JOURNAL OF
EMERGING INVESTIGATORS

Article

Investigating the connection between free word association and demographics

Vincent Chia[1], Nicholas Chia[1]
[1] Carmel High School, Carmel, Indiana

SUMMARY

Free word association (FWA) has been used to analyze cultures, thoughts, beliefs, and demographics across various fields. FWAs are a widely used scientific tool to quickly view a subject's beliefs, biases, and opinions that are often repressed or difficult to detect in short interviews or surveys. Here, we explored the relationship between FWA and demographics through neural network analysis. We hypothesized that neural network analysis of FWA could accurately predict a participant's age, gender, first language, and current country based on their FWA responses to a random cue word from a set of 12,292 cues selected from prior FWA studies. Using the "Small World of Words" dataset containing over 1.2 million FWAs, we created a prediction model and evaluated for accuracy across the four demographic variables. The study employed an existing linguistic neural network, Large Language Model Meta AI 2 (LLaMA 2), which was fine-tuned to predict demographics from FWAs. The trained model demonstrated noteworthy accuracy predicting first language (63.6%), current country (58.4%), and age (median distance of nine years from predicted to actual age), but demonstrated a fluctuating accuracy across generation parameters when predicting gender. Our findings suggest a correlation between FWAs and demographics, aligning with previous research on FWA reflecting geographical differences, cultural beliefs, and age-related patterns. The study demonstrates the potential of using FWA and neural networks to identify demographic information more efficiently than other large scale data collection methods such as surveys.

INTRODUCTION

Free word association (FWA), a technique in which participants respond to a cue word with the first word or phrase that comes to mind, has been widely employed to analyze cultures, thoughts, beliefs, and demographics across diverse fields such as psychology, communication, cultural exploration, and marketing (1). Numerous studies have utilized FWA to delve into the intricacies of memory, cognition, beliefs, biases, and culture, revealing that underlying beliefs are not only reflected in but also extractable from word associations, and that polarizing word associations could accurately and efficiently determine a participant's stance on a social issue (1, 2). Moreover, neural network analysis has been successfully used to extract the connection between FWA and a participant's belief on a large social issue (2). The validity of FWA as a method of cultural analysis has been established through past studies that compared trends in FWA to findings from cultural studies. One such study conducted experiments in Korea, Japan, Thailand, and France; researchers compared the FWA to "rice" and "good rice" and utilized FWA's cultural reflectiveness to analyze and highlight different values each respective country placed on rice (1, 3). Together, these studies underscore the utility of FWA research in the context of exploring cultures and groups of different demographics (2, 3).

Past studies employing increasingly larger participant sets and numerous cue words have necessitated the use of computational methods, such as neural networks, to facilitate increasingly complex data analysis and modeling (4). A study modeling how FWA are produced found FWAs are not absolute but shaped by both life experiences and recent experiences (4). FWA and neural networks have been shown to reflect information that would otherwise be difficult to observe, such as mental organization, emotional and energy level patterns, and determining subsequent FWA from an initial string, enabling models to map abstract associations and thought patterns that would otherwise be challenging to map (5). Together with a recent study that demonstrated the success of FWA based models in predicting the beliefs of participants, FWA analysis through computation models or neural networks is shown to be capable of reflecting a wide range of information including beliefs, thought patterns, and subsequent word associations (2,4,5).

Prior studies have employed FWA to investigate how beliefs and consumer behavior vary by geographical location. An example of this can be seen in the aforementioned "good rice" study, where researchers compared the FWA in four countries in response to the cues "rice" and "good rice" (3). Findings from that study revealed that each country had a different emphasis on what values constitute good rice, demonstrating the uniqueness of FWA by region and the capability of FWA to provide insight into beliefs by geographical location (3). Additionally, FWAs collected from six European regions with the prompt "traditional, in a food context" revealed defining characteristics for each region

Model	top-P Value	Temperature
Model 1	0.66	1.75
Model 2	0.66	1.00
Model 3	0.66	1.50
Model 4	0.80	1.00
Model 5	0.50	1.00

Table 1: Model Parameters. Generation parameter values for the five prediction models. The table shows the top *p*-value and temperature settings used for models 1-5.

