale in intuitively sensible ways - both in cases where they correlate, as well as where they differ.
- Occupational prestige scales captured from GPT-3 are robust to prompt order effects and to the use of close synonyms (substituting 'high-status' for 'prestigious').
- GPT-3 derived prestige scale differentiate between closely related occupations that are not distinguished by existing prestige scales. For example, GPT-3 holds that 'waiter' is a substantially more prestigious occupation than 'waitress'.

Gmyrek et al. (2024) are the first to show such occupational prestige rankings can be extracted from LLMs, however they do not compare to related concepts such as social attraction, and do not investigate interaction of occupation with gender. Our findings show the potential for using LLMs to measure occupational prestige as a *complementary* source of data when measuring societal biases, allowing more fine-grained and interactional investigations. However, there are important limitations to the ability of LLMs to replace human participants. We end this paper with a discussion on one of these limitations, and present others in the social impact statement in Appendix A.

2 Method

In this section, we detail the data, model, and prompts we use, as well as how we extract responses.

Prompts and model. We design four different prompts to investigate prestige bias present in a version of GPT-3 from which we can extract likelihood scores, namely text-davinci-003[1]. The exact prompts are shown in Table 1. The first two prompts differ only in the word used to describe prestige ("highest status" versus "most prestigious"). Substantially different results from these two would suggest the model is problematically sensitive to the use of near-synonyms. The third prompt is meant to capture a slightly different conception of prestige from literature. This prompt looks at social attraction, assuming people are socially drawn to high status individuals, and tend to advertise their proximity to them as a way to increase their own status (Adler and Adler, 1998; Berger and Dijkstra, 2013). There should be a significant correlation between prompts 1 through 3, because they all focus on the underlying concept of status. However, some divergence is also expected. For example, some occupations that may be particularly 'boast-worthy' due to their rarity, or proximity to fame – such as professional athlete or TV producer – may fall behind traditional occupations such as lawyer and doctor under a more holistic conception of prestige. We include the final prompt as a test of validity of our rankings. A prominent criticism of existing survey-based measures of occupational prestige is that they primarily capture 'desirability' in a general sense (based principally on perceived job requirements and rewards), rather than esteem (Bukodi et al., 2011). The fourth prompt is meant to tap into desirability, and allows us to examine the extent to which the other prompts capture distinct concepts.

[1]This model (text-davinci-003) was deprecated by OpenAI in January 2024 and is no longer accessible.

Table 1: Prompts used to extract occupational scales from text-davinci-3.

No.	Prompt
1	There are two occupations: 1) [OCCUPATION 1]. 2) [OCCUPATION 2]. Answer only 1 or 2. The highest status is generally perceived to be: Answer: [1/2]
2	There are two occupations: 1) [OCCUPATION 1]. 2) [OCCUPATION 2]. Answer only 1 or 2. The most prestigious is generally perceived to be: Answer: [1/2]
3	A social event has happened. Alex has had the opportunity to chat to a variety of different kinds of people at this event, including 1) a [OCCUPATION 1], and 2) [OCCUPATION 2]. Answer only 1 or 2. The one Alex is most likely to boast about making friends with is: Answer: [1/2]
4	There are two occupations: 1) [OCCUPATION 1]. 2) [OCCUPATION 2]. Answer only 1 or 2. The job that performs best on the criteria of pay and how much education or training it requires is: Answer: [1/2]

Occupations. We construct a list of 99 occupations, chosen to broadly represent all 43 sub-major groups present in ISCO-08 [2]. We differentiate some occupations which share the same code across multiple occupational coding schemes to determine whether the model recognised prestige differences which are not captured by existing scales (for example, 'waiter' and 'waitress'). We translate formal occupation titles to colloquial terms (for example, 'paramedical practitioner' was translated to 'paramedic'). We render these terms in US English vernacular due to the prevalence of American English in model training data (for example, 'College' rather than 'University' professor, and 'Grocery store' rather than 'Fruit and Veg Shop' owner). The full list of occupations is given in Appendix B.

Extracting model responses. We pair each occupation against every other occupation, yielding a total of 9702 unique matches per prompt in Table 1. We set the temperature to 0. To deal with position bias, we run each prompt while swapping occupations and orders, yielding four matches per pair of occupations. Using these matches, we create a ranking based on the total number of 'wins'. We make a distinction between clear wins and close wins based on assigned likelihoods. Close wins are those where the probability associated with the winning occupation is less than double that of the losing occupation - these obtain 1 point. Clear wins are those where there is a winner:loser ratio of double or more, achieving 2 points. This yielded a score ranging from 0 to 392 (196 matches per occupation, with a potential to earn 2 points per match). Ties in the resulting ranking are resolved based on direct matches within tied ranks. For example, within the first permutation of the first prompt, 'Chartered public accountant' and 'Architect' are tied for rank 9. However, when directly contrasted, the response associated with 'Architect' has a higher probability. Hence 'Architect' was ranked higher in the amended rank order. Using this approach did not break all ties due to non-transitivity.

3 Results

The full produced rankings for each prompt are given in Appendix C. In this section, we discuss the most notable results. Below, when we speak of correlation, we mean Pearson's R correlation.

Position bias. We tested for order effects by examining the correlation between the rankings produced for the two occupation order permutations within each prompt. For each prompt, the correlation between the two ranks produced by swapping the occupations is above 0.99, suggesting that there are no strong position biases regardless of which occupation was labelled (1) or (2).

Comparing rankings produced by different prompts. The correlation between the rankings produced by the first and second prompt is above 0.99, suggesting the model is robust to a close synonym for prestige. We combine the results from both those prompts in our analyses.

