AMERICAN UNIVERSITY OF BEIRUT

QUANTIFYING GENDER AND POLITICAL BIAS IN ARABIC WORD EMBEDDINGS

by
GHUMDAN HATIM AL-SABAHI

A project
submitted in partial fulfillment of the requirements
for the degree of Master of Science
to the Department of Computer Science
of the Faculty of Arts And Sciences
at the American University of Beirut

Beirut, Lebanon
September 2022

AMERICAN UNIVERSITY OF BEIRUT

QUANTIFYING GENDER AND POLITICAL BIAS IN
ARABIC WORD EMBEDDINGS

by
GHUMDAN HATIM AL-SABAHI

Approved by:

	Signature
Fatima K. Abu Salem, Associate Professor Dr. Full Name, rank Department of Computer Science	First Reader
	Signature
Shady Elbassuoni, Associate Professor Dr. Full Name, rank Department of Computer Science	Second Reader

Date of project presentation: September 5th, 2022

AMERICAN UNIVERSITY OF BEIRUT

PROJECT RELEASE FORM

Student Name:	Al-Sabahi	Ghumdan	Hatim
	Last	First	Middle

I authorize the American University of Beirut, to: (a) reproduce hard or electronic copies of my project; (b) include such copies in the archives and digital repositories of the University; and (c) make freely available such copies to third parties for research or educational purposes:

- ☐ As of the date of submission
- ☒ One year from the date of submission of my project.
- ☐ Two years from the date of submission of my project.
- ☐ Three years from the date of submission of my project.

ghumdan	September 15, 2022
Signature	Date

ACKNOWLEDGEMENTS

I would like to thank my professors, Dr. Fatima Abu Salem and Dr. Shady Elbassuoni, for their patient guidance and eager support throughout the last year. I would also want to thank the MEPI TLG for providing me with the opportunity to work on this project on the topic "Quantifying gender and political bias in Arabic word embeddings," which also allowed me to perform a lot of research and learn about new topics, such as word embeddings. Finally, I would like to thank my parents and my wife for their continuous encouragement during my studies.

ABSTRACT
OF THE PROJECT OF

Ghumdan Hatim Al-Sabahi for Master of Science
Major: Computer Science

Title: Quantifying Gender and Political Bias in Arabic Word Embeddings

Word embeddings are a breakthrough in the world of artificial intelligence. They replaced the one hot encoding that is used in many Natural Language Processing (NLP) systems such as sentiment analysis, recommendation systems, and so on. In word embeddings, each word is represented as a vector with related words clustered together. In other words, words that are close in vector space should have similar meanings. Recent research, however, has revealed that these word embeddings contain biases towards specific groups that are transferred from our culture to machines. However, the majority of such research has been conducted for English word embeddings. Other research on languages that incorporate grammatical gender terms have adjusted the bias test to accommodate for gendered words. However, little has been done on the Arabic language. In this study, we focus on quantifying gender and political bias in Twitter, Wikipedia, and two Lebanese newspaper corpora, all of which were trained using the CBOW algorithm. In the Twitter and Wikipedia models, we examine the relation of male and female terms with various categories, including strength, weakness, career, family, domestic work, science, art, money & business, and beauty & appearance. Furthermore, we investigate the relationships between "Palestine" and "Israel" in all of our embeddings with "occupation", "resistance", "peace", and "violence" & "terrorism". We rely on manual translation and evaluation due to a scarcity of Arabic language literature. Our findings reveal that some stereotypes, such as the connection of females with domestic work and art as well as males with strength and money & business, are expressed in our embeddings. In terms of political categories, the Lebanese newspapers examined have long portrayed Israel using terms associated with "occupation and violence" & "terrorism", whereas Palestinians have long been associated with "resistance". Furthermore, we investigate the political bias in greater depth across decades to demonstrate how newspapers' opinions have evolved over time.

TABLE OF CONTENTS

ILLUSTRATIONS

TABLES

ABBREVIATIONS

D_W: Domestic Work
M&B: Money and business
B&A: Beauty and appearance
V&T: Violence and terrorism
CBOW: Continuous Bag of Words

CHAPTER I

INTRODUCTION

A. Motivation

Biases and stereotypes have continued to exist in our societies throughout history. Women, for example, have been related to certain jobs or housekeeping, African Americans have been linked to criminal activities, and much more. These biases are visible on a daily basis in two forms: verbal and written. These data are the primary sources that feed NLP systems. Since AI is derived from mathematics, people assumed it would not contain any bias since math is neutral. People eventually discovered that the output of these AI models was skewed towards certain groups. For example, when translating "he is a nurse" to Arabic using Google Translate, the output is 'هو نارس,' which has no meaning. On the other hand, when we translate the statement "she is a nurse," we obtain 'انها ممرضه ,' which is accurate.[1] The 'Reviewed by contributors' mark appears next to these translations.[2] We can see here that Google Translate has a gender bias toward women using the word "nurse," despite the fact that the term "nurse" can refer to either a male or a female. In other words, the word nurse was not associated with the pronoun "he." This bias was transferred to machines from our cultures.

AI has extended across several industries since its birth. With this expansion, there is widespread fear that the biases and stereotypes we live with and fight against will be passed down to future generations via robots. Furthermore, these AI systems can detect

[1] The example provided was last captured on September 9th, 2022.
[2] Reviewed by contributors: Translation was marked as correct by Google Translate users.

hidden biases and stereotypes that we are currently unaware of but will become more evident in the future.

B. Literature Review

One of the earliest studies about gender bias in word embeddings was done by (Bolukbasi et al., 2016) in their paper called "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings." Their work shed a light on the importance of quantifying gender bias and debiasing word embeddings. There have been several studies to detect biases or stereotypes towards particular groups through word embeddings that followed (Bolukbasi et al., 2016). (Caliskan et al., 2017) have shown that a machine learning model, GloVe, trained on data from the web, has biases towards race and gender. They have also introduced a Word Embedding Association Test (WEAT) analogous to the Implicit Association Test (IAT) used in psychology. (Garg et al., 2018) has studied gender and ethnic biases for the past 100 years in the United States. They found out that the embeddings captured the women's movement. Also, (Wevers, 2019) compared the gender bias towards occupations, psychological states, and social life in six Dutch newspapers over four decades. Further studies (Mendelsohn et al., 2020), showed how LGBTQ groups are always associated with dehumanizing languages.

All the previously mentioned studies were conducted on languages that do not contain grammatical gendered words. However, there have been few studies that tackled gender bias in languages with grammatical gender. (Zhou et al., 2019) provided the MWEAT which is a modified version of the WEAT test that takes grammatical gendered words into an account and applies them to the French and Spanish languages. Moreover, (Lauscher et al., 2020) investigated bias found in Arabic word embeddings using several tests applied on embeddings. They added the male and female versions of the words to the target lists. Alternatively, (Chen et al., 2021) extend the method used in (Bolukbasi et al., 2016) to account for languages with grammatical gender. Their analysis was done in nine languages, where five languages have grammatical gender including Arabic. No other studies about biases through word embeddings were published in the Arabic language to the best of our knowledge. We will be studying gender bias in the context of the Lebanese media, Twitter, and Wikipedia through word embeddings.