https://doi.org/10.59720/24-136

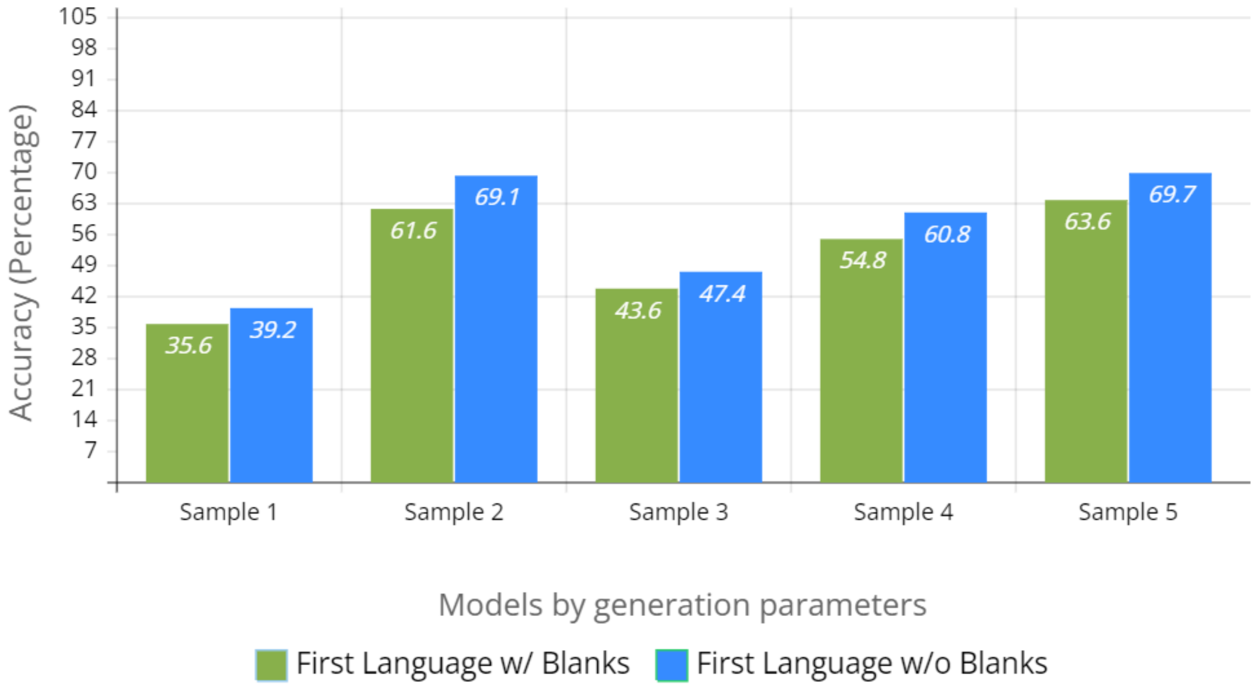

Figure 1: First Language Accuracy Graph. Percent accuracy of each sample model in predicting first language accuracy from FWA data. Analysis was conducted either including (green) or excluding (blue) blank responses.

(6). In particular, the study found southern regions exhibited a tendency towards words pertaining to heritage, culture or history, while the northern regions gravitated towards words pertaining to convenience, health or appropriateness (6). Together, the two studies demonstrated that geographical location is reflected in FWA (3, 6).

A study that mapped FWAs on responses from individuals aged 10-84 using a neural network analysis found that the pattern of word associations across differing ages showed distinct differences and maintained independence throughout the lifespan, demonstrating FDA's ability to reflect age (7). Involving 8,000 participants and over 400 cue words per age group, the study found that the lifetime variability of word associations is primarily determined by associative changes in the least well-connected words (7). This indicates that the identification of age through FWA is likely feasible (7).

Gender differences in word meaning have also been observed in recent research. A recent study investigated gender differences in word associations and meanings using constrained word association tasks and association strength judgments. They found evidence for gender-specific meanings in a substantial fraction of the 42 words studied, 29% in a word association task and 31% in an association strength judgment task (8). Notably, these differences were also found in words without obvious gender connotations and included seemingly neutral words as well. The study also found evidence for gender-specific concepts in 46% of the words that mapped onto multiple concepts. These findings suggest that individuals of different genders who speak the same language have slightly varied representations of words (8). This research indicates that gender could be a significant factor in predicting variations in word meanings and associations.

Despite the considerable literature on the use of FWA to extract opinions, beliefs, memory, cognition, and culture, we are not aware of any research on the use of FWA to predict the demographics of the FWA participant. We aimed to test the feasibility of predicting a participant's age, gender, first language, and current country through neural network analysis of their FWAs, serving as a culmination of prior research on the variation in FWA patterns by demographics. We hypothesized that neural network analysis of FWA could accurately predict these demographic variables based on responses to random cue words. The trained model demonstrated accuracy in predicting first language, current country, and age, but showed poor results for gender prediction. Our findings show a correlation between FWAs and demographics, consistent with previous research on FWA concerning geographical location, cultural beliefs, and age-related patterns. We demonstrate how FWA and neural networks can be utilized to gain demographic insights and enable more efficient collection of high-quality large-scale data compared to traditional methods like surveys or interviews.

RESULTS

The objective of this study was to create an experiment to predict participants' demographics based on their FWA responses. We analyzed the "Small World of Words" English 2018 dataset, containing over 1.2 million Free Word Associations (FWAs), using an existing linguistic neural network, Language Model Meta AI 2 (LLaMA 2) (9). In order to further process the dataset for the current studies purposes, the data was split into training and testing groups following an 80/20 split.

We evaluated the model's accuracy in predicting four

https://doi.org/10.59720/24-136

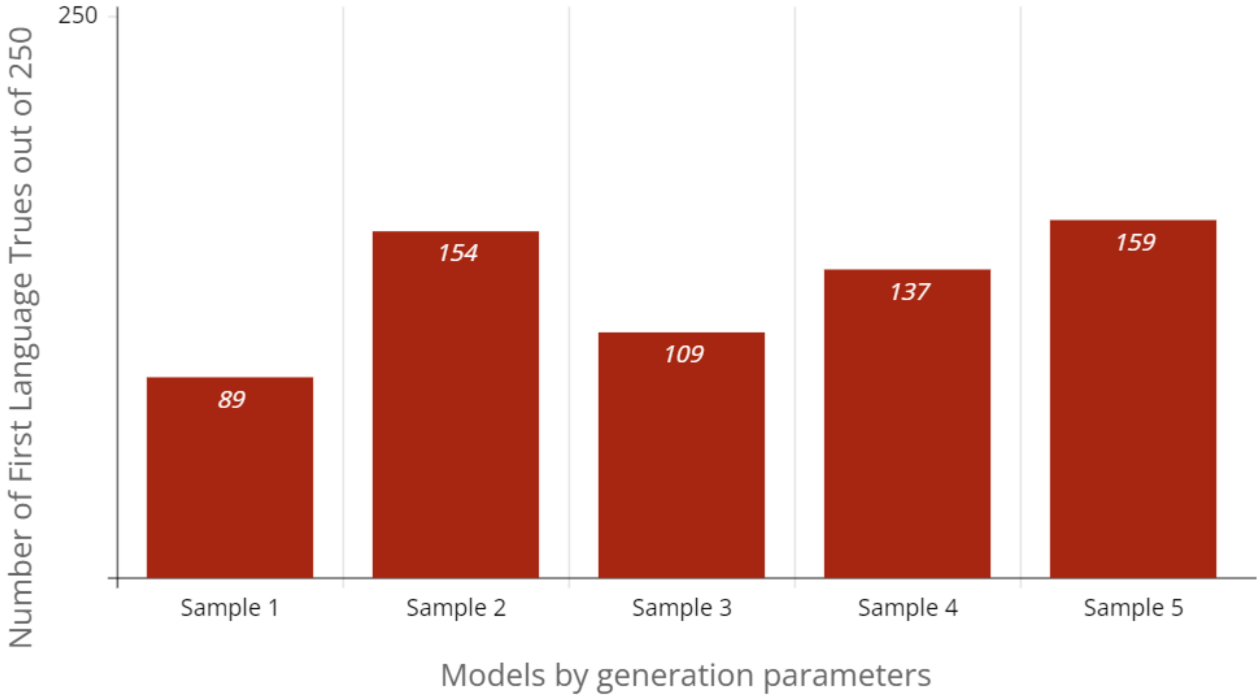

Figure 2: First Language Accuracy Graph. Total number of trues in each sample model for first language prediction. The maximum number of trues for each sample model was 250.

demographic variables: first language, current country, gender, and age. To maximize accuracy and test the limits of the model, we modified generation parameters of top *p*-value and temperature to create five prediction models. The five models represented the untrained model's highest accuracy parameters, with models 1-4 representing slight modifications to a parameter(s) (**Table 1**).

