What Do Language Models Know About the Past?

Probing Collective Mentalities With GenAI

by Tom Einhorn Yash Mali Laura K. Nelson (University of British Columbia)

on American Sociological Association Annual Meeting
New York City, NY
August, 2026

What is a Fact?

» What is a *fact* in the age of Generative AI?

Evidence-based

Verified data and verifiable measurements.

Institutional

Defined by established social structures and laws.

Interpretive

Multiple valid perspectives dependant on analysis.

» What is a *fact* in the age of Generative AI?

Evidence-based

Verified data and verifiable measurements.

Institutional

Defined by established social structures and laws.

Interpretive

Multiple valid perspectives dependant on analysis.

Generative

Created or simulated by computational models.

» Emerging facts

Emergent facts *are the result of dynamic interactions between training data, model processes and user instructions. They are not pre-existing truths, but computationally constructed outputs that reflect probabilistic patterns and contextual inferences drawn from the model's training. Emergent facts are characterised by their dependence on the specific prompt, the variability of the training data, and the inherent probabilistic nature of the model. They are plausible but unverified, contextually adaptive, highly variable and opaque in their derivation.*

— Dierickx et al., *What is a fact in the age of generative AI? Fact-checking as an epistemological lens*
Information, Communication, and Society (2026)
p. 10

» Emerging facts

Emergent facts *are the result of dynamic interactions between training data, model processes and user instructions. They are not pre-existing truths, but computationally constructed outputs that reflect probabilistic patterns and contextual inferences drawn from the model's training. Emergent facts are characterised by their dependence on the specific prompt, the variability of the training data, and the inherent probabilistic nature of the model. They are* ***plausible but unverified, contextually adaptive, highly variable and opaque in their derivation****.*

— Dierickx et al., *What is a fact in the age of generative AI? Fact-checking as an epistemological lens*
Information, Communication, and Society (2026)
p. 10

» Emergent facts in Archival Theory

Ginzburg's work on distance, perspective, and evidence — most fully developed in Wooden Eyes and Clues (Ginzburg, 1989; 2001) — offers a crucial but underused framework, particularly for thinking about how knowledge is produced when direct access to social reality is impossible. Although developed as a work of historical scholarship, his reflections are highly relevant for contemporary sociology and, in fact, ever more salient as the discipline increasingly relies on computational methods and explores generative artificial intelligence as a form of research methodology. If we read Ginzburg's intervention as a theory of knowledge under conditions of unavoidable mediation, it can make explicit sociology's longstanding but often implicit reliance on evidential modes of reasoning.

— Skarpelis, *From Archives to Algorithms: Distance, Evidence, and Inference*
Sociologica (2026)
p. 28

» Emergent facts in Archival Theory

Ginzburg's work on distance, perspective, and evidence — most fully developed in Wooden Eyes and Clues (Ginzburg, 1989; 2001) — offers a crucial but underused framework, particularly for thinking about ***how knowledge is produced when direct access to social reality is impossible****. Although developed as a work of historical scholarship, his reflections are highly relevant for contemporary sociology and, in fact, ever more salient as the discipline increasingly relies on computational methods and explores generative artificial intelligence as a form of research methodology. If we read Ginzburg's intervention as a theory of knowledge under conditions of unavoidable mediation, it can make explicit sociology's longstanding but often implicit reliance on evidential modes of reasoning.*

— Skarpelis, *From Archives to Algorithms: Distance, Evidence, and Inference*
Sociologica (2026)
p. 28

» Inaccessible knowledge

Some knowledge cannot be recovered by ordinary research methods:

- Never recorded, or the records are lost
- Exists only as shared, tacit cultural understanding of the kind that is not explicitly written down
- No surviving informants to ask

The past is a paradigmatic case.
We cannot interview the dead, but the traces of past collective mentalities may be encoded in the textual record, the same record that language models have been trained on.

» Inaccessible knowledge

Some knowledge cannot be recovered by ordinary research methods:

- Never recorded, or the records are lost
- Exists only as shared, tacit cultural understanding of the kind that is not explicitly written down
- No surviving informants to ask

The past is a paradigmatic case.
We cannot interview the dead, but the traces of past collective mentalities may be encoded in the textual record, the same record that language models have been trained on.

Can we use language models to probe these textually encoded collective mentalities?

» Diffuse, perspectival knowledge

Documented facts:
Recorded and accessible.

- Census data
- Birth and death records
- Wage tables

Diffuse, perspectival knowledge:
Collectively shared and understood, but rarely recorded.

- How prestigious was a “clerk” in 1900?
- What counted as “respectable” work in 1880s?
- How did status hierarchies varied by location or class?

» Diffuse, perspectival knowledge

Documented facts:
Recorded and accessible.

- Census data
- Birth and death records
- Wage tables

Diffuse, perspectival knowledge:
Collectively shared and understood, but rarely recorded.

- How prestigious was a "clerk" in 1900?
- What counted as "respectable" work in 1880s?
- How did status hierarchies varied by location or class?

Occupational prestige constitutes this kind of diffused, collective mental structure — temporally variable cultural knowledge widely shared by contemporaries but rarely explicitly recorded.

» Emergent facts as a bridge to inaccessible knowledge

LLMs trained on historical texts should enable scientists to more directly measure the responses of diverse past societies by simulating their responses on modern psychological instruments and behavioral measures. Such HLLMs should contain a trace, so to speak, of the collective mentality of the historical people whose writings were used to build them. Just as LLMs trained on text produced by contemporary populations are able to reproduce the psychological responses of those populations, LLMs trained on large corpora of historical text should reasonably reproduce the psychological responses of the historical populations that produced those texts. Drawing on data sources including fiction, plays, diaries, letters, and scholarly texts, it may be possible to gain novel insight into the thinking of populations that are no longer living.

— Varnum et al., *Large Language Models based on historical text could offer informative tools for behavioral science*
Psychological and Cognitive Sciences (2024)
p. 2

» Emergent facts as a bridge to inaccessible knowledge

LLMs trained on historical texts should enable scientists to more directly measure the responses of diverse past societies by simulating their responses on modern psychological instruments and behavioral measures. ***Such HLLMs should contain a trace, so to speak, of the collective mentality of the historical people whose writings were used to build them****. Just as LLMs trained on text produced by contemporary populations are able to reproduce the psychological responses of those populations, LLMs trained on large corpora of historical text should reasonably reproduce the psychological responses of the historical populations that produced those texts.* ***Drawing on data sources including fiction, plays, diaries, letters, and scholarly texts, it may be possible to gain novel insight into the thinking of populations that are no longer living****.*

— Varnum et al., *Large Language Models based on historical text could offer informative tools for behavioral science*
Psychological and Cognitive Sciences (2024)
p. 2

Historical Prestige Classifications

» Historical inccessibility

In historical research which pre-dates the era of large and complex social survey data collections, detailed data about occupations and their incumbents is not generally available on a large scale. Until recently, historical datasets with occupational data were often limited in their geographical, temporal, or socioeconomic coverage due to the labor intensive nature of gathering occupational records (cf. Miles and Vincent 1993). Digitization of historical documents, such as census records, parish records, and birth, marriage, and death certificates, now provides access to wider-ranging data on occupations (cf. NAPP 2008), but such records are often limited to the occupational title and lack systematic information about the features of occupations which might be used to explain the social position of occupations and their incumbents.