CHAPTER II

METHODOLOGY

A. Datasets

We performed our gender bias experiments over the Twitter and Wikipedia word embeddings. As for the political bias, we performed the experiment over Twitter, Wikipedia, and two Lebanese newspapers word embeddings. The Twitter and Wikipedia embeddings were trained by (Soliman et al., 2017)[3] while the Lebanese newspapers embeddings were trained by (Doughman et al., 2020).

1. Twitter

The Twitter corpus consists of 77,600,000 Arabic tweets shared in the period between 2008 and 2016. These tweets were written using many Arabic dialects and sub-dialects. All the data was trained using the n-grams CBOW algorithm.

2. Wikipedia

This dataset is the 2017 Arabic dump of Wikipedia. All articles were divided into paragraphs giving a total of 1,800,000 paragraphs. All this data was also trained using n-grams CBOW algorithm

3. Lebanese Newspapers

We performed the analysis on As-Safir and An-Nahar newspapers. Both newspapers have been leading in Lebanon, but As-Safir closed in December 2016.

---

[3] Twitter and Wikipedia models can be downloaded from https://github.com/bakrianoo/aravec

a. As-Safir:

- 12058 issues
- 15.2 ~ 16 pages per issue
- 1974 – 2011 timeframe
- Range of 37 years

b. Nahar:

- 23,112 issues
- 11.9 ~ 12 pages per issue
- 1933 – 2009
- Range of 76 years

Both newspapers have CBOW word embeddings trained on the entire corpus as well as embeddings trained on a decade level.

B. Experimental Setup

We must determine a collection of key terms that best describe the group in order to quantify any bias toward groups like females or Palestinians. Then, we must create categories, such as "strength" or "peace," against which bias must be measured. These categories also contain a set of words. We next run our bias test on each group against the corresponding category.

1. Forming Groups

In our study, the groups were divided into two gender groups, male and female, and two political groups, Palestinians and Israel. Gender sets are composed of gender traits and pronouns. For example, the male group's set was filled with male attributes such as man and father, as well as gender pronouns such as he and masculine they. The female group's set was filled with the feminine equivalents of the male group's keywords. In terms of the political groups, we populated the sets with commonly used terminologies in the Arab world. The Palestinian group, for example, contained terminologies like Palestine and Muslims, whereas the Israel group contained terms like Zionist and Jewish. The Appendix provides all the keywords in each category.

2. Forming Categories

Due to the lack of word categorization in Arabic literature, we relied on translating English categorized words using Google Translate. Each term was translated and then reviewed before adding it to the set. The English word categorization tool is called Empath, and it was done by (Fast et al., 2016)[4]. The final step was to exclude words that convey more than one meaning. For example, the word weak means 'ضعيف' in Arabic which can also indicate a loss in weight, so it was removed from the weakness set. We analyzed gender bias over nine categories. The following table shows the categories and some sample words.

[4] You can find the repository for the English terms in https://github.com/Ejhfast/empath-client

Category	Sample Keywords
Strength	Strong, capability, effort
Weakness	Weakness, tired, exhausted
Career	Office, business, salary, manager, doctor
Family	Family, relatives, marriage, divorce
Domestic Work (D_W)	Cooking, cleaning, bread
Science	Engineering, medicine, researcher
Art	Creative, artist, exhibition, designer
Money & Business (M&B)	Company, economics, budget, dollar
Beauty & Appearance (B&A)	Beautiful, tall, blond, flirt

Table 1: Gender categories with samples

Regarding political bias, we constructed 4 categories as shown in the table below.

Category	Sample Keywords
Occupation	Raid, missile, explosion
Resistance	Suffering, confrontation, adaptation
Peace	Safe, peaceful, quiet
Violence & Terrorism (V&T)	Massacre, terrorist, crime

Table 2: Political categories with samples

3. Bias Test

We used the bias tests used in (Bolukbasi et al., 2016) and (Chen et al., 2021) for political bias and gender bias respectively.

a. Political Bias

We first need to identify our political direction. In other words, we are seeking a vector that can capture all the political group terms used. To construct this vector, we take the vector of the first word from the Israel list minus the vector of the first word from the Palestinian list.

$$\overrightarrow{JEWS} - \overrightarrow{MUSLIMS}$$

We do this for all the terms in the list, and then we aggregate these vectors and apply principal component analysis, PCA, and take the first component. PCA is a dimensionality reduction technique where it reduces higher dimensions into lower dimensions. The intuition behind applying PCA is to have one direction that accounts for all the variation in our calculated vectors. We will refer to this vector as $\vec{g}$. Next, we apply the below direct bias test for each of our categories.

$$Direct\ bias = \frac{1}{|N|} \sum_{w \in N} \cos\left(\vec{w}, \vec{g}\right)$$

where N is the political category set, and cos(w,g) is the cosine similarity between the word vector w and the political direction g. This equation yields a result between -1 and 1. A positive result indicates a bias toward Israel while a negative result indicates a bias toward Palestinians.

b. Gender Bias

We apply the same method described above to get the gender direction. We take a vector from our female category minus the corresponding vector from the male. We then apply PCA to the aggregated vectors. Next, we apply a modified version of the bias test to account for gendered words.

$$Direct\ bias = \frac{1}{|N|} \sum_{w \in N} \frac{freq_m * \cos(\overrightarrow{w_m}, \vec{g}) + freq_f * \cos\left(\overrightarrow{w_f}, \vec{g}\right)}{freq_m + freq_f}$$

where N is the gender category set, cos(w,g) is the cosine similarity between the word vector w and the gender direction g, w_m is the masculine term, w_f is the feminine term, $freq_m$ and $freq_f$ are the frequency of the masculine and feminine term respectively. The intuition behind taking the weighted average is that we need to consider the word frequency as we believe it has a value in the bias test.

CHAPTER III

RESULTS

A. Gendered Discourse

A positive value indicates a bias towards females while a negative value indicates a bias towards males.

