Bourdieu Vectors: A Framework for Quantifying Taste in Social Space

Simon Hecht
Vienna
simon@hecht.dev

***Abstract*—This paper introduces *Bourdieu Vectors*, a computational publicly accessible framework (https://bourdieuvectors.com) for estimating the social, cultural, economic, and symbolic capital of individuals based on the cultural practices they engage in. Each practice (e.g., football, opera, cooking) is represented as a multidimensional vector capturing likely levels of various forms of capital. Vectors are generated via prompt-based querying of the Gemini-2.5-Flash-Lite model.**

Bourdieu Vectors enable systematic analysis of social practices, measurement of similarity between practices, construction of individual or group taste profiles, and generation of personalized recommendations. The framework is intended for use in social sciences, digital humanities, market research, and cultural analytics, emphasizing transparency, reproducibility, and ethical responsibility.

***Index Terms*—quantifying taste, vectors, embeddings, social space, Bourdieu, social capital, cultural analytics, recommendation systems**

I. Introduction

Pierre Bourdieu's theory of social space conceptualizes society as a multidimensional field structured by forms of capital (economic, cultural, social, symbolic) and the distribution of practices, preferences, and resources [Bourdieu, 1984], [Bourdieu, 1986]. Traditional operationalizations often focused on the characteristics of the practices themselves.

Bourdieu Vectors shifts the perspective to the individual, estimating the likely capital profile of someone engaging in a given practice [Bourdieu, 1984], [Bourdieu, 1986]. This enables translation of qualitative descriptions of cultural participation into quantitative vectors, allowing systematic comparison, visualization, computational modeling, and analysis of social tastes.

II. Methodology

The approach builds upon Pierre Bourdieu's theory of capitals, which conceptualizes social position as determined by the distribution of different forms of capital—cultural, economic, social, and symbolic [Bourdieu, 1986]. To operationalize these concepts computationally, each cultural practice is represented as a multidimensional capital vector. Each dimension (section A) corresponds to a specific subcomponent of capital (e.g., educational attainment, material resources, prestige, network influence), and takes a continuous value between 0.0 and 1.0, reflecting its relative strength.

For each text describing a cultural or social practice, the Gemini-2.5-Flash-Lite Large Language Model (LLM) is prompted (section B) to estimate the capital profile of a hypothetical individual engaging in that practice. The model is instructed to assign numerical scores between 0.0 and 1.0 for each capital dimension based on explicit textual cues. It also provides aggregated "overview" scores summarizing the overall levels of cultural, economic, social, and symbolic capital. Optionally, a contextual prompt (e.g., demographic or situational background) can be supplied to guide the model toward more contextually grounded interpretations.

All model outputs are structured as JSON-formatted vectors. This procedure yields a machine-interpretable representation of how various cultural activities map onto Bourdieu's multidimensional capital space.

These capital vectors can subsequently be used for clustering, similarity analysis, and the visualization of social fields. By embedding practices in a shared semantic-capital space, the method allows for the empirical exploration of distinctions, hierarchies, and affinities between social activities.

Applications include:

- *Measuring similarity:* Quantifying how similar two practices are using metrics such as cosine similarity (e.g., "American football" vs. "basketball").
- *Generating recommendations:* Suggesting complementary interests, activities, or cultural events based on a user's taste profile.
- *Interest profiling:* Averaging multiple vectors to construct an individual's or group's overall cultural, social, and economic taste profile.
- *Cultural analysis:* Comparing practices across demographic segments, regions, or historical periods.
- *Market research:* Segmenting audiences and quantifying product appeal based on taste vectors.
- *Education and policy:* Understanding access to different forms of capital and guiding interventions.

III. Datasets Provided

The following datasets are provided[1] for testing and analysis of Bourdieu Vectors:

- **bourdieuvectors-ds-50-high-low-eyebrow (Version 1.0.0)**: Contains 50 social preferences, activities, or likes, divided equally into 25 high and 25 low samples. Each set also distinguishes high eyebrow vs. low eyebrow samples. Language: en.
- **bourdieuvectors-ds-75-high-mid-low-eyebrow (Version 1.0.0)**: Contains 75 social preferences, activities, or likes, divided equally into high, mid, and low samples. Language: en.
- **bourdieuvectors-ds-1000-high-mid-low-eyebrow (Version 1.0.1)**: Contains 1000 social preferences, activities,

[1] https://bourdieuvectors.com/datasets/

or likes, divided into high (334), mid (333), and low (333) samples. Language: en.
- **bourdieuvectors-ds-100 (Version 1.0.0)**: Contains 100 social preferences, activities, or likes. Language: en.
- **bourdieuvectors-ds-100-bourdieu-preferences (Version 1.0.0)**: Contains 100 social preferences, activities, or likes specifically categorized as: 10 economic_capital_low, 10 economic_capital_high, 10 cultural_capital_low, 10 cultural_capital_high, 10 social_capital_low, 10 social_capital_high, 10 symbolic_capital_low, 10 symbolic_capital_high, 10 habitus_alignment_aligned, and 10 habitus_alignment_misaligned. Language: en.
- **bourdieuvectors-ds-250 (Version 1.0.0)**: Contains 250 social preferences, activities, or likes. Language: en.