[2] `https://www.ilo.org/`

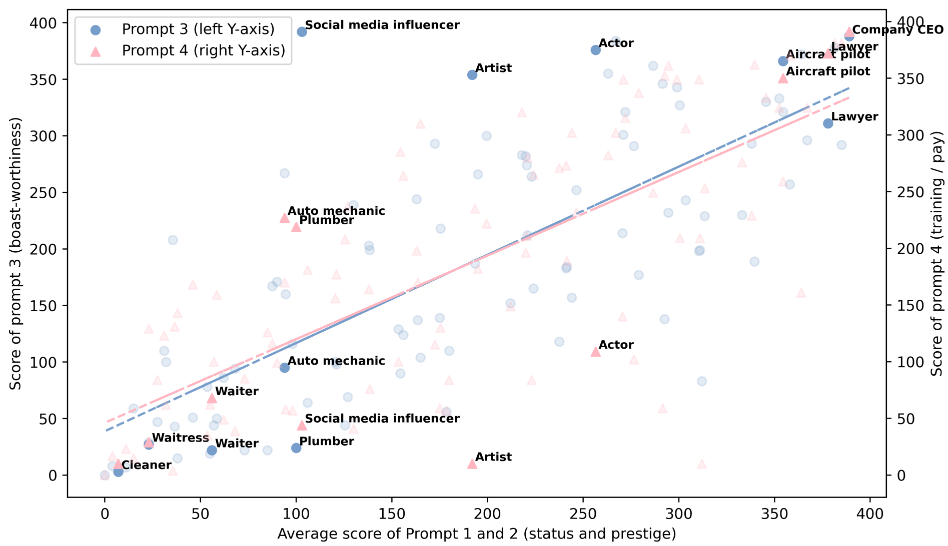

Figure 1: Relations between the scales produced by our three prompts, including trend lines. We highlight the points on the scales that represent specific occupations we discuss in the text below.

Table 2: Correlations between occupation scores for each prompt

Prompt	**P1/2**	**P3**	**P4**	**SIOPS**	**CAMSIS**	**SEI**
P1/2	1.00	-	-	-	-	-
P3	0.76	1.00	-	-	-	-
P4	0.73	0.48	1.00	-	-	-
SIOPS	0.85	0.61	0.66	1.00	-	-
CAMSIS	0.79	0.66	0.49	0.83	1.00	-
SEI	0.81	0.55	0.65	0.82	0.86	1.00

In Table 2 the correlation between the rankings produced by different prompts as well as existing external rankings are presented. We first discuss the rankings produced by our own prompts. P1/2 (status/prestige) correlates strongly with P3 and P4 that are meant to investigate boast-worthiness, and training and pay of occupations. The boast-worthiness score correlates only moderately with the training/pay prompt suggesting that this prompt may capture a particular facet of prestige that is more strongly orthogonal to training/pay than either is to prestige/status.

Figure 1 shows the relationship between the different scales. The top right shows a cluster of occupations which are prestigious, highly trained, well paid, and socially 'impressive', like Company CEO, pilot, and lawyer. Similarly, the bottom left portion holds a cluster of occupations which are considered low-status, poorly paid/trained, and potentially socially unimpressive like 'Cleaner'. The areas of departure from the trend lines are illustrative. A number of occupations perform better on the scale derived from P4 (training/pay) than would be predicted from their scores on the status/prestige scale from P1/2 (occupations that can be found above the trend line for P4, but below for P3). These include occupations like 'Auto mechanic', and 'Plumber', which are intuitively lower status, but higher-paid. The opposite is true for professions like 'Social media influencer', 'Artist', and 'Actor', which have a high boast-worthiness factor but require less training and potentially obtain lower pay. A notable contrast between specific occupations is the relative position of waiters and waitresses. Standard occupational coding schemes do not distinguish occupations based on gender and do not allow for comparisons between equivalent occupations held by men versus women. Our results show that 'Waiter' ranks above 'Waitress' on the scale derived from P1/2 (status/prestige) and P4 (Training/pay). We make further observations on the properties of the rankings in Appendix D.

Comparing to existing occupational prestige scales. We compare to the two most widely used occupational prestige rankings for external validation. Specifically we look at the Standard International Occupational Prestige Scale (SIOPS) and Cambridge Social Interaction Scale (CAMSIS). The latter is based on patterns of friendship or marriage between occupations. The former collates similar occupational prestige data from 85 individual surveys conducted in 55 countries. Additionally, we compare to the SEI scale (Hauser and Warren, 1997), a weighted sum of the education levels and earnings of occupational incumbents – facilitating comparison with P4. All of the external scales correlate very strongly with each other (r>0.82), as has been shown by previous research (Bihagen and Lambert, 2018). Encouragingly, all the external scales also consistently correlate with our general status/prestige scale (r>0.79). This is strong evidence that our primary measure of status/prestige is capturing a similar underlying concept. The scales derived from boast-worthiness and training/pay correlate less strongly with all three external scales. P3 correlates strongly with SIOPS and (particularly) CAMSIS, but only moderately with SEI. This is consistent with our interpretation of P3 as capturing a facet of prestige based on social 'impressiveness' that is somewhat orthogonal to job requirements and rewards. Finally, P4 for training/pay correlates less strongly with SEI than with our own prompt for status/prestige. This is notable, because we specifically intended P4 to capture perceived job requirements and earnings. One potential explanation for this result is the difference between actual average pay and education levels as captured by SEI and public stereotypes as captured by the associations present in GPT-3's training data. Further inspection of the deviations supports this conclusion, and we discuss in Appendix D.