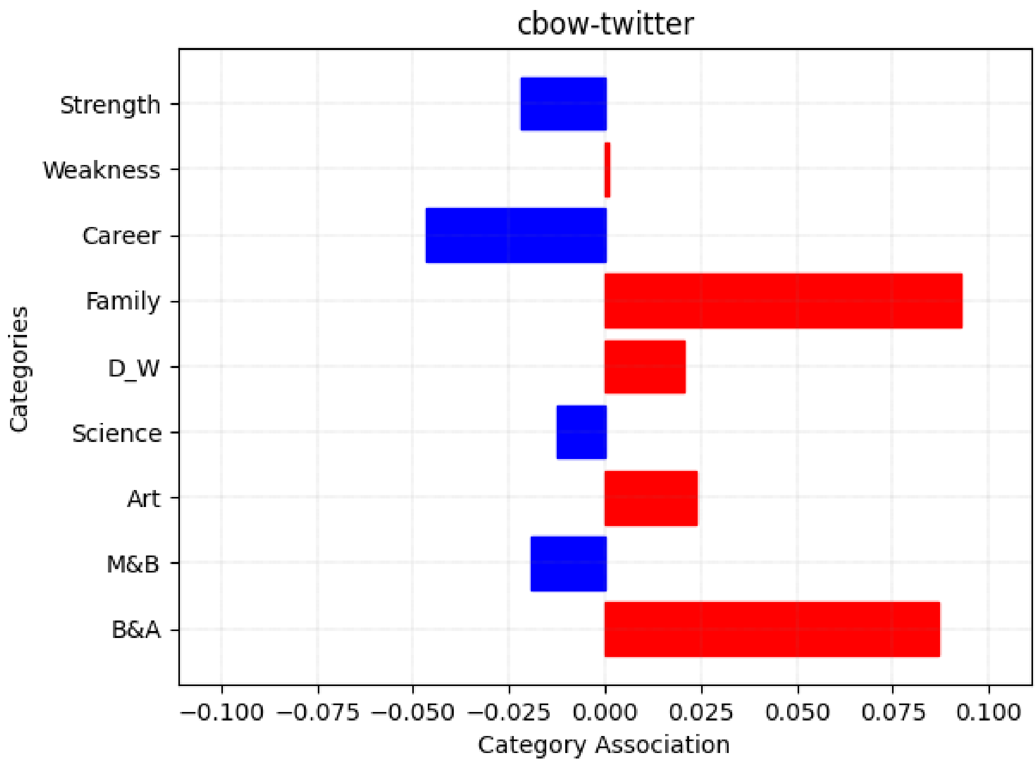

Figure 1: Gender bias on the Twitter word embedding.

Based on our study of the Twitter word embedding, males are mostly linked with careers and only tangentially with strength, science, and M&B. Females, on the other hand, are primarily linked with family, B&A, and only tangentially with D_W and Art. Even though females are rarely connected with weakness, due to its small scale, it might be viewed as a balance.

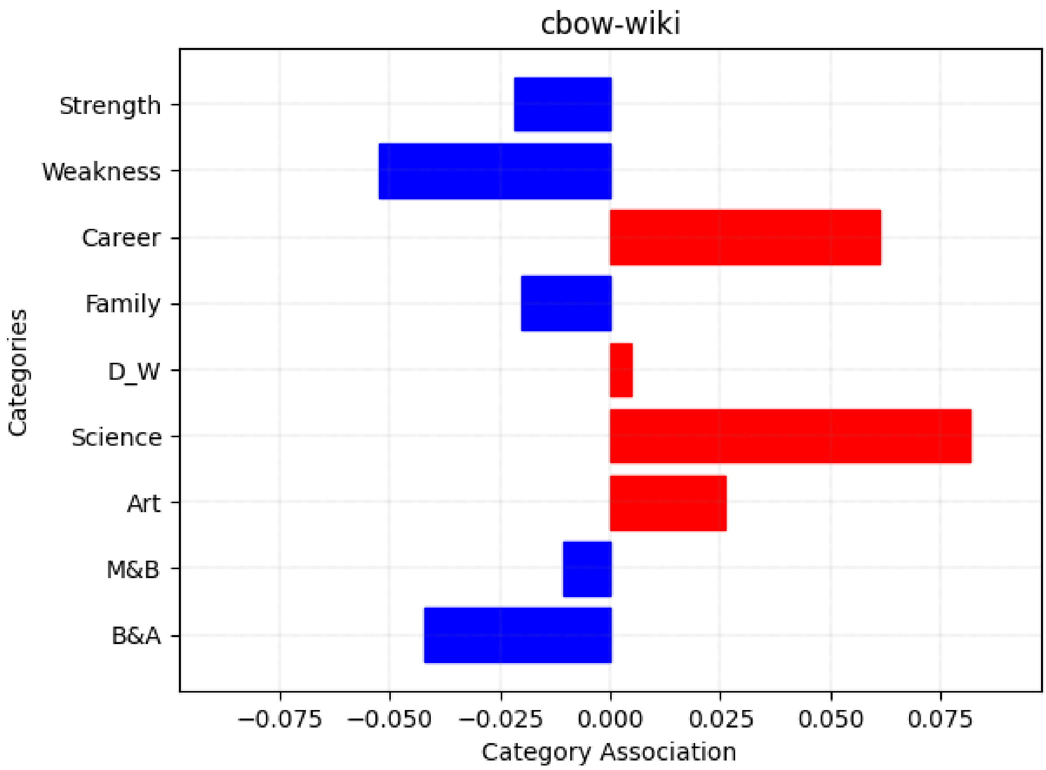

Figure 2: Gender bias on the Wikipedia word embedding

As per our research, males are strongly related with weakness and B&A and only occasionally associated with strength, family, and M&B. Females, on the other hand, have a stronger association with career and science and a weaker association with the arts. Regardless of the fact that D_W is seldom ever connected with females, its tiny magnitude allows it to be seen as a balance.

B. Political Discourse

A positive value indicates a bias toward Israel while a negative value indicates a bias toward Palestinians.