The datasets were manually compiled with the help of the ChatGPT5-Mini LLM model. For example, a prompt used to generate parts of the datasets was:

```
Create a dataset of 50 social preferences,
activities, or likes, divided equally into
25 samples associated with upper-class and
25 samples associated with lower-class.
```

These datasets support testing, validation, and demonstration of Bourdieu Vector analyses across cultural, social, economic, symbolic, and habitus dimensions.

IV. Case Study: Comparing Soccer, Tennis, and Hiking Vectors

To illustrate the utility of Bourdieu Vectors, we compare the social profiles associated with three activities: *soccer*, *tennis*, and *hiking*. Vectors are generated using the Gemini-2.5-Flash-Lite model, capturing likely levels of cultural, social, economic, and symbolic capital for individuals engaging in each activity.

A. Vector Comparison

Table I shows a selection of key dimensions from the soccer, tennis, and hiking vectors.

Dimension	Soccer	Tennis	Hiking
Cultural Entertainment	0.6	0.5	0.3
Sport Competition	0.8	0.6	0.3
Sport Fitness	0.7	0.7	0.5
Sport Team	0.8	0.3	0.1
Sport Individual	0.2	0.4	0.4
Economic Assets	0.3	0.4	0.2
Overall Cultural Capital	0.26	0.33	0.25
Overall Economic Capital	0.23	0.33	0.20
Overall Habitus Alignment	0.6	0.5	0.35
Overall Social Capital	0.17	0.3	0.10
Overall Symbolic Capital	0.2	0.4	0.20

TABLE I
SELECTED BOURDIEU VECTOR DIMENSIONS FOR SOCCER, TENNIS, AND HIKING.

B. Interpretation

The three activities show distinct profiles:

- *Soccer:* Emphasizes team orientation, sport fitness, and social capital, reflecting its collaborative and community-based nature. Cultural and symbolic capital are also relatively high.
- *Tennis:* Highlights individual performance and higher economic and symbolic capital, consistent with a more resource-intensive and competitive activity.
- *Hiking:* Prioritizes outdoor and fitness engagement with moderate cultural capital, but minimal social and symbolic capital. Environmental and preventive health associations are stronger, aligning with a solitary, wellness-oriented activity.

For an individual participating in all three activities, averaging the vectors provides a composite profile:

- *Habitus and motivation:* High across all three activities, indicating a physically active, goal-oriented, and health-conscious lifestyle.
- *Social and symbolic capital:* Soccer contributes team and local networks, tennis adds symbolic and economic capital, while hiking contributes little social or symbolic capital but emphasizes environmental and preventive health dimensions.
- *Overall profile:* Balanced cultural and habitus alignment, moderate economic capital (driven by tennis), low-to-moderate social capital, and diversified symbolic capital. The combination reflects team sports, individual sports, and outdoor wellness activities.

V. Results

This section presents the evaluation results of the *bourdieuvectors-ds-1000-high-mid-low-eyebrow* (v1.0.1) dataset conducted with Bourdieu Vectors v1.0.10. The analysis focuses on the separability and internal cohesion of the three hypothesized groups: *High* (cultural and economic capital), *Middle*, and *Low* (cultural and economic capital).

A. Visual Inspection and Capital Distribution

Dimensionality reduction and bi-dimensional capital distribution plots in Figure 1 reveal a clear systematic separation of the groups based on the main capital axes. The Principal Component Analysis (PCA) shows that the highest variance (Component 1) is strongly correlated with the overall capital level, with the *Low* group clustering negatively and the *High* group clustering positively along this axis. The radar scatter chart (Figure 2) confirms that the primary distinction is driven by high values of *overall_economic_capital* and *overall_cultural_capital*, which are predominantly occupied by the *High* group. Conversely, the *Low* group consistently maps to the lowest values across all capital dimensions.

B. Cluster Separability and Cohesion

1) Internal Cohesion and Pairwise Similarity: The analysis of intra-cluster similarity (Table II) demonstrates strong internal cohesion within the groups, particularly in terms of directional alignment among the vectors. This table shows the average similarity and variance for each metric within clusters, providing an overview of how cohesive the clusters are.

