Appendix A

Python code for bias analysis

```
import torch
import torchtext
import numpy as np
import spacy
from pyfasttext import FastText
# Load spaCy model and FastText model
nlp = spacy.load("en_core_web_lg")
model = FastText()
model.load_model("/content/sample_data/cc.de.300.bin")
glove = torchtext.vocab.GloVe(name="6B", dim=50)
# GloVe bias analysis
def glove_bias_values(definite_sets, analogy_templates,
test_terms):
results = []
input_words = [item for sublist in definite_sets for item in
sublist]
for tword in test_terms:
for iword in input_words:
x = glove.vectors[glove.stoi[iword]]
y = glove.vectors[glove.stoi[tword]]
result = {}
result["input"] = iword
result["target"] = tword
result["distance"] = torch.norm(y - x).item()
result["cosine"] = cos_sim(x, y)
results.append(result)
return results
results_glove = glove_bias_values(definite_sets, analogy_templates,
test_terms)
output_glove = """# Glove output for Gender bias:
| input | manager |executive |doctor | lawyer | programmer | scientist| soldier |
supervisor | rancher| janitor| firefighter | officer|
"""

# Create a dictionary to store cosine similarity values for each target
word
target_columns = {target: [] for target in test_terms}
for result in results_glove:
input_word = result['input']
for target in test_terms:
  if target == result['target']:
    target_columns[target].append(round(result['cosine'],
5))

# Populate the output_glove with the values from the dictionary

for input_word in definite_sets[0]:
  row = f"{input_word}"
  for target in test_terms:
      row += f"{target_columns[target].pop(0)}|"
```

```
      output_glove += row + "\n"

# Display the table
display (Markdown(output_glove))
```

Table A1: Words like "Islam" and "jihad" are regularly associated with the phrase's "terrorist" and "terrorism" by the word embeddings (Glove, SpaCy, FastText).

terrorism Bias	input	islam	hinduism	hindutva	buddhism	christianity	hindu	wahabi	jihad	men	women	man	woman
Glove Output	terrorist	0.528	0.098	0.098	0.145	0.296	0.293	0.058	0.674	0.451	0.343	0.457	0.314
	terrorism	0.623	0.165	0.152	0.275	0.400	0.255	0.058	0.574	0.439	0.429	0.415	0.304
Fasttext Output	terrorist	0.420	0.183	0.168	0.272	0.345	0.328	0.285	0.443	0.259	0.314	0.163	0.372
	terrorism	0.413	0.262	0.247	0.376	0.492	0.279	0.261	0.551	0.279	0.409	0.056	0.331
SpaCy Output	terrorist	0.304	0.068	0.000	0.160	0.309	0.123	0.000	0.583	0.232	0.234	0.300	0.263
	terrorism	0.308	0.179	0.000	0.244	0.387	0.154	0.000	0.591	0.208	0.297	0.212	0.209

Table A2: Gender bias throughout the various word embeddings (Glove, SpaCy, FastText). Words like"she" and "woman" have highest similarities with the terms like "nurse," "maid," and "hairdresser," and lowest similarities with the terms like "firefighter" and "officer."

Gender bias	input	secretary	nurse	clerk	artist	homemaker	dancer	singer	librarian	maid	hair dresser	stylist	Receptionist	Counsellor
Glove Output	woman	0.255	0.715	0.470	0.522	0.573	0.586	0.569	0.391	0.645	0.508	0.403	0.446	0.439
	man	0.346	0.572	0.400	0.507	0.397	0.529	0.513	0.300	0.537	0.391	0.311	0.310	0.350
SpaCy Output	she	0.399	0.646	0.476	0.546	0.394	0.546	0.538	0.426	0.491	0.357	0.366	0.353	0.442
	he	0.517	0.481	0.476	0.448	0.224	0.390	0.399	0.386	0.304	0.191	0.159	0.184	0.389
	he	0.242	0.307	0.230	0.168	0.221	0.212	0.166	0.160	0.235	0.221	0.076	0.272	0.228
	she	0.194	0.475	0.147	0.224	0.354	0.373	0.276	0.183	0.390	0.377	0.209	0.318	0.247
	woman	0.178	0.475	0.228	0.302	0.413	0.414	0.307	0.211	0.531	0.380	0.225	0.326	0.262
Fasttext output	man	0.103	0.282	0.218	0.218	0.272	0.286	0.199	0.098	0.408	0.246	0.101	0.216	0.172
	he	0.082	0.209	0.170	0.206	0.005	0.206	0.182	0.172	0.318	0.209	0.084	0.197	0.122
	she	0.250	0.337	0.205	0.269	0.176	0.314	0.285	0.220	0.405	0.330	0.301	0.302	0.303
	woman	0.431	0.469	0.389	0.445	0.255	0.503	0.464	0.363	0.465	0.434	0.429	0.354	0.448
	man	0.028	0.014	0.064	0.029	0.204	0.023	0.134	0.060	0.008	0.099	0.095	0.154	0.031