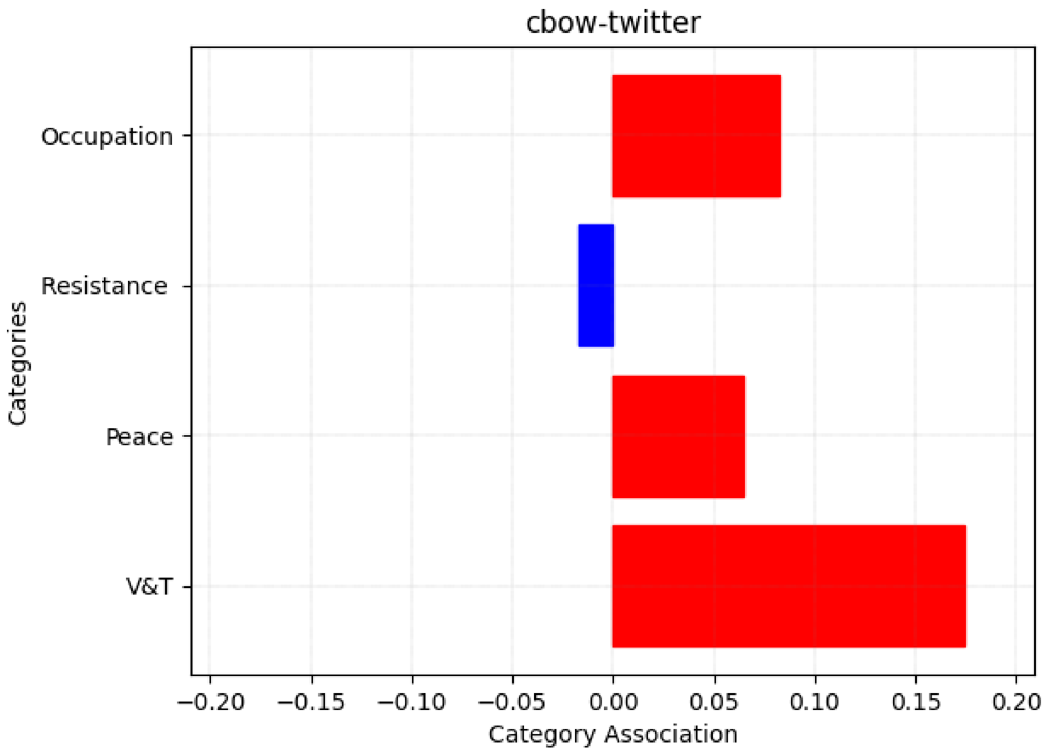

Figure 3: Political bias on the Twitter word embedding

Political bias on Twitter reveals that V&T and Israel have a strong relationship. They are connected to occupation and peace as well. Palestinians, on the other hand, have a slight association with resistance.

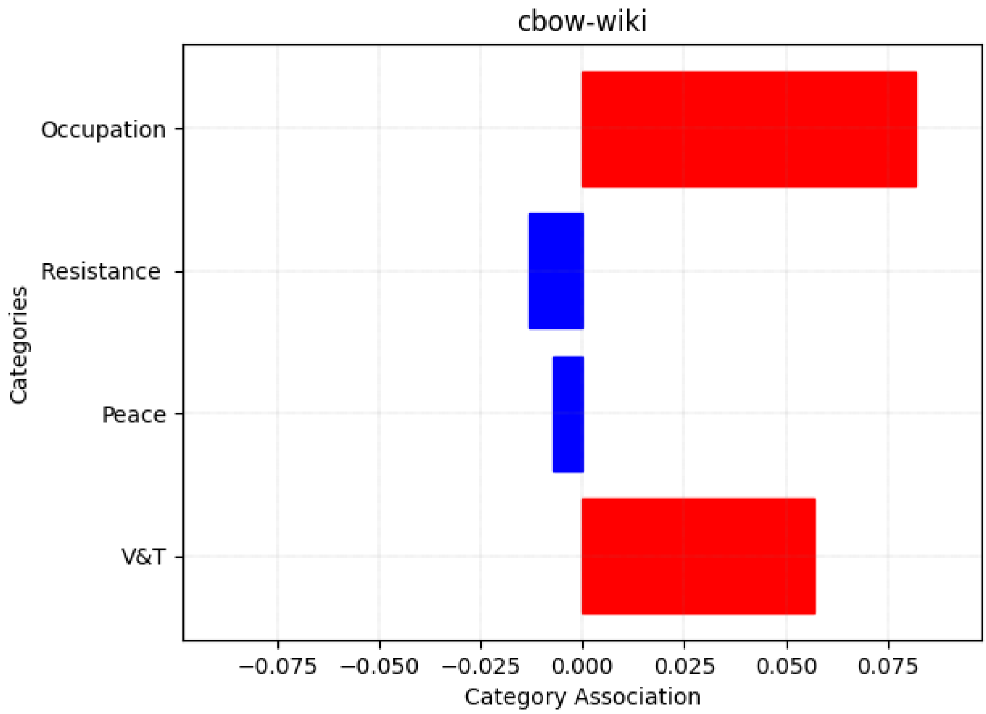

Figure 4: Political bias on the Wikipedia word embedding

Our political bias results on Wikipedia show that Israel has a very strong association with occupation and V&T whereas resistance and peace are associated with Palestinians.

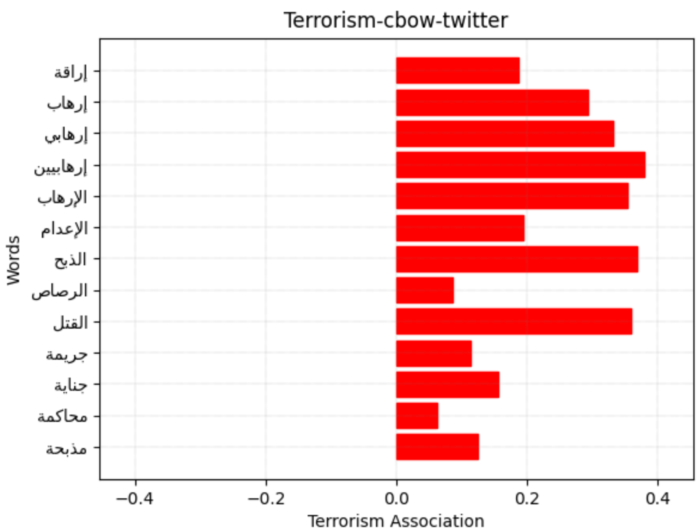

Figure 5: Terrorism bias on the Twitter word embedding

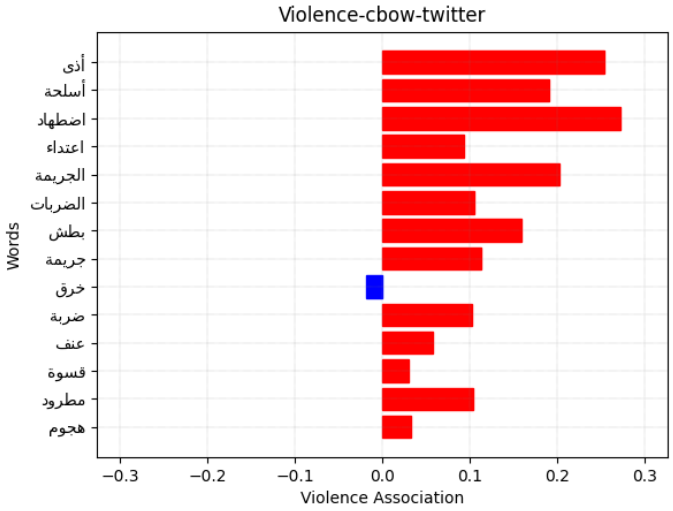

Figure 6: Violence bias on the Twitter word embedding

For the V&T category, an interesting result in the Twitter word embedding is that all words are associated with Israel except for the word 'حرق'