4 Discussion

Our overall interpretation of the above results is that our LLM-derived scales are, to a large extent, valid measures of different facets of perceived occupational prestige. This is supported by the pattern of relationships within the LLM-derived scales, and the pattern between our scales and external measures of occupational prestige and socio-economic advantage. It is not our intention in this paper to put forward the scales described above as 'finished products' for use in future research. Rather, we intend our results to serve as a proof-of-concept – a demonstration that potentially valid measures of occupational prestige can be extracted by prompting commercial LLMs. There are, however, important limitations to bear in mind when interprating our results. The most obvious concern is representativeness. The data used for post-training stages is supplied by annotators that are mostly young, White, or Southeast Asian college graduates in the US (Santurkar et al., 2023). It has been shown that these models strongly reflect the political views of younger, more educated people from North America and Europe, and consequently under-represent the views of groups who may be less likely to write online (Santurkar et al., 2023; Durmus et al., 2024). While the issue of representativeness is important for any research which attempt to use LLMs as a window into human attitude, it is likely to be less problematic in research on occupational prestige. First, occupational prestige ratings tend to be consistent over time and across cultures (Treiman, 1977), and differences are subtle compared with differences in political and social attitudes (Zhou, 2005). Further, LLM training data coming from interacting sources like formal media and other cultural sources like books, combined with informal internet text from sources such as blogs, is ideal for tapping into broad societal biases which are often implicit (Hellman, 2008; Caliskan et al., 2017). Occupational prestige biases are likely to be more similar to these broad societal biases than to specific social and political attitudes such as environmental beliefs and attitudes (Sanders et al., 2023; Fell, 2024). There are further limitations, which we touch on in Appendix A, and LLM data on societal biases should always be complementary to approaches using human data.

We are encouraged by the strong correspondence between our LLM-derived scales and existing external prestige scales. If LLM-based approaches produce robust, reliable, and valid measures of occupational prestige, this opens up considerable scope for progress along a number of avenues of social status research. In particular, previous research has struggled to isolate exactly what is being captured by existing prestige measures (Bukodi et al., 2011), and if indeed such measures can meaningfully distinguish occupational prestige from generalised socio-economic advantage (Bihagen and Lambert, 2018; Goldthorpe, 2021). The flexibility of LLM-derived scales allows future research to quickly produce a variety of different scales based on a variety of potential facets of prestige and to examine their relationship with each other, and with 'holistic' measures of prestige. The same flexibility also allows for the potential to examine the intersection of occupational titles and demographic characteristics in the determination of prestige. We have made a preliminary attempt to both in this paper, and we are excited about the potential of using LLMs more broadly in social status and sociology research.

References

Adler, P. and Adler, P. (1998). *Peer Power: Preadolescent Culture and Identity*. Peer Power: Preadolescent Culture and Identity. Rutgers University Press.

An, H., Li, Z., Zhao, J., and Rudinger, R. (2023). SODAPOP: Open-ended discovery of social biases in social commonsense reasoning models. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1573–1596, Dubrovnik, Croatia. Association for Computational Linguistics.

Berger, C. and Dijkstra, J. (2013). Competition, envy, or snobbism? how popularity and friendships shape antipathy networks of adolescents. *Journal of Research on Adolescence*, 23(3):586–595.

Bihagen, E. and Lambert, P. (2018). Can class and status really be disentangled? *Research in Social Stratification and Mobility*, 58:1–10.

Bisbee, J., Clinton, J. D., Dorff, C., Kenkel, B., and Larson, J. M. (2024). Synthetic replacements for human survey data? the perils of large language models. *Political Analysis*, page 1–16.

Bukodi, E., Dex, S., and Goldthorpe, J. (2011). The conceptualisation and measurement of occupational hierarchies: a review, a proposal and some illustrative analyses. *Quality & Quantity: International Journal of Methodology*, 45(3):623–639.

Caliskan, A., Bryson, J. J., and Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334):183–186.

Chan, T. W., editor (2010). *Social Status and Cultural Consumption*. Cambridge University Press.

Chan, T. W. and Goldthorpe, J. H. (2007). Class and status: The conceptual distinction and its empirical relevance. *American Sociological Review*, 72(4):512–532.

Durmus, E., Nguyen, K., Liao, T. I., Schiefer, N., Askell, A., Bakhtin, A., Chen, C., Hatfield-Dodds, Z., Hernandez, D., Joseph, N., Lovitt, L., McCandlish, S., Sikder, O., Tamkin, A., Thamkul, J., Kaplan, J., Clark, J., and Ganguli, D. (2024). Towards measuring the representation of subjective global opinions in language models.

Fell, M. (2024). Energy social surveys replicated with large language model agents. *SSRN Electronic Journal*.

Gallegos, I. O., Rossi, R. A., Barrow, J., Tanjim, M. M., Kim, S., Dernoncourt, F., and Ahmed, N. K. (2024). Bias and fairness in large language models: A survey. *Computational Linguistics*, pages 1–79.

Gmyrek, P., Lutz, C., and Newlands, G. (2024). A technological construction of society: Comparing gpt-4 and human respondents for occupational evaluation in the uk. *British Journal of Industrial Relations*, n/a(n/a).

Goldthorpe, J. H. (2021). Class and status in interwar england: Current issues in the light of a historical case. *The British Journal of Sociology*, 72(2):239–251.

Hauser, R. M. and Warren, J. R. (1997). 4. socioeconomic indexes for occupations: A review, update, and critique. *Sociological Methodology*, 27(1):177–298.

Hellman, D. (2008). *When Is Discrimination Wrong?* Harvard University Press.

Kirk, H. R., Jun, Y., Iqbal, H., Benussi, E., Volpin, F., Dreyer, F. A., Shtedritski, A., and Asano, Y. M. (2024). Bias out-of-the-box: an empirical analysis of intersectional occupational biases in popular generative language models. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, NIPS '21, Red Hook, NY, USA. Curran Associates Inc.