— Lambert et al., *The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Comparative Research*
Historical Methods: A Journal of Quantitative and Interdisciplinary History (2013)
p. 77

» Historical inccessibility

***In historical research which pre-dates the era of large and complex social survey data collections, detailed data about occupations and their incumbents is not generally available on a large scale**. Until recently, historical datasets with occupational data were often limited in their geographical, temporal, or socioeconomic coverage due to the labor intensive nature of gathering occupational records (cf. Miles and Vincent 1993). Digitization of historical documents, such as census records, parish records, and birth, marriage, and death certificates, now provides access to wider-ranging data on occupations (cf. NAPP 2008), but such records are often limited to the occupational title and lack systematic information about the features of occupations which might be used to explain the social position of occupations and their incumbents.*

— Lambert et al., *The Construction of HISCAM: A Stratification Scale Based on Social Interactions for Historical Comparative Research*
Historical Methods: A Journal of Quantitative and Interdisciplinary History (2013)
p. 77

Table 2. Occupational Standing Variables included in IPUMS

Variable name in IPUMS	Label	Years available in IPUMS	Occupational classification basis	Basis of score	Source data	Score varies across censuses
SEI	Duncan Socioeconomic Index	1850-present	1950	Income, Education, Prestige	1950 census, 1947 survey [5]	No
NPBOSS50	Nam-Powers-Boyd Occupational Status Score	1850-present	1950	Earnings, Education	Each census	Yes (post-1950)
PRESGL	Siegel Prestige Score	1850-present	1950	Prestige	1960s surveys [5]	No
EDSCOR50	Occupational Education Score	1850-present	1950	Education	Each census	Yes (post-1950)
ERSCOR50	Occupational Earnings Score	1850-present	1950	Earnings	Each census	Yes (post-1950)
OCCSCORE	Occupational Income Score	1850-present	1950	Income	1950 census	No
HWSEI	Hauser-Warren Socioeconomic Index	1950-present	1990	Earnings, Education, Prestige	1990 census 1989 GSS [5]	No
NPBOSS90	Nam-Powers-Boyd Occupational Status Score	1950-present	1990	Earnings, Education	Each census	Yes (post-1950)
PRENT	Nakao-Treas Prestige Score	1950-present	1990	Prestige	1989 GSS [5]	No
EDSCOR90	Occupational Education Score	1950-present	1990	Education	Each census	Yes (post-1950)
ERSCOR90	Occupational Earnings Score	1950-present	1990	Earnings	Each census	Yes (post-1950)

Table 2. Occupational Standing Variables included in IPUMS

Variable name in IPUMS	Label	Years available in IPUMS	Occupational classification basis	Basis of score	Source data	Score varies across censuses
SEI	Duncan Socioeconomic Index	1850-present	1950	Income, Education, Prestige	1950 census, 1947 survey [3]	No
NPBOSS50	Nam-Powers-Boyd Occupational Status Score	1850-present	1950	Earnings, Education	Each census	Yes (post-1950)
PRESGL	Siegel Prestige Score	1850-present	1950	Prestige	1960s surveys [3]	No
EDSCOR50	Occupational	1850-	1950	Education	Each census	Yes (post-1950)
			1950	Earnings	Each census	Yes (post-1950)
OCCSCORE	Occupational Income Score	1850-present	1950	Income	1950 census	No
HWSEI	Hauser-Warren Socioeconomic Index	1950-present	1990	Earnings, Education, Prestige	1990 census 1989 GSS [3]	No
NPBOSS90	Nam-Powers-Boyd Occupational Status Score	1950-present	1990	Earnings, Education	Each census	Yes (post-1950)
PRENT	Nakao-Treas Prestige Score	1950-present	1990	Prestige	1989 GSS [3]	No
EDSCOR90	Occupational Education Score	1950-present	1990	Education	Each census	Yes (post-1950)
ERSCOR90	Occupational Earnings Score	1950-present	1990	Earnings	Each census	Yes (post-1950)

1950–1990

Table 2. Occupational Standing Variables included in IPUMS

Variable name in IPUMS	Label	Years available in IPUMS	Occupational classification basis	Basis of score	Source data	Score varies across censuses
SEI	Duncan Socioeconomic Index	1850-present	1950	Income, Education, Prestige	1950 census, 1947 survey [5]	No
NPBOSS50	Nam-Powers-Boyd Occupational Status Score	1850-present	1950	Earnings, Education	Each census	Yes (post-1950)
PRESGL	Siegel Prestige Score	1850-present	1950	Prestige	1960s surveys [5]	No
EDSCOR50	Occupational [illegible]	1850-[illegible]	1950	Education	Each census	Yes (post-1950)
[illegible]	[illegible]	[illegible]	1950	Earnings	Each census	Yes (post-1950)
OCCSCORE	Occupational Income Score	1850-present	1950	Income	1950 census	No
HWSEI	Hauser-Warren Socioeconomic Index	1950-present	1990	Earnings, Education, Prestige	1990 census 1989 GSS [3]	No
NPBOSS90	Nam-Powers-Boyd Occupational Status Score	1950-present	1990	Earnings, Education	Each census	Yes (post-1950)
PRENT	Nakao-Treas Prestige Score	1950-present	1990	Prestige	1989 GSS [3]	No
EDSCOR90	Occupational Education Score	1950-present	1990	Education	Each census	Yes (post-1950)
ERSCOR90	Occupational Earnings Score	1950-present	1990	Earnings	Each census	Yes (post-1950)

1950–1990

Even existing historical schemes encode a point of view:

"First of all, women: **I lumped all women** under a category called **'female occupations'**; this was the only solution to the problem created by the **looseness** of female occupational designations."

Occupational Classification in History/ Katz, 1972

"Thirty additional titles were chosen from the list of intermediate occupational titles for the 1940 U.S. census by selecting 'the most frequently reported occupations at what we arbitrarily judged to be all status levels' and **excluding 'women's occupations'**."