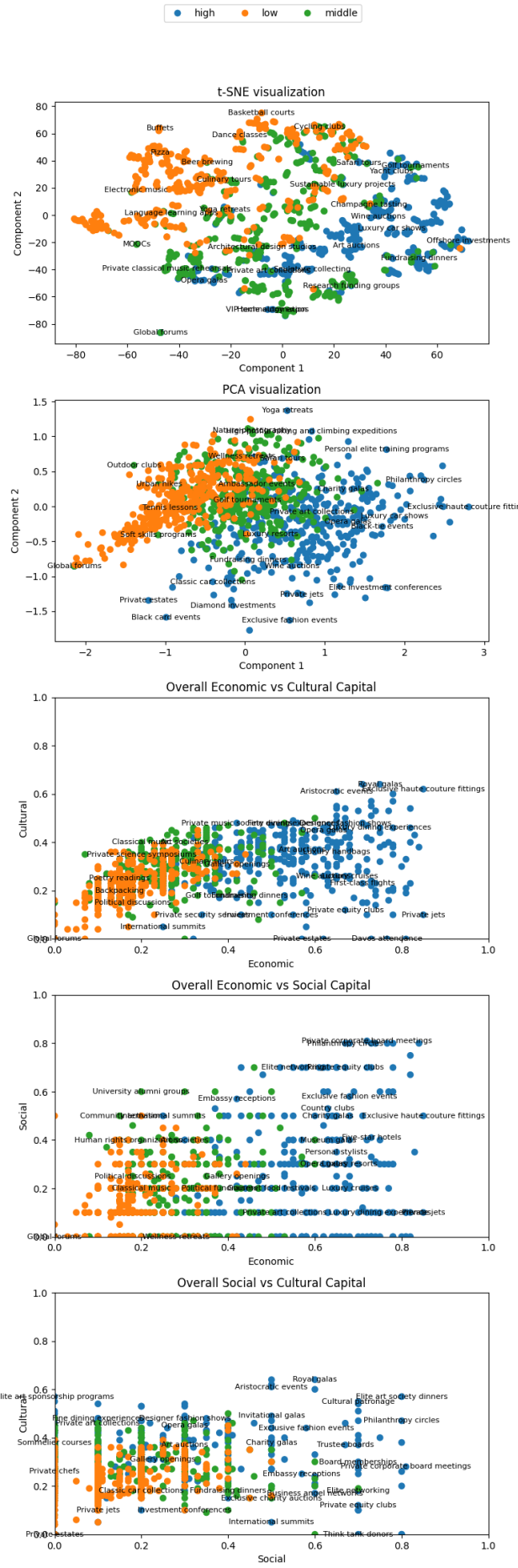

Fig. 1. An

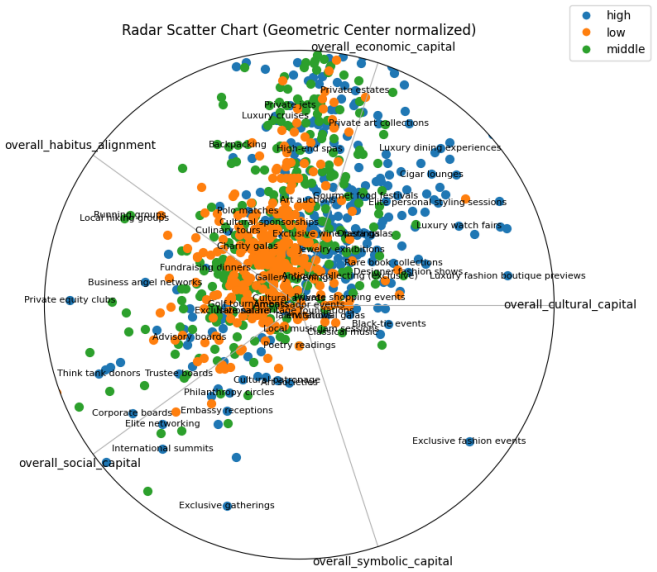

Fig. 2. An

TABLE II
INTRA-CLUSTER SIMILARITY AND VARIANCE

Metric	Average Similarity	Variance
Cosine Similarity	0.782	0.023
Pearson's Correlation	0.416	0.050
Euclidean Similarity	0.360	0.005

2) Inter-Group Similarity: The pairwise inter-group similarity (Table III) shows how closely the different clusters relate to one another across three metrics:

- **Cosine similarity:** Measures the angle between vectors, indicating how similar their directions are.
- **Pearson's correlation:** Measures the linear relationship between vectors, capturing correlated patterns.
- **Euclidean similarity:** Measures the inverse distance between vectors, reflecting overall spatial closeness.