For first language prediction, model 5 achieved the highest accuracy at 63.6% (**Figures 1, 2**). The other models showed varying levels of accuracy, with model 2 performing second-best in this category at 61.6%.

In predicting the current country, model 5 again demonstrated the highest accuracy at 58.4% (**Figures 3, 4**). The performance of the other models followed a similar pattern to that observed in first language prediction, with model 2 showing the second-highest accuracy at 55.2%.

Age prediction accuracy remained relatively consistent across all models, with models 2 and 5 having the highest average accuracy- median distance of 9 and 12 respectively (**Figure 5**). Age accuracy was measured by distance to the correct age, which differed from the other three demographics.

Gender prediction accuracy showed unexpected variations compared to the other demographics. Model 5, the most accurate in other demographics, became the second least accurate with an accuracy of 58.3%, while model 1, previously the least accurate or second least accurate in other demographics, became the most accurate at 64.8% (**Figures 6, 7**).

Across all four demographic predictions, model 2 demonstrated consistent performance, achieving close to the highest accuracy in each category (**Figures 1, 3, 5, 6**). The total number of correct gender predictions ("trues") for each model showed model 1 having the highest count (**Figure 7**). Similar "true" counts for first language and current country predictions revealed that model 5 performed best in these categories (**Figures 2, 4**).

DISCUSSION

In this study, we hypothesized that neural network analysis of FWA and demographics could be used to predict participants' demographics based solely on FWA. The findings from this study suggest that neural network analysis is capable of predicting demographics from FWAs. We utilized the "Small World of Words" dataset containing over 1.2 million FWAs and employed an existing linguistic neural network, Large Language Model Meta AI 2 (LLaMA 2), which was fine-tuned to predict demographics from FWAs. The trained model demonstrated noteworthy accuracy predicting first language (63.6%), current country (58.4%), and age (median distance of nine years from predicted to actual age), but demonstrated a fluctuating accuracy across generation parameters when predicting gender. Our findings suggest a correlation between FWAs and demographics, aligning with previous research on FWA reflecting geographical differences, cultural beliefs, and age-related patterns.

The high accuracy in predicting first language and current country aligns with previous findings that FWA reflects geographical differences and cultural beliefs (3, 6). Age prediction accuracy (median distance of nine) showed relatively consistent trends across all generation parameters. Additionally, the median distance of nine years suggests that the model began to learn and predict age from FWA. These results are consistent with prior findings by researchers analyzing FWA patterns by age group, providing additional evidence of an FWA-age correlation (7).

The inconsistent gender prediction accuracy raised new

JOURNAL OF
EMERGING INVESTIGATORS
https://doi.org/10.59720/24-136

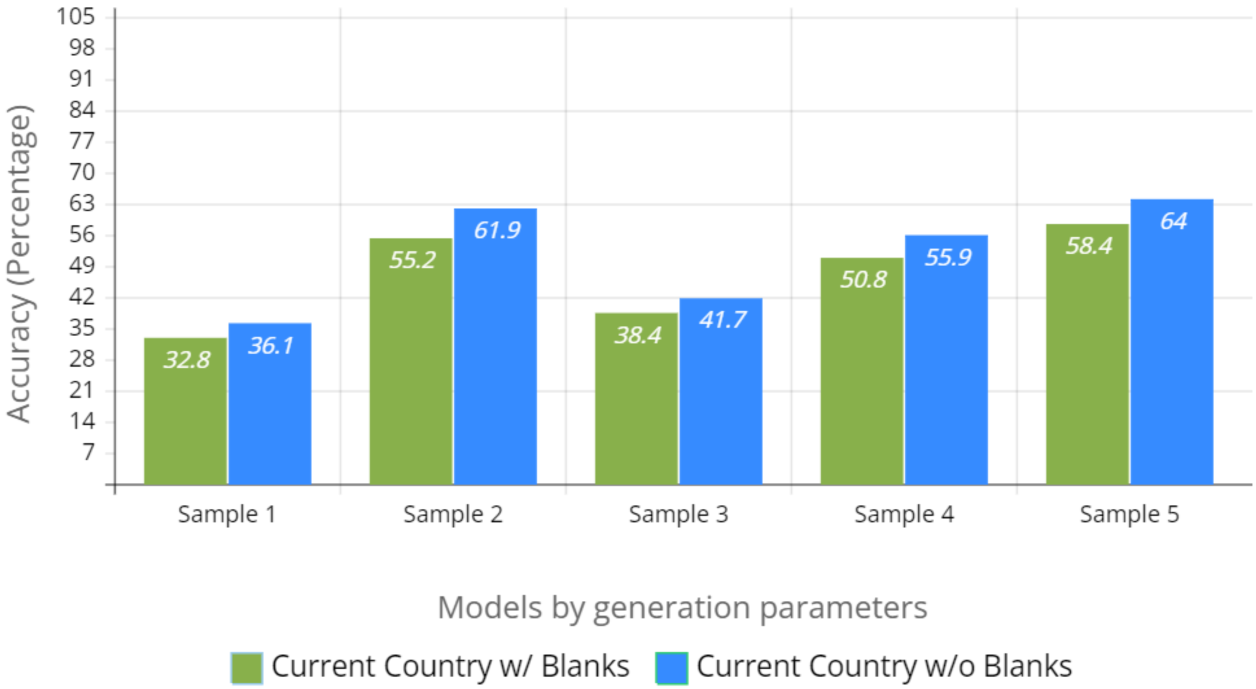

Figure 3: Current Country Accuracy Graph. Percent accuracy of each sample model in predicting current country accuracy from FWA data. Analysis was conducted either including (green) or excluding (blue) blank responses.