Jobs and Occupations: A Popular Evaluation/ Hatt & North, 1947

Even existing historical schemes encode a point of view:

"First of all, women: **I lumped all women** under a category called **'female occupations'**; this was the only solution to the problem created by the **looseness** of female occupational designations."

Occupational Classification in History/ Katz, 1972

"Thirty additional titles were chosen from the list of intermediate occupational titles for the 1940 U.S. census by selecting 'the most frequently reported occupations at what we arbitrarily judged to be all status levels' and **excluding 'women's occupations'**."

Jobs and Occupations: A Popular Evaluation/ Hatt & North, 1947

Any prestige ranking is always *someone's* ranking.

Can we prompt LLMs to generate occupational prestige measures?

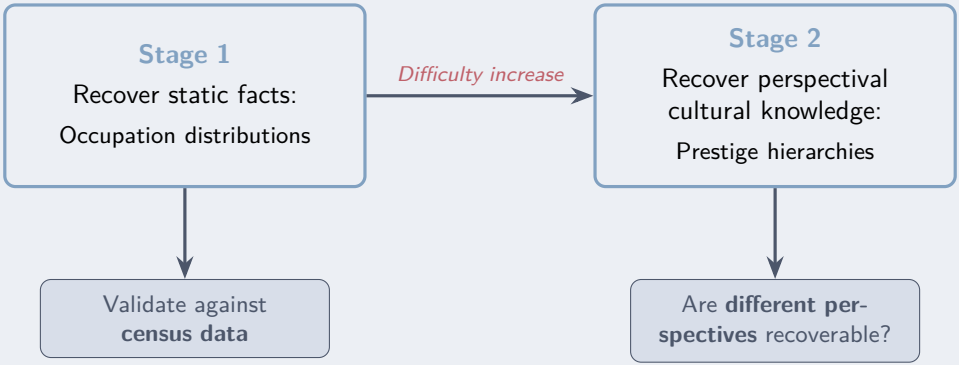
» Research team

Laura Nelson
Associate Professor, Sociology
UBC

Yash Mali
BA, Computer Science
UBC

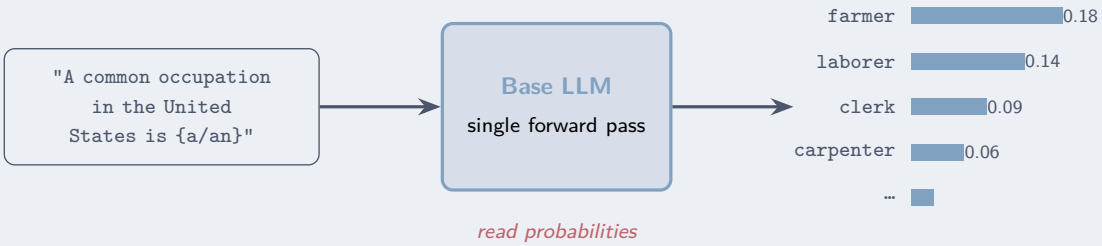
Tom Einhorn
PhD Candidate, Sociology
UBC

» Research design



Stage 1: Recovering static facts

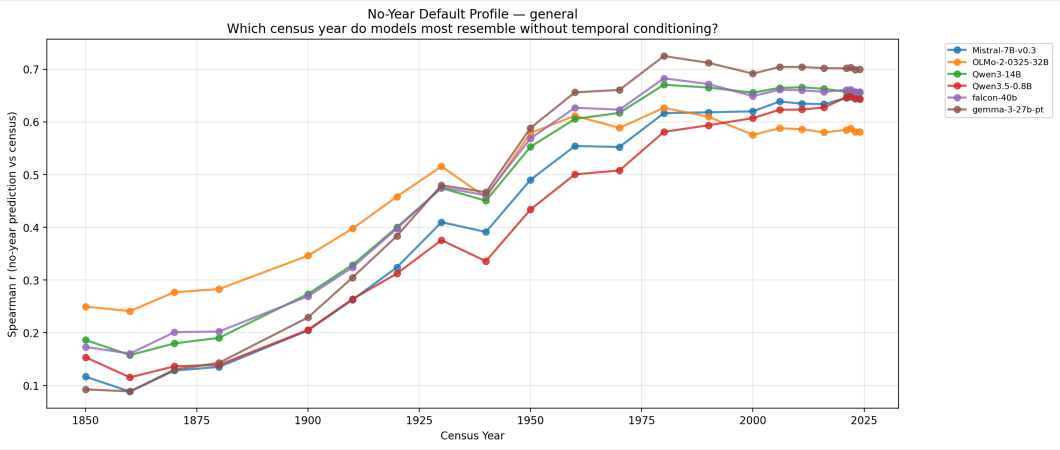
» Probing token probabilities



The model doesn't *generate* a completion. Instead, for each occupation in a fixed slate, we read its probability as a completion to the prompt.

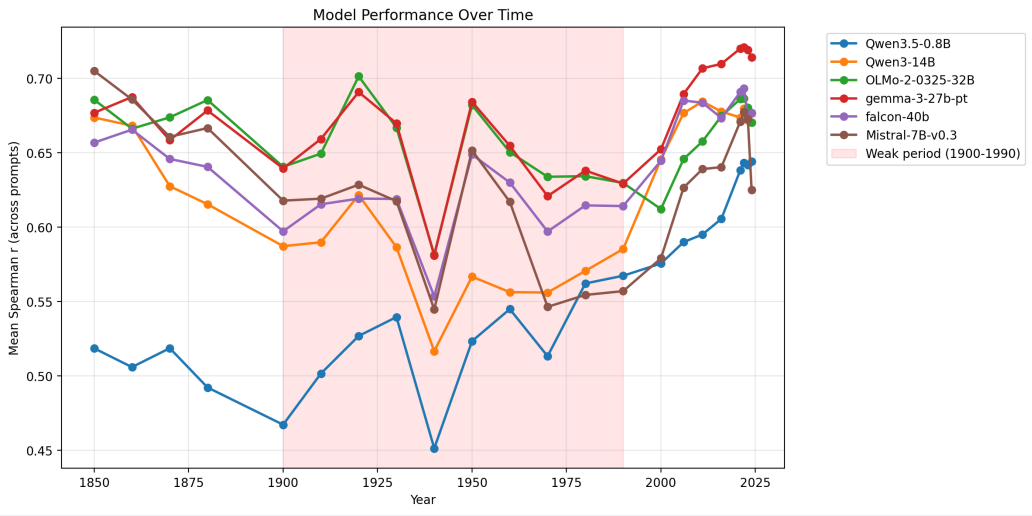
» “A common occupation in the United States is”

» “A common occupation in the United States is”



» “In {year}, a common occupation in the United States was”