High similarity values indicate that the groups share similar patterns or distributions, while lower values suggest greater distinction between clusters. This helps assess whether the model effectively separates different cultural events or topics within the embedding space.

TABLE III
INTER-GROUP SIMILARITY

Group Comparison	Cosine	Pearson's	Euclidean
high_middle	0.792	0.409	0.331
high_low	0.723	0.279	0.305
middle_low	0.781	0.407	0.368

Overall, the *High* and *Low* clusters are the most distinct, exhibiting the lowest similarity across all metrics. The *Middle* group shows moderate similarity with both the *High* and *Low* groups, indicating its function as an intermediate cluster within the embedding space.

3) Statistical Significance and Separation Quality:

a) PERMANOVA Results.: The Permutational Multivariate Analysis of Variance (PERMANOVA) was applied to

assess whether the multivariate distances between groups differ significantly from what would be expected by chance. Unlike traditional ANOVA, which tests for differences in the means of a single variable, PERMANOVA operates on distance or dissimilarity matrices derived from multiple variables simultaneously. This makes it particularly suitable for high-dimensional data such as embeddings, microbiome profiles, or other multivariate measurements.

The test works by permuting the group labels many times and recalculating the test statistic for each permutation. This process produces a null distribution, allowing for the computation of a p-value that reflects the probability of observing the measured between-group differences under the null hypothesis of no group effect. A low p-value (typically ≤ 0.05) indicates that the observed grouping structure is unlikely to have arisen by chance, signifying significant group separation with respect to the chosen distance metric.

TABLE IV
PERMANOVA RESULTS COMPARING GROUP DIFFERENCES.

Metric	**Value**
Test statistic (pseudo-F)	151.0453
Sample size (n)	1000
Number of groups	3
p-value	0.001
Number of permutations	999

The PERMANOVA results indicate a highly significant difference between the group distances ($p = 0.001$), confirming that the observed separation between clusters is not due to random variation.

b) Silhouette Score per Group.: The Silhouette Score measures how well each cluster's points are separated from those of other clusters.

- Score ≈ 1: very well separated
- Score ≈ 0: clusters overlap strongly
- Score < 0: points are closer to other groups than their own

TABLE V
SILHOUETTE SCORE PER GROUP

Group	Silhouette Score
High	0.145
Low	0.004
Middle	0.149

The Silhouette Scores indicate limited cluster separability across groups. All scores are positive or near zero, suggesting weak but existent structure within the data. The *Low* group shows a near-zero score (0.004), implying considerable overlap with neighboring clusters. In contrast, the *Middle* group achieves the highest score (0.149), reflecting slightly stronger internal cohesion and distinction. Overall, the modest values across groups suggest that the clusters are only weakly defined.

C. Exploratory Cluster Analysis

The clustering analysis detected three clusters, which aligns with the hypothesized group structure (*High*, *Middle*, *Low*). The Adjusted Rand Index (ARI = 0.302) and Normalized Mutual Information (NMI = 0.316) indicate moderate agreement between the detected clusters and the ground truth labels, suggesting that the clustering algorithm captures meaningful groupings but does not perfectly recover the true labels. The Explained Variance Ratio (0.326) shows that approximately one-third of the total variance in the data is explained by the clustering solution, indicating that while the clusters capture some of the underlying structure, substantial variability remains unaccounted for. Overall, these results reflect a partial but interpretable recovery of the underlying group structure, consistent with the expected gradation between high, middle, and low capital groups.

TABLE VI
EXPLORATORY CLUSTERING METRICS

Metric	Value
Number of Clusters	3
Adjusted Rand Index (ARI)	0.302
Normalized Mutual Information (NMI)	0.316
Explained Variance Ratio	0.326

D. Multivariate Analysis of Variance (MANOVA)

MANOVA confirmed significant differences across multiple dimensions. Overall tests (Wilks' lambda, Pillai's trace, Hotelling-Lawley trace, and Roy's greatest root) yielded extremely low p-values ($< 10^{-199}$), indicating robust group separation in the multivariate space. Posthoc hierarchy tests show that 100% of the overall dimensions and 81.6% of selected dimensions adhere to the expected hierarchical ordering, validating the theoretical capital distinctions.