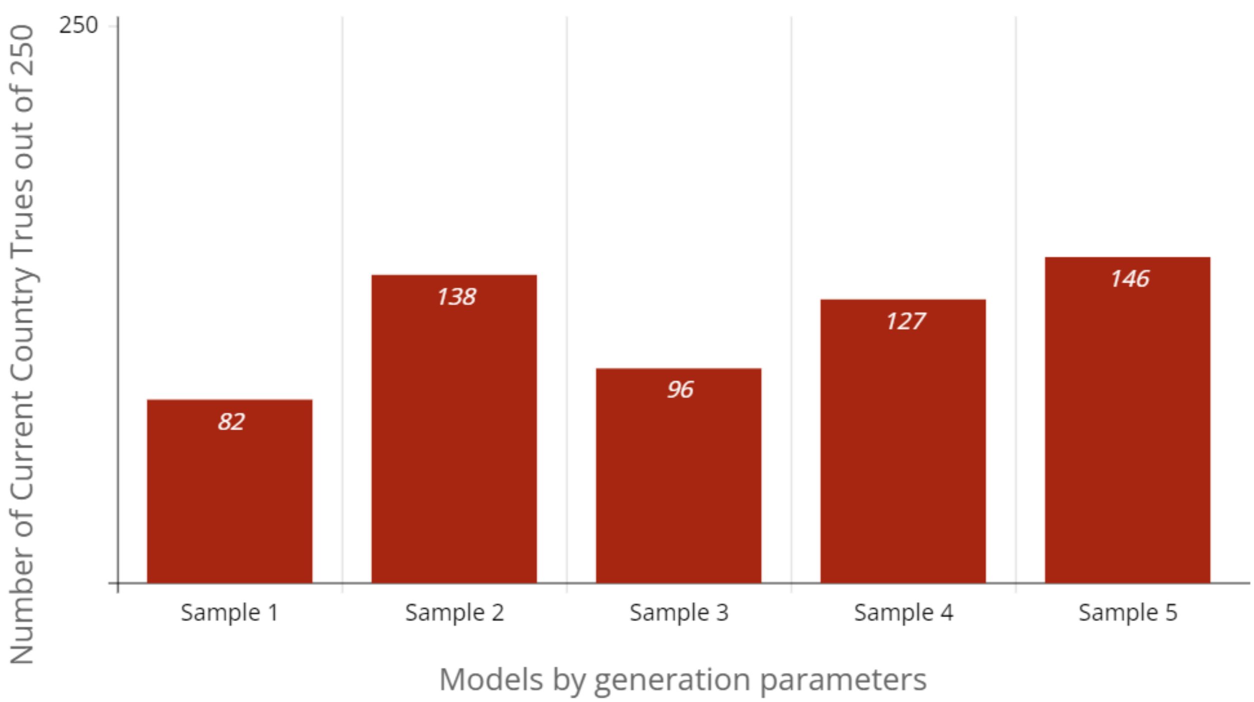

Figure 4: Current Country Trues Graph. Total number of trues in each sample model for current country prediction. The maximum number of trues for each sample model was 250.

https://doi.org/10.59720/24-136

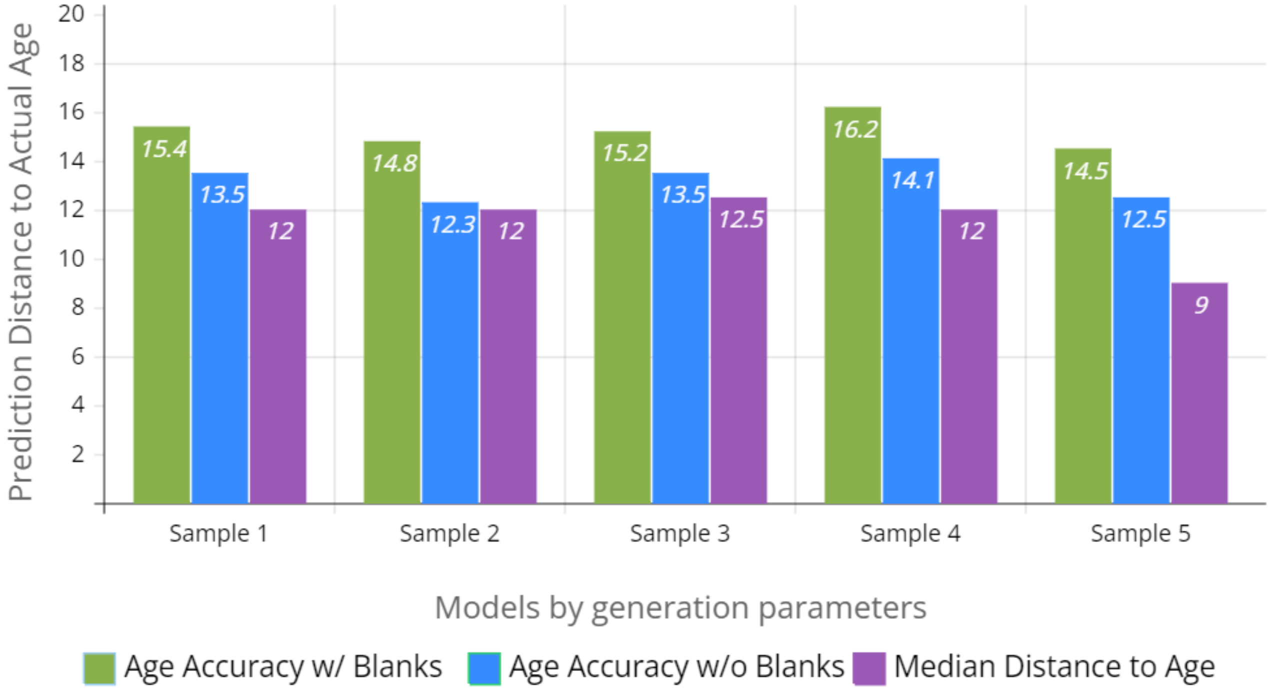

Figure 5: Age Accuracy Graph. Distance from actual age to predicted age from each sample model in predicting first language accuracy from FWA data. Analysis was conducted either with including blank responses (green), excluding blanks responses (blue), or the median distance to age excluding blanks (purple).