» “In {year}, a common occupation in the United States was”



» “In {year}, a common occupation in the United States was”

» Culturally-anchored prompt

Each query wraps the occupation in period-specific framing, so the model reasons *from within* the era:

Base prompt

```
In {year}, {role} was {name}. A popular {media_type} was {media}.
{event}. A common occupation in the United States was
```

» Culturally-anchored prompt

Each query wraps the occupation in period-specific framing, so the model reasons *from within* the era:

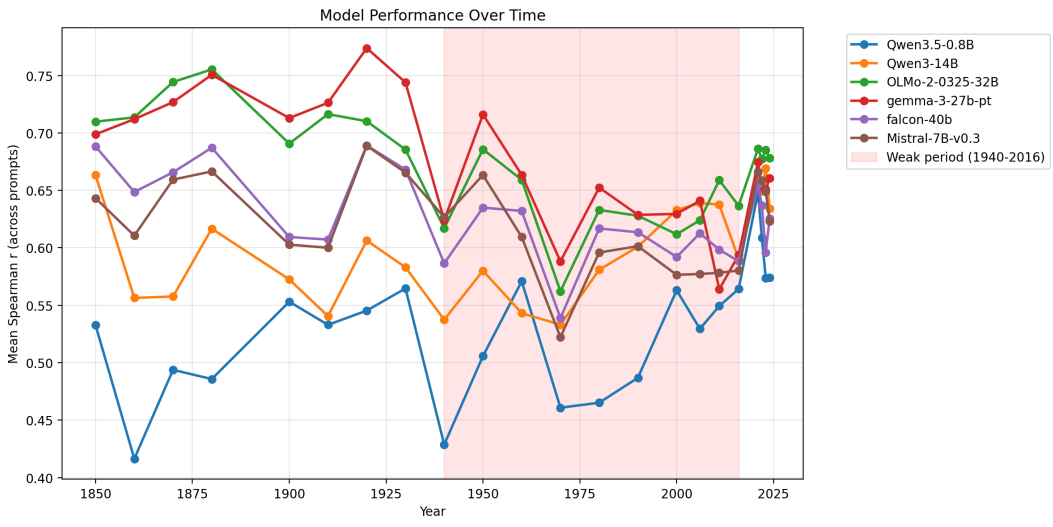
Base prompt

```
In {year}, {role} was {name}. A popular {media_type} was {media}.
{event}. A common occupation in the United States was
```

Example prompt

```
In 1850, Nathan K. Hall was Postmaster General, The Wide, Wide World
was being widely read, and the State of California is admitted into the
union. A common occupation in the United States was
```

» Culturally-anchored prompt



» Stage 1: Recovering static facts

Can base models reproduce historical U.S. occupation distributions, validated against census data?

What we found

- By default, models show a strong presentist bias.
- When prompted with year, recovery of historical census distributions improves.
- Culturally-anchored prompts perform better pre-1940, moving the error closer to the present.
- The remaining error is systematic and interpretable, not random.

Recovery with a systematic and interpretable error provides confidence in continuing to Stage 2.

Stage 2: Diffused, perspectival facts

» Stage 2: Diffused, perspectival facts

From a documented fact to diffused, perspectival facts: occupational **prestige hierarchies**, 1850–1940.

Unbounded base models

OLMo-2,
Mistral-7B,
Qwen3-14B,
Qwen3.5-0.8B,
Falcon-40B,
Gemma-3-27B

trained on textual records including modern scholarship.

Plus one bounded model

talkie-1930
trained only on text through 1931.

Not contaminated by present-day knowledge of the past, allowing us to ask what the model knows without the benefit of hindsight.

» Attempt 1: Structured completion task

We first asked models to generate a classification scheme via chained completion.

Validation

Output validated manually against six historical classifications, spanning 1852–1947.

» Completion task

"In 1850, Nathan K. Hall was Postmaster General, The Wide, Wide World was being widely read, and the State of California is admitted into the union. It was a period in which the social and economic character of the nation was in flux, and with it, the way people understood the status of different kinds of work.

During this time, occupations were classified by their social status. The principles that guided this classification were"

» Completion task

"In 1850, Nathan K. Hall was Postmaster General, The Wide, Wide World was being widely read, and the State of California is admitted into the union. It was a period in which the social and economic character of the nation was in flux, and with it, the way people understood the status of different kinds of work.

During this time, occupations were classified by their social status. The principles that guided this classification were"

Qwen3-14B: based on the idea that some jobs were more important than others. This was a time when the concept of the "gentleman" was still prevalent, and the idea that certain professions were more suitable for men of higher social standing.

» Completion task

Model	Professional / Professor	Manual / Wage / Manufacturing
Qwen3-14B	College professor, Judge, Lawyer, Physician, Chemist, Architect, Civil engineer, Software engineer, Nuclear engineer, Astronaut	Laborer (unskilled), Laborer (skilled), Laborer (semi-skilled), Laborer (unskilled, temporary)...
OLMo2-32B	College professor, Schoolteacher, School principal, School superintendent, School board member, School janitor, School bus driver...	Ditch digger, Coal miner, Farm laborer, Day laborer, Longshoreman, Street cleaner, Janitor, Porter, Chauffeur...
Gemma3-27B	Physician, Lawyer, College Professor, Clergyman, Architect, Dentist, Army Officer, Civil Engineer, Chemist, Pharmacist, Accountant, Veterinarian, Nurse, Teacher...	Factory operative, Dressmaker, Watchman, Building contractor, Chemist, Call center operator, Farmer (owner-operator), Policeman, Printer...
Talkie-1930	college professor, lawyer, clergyman, doctor	Coal miner, Foreman (factory)

» Completion task

Model	Professional / Professor	Manual / Wage / Manufacturing
Qwen3-14B	College professor, Judge, Lawyer, Physician, Chemist, Architect, Civil engineer, Software engineer, Nuclear engineer, Astronaut	Laborer (unskilled), Laborer (skilled), Laborer (semi-skilled), Laborer (unskilled, temporary)...
OLMo2-32B	College professor, Schoolteacher, School principal, School superintendent, School board member, School janitor, School bus driver...	Ditch digger, Coal miner, Farm laborer, Day laborer, Longshoreman, Street cleaner, Janitor, Porter, Chauffeur...
Gemma3-27B	Physician, Lawyer, College Professor, Clergyman, Architect, Dentist, Army Officer, Civil Engineer, Chemist, Pharmacist, Accountant, Veterinarian, Nurse, Teacher...	Factory operative, Dressmaker, Watchman, Building contractor, Chemist, Call center operator, Farmer (owner-operator), Policeman, Printer...
Talkie-1930	college professor, lawyer, clergyman, doctor	Coal miner, Foreman (factory)

» Attempt 2: continuous variable task

This is also a completion task — but instead of parsing generated text, we read the model's probabilities over a fixed occupation list.