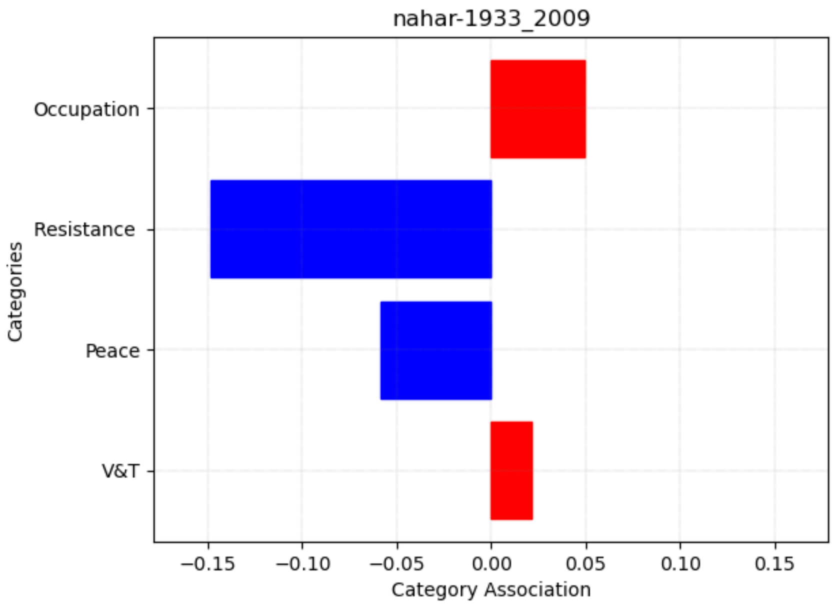

Figure 7: Political bias on the An-Nahar newspaper word embedding

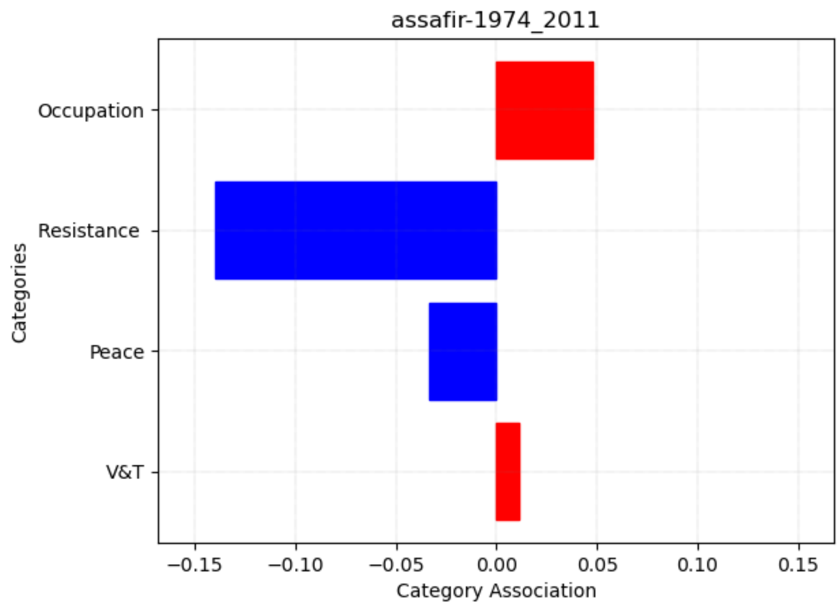

Figure 8: Political bias on the As-Safir newspaper word embedding

Both newspapers exhibit the agreement on the direction and almost the magnitude of bias towards each political group. Israel is associated with occupation and V&T while Palestinians are highly associated with resistance and slightly associated with peace.

C. Political Discourse (Temporal)

A positive value indicates a bias toward Israel while a negative value indicates a bias toward Palestinians. Note that As-Safir did not exist in the decades between the period of 1933 – 1972. Therefore, a value of 0 was assigned to it in these decades.

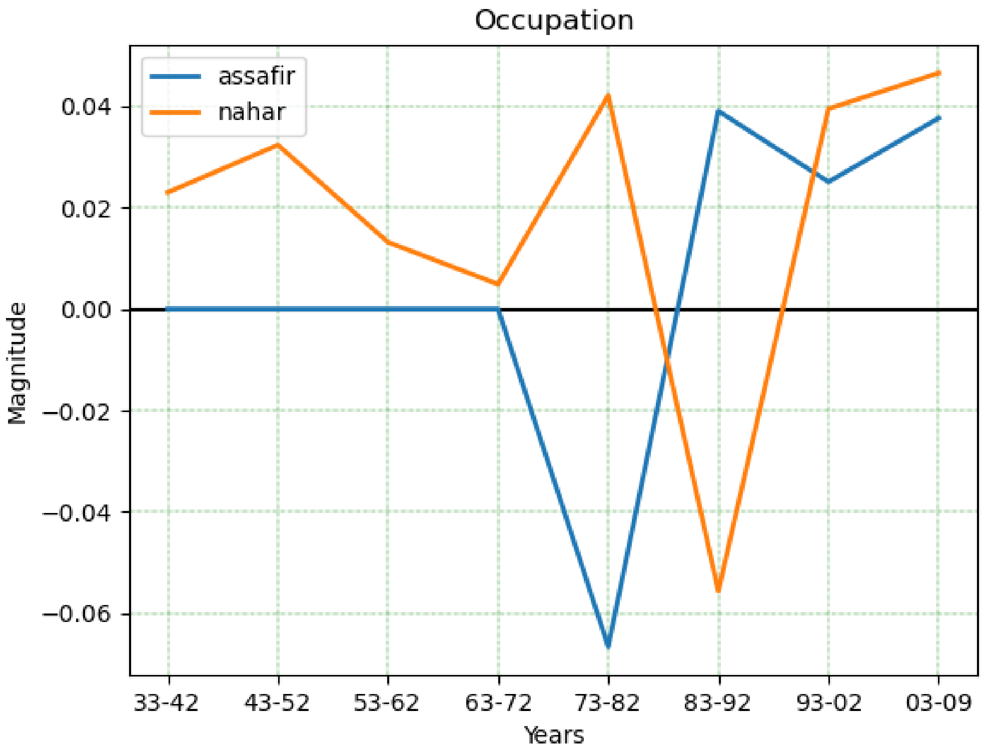

Figure 9: Occupation trends in the newspapers' word embeddings.

Israel has been linked to occupation from 1933 to 1972, according to An-Nahar. When As-Safir was first published, it held a different occupation interpretation from An-Nahar from 1973 through 1992. Beginning in 1993, both publications shared the same perspective.