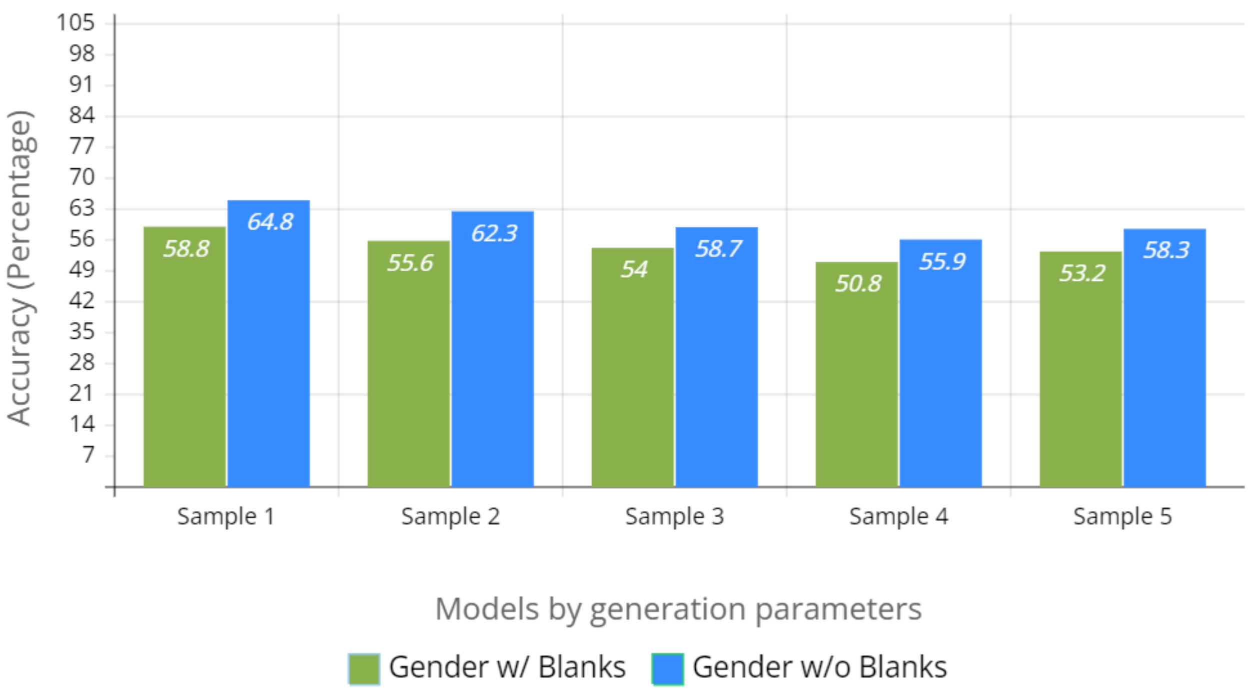

Figure 6: Gender Accuracy Graph. Percent accuracy of each sample model in predicting gender from FWA data. Analysis was conducted either including (green) or excluding (blue) blank responses.

https://doi.org/10.59720/24-136

questions. Optimizing for other demographics seemed to inversely affect gender accuracy, as seen in models 1 and 5. Gender also appears to be impacted most by differing temperatures and varying top *p*-values. Temperature controls the randomness of a model, a temperature of zero will have the exact same results given the same input, while top *p*-values control the amount of answers a model selects its output from, with higher value leads to more creative answers. Low temperatures in general corresponded to lower gender prediction accuracy while higher top *p*-values correspond to higher accuracy. A likely explanation for this is that a larger context window (top-P) and higher temperature allow for the model to avoid stereotyping and overgeneralization. However, model 2's stable accuracy in all demographics warrants further investigation.

The main limitation we faced was a lack of computing resources, which restricted the amount of model training, optimization, and depth of analysis. Hence, the model should be considered a proof-of-concept rather than a fully optimized predictor. Using the "Small World of Words" dataset enabled a large sample size; however, the dataset itself was not perfect and may have introduced biases. For example, the dataset contained an overrepresentation of participants with a bachelor's degree or higher and the focus on quantity meant a decreased control over environmental biases, affecting the FWAs' reflection of real-world scenarios (9). Additionally, the dataset changed any British spelling to their American counterparts, which could have diminished accuracy in first language and current country.

If increased computing resources were available, it would be beneficial to more rigorously test different training methods on a large scale and to explore further prompt engineering to optimize results. However, ethical concerns around privacy must be considered, as malicious parties could potentially misuse high-accuracy prediction models to glean large amounts of information (1). Given the connection found between FWA and demographics, it is possible that a more highly funded group could fine-tune a model for the express purpose of extracting key demographic data from publicly available messages or ads.

Future research should focus on increasing accuracy through further fine-tuning of the model, exploration into the optimization of prompts and generation parameters, and exploration into the model's reasoning behind its predictions, specifically the weights assigned to connections. These would aid in creating a more accurate model, and through the analysis of weights of an accurate prediction model, capable of highlighting the relationship between demographics and word associations. Understanding these relationships could help identify which word associations are most predictive of demographic factors, potentially improving the efficiency of FWA as a demographic analysis tool. The impact of utilizing an optimized model based on FWA to extract information allows for a large-scale collection of opinions that represent underlying beliefs that might otherwise not be expressed in different data collection methods; ease of access to such information could aid a range of fields that aim to impact or study large populations, including but not limited to public policy, marketing, or sociology (1,10).