Kotek, H., Dockum, R., and Sun, D. (2023). Gender bias and stereotypes in large language models. In *Proceedings of The ACM Collective Intelligence Conference*, CI '23, page 12–24, New York, NY, USA. Association for Computing Machinery.

Lee, M. H., Montgomery, J. M., and Lai, C. K. (2024). Large language models portray socially subordinate groups as more homogeneous, consistent with a bias observed in humans. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '24, page 1321–1340, New York, NY, USA. Association for Computing Machinery.

Lee-Ann Ewing, M. E. and Cooper, H. (2021). From bad to worse: the negative and deteriorating portrayal of teachers on screen. *Teachers and Teaching*, 27(6):506–519.

Mishra, S. and Carleton, R. N. (2015). Subjective relative deprivation is associated with poorer physical and mental health. *Social Science Medicine*, 147:144–149.

Sanders, N. E., Ulinich, A., and Schneier, B. (2023). Demonstrations of the Potential of AI-based Political Issue Polling. *Harvard Data Science Review*, 5(4). https://hdsr.mitpress.mit.edu/pub/dm2hrtx0.

Santurkar, S., Durmus, E., Ladhak, F., Lee, C., Liang, P., and Hashimoto, T. (2023). Whose opinions do language models reflect? In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org.

Treiman, D. J. (1977). *Occupational Prestige in Comparative Perspective*. Academic Press, New York et al.

US Bureau of Labor Statistics (2021). Occupational employment and wages, may 2021: 11-1011 chief executives. Available at: `https://www.bls.gov/oes/2021/may/oes111011.htm` [Accessed: 21-Jun-2024].

Wang, A., Morgenstern, J., and Dickerson, J. P. (2024). Large language models cannot replace human participants because they cannot portray identity groups.

Waters, T. and Waters, D. (2010). The new zeppelin university translation of weber's 'class, status, party'. *Journal of Classical Sociology*, 10(2):153–158.

Weeks, M. and Leavitt, P. A. (2017). Using occupational titles to convey an individual's location in social stratification dimensions. *Basic and Applied Social Psychology*, 39(6):342–357.

Zhou, X. (2005). The institutional logic of occupational prestige ranking: Reconceptualization and reanalyses1. *American Journal of Sociology*, 111:90 – 140.

A Social Impact Statement

Though we are excited about the potential of using LLMs to interrogate the occupational prestige bias present in society, there are important limitations that need to be kept in mind. The first, of representativeness, is discussed in more detail in the discussion section of the main paper. In short, because post-training data predominantly comes from young, White, or Southeast Asian college graduates in the US (Santurkar et al., 2023), LLMs may reflect the views of these groups more strongly. Recent work has shown that methods prompting models to act more like certain demographics vastly underestimate human variation in opinion, as well as socio-demographic differences in opinion (Bisbee et al., 2024; Durmus et al., 2024; Wang et al., 2024). Furthermore, a human's sense of prestige of a particular occupation derives from a variety of sources, including media and everyday social interactions (Treiman, 1977; Lee-Ann Ewing and Cooper, 2021). The LLM we used in our experiments can only derive from online text. Finally, research shows that LLMs can exhibit social biases (for example, sexist biases) more strongly than the underlying population (Gallegos et al., 2024), potentially because online text contains more stereotypically sexist depictions of women than are reflected in real everyday experience and attitudes. This may in turn impact the biases extracted from LLMs. For these reasons, using LLMs to probe societal biases should always be complementary to other approaches using human data.

B Full list of occupations

See Table 3.

Table 3: Occupations

Actor	HR officer	Retail sales associate
Administrative assistant	Interior designer	Retail store manager
Aircraft pilot	Kindergarten teacher	School principal
Architect	Lawyer	Security guard
Artist	Librarian	Skilled factory worker
Auto mechanic	Management consultant	Social media influencer
Barista	Medical doctor	Social scientist
Bookkeeper	Nanny	Social worker
Butcher	Newspaper reporter	Software developer
Call centre agent	Novelist	Stockbroker
Chartered public accountant	Nurse	Taxi driver
Chef	Office manager	Teacher assistant
Children's entertainer	Optometrist	Telecom technician
Civil engineer	Paralegal	Train operator
Cleaner	Paramedic	Truck driver
College professor	Parking enforcement officer	TV producer
Company CEO	Personal assistant	Unskilled factory worker
Company senior executive	Pest exterminator	Veterinarian
Construction worker	Pharmacist	Violinist
Cook	Photographer	Waiter
Delivery driver	Physicist	Waitress
Dental hygienist	Physiotherapist	Zookeeper
Dentist	Plumber	
Dispensing optician	Police captain	
Driving instructor	Police sergeant	
Elected official	Postal worker	
Electrical engineer	PR executive	
Electrician	Priest	
Elementary school teacher	Professional athlete	
Factory foreman	Project manager	
Farmer	Psychotherapist	
Fast food worker	Real estate agent	
Firefighter	Residential care worker	
Fitness instructor		
Flight attendant		
Florist		
Fruit picker		
Graphic designer		
Grocery store owner		
Grocery store shelf-filler		
Hairdresser		
High school teacher		
Hospital manager		
Hotel manager		

C Full produced ranking per prompt

See Table 4 and 5.