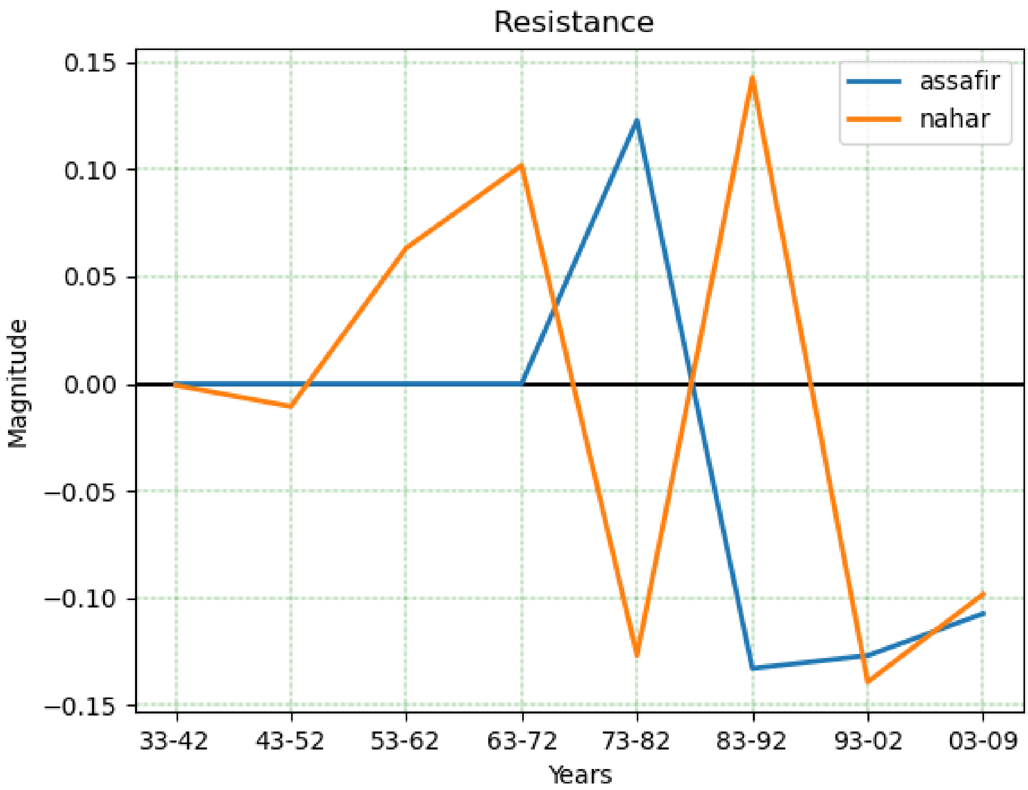

Figure 10: Resistance trends in the newspapers' word embeddings.

There has been a balance towards resistance in the first two decades. Then, Israel has been linked to resistance from 1953 to 1972, according to An-Nahar. In contrast to An-Nahar from 1973 to 1992, As-Safir had a distinct resistance interpretation at the time of its initial publication. Both newspapers began to adopt the same viewpoint around 1993.

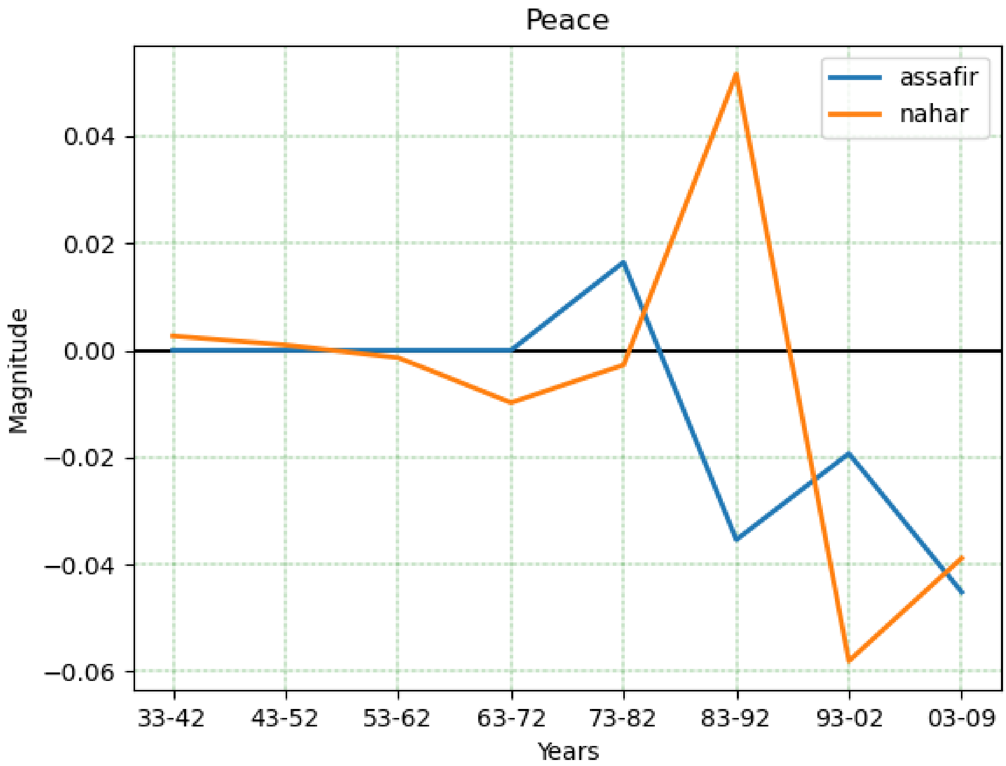

Figure 11: Peace trends in the newspapers' word embeddings.

There has been a balance towards peace in the first three decades. Then, Palestine has been linked to peace from 1963 to 1972, according to An-Nahar. As-Safir had a clear peace interpretation when it was first published, unlike An-Nahar from 1973 until 1992. Around 1993, the two newspapers started to express the same opinion.

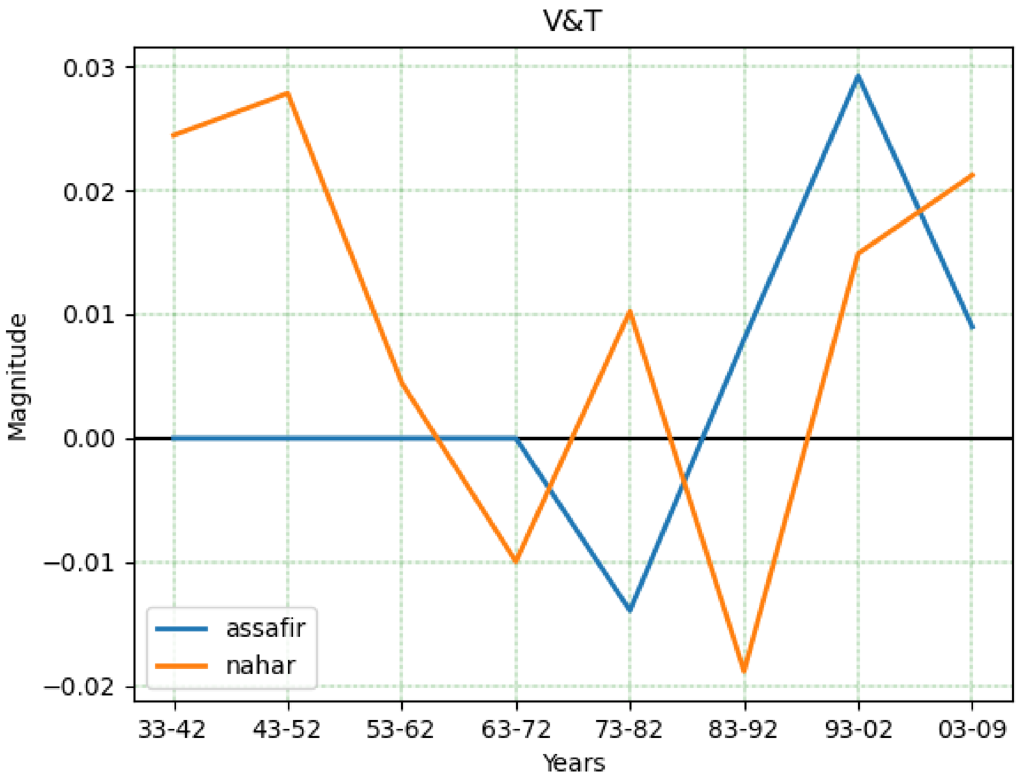

Figure 12: V&T trends in the newspapers' word embeddings.