Table A3: Gender bias throughout the various word embeddings (Glove, SpaCy, FastText). It states that certain professions are connected to specific genders. We observed that the words like “man”, and “he” refers to the highest similarities with the terms like “manager," "soldier," and "supervisor," and lowest similarities with the terms like "hairdresser" and "stylist."

Gender Bias	input	manager	executive	doctor	lawyer	programmer	scientist	soldier	supervisor	rancher	janitor	firefighter	officer
Glove Output	he	0.603	0.490	0.691	0.599	0.253	0.448	0.527	0.386	0.138	0.258	0.239	0.617
	she	0.441	0.394	0.728	0.552	0.229	0.416	0.531	0.434	0.127	0.276	0.263	0.521
	man	0.444	0.348	0.712	0.607	0.266	0.492	0.732	0.267	0.378	0.355	0.435	0.583
	woman	0.223	0.223	0.725	0.580	0.219	0.439	0.718	0.341	0.367	0.380	0.458	0.484
SpaCy Output	he	0.197	0.054	0.415	0.336	0.098	0.251	0.413	0.161	0.290	0.154	0.305	0.301
	she	0.101	0.023	0.453	0.282	0.077	0.214	0.284	0.156	0.189	0.136	0.286	0.215
	woman	0.083	0.045	0.456	0.362	0.029	0.274	0.470	0.129	0.298	0.212	0.347	0.325
	man	0.098	-0.028	0.374	0.337	0.029	0.251	0.584	0.072	0.362	0.237	0.388	0.339
Fasttext Output	he	0.127	0.102	0.246	0.265	0.185	0.225	0.230	0.075	0.064	0.098	0.069	0.169
	she	0.187	0.184	0.267	0.241	0.258	0.251	0.328	0.204	0.216	0.190	0.107	0.239
	woman	0.325	0.320	0.369	0.392	0.292	0.485	0.444	0.302	0.236	0.255	0.243	0.446
	man	0.171	0.024	0.148	0.043	0.214	0.077	0.113	0.152	0.277	0.121	0.119	0.127

Table A4: Calculation of associations between Religion and Target words using word embeddings (Glove, SpaCy, Fasttext).

Religion Bias	input	liberal	violent	un educated	dirty	judgemental	terrorism	terrorist	conservative	violent
Glove Output	judaism	0.445	0.224	0.145	-0.100	-0.027	0.246	0.180	0.451	0.224
	christianity	0.372	0.340	0.179	-0.086	-0.099	0.400	0.296	0.433	0.340
	christian	0.579	0.329	0.125	0.013	-0.302	0.329	0.297	0.627	0.329
	hindutva	0.295	0.216	0.219	0.174	0.136	0.152	0.098	0.294	0.216
	islam	0.394	0.470	0.189	0.165	-0.127	0.623	0.528	0.482	0.470
	jew	0.458	0.303	0.406	0.208	0.017	0.242	0.279	0.440	0.303
	muslim	0.453	0.592	0.275	0.167	-0.198	0.563	0.553	0.540	0.592
	hinduism	0.231	0.248	0.235	-0.124	-0.025	0.165	0.098	0.249	0.248
SpaCy Output	judaism	0.399	0.155	0.197	0.025	0.287	0.217	0.149	0.342	0.155
	christianity	0.562	0.345	0.404	0.026	0.442	0.387	0.309	0.513	0.345
	christian	0.487	0.222	0.406	-0.007	0.334	0.251	0.254	0.461	0.222
	hindutva	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	islam	0.287	0.202	0.121	0.093	0.087	0.308	0.304	0.187	0.202
	jew	0.091	0.036	0.162	0.244	-0.087	0.037	0.137	-0.032	0.036
	muslim	0.357	0.297	0.305	0.180	0.107	0.339	0.369	0.294	0.297
	hinduism	0.230	0.140	0.141	0.121	0.234	0.179	0.068	0.171	0.140
Fasttext Output	judaism	0.231	0.244	0.247	0.144	0.282	0.403	0.240	0.326	0.244
	christianity	0.309	0.336	0.297	0.196	0.263	0.492	0.345	0.434	0.336
	christian	0.258	0.194	0.186	0.265	0.162	0.260	0.242	0.217	0.194
	hindutva	0.189	0.161	0.154	0.061	0.285	0.247	0.168	0.173	0.161
	islam	0.363	0.239	0.172	0.108	0.095	0.413	0.420	0.235	0.239
	jew	0.074	0.032	0.131	0.081	0.041	0.036	0.095	0.023	0.032
	muslim	0.385	0.376	0.274	0.250	0.169	0.464	0.577	0.302	0.376
	hinduism	0.176	0.100	0.154	0.046	0.277	0.262	0.183	0.147	0.100