Validation

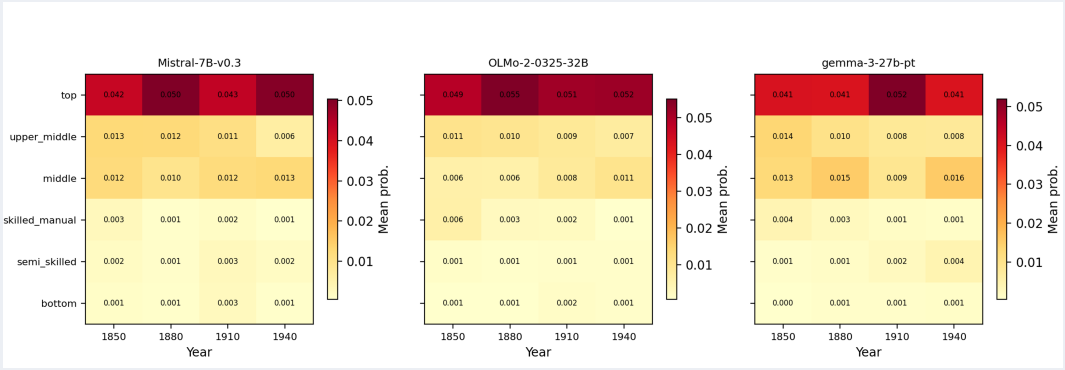
- **Base (US) set:** an ordinal prestige list built from seven historical classifications covering the period (1852–1947).
  Six tiers containing 92 occupations.
- **UK set:** continuous HISCAM scores.
  103 occupations derived from stratified random sample of HISCAM.

Scored by Spearman correlation against each scheme.

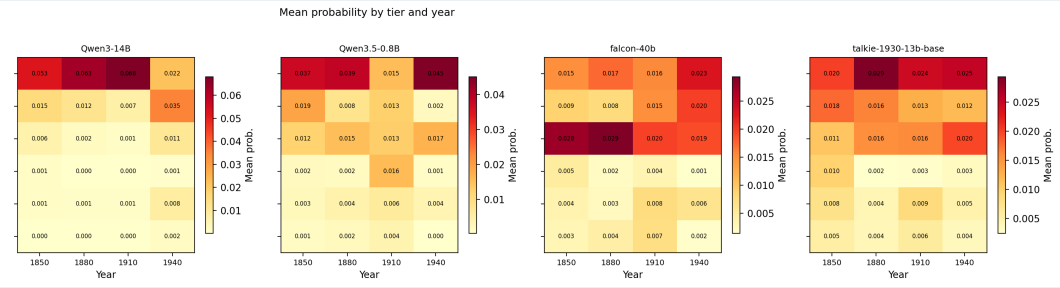
Prompt

```
[…] An occupation near the top of the status hierarchy in the
United States is
```

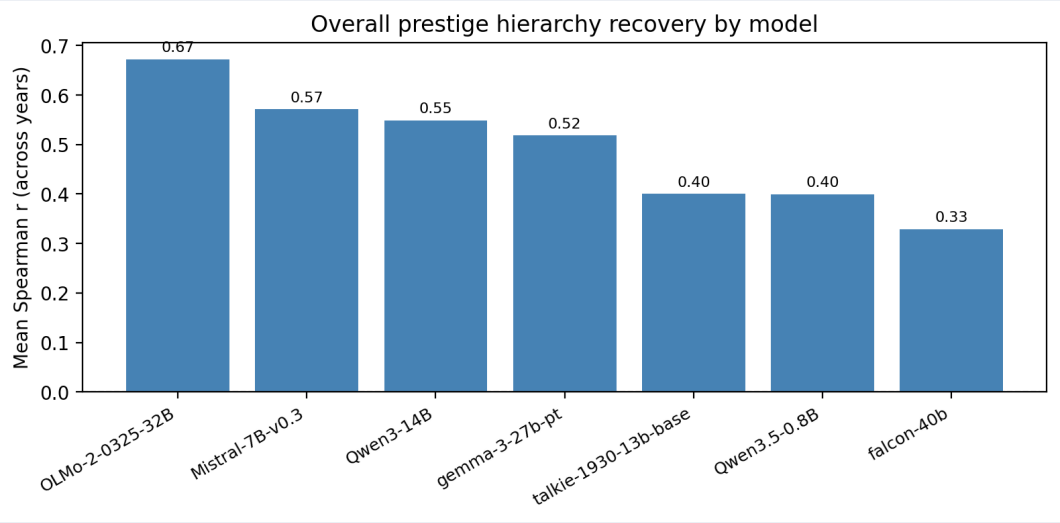
» Continuous variable task: Tier recovery (US set)



» Continuous variable task: Tier recovery (US set)



» Continuous variable task: Tier recovery (US set)



» Continuous variable task: Variations

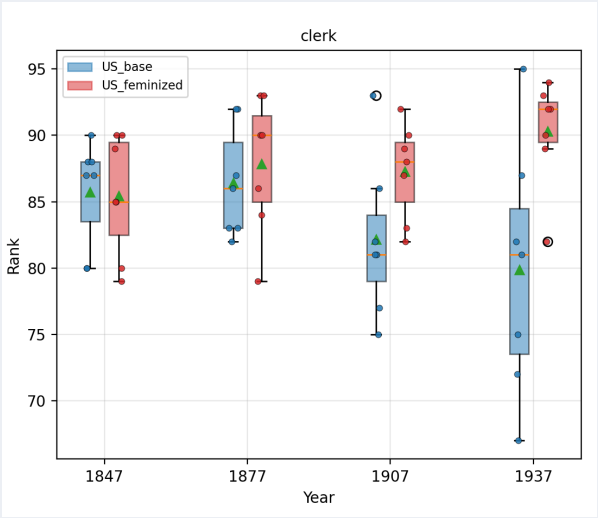
Feminzed

`[…] An occupation near the top of the status hierarchy for a woman in {Country} is`