Israel has been linked to V&T from 1933 to 1962 while Palestinians have been linked to V&T from 1963 to 1972, according to An-Nahar. The V&T interpretation of As-Safir at the time of its initial publication was distinct from that of An-Nahar from 1973 to 1992. Both newspapers began to have the same view in 1993.

CHAPTER VI

DISCUSSION

A. Hypotheses

- We hypothesize that strength, career, science, and money & business have been associated with males.
- We hypothesize that weakness, family, domestic work, art, and beauty & appearance have been associated with females.
- We hypothesize that Palestine has been associated with peace and resistance while Israel has been associated with occupation and violence & terrorism.

B. Gender Discourse

As most Twitter users communicate their opinions through tweets, our gender assumptions, which reflect the thinking of the majority of the Arab World, are confirmed by the data from Twitter. Wikipedia, on the other hand, shows a strong correlation between career and science for females and weakness, family, and B&A for males. Wikipedia includes a substantial number of Arabic pages concerning feminism and gender equality since Wikipedia authors frequently write to dispel preconceived notions. These subjects could have pushed males more in the direction of weakness, family, and B&A while influencing women's career and science associations. Males are indeed connected with strength and M&B, while females are associated with art and domestic work, according to both Wikipedia and Twitter.

C. Political Discourse

According to Wikipedia and Twitter, whereas Palestine has been linked to resistance, Israel has been linked to occupation and V&T. Twitter indicates that Israel is more linked with peace, in contrast to Wikipedia, which indicates the reverse. Political articles on Wikipedia are published to summarize and record historical events. These incidents get Palestine closer to peace. Twitter users, in contrast, typically tweet as soon as an incident takes place. The data from 2008 to 2016 were used to train the Twitter model. Israel held several peace negotiations throughout that time, including one in Washington, D.C., and Sharm el-Sheikh in September 2010. Users tweet about such events instantly.

The high magnitudes of words in the V&T category should be noted. While gender and political entries on Wikipedia appear to be evenly distributed, Arabs prefer to discuss politics on Twitter more frequently than gender issues.

Results from Lebanese newspapers reveal that both newspapers' viewpoints are generally in agreement. The fact that Israel and Lebanon have repeatedly engaged in armed conflict has led to a strong association between resistance and the Palestinians.

D. Political Discourse (Temporal)

Our simulations accurately represented the fact that there was no mutually acceptable peace from 1943 to 1962. Additionally, the First Intifada was overstated in our model since both publications gave resistance a very high magnitude despite the fact that they had the exact opposite perspective. The occupation was closely associated with Israel when the state of Israel was established due to a large number of Palestinians refugees facing exodus. Furthermore, Palestinian civilian casualties have outnumbered

Israeli civilian casualties, which may justify the newspapers associating Israel very close relationship with V&T during the past two decades.

CHAPTER V

CONCLUSION

Our research has some restrictions that we are unable to overcome. First, just two newspapers serve as our data providers. As a result, our evaluation of the media bias in Lebanon is rendered exclusively through the lens of these newspapers, which typically convey the opinions of journalists, rather than broadcasted media like TV and radio, other newspapers, and the opinions of other people. Additionally, OCR software was used to capture our data. This recognition program commonly yields inaccurate results because of spelling errors. Another issue is that the terms used for each category will limit and make our techniques of assessing bias unrepresentative.

On models trained on annual data with more categories, more research may be done. To explore biases across countries and people, the Twitter model may also be split into separate models for each dialect. One can do more research using additional terms with various dialects and more group attributes because the terms in the groups and categories have a significant impact on our findings.

APPENDIX

Gender Bias

Males= ['هو', 'اب', 'هم', 'ولد', 'عليه', 'اخ', 'به', 'رجل', 'سيد', 'ذكر', 'صبي', 'صبيان', 'شباب', 'عنه', 'له']

Females= ['هي', 'ام', 'هن', 'بنت', 'عليها', 'اخت', 'بها', 'امراة', 'سيدة', 'انثى', 'صبية', 'صبايا', 'بنات', 'عنها', 'لها']

StrengthM=["جهد","قوة","قدرة","قدرات","طاقة","فاعلية","كفاءة","قوي","جدارة","فعال","مجتهد","يجتهد"]

StrengthF=["جهد","قوة","قدرة","قدرات","طاقة","فاعلية","كفاءة","قوية","جدارة","فعالة","مجتهدة","تجتهد"]

WeaknessM=["مرهق","متعب","ضعف","متذمر","مجروح"]

WeaknessF=["مرهقة","متعبة","ضعف","متذمرة","مجروحة"]

CareerM=["مهنة", "تجارة", "عمل","مكتب","معاش","راتب","محاسب","مؤسسة", "مصمم", "مدير", "مشرف", "وظيفة","تدريب","يتدرب","طبيب","موظف","ممرض","تقاعد", "أخصائي", "اعمال","محامي","شغل","كاتب","سكرتير","فنان","ممثل""]

CareerF=["مهنة", "تجارة", "عمل","مكتب","معاش","راتب","محاسبة","مؤسسة", "مصممة", "مديرة", "مشرفة", "وظيفة","تدريب","تتدرب","طبيبة","موظفة","ممرضة","تقاعد", "أخصائية", "اعمال","محامية","شغل","كاتبة","سكرتيرة","فنانة","ممثلة""]

FamilyM=["طلاق","طفل","طفولة","ولد","زوج","زواج","عم","خال","زوج","يربي","اب","أقارب","أسرة","عائلة"]

FamilyF=["طلاق","طفلة","طفولة","بنت","زوجة","زواج","عمة","خالة","زوجة","تربي","ام","أقارب","أسرة","عائلة""]

DomesticWorkM=["مهمة","خبز","طبخ","مطبخ","يطبخ","ينظف", "متسخ","مرتب","تنشيف"]