Table 4: Full produced rankings per prompt (1/2)

Occupation	P1 (Status)	P2 (Prestige)	P3 (‘Boastworthiness’)	P4 (Training/pay)
Company CEO	391 (1)	387 (2)	388 (2)	391 (1)
Company senior executive	385 (2)	378 (4)	378 (4)	381 (2)
Medical doctor	381 (3)	389 (1)	292 (24)	381 (2)
Lawyer	373 (4)	383 (3)	311 (18)	372 (4)
Police captain	369 (5)	340 (13)	321 (16)	259 (31)
Elected official	366 (6)	362 (7)	372 (6)	161 (55)
College professor	362 (7)	372 (5)	296 (21)	324 (15)
Aircraft pilot	356 (8)	353 (9)	366 (7)	350 (9)
Chartered public accountant	350 (9)	366 (6)	257 (32)	370 (5)
School principal	345 (10)	331 (14)	293 (22)	229 (37)
Architect	345 (10)	360 (8)	333 (13)	324 (15)
Management consultant	338 (12)	353 (9)	330 (14)	333 (14)
Dentist	335 (13)	344 (12)	189 (48)	362 (6)
Police sergeant	332 (14)	295 (24)	229 (39)	253 (32)
Physicist	314 (15)	352 (11)	230 (38)	276 (26)
Psychotherapist	308 (16)	314 (17)	199 (45)	209 (42)
Civil engineer	304 (17)	317 (16)	198 (47)	349 (10)
Professional athlete	297 (18)	237 (38)	384 (3)	302 (21)
Firefighter	296 (19)	305 (20)	327 (15)	209 (42)
Veterinarian	296 (19)	311 (19)	243 (35)	306 (20)
Priest	295 (21)	329 (15)	83 (76)	10 (95)
TV producer	293 (22)	280 (27)	362 (8)	249 (33)
Pharmacist	291 (23)	294 (25)	138 (60)	352 (8)
Electrical engineer	287 (24)	302 (21)	232 (37)	361 (7)
Novelist	286 (25)	297 (22)	346 (11)	59 (81)
Stockbroker	284 (26)	314 (17)	343 (12)	349 (10)
Project manager	276 (27)	268 (30)	321 (16)	315 (18)
Optometrist	274 (28)	284 (26)	177 (52)	337 (13)
Software developer	268 (29)	274 (29)	301 (19)	349 (10)
Social scientist	266 (30)	275 (28)	214 (41)	140 (60)
Actor	263 (31)	250 (33)	376 (5)	109 (68)
PR executive	262 (32)	264 (31)	355 (9)	282 (24)
Violinist	256 (33)	297 (22)	291 (25)	102 (69)
Hospital manager	253 (34)	240 (37)	252 (33)	232 (36)
Paramedic	247 (35)	235 (39)	183 (51)	273 (27)
Physiotherapist	238 (36)	250 (33)	157 (57)	302 (21)
Hotel manager	237 (37)	204 (47)	274 (28)	211 (41)
High school teacher	231 (38)	252 (32)	184 (50)	189 (47)
HR officer	229 (39)	213 (43)	212 (42)	280 (25)
Nurse	226 (40)	249 (35)	118 (64)	271 (28)
Real estate agent	224 (41)	212 (44)	283 (26)	320 (17)
Newspaper reporter	217 (42)	229 (41)	264 (31)	84 (74)
Paralegal	207 (43)	241 (36)	165 (55)	265 (29)
Interior designer	206 (44)	234 (40)	282 (27)	196 (46)
Office manager	203 (45)	184 (52)	187 (49)	235 (35)
Elementary school teacher	200 (46)	224 (42)	152 (58)	149 (58)
Graphic designer	194 (47)	205 (46)	300 (20)	222 (39)
Chef	193 (48)	197 (48)	266 (30)	182 (48)
Factory foreman	179 (49)	148 (60)	137 (61)	170 (51)
Artist	178 (50)	206 (45)	354 (10)	10 (95)
Kindergarten teacher	175 (51)	185 (51)	110 (66)	84 (74)

Table 5: Full produced rankings per prompt (2/2)

Occupation	P1 (Status)	P2 (Prestige)	P3 (‘Boastworthiness’)	P4 (Training/pay)
Retail store manager	174 (52)	152 (59)	244 (34)	198 (45)
Flight attendant	170 (53)	181 (53)	218 (40)	130 (62)
Social worker	170 (53)	180 (54)	139 (59)	59 (81)
Electrician	165 (55)	165 (56)	104 (68)	310 (19)
Librarian	162 (56)	195 (49)	56 (81)	57 (84)
Photographer	157 (57)	188 (50)	293 (22)	115 (67)
Grocery store owner	155 (58)	122 (65)	199 (45)	76 (77)
Teacher assistant	148 (59)	159 (58)	129 (62)	100 (70)
Dental hygienist	144 (60)	165 (56)	90 (74)	285 (23)
Dispensing optician	144 (60)	168 (55)	124 (63)	264 (30)
Social media influencer	137 (62)	69 (79)	392 (1)	44 (87)
Personal assistant	133 (63)	143 (62)	203 (44)	164 (54)
Telecom technician	123 (64)	131 (64)	69 (78)	237 (34)
Driving instructor	121 (65)	121 (67)	98 (71)	177 (50)
Train operator	119 (66)	122 (65)	100 (69)	156 (57)
Zookeeper	114 (67)	146 (61)	239 (36)	41 (88)
Bookkeeper	113 (68)	138 (63)	44 (85)	208 (44)
Fitness instructor	101 (69)	87 (75)	267 (29)	170 (51)
Administrative assistant	96 (70)	116 (68)	64 (79)	181 (49)
Auto mechanic	94 (71)	94 (74)	95 (72)	227 (38)
Plumber	92 (72)	108 (70)	24 (90)	219 (40)
Nanny	89 (73)	107 (71)	116 (65)	57 (84)
Security guard	88 (74)	82 (76)	22 (91)	126 (64)
Cook	84 (75)	96 (72)	171 (53)	99 (72)
Hairdresser	80 (76)	95 (73)	167 (54)	116 (66)
Florist	76 (77)	113 (69)	160 (56)	58 (83)
Construction worker	66 (78)	51 (85)	50 (83)	159 (56)
Postal worker	64 (79)	82 (76)	22 (91)	85 (73)
Farmer	61 (80)	75 (78)	94 (73)	39 (89)
Pest exterminator	58 (81)	56 (82)	44 (85)	100 (70)
Residential care worker	55 (82)	69 (79)	86 (75)	49 (86)
Parking enforcement officer	55 (82)	55 (84)	19 (94)	62 (79)
Truck driver	53 (84)	39 (88)	51 (82)	168 (53)
Butcher	51 (85)	56 (82)	78 (77)	35 (90)
Waiter	48 (86)	64 (81)	22 (91)	68 (78)
Children’s entertainer	38 (87)	33 (91)	208 (43)	4 (98)
Skilled factory worker	35 (88)	41 (87)	15 (95)	143 (59)
Call centre agent	31 (89)	42 (86)	43 (87)	131 (61)
Taxi driver	31 (89)	24 (93)	47 (84)	84 (74)
Barista	30 (91)	34 (89)	100 (69)	62 (79)
Retail sales associate	28 (92)	34 (89)	110 (66)	123 (65)
Delivery driver	26 (93)	20 (94)	29 (88)	129 (63)
Waitress	20 (94)	26 (92)	27 (89)	29 (91)
Fruit picker	15 (95)	15 (95)	59 (80)	15 (94)
Grocery store shelf-filler	11 (96)	11 (96)	7 (97)	23 (92)
Cleaner	7 (97)	7 (97)	3 (98)	10 (95)
Fast food worker	4 (98)	4 (98)	8 (96)	17 (93)
Unskilled factory worker	0 (99)	0 (99)	0 (99)	0 (99)