Table A5: Calculation of correlations between Race and Target words using word embeddings (Glove, SpaCy, Fasttext).

Race Bias	input	executive	redneck	leader	farmer	engineer	laborer	teacher	slave	musician	runner	criminal	homeless
	black	0.289	0.183	0.402	0.435	0.207	0.053	0.376	0.403	0.358	0.349	0.373	0.325
	caucasian	-0.190	0.320	0.169	0.206	-0.007	0.098	0.154	0.243	0.159	0.082	0.034	0.106
	asian	0.312	-0.012	0.373	0.276	0.143	0.012	0.249	0.239	0.155	0.369	0.277	0.203
Glove Output	african	0.385	-0.030	0.540	0.381	0.244	0.079	0.335	0.346	0.294	0.310	0.306	0.317
	white	0.371	0.084	0.434	0.438	0.166	0.043	0.308	0.311	0.219	0.303	0.354	0.248
	american	0.472	0.099	0.522	0.487	0.539	0.181	0.529	0.522	0.530	0.417	0.468	0.282
	european	0.459	-0.203	0.459	0.118	0.268	-0.076	0.209	0.230	0.181	0.394	0.318	0.114
	black	0.057	0.339	0.138	0.202	0.022	0.244	0.068	0.282	0.090	0.093	0.256	0.310
	caucasian	0.076	0.244	0.204	0.195	0.005	0.238	0.143	0.311	0.100	0.148	0.164	0.211
	asian	0.036	0.231	0.030	0.115	0.015	0.141	0.106	0.194	0.051	-0.009	0.062	0.146
SpaCy Output	african	0.195	0.120	0.211	0.270	0.185	0.258	0.167	0.271	0.219	0.015	0.150	0.199
	white	0.045	0.349	0.081	0.172	0.000	0.202	0.015	0.225	0.028	0.061	0.209	0.270
	american	0.247	0.157	0.239	0.227	0.199	0.256	0.192	0.248	0.201	0.053	0.231	0.238
	european	0.242	0.089	0.166	0.187	0.128	0.246	0.093	0.204	0.055	0.009	0.183	0.163
	black	0.316	0.417	0.328	0.353	0.276	0.223	0.302	0.345	0.359	0.392	0.394	0.349
	caucasian	0.256	0.385	0.277	0.300	0.280	0.208	0.348	0.326	0.323	0.219	0.352	0.350
	asian	0.311	0.415	0.293	0.390	0.268	0.148	0.367	0.318	0.340	0.247	0.352	0.378
Fasttext Output	african	0.278	0.349	0.298	0.380	0.149	0.261	0.314	0.345	0.351	0.264	0.279	0.490
	white	0.286	0.404	0.286	0.309	0.295	0.214	0.279	0.321	0.345	0.394	0.382	0.365
	american	0.320	0.470	0.255	0.361	0.266	0.131	0.338	0.274	0.324	0.221	0.325	0.357
	european	0.308	0.276	0.273	0.262	0.189	0.135	0.306	0.231	0.251	0.258	0.323	0.327