DomesticWorkF=["مهمة", "خبز","طبخ","مطبخ","تطبخ","تنظف", "متسخ","مرتب","تنشيف"]

```
ScienceM=["هندسة", "كيمياء",
"بيانات","رياضيات","علم","طب","عالم","اكتشاف","يكتشف","أحياء","فيزياء",
"ابتكار","علوم","اختراع","يخترع","كمبيوتر","حاسوب","معمل", "بحث",
"باحث","مهندس","دكتوراة","دكتور","طبيب","مخترع"]
ScienceF=["هندسة", "كيمياء",
"بيانات","رياضيات","علم","طب","عالمة","اكتشاف","تكتشف","أحياء","فيزياء",
"ابتكار","علوم","اختراع","تخترع","كمبيوتر","حاسوب","معمل", "بحث",
"باحثة","مهندسة","دكتوراة","دكتورة","طبيبة","مخترعة"]
ArtM=["إبداع", "مبدع",
"فني","فنان","معرض","تصوير","مصور","رسم","رسام","تلوين","مسرح","نحت", "نحات",
"مزخرف","موهبة","موهوب","موسيقى","أغنية","رقص","تصميم","مصمم",
"تزيين","فن","غناء","أزياء","طرب","مطرب"]
ArtF=["إبداع", "مبدعة",
"فني","فنانة","معرض","تصوير","مصورة","رسم","رسامة","تلوين","مسرح","نحت", "نحاتة",
"مزخرف","موهبة","موهوبة","موسيقى","أغنية","رقص","تصميم","مصممة",
"تزيين","فن","غناء","أزياء","طرب","مطربة"]
MoneyAndBusinessM=["بنك", "مصرف",
"ديون","مليونير","بضائع","مال","مصاري","دولار","يورو","ليرة","تجارة","محاسب", "شيكات",
"راتب","قرض","ميزانية","مشروع","ثروة","شيك","تمويل","صفقة","إستثمار","ربح",
"أعمال","مشروع","إقتصاد","شراكة","تسويق","مدير","صناعة"]
MoneyAndBusinessF=["بنك", "مصرف",
"ديون","مليونيرة","بضائع","مال","مصاري","دولار","يورو","ليرة","تجارة","محاسبة", "شيكات",
"راتب","قرض","ميزانية","مشروع","ثروة","شيك","تمويل","صفقة","إستثمار","ربح",
"أعمال","مشروع","إقتصاد","شراكة","تسويق","مديرة","صناعة"]
```

BeautyAndAppearanceM=["فاتن",

"أناقة","جمال","جذاب","معجب","مذهل","طويل","قصير","وزن","أشقر", "أنيق",

"أسمر","بشرة","تسريحة","مثير","لياقة","لباس", "شهوة", "قبلة","رومانسي","يتغزل","ملفت","يرتدي"]

BeautyAndAppearanceF=["فاتنة",

"أناقة","جمال","جذابة","معجبة","مذهلة","طويلة","قصيرة","وزن","شقراء","أنيقة","سمراء","بشرة","تسر

يحة","مثيرة","لياقة","لباس", "شهوة","قبلة","رومانسي","تتغزل","ملفت","يرتدي"]

<u>Pollical Bias</u>

Israel= ['صهيوني', 'إسرائيلي', 'الإسرائيليين', 'يهود', 'إسرائيل', 'أبيب', 'اليهود', 'شارون']

Palestine= ['المسلمين', 'الفلسطينيون', 'فلسطينيون', 'حماس', 'فلسطين', 'فلسطيني', 'التحرير', 'الفلسطينية']

Occupation= [' تفتيش', 'إخلاء', 'مستوطنة', 'المستوطنين', 'مستوطن', 'احتل', 'هدم', 'احتلال', 'انفصال',

'الهدم', 'المستوطنات', 'عزل', 'سجناء', 'انسحاب', 'صاروخ', 'الهجمات', 'الصواريخ', 'قصف', 'انفجار', 'غارة']

Resistance =[' معاناة', 'كفاح', 'نضال', 'صراع', 'تحمل', 'الصمود', 'المقاومة', 'صمود', 'التأقلم', 'المواجهة',

'تأقلم', 'مواجهة', 'مقاومة']

Peace=["مسالم","سلام","السلمي","سلمي","امن","أمان","هادئ","هدوء","سكون","سلمية"]

V&T= [' أذى', 'أسلحة', 'اضطهاد', 'اعتداء', 'الجريمة', 'الضربات', 'بطش', 'جريمة', 'خرق', 'ضربة', 'عنف',

'قسوة', 'مطرود', 'هجوم', 'إراقة', 'إرهاب', 'إرهابي', 'إرهابيين', 'الإرهاب', 'الإعدام', 'الذبح', 'الرصاص', 'القتل',

'جريمة', 'جناية', 'محاكمة', 'مذبحة']

<u>GitHub Repository</u>: https://github.com/gha30/genderAndPoliticalBias

REFERENCES / BIBLIOGRAPHY

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings* (arXiv:1607.06520). arXiv. http://arxiv.org/abs/1607.06520

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183. https://doi.org/10.1126/science.aal4230

Chen, Y., Mahoney, C., Grasso, I., Wali, E., Matthews, A., Middleton, T., Njie, M., & Matthews, J. (2021). Gender Bias and Under-Representation in Natural Language Processing Across Human Languages. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 24–34. https://doi.org/10.1145/3461702.3462530

Fast, E., Chen, B., & Bernstein, M. S. (2016). Empath: Understanding Topic Signals in Large-Scale Text. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 4647–4657. https://doi.org/10.1145/2858036.2858535

Garg, N., Schiebinger, L., Jurafsky, D., & Zou, J. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, *115*(16), E3635–E3644. https://doi.org/10.1073/pnas.1720347115

Lauscher, A., Takieddin, R., Ponzetto, S. P., & Glavaš, G. (2020). AraWEAT: Multidimensional Analysis of Biases in Arabic Word Embeddings. *ArXiv:2011.01575 [Cs]*. http://arxiv.org/abs/2011.01575

Mendelsohn, J., Tsvetkov, Y., & Jurafsky, D. (2020). A Framework for the Computational Linguistic Analysis of Dehumanization. *Frontiers in Artificial Intelligence*, *3*, 55. https://doi.org/10.3389/frai.2020.00055

Soliman, A. B., Eissa, K., & El-Beltagy, S. R. (2017). AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP. *Procedia Computer Science, 117*, 256–265. https://doi.org/10.1016/j.procs.2017.10.117

Wevers, M. (2019). Using Word Embeddings to Examine Gender Bias in Dutch Newspapers, 1950-1990. *ArXiv:1907.08922 [Cs, Stat]*. http://arxiv.org/abs/1907.08922

Zhou, P., Shi, W., Zhao, J., Huang, K.-H., Chen, M., Cotterell, R., & Chang, K.-W. (2019). Examining Gender Bias in Languages with Grammatical Gender. *ArXiv:1909.02224 [Cs]*. http://arxiv.org/abs/1909.02224

Doughman, J., Salem, F. A., & Elbassuoni, S. (2020). *Time-Aware Word Embeddings for Three Lebanese News Archives*. 9.