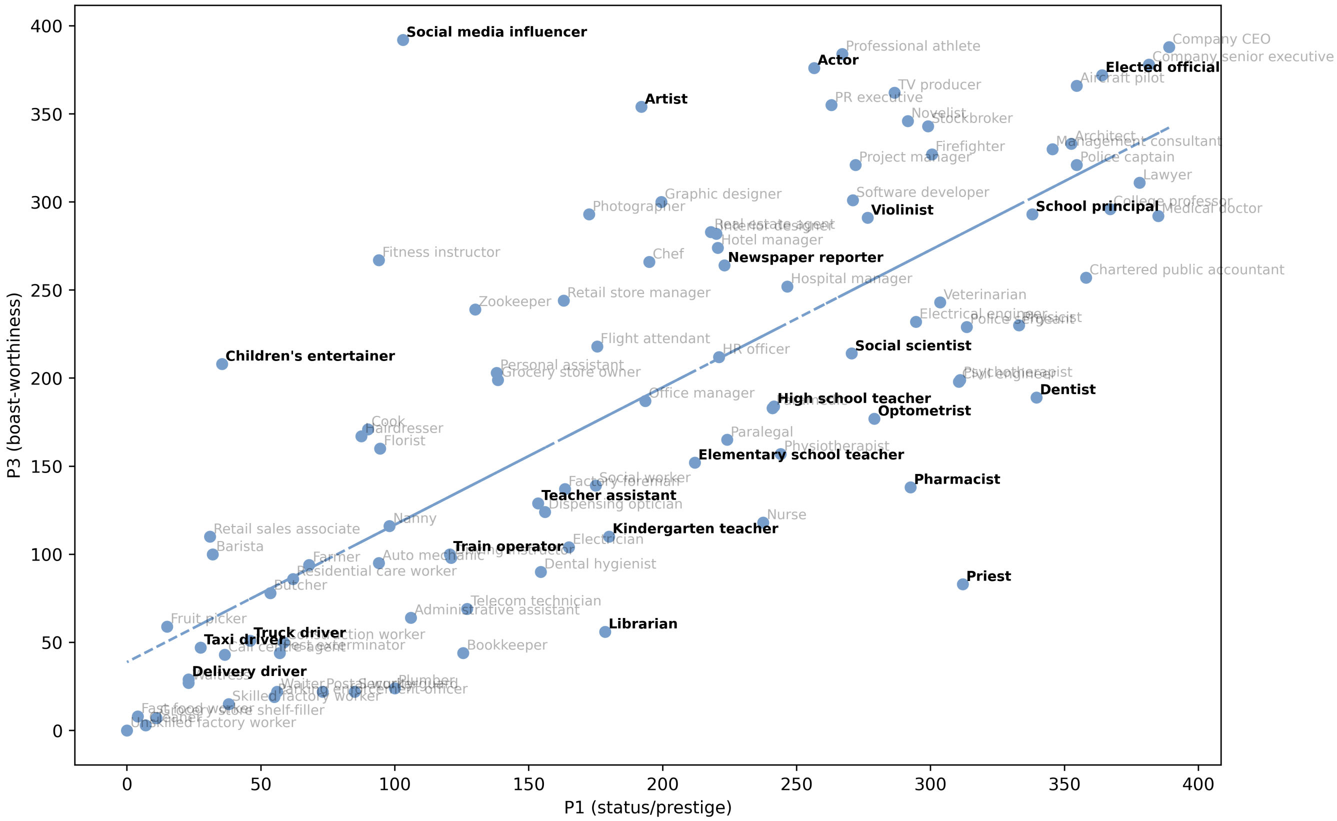

Figure 2: Relations between the scales produced by our P1/2 (status/prestige) and P3 (boast-worthiness), including trend lines.