E. Feature Contribution Analysis

The Feature Contribution Analysis identifies the dimensions most critical for group separation, based on the positive PERMANOVA f-drop values. Dimensions associated with economic capital show the strongest positive contributions to group distinctiveness:

1) *economic_capital_economic_luxury* ($\Delta F = 8.933$)
2) *cultural_capital_cultural_art* ($\Delta F = 7.462$)
3) *economic_capital_economic_investments* ($\Delta F = 5.939$)
4) *cultural_capital_cultural_culinary* ($\Delta F = 5.720$)
5) *economic_capital_economic_wealth* ($\Delta F = 5.041$)
6) *cultural_capital_cultural_entertainment* ($\Delta F = 4.470$)
7) *cultural_capital_media_social* ($\Delta F = 4.139$)
8) *economic_capital_power_economic* ($\Delta F = 4.096$)
9) *cultural_capital_media_entertainment* ($\Delta F = 3.511$)
10) *economic_capital_economic_assets* ($\Delta F = 3.311$)

These results indicate that both economic and cultural capital dimensions substantially contribute to group differentiation. The strongest effects are observed for *economic_capital_economic_luxury* and *cultural_capital_cultural_art*, highlighting that financial exclusivity and artistic engagement are key discriminators among the groups. Other dimensions—particularly those tied to investments, culinary culture, and general economic wealth—also meaningfully contribute to separation.

At the same time, several dimensions show only modest effects (e.g., *social_capital_community_local*, $\Delta F = 2.063$; *economic_capital_economic_income*, $\Delta F = 1.899$; *cultural_capital_educational_formal*, $\Delta F = 1.254$), indicating that their influence on group distinction is comparatively limited.

Moreover, other capital forms—such as *symbolic_capital_power_professional* ($\Delta F = 0.518$)—also exert some influence, but their overall contribution remains weak. This suggests that while economic and cultural capitals are the main axes of differentiation, social and symbolic capitals provide additional, though less pronounced, layers of structure within the data.

F. Robustness

The Bootstrap Robustness analysis confirmed the stability of the cluster structure under resampling (80%, $n = 100$ runs). The correlation of the radar means was exceptionally high (mean = 1.000, SD = 0.000, range = 0.999–1.000), indicating perfect consistency across resampled subsets. Similarly, the resampled PERMANOVA pseudo-F mean (121.654, SD = 5.575) remained close to the original value, demonstrating that both the feature importance and group separation are highly stable and reliable under data resampling.

VI. Ethical Considerations

The framework avoids profiling based on political ideology or other sensitive attributes. Scores are strictly limited to estimated capital levels, and no personally identifiable information is used.

Additional considerations:

- **Bias Awareness:** LLM-based estimations may reflect historical or cultural biases; scores are influenced by typical social patterns rather than individual characteristics.
- **Transparency:** All prompts, aggregation formulas, and weighting schemes are openly documented on the project website, ensuring full methodological traceability.
- **Avoiding stereotyping:** Scores reflect likely capital levels associated with practices, not labels for individuals.

VII. Discussion

The results demonstrate that Bourdieu Vectors offer a robust computational framework for operationalizing Pierre Bourdieu's theory of social space through vector representations of cultural practices. The framework effectively quantifies and visualizes the relationships between practices, capturing variations in cultural, economic, social, and symbolic capital with statistically significant group separation (PERMANOVA: pseudo-F = 151.0453, $p = 0.001$).

The clustering and exploratory analysis revealed that while the *High*, *Middle*, and *Low* groups differ significantly in multivariate space, the boundaries between them are gradual rather than discrete. Moderate cluster validation metrics (ARI = 0.302, NMI = 0.316, Explained Variance Ratio = 0.326) do not undermine these distinctions but rather empirically reflect the continuous and relational structure of social space, as posited by Bourdieu. The *Middle* cluster functions as a transitional zone between the poles of cultural and economic privilege, confirming the theory's expectation of graded rather than rigid social differentiation.

Silhouette Scores further underscore this continuum-like distribution. While the *High* and *Middle* groups exhibit modest internal cohesion, the *Low* cluster shows near-zero separation, highlighting substantial overlap with adjacent clusters. These results suggest that real-world taste and capital distributions are rarely rigid, and that gradational overlap is a meaningful empirical feature rather than a methodological limitation.

MANOVA and posthoc hierarchy analyses corroborate the robustness of these distinctions across multiple dimensions. Overall tests (Wilks' lambda, Pillai's trace, Hotelling-Lawley trace, Roy's greatest root) yielded extremely low p-values ($< 10^{-199}$), confirming statistically significant differentiation across the full set of capital dimensions. Posthoc analyses indicate that 100% of the main dimensions and 81.6% of selected dimensions adhere to the expected hierarchical ordering, providing strong support for the theoretically predicted structure of social space.

The Feature Contribution Analysis highlights the dominant influence of economic and cultural capital dimensions in defining group separability. While economic and cultural capital emerged as the most discriminative features, social and symbolic capital contributed less consistently. This may partly reflect methodological limitations of the LLM in extrapolating abstract relational qualities, but it also invites a theoretical interpretation: in contemporary digital social spaces, symbolic distinction is often mediated through economic and cultural markers, whereas social and symbolic capital operate more diffusely across groups. Thus, the weaker signals for these dimensions may illuminate genuine patterns of cross-cutting participation rather than model failure.