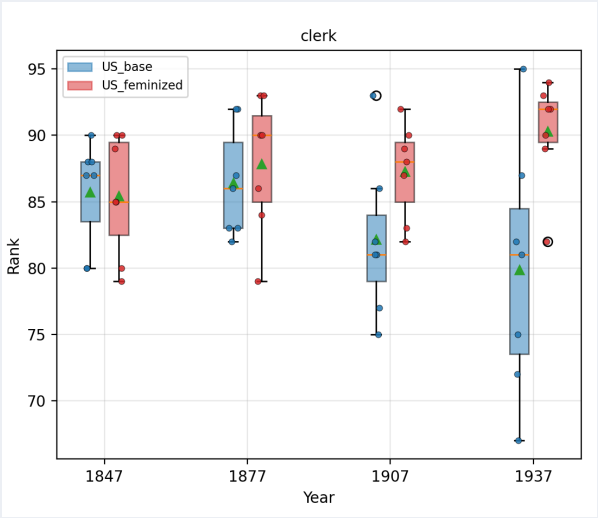
Contextual shift

`[…] An occupation near the top of the status hierarchy` *`(feminizaiton: for a woman)`* `in Britain is`

» Continuous variable task: Feminization



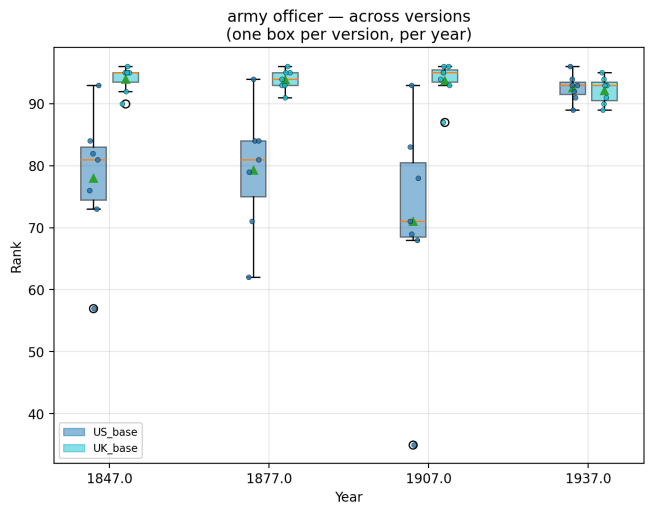
» Continuous variable task: Feminization



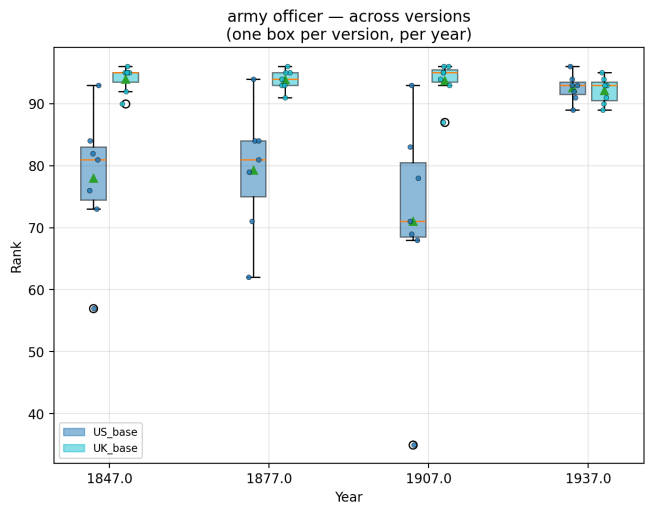
Tentative interpretation

As clerk becomes feminized, status increases for women but decreases overall

» Continuous variable task: Geographical context



» Continuous variable task: Geographical context



Tentative interpretation

Differences may reflect the nature of pre-WWI British Army as more aristocratic

» Stage 2: Diffused, perspectival facts

Can base models recover diffused occupational prestige hierarchies?

What we found

- Base models struggle to recover a full classification via chained completion, demonstrating drift, degeneration, and failed output.
- Probing token probabilities is far more effective than parsing generated text.
- Perspective can be shifted by prompting: feminized and contextual prompts move the recovered hierarchy in systematic and interpretable ways.
- A bounded model (talkie-1930) produces *different flavored* but not significantly better output.
- Recovered rankings correlate with ground truth, but may track occupational salience as much as prestige.

Conclusions

» Methodological provocations

* Generated collective mentalities > in-silico sampling:
  LLMs as a synthesizers of aggregate mentalities rather than simulated study participants.

* Probing probabilities > parsing generated output:
  The signal is in the distribution, not the surface text.

* Base models > aligned models:
  Alignment training distorts historical signal.

» Computational simulations

Generate collective mentalities

by probing token probabilities

under varying conditions

using base models only.

» Open Questions

Audiences

What audiences or venues might be interested in this conversation?

Terminology

What should we name the type of fact that LLMs produce?
Emergent fact? Synthetic fact? A posteriori fact? Something else?

Thank you

What Do Language Models Know About the Past?

Probing Collective Mentalities With GenAI

Tom Einhorn Yash Mali Laura K. Nelson

University of British Columbia

Scan for contact info

Appendix

» talkie-1930

talkie differs from the other models on two dimensions at once.

Historically bounded corpus

Trained only on text published before 1931 — no later scholarship or historical reinterpretation.

Slghtly smaller training corpus

Trained on fewer tokens than most modern web-scale open-class language models.

What does this mean?

- We cannot attribute talkie's behaviour solely to the historical cutoff.
- Differences may also reflect the amount of training data.

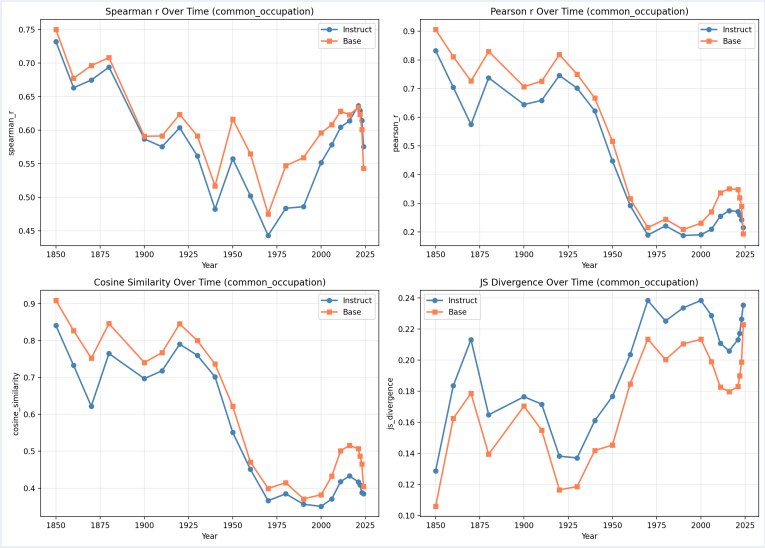
Boundedness and corpus size remain confounded in our comparison.