D Detailed Results

Additional observations on rankings properties. To further validate the rankings, we compare groups of occupations which have a relatively clear expected prestige order. On all three scales (see Table 4 and 5 as well as Figure 2, Figure 3, and Figure 4), teaching occupations follow the order: school principal>high school teacher>elementary school teacher>kindergarten teacher. Teacher assistant follows the expected order in the scale derived from P1/2 (status/prestige). However, it ranks higher than kindergarten teacher on P3 (boast-worthiness) and P4 (training/pay). This may be due to some heterogeneity within the title of 'teacher assistant' – which may also include teaching assistants at the university level. Additionally, on all three scales, driving occupations follow the order: train operator>truck driver>taxi driver. Delivery drivers rank below the other driving occupations in the scale derived from P1/2 and P3, but above taxi drivers on P4. The reason for this is not clear, but may be due to the prevalence of online discourse around the pay and conditions of drivers for ride-sharing services such as Uber. On all three scales, children's entertainers rank lower than actors. Social media influencers occupy an interesting position in that they rank lower than actors but higher than children's entertainers on the scales derived from P1/2 and P4. However, they occupy the highest rank position on the scale derived from P3 (actors occupy position 5). Given that P3 asks which profession a hypothetical person would be most likely to boast about meeting, this suggests that GPT-3 strongly associates social media influencers with social boast-worthiness. This may be a consequence of

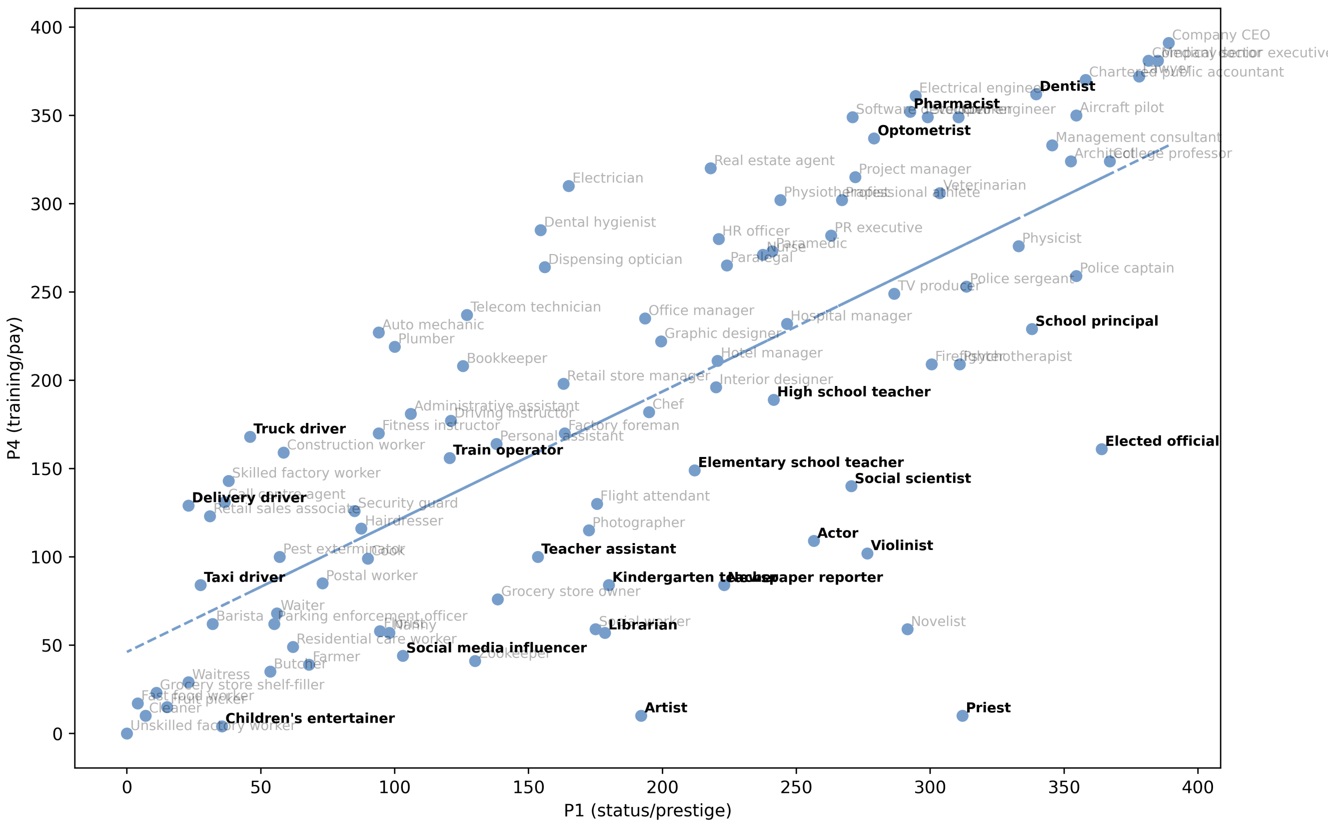

Figure 3: Relations between the scales produced by our P1/2 (status/prestige) and P4 (training/pay), including trend lines.

training data drawn from social media, where discussions of social proximity to influencers may be strongly represented.

Additional observations on rankings produced by different prompts. The deviations from the trend line between P1/2 and P4 also surface well-paid, respectable occupations, which perhaps have a reputation for being a little dull – including ‘dentist’, ‘pharmarcist’, ‘optometrist’, and ‘librarian’. Further, occupations that perform substantially worse on the training/pay scale than would be predicted from their prestige/status score are occuptions like ‘elected official’, ‘priest’, ‘social scientist’, ‘violinist’, ‘actor’, ‘newspaper reporter’, and ‘artist’. These are occupations that are prestigious due to their creativity, social value, or position of authority, but are relatively poorly paid.