MATERIALS AND METHODS

The objective of this study was to predict participant demographics using their FWA responses by using a pre-existing large language model (LLaMA2) and the "Small World of Words" dataset (9,12). The "Small World of Words" dataset (SNOW-EN 18), compiled through online crowdsourcing, contains approximately 1.2 million cues given from a list of 12,000 cue words collected from previous FWA

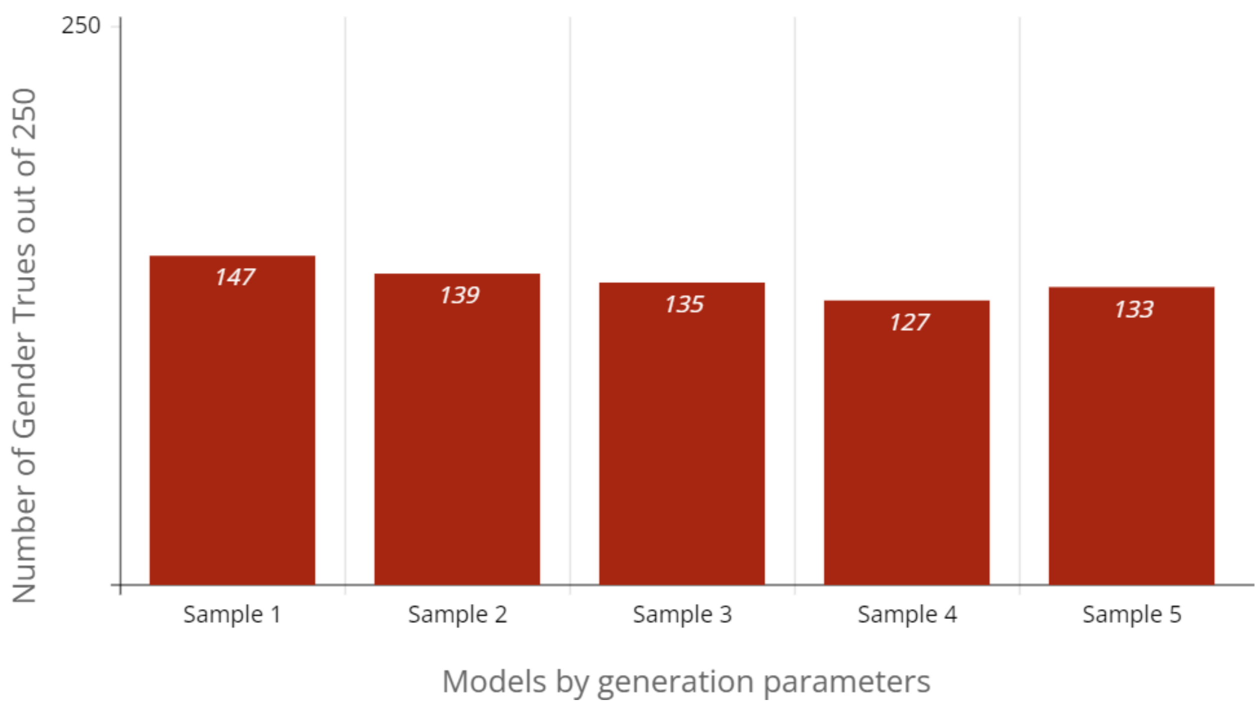

Figure 7: Gender Trues Graph. Total number of trues in each sample model for gender prediction. The maximum number of trues for each sample model was 250.

studies with three response spaces for FWA per cue. The only prerequisite for study participation was English fluency (9). The dataset included participant ID, cue word, FWA response spaces (a maximum of three responses per cue), first language, current country, age, gender, and timestamp. Participants were anonymous in the published dataset. This study did not attempt to re-identify participants and received ethical approval from the Carmel Clay Schools Ethics Review Committee prior to data retrieval. In addition, data preprocessing removed participant identification numbers, further increasing anonymity of the participants (9).

LLaMA 2 was selected as a base model due to its availability for research use as a state-of-the-art large language model. This study utilized the 7B parameter version, comprising multiple stacked transformer blocks. Each block incorporated a multi-head self-attention mechanism and feed-forward neural networks (12).

The dataset was preprocessed, split into 80% training and 20% testing sets, and formatted for LLaMA 2 compatibility (9). The data preprocessing of this dataset was completed by the researchers who published the SNOW-EN 18 Dataset; their process involved correcting responses with capitalization, double spaces, and removing “a” and “the”. In addition, the researchers compiling the SNOW-EN 18 Dataset removed certain responses to cues that utilized a non-American spelling, and similarly replaced non-American responses with American counterparts (9).

LLaMA 2 was fine-tuned on the training data over five epochs to predict participant demographics from FWAs. The training process involved identifying the generation settings that produced responses in the correct format (age, current country, first language, and gender) and contained answers within feasibility of the training set. An example of an incorrectly-formatted output outside of feasibility is: “Age: 124, Current Country: cats are known to be house pets throughout North America, First Language: N/A, Gender: N/A”. Since this response contains no feasible information pertaining to participant demographic, it was inputted as blank for our analysis.

Following the training process, the fine-tuned LLaMA 2 model was prompted with the cue word and subsequent word associations from the testing portion of the dataset. The code used to preprocess the dataset, the code used to process the output of the fine-tuned model, and the data used to identify training parameters has been included in the appendices. Inputted word associations remained the same for all settings and consisted of 250-word sets selected from every 40th data set across 10,000 data sets. Prediction accuracy was quantitatively analyzed. Age accuracy was measured by the difference between predicted and actual age. First language, current country, and gender accuracies were measured by percentage of correct responses.

The analysis was carried out with Python (3.12.4) and utilized pandas, pyperclip, pyautogui packages. This study includes several appendices that provide detailed information on various aspects of the methodology: **Appendix A** details the process of converting the SNOW EN 2018 dataset into the training set format; **Appendix B** describes the method for selecting and formatting the test set; **Appendix C** outlines the automated input and output collection process for the web service running the study’s model; **Appendix D** explains the pre-processing steps for formatting the output data; **Appendix E** details the process of isolating predictions and handling incorrect formats in the dataset; **Appendix F** describes the method for pre-selecting correct demographics from the test set; **Appendix G** outlines the process of compiling correct demographics into an Excel file for comparison; **Appendix H** details what and how training parameters were selected.