Bootstrap Robustness analysis confirmed the model's stability under repeated resampling, with near-perfect radar mean correlations (1.000 ± 0.000) and consistent pseudo-F values across iterations. This supports the reproducibility and methodological soundness of embedding-based approaches for representing social space.

To move beyond static representations, a weighted, context-specific practice vector could be defined as

$$\mathbf{V}_{\text{Total}} = w_{\text{Core}} \cdot \mathbf{V}_{\text{Core}} + w_{\text{Context}} \cdot \mathbf{V}_{\text{Context}},$$

where $\mathbf{V}_{\text{Core}}$ is the LLM-generated core practice vector and $\mathbf{V}_{\text{Context}}$ encodes context-specific information (e.g., subscription vs. public consumption). This formulation allows the model to dynamically adjust core capital scores to reflect modes of participation, aligning with Bourdieu's relational understanding of social space.

The framework has inherent limitations due to the nature of LLM-based estimation. The vectors inevitably reflect social biases, stereotypes, and hierarchical assumptions present in the training corpora. However, rather than being purely methodological weaknesses, these biases reveal the internalized classifications of society that shape cultural perception and symbolic distinction. The model thus functions as a sociological lens

on collective perception, complementing rather than replacing empirical measures of individual capital.

Overall, Bourdieu Vectors bridge qualitative sociological theory and computational modeling. By encoding cultural practices into interpretable, quantitative representations, the framework enables large-scale, data-driven investigations into how taste operates as a marker of social distinction in contemporary digital societies. The integration of exploratory clustering, MANOVA, and posthoc hierarchy analyses reinforces the validity of the framework, demonstrating both statistically and conceptually meaningful patterns of social differentiation while acknowledging the nuanced gradation and relational complexity inherent in real-world social space.

VIII. Summary

Bourdieu Vectors present a novel, transparent, and reproducible method for quantifying social differentiation through the lens of Pierre Bourdieu's theory of capitals. By leveraging large language model–based estimations of cultural, economic, social, and symbolic capital, the framework translates qualitative cultural practices into interpretable, quantitative embeddings suitable for rigorous empirical analysis.

Evaluation on the *bourdieuvectors-ds-1000-high-mid-low-eyebrow* dataset demonstrates that the method captures meaningful and statistically significant distinctions between social groups (PERMANOVA, $p = 0.001$). These results show that while the *High*, *Middle*, and *Low* groups are clearly differentiated in multidimensional capital space, their boundaries remain fluid and gradational—reflecting Bourdieu's conception of social space as continuous rather than categorical. Moderate cluster cohesion metrics (ARI = 0.302, NMI = 0.316) support this interpretation, indicating structured yet overlapping regions of taste and capital distribution.

Feature contribution analysis highlights the dominant roles of economic and cultural capital in defining social distinctions, with symbolic and social capital adding more subtle layers of variation. Robustness tests confirm the stability and reproducibility of the framework across repeated resampling, reinforcing its methodological reliability.

The transparency of the approach, along with its openly available datasets, makes Bourdieu Vectors an adaptable tool for social science, digital humanities, and cultural analytics research. Future developments may include multimodal extensions, dynamic or longitudinal modeling of cultural change, and integration with empirical survey or behavioral data to contextualize habitus and lifestyle trajectories.

In essence, Bourdieu Vectors operationalize the relational logic of social space through computational means. They provide a scalable bridge between qualitative sociological theory and quantitative modeling, enabling systematic visualization, comparison, and analysis of how taste and capital structure the symbolic order of contemporary digital societies.

IX. Full Disclosure

This paper has been prepared with the assistance of ChatGPT, which was used to help formulate and refine the wording of the text. The underlying ideas, analyses, and interpretations are entirely my own.

References

[Bourdieu, 1984] Bourdieu, P. (1984). *Distinction: A Social Critique of the Judgement of Taste*. Harvard University Press.

[Bourdieu, 1986] Bourdieu, P. (1986). The forms of capital. *Handbook of Theory and Research for the Sociology of Education*, pages 241–258.