» Model specifications

Model	Params	Corpus cutoff	Developer
OLMo-2-0325-32B	32B	Dec. 2023	Allen AI
Falcon-40B	40B	not specified	TII (Abu Dhabi)
gemma-3-27b-pt	27B	not specified	Google DeepMind
Qwen3-14B	14B	not specified	Alibaba
Mistral-7B-v0.3	7B	not specified	Mistral AI
Qwen3.5-0.8B	0.8B	not specified	Alibaba
talkie-1930-13b-base	13B	before 1931	Contextual AI

All base (non-instruction-tuned) models.
Except for talkie, all models are open weight, web-scale models.

» Base vs. aligned models



» Are we measuring prestige or salience?

The probability a model assigns to an occupation may confound how available that occupation with how esteemed it was.

Evidence

- Status-framed and no-status-framed prompts correlate at $r \approx +0.43$.
- Each correlates with census frequency at $r \approx +0.24$ (status prompted) and $r \approx +0.33$ (none status prompted).
- The two task framings resemble each other more than either resembles the census.

Interpretation

Prompt framing exerts influence, but a shared salience component likely underlies both.

» Prompt inventory

Cultural anchors held constant across variants, so each contrast varies one thing only.

Base

```
[…] An occupation near the top of the status hierarchy in the United States is
```

Feminized

```
[…] An occupation near the top of the status hierarchy for a woman in the United States is
```

Contextual shift

```
[…] An occupation near the top of the status hierarchy in Britain is
```

No-status control

```
[…] A common occupation in the United States is
```

» Occupation list construction

US set: 92 occupations

- Drawn from the seven historical schemes.
- Retained where occupations appearing across multiple classifications.
- Selected to for both stability and change.
- Six ordinal tiers.

UK set: 103 occupations

- 10% Stratified random sample from HISCAM.
- Continuous scale, 0–100.
- Stratified by percentile.

Variant handling

Each occupation carries 2–3 label variants, drawn only from confirmed historical synonyms and existing pipeline mappings — no invented synonyms. Scores averaged across variants.

» Reveiwed historical classifications

Year	Source
1852 [1972]	Katz
1897	Hunt
1917	Edwards
1925	Counts
1933	Edwards
1940/43	Smith
1947	North & Hatt

» Anachronisms

Deliberately anachronistic occupations are included as a sanity check on temporal anchoring.

Treatment

- Scored alongside in-period occupations.
- Excluded from Spearman correlation (no ground-truth tier).
- Retained for qualitative inspection only.

Findings

- Anachronism probabilities are uniformly low across models and periods.
- "Astronaut" is the most confusable, perhaps drawing from early 20th century science fiction.

» Technical implementation

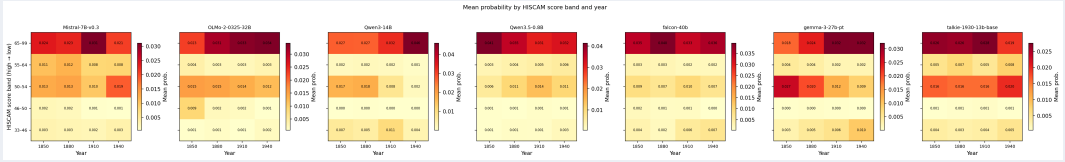
Scoring

- Forward-push log-probability extraction over a fixed occupation list.
- Multi-token occupations scored by aggregation rule.
- Scores normalised (& ranked) before correlation.

talkie-1930 requires a separate path

- Custom loader + wrapper bypassed to reach `model.model` directly.
- Autoregressive loop for log-probability scoring rather than a single forward pass.
- Hard 2048-token context window enforced at the rotary embedding level, never used.
- Requires `temperature` > 0; degenerate at 0.

» UK results (HISCAM)



Notes

Continuous 0–100 scale over 103 occupations. Models struggled with this recovery more so than with the US set, perhaps reflecting a US-bias. OLMo leads.

» Six task versions compared

Prompt	Ground truth	Variant	Leading model	Lead r value
US	tiers	base	OLMo	+0.67
US	tiers	feminized	Qwen-14B	+0.45
UK	tiers	base	OLMo	+0.61
UK	tiers	feminized	talkie-1930	+0.40
UK	HISCAM	base	Qwen-14B	+0.37
UK	HISCAM	feminized	falcon-40b	+0.33

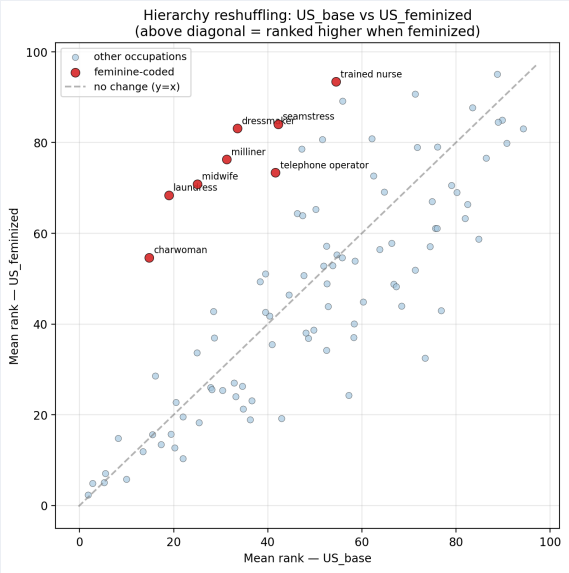
- Overall strongest result: OLMo on the US base task ($r \approx +0.67$).
- Overall weakest result: Qwen-0.8B on the UK feminized task ($r \approx +0.29$).
- OLMo in the top 3 for all base tasks, but struggles with shifting perspectives.

» Why probe rather than fine-tune?

Rationale

- The question is what models already encode from the historical record without the need for additional training.
- Probing is cheap and reproducible. Others can apply it to off-the-shelf checkpoints.
- talkie-1930 represents and adjacent solution: bounded models. We use it as a control.

» Continuous variable task: Feminization



» Structured completion task

We first asked models to generate a classification scheme via chained completion.

Validation

Output validated manually against six historical classifications, spanning 1852–1947.

» Completion task

"In 1850, Nathan K. Hall was Postmaster General, The Wide, Wide World was being widely read, and the State of California is admitted into the union. It was a period in which the social and economic character of the nation was in flux, and with it, the way people understood the status of different kinds of work.

During this time, occupations were classified by their social status. The principles that guided this classification were"

» Completion task

"In 1850, Nathan K. Hall was Postmaster General, The Wide, Wide World was being widely read, and the State of California is admitted into the union. It was a period in which the social and economic character of the nation was in flux, and with it, the way people understood the status of different kinds of work.

During this time, occupations were classified by their social status. The principles that guided this classification were"

Qwen3-14B: based on the idea that some jobs were more important than others. This was a time when the concept of the "gentleman" was still prevalent, and the idea that certain professions were more suitable for men of higher social standing.

