Large Language Model Prediction Ability: Narrative vs Variable

Joy Qu
Yale University
joy.qu@yale.edu

December 2025

1 Motivation

Rsearch Question: Are large language models (LLMs) intrinsically better at predicting outcomes from narrative descriptions than from explicitly defined variables?

1. **Social Science and Policy Use**: Many real-world applications of LLMs rely on qualitative descriptions, not clean tabular data. If LLMs systematically perform better on narrative inputs, this would support their use as tools for prediction on qualitative data. On the other hand, if variable-based representations perform equally well or better, it strengthens the case for continued reliance on traditional variable-based models.

2.

2 Hypothesis

We hypothesize that the narrative model will perform better than the variable model by the following three arguments:

1. **Latent information**
Narrative descriptions may encode latent traits and constraints that are not directly observed in the variables. When variables are explicitly defined, they necessarily reflect the researcher's choice of dimensions, potentially omitting relevant but unmeasured factors. However, narratives allow these latent characteristics to be expressed implicitly through context, routines, and relationships. If LLMs are able to generate and infer latent information from text, we hypothesize narrative-based inputs may produce superior predictions even when the observable variables appear comparable.

2. **Prior information**
LLMs are trained primarily on natural language and therefore learn common patterns that link everyday behaviors, social settings, and typical outcomes, which is the prior. Narrative descriptions resemble the kinds of situations the model encoun-

tered during training, making it easier for the model to recognize familiar scenarios and apply these learned patterns. By contrast, a variable model may need to translate numerical or categorical inputs into meaningful contexts, which may weaken its ability to use this background knowledge. As a result, we hypothesize stronger performance on narrative inputs may reflect better alignment between the model's training experience and the input.

3. **Variable interaction**
   Narratives may implicitly encode complex, nonlinear interactions among variables that are difficult to specify explicitly. In tabular data for the variable model, interactions must be pre-defined or estimated from limited samples, often leading to misspecification or over-simplification. Narrative descriptions naturally combine multiple dimensions such as age, work environment, family structure, and routines into coherent scenarios. If LLMs can internally represent and exploit these interactions, narrative-based prediction may outperform variable-based approaches that fail to capture higher-order dependencies.

3 Methodology

Variable Model
We use data from the 2024 General Social Survey (GSS). At a high level, we use an LLM to predict political ideology (polviews) from structured demographic variables.

We load an unlabeled GSS random sample of 100 people, formats each respondent's demographics into a standardized prompt, and queries the model to place the individual on a 7-point liberal–conservative scale. We treat the LLM as a variable-based prediction model, producing ideology estimates directly from coded covariates.

We run this on different number of categories of the outcome variable, starting with the default GSS of seven categories for polviews:
GSS outcome variable: POLLVIEWS Measures the individual's political ideology on a discrete 7-point scale 1 = Extremely liberal 2 = Liberal 3 = Slightly liberal 4 = Moderate 5 = Slightly conservative 6 = Conservative 7 = Extremely conservative

For predictors, we start with 7 predictors which were manually chosen.
Age (AGE): (numeric age in years: 18–89; 89 = topcoded; 0/99 = missing)
Gender (SEX): (1 = Male, 2 = Female)
Race(RACE): (1 = White, 2 = Black, 3 = Other)
Education (EDUC): (years of schooling; e.g. 12 = high school, 16 = college, 18 = master's, 19–20 = professional/PhD)
Marital Status (MARITAL): (1 = Married, 2 = Widowed, 3 = Divorced, 4 = Separated, 5 = Never married)
Occupation (OCC10): (0010–0950 = Management/Professional, 1000–1240 = Service, 1300–1965 = Sales/Office, 2000–3955 = Construction/Maintenance, 4000–5940 = Production/Transportation, 5950–9750 = Military)
Region (REGION): (1 = Northeast, 2 = Midwest, 3 = South, 4 = West)

Then, we collapse the outcome variable into binary, three, four and five categories as so:

Binary Polviews
I collapse the 7-category POLVIEWS into a binary variable, where 1,2,3,4 are classified as 0 = not conservative and 5,6,7 are classified as 1 = conservative.

Three Category Polviews
I collapse the 7-category POLVIEWS into a three-category variable, where 1,2,3 are classified as 1= liberal, 4 as 2 = moderate, 5,6,7 as 3 =conservative.

Four Category Polviews
I collapse the 7-category POLVIEWS into a four-category variable, where 1,2 are classified as 1 = extremely liberal, 3 as 2 = slightly liberal, 4 as 3 = moderate, and 5,6,7 as 4 = conservative.

Five Category Polviews
I collapse the 7-category POLVIEWS into a five-category variable, where 1 is classified as 1 = extremely liberal, 2,3 as 2 = liberal, 4 as 3 = moderate, and 5,6 as 4 = conservative, 7 as 5 = extremely conservative.

Next, we modify the number of predictors.

3 predictor model: age, race, sex

Narrative Model

We use an LLM to generate vignettes for the same 100-person sample as used in the variable model with the following prompt:

Then, we run Python code using API key to query the LLM on an individual by individual basis in the sample. In other words, the LLM is queried one time per person in the sample to generate a prediction for that person.

To test each hypothesis argument, we run various prompts to query the LLM in identical fashion for both the variable and narrative models.

1. **Latent information prompt**:

4 Results

5 Implications

6 Conclusion