Appendix A
Vector Dimensions

The following dimensions are used for each Bourdieu Vector:

```
cultural_capital_cultural_art,
cultural_capital_cultural_culinary,
cultural_capital_cultural_design,
cultural_capital_cultural_entertainment,
cultural_capital_cultural_fashion,
cultural_capital_cultural_literature,
cultural_capital_cultural_music,
cultural_capital_cultural_theater,
cultural_capital_educational_certifications,
cultural_capital_educational_formal,
cultural_capital_educational_informal,
cultural_capital_educational_teaching,
cultural_capital_educational_training,
cultural_capital_environmental_conservation,
cultural_capital_environmental_
sustainability,
cultural_capital_innovation_creativity,
cultural_capital_innovation_openness,
cultural_capital_lifestyle_gaming,
cultural_capital_lifestyle_outdoor,
cultural_capital_lifestyle_travel,
cultural_capital_media_digital,
cultural_capital_media_entertainment,
cultural_capital_media_journalism,
cultural_capital_media_social,
cultural_capital_media_traditional,
cultural_capital_medical_fitness,
cultural_capital_medical_healthcare,
cultural_capital_medical_prevention,
cultural_capital_medical_research,
cultural_capital_religious_affiliation,
cultural_capital_religious_community,
cultural_capital_religious_practice,
cultural_capital_religious_rituals,
cultural_capital_sport_competition,
cultural_capital_sport_fitness,
cultural_capital_sport_individual,
cultural_capital_sport_outdoor,
cultural_capital_sport_team,
economic_capital_economic_assets,
economic_capital_economic_entrepreneurship,
economic_capital_economic_income,
economic_capital_economic_investments,
economic_capital_economic_luxury,
economic_capital_economic_wealth,
economic_capital_power_economic,
habitus_subjective_emotional_intelligence,
```

```
habitus_subjective_lifestyle_alignment,
habitus_subjective_motivation,
habitus_subjective_resilience,
habitus_subjective_satisfaction,
habitus_subjective_selfesteem,
habitus_subjective_values,
overall_cultural_capital,
overall_economic_capital,
overall_habitus_alignment,
overall_social_capital,
overall_symbolic_capital,
social_capital_community_local,
social_capital_community_volunteering,
social_capital_legal_judiciary,
social_capital_legal_law_enforcement,
social_capital_legal_legislation,
social_capital_legal_profession,
social_capital_network_professional,
social_capital_network_social,
social_capital_political_activism,
social_capital_political_engagement,
social_capital_political_government,
social_capital_political_parties,
social_capital_power_political,
symbolic_capital_power_cultural,
symbolic_capital_power_educational,
symbolic_capital_power_professional,
symbolic_capital_power_public,
symbolic_capital_power_social,
symbolic_capital_power_symbolic
```

APPENDIX B
PROMPT FOR GEMINI-2.5-FLASH

The prompt used to generate Bourdieu Vectors:

```
You are an expert in sociology and cultural
    analysis, familiar with Pierre Bourdieu's
    theory of social space.
The social space can be described as a
    multidimensional field structured by forms
     of capital and practices.

You will be given a text describing a cultural
     practice, preference (something someone
    likes), or activity.
Your task is to estimate the social profile of
     a person who likes this practice or
    preference by evaluating how much of each
    type of capital they likely possess,
    according to the predefined dimensions in
    'vector_dimensions'.

Rules for scoring:

1. For each dimension, assign a score between
    0.0 (very low capital) and 1.0 (very high
    capital) that reflects the likely level of
     that capital in a person who likes this
    activity or preference.
2. Only assign high scores (>0.5) if the text
    implies that individuals who likes in this
     practice are likely to have significant
    capital in that dimension.
3. Consider context: hobbies, interests, or
    popular activities may indicate moderate
    or low capital unless explicitly
    associated with high social, economic,
    cultural, or symbolic status.
4. Assume an individual-level profile. Do not
    infer institutional or organizational
    involvement unless explicitly stated.
5. Map the text carefully to each subdimension
    ; assign 0.0 if the implied capital is
    unclear.
6. Compute 'overview_*' dimensions as the mean
     of the related subdimensions; do not
    assign overview scores arbitrarily.
7. Do not inflate scores based on popularity,
    media attention, or general familiarity.
8. Return only valid JSON with one score per
    dimension and the two letter iso language
    code of the cultural practice - no
    explanation, no extra text.

Text:
{social_practice}
```

The prompt is extended when a context is set:

```
Context of the practice or individuals:
{social_practice_context}
```

Note that the dimensions are included in the JSON response schema.