» Completion task

“Under this framework, the classification comprised the following groups, listed from the highest to the lowest social status:

1.”

» Completion task

"Under this framework, the classification comprised the following groups, listed from the highest to the lowest social status:

1."

Qwen3-14B:

1. The learned professions (lawyers, doctors, clergymen)
2. The liberal professions (teachers, writers, editors)
3. The mechanical professions (engineers, architects, machinists)
4. The commercial professions (merchants, bankers, traders)
5. The agricultural professions (farmers, ranchers, laborers)
6. The manual professions (carpenters, blacksmiths, laborers)
7. The domestic professions (housekeepers, cooks, servants)
8. The laboring professions (unskilled laborers, miners, dockworkers)
10. The menial professions (street sweepers, porters, beggars)

» Completion task

"In 1850, Nathan K. Hall was Postmaster General, The Wide, Wide World was being widely read, and the State of California is admitted into the union. It was a period in which the social and economic character of the nation was in flux, and with it, the way people understood the status of different kinds of work.

During this time, occupations were classified by their social status. The principles that guided this classification were"

» Completion task

"In 1850, Nathan K. Hall was Postmaster General, The Wide, Wide World was being widely read, and the State of California is admitted into the union. It was a period in which the social and economic character of the nation was in flux, and with it, the way people understood the status of different kinds of work.

During this time, occupations were classified by their social status. The principles that guided this classification were"

talkie-1930-13b-base: **the status of the work, the nature of society, and the nature of the world in which the work was done. The world of work was considered a world of folk who were socially inferior to the professional man and the man of property**. In that world, the wage earner held a lower social standing than the small shopkeeper. The small shopkeeper held a lower social standing than the man of property. The man of property held a social standing higher than the man of the professions. The man of the professions was the highest in the scale. The reason for this classification is clear. The professional man was considered a man of superior education. The man of property was considered a man of superior wealth. The man of property and the professional man were men of leisure. Leisure is a natural concomitant of wealth and superior education. [...]

» Completion task: Examples

1850

Qwen3-14B

The learned professions
The liberal professions
The mechanical professions
The commercial professions
The agricultural professions
The manual professions
The domestic professions
The laboring professions
The menial professions

OLMo2-32B

Clergy
Physicians
Lawyers
Professors
Merchants
Bankers
Mechanics
Farmers
Laborers

Gemma3-27B

Professional
Managerial
Clerical
Sales
Skilled
Semi-skilled
Unskilled

talkie-1930

The man of property
Professional man
Small shopkeeper
Wage earner
Vagabond

1940

Qwen3-14B

Professions
White-collar workers
Skilled manual workers
Unskilled manual workers

OLMo2-32B

Professional occupations
Technical occupations
Clerical occupations
Sales occupations
Service occupations
Agricultural occupations
Labor occupations

Gemma3-27B

Professional, technical, and kindred workers
Managers, officials, and proprietors, except farm
Clerical and kindred workers
Sales workers
Craftsmen, foremen, and kindred workers
Operatives and kindred workers
Private household workers
Service workers, except private household
Farm laborers and foremen
Laborers, except farm and mine

» Completion task: Examples

1850

Qwen3-14B

The learned professions
The liberal professions
The mechanical professions
The commercial professions
The agricultural professions
The manual professions
The domestic professions
The laboring professions
The menial professions

OLMo2-32B

Clergy
Physicians
Lawyers
Professors
Merchants
Bankers
Mechanics
Farmers
Laborers

Gemma3-27B

Professional
Managerial
Clerical
Sales
Skilled
Semi-skilled
Unskilled

talkie-1930

The man of property
Professional man
Small shopkeeper
Wage earner
Vagabond

1940

Qwen3-14B

Professions
White-collar workers
Skilled manual workers
Unskilled manual workers

OLMo2-32B

Professional occupations
Technical occupations
Clerical occupations
Sales occupations
Service occupations
Agricultural occupations
Labor occupations

Gemma3-27B

Professional, technical, and kindred workers
Managers, officials, and proprietors, except farm
Clerical and kindred workers
Sales workers
Craftsmen, foremen, and kindred workers
Operatives and kindred workers
Private household workers
Service workers, except private household
Farm laborers and foremen
Laborers, except farm and mine

» Completion task

Under this framework, the classification comprised the following groups, listed from the highest to the lowest social status:

{generated categories}

Some of the occupations included in this classification were: ditch digger, insurance agent, hotel keeper, watchman, bootblack, college professor [...]

Examples of occupations classified under {category} included:

» Completion task

Model	Professional / Professor	Manual / Wage / Manufacturing
Qwen3-14B	College professor, Judge, Lawyer, Physician, Chemist, Architect, Civil engineer, Software engineer, Nuclear engineer, Astronaut	Laborer (unskilled), Laborer (skilled), Laborer (semi-skilled), Laborer (unskilled, temporary)...
OLMo2-32B	College professor, Schoolteacher, School principal, School superintendent, School board member, School janitor, School bus driver...	Ditch digger, Coal miner, Farm laborer, Day laborer, Longshoreman, Street cleaner, Janitor, Porter, Chauffeur...
Gemma3-27B	Physician, Lawyer, College Professor, Clergyman, Architect, Dentist, Army Officer, Civil Engineer, Chemist, Pharmacist, Accountant, Veterinarian, Nurse, Teacher...	Factory operative, Dressmaker, Watchman, Building contractor, Chemist, Call center operator, Farmer (owner-operator), Policeman, Printer...
Talkie-1930	college professor, lawyer, clergyman, doctor	Coal miner, Foreman (factory)

» Completion task

Model	Professional / Professor	Manual / Wage / Manufacturing
Qwen3-14B	College professor, Judge, Lawyer, Physician, Chemist, Architect, Civil engineer, Software engineer, Nuclear engineer, Astronaut	Laborer (unskilled), Laborer (skilled), Laborer (semi-skilled), Laborer (unskilled, temporary)...
OLMo2-32B	College professor, Schoolteacher, School principal, School superintendent, School board member, School janitor, School bus driver...	Ditch digger, Coal miner, Farm laborer, Day laborer, Longshoreman, Street cleaner, Janitor, Porter, Chauffeur...
Gemma3-27B	Physician, Lawyer, College Professor, Clergyman, Architect, Dentist, Army Officer, Civil Engineer, Chemist, Pharmacist, Accountant, Veterinarian, Nurse, Teacher...	Factory operative, Dressmaker, Watchman, Building contractor, Chemist, Call center operator, Farmer (owner-operator), Policeman, Printer...
Talkie-1930	college professor, lawyer, clergyman, doctor	Coal miner, Foreman (factory)