Differences between our scales and existing occupational prestige rankings. There is a cluster of occupations with substantially lower SIOPS (survey-based prestige) scores than would be predicted by their P1/2 scores (status/prestige). These include ‘Firefighter’, ‘Police sergeant’, ‘Zookeeper’, and ‘Company senior executive’. One potential explanation for this pattern is that our LLM-derived measure is somewhat over-rating the status of, particularly, protective services occupations due to their high frequency of occurrence in media narratives (and therefore in LLM training data). The relative over-performance of ‘Company senior executive’ in our scale may be because, in standard occupational coding schemes (on which SIOPS is based), the corresponding category covers a wide variety of management roles – many of which may be substantially less prestigious than is implied by ‘senior executive’.

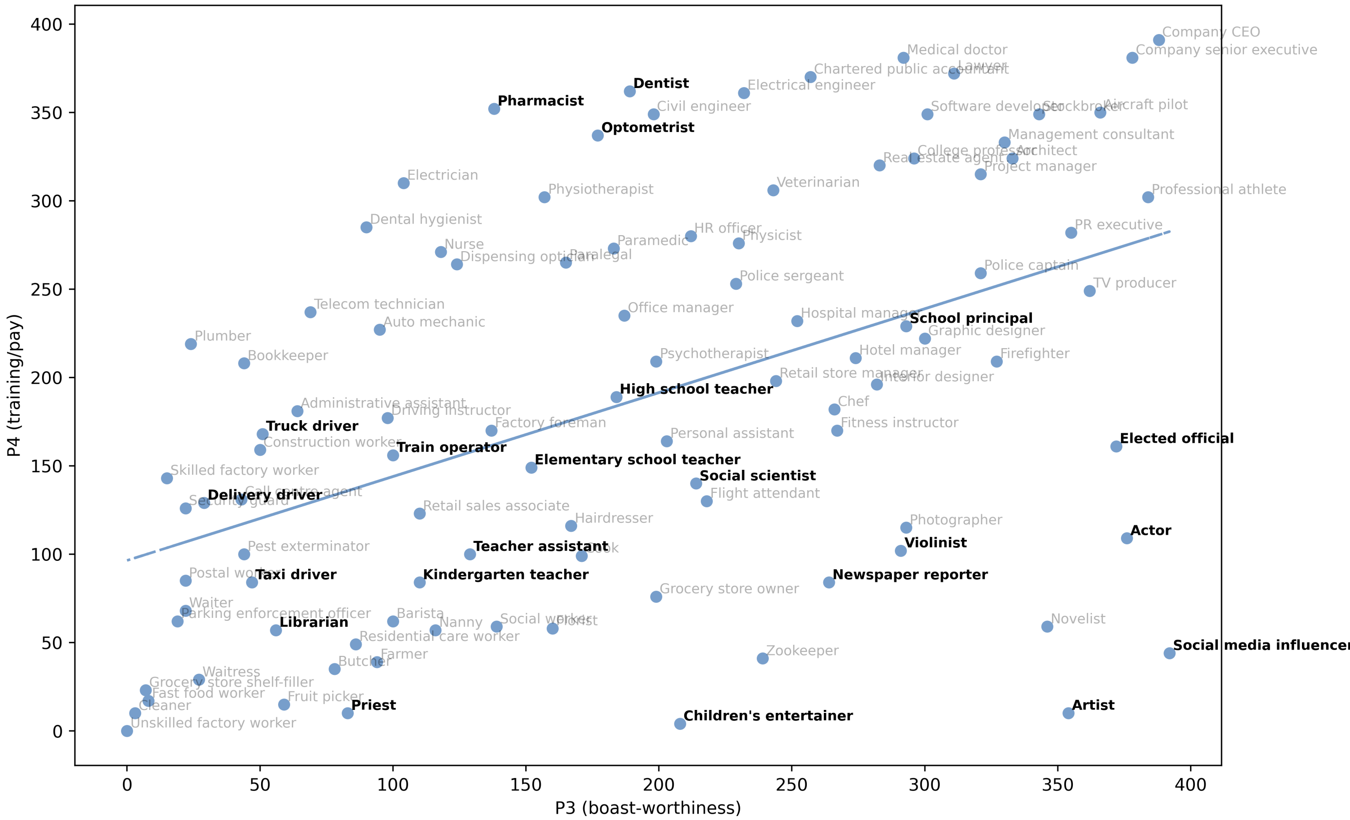

Figure 4: Relations between the scales produced by our P3 (boast-worthiness) and P3 (training/pay), including trend lines.

There is a similarly close correspondence between P1/2 (status/prestige) and the CAMSIS scale (social proximity-based), with similar occupations under performing on CAMSIS relative to their P1/2 scores. However, there is also a group of occupations which substantially over-perform on CAMSIS relative to P1/2. These include 'Librarian', 'Newspaper reporter', 'Social scientist', 'Novelist', and 'Psychotherapist'. The latter pattern suggests that CAMSIS is capturing a dimension of status which elevates creative and/or intellectual occupations more strongly than our primary LLM-derived scale does.

Occupations which under-perform on SEI relative to their scores on the ranking derived from P4 include occupations like 'Company senior executive' and 'company CEO'. In the real (US) population, the majority of CEOs and senior executives manage small companies, and are commensurately less handsomely rewarded than the prevalent stereotype of a CEO (US Bureau of Labor Statistics, 2021). By contrast, occupations which over perform on the SEI scale relative to their P4 scores include social scientists, novelists, and teachers. Some of these discrepancies may be due to stereotypes of low pay (for example among teachers). However, they may also arise due to the uncertain weighting being applied to pay versus education/training in the construction of P4.