Received: May 11, 2024
Accepted: August 29, 2024
Published: February 11, 2025

REFERENCES

1. Stacy, Alan W., *et al.* “Word association tests of associative memory and implicit processes: Theoretical and assessment issues.” *Handbook of Implicit Cognition and Addiction*, 2006, pp. 75-90. https://doi.org/10.4135/9781412976237.n6
2. File, Balint, *et al.* “Emergence of polarized opinions from free association networks.” *Behavior Research Methods*, vol. 51, no. 1, Aug. 2018, pp. 280-294. https://doi.org/10.3758/s13428-018-1090-z
3. Son, J., *et al.* “Understanding the effect of culture on food representations using word associations: The case of ‘rice’ and ‘good rice’.” *Food Quality and Preference*, vol. 31, 2014, pp. 38-48. https://doi.org/10.36923/jicc.v5i1.401
4. Nelson, Douglas L., *et al.* “What is free association and what does it measure?” *Memory & Cognition*, vol. 28, no. 6, 2000, pp. 887-899. https://doi.org/10.3758/BF03209337
5. Dover, Yaniv, and Zachary Moore. “Using free association networks to extract characteristic patterns of affect dynamics.” *Proceedings of the Royal Society A*, vol. 476, no. 2236, 2020. https://doi.org/10.1098/rspa.2019.0647
6. Guerrero, Luis, *et al.* “Perception of traditional food products in six European regions using free word association.” *Food Quality and Preference*, vol. 21, no. 2, 2010, pp. 225-233. https://doi.org/10.1016/j.foodqual.2009.06.003
7. Dubossarsky, Haim, *et al.* “Quantifying the structure of free association networks across the life span.” *Developmental Psychology*, vol. 53, no. 8, 2017, pp. 1560-1570. https://doi.org/10.1037/dev0000347
8. De Deyne, S., *et al.* “Common words, uncommon meanings: Evidence for widespread gender differences in word meaning.” *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 45, 2023. escholarship.org/uc/item/191089bx
9. De Deyne, Simon, *et al.* “The ‘Small world of words’ English word association norms for over 12,000 cue words.” *Behavior Research Methods*, vol. 51, no. 3, 2018, pp. 987-1006. https://doi.org/10.3758/s13428-018-1115-7
10. Valba, Oleg, *et al.* “Analysis of English free association network reveals mechanisms of efficient solution of remote association tests.” *PLOS ONE*, vol. 16, no. 4, 2021. https://doi.org/10.1371/journal.pone.0248986
11. Viana, Filipy M., *et al.* “Applying free word association to understand the perception of fish as a meal by Brazilians with different consumption frequencies.” *Journal of Sensory Studies*, vol. 36, no. 2, 2020. https://doi.org/10.1111/joss.12628
12. Touvron, Hugo, *et al.* “Llama 2: Open Foundation and

https://doi.org/10.59720/24-136

Fine-Tuned Chat Models." *Preprint at arXiv,* https://doi.org/10.48550/arXiv.2302.13971

JOURNAL OF
EMERGING INVESTIGATORS

Appendix A: SNOW EN 2018 to Training Set

Removes all other information from the training set except for: Cue word, FWA 1-3, and demographics.

```
import pandas as pd
from io import StringIO
import json

#Sample Data
data_string = """
"1",29,3,33,"Fe","United States","Australia",NA,2011-08-12 02:19:38,"although","nevertheless","yet","but"
"2",30,3,33,"Fe","United States","Australia",NA,2011-08-12 02:19:38,"deal","no","cards","shake"
"3",31,3,33,"Fe","United States","Australia",NA,2011-08-12 02:19:38,"music","notes","band","rhythm"
"4",32,3,33,"Fe","United States","Australia",NA,2011-08-12 02:19:38,"inform","tell","rat on",NA
"""
df = pd.read_csv(StringIO(data_string), dtype=str)


def format_data(row):
    cue_word = row[9]
    word_1 = row[10]
    word_2 = row[11] if pd.notna(row[11]) else ""
    word_3 = row[12] if pd.notna(row[12]) else ""
    age = row[3]
    gender = row[4]
    first_language = row[5]
    country = row[6]

    formatted_string = f'Analyze this data to predict the Age, Gender, Country, and Country of Origin given just word associations. ' \
                       f'### Instruction: {cue_word}, {word_1}, {word_2}, {word_3}. ' \
                       f'### Response: Age: {age}, Gender: {gender}, First Language: {first_language}, Country: {country}.'

    return {"text": formatted_string}

formatted_data = df.apply(format_data, axis=1).tolist()

output_file_path = 'Filepath/FolderName/FileName'
with open(output_file_path, 'w', encoding='utf-8') as output_file:
    for entry in formatted_data:
        json_line = json.dumps(entry, ensure_ascii=False)
        output_file.write(json_line + '\n')

print(f'Data written to {output_file_path}')
```

Appendix B: Test Set Selection and Formatting

Selects from every 20th line of the test set, in order to more efficiently test a larger selection of participants. (A single participants responses were filed sequentially in the dataset)

```
import pandas as pd
from io import StringIO

#Sample Data
data_string = """
"1000000",1174724,102810,40,"Ma","Australia","Australia",3,2015-03-10 20:15:12,"carry
out","construct","embark","employ"
"1000001",1174725,102810,40,"Ma","Australia","Australia",3,2015-03-10 20:15:12,"beater","mixer","abuser",NA
"1000002",1174726,102810,40,"Ma","Australia","Australia",3,2015-03-10
20:15:12,"newsletter","pamphlet","broadsheet","gazette"
"1000003",1174727,102810,40,"Ma","Australia","Australia",3,2015-03-10 20:15:12,"yak","animal","Herder","bovine"
"1000004",1174728,102810,40,"Ma","Australia","Australia",3,2015-03-10
20:15:12,"deceptive","fraudulent","lying","misrepresentation"
"""

df = pd.read_csv(StringIO(data_string), dtype=str)

def format_data(row):
   cue_word = row[9]
   word_1 = row[10]
   word_2 = row[11] if pd.notna(row[11]) else ""
   word_3 = row[12] if pd.notna(row[12]) else ""

   formatted_string = f' "{cue_word}, {word_1}, {word_2}, {word_3}.",'

   print(formatted_string)

df_selected = df.iloc[::20]
df_selected.apply(format_data, axis=1)
```