APPENDIX C
APPENDIX: FULL VECTORS FOR SOCCER, TENNIS, AND HIKING

Dimension	Soccer	Tennis	Hiking
cultural_capital_cultural$_{a}rt$	0.0	0.3	0.1
cultural_capital_cultural_culinary	0.2	0.2	0.2
cultural_capital_cultural_design	0.0	0.3	0.1
cultural_capital_cultural_entertainment	0.6	0.5	0.3
cultural_capital_cultural_fashion	0.1	0.4	0.1
cultural_capital_cultural_literature	0.0	0.2	0.1
cultural_capital_cultural_music	0.3	0.4	0.2
cultural_capital_cultural_theater	0.0	0.2	0.1
cultural_capital_educational_certifications	0.2	0.3	0.2
cultural_capital_educational_formal	0.2	0.3	0.2
cultural_capital_educational_informal	0.3	0.3	0.4
cultural_capital_educational_teaching	0.1	0.3	0.1
cultural_capital_educational_training	0.2	0.3	0.2
cultural_capital_environmental_conservation	0.0	0.1	0.4
cultural_capital_environmental_sustainability	0.0	0.1	0.3
cultural_capital_innovation_creativity	0.3	0.4	0.3
cultural_capital_innovation_openness	0.4	0.4	0.4
cultural_capital_lifestyle_gaming	0.5	0.2	0.1
cultural_capital_lifestyle_outdoor	0.4	0.5	0.7
cultural_capital_lifestyle_travel	0.3	0.5	0.4
cultural_capital_media_digital	0.6	0.4	0.2
cultural_capital_media_entertainment	0.7	0.5	0.2
cultural_capital_media_journalism	0.2	0.3	0.1
cultural_capital_media_social	0.6	0.4	0.2
cultural_capital_media_traditional	0.4	0.4	0.1
cultural_capital_medical_fitness	0.7	0.7	0.4
cultural_capital_medical_healthcare	0.1	0.3	0.2
cultural_capital_medical_prevention	0.2	0.5	0.3
cultural_capital_medical_research	0.0	0.2	0.1
cultural_capital_religious_affiliation	0.1	0.1	0.1
cultural_capital_religious_community	0.1	0.1	0.1
cultural_capital_religious_practice	0.1	0.1	0.1
cultural_capital_religious_rituals	0.1	0.1	0.1
cultural_capital_sport_competition	0.8	0.6	0.3
cultural_capital_sport_fitness	0.7	0.7	0.5
cultural_capital_sport_individual	0.2	0.4	0.4
cultural_capital_sport_outdoor	0.4	0.6	0.7
cultural_capital_sport_team	0.8	0.3	0.1
economic_capital_economic_assets	0.3	0.4	0.2
economic_capital_economic_entrepreneurship	0.1	0.3	0.1
economic_capital_economic_income	0.3	0.4	0.3
economic_capital_economic_investments	0.2	0.3	0.2
economic_capital_economic_luxury	0.1	0.3	0.1
economic_capital_economic_wealth	0.3	0.4	0.2
economic_capital_power_economic	0.3	0.4	0.2
habitus_subjective_emotional_intelligence	0.5	0.5	0.3
habitus_subjective_lifestyle_alignment	0.6	0.5	0.4
habitus_subjective_motivation	0.6	0.5	0.4
habitus_subjective_resilience	0.5	0.5	0.4
habitus_subjective_satisfaction	0.5	0.5	0.3
habitus_subjective_selfesteem	0.5	0.5	0.3
habitus_subjective_values	0.5	0.5	0.3
overall_cultural_capital	0.26	0.33	0.25
overall_economic_capital	0.23	0.33	0.20
overall_habitus_alignment	0.6	0.5	0.35
overall_social_capital	0.17	0.3	0.10
overall_symbolic_capital	0.2	0.4	0.20
social_capital_community_local	0.3	0.3	0.2
social_capital_community_volunteering	0.1	0.3	0.2
social_capital_legal_judiciary	0.0	0.1	0.1
social_capital_legal_law_enforcement	0.1	0.1	0.1
social_capital_legal_legislation	0.0	0.1	0.1
social_capital_legal_profession	0.0	0.1	0.1
social_capital_network_professional	0.2	0.3	0.1
social_capital_network_social	0.4	0.4	0.2
social_capital_political_activism	0.1	0.2	0.1
social_capital_political_engagement	0.2	0.2	0.1
social_capital_political_government	0.1	0.1	0.1
social_capital_political_parties	0.1	0.1	0.1
social_capital_power_political	0.1	0.1	0.1
symbolic_capital_power_cultural	0.3	0.4	0.1
symbolic_capital_power_educational	0.2	0.4	0.1
symbolic_capital_power_professional	0.2	0.4	0.1
symbolic_capital_power_public	0.2	0.4	0.1
symbolic_capital_power_social	0.3	0.4	0.1
symbolic_capital_power_symbolic	0.2	0.4	0.1

TABLE VII

FULL BOURDIEU VECTORS FOR SOCCER, TENNIS, AND HIKING WITH UPDATED VALUES FROM MODEL *bourdieuvectors 1.0.10*, INCLUDING ALL 76 STANDARDIZED DIMENSIONS.