Appendix C: Automated Input and Output Collection

Code to input the FWA and collect the output one by one into the web service running this studies model. (together.ai)

```
import pyperclip
import pyautogui
import time

def record_textbox_content():

    pyautogui.hotkey('ctrl', 'a')
    pyautogui.hotkey('ctrl', 'c')
    time.sleep(.25)
    clipboard_content = pyperclip.paste()
    return clipboard_content

def clear_and_input_new_data(new_data):
    pyautogui.hotkey('ctrl', 'a')
    pyautogui.press('delete')
    time.sleep(.25)
    pyautogui.write(new_data)

# Sample data
try:
    data_list = [
        "beater, mixer, abuser, .",
        "flashing, lights, police, strobe.",
        "probation, jail, criminal, .",
        "squawk, bird, loud, annoying.",
    ]
    submit_button_x = 3160
    submit_button_y = 2000

    for data in data_list:
        time.sleep(.5)
        clear_and_input_new_data(data)


        pyautogui.click(submit_button_x, submit_button_y)
        time.sleep(2.5)

        textbox_content = record_textbox_content()
        print(textbox_content)
```

```
finally:
    time.sleep(6)
    pyautogui.hotkey('ctrl', 'w')
```

Appendix D: Pre-processing Output Data Formatting

Cleans the output data into a easier to process format

```
#Sample Text
raw_text = """
beater, mixer, abuser, .<s> февраль.February. February is the month of love. February is also the shortest month of the year
flashing, lights, police, strobe. ### Response: Age: 32, Gender: Fe, First Language: United States, Country: United States.<s>
probation, jail, criminal, .<s> живелон.I'm sorry, I don't speak Spanish.(of a person, usually male,
squawk, bird, loud, annoying. ### Response: Age: 26, Gender: Fe, First Language: United States, Country: United States.<s>
"""
raw_text_lines = raw_text.strip().split('\n')


quoted_text_lines = ['"' + line.strip() + '",' for line in raw_text_lines]

quoted_text = '\n'.join(quoted_text_lines)

print (quoted_text)
```

Appendix E: Output Data Formatting

Isolates predictions and creates blanks in the dataset if the format is incorrect. Removes any obviously wrong predictions.

```
import re
import pandas as pd

#Sample Data
responses = [
"beater, mixer, abuser, .<s> февраль.February. February is the month of love. February is also the shortest month of the year",
"flashing, lights, police, strobe. ### Response: Age: 32, Gender: Fe, First Language: United States, Country: United States.<s> Anal",
"probation, jail, criminal, .<s> живелон.I'm sorry, I don't speak Spanish.(of a person, usually male,",
"squawk, bird, loud, annoying. ### Response: Age: 26, Gender: Fe, First Language: United States, Country: United States.<s> Anal",
]

pattern = r"Age: (\d+), Gender: (\w+), First Language: ([\w\s]+), Country: ([\w\s]+)"

ages = []
genders = []
first_languages = []
countries = []

for response in responses:
  match = re.search(pattern, response)
  if match:
    age = match.group(1)
    gender = match.group(2)
    first_language = match.group(3)
    country = match.group(4)

    ages.append(age)
    genders.append(gender)
    first_languages.append(first_language)
    countries.append(country)
  else:
    ages.append("")
    genders.append("")
    first_languages.append("")
    countries.append("")

data = {
```

```
    "Age": ages,
    "Gender": genders,
    "First Language": first_languages,
    "Country of Origin": countries
}

df = pd.DataFrame(data)

excel_file = "outputfile.xlsx"
df.to_excel(excel_file, index=False)

print("Data saved to:", excel_file)
```

Appendix F: Correct Demographic Pre-Selection from Test Set

Selected the same lines as the code in Appendix A: captured only the correct answers.

```
import pandas as pd
from io import StringIO

#Sample Data
data_string = """
"1000000",1174724,102810,40,"Ma","Australia","Australia",3,2015-03-10 20:15:12,"carry
out","construct","embark","employ"
"1000001",1174725,102810,40,"Ma","Australia","Australia",3,2015-03-10 20:15:12,"beater","mixer","abuser",NA
"1000002",1174726,102810,40,"Ma","Australia","Australia",3,2015-03-10
20:15:12,"newsletter","pamphlet","broadsheet","gazette"
"1000003",1174727,102810,40,"Ma","Australia","Australia",3,2015-03-10 20:15:12,"yak","animal","Herder","bovine"
"""

df = pd.read_csv(StringIO(data_string), dtype=str)


def format_data(row):
    age = row[3]
    gender = row[4]
    first_language = row[5]
    country = row[6]

    formatted_string = f' "{age}, {gender}, {first_language}, {country}",'

    print(formatted_string)


df_selected = df.iloc[::20]
df_selected.apply(format_data, axis=1)
```

Appendix G: Correct Demographic Compilation

Formatted the correct answers into an excel file for easier comparison

```
import pandas as pd

# Sample data
responses = [
    "40, Ma, Australia, Australia",
 "37, Ma, United States, nan",
 "17, Ma, United States, United States",
 "30, Ma, New Zealand, New Zealand",
 "34, Fe, Canada, nan"
]


response_data = [response.split(", ") for response in responses]


df = pd.DataFrame(response_data, columns=["Age", "Gender", "First Language", "Country of Origin"])

excel_file = "outputfile.xlsx"
df.to_excel(excel_file, index=False)

print("Data saved to:", excel_file)
```

Appendix H: Training Parameter Selection

Parameter selection and justification

The training of the LLaMA 2 utilized a batch size was set to 32, learning rate was set to 1E-4, epochs were set to 5 due to computational constraints. Fine-tuning settings were determined through small scale testing and were optimized for lower training loss and validation loss seen below. Small scale tests utilized 10,000 sets of FWA and demographics with a 70/30 ratio between training and testing.

Test Number	Batch Size	Learning Rate	Training Loss after Epoch 5	Validation Loss after Epoch 5
Test 1	32	1E-4	0.55	0.85
Test 2	16	1E-4	0.50	0.95
Test 3	64	1E-4	0.60	0.90
Test 4	32	2E-4	0.45	1.00
Test 5	32	5E-5	0